Stochastic volatility and the pricing of financial derivatives

van der Ploeg, A.P.C.

Publication date
2006

Document Version
Final published version

Stochastic Volatility
and the
Pricing of Financial Derivatives

ISBN 90 5170 577 8

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 366 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Stochastic Volatility and the Pricing of Financial Derivatives

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. mr. P.F. van der Heijden

ten overstaan van een door het college voor promoties ingestelde commissie,

in het openbaar te verdedigen in de Aula der Universiteit

op vrijdag 10 februari 2006, te 14.00 uur

door

Antoine Petrus Cornelius van der Ploeg

geboren te Noordwijk aan Zee

Promotiecommissie

Promotors:

Prof. dr. H.P. Boswijk

Prof. dr. F.C.J.M. de Jong

Overige leden:

Prof. dr. P.C. Schotman

Prof. dr. S.J. Koopman

Prof. dr. ir. J.G. de Gooijer

Dr. J.J.A.G. Driessen

Dr. K.J. van Garderen

Faculteit der Economische Wetenschappen en Econometrie

Acknowledgements

This dissertation forms the completion of five years of doctoral research at the Department of Quantitative Economics of the University of Amsterdam, and the Tinbergen Institute. I would like to thank a number of people who have either directly or indirectly contributed to this thesis.

First of all, I thank my promoters Peter Boswijk and Frank de Jong for having given me the opportunity to become a doctor and to follow my own research path, but in particular for their inspiring and very pleasant guidance during my PhD. I also thank the other members of the promotion committee, Siem Jan Koopman, Joost Driessen, Peter Schotman, Kees Jan van Garderen and Jan de Gooijer, for their interest and time they have put in reading and evaluating this thesis. Special thanks goes to Joost Driessen for the always stimulating discussions we had over the past few years. I also thank the people I talked with on various conferences. Besides having been a great stimulant, they have helped further shape my ideas and research.

I am grateful to the people I worked together with in teaching several courses, including Roald Ramer, Frank Kleibergen, Risto Heijmans, Jan Bogers and Thije Euverman. This team work formed a welcome supplement to doing so much research on my own. Thanks again to Peter Boswijk for also having given me the responsibility of teaching the intensive international course Stochastic Calculus at the Amsterdam Graduate Business School on my own. I also thank my fellow PhD-ers and my other colleagues of the Department of Quantitative Economics and the Finance Group (thanks for the nice parties, and –of course- for all the coffee!) for having provided a pleasant environment to work in. I thank my neighboring-colleague Herman ten Napel for having often reminded me that it was really time to go home when I was still working enthusiastically late in the evening.

Many thanks goes to my good friend Yiu Chung Cheung. Having known him since our Econometrics undergraduate, and having jointly started the M.Phil. educational program at the Tinbergen Institute and our PhDs, we have developed a personal (non fixed-income!) bond for life. Not only has it been a great pleasure to share our PhD ups and downs with each other, but also on the private side. Special thanks goes to Hanneke Hoppenbrouwer as well. Having known her for many years now, I am grateful for her friendship, the good times we always share, and the positive support she always provides. I thank my other friends as well, in particular Harry Laats, Ramon de Zwart, Dick van de Berg, Sander van Riel, Ana van Walem, Mark Schilstra and the people from Het Pannekoekenhuis in Noordwijk and Scheveningen.

I owe many thanks to my parents and sister. They are always available whenever I need them, and they have always stimulated me on my personal path through life. Their love has been -and still is- an invaluable support. Last but certainly not least, I thank Nico for him being such a wonderful, enthusiastic, loving and caring person.

Antoine P.C. van der Ploeg
Amsterdam, September 2005
(Contact address: apcvanderploeg@hotmail.com)

Contents

- I -
Stochastic Volatility and the Pricing of Financial Derivatives: An Introduction

-II-
The Multifactor Affine Stochastic Volatility Derivative Pricing Model

-III-
A State Space Approach to the Estimation of the Multifactor Affine Stochastic Volatility Derivative Pricing Model

- IV -
Monte Carlo and Empirical Results for Ornstein-Uhlenbeck SV: Examining UK Financial Markets

- V-
Exploiting Level-Dependent Volatility-of-Volatility: An Extended Kalman Filter Approach to the Estimation of the Multifactor Affine SV Derivative Pricing Model

- VI -
Leverage Effect and the Heston (1993) Model

- I -

Stochastic Volatility and the Pricing of Financial Derivatives: An Introduction

1. Motivation

Volatility, i.e., the uncertain fluctuations in the prices of financial assets such as stocks, is of paramount importance in financial markets theory and practice. Volatility is central in asset allocation decisions, risk management and asset pricing. Volatility characterizes the risk of investments. Moreover, the volatility of an asset is a major determinant of the price of a financial derivative.

A *financial derivative* is a contract that derives its value from another, underlying asset. The holder of a European call option for example, has the right (but not the obligation) to buy stock for a certain contract-specified price, the strike price, at some date in the future, the expiration or maturity date. A put option gives the right to sell. The value of this right is the option price or premium. Apart from these vanilla calls and puts, numerous other option-style products are traded in financial markets. These assets are collectively dubbed *exotic options*. Nowadays, a multimillion industry in these derivatives exists, and financial engineers continually invent new products to extend a bank's product range. These products are often tailored to specific speculative and hedging needs of a bank's clients, and are typically sold over the counter, rather than at organized markets. [1]

Black and Scholes (1973)
A bank or other financial institution wishing to issue a new derivative wants to know the fair price that can be asked for this product. A *derivative pricing model* serves this purpose. In practice, the most widely applied model is still the celebrated Black-Scholes (1973) model, in which the underlying asset price follows a geometric Brownian motion. The main advantage of the Black-Scholes (BS) model is that it leads to closed-form formulas for many of the standard option products (e.g., the famous Black-Scholes formula). Moreover, the model depends on just one parameter, the underlying asset's volatility, which the BS model assumes *constant*. As such, once this volatility has been estimated from e.g. past stock returns or the prices of market-traded liquid options, any exotic derivative can subsequently be priced by Monte Carlo simulation.

As convenient as this is, the Black-Scholes model is severely *misspecified*: Volatility changes over time, volatility is *stochastic*. Periods of high and low asset

[1] An excellent introduction to financial derivatives and their uses in practice is Hull (2003).

volatility alternate in the practice of financial markets, a phenomenon known as volatility clustering. [2] When a 3D-plot is made of Black-Scholes implied volatilities of empirical call and put options against moneyness and remaining time to maturity, the resulting *implied volatility surface* is far from flat, whereas a flat surface is what the Black-Scholes theory predicts. [3] Instead, for given maturity one observes a smile or skew-shaped pattern, known as the *volatility smile* or *skew*. For given moneyness one observes a variety of shapes of the so-called *volatility term structure*. A stochastic volatility (SV) model generates a smile-shaped Black-Scholes implied volatility curve. An SV model with so-called *leverage effect* generates a volatility skew. [4]

Derivative pricing under SV

As stochastic volatility provides a more realistic data description, a derivative pricing model in which the underlying asset is subject to SV has a clear potential of generating more accurate exotic option prices. [5] In here should lie the main reason why a financial institution should have an interest in SV derivative pricing models. [6]

Stochastic volatility makes derivative pricing more complex however. Due to SV, the market is no longer *complete*, in the sense that derivative payoffs cannot perfectly be hedged by trading in the underlying stock and risk-free borrowing and lending only. Other derivatives are needed to "complete" the market, i.e., to be able to hedge price changes due to volatility changes. Stated differently, the assumption of no arbitrage only (and hence *risk-neutral valuation*) no longer results in unique option prices. In contrast to the BS model, the option prices resulting from an SV derivative pricing model are not preference-free: The risk-neutral probability measure used for derivative pricing depends on the attitude of investors towards volatility risk, as reflected in the *market price of volatility risk*. This attitude is implicitly present in the market prices of options, and needs to be distilled out from data to be able to recover the risk-neutral measure the market uses for pricing. Once this has been done, exotic derivatives can next be priced by risk-neutral valuation as usual.

Paying a price for volatility risk

The issue whether investors are compensated for the risk associated with uncertain volatility fluctuations has recently attracted a lot of academic attention. Conventional asset pricing theory states that if volatility risk is systematic, investors should be compensated for it. Recent empirical evidence indicates that

---

[2] See sections 2.1, 3.2.1, 3.2.2 and 3.3.2 of chapter III for a discussion of GARCH, stochastic volatility and realized volatility models, their estimation methods and research in this area.

[3] The *BS implied volatility* of an option is that stock volatility for which the observed market price of the option equals the theoretical Black-Scholes price. An option's *moneyness* is defined here as the ratio of the strike price and the current stock price. A currently *at-the-money* option has a moneyness of 1.

[4] The *leverage effect* (Black (1976)) translates into a negative correlation between stock returns and volatility changes. Renault and Touzi (1996) show that SV implies a volatility smile. See chapter VI.

[5] This holds for vanilla options as well, though in practice their prices are typically determined by interpolation on the implied volatility surface. A more sophisticated model than the BS model is thus in principle not needed for pricing vanilla options.

[6] Examples of SV derivative pricing models include the models by Hull and White (1987), Scott (1987), Wiggins (1987), Melino and Turnbull (1990), Stein and Stein (1991), Ball and Roma (1994), the influential Heston (1993) model (see chapter VI), and the very general class of affine-jump diffusion models of Duffie, Pan and Singleton (2000).

investors are willing to pay a premium for bearing market volatility risk. In other words, the *volatility risk premium* is found to be *negative.* [7]

At first sight, a negative volatility risk premium seems counterintuitive. Why would investors be willing to pay for volatility risk? Should they not receive a reward instead? One of our main aims is to provide a possible economic rationale for negative volatility risk compensation.

Estimating SV derivative pricing models
Although an SV derivative pricing model is more realistic than the Black-Scholes model (and hence is likely to generate more reliable prices), the econometric estimation of these models is complicated and far from trivial. First, volatility cannot be observed directly, but is somewhere hidden in the background instead. Volatility is *latent*. As such, it needs to be "integrated out" during estimation. This is not easily possible in general and requires intensive computations. [8] Second, as option prices can only be obtained by simulation, computationally demanding procedures result. This makes a fast implementation impossible (at least, so it seems). A commonly applied estimation method is *calibration.* [9] Calibration performs parameter estimation by minimizing the sum of squared deviations between the market-observed and theoretical option prices on a particular day, essentially by a simulation-based non-linear least squares procedure. Although seemingly attractive, this method disregards the time-series dimension of the model completely and is therefore inherently inconsistent with the dynamic principles of the model. Moreover, the estimates depend on today's information only. Other estimation methods based on information in time series of stock returns only have as drawback that the market price of volatility risk cannot be estimated, whereas this is a necessary input for pricing derivatives.

A clear need thus exists for estimation methods for SV derivative pricing models that *combine* stock and option time-series data in a dynamically consistent manner. Given the complexity of this problem, only recently a number of estimation methods has appeared in the literature, see Chernov and Ghysels (2000), Pan (2002) and Jones (2003). Chernov and Ghysels (2000) and Pan (2002) consider method of moments-based approaches, Jones (2003) a Bayesian approach. From a practical point of view, the main disadvantage of these methods is that they require long, demanding simulations. This makes a fast implementation impossible. In addition, these methods do not directly yield a volatility forecast. [10]

Another central aim of this thesis is to develop a different, transparent but above all *fast* estimation method for affine SV derivative pricing models that circumvents long simulations during estimation. As such, especially for practical (banking) purposes our method forms an attractive alternative to the other methods. Our method is based on a *state space approach* of the problem and the *Kalman filter*. As volatility is latent, the state space framework is naturally suited for estimation. As opposed to moment-based approaches, state space models

[7] See section 6.3 of chapter II for a discussion of the recent evidence in Chernov and Ghysels (2000), Coval and Shumway (2001), Buraschi and Jackwerth (2001), Pan (2002), Driessen and Maenhout (2003), Jones (2003), Bakshi and Kapadia (2003) and Carr and Wu (2004).
[8] See section 2.1 of chapter III for current estimation methods for SV models.
[9] E.g., Bates (1996b, 2000), Bakshi, Cao and Chen (1997), Duffie et al. (2000). Calibration is also popular in banking practice. For more discussion, see section 2.1 of chapter III.
[10] See section 2.1.1 of chapter III for an explanation and discussion of these approaches.

truly treat volatility as unobserved and readily deliver a volatility estimate (and forecast). Using a number of approximations, we derive a discrete-time linear state space model from our continuous-time SV derivative pricing model, which can be estimated by quasi-maximum likelihood. [11]

A related aim is to investigate what data is best suited for the estimation of SV derivative pricing models. What type of data yields the most reliable parameter and volatility estimates? Is this daily squared return data, high-frequency intraday data (i.e. *realized volatility* data), option data, a combination of daily squared return and option data, or a combination of realized volatility and option data? To the best of our knowledge, combining realized volatility and option data in an estimation strategy of SV derivative pricing models is also new to the literature. [12]

A multifactor SV model

As accumulating empirical evidence indicates that stock volatilities are driven by more than one latent factor in practice, we consider a *multifactor SV model*. [13] Specifically, we assume the conditional variance of the stock returns to be driven by an affine function of an arbitrary number of latent volatility factors, which follow mean-reverting affine Markov diffusions.

The extension towards multiple volatility factors is well motivated. As Meddahi (2002), Chernov, Gallant, Ghysels and Tauchen (2003) and Tauchen (2004) argue, one of the reasons for the poor fit of 1-factor SV diffusion models is that these models do not allow for a joint adequate fit of a fat-tailed return distribution and large volatility persistence, which are both stylized facts of asset return data. [14] Introducing a second volatility factor (or adding jumps) breaks this link. Given that volatility is typically found to be so persistent, Eraker, Johannes and Polson (2003) argue that 1-factor SV models do not allow for sufficiently fast-changing volatility, which is unrealistic in times of sudden market stress. They advocate adding jumps to volatility. A possible alternative is a second volatility factor.

Evidence from option markets that points towards multiple volatility factors comes from Cont and Fonseca (2002). These authors show that the daily fluctuations in the implied volatility surface of S&P500 and FTSE100-index options can satisfactorily be described by three underlying, abstract principal component processes. Important additional evidence against 1-factor SV models comes from studies that combine stock and option data for the estimation of SV option pricing models (Chernov and Ghysels (2000), Pan (2002), Jones (2003)). Besides time-series data on the underlying stock, typically one additional short-maturity at-the-money option series is used for estimation. The estimated 1-factor models tend to overprice longer-dated options out of sample. Pan (2002) for example, mentions that multiple SV factors may be necessary to improve on the fit of the volatility term structure, a point also emphasized by Bates (1996b, 2000).

Among the various multifactor volatility models that have recently been proposed are the 2-factor GARCH model of Engle and Lee (1999), the affine and non-affine

[11] For references on state space models and related methods, see sections 2 and 3.2.1 of chapter III.
[12] Section 3.3 of chapter III introduces *realized volatility* and provides an overview of research.
[13] See Engle and Lee (1999), Gallant et al. (1999), Barndorff-Nielsen and Shephard (2001b, 2002), Alizadeh et al. (2002), Cont and Fonseca (2002), Pan (2002), Eraker et al. (2003) and Chernov et al. (2003) for recent empirical evidence on the existence of multiple volatility factors.
[14] Volatility is said to be *persistent* if it shows close-to random walk behavior.

(e.g. logarithmic) SV models considered by Gallant, Hsu and Tauchen (1999), Alizadeh, Brandt and Diebold (2002), Chernov et al. (2003), the Lévy process-driven models of Barndorff-Nielsen and Shephard (2001b, 2002) and the continuous-time pure jump Lévy process-driven ARMA SV models (CARMA) of Brockwell and Davis (2001) and Brockwell (2001). Meddahi (2001) introduces the eigenfunction approach for volatility modeling. His general class of Eigenfunction SV models includes the (multifactor) logarithmic and affine SV models as special cases. Most of these papers focus on the stock price dynamics and, as such, only use stock data for estimation. In a more direct option-pricing context, Bates (2000) proposes a 2-factor SV model with jumps in returns. Bates uses calibration to a cross-section of option prices for estimation. Duffie et al. (2000) introduce the very general class of affine jump-diffusions models for option pricing. Our multifactor model fits in this class.

Empirical implementations of multifactor volatility models show the following. Engle and Lee (1999) decompose the volatility into a permanent and transitory component, of which the latter one is mean reverting towards the former. US and Japanese stock market data support this decomposition, and reinforce the common finding of persistent volatility. The empirical results in Gallant et al. (1999) of a 2-factor non-affine volatility model show an improvement in fit of S&P500 returns over a 1-factor model. Using daily price-range data on five major US dollar exchange rates in a Kalman filter-based QML estimation strategy of non-affine SV models, Alizadeh et al. (2002) find strong evidence of two volatility factors, one very persistent and the other fast mean-reverting. Chernov et al. (2003) find that a 2-factor logarithmic SV stock price model with leverage (but without jumps) yields a better fit of daily 1953-1999 DJIA stock-index data, than do 1 or 2-factor affine SV models, or SV models with jumps. They find the first factor to be very persistent and the second quickly mean-reverting. The persistent factor does not feature *level-dependent volatility*, the other does. [15]

Investigating the information in option data

So far multifactor volatility models have been estimated (in a dynamically consistent manner) using stock return data only. As mentioned earlier, the value of an option is largely determined by the volatility of the underlying asset. As such, an evident additional, very rich source of volatility information is formed by option prices. Indeed, Tauchen (2004) points out that daily returns (even very long time series) are just not sufficiently informative to discriminate between different competing models, which appear to fit US stock-index returns about equally well. [16]

Another central aim of this thesis is to provide further empirical evidence on the existence of multiple volatility factors, by using stock *and* option data for estimation. The data we analyze consists of four time series of FTSE100-index returns and at-the-money index options of different maturities. Focusing on the affine class of SV models assuming an arbitrary number of volatility factors (but no jumps), we search for the model that best fits the joint data.

[15] A volatility factor features *level-dependent volatility* if its volatility depends on the current level of the factor. Relatedly, *level-dependent volatility-of-volatility* or *volatility feedback* means that the volatility of stock volatility fluctuations depends on the level of the volatility: A large (small) current volatility implies more (less) turbulent volatility fluctuations. See section 2 of chapter II and chapter V for more details.

[16] These are the (continuous-path) 2-factor logarithmic (i.e. exponential-linear) SV model considered in Chernov et al. (2003), the 1-factor affine SV model with price jumps considered in Andersen, Benzoni and Lund (2002) and the 1-factor affine SV model with both jumps in prices and volatility as considered in Eraker et al. (2003).

Interpreting the volatility factors
Driven by the recent evidence just discussed, our first interest is in how many volatility factors are needed to obtain an adequate description of the volatility dynamics present in the joint FTSE100 stock-index and option data. A second interest is, if we indeed find multiple SV factors, which of these factors feature level-dependent volatility and which do not. A third interest is in possible interpretations of these hidden volatility factors. What is their role in the stock volatility evolution? How do the factors affect the shape and dynamics of the volatility term structure over time? Relatedly, how does each of the unobserved volatility factors impact on the prices of options of different maturity?

1.1 Summary of main goals

To summarize, this thesis considers the pricing and hedging of financial derivatives written on assets which value is subject to stochastic volatility, in a multifactor affine SV derivative pricing model. We develop a state space, Kalman filter-based QML estimation strategy for estimating this model that circumvents simulation of option prices during estimation. Our method is fast. We show how to include stock return, realized volatility and option data (jointly) for estimation. We investigate what type of data yields the best estimation results, regarding bias, efficiency and volatility filtering. Our empirical work fits our model to time-series cross-section data consisting of FTSE100-index returns and FTSE100-index options. We search for the derivative pricing model within the affine SV class that best fits the data. We give interpretations to the volatility factors, and explain how they impact on the prices of options of different maturity. We study the nature of volatility risk, and provide an economic justification for why the volatility risk premium is found to be negative in practice.

2. Outline

The outline of the remainder of this thesis is as follows.

Chapter II, *The Multifactor Affine Stochastic Volatility Derivative Pricing Model*, considers the theoretical framework of our asset pricing model and its implications regarding derivative pricing, hedging and risk compensation.

The chapter starts with discussing the stock market setting of the model. The pricing and hedging of arbitrary complex financial derivatives receives attention, and we derive pricing formulas for European call and put options. We consider asset returns, define the volatility risk premium, and relate to classical asset pricing theories such as the Capital Asset Pricing Model and Arbitrage Pricing Theory. We develop a beta concept for derivatives similar to the CAPM beta, leading to an expected return – beta relationship for derivatives.

To set the stage for our theory on volatility risk compensation, we consider investment strategies dominated by volatility risk: *straddles*, *variance swaps* and *delta hedging* a derivative. Having discussed the empirical evidence, we introduce investor attitude into the model. Connecting to consumption-based asset pricing theory and the permanent-income hypothesis from macroeconomics, we provide an economic rationale for why the volatility risk premium is negative. The chapter ends with outlining how the model can be transformed towards derivative pricing and hedging in foreign exchange markets.

Chapter III, *A State Space Approach to the Estimation of the Multifactor Affine Stochastic Volatility Derivative Pricing Model*, proposes a first, Kalman filter-based QML estimation strategy for our derivative pricing model.

The chapter starts with an overview of current estimation methods for SV models. Here the emphasis is on recently developed methods that combine stock and option data for estimation and we argue why this is important. We discuss drawbacks of the existing methods and motivate why we advocate a state space approach. We give a review of the linear state space representation, the Kalman filter, QML estimation and diagnostic checking in state space models.

We next derive a discrete-time linear state space model from the derivative pricing model. We relate our way of extracting information from squared stock returns to earlier developed state space approaches for estimating the autoregressive SV model for stock prices. We highlight connections and differences with GARCH models. We discuss what realized volatility is, give an overview of the main strands of research towards this recent novelty in financial econometrics and explain how to incorporate high-frequency intraday data in the state space model. We then turn to our method for extracting information from option prices. Using a specific linear approximation of the call price formula, we show how information from option markets can be included in a linear state space model. A discussion of higher-order approximations is next. The chapter ends with a number of resulting possible *unconditional state space models* that are appropriate for parameter estimation and volatility extraction, and we argue why we favor one above the other.

Chapter IV, *Monte Carlo and Empirical Results for Ornstein-Uhlenbeck SV: Examining UK Financial Markets*, considers a special case of our model, the Ornstein-Uhlenbeck (OU) SV case. This chapter assumes the volatility-of-volatility to not depend on the volatility level itself. For the OU SV case, the unconditional state space model is excellently suited for estimation. As Kalman filter QML yields consistent estimates in such a linear framework, OU SV allows us to focus on the quality of e.g. the call price approximation.

The chapter starts with an explorative analysis of the FTSE100-index data. Using the information in this data, we next simulate data from the 1-factor OU SV model and examine the performance of the unconditional state space model as a means for parameter estimation and volatility filtering. We use (combinations of) squared return, realized volatility and option data for estimation.

We next confront the 1-factor OU SV model to the FTSE100-index data. Initially we use only part of the dataset for estimation. We interpret the results, compare GARCH with SV volatilities, study compensation for FTSE100-index volatility risk, and show that the 1-factor OU SV model is heavily misspecified for the full set of data. As the main misspecification appears inadequate modeling of the volatility dynamics present in the data, this naturally motivates a consideration of multiple OU volatility factors. We interpret the factors in a number of ways, consider their risk premia and implied expected straddle returns, and show how each SV factor impacts on option prices. Diagnostic checking reveals that a major shortcoming of OU SV is its negligence of level-dependent volatility-of-volatility.

Chapter V, *Exploiting Level-Dependent Volatility-of-Volatility: An Extended Kalman Filter Approach to the Estimation of the Multifactor Affine Stochastic Volatility Derivative Pricing Model*, considers affine SV processes that model level-dependent volatility-of-volatility. To better exploit the characteristics of such SV processes, we propose a second estimation method for our derivative pricing model based on the *Extended Kalman filter*.

The chapter starts with introducing the *conditional state space model* and the Extended Kalman filter. Apart from the approximations performed to arrive at a linear state space representation, Extended Kalman filter QML generates inconsistent estimates. As such, we first explore the performance of the estimation method by simulation for the 1-factor CIR SV model (i.e. the Heston (1993) model without leverage effect).

We next confront the 1-factor CIR SV model with associated estimation method to the FTSE100-index data. We compare the results to the 1-factor OU SV results, focus on the main differences and perform a specification analysis. The 1-factor CIR SV model lacks sufficient volatility dynamics to model the full set of data (which is not surprising given the OU SV results). Estimating various 2-factor affine SV specifications still reveals an inadequate modelling of volatility behavior. Extending to various 3-factor affine SV specifications shows that the most suitable model for the joint FTSE100 stock-index and option data within the affine class is a 3-factor SV model with one CIR and two affine independent volatility factors. As diagnostic checks reveal however, the model can still be improved upon in several ways, one of which is modelling of the leverage effect.

Chapter VI, *Leverage Effect and the Heston (1993) Model,* takes the leverage effect into account in a 1-factor affine SV derivative pricing model, of which the Heston (1993) model is a special case.

Chapters II - V ignore the leverage effect because it complicates matters greatly, mainly with regard to our estimation method. Nonetheless, a valid question to pose is if having neglected leverage has influenced the FTSE100-index estimation results reported in chapters IV - V in any major way. This chapter reports Monte Carlo evidence suggesting that the results seem not much distorted towards having neglected leverage, given our focus on at-the-money options only in the empirical analysis of the UK market. We concentrate hereby on the Heston (1993) model. The chapter also shows how to value and hedge derivatives in the 1-factor SV model with leverage effect, considers risk compensation and explains the differences induced by leverage as opposed to assuming it absent.

Chapter VII, *Summary and Directions for Future Research*, provides a brief summary of our findings regarding the main aims of this thesis, and indicates directions for future research.

Appendix A serves as a convenient reference regarding matrix definitions associated with the statistical properties of the volatility factors and the affine SV specification.

Appendix B derives and explores the main statistical properties of the volatility factors and the multifactor affine SV process. These properties are extensively referred to in the main text.

3. Overviews of research

At several locations in this thesis we provide overviews of literature and recent evidence on certain research topics. These include the following.

For multifactor SV models, see section 1 of this chapter. For market price of volatility risk assumptions in the literature, see section 6.1 of chapter II. For empirical evidence on volatility risk compensation, see section 6.3 of chapter II. For estimation methods for SV stock price models, see section 2.1 of chapter III. For Kalman filter-based approaches to estimating SV stock price models, see section 3.2.1 of chapter III. For state space approaches to estimating affine models of the term structure of interest rates, see section 2.1.2 of chapter III. For GARCH models, see section 3.2.2 of chapter III. For methods that combine stock and option data for estimating derivative pricing models, see section 2.1.1 of chapter III. For realized volatility research, see section 3.3.2 of chapter III. For estimated Heston (1993) model parameters by researchers in the literature, see section 5.1.1. of chapter VI. For statistical properties of affine SV processes (and of affine models of the term structure of interest rates), see appendix B.

Programming
The computations in this thesis have been carried out with the excellent object-orientated matrix programming language *Ox* (which is part of *OxMetrics*). See http://www.doornik.com and http://www.timberlake.co.uk for more details on this C/C++ related programming language.

- II -

The Multifactor Affine Stochastic Volatility Derivative Pricing Model

1. Introduction

In this chapter we propose a derivative pricing model that plays a central role in the chapters to come. We label it the *multifactor affine stochastic volatility derivative pricing model*. [1] In this model investors are exposed to two sources of risk, stock price fluctuations and stock volatility changes. This chapter discusses the theoretical framework of the model and its implications.

The outline is as follows. In section 2 we describe the stock market setting of our financial markets model. In our market investors can trade in two basic assets, a bond and a dividend-paying stock. The volatility of the stock price is assumed to be stochastic. It is determined by a number of unobserved factors, which follow mean-reverting Markov diffusions. In addition to the stock and bond, investors may also trade in derivative assets written on the stock, to meet their speculative needs or for hedging purposes.

Section 3 is devoted to the pricing of derivative securities, assuming the absence of arbitrage opportunities in the market. We consider the partial-differential-equation approach, the martingale approach and the stochastic-discount-factor approach to contingent claim valuation. We derive pricing formulas for European call and put options, and forward contracts.

Section 4 deals with hedging derivative products. As the market is incomplete, meaning that a trading strategy in the bond and stock is not sufficient to neutralize all derivatives risk, derivatives themselves must be incorporated in the hedging strategy to achieve a perfect hedge. For path-independent derivatives we derive an explicit hedging strategy in the 1-factor SV special case of the model.

Section 5 looks at asset returns and derivative betas. We develop a beta concept for derivatives, which is related to the Capital-Asset-Pricing-Model beta. It appears that a similar reward-risk relationship as the expected return - beta relationship implied by the CAPM can be derived for derivative investments in our model. Derivative returns are partly determined by a premium for volatility risk. An excursion to Arbitrage Pricing Theory reveals why we could have expected such a reward-risk relationship to hold in our arbitrage-free market beforehand. We explain the concept of spanning derivatives by other derivatives from mimicking the unobserved volatility risk premium by risk premia on traded

[1] Our model fits in the very general class of affine jump-diffusion models for asset pricing, as considered in Duffie, Pan and Singleton (2000). Alternative multifactor models for stock price dynamics are discussed in e.g., Barndorff-Nielsen and Shephard (2001b, 2002) and Chernov et al. (2003).

assets. This is closely related to market completeness, hedging and redundancy of derivatives. We also provide an alternative derivation of the derivative-return formula based on the hedging strategy. We end section 5 by discussing the link between asset returns and the stochastic discount factor.

Section 6 is devoted to volatility risk and the empirical evidence on whether investors receive a premium for (market) volatility risk in practice. Volatility-risk compensation has been attracting a lot of academic attention lately. We review the assumptions made in the literature with regard to the market price of volatility risk. We next consider several trading strategies that (partly) hedge stock risk, but not volatility risk: straddles, variance swaps and delta-hedging a derivative. If there exists a volatility risk premium in practice, it should be apparent from considering the returns on such investments. An overview of the recent empirical evidence on volatility-risk compensation is thereafter. It appears that investors seem to be willing to pay for undiversifiable market-volatility risk.

Section 7 provides an economic justification for the likely existence of a *negative* volatility risk premium. We first extend our model by incorporating a more explicit characterization of an investor and his equilibrium behaviour. We then relate expected derivative returns to an investor's optimal consumption pattern. To pursue a smooth consumption pattern over time, we argue that an investor is likely to be willing to pay a premium for consumption insurance. As delta-neutral positive-vega derivatives provide such insurance, we show that this premium takes the form of a negative market-volatility risk premium.

Section 8 considers foreign exchange markets. Although the context in the previous sections deals with a stock market setting, we explain how to easily transform to a context suitable for derivative pricing and hedging in foreign exchange markets. No additional mathematical analysis is required.

Section 9 discusses some first thoughts on how to implement our theoretical model for practical purposes as e.g., the pricing and hedging of new-to-be-issued exotic over-the-counter derivative products by a financial institution. An appendix concludes this chapter.

2. Stock market setting of the model

An arbitrage-free perfect market

Consider an arbitrage-free financial market in which trading takes place in continuous time. The calendar time is denoted by t, with $t \geq 0$. Time is measured in years. The market is assumed informationally efficient and frictionless, in the sense that there are no transaction costs, taxes or any short-sale restrictions, assets are infinitely divisible, and borrowing and lending occurs at the same rate. The market is competitive such that a single investor cannot influence prices by his individual trades. In this financial market uncertainty is resolved by an $(n+1)$-dimensional standard Brownian motion process $\{\mathbf{W}_t; t \geq 0\}$, given by

$$\mathbf{W}_t = (W_{S,t}, \mathbf{W}_{x,t}')'; \quad \mathbf{W}_{x,t} = (W_{1t}, .., W_{nt})'. \qquad (\mathbb{P}) \qquad (2.1)$$

$\mathbf{W}$ is defined on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\geq 0}, \mathbb{P})$, satisfying the "usual conditions". [2] Probability measure $\mathbb{P}$ represents the *objective* or *market measure*. The natural Brownian filtration $\{\mathcal{F}_t; t \geq 0\}$ represents the information accruing to all market participants as time goes by.

Basic assets: stock and bond

Two *primitive* or *basic assets* are traded in this market, which price processes are exogenously given. Investors can deposit part of their wealth in a risk-free savings (or money-market) account, or invest in a so-called *cash bond* B, which value process $\{B_t; t \geq 0\}$ is characterized by the ordinary differential equation

$$dB_t = r_t B_t dt \Leftrightarrow B_t = B_0 \exp\left(\int_0^t r_s ds\right), \tag{2.2}$$

in which $\{r_t; t \geq 0\}$ represents some deterministic short interest-rate process. [3]

Besides the bond a risky stock S is traded that pays dividends at a continuous rate. The ex-dividend stock price process is denoted by $\{S_t; t \geq 0\}$. The stock pays a continuously compounded dividend yield of $q_t \geq 0$ per annum at time t. The dividend payment in the time interval $[t, t + dt]$ of infinitesimal length equals $q_t S_t dt$. Under $\mathbb{P}$ the stock price evolves according to the stochastic differential equation (SDE)

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_{S,t}, \qquad (\mathbb{P}) \tag{2.3}$$

in which $\{\mu_t; t \geq 0\}$ governs the drift [4] and $\{\sigma_t; t \geq 0\}$ is a stochastic volatility process. Notice that S_t is not the price of a tradable asset, as dividends are automatically received as soon as the stock is bought. The dividend payment $q_t S_t dt$ instantaneously buys $q_t dt$ additional units of stock. By continuous immediate reinvestment of the dividends a tradable asset is obtained however. This asset will be referred to as the *reinvestment portfolio*, S^r. As explained in the appendix, its value process $\{S_t^r;\ t \geq 0\}$ satisfies

$$dS_t^r = (\mu_t + q_t) S_t^r dt + \sigma_t S_t^r dW_{S,t}. \qquad (\mathbb{P}) \tag{2.4}$$

This SDE has a natural interpretation: On average, the value of the reinvestment portfolio appreciates with the growth rate μ_t of the stock price plus the dividend yield q_t. (Notice that risk-averse investors typically require that $\mu_t + q_t \geq r_t$.) The market consisting of the bond and tradable reinvestment portfolio is arbitrage-free, as shown in the appendix.

Stochastic volatility specification

The instantaneous stock variance is driven by n possibly correlated, unobserved factors $\mathbf{x} = (x_1, .., x_n)'$ in an affine way,

$$\sigma_t^2 = \delta_0 + \boldsymbol{\delta}' \mathbf{x}_t, \tag{2.5}$$

[2] See, e.g., Protter (1990).

[3] Allowing interest rates to vary randomly over time would further complicate the mathematical analysis and would moreover introduce additional issues such as interest-rate risk compensation. Bakshi et al. (1997) show that modeling interest rates as being stochastic does not significantly improve the performance of both the Black-Scholes model and the Heston (1993) model extended with jumps in returns, as opposed to assuming constant interest rates.

[4] We assume a general drift specification of the form $\mu_t = \mu(t, S_t, \mathbf{x}_t)$.

in which δ_0 and $(nx1)\boldsymbol{\delta} = (\delta_1,..,\delta_n)'$ are positively valued. Motivations for considering a multifactor stochastic volatility (SV) specification were given before in chapter I. Under $\mathbb{P}$ the latent factors evolve according to stationary mean-reverting affine Markov diffusions, [5]

$$d\mathbf{x}_t = \mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t)dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t}, \qquad (\mathbb{P}) \qquad (2.6)$$

in which $(nx1)\boldsymbol{\theta} > \mathbf{0}$ represents the mean of the factors, $(nxn)\mathbf{K}_d, \boldsymbol{\Sigma}$ are matrices of constants with $\mathbf{K}_d = \text{diag}[k_1,..,k_n]$ being diagonal and positive definite (the stationarity condition), and $\boldsymbol{\Lambda}_t$ is a diagonal matrix given by

$$\boldsymbol{\Lambda}_t = \text{diag}\left[\sqrt{\alpha_1 + \boldsymbol{\beta}_1'\mathbf{x}_t},..,\sqrt{\alpha_n + \boldsymbol{\beta}_n'\mathbf{x}_t}\right], \qquad (2.7)$$

in which $\boldsymbol{\alpha} = (\alpha_1,..,\alpha_n)'$ and $\boldsymbol{\beta}_i = (\beta_{i1},..,\beta_{in})'$, $i = 1,..,n$ are $(nx1)$ vectors of positive real-valued constants. [6] The matrix $\mathbf{K}_d$ governs the *speed of adjustment* of the factors towards their mean $\boldsymbol{\theta}$. We assume the dynamics of the factors to be well defined, which requires $\alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_t \geq 0$ for all i and t. [7]

The Heston (1993) volatility specification (without leverage effect) in which the latent factor follows a square-root (i.e. CIR) process is a special case of this model. [8] Note that if the factors follow a multifactor Ornstein-Uhlenbeck process (in case $\boldsymbol{\beta}_i = \mathbf{0}\ \forall i$), the stock variance is not guaranteed to stay positive at all times. Although this is theoretically inconsistent, the OU assumption yields large analytical tractability. [9]

The volatility specification allows for *level-dependent volatility-of-volatility* or *volatility feedback*: If the current volatility is high (low), then the volatility of the volatility is currently high (low). [10] In other words, fluctuations in the volatility level depend on the state of the volatility. A thorough investigation of the statistical properties of this multifactor SV process can be found in appendix B. These properties will be used extensively in the sequel.

Notice that the Brownian motions W_S and $\mathbf{W}_x$ are (assumed) independent, such that stock price movements occur independently from volatility movements. The

[5] Duffie and Kan (1996) introduce this multivariate system of affine SDEs in *yield-curve* modeling.

[6] The theoretical specification of the SV process is clearly very general. Econometric estimation will require certain parameter restrictions to be imposed to avoid identification problems. These will depend on both the specific special case one considers and the particular estimation method.

[7] Parameter restrictions that ensure a unique strong solution to the SDE (2.5) are in Duffie and Kan (1996) and Dai and Singleton (2000).

[8] The volatility factor then follows $dx_t = k(\theta - x_t)dt + \sigma\sqrt{x_t}dW_{x,t}$, and $\sigma_t^2 = x_t$.

[9] Similarly, in the interest-rate literature the Gaussian OU assumption has often been used as a model for the (non-negative) short rate, see e.g. Vasicek (1977), Langetieg (1980), the continuous-time limit of the Ho-Lee (1986) model and de Jong (2000). Stein and Stein (1991) and Scott (1987) assume a 1-factor OU process for stock volatility, not the least because of its analytical tractability. A similar argument applies here: In our multifactor SV estimations to be discussed in chapter IV this tractability is more than welcome. The OU assumption ought to be considered as a first approximation. More recently, Bakshi and Kapadia (2003) also assume an OU process for the stock volatility in their proposition 2. Barndorff-Nielsen and Shephard (2001b, 2002) and Nicolato and Venardos (2002) discuss modeling the stock variance as a sum of independent non-Gaussian positive OU processes driven by positive-increment Lévy processes. See section 3.3.4 of the next chapter for details.

[10] Recent time-series evidence that supports this assumption is in Gallant, Hsu and Tauchen (1999), Pan (2002) and Jones (2003).

leverage effect (Black (1976)) is thus not modeled. [11] Although this is generally not realistic for stock markets, for foreign exchange markets this seems an adequate assumption. We refer to section 8 for a discussion.

Exposures to risk
Investors are exposed to two sources of risk in this market: stock price fluctuations and volatility changes. Risk-averse investors require compensation for bearing these risks in the form of expected asset returns. These returns are determined by the market prices of the risk sources; see section 5 for details. The market price of risk associated with stock price fluctuations, $\gamma_{S,t}$, is given by the *Sharpe ratio* (see appendix IIa),

$$\gamma_{S,t} = \frac{\mu_t + q_t - r_t}{\sigma_t}. \tag{2.8}$$

It equals the risk premium on the reinvestment portfolio expressed per unit of risk, as measured by its standard deviation. A specification of the market price of volatility risk is given shortly.

Derivative assets
Apart from the stock and bond (the basic assets) investors can trade in *derivative assets* written on the stock, like options and forward contracts. These assets derive their value from the underlying stock. Such value is therefore endogenously determined.

The condition of no arbitrage imposes certain restrictions on the price processes of these derivative securities. However, as explained and further elaborated upon in section 4.2, due to stochastic volatility the market consisting of the stock and bond only is *incomplete* in the sense that not all derivatives can be perfectly hedged by a dynamic self-financing trading strategy in the basic assets. Although the stock risk inherent in derivatives can be hedged, the inherent volatility risk cannot be hedged. Other derivatives are needed to "complete" the market; that is, to be able to neutralize all risk. Stated differently, derivatives are typically *non-redundant* assets in this market. Therefore, imposing the condition of no arbitrage is not sufficient for generating unique derivative prices. [12] As opposed to a complete-markets setting like Black-Scholes (1973), derivative prices are no longer preference-free, but instead do depend on the attitude of investors towards volatility risk. This attitude is reflected in the *market price of volatility risk*. Together with the market price of stock risk, the price of volatility risk fully determines the *risk-neutral pricing measure* $\mathbb{Q}$. Different assumptions result in different prices for derivatives that cannot be fully hedged by trading in the stock and bond only.

In order to guarantee unique derivative prices it is necessary to make additional assumptions besides no arbitrage. Rather than yet now modeling investor behavior explicitly however, we first directly focus on its implication; i.e., a certain specification for the market price of volatility risk. (A more explicit

---

[11] Modeling the leverage effect in this multifactor SV setting complicates matters dramatically, mainly with regard to our estimation method to be developed in the next chapter. Nonetheless, we hope to cover it in future research. In chapter VI we consider the Heston (1993) model, which has 1 SV factor and which allows for leverage.

[12] Exceptions are derivatives which payoff can be replicated by a generating strategy in the stock and bond, e.g. forward contracts. Their price equals the cost of the replicating bond-stock portfolio.

inclusion of investor preferences is deferred to section 7, where we propose an economic theory for volatility-risk compensation.) Following the literature on the term structure of interest rates (see e.g. Duffie and Kan (1996), de Jong (2000) and Dai and Singleton (2000)), we model the market price of risk for factor x_i -denoted by γ_{it} - as being proportional to its instantaneous standard deviation,

$$\gamma_{it} = \gamma_i \sqrt{\alpha_i + \boldsymbol{\beta}_i ' \mathbf{x}_t}\ , \qquad (2.9)$$

in which $\gamma_i \in \mathbb{R}$. The market price of volatility risk may thus be represented by the vector $\boldsymbol{\gamma}_{x,t} = (\gamma_{1t},..,\gamma_{nt})'$ given by

$$\boldsymbol{\gamma}_{x,t} = \boldsymbol{\Lambda}_t \boldsymbol{\gamma}\ , \qquad (2.10)$$

in which $\boldsymbol{\gamma} = (\gamma_1,..,\gamma_n)'$. The price of volatility risk is thus a fixed multiple of the volatility function of the risk source. It goes to zero if volatility risk goes to zero, as it should by no-arbitrage requirements. [13] Given this specification of the price of volatility risk, derivative prices can now uniquely be determined. In section 3 we return to derivative pricing.

The model under $\mathbb{Q}$

Given the prices of stock and volatility risk, from Girsanov's theorem [14], changing from the market measure $\mathbb{P}$ to the artificial pricing measure $\mathbb{Q}$ is governed by the transformation

$$d\tilde{\mathbf{W}}_t = d\mathbf{W}_t + \boldsymbol{\gamma}_t dt \Leftrightarrow \tilde{\mathbf{W}}_t = \mathbf{W}_t + \int_0^t \boldsymbol{\gamma}_u du\ , \qquad (2.11)$$

in which $\boldsymbol{\gamma}_t \equiv (\gamma_{S,t}, \boldsymbol{\gamma}_{x,t}')'$ is the vector of prices of risk. [15] The process $\{\tilde{\mathbf{W}}_t; t \geq 0\}$ with $\tilde{\mathbf{W}}_t = (\tilde{W}_{S,t}, \tilde{\mathbf{W}}_{x,t}')'$ is an $(n+1)$-dimensional drifting Brownian motion under measure $\mathbb{P}$. Under the risk-neutral probability measure $\mathbb{Q}$, which is characterized by the Radon-Nikodym derivative

$$L_T \equiv \frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(-\int_0^T \boldsymbol{\gamma}_u ' d\mathbf{W}_u - \tfrac{1}{2} \int_0^T \boldsymbol{\gamma}_u ' \boldsymbol{\gamma}_u dt \right), \qquad (2.12)$$

$\{\tilde{\mathbf{W}}_t; t \geq 0\}$ is a standard Brownian motion however. [16]

Under $\mathbb{Q}$ the ex-dividend stock price follows

$$dS_t = (r_t - q_t) S_t dt + \sigma_t S_t d\tilde{W}_{S,t}\ , \qquad (\mathbb{Q}) \qquad (2.13)$$

whereas the tradable reinvestment portfolio follows $dS_t^r = r_t S_t^r dt + \sigma_t S_t^r d\tilde{W}_{S,t}$. One advantage of the assumed form for the market price of volatility risk is the fact that it delivers the same type of mean-reverting SDE for the factors under $\mathbb{Q}$ as

---

[13] In the literature on *affine* models of the term structure of interest rates, people have recently criticized this assumption, as the risk compensation cannot switch sign over time; see e.g. Duffee (2002). Duffee extends the affine class to the *essentially affine* class, which allows for such switching. He finds that this increased flexibility results in a better fit of empirical bond returns and improved forecasts of future Treasury yields.

[14] See, e.g., Duffie (2001) or Etheridge (2002).

[15] We assume that the specifications for $\{\mu_t\},\{q_t\}$ and $\{r_t\}$ are such that the Novikov condition $\mathbb{E}_{\mathbb{P}}[\exp(\frac{1}{2}\int_0^T \boldsymbol{\gamma}_u ' \boldsymbol{\gamma}_u du)] < \infty$ necessary for Girsanov's theorem to hold is satisfied.

[16] In particular, for all Borel sets $\mathrm{B} \subset \mathbb{R}^{n+1}$, $\mathbb{Q}[\tilde{\mathbf{W}}_t \in \mathrm{B}] \equiv \mathbb{E}_{\mathbb{P}}[L_T 1_{\mathrm{B}}(\tilde{\mathbf{W}}_t)]$ where $1_{\mathrm{B}}(.)$ denotes the indicator function of the set B.

under $\mathbb{P}$. [17] Although the volatility function remains the same, both the speed of adjustment towards the mean and the mean itself differ under $\mathbb{P}$ and $\mathbb{Q}$. Some algebraic manipulations show that the factors obey the following SDE under $\mathbb{Q}$:

$$d\mathbf{x}_t = \tilde{\mathbf{K}}(\tilde{\boldsymbol{\theta}} - \mathbf{x}_t)dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\tilde{\mathbf{W}}_{x,t}, \qquad (\mathbb{Q}) \qquad (2.14)$$

with

$$\tilde{\mathbf{K}} \equiv \mathbf{K}_d + \boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathcal{B}' \qquad (2.15)$$
$$\tilde{\boldsymbol{\theta}} \equiv \tilde{\mathbf{K}}^{-1}(\mathbf{K}_d\boldsymbol{\theta} - \boldsymbol{\Sigma}\boldsymbol{\Gamma}\boldsymbol{\alpha}),$$

and

$$(nx1)\,\boldsymbol{\alpha} \equiv (\alpha_1,..,\alpha_n)', \quad (nxn)\mathcal{B} \equiv \left[\boldsymbol{\beta}_1,..,\boldsymbol{\beta}_n\right], \quad (nxn)\boldsymbol{\Gamma} \equiv \text{diag}\left[\gamma_1,..,\gamma_n\right],$$

and in which we assume the inverse of the matrix $\tilde{\mathbf{K}}$ to exist. This completes the description of the multifactor affine SV derivative pricing model.

3. Derivative pricing

In this section we consider the pricing of derivative securities written on the stock by enforcing the condition of no arbitrage. Section 3.1 describes the *partial-differential-equation*, the *martingale* and *stochastic-discount-factor approaches* to contingent claim valuation. Their intimate connections are highlighted. In section 3.2 we derive pricing formulas for European call and put options that hold in our market. In everyday banking practice these *vanilla* options are often combined with a stock position to create a wide range of different payoffs, to meet speculative or hedging needs of clients. Such *packages* are the simplest form of *exotic* options. We refer to Hull (2003) for further discussion.

3.1 Valuation of a general contingent claim

Consider a general tradable European-style derivative F written on the stock S that pays off an amount $F_T = f(S_T)$ at some fixed future maturity date T. For now we assume that this payoff is a function $f(.)$ of the terminal stock price S_T only. (*Path-dependent* derivatives, which payoff depends on the path followed by the stock price during the life of the contract, are considered at the end of this section.) Our aim is to derive the time-t price, F_t, of such a claim for $t < T$. Let us assume that this price is given by some function $F_t = F(t, S_t, \mathbf{x}_t)$ of t, S_t and $\mathbf{x}_t$ only. [18] Itô's formula [19] then yields the SDE the derivative price follows under the market measure $\mathbb{P}$: [20]

[17] Chernov and Ghysels (2000), Pan (2002) and Jones (2003) make similar assumptions.

[18] Notice that we –moreover- implicitly assume that such a price *exists*. Currently we do not know if this is the case, let alone if it is *unique*.

[19] The following should be noted. Whenever we perform general derivations in this thesis, the implicit assumption is that the processes and functions involved are such that technical conditions as e.g. Itô and Riemann-integrability conditions are satisfied. We will not repeat this all the time to enhance the readability of this thesis.

[20] $\text{tr}(.)$ denotes the trace of a matrix. The first equality uses the result

$$d\mathbf{x}_t'\frac{\partial^2 F_t}{\partial\mathbf{x}_t\partial\mathbf{x}_t'}d\mathbf{x}_t = d\mathbf{W}_{x,t}'\boldsymbol{\Lambda}_t\boldsymbol{\Sigma}'\frac{\partial^2 F_t}{\partial\mathbf{x}_t\partial\mathbf{x}_t'}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t} = \text{tr}\left(\boldsymbol{\Lambda}_t\boldsymbol{\Sigma}'\frac{\partial^2 F_t}{\partial\mathbf{x}_t\partial\mathbf{x}_t'}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_t\right)dt.$$

$$dF_t = \left[\frac{\partial F_t}{\partial t} + \tfrac{1}{2}\sigma_t^2 S_t^2 \frac{\partial^2 F_t}{\partial S_t^2} + \tfrac{1}{2}\,\mathrm{tr}\left(\mathbf{\Lambda}_t \mathbf{\Sigma}' \frac{\partial^2 F_t}{\partial \mathbf{x}_t \partial \mathbf{x}_t'} \mathbf{\Sigma\Lambda}_t \right) \right] dt + \frac{\partial F_t}{\partial S_t} dS_t + \frac{\partial F_t}{\partial \mathbf{x}_t'} d\mathbf{x}_t$$

$$= \left[\frac{\partial F_t}{\partial t} + \mu_t S_t \frac{\partial F_t}{\partial S_t} + \frac{\partial F_t}{\partial \mathbf{x}_t'} \mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t) + \tfrac{1}{2}\sigma_t^2 S_t^2 \frac{\partial^2 F_t}{\partial S_t^2} + \tfrac{1}{2}\,\mathrm{tr}\left(\mathbf{\Lambda}_t \mathbf{\Sigma}' \frac{\partial^2 F_t}{\partial \mathbf{x}_t \partial \mathbf{x}_t'} \mathbf{\Sigma\Lambda}_t \right) \right] dt$$

$$+ \frac{\partial F_t}{\partial S_t} \sigma_t S_t dW_{S,t} + \frac{\partial F_t}{\partial \mathbf{x}_t'} \mathbf{\Sigma\Lambda}_t d\mathbf{W}_{x,t}\,. \qquad (\mathbb{P}) \qquad (3.1)$$

We need to rule out arbitrage strategies that include this derivative instrument. The SDE (3.1) is not the SDE the arbitrage-free derivative price follows, as we have not yet enforced the no-arbitrage condition in its derivation. That is our next aim. Under the risk-neutral measure $\mathbb{Q}$ the derivative price follows

$$dF_t = \left(\frac{\partial F_t}{\partial t} + (\mu_t - \gamma_{S,t}\sigma_t) S_t \frac{\partial F_t}{\partial S_t} + \frac{\partial F_t}{\partial \mathbf{x}_t'} \left[\mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t) - \mathbf{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t} \right] + \tfrac{1}{2}\sigma_t^2 S_t^2 \frac{\partial^2 F_t}{\partial S_t^2} \right) dt$$

$$+ \tfrac{1}{2}\,\mathrm{tr}\left(\mathbf{\Lambda}_t \mathbf{\Sigma}' \frac{\partial^2 F_t}{\partial \mathbf{x}_t \partial \mathbf{x}_t'} \mathbf{\Sigma\Lambda}_t \right) dt + \frac{\partial F_t}{\partial S_t} \sigma_t S_t d\tilde{W}_{S,t} + \frac{\partial F_t}{\partial \mathbf{x}_t'} \mathbf{\Sigma\Lambda}_t d\tilde{\mathbf{W}}_{x,t}\,. \qquad (\mathbb{Q}) \qquad (3.2)$$

The *Fundamental Theorem of Asset Pricing* states that the prices of all tradable assets expressed in terms of the bond price should be martingales under $\mathbb{Q}$ if arbitrage is to be excluded. [21] For our choice of $\mathbb{Q}$, we prove in appendix Ia that the relative prices of the bond and reinvestment portfolio (the stock with dividends reinvested) are indeed $\mathbb{Q}$-martingales. Under $\mathbb{Q}$ the relative derivative price evolves according to

$$d(B_t^{-1}F_t) = B_t^{-1} dF_t - r_t B_t^{-1} F_t dt \qquad (\mathbb{Q}) \qquad (3.3)$$

$$= B_t^{-1} \left(\frac{\partial F_t}{\partial t} + .. + \tfrac{1}{2}\,\mathrm{tr}\left(\mathbf{\Lambda}_t \mathbf{\Sigma}' \frac{\partial^2 F_t}{\partial \mathbf{x}_t \partial \mathbf{x}_t'} \mathbf{\Sigma\Lambda}_t \right) - r_t F_t \right) dt$$

$$+ B_t^{-1} \frac{\partial F_t}{\partial S_t} \sigma_t S_t d\tilde{W}_{S,t} + B_t^{-1} \frac{\partial F_t}{\partial \mathbf{x}_t'} \mathbf{\Sigma\Lambda}_t d\tilde{\mathbf{W}}_{x,t}\,,$$

in which the expression for the dots is as in (3.2). As martingales are driftless, imposing no-arbitrage implies that the drift of this SDE must be zero. This results in the following restriction:

$$r_t F_t = \frac{\partial F_t}{\partial t} + (\mu_t - \gamma_{S,t}\sigma_t) S_t \frac{\partial F_t}{\partial S_t} + \frac{\partial F_t}{\partial \mathbf{x}_t'} \left[\mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t) - \mathbf{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t} \right] \qquad (3.4)$$

---

21 Harrison and Kreps (1979), Harrison and Pliska (1981), and Delbaen and Schachermayer (1994). The *First Fundamental Theorem of Asset Pricing* states that the absence of arbitrage is equivalent to the existence of at least one probability measure $\mathbb{Q}$ -equivalent to the market measure $\mathbb{P}$-, under which the prices of all tradable assets relative to the price of some numéraire asset, are $\mathbb{Q}$-martingales. Such a $\mathbb{Q}$ is called an *equivalent martingale measure* (EMM). Each asset characterized by –at all times- a strictly positive price can act as a numéraire asset. The martingale measure(s) is (are) numéraire-dependent. The conventional numéraire-asset choice is the cash bond, and the associated EMM(s) –if it exist- is commonly called the *risk-neutral measure*. The *Second Fundamental Theorem of Asset Pricing* states that the EMM $\mathbb{Q}$ is *unique* if and only if the market is *complete*; that is –essentially-, if all risks can be hedged by trading in the primitive (i.e. exogenously given) assets.

$$+\tfrac{1}{2}\sigma_t^2 S_t^2 \frac{\partial^2 F_t}{\partial S_t^2} + \tfrac{1}{2}\operatorname{tr}\left(\mathbf{\Lambda}_t \mathbf{\Sigma}' \frac{\partial^2 F_t}{\partial \mathbf{x}_t \partial \mathbf{x}_t'} \mathbf{\Sigma\Lambda}_t\right).$$

From (3.2), under the risk-neutral measure the arbitrage-free derivative price thus follows

$$dF_t = r_t F_t dt + \frac{\partial F_t}{\partial S_t}\sigma_t S_t d\tilde{W}_{S,t} + \frac{\partial F_t}{\partial \mathbf{x}_t'}\mathbf{\Sigma\Lambda}_t d\tilde{\mathbf{W}}_{x,t}. \qquad (\mathbb{Q}) \qquad (3.5)$$

Recall that the reinvestment portfolio follows $dS_t^r = r_t S_t^r dt + \sigma_t S_t^r d\tilde{W}_{S,t}$ under $\mathbb{Q}$. It is clear that enforcing no-arbitrage results in all tradable assets earning the same rate (on average) as the numéraire asset; i.e., the risk-free rate of interest r_t. Under the artificial probability measure $\mathbb{Q}$ investors thus demonstrate risk-neutral behavior: Although trading in the assets entails risk, as compensation for bearing this risk they only require the risk-free rate of return. They are thus indifferent towards the risks involved. Hence the name *risk-neutral* measure for $\mathbb{Q}$.

Finally, transforming back to $\mathbb{P}$ using (2.11) yields the SDE the arbitrage-free derivative price follows under the market measure:

$$dF_t = \left[r_t + \left(\frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t}\right)\sigma_t \gamma_{S,t} + \left(\frac{1}{F_t}\frac{\partial F_t}{\partial \mathbf{x}_t'}\right)\mathbf{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t}\right]F_t dt + \frac{\partial F_t}{\partial S_t}\sigma_t S_t dW_{S,t} + \frac{\partial F_t}{\partial \mathbf{x}_t'}\mathbf{\Sigma\Lambda}_t d\mathbf{W}_{x,t}.$$

$$(\mathbb{P}) \qquad (3.6)$$

Partial-differential-equation approach to derivative pricing

The no-arbitrage restriction (3.4) implicitly defines a second-order or parabolic partial differential equation (PDE) that the derivative price $F_t = F(t, S_t, \mathbf{x}_t)$ must satisfy for $t < T$. The boundary condition of this PDE is formed by the derivative-specific payoff $F_T = f(S_T)$, which is realized at maturity T. Together they uniquely define the derivative price at time t.

As mentioned in section 2, notice that this price is indeed not preference-free: Although we may simplify $\mu_t - \gamma_{S,t}\sigma_t = r_t - q_t$ in (3.4) such that compensation for stock price uncertainty (μ_t) has no derivative-pricing implications, the attitude of investors towards volatility risk (as reflected in the price of volatility risk, $\boldsymbol{\gamma}_{x,t}$) does matter.

The PDE and hence the derivative price can in principle be solved for by engaging in numerical methods such as finite difference methods. [22] However, even if there is only one latent factor the computations are cumbersome (as this already requires a 3-dimensional grid of (t, S_t, x_t) to evaluate the derivative price on), and their number rapidly increases when the number of factors increases.

Martingale approach

Fortunately, the derivative price can also be obtained from a different angle. If we integrate the SDE (3.3) over the interval $[t, T]$ with the no-arbitrage restriction (3.4) imposed, we obtain

[22] Finite difference methods replace differentials by differences. See e.g. Hull (2003) for an introduction to these methods.

$$B_T^{-1}F_T = B_t^{-1}F_t + \int_t^T B_u^{-1} \frac{\partial F_u}{\partial S_u} \sigma_u S_u d\tilde{W}_{S,u} + \int_t^T B_u^{-1} \frac{\partial F_u}{\partial \mathbf{x}_u'} \boldsymbol{\Sigma\Lambda}_u d\tilde{\mathbf{W}}_{x,u} \,.\, (\mathbb{Q}) \tag{3.7}$$

Taking conditional $\mathbb{Q}$-expectations yields

$$\mathbb{E}_{\mathbb{Q}}\left[B_T^{-1}F_T \mid \mathcal{F}_t\right] = B_t^{-1}F_t + \mathbb{E}_{\mathbb{Q}}\left[\int_t^T B_u^{-1} \frac{\partial F_u}{\partial S_u} \sigma_u S_u d\tilde{W}_{S,u} \mid \mathcal{F}_t\right] + \mathbb{E}_{\mathbb{Q}}\left[\int_t^T B_u^{-1} \frac{\partial F_u}{\partial \mathbf{x}_u'} \boldsymbol{\Sigma\Lambda}_u d\tilde{\mathbf{W}}_{x,u} \mid \mathcal{F}_t\right]. \tag{3.8}$$

As the Itô-integrals involved in (3.8) are $(\mathbb{Q},\{\mathcal{F}_t\})-$ martingales, we subsequently obtain the following *risk-neutral valuation formula* for the derivative price:

$$F_t = B_t\, \mathbb{E}_{\mathbb{Q}}\left[B_T^{-1}F_T \mid \mathcal{F}_t\right] = \exp\left(-\int_t^T r_u du\right)\mathbb{E}_{\mathbb{Q}}\left[F_T \mid \mathcal{F}_t\right], \tag{3.9}$$

in which the second equality uses the assumption of deterministic interest rates. The fact that the discounted conditional expectation on the right-hand side of (3.9) is a function of $(t,S_t,\mathbf{x}_t)$ only (and does not depend on the history $\mathcal{F}_{t^-}$), is a consequence of the assumption of a Markov process for $\mathbf{x}$, and the assumption of the payoff function to be of the form $F_T = f(S_T)$.

The solution to the (deterministic) PDE (3.4) with boundary condition $F_T = f(S_T)$ is thus given by the pricing function $F_t = F(t,S_t,\mathbf{x}_t)$, which can be expressed as the conditional expectation (3.9). We have essentially proven the *Feynman-Kac Stochastic Representation Theorem* in our setting (see, e.g., Duffie (2001)).

Besides solving the PDE, the derivative price can thus also be computed as a discounted probability-weighted average of all possible payoffs that may be realized under the risk-neutral measure. This price can generally only be obtained by Monte Carlo simulation, if the model parameters are known. Notice that as investors are risk neutral under $\mathbb{Q}$, it makes sense that the proper (deterministic) discount factor is derived from the risk-free interest rate.

Stochastic-discount-factor approach

As a third possible way to compute the derivative price, an implication of Girsanov's theorem (see e.g. Shreve (1997)) is that the derivative price (3.9) can also be rewritten as a certain discounted payoff under the market measure $\mathbb{P}$. Specifically,

$$F_t = \exp\left(-\int_t^T r_u du\right)\mathbb{E}_{\mathbb{P}}\left[\frac{L_T}{L_t}F_T \mid \mathcal{F}_t\right] = \mathbb{E}_{\mathbb{P}}\left[\frac{M_T}{M_t}F_T \mid \mathcal{F}_t\right], \tag{3.10}$$

in which $\{L_t; t \geq 0\}$ is the *Radon-Nikodym process* defined by

$$L_t = \left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = \exp\left[-\int_0^t \boldsymbol{\gamma}_u' d\mathbf{W}_u - \tfrac{1}{2}\int_0^t \boldsymbol{\gamma}_u' \boldsymbol{\gamma}_u du\right], \tag{3.11}$$

and the process $\{M_t; t \geq 0\}$ with $M_t \equiv B_t^{-1}L_t$ is known as the *stochastic discount factor* (SDF) process. Commonly used alternative names are *state-price deflator*, *state-price density*, *pricing kernel* or *marginal rate of substitution* (see also section 7). Notice that $M_tF_t = \mathbb{E}_{\mathbb{P}}[M_TF_T \mid \mathcal{F}_t]$, which illustrates that the product of the stochastic discount factor and the price of any tradable asset is a $(\mathbb{P},\{\mathcal{F}_t\})-$martingale.

Formula (3.10) shows that the derivative price can also be computed as the expected value of its discounted payoff under the market measure $\mathbb{P}$, but now using a *stochastic* discount factor M_T / M_t that involves the market prices of stock and volatility risk. Notice that if these prices of risk were both equal to zero, then the discount factor reduces to $\exp[-\bar{r}_t(T-t)]$ with $\bar{r}_t$ as in (3.14); i.e., the one normally used for discounting non-risky payoffs. This implies that investors would be risk-neutral under $\mathbb{P}$ in that case. This is clearly not realistic in practice; investors are typically risk-averse and therefore require compensation for bearing (non-diversifiable) risks.

The stochastic discount factor thus clearly corrects for risk. Notice that the SDF is not asset-specific. It carries all pricing information and prices all assets in the economy if there is no arbitrage (see e.g. Cochrane (2001)). We have basically illustrated the fact that the First Fundamental Theorem of Asset Pricing –which provides us with (3.9)- can essentially also be stated as follows: The absence of arbitrage is equivalent to the existence of at least one (strictly positive) stochastic discount factor that prices all payoffs in the market. It will be clear that the Second Fundamental Theorem can alternatively be stated as: The SDF is unique if and only if the market is complete.

Path-dependent derivatives

So far we have assumed payoff functions of the form $F_T = f(S_T)$. Path-dependent derivatives are characterized by payoffs given by $F_T = f(S_u; 0 \le u \le T)$. Prominent examples are *forward start options*, *barrier, Asian* and *lookback options*. The pricing function of these derivatives is typically *not* of the form $F_t = F(t, S_t, \mathbf{x}_t)$.

Consider for example a forward start call, in which the holder receives at some future date T_1 (at no extra cost) a European call option with strike $K = S_{T_1}$ (i.e., at-the-money) and maturity $T_2 > T_1$. For $t \in [T_1, T_2]$ the value F_t of the forward start thus equals the value of a European call option with a fixed strike S_{T_1}, as S_{T_1} is part of the current information set $\mathcal{F}_t$. From (3.19) it is clear that $F_t = F(t, S_t, S_{T_1}, \mathbf{x}_t)$ for $t \in [T_1, T_2]$.

The PDE (3.4) is derived from the assumption that $F_t = F(t, S_t, \mathbf{x}_t)$ using Itô's lemma. As a result this PDE can in general not be used for pricing such path-dependent derivatives. [23] However, as the Fundamental Theorem of Asset Pricing states that relative asset prices should all be $(\mathbb{Q},\{\mathcal{F}_t\})-$martingales if there is to be no arbitrage, it immediately follows that $F_t / B_t = \mathbb{E}_{\mathbb{Q}}[F_T / B_T \mid \mathcal{F}_t]$. This leads directly to the risk-neutral valuation formula (3.9), and thus (3.10). Hence, these formulas can still be used for pricing path-dependent derivatives.

Let us return to the forward-start example. For $t \in [0, T_1)$ the strike price S_{T_1} is still uncertain. The risk-neutral-valuation formula (3.9) nevertheless applies. Due to the Markov property it now does hold that $F_t = F(t, S_t, \mathbf{x}_t)$.

[23] For some path-dependent claims a different PDE can be obtained that the price should satisfy. However, as stressed by Hull (2003), finite difference methods are very difficult to apply in those cases.

3.2 Call and put option valuation

In this section we value some specific derivative instruments. Consider a European call option C written on the stock S that has strike price K and maturity $T > t$. Its payoff function is given by $C_T = \max\{0, S_T - K\}$. From the risk-neutral valuation formula (3.9), its arbitrage-free time-t price C_t is given by

$$C_t = B_t\, \mathbb{E}_{\mathbb{Q}}\left[B_T^{-1} C_T \mid \mathcal{F}_t\right] = \exp\left(-\int_t^T r_u du\right) \mathbb{E}_{\mathbb{Q}}\left[\max\{0, S_T - K\} \mid \mathcal{F}_t\right]. \qquad (3.12)$$

To compute this expectation requires the conditional distribution of S_T under $\mathbb{Q}$. Itô's lemma yields the SDE for $\ln S_t$ from (2.13). If we next integrate the increments $d\ln S_t$ over $[t,T]$ and then take exponents, results in

$$S_T = S_t \exp\left[(\bar{r}_t - \bar{q}_t - \tfrac{1}{2}\bar{\sigma}_t^2)\tau_t + \int_t^T \sigma_u d\tilde{W}_{S,u}\right], \qquad (\mathbb{Q}) \qquad (3.13)$$

in which we define the current time to maturity τ_t, the average interest rate ($\bar{r}_t$), dividend yield ($\bar{q}_t$), and variance ($\bar{\sigma}_t^2$) -all over the remaining option's life-, by:

$$\tau_t \equiv T - t, \quad \bar{r}_t \equiv \frac{1}{\tau_t}\int_t^T r_u du, \quad \bar{q}_t \equiv \frac{1}{\tau_t}\int_t^T q_u du, \quad \bar{\sigma}_t^2 \equiv \frac{1}{\tau_t}\int_t^T \sigma_u^2 du \;. \qquad (3.14)$$

Recall the assumption of independence between the volatility and stock-price driving $\mathbb{P}$-Brownian motions $\mathbf{W}_x$ and W_S. This independence is preserved under $\mathbb{Q}$, such that $\tilde{\mathbf{W}}_x$ and $\tilde{W}_S$ are also independent. [24] Therefore, conditioning on the sample path of $\tilde{\mathbf{W}}_{x,u}$ for $t \le u \le T$ implies that the volatility path $\{\sigma_u; t \le u \le T\}$ is known, such that the Itô integral in (3.13) is conditionally normally distributed,

$$\int_t^T \sigma_u d\tilde{W}_{S,u} \mid \{\tilde{\mathbf{W}}_{x,u}\} \sim \mathcal{N}\left(0, \int_t^T \sigma_u^2 du\right), \qquad (\mathbb{Q}) \qquad (3.15)$$

in which $\{\tilde{\mathbf{W}}_{x,u}\}$ is short-hand notation for $\{\tilde{\mathbf{W}}_{x,u}; t \le u \le T\}$. In turn this implies that $S_T \mid \mathcal{F}_t, \{\tilde{\mathbf{W}}_{x,u}\}$ is lognormally distributed. We may thus write

$$S_T \mid \mathcal{F}_t, \{\tilde{\mathbf{W}}_{x,u}\} = S_t \exp\left[(\bar{r}_t - \bar{q}_t - \tfrac{1}{2}\bar{\sigma}_t^2)\tau_t + \bar{\sigma}_t\sqrt{\tau_t}\, \varepsilon\right], \qquad (\mathbb{Q}) \qquad (3.16)$$

with $\varepsilon \sim \mathcal{N}(0,1)$ under $\mathbb{Q}$. Next, express the conditional call payoff in terms of ε:

$$\max\{0, S_T - K\} \mid \mathcal{F}_t, \{\tilde{\mathbf{W}}_{x,u}\} = \begin{cases} S_t \exp\left[(\bar{r}_t - \bar{q}_t - \tfrac{1}{2}\bar{\sigma}_t^2)\tau_t + \bar{\sigma}_t\sqrt{\tau_t}\, \varepsilon\right] - K, & \text{if } \varepsilon > -d_{2t} \\ 0, \quad \text{if } \varepsilon \le -d_{2t} \end{cases}$$

in which

$$d_{2t} \equiv d_{1t} - \bar{\sigma}_t\sqrt{\tau_t}, \qquad d_{1t} \equiv \frac{\ln\dfrac{S_t}{K} + (\bar{r}_t - \bar{q}_t + \tfrac{1}{2}\bar{\sigma}_t^2)\tau_t}{\bar{\sigma}_t\sqrt{\tau_t}}. \qquad (3.17)$$

---

[24] This is so as the probability measures $\mathbb{P}$ and $\mathbb{Q}$ are *equivalent*; i.e., they agree on the same null sets.

Given these results, we obtain

$$\mathbb{E}_{\mathbb{Q}}\left[\max\{0, S_T - K\} \mid \mathcal{F}_t, \{\tilde{\mathbf{W}}_{x,u}\}\right]$$

$$= S_t \exp\left[(\bar{r}_t - \bar{q}_t)\tau_t\right] \int_{-d_{2t}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\left(\varepsilon - \bar{\sigma}_t\sqrt{\tau_t}\right)^2\right] d\varepsilon - K\,\Phi(d_{2t})$$

$$= S_t \exp\left[(\bar{r}_t - \bar{q}_t)\tau_t\right]\, \Phi(d_{1t}) - K\,\Phi(d_{2t}), \qquad (3.18)$$

in which $\Phi(.)$ denotes the cumulative standard Gaussian distribution function, and where the last equality follows after making the change-of-variable towards $\xi \equiv \varepsilon - \bar{\sigma}_t\sqrt{\tau_t}$, and performing some further manipulations.

The arbitrage-free call price at time t can subsequently be obtained by invoking the law of iterated expectations:

$$C_t = \exp(-\bar{r}_t\tau_t)\ \mathbb{E}_{\mathbb{Q}}\left(\mathbb{E}_{\mathbb{Q}}\left[\max\{0, S_T - K\} \mid \mathcal{F}_t, \{\tilde{\mathbf{W}}_{x,u}\}\right] \mid \mathcal{F}_t\right)$$

$$= \mathbb{E}_{\mathbb{Q}}\left[S_t \exp(-\bar{q}_t\tau_t)\,\Phi(d_{1t}) - K\exp(-\bar{r}_t\tau_t)\,\Phi(d_{2t}) \mid \mathcal{F}_t\right]$$

$$= \mathbb{E}_{\mathbb{Q}}\left[BS(S_t, K, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t\right], \qquad (3.19)$$

where

$$BS(S_t, K, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2) \equiv S_t \exp(-\bar{q}_t\tau_t)\,\Phi(d_{1t}) - K\exp(-\bar{r}_t\tau_t)\,\Phi(d_{2t}) \qquad (3.20)$$

stands for the conventional Black-Scholes (BS) call price adjusted for continuous dividend payments, evaluated in the arguments S_t, K, τ_t, $\bar{r}_t, \bar{q}_t$ and $\bar{\sigma}_t^2$.

If $F_{t,T}$ represents the fair time-t forward price of the forward contract written on S that has the same maturity $\tau_t = T - t$ as the call option, i.e., [25]

$$F_{t,T} = S_t \exp\left[(\bar{r}_t - \bar{q}_t)\tau_t\right], \qquad (3.21)$$

then the call valuation formula can conveniently be rewritten as

$$C_t = \mathbb{E}_{\mathbb{Q}}\left[BS(F_{t,T}, K, \tau_t, \bar{r}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t\right], \qquad (3.22)$$

with

$$BS(F_{t,T}, K, \tau_t, \bar{r}_t, \bar{\sigma}_t^2) \equiv \exp(-\bar{r}_t\tau_t)\left[F_{t,T}\,\Phi(d_{1t}) - K\,\Phi(d_{2t})\right], \qquad (3.23)$$

$$d_{1t} = \frac{\ln\frac{F_{t,T}}{K} + \frac{1}{2}\bar{\sigma}_t^2\tau_t}{\bar{\sigma}_t\sqrt{\tau_t}}, \quad d_{2t} = d_{1t} - \bar{\sigma}_t\sqrt{\tau_t}\ . \qquad (3.24)$$

[25] This forward price is derived from no-arbitrage arguments as follows. Consider the following portfolio strategy: At time t we buy a number of $\exp(-\bar{q}_t\tau_t)$ stocks, and reinvest all dividends received immediately in the stock as time goes by. Furthermore, we borrow an amount of $\exp(-\bar{r}_t\tau_t)F_{t,T}$ against the short rate $\{r_t\}$. The forward contract expires at maturity T and results in a payoff of $S_T - F_{t,T}$. This payoff is exactly replicated by the payoff of liquidating the portfolio at time T: At time T we have exactly one stock in our portfolio which we sell for S_T in the market, whereas the cash has to be paid back with interest, which results in a negative cash flow of $-F_{t,T}$. By imposing no-arbitrage, the initial cash flow of entering the forward contract at time t must match the cost of setting up the portfolio. As it costs nothing to enter the forward contract at time t as it is an obligation and not a right, this leads us to the restriction $0 \equiv -\exp(-\bar{q}_t\tau_t)S_t + \exp(-\bar{r}_t\tau_t)F_{t,T}$, such that the arbitrage-free time-t forward price equals $F_{t,T} = S_t\exp[(\bar{r}_t - \bar{q}_t)\tau_t]$.

The advantage of this latter expression is that the average dividend yield, which may be hard to estimate in practice, does not play a role anymore in this formula.

The analysis shows that in our multifactor SV model, the value of a call option is obtained as an expectation of the Black-Scholes call price, where the expectation is taken over all possible volatility paths that may realize over the remaining life of the option, under the risk-neutral measure. Notice that although we allow for multifactor SV, a similar type of valuation formula results as in the 1-factor SV model considered by Hull and White (1987). The reason is twofold. First, the volatility and stock price processes are assumed independent (the leverage effect is not modeled). Second, the stock price follows a geometric Brownian motion-type SDE under $\mathbb{Q}$.

Put option valuation
European put options can be valued by no-arbitrage arguments as follows. We focus on a put P with the same strike K and remaining maturity $\tau_t = T - t$ as the call option above. Consider the following trading strategy. At time $t < T$ we buy one put at price P_t, and $\exp(-\bar{q}_t \tau_t)$ stocks for price S_t each. All dividends received are immediately reinvested in the stock as time goes by, such that at maturity T we have exactly one stock in our portfolio. Consider next the strategy of buying one call for C_t at time t, and investing an amount of money equal to $K\exp(-\bar{r}_t \tau_t)$ in the cash bond. This amount has grown to K at time T. As both strategies yield exactly the same payoff at time T, irrespective of specific stock-price movements, it must be the case –by no-arbitrage- that the costs of setting up each strategy exactly match. Therefore, the price of the put satisfies

$$P_t = C_t + \exp(-\bar{r}_t \tau_t) K - \exp(-\bar{q}_t \tau_t) S_t . \qquad (3.25)$$

This is the so-called *put-call parity* relation that holds in our market. Using the forward price from (3.21), the put price can also be computed from $P_t = C_t + \exp(-\bar{r}_t \tau_t)(K - F_{t,T})$.

Deterministic and constant-volatility cases: Black-Scholes option prices
For future reference, a final remark. If volatility is *deterministic* instead of stochastic, it follows from our analysis that the call price is given by equations (3.20) and (3.17), or, in terms of the forward price, by (3.23)-(3.24). If volatility is further restricted to be *constant,* such that the ex-dividend stock price (2.3) evolves as $dS_t = \mu_t S_t dt + \sigma S_t dW_{S,t}$ for some $\sigma > 0$ under $\mathbb{P}$, our model essentially reduces to the conventional Black-Scholes model. In that case the call price is given by either equations (3.20) and (3.17), or, in terms of the forward price, by equations (3.23)-(3.24), but with the random $\bar{\sigma}_t$ replaced by the parameter σ. In both cases the put option value follows again from the put-call parity (3.25), which remains in tact in case of deterministic and constant volatility, like the formula for the forward price (3.21) does.

4. Hedging derivatives

If our model is to be used by a financial institution in practice, then *hedging* becomes important. Suppose there are a number of underlying risk factors that determine the value of a certain position in assets. Uncertain movements in these risk factors result in value changes, which may be (temporarily) undesirable.

Hedging aims at building protection against such unwanted value changes by taking in an offsetting position in other assets, which is subject to the same risk factors. One may desire to hedge against all risk factors simultaneously –if possible-, or against a subset of them. In the latter case the risk is only partly neutralized. In general, an offsetting hedge position only temporarily provides protection. If prolonged protection is desired, the hedge needs to be rebalanced.

In section 4.1 we first consider measures to quantify the various exposures to risk inherent in derivative securities. Section 4.2 takes a closer look at the hedging problem in our incomplete-markets model. In section 4.3 we derive an explicit hedging strategy for derivatives with payoff function $F_T = f(S_T)$ in the 1-factor SV special case of our model. We also comment on hedging path-dependent derivatives, and hedging in practice.

4.1 Delta, vega and other Greeks

Consider a general European-style derivative F. This derivative may or may not be path-dependent. The arbitrage-free time-t price F_t of this derivative is given by the risk-neutral valuation formula (3.9), or, equivalently, by the SDF formula (3.10). This price is perceptive to the two sources of randomness in our model: stock price fluctuations and latent factor changes; i.e., volatility movements.

A common way of measuring derivatives risk is by focusing on derivative-price sensitivities. These sensitivities are known as the *Greeks*; we refer to Hull (2003) for details. The *delta* (Δ) of derivative F is defined as its price sensitivity with respect to stock price movements, whereas its *gamma* (Γ) measures the rate of change in its delta. That is,

$$\Delta_{F,t} \equiv \frac{\partial F_t}{\partial S_t}, \qquad \Gamma_{F,t} \equiv \frac{\partial \Delta_{F,t}}{\partial S_t} = \frac{\partial^2 F_t}{\partial S_t^2}. \tag{4.1}$$

Both are obviously varying over time. In our multifactor SV setting, it is most convenient to define the *vega* ($\mathcal{V}$) of derivative F as the derivative-price sensitivity towards changes in the stock-variance-driving factors: [26]

$$\mathcal{V}_{F,t} \equiv \frac{\partial F_t}{\partial \mathbf{x}_t}. \tag{4.2}$$

Notice that the vega is an ($nx1$) vector. In the 1-factor SV special case with imposed restrictions $\delta_0 = 0,\ \delta_1 = 1$ such that $\sigma_t^2 = x_t$, the vega equals $\mathcal{V}_{F,t} = \partial F_t / \partial \sigma_t^2$, and measures price changes as a result of changes in the stock variance. Its price sensitivity with respect to volatility movements is then given by $\partial F_t / \partial \sigma_t = 2\sigma_t \mathcal{V}_{F,t}$. Other Greeks are *rho*, which is defined as $\text{rho}_{F,t} \equiv \partial F_t / \partial r_t$, and *theta* ($\Theta$), defined as $\Theta_{F,t} \equiv \partial F_t / \partial t$. Less commonly known is *epsilon,* which measures dividend risk: $\text{epsilon}_{F,t} \equiv \partial F_t / \partial q_t$.

---

[26] In the special case of $F_t = F(t, S_t, \mathbf{x}_t)$ with $n \geq 2$ factors, notice that

$$\mathcal{V}_{F,t} = \frac{\partial F_t}{\partial \mathbf{x}_t} \neq \frac{\partial \hat{F}_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \mathbf{x}_t} = \frac{\partial \hat{F}_t}{\partial \sigma_t^2} \boldsymbol{\delta} = \frac{1}{2\sigma_t} \left(\frac{\partial \hat{F}_t}{\partial \sigma_t} \right) \boldsymbol{\delta},$$

as it can be shown that it does *not* hold that F_t can be (re)written as a function of $\sigma_t^2 = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t$ directly; i.e., $F_t = F(t, S_t, \mathbf{x}_t) \neq \hat{F}(t, S_t, \sigma_t^2)$. Unfortunate as this is, as it would facilitate and further embellish the analysis in the coming sections.

4.2 Hedging in incomplete markets

Due to stochastic volatility our market is *incomplete*, in the sense that most derivatives cannot be hedged by a conventional *delta-hedging strategy* in the stock (cum dividends) and bond only. Those that can be hedged in this way have a price that does not depend on the volatility-driving factors $\mathbf{x}$; e.g., a forward contract. To be able to hedge a general derivative however, requires adding other derivatives to the set of hedging instruments. In a sense one then completes the market.

Let us explain the hedging problem in more detail. For simplicity we focus on a derivative with payoff $F_T = f(S_T)$. As explained in section 3.1, its time-t price F_t is given by some function $F(t,S_t,\mathbf{x}_t)$ of t, S_t and $\mathbf{x}_t$ only. A change in this price is either caused by stock price fluctuations or volatility movements (or just by the passage of time). Although such a derivative position can be made instantaneously *delta neutral* (i.e., a zero delta) by taking in an offsetting position in the stock, *vega neutrality* cannot be achieved with a stock investment only, as price changes due to volatility shocks cannot be hedged with it. Hence, a delta-hedging strategy will only partly neutralize risk, and will therefore exhibit a tracking error. (In section 6.2.3 we return to this issue when discussing *delta-hedged gains*.) In more technical terms, the price change dF_t (i.e. infinitesimal increment) which ought to be hedged against, depends on both the stock price and factor changes dS_t and $d\mathbf{x}_t$, and thus on the Brownian increments $dW_{S,t}$ and $d\mathbf{W}_{x,t}$; recall the SDEs (3.1) and (3.6). With a stock investment one can only hedge against the risky W_S, but not against the risky $\mathbf{W}_x$, which is clear from reconsidering the stock price SDE (2.3).

To be able to hedge against volatility risk, we need to rely on other derivatives. These derivatives essentially proxy for the *non-tradable* unobserved *volatility*. The principle is as follows. Recall that $F_t = F(t,S_t,\mathbf{x}_t)$, such that if we observe the prices of n derivatives in the market, then the current factor value $\mathbf{x}_t$ can in principle be backed out by an inverse transformation. The latent volatility-driving factors $\mathbf{x}_t$ then become essentially observable. Hence, these derivatives indirectly make volatility a tradable asset. Therefore, in our market, a hedging strategy for a derivative which price is sensitive to the $(n+1)$ risk factors W_S and $\mathbf{W}_x$ through S and $\mathbf{x}$, incorporates both the bond, stock and n other derivatives. [27]

4.3 Example: hedging in 1-factor SV models

As an example, consider hedging in the 1-factor SV special case of our model. For simplicity we set $\delta_0 = 0$ and $\delta_1 = 1$ such that $\sigma_t^2 = x_t$. The latent factor evolves as

$$dx_t = k(\theta - x_t)dt + \sigma\lambda_t dW_{x,t} \qquad (\mathbb{P}) \qquad (4.3)$$

under $\mathbb{P}$, with $\lambda_t \equiv \sqrt{\alpha + \beta x_t}$.

[27] These n derivatives ought to be chosen in such a way that together they are able to fully hedge the $\mathbf{W}_x$-risk. Thus they ought to be "linear-independent" assets, in the sense that their vegas must be linearly independent. Intuitively, we cannot hedge more risk with two assets that equally respond to the $\mathbf{W}_x$-risk as we can with just one of these assets. Moreover, the maturity of each of these n derivatives ought to be at least as large as the maturity of the derivative that we want to hedge.

4.3.1 Hedging path-independent derivatives

In this section our interest is in hedging a general European-type derivative F written on the stock S that pays off an amount $F_{T_F} = f(S_{T_F})$ at its maturity T_F. Notice that we explicitly assume that this payoff is a function of the terminal stock price only. In section 4.3.2 we cover the hedging of path-dependent derivatives.

From (3.6) the arbitrage-free price $F_t = F(t, S_t, x_t)$ evolves according to the SDE

$$dF_t = \mu_{F,t} F_t dt + \Delta_{F,t} \sigma_t S_t dW_{S,t} + \mathcal{V}_{F,t} \sigma \lambda_t dW_{x,t} , \qquad (\mathbb{P}) \qquad (4.4)$$

with $\mu_{F,t} \equiv r_t + (S_t / F_t)\Delta_{F,t}\sigma_t\gamma_{S,t} + (1/F_t)\mathcal{V}_{F,t}\sigma\lambda_t\gamma_{x,t}$. Consider next another, longer-maturity European-type derivative G that pays off $G_{T_G} = g(S_{T_G})$ at its maturity $T_G > T_F$. Derivative G will serve as a hedging instrument for F. Its arbitrage-free price $G_t = G(t, S_t, x_t)$ follows

$$dG_t = \mu_{G,t} G_t dt + \Delta_{G,t} \sigma_t S_t dW_{S,t} + \mathcal{V}_{G,t} \sigma \lambda_t dW_{x,t} , \qquad (\mathbb{P}) \qquad (4.5)$$

in which $\mu_{G,t} \equiv r_t + (S_t / G_t)\Delta_{G,t}\sigma_t\gamma_{S,t} + (1/G_t)\mathcal{V}_{G,t}\sigma\lambda_t\gamma_{x,t}$.

Our aim is now to design a *dynamic self-financing hedging strategy* for derivative F, which takes both the bond B, the dividend-paying stock S, and derivative G into account. Derivative payoffs are stated in terms of the ex-dividend stock price. As such derivative prices are functions of S_t, and not of the reinvestment-portfolio value S_t^r. Now recall that the process $\{S_t\}$ does not represent the price of a tradable asset. The desired trading strategy must therefore be in terms of the reinvestment portfolio S^r, as this *is* a tradable asset. Moreover, as we desire a *self-financing* strategy which, by definition, does not allow for an outflow of wealth, the automatically received stock dividends need to be immediately reinvested again, which makes the reinvestment portfolio a natural candidate to hedge stock price risk with. Recall that the bond price evolves as $dB_t = r_t B_t dt$, the stock price as $dS_t = \mu_t S_t dt + \sigma_t S_t dW_{S,t}$, and the value of the reinvestment portfolio as $dS_t^r = (\mu_t + q_t) S_t^r dt + \sigma_t S_t^r dW_{S,t}$, all under measure $\mathbb{P}$.

A self-financing hedging strategy

Assume that we start hedging F at time 0, for maturity T_F. The dynamic hedging strategy is given by an $\mathcal{F}_t$-adapted triplet $(\psi_t, \phi_t^r, \xi_t)$ for $t \in [0, T_F]$. Here, ψ_t governs the bond holding, ϕ_t^r is the number of reinvestment portfolios, and ξ_t is the number of $G-$derivatives to hold in the hedge portfolio at time t. Recall that starting off with 1 stock at time 0 followed by continuously reinvesting received dividends, implies that the time-t value of the reinvestment portfolio is given by $S_t^r = \exp(\bar{q}_{0t} t) S_t$, where $\bar{q}_{0t}$ is the average dividend yield over $[0, t]$.

Our object now is to determine the positions $(\psi_t, \phi_t^r, \xi_t)$ to hold in each instrument at each point in time, such that the hedging strategy is self-financing and replicates the payoff of derivative F. The value of the hedge portfolio at time t, V_t, is given by

$$V_t = \psi_t B_t + \phi_t^r S_t^r + \xi_t G_t . \qquad (4.6)$$

We want the strategy to be *replicating*. Therefore, we impose the restriction that $V_{T_F} = F_{T_F}$. Moreover, we require it to be *self-financing*, such that the infinitesimal increment in the hedge portfolio value must satisfy

$$dV_t = \psi_t dB_t + \phi_t^r dS_t^r + \xi_t dG_t \Leftrightarrow V_t = V_0 + \int_0^t \psi_u dB_u + \int_0^t \phi_u^r dS_u^r + \int_0^t \xi_u dG_u \; . \quad (4.7)$$

Then, by no arbitrage, it must hold that the value of the hedging strategy equals the derivative value at all times, i.e. $V_t = F_t \;\; \forall t$. It then automatically follows that their increments must match, such that $dV_t = dF_t$. In other words, the infinitesimal change in the derivative price is *perfectly* hedged by the change in value of the hedge portfolio, as required. The expressions for the instrument holdings ψ_t, ϕ_t^r and ξ_t then follow from equating the coefficients of the dt, $dW_{x,t}$ and $dW_{S,t}$ terms in the equality $dV_t = dF_t$. This equality is given by

$$\begin{aligned} dV_t &= \psi_t dB_t + \phi_t^r dS_t^r + \xi_t dG_t \\ &= \left[\psi_t r_t B_t + \phi_t^r(\mu_t + q_t)S_t^r + \xi_t \mu_{G,t} G_t\right]dt + \left[\phi_t^r \sigma_t S_t^r + \xi_t \Delta_{G,t}\sigma_t S_t\right]dW_{S,t} + \left[\xi_t \mathcal{V}_{G,t}\sigma\lambda_t\right]dW_{x,t} \\ &\equiv \mu_{F,t}F_t dt + \Delta_{F,t}\sigma_t S_t dW_{S,t} + \mathcal{V}_{F,t}\sigma\lambda_t dW_{x,t} \; . \qquad (4.8) \end{aligned}$$

Specifically, we obtain that the number of G-derivatives to hold in the hedge portfolio at time t equals the ratio of vegas (as long as $\mathcal{V}_{G,t} \neq 0$),

$$\xi_t = \mathcal{V}_{F,t} / \mathcal{V}_{G,t} \; , \qquad (4.9)$$

whereas the number of reinvestment portfolios should equal

$$\phi_t^r = \frac{S_t}{S_t^r}\left[\Delta_{F,t} - \xi_t \Delta_{G,t}\right] = \exp(-\bar{q}_{0t}t)\left[\Delta_{F,t} - \frac{\mathcal{V}_{F,t}}{\mathcal{V}_{G,t}}\Delta_{G,t}\right]. \qquad (4.10)$$

Notice that since $\phi_t^r S_t^r = [\Delta_{F,t} - (\mathcal{V}_{F,t} / \mathcal{V}_{G,t})\Delta_{G,t}]S_t$, a self-financing strategy of ϕ_t^r invested in the reinvestment portfolio corresponds to a strategy of $\phi_t \equiv \Delta_{F,t} - (\mathcal{V}_{F,t} / \mathcal{V}_{G,t})\Delta_{G,t} = \exp(\bar{q}_{0t}t)\phi_t^r$ invested in the stock, as $\phi_t^r S_t^r = \phi_t S_t$. Finally, equating the dt terms yields for the bond holding

$$\psi_t = \frac{1}{r_t B_t}\left[\mu_{F,t}F_t - \xi_t \mu_{G,t} G_t - \phi_t^r(\mu_t + q_t)S_t^r\right], \qquad (4.11)$$

with ξ_t and ϕ_t^r as above. For an alternative expression, recall that $V_t = \psi_t B_t + \phi_t^r S_t^r + \xi_t G_t \equiv F_t$ by no arbitrage, such that ψ_t can also be computed as $\psi_t = B_t^{-1}[F_t - \xi_t G_t - \phi_t^r S_t^r] = B_t^{-1}[F_t - \xi_t G_t - \phi_t S_t]$, which is more convenient. It requires some algebra to show that both expressions for ψ_t are indeed the same.

“Spanning” and redundancy

We have shown that we are able to perfectly hedge a general derivative F with payoff function $F_{T_F} = f(S_{T_F})$ by a continuous-time self-financing trading strategy in the bond, stock and some longer-maturity derivative G having payoff function $G_{T_F} = g(S_{T_F})$. In particular, if we are *long* F we must *short* the hedging strategy (and vice versa), to be sure that at maturity we end up with a zero net cash position. Derivative G assumes the role of hedging that part of F's value change due to unobserved volatility movements. Taking derivative G as a hedging instrument into account ensures that we have essentially completed the market, in the sense that shorter-maturity derivatives are now *attainable* by a *generating* strategy in the bond, stock and derivative G .

We might even state this differently. Together the bond, stock and derivative G essentially *span* the space of derivative instruments that have a shorter maturity than G: their payoffs can all be perfectly replicated by a self-financing trading strategy. The stock and bond only do not span that space, as these assets do not allow volatility risk to be "traded". This is basically the reason why the bond-stock market is said to be incomplete. Notice furthermore that these shorter-maturity derivatives are *redundant* assets as soon as we take derivative G into account.

4.3.2 Hedging path-dependent derivatives

Derivatives with payoff function $F_T = f(S_T)$ can perfectly be hedged by implementing the dynamic trading strategy (4.9)-(4.11). The strategy hinges on the fact that for such derivatives $F_t = F(t, S_t, x_t)$. But what about hedging path-dependent derivatives? It turns out that although for these derivatives a perfect hedging strategy exists, explicit general expressions for the instrument holdings can unfortunately not be given.

Consider hedging a general derivative F that is now allowed to be path-dependent. It pays off F_{T_F} at its maturity T_F. As before, our hedging instruments consist of the bond B, the reinvestment portfolio S^r, and some longer-maturity derivative G that pays off $G_{T_G} = g(S_{T_G})$ at its maturity $T_G > T_F$, such that its price is given by $G_t = G(t, S_t, x_t)$. (Note the explicit assumption of $G_{T_G} = g(S_{T_G})$.)

Collect the prices of the latter two assets in the vector $\mathbf{H}_t = (S_t^r, G_t)'$, and collect their bond-discounted prices in $\mathbf{Z}_t = (Z_t^r, Z_t^G)' \equiv B_t^{-1}\mathbf{H}_t$. The SDE G_t obeys under $\mathbb{Q}$ follows from (3.5). Moreover, S_t^r satisfies $dS_t^r = r_t S_t^r dt + \sigma_t S_t^r d\tilde{W}_{S,t}$ under $\mathbb{Q}$. As such

$$\begin{pmatrix} dS_t^r \\ dG_t \end{pmatrix} = r_t \begin{pmatrix} S_t^r \\ G_t \end{pmatrix} dt + \begin{pmatrix} \sigma_t S_t^r & 0 \\ \Delta_{G,t}\sigma_t S_t & \mathcal{V}_{G,t}\sigma\lambda_t \end{pmatrix} \begin{pmatrix} d\tilde{W}_{S,t} \\ d\tilde{W}_{x,t} \end{pmatrix}, \qquad (\mathbb{Q}) \qquad (4.12)$$

or $d\mathbf{H}_t = r_t \mathbf{H}_t dt + \mathbf{\Sigma}_t^H d\tilde{\mathbf{W}}_t$ with $\mathbf{\Sigma}_t^H$ the 2x2 matrix in (4.12). It follows that under the risk-neutral measure, $d\mathbf{Z}_t = d(B_t^{-1}\mathbf{H}_t) = B_t^{-1}d\mathbf{H}_t + \mathbf{H}_t dB_t^{-1} = B_t^{-1}\mathbf{\Sigma}_t^H d\tilde{\mathbf{W}}_t$. The relative hedge-instrument prices $\mathbf{Z}$ thus follow a 2-dimensional $\mathbb{Q}$-martingale process with respect to the natural filtration $\{\mathcal{F}_t\}$ of the 2-dimensional standard $\mathbb{Q}$-Brownian motion $\tilde{\mathbf{W}}$. By the tower property of conditional expectations, the discounted claim process $\{Б_t; 0 \le t \le T_F\}$, defined by

$$Б_t \equiv \mathbb{E}_{\mathbb{Q}}\left[B_{T_F}^{-1} F_{T_F} \mid \mathcal{F}_t \right], \qquad (4.13)$$

is a 1-dimensional $(\mathbb{Q}, \{\mathcal{F}_t\})-$martingale. It then follows from the *multifactor Brownian martingale representation theorem* (MRT; see e.g. Etheridge (2002)), that there exists a unique $\{\mathcal{F}_t\}-$previsible [28] 2-dimensional process $\{\boldsymbol{\varphi}_t; 0 \le t \le T_F\}$ such that

$$Б_t = Б_0 + \int_0^t \boldsymbol{\varphi}_u' d\mathbf{Z}_u \Leftrightarrow dБ_t = \boldsymbol{\varphi}_t' d\mathbf{Z}_t. \qquad (4.14)$$

In words, for all t the increment in the martingale Б is a unique previsible linear combination of the increments in the martingales Z^r and Z^G.

[28] As our model does not allow for jumps such that all sample paths are continuous, there is no distinction between $\{\mathcal{F}_t\}-$predictability and $\{\mathcal{F}_t\}-$adaptedness.

A self-financing hedging strategy
Given these preliminaries, consider next the following hedging strategy for derivative F. At time t hold

$$\begin{cases} \boldsymbol{\varphi}_t \text{ units of the instruments } \mathbf{H} = (S^r, G)' \\ \psi_t \equiv \text{Б}_t - \boldsymbol{\varphi}_t{}'\mathbf{Z}_t \text{ units of the cash bond } B \end{cases} \tag{4.15}$$

in the hedging portfolio. The value V_t of the hedge at time t equals $V_t = \boldsymbol{\varphi}_t{}'\mathbf{H}_t + \psi_t B_t = B_t \text{Б}_t$. In particular, at the expiration date T_F of derivative F we have $V_{T_F} = B_{T_F}\text{Б}_{T_F} = B_{T_F}\mathbb{E}_{\mathbb{Q}}[B_{T_F}^{-1}F_{T_F} \mid \mathcal{F}_{T_F}] = F_{T_F}$, as the payoff F_{T_F} is $\mathcal{F}_{T_F}$-measurable. The hedge strategy (4.15) thus replicates the payoff of derivative F. It is moreover self-financing:

$$\begin{aligned} dV_t &= B_t d\text{Б}_t + \text{Б}_t dB_t \\ &= B_t\boldsymbol{\varphi}_t{}' d\mathbf{Z_t} + (\psi_t + \boldsymbol{\varphi}_t{}'\mathbf{Z}_t)dB_t \\ &= B_t\boldsymbol{\varphi}_t{}'\left[B_t^{-1}d\mathbf{H}_t - r_t B_t^{-1}\mathbf{H}_t dt\right] + \psi_t dB_t + r_t B_t \boldsymbol{\varphi}_t{}'\mathbf{Z}_t dt \\ &= \boldsymbol{\varphi}_t{}' d\mathbf{H}_t + \psi_t dB_t. \end{aligned} \tag{4.16}$$

This analysis shows that a unique self-financing perfect hedging strategy *exists* for each and every derivative in our model, be it path-dependent or not. [29] Our proof is based on an application of the MRT. Although a strong result, its weakness is that the MRT is not constructive, in the sense that the MRT only tells us that *some* unique process $\{\boldsymbol{\varphi}_t\}$ exists such that (4.14) holds, without giving guidance how to find it. Obviously, if we cannot find the process $\{\boldsymbol{\varphi}_t\}$, we cannot physically implement the hedging strategy (4.15).

For derivatives characterized by a payoff function of the form $F_T = f(S_T)$ we derived the explicit expressions for $(\boldsymbol{\varphi}_t, \psi_t)$ in the previous section; i.e., (4.9)-(4.11). Although for path-dependent derivatives a perfect hedging strategy thus exists, whether one succeeds in finding it, will typically depend on both the specific derivative and the creativity of the financial engineer. In light of our discussion in the next subsection, instead of trying to find the perfect hedging strategy for path-dependent derivatives, a satisfactory alternative might be to resort to an approximate strategy; e.g., the strategy (4.9)-(4.11).

4.3.3 Hedging in practice

The perfect hedging strategy outlined in section (4.3.1) requires *continuous rebalancing* of the hedge portfolio -as expressed by the time-varying instrument holdings $(\psi_t, \phi_t^r, \xi_t)$. In our frictionless market this poses no problem. However, in practice where markets are less than perfect, continuous rebalancing is obviously infeasible if only due to *transaction costs* or *market-microstructure effects*. As a result a financial institution faces a trade-off between being prone to sufficiently

---

[29] Notice that no-arbitrage thus dictates $F_t \equiv V_t = B_t\text{Б}_t = B_t\mathbb{E}_{\mathbb{Q}}[B_{T_F}^{-1}F_{T_F} \mid \mathcal{F}_t]$. Our analysis essentially reveals the following. Suppose we extend our set of basic assets (bond and stock) with derivative G. That is, suppose its price process (4.5) is now exogenously given (and hence is not determined within the model). Suppose next that we want to price a new-to-be-issued (shorter-maturity) derivative F. Since we can replicate F's payoff by a self-financing trading strategy in (B, S^r, G) as we just proved, its price must be equal to the cost of the replicating portfolio, V_t, by no-arbitrage. This next leads us to the risk-neutral valuation formula $F_t = V_t = B_t\mathbb{E}_{\mathbb{Q}}[B_T^{-1}F_T \mid \mathcal{F}_t]$ (where $\mathbb{Q}$ is the measure under which $\mathbf{Z}_t = (B_t^{-1}S_t^r, B_t^{-1}G_t)'$ is a martingale), and the notion that the relative derivative price $B_t^{-1}F_t$ is itself a $\mathbb{Q}$-martingale. Moreover, the market is now complete. Indeed, we have essentially proven the Fundamental Theorem of Asset Pricing in this case.

limited risk and the cost of hedging. Rebalancing at discrete intervals will not eliminate all risk: delta and vega-neutrality can only be achieved for a short while, after which rebalancing is needed. For example, if gamma is large (in absolute value) then delta changes quickly, and the hedge portfolio ought to be sooner adjusted in order to maintain sufficiently limited protection against stock price fluctuations. [30]

5. Derivative betas and asset returns

In our model investors are exposed to two sources of risk: stock price fluctuations and volatility changes. Risk-averse investors require a reward for bearing (undiversifiable) risks. Such compensation takes the form of expected asset returns. These returns are determined by the market prices of stock and volatility risk. In this section we explore the various risk premia earned on assets traded in our market. It appears that in our model, a similar expected return - beta relationship as implied by the *Capital Asset Pricing Model* (CAPM [31]) can be derived for derivatives.

In section 5.1 we start with developing a concept that is related to the CAPM beta. We label it the *derivative betas*. A derivative's stock and volatility beta have a natural interpretation in terms of expected derivative returns, as explained in section 5.2. In section 5.3 we relate our derivative pricing model to Ross's (1976) *Arbitrage Pricing Theory* (APT). We explain why the specific form of the derivative-return formula could be expected in advance. We show how to mimic the premium for volatility risk by risk premia on traded assets. The link to APT proves valuable in later sections. In section 5.4 we derive the derivative-return formula from a different viewpoint: the hedging strategy. In section 5.5 we explore the relation between derivative returns and the stochastic discount factor. It paves the way for the connection with consumption-based asset pricing theory and the premium for volatility risk, both to be developed in section 7.

5.1 CAPM and stock and volatility betas

The expected return – beta, or reward-risk relationship predicted by the CAPM is central in modern finance. The conventional CAPM implies that the risk premium on an individual asset equals the product of its *beta* and the risk premium on the market portfolio, which comprises all assets. In symbols: $\mathbb{E}[r_i] - r_f = \beta_i(\mathbb{E}[r_m] - r_f)$. Here, r_i is the return on asset i, r_f is the risk-free rate, r_m is the return on the market portfolio, and β_i is the beta of security i, which is defined as $\beta_i \equiv \text{cov}[r_i, r_m]/\text{var}[r_m]$. The asset beta measures the extent in which the asset and market returns move together. The CAPM implies that the appropriate asset risk premium demanded by investors is determined by the extent in which this

[30] In an interesting paper, Figlewski (1989) explores by simulation in the BS framework the impact of various market imperfections, if the usual delta-hedging strategy is rebalanced only daily. He finds that daily rebalancing has a considerable impact on its risk, but also leads to substantial transaction costs. Bertsimas et al. (2000) derive the asymptotic distribution of the tracking error incurred when discrete-time delta-hedging a call option in the BS model. Although its distribution is symmetric with zero mean, the root-mean-squared tracking error is of order $N^{-1/2}$, where N is the number of rebalancing times. Bakshi and Kapadia (2003) allow for 1-factor SV and delta-hedge a call with a stock and bond investment only. They show that the discrete *delta-hedged gains* -as they call them- is centered around zero unless volatility risk is priced (i.e. a non-zero market price of volatility risk). See also section 6.2.3.

[31] Sharpe (1964), Lintner (1965), Mossin (1966). See e.g. Bodie Kane and Marcus (1996) for a standard textbook treatment of the CAPM.

asset contributes to the overall risk of the fully diversified market portfolio. The market portfolio yields a better risk-return profile than holding the individual assets -the diversification effect. All *idiosyncratic* (i.e. firm-specific) risk has been diversified away; only *systematic* (i.e. market) risk remains. Therefore, investors do not care about an asset's risk as measured by its standard deviation of returns; it is its beta that matters.

Based on these insights, our aim is to devise an analogy between our derivative pricing model and the CAPM. In our market derivative returns are influenced by two underlying "systematic" risk factors, stock price and latent factor movements. For simplicity, in the remainder of this chapter we focus on derivatives characterized by $F_t = F(t, S_t, \mathbf{x}_t)$; i.e., path-independent derivatives.

Similar to the CAPM beta, we define the *stock beta* of a derivative F, denoted by $\beta_{F,S}$ [32], as the covariance between the spot derivative and stock returns, as a fraction of the instantaneous stock return variance, given current information:

$$\beta_{F,S,t} \equiv \text{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, \frac{dS_t}{S_t} \mid \mathcal{F}_t\right] / \text{var}_{\mathbb{P}}\left[\frac{dS_t}{S_t} \mid \mathcal{F}_t\right] = \frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t} = \frac{S_t}{F_t}\Delta_{F,t}. \quad (5.1)$$

We refer to the appendix for a derivation. The stock beta clearly varies over time. It equals the elasticity of the pricing function $F(t, S_t, \mathbf{x}_t)$ with respect to S_t, or the product of leverage S_t / F_t and delta. It is a unitless measure. [33] Moreover, note that the beta of the "systematic" risk factor S equals 1.

To develop a beta measure with respect to the hidden factors $\mathbf{x}$, it is again natural to focus on the derivative return. In this case however, we focus on its sensitivity towards *absolute* instead of relative factor changes. As the factors drive the stock volatility, and as absolute volatility changes directly affect the derivative price and hence return, this seems most natural. Another argument is that the factors are not assets, such that there seems no need for focusing on relative changes. [34] As such, we define F's *volatility beta* as

$$\boldsymbol{\beta}_{F,x,t} \equiv \left(\text{var}_{\mathbb{P}}\left[d\mathbf{x}_t \mid \mathcal{F}_t\right]\right)^{-1} \text{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, d\mathbf{x}_t \mid \mathcal{F}_t\right] = \frac{1}{F_t}\frac{\partial F_t}{\partial \mathbf{x}_t} = \frac{1}{F_t}\mathcal{V}_{F,t}, \quad (5.2)$$

with $\boldsymbol{\beta}_{F,x,t} = (\beta_{F,1,t}, \ldots, \beta_{F,n,t})'$. A derivation is again in the appendix. This beta essentially equals the covariance between the spot derivative return and absolute factor changes, corrected for the instantaneous variation in factor value changes. The beta of F with respect to volatility-driving factor x_i is given by $\beta_{F,i,t} = (1/F_t)(\partial F_t / \partial x_{it})$, $i = 1,..,n$. It represents the vega with respect to factor x_i, scaled by the derivative price. Notice that if $\boldsymbol{\Sigma}$ is diagonal such that $\text{var}_{\mathbb{P}}[d\mathbf{x}_t \mid \mathcal{F}_t]$ is diagonal, then $\beta_{F,i,t} = \text{cov}_{\mathbb{P}}[dF_t / F_t, dx_{it} \mid \mathcal{F}_t] / \text{var}_{\mathbb{P}}[dx_{it} \mid \mathcal{F}_t]$.

---

[32] The symbols $\beta_{F,S}$ and $\boldsymbol{\beta}_{F,x} = (\beta_{F,1}, \ldots, \beta_{F,n})'$ for the derivative betas ought not be confused with the symbol for the parameter vector $\boldsymbol{\beta}_i$ which consists of parameters β_{ij} $(i, j = 1,..,n)$, and which occurs in the SDE for the factors $\mathbf{x}$.

[33] This obviously follows from focusing on *returns* instead of *absolute changes*, which is most natural. As prices in economic models are generally expressed in terms of some numéraire asset (e.g. the cash bond or the currency "money"), by focusing on asset returns this unit of measurement drops out.

[34] Recall that $\mathbf{x}_t$ governs the variance of the instantaneous stock *return*, i.e. $\text{var}_{\mathbb{P}}[dS_t / S_t \mid \mathcal{F}_t] = \sigma_t^2 dt = (\delta_0 + \boldsymbol{\delta}'\mathbf{x}_t)dt$ and is therefore already numéraire-free.

The SDE (3.6) the arbitrage-free derivative price follows can be rewritten in terms of the stock and volatility betas, as follows:

$$\frac{dF_t}{F_t} = \left[r_t + \beta_{F,S,t} \left(\sigma_t \gamma_{S,t} \right) + \boldsymbol{\beta}_{F,x,t}{}' \left(\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t} \right) \right] dt + \beta_{F,S,t} \left(\sigma_t dW_{S,t} \right) + \boldsymbol{\beta}_{F,x,t}{}' \left(\boldsymbol{\Sigma\Lambda}_t d\mathbf{W}_{x,t} \right). \quad (\mathbb{P}) \qquad (5.3)$$

This equation makes perfect sense now: Given time-t information, the spot derivative return over the next infinitesimal time interval consists of a deterministic (i.e. adapted) part (see next section) and a random part, which decomposes into two components. The first component consists of the derivative's sensitivity towards stock price fluctuations as measured by its beta $\beta_{F,S,t}$, multiplied by the exogenous shock $\sigma_t dW_{S,t}$ occurring in the instantaneous stock return dS_t / S_t. The second component equals the product of the volatility beta $\boldsymbol{\beta}_{F,x,t}$ and the shock $\boldsymbol{\Sigma\Lambda}_t d\mathbf{W}_{x,t}$ occurring in the factor change $d\mathbf{x}_t$. The shocks to the derivative return have a very natural interpretation now.

5.2 Asset returns and the volatility risk premium

In this section we explore the returns investors are expected to earn when investing in the various assets traded in our market.

Consider first the cash bond. Its price varies deterministically according to $dB_t = r_t B_t dt$. The instantaneous bond return dB_t / B_t is given by the non-random quantity $r_t dt$. It represents the *time value of money* or "*market price of time*": By investing in the bond and thereby postponing consumption, an investor is guaranteed to make a risk-free rate of return equal to the bond rate.

Consider next the dividend-paying stock. The stock price itself does not represent the value of a tradable asset. The reinvestment portfolio is tradable however. In terms of the market price of stock risk (2.8), under $\mathbb{P}$ its value evolves as $dS_t^r = (r_t + \sigma_t \gamma_{S,t}) S_t^r dt + \sigma_t S_t^r dW_{S,t}$. This investment's expected spot return equals

$$\mathbb{E}_{\mathbb{P}} \left[\frac{dS_t^r}{S_t^r} \mid \mathcal{F}_t \right] = (r_t + \sigma_t \gamma_{S,t}) dt = (\mu_t + q_t) dt. \qquad (5.4)$$

It provides a clear interpretation of the market price of stock risk, $\gamma_{S,t}$. The risk premium on this asset equals the product of the amount of risk currently inherent in the stock price (as measured by the volatility of its fluctuations, σ_t), and the price of this risk, measured by $\gamma_{S,t}$. Notice that this risk premium may depend on the current level of the stock volatility, as we assumed $\mu_t = \mu(t, S_t, \mathbf{x}_t)$ in general terms. It may be argued that investors demand a larger (smaller) compensation when risk (i.e. volatility) is high (low), which would require $\partial \mu_t / \partial \mathbf{x}_t > \mathbf{0}$.

Consider finally a general derivative F. From (5.3) the expected spot derivative return equals

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left[\frac{dF_t}{F_t} \mid \mathcal{F}_t \right] &= \left[r_t + \beta_{F,S,t} \left(\sigma_t \gamma_{S,t} \right) + \boldsymbol{\beta}_{F,x,t}{}' \left(\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t} \right) \right] dt \qquad (5.5) \\ &= \left[r_t + \beta_{F,S,t} \left(\mu_t + q_t - r_t \right) + \boldsymbol{\beta}_{F,x,t}{}' \left(\boldsymbol{\Sigma\Lambda}_t^2 \boldsymbol{\gamma} \right) \right] dt, \end{aligned}$$

where the latter equality follows from substituting the expressions for the market prices of stock and volatility risk, $\gamma_{S,t} = (\mu_t + q_t - r_t) / \sigma_t$ and $\boldsymbol{\gamma}_{x,t} = \boldsymbol{\Lambda}_t \boldsymbol{\gamma}$. This is the *expected return - beta* or reward–risk relationship that holds for derivative securities in our market. It says that the return an investor is expected to earn consists of three components (of which the third consists of n separate components). The risk-free rate is earned as compensation for delayed consumption by investing in the derivative instead of immediate expenditure, i.e. the time value of money. Both other components are compensation for being exposed to the risks underlying the derivative. Compensation for stock risk equals the product of the risk premium on the dividend-paying stock and the amount of stock risk inherent in the derivative, as measured by its beta $\beta_{F,S,t}$. Compensation for volatility risk equals the product of the volatility beta $\boldsymbol{\beta}_{F,x,t}$ and a vector, which we will refer to as the *volatility risk premium:* $\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t} = \boldsymbol{\Sigma\Lambda}_t^2 \boldsymbol{\gamma}$.

The volatility risk premium consists of n separate *factor risk premia*, one for each factor x_i. [35] The premium depends on the current state of the volatility, unless the factors follow an OU process, for which it is constant. If there is currently no volatility risk such that the volatility-of-volatility is zero, the volatility risk premium is zero, as required by no arbitrage. Depending on the value of $\boldsymbol{\gamma}$, the volatility risk premium can be zero, negative or positive. It is of obvious interest to investigate both the sign and magnitude of the volatility risk premium in real-world financial markets. We refer to sections 6 and 7 for further discussion.

5.3 Asset returns and the relation with Arbitrage Pricing Theory

In our market, enforcing no-arbitrage results in the reward–risk relationship (5.5) for investing in derivatives. This relationship is similar in spirit to what Ross's (1976) *Arbitrage Pricing Theory* (APT) predicts. So could we have expected a relation as (5.5) to hold in our market beforehand? At first sight our derivative pricing and the APT setting seem only superficially related. In this section we explain the correspondences. Perhaps not surprisingly, it is the arbitrage argument that provides the connection. We also show how the volatility risk premium can be mimicked by risk premia on traded assets. As in sections 5.2 and 5.3 where we connected to the CAPM, the analysis below provides valuable insights in continuous-time derivative pricing theory in general, and its links with more traditional asset pricing models.

APT and factor-mimicking portfolios

Ross's (1976) Arbitrage Pricing Theory assumes that a set of abstract *common factors* $\mathbf{f}$ drives security returns in a perfect, competitive market. Specifically, Ross assumes the 1-period return on asset i, r_i, to be generated by

$$r_i = a_i + \mathbf{b}_i{}'\mathbf{f} + e_i, \qquad (\mathbb{P}) \qquad (5.6)$$

in which $e_i \mid \mathbf{f} \sim (0, \sigma_i^2)$, and a_i and $\mathbf{b}_i$ are constants with the vector $\mathbf{b}_i$ measuring the asset's sensitivity to the systematic factors $\mathbf{f}$. These sensitivities are also known as *factor loadings*. The error term e_i measures the idiosyncratic or specific risk in the security return that cannot be attributed to the systematic factors that drive the economy. If the e_i's are sufficiently uncorrelated across assets, then the specific risk of a large well-diversified portfolio can be shown to approach zero, essentially by applying the weak law of large numbers. Notice that

[35] Notice that the volatility risk premium equals the product of the volatility function $\boldsymbol{\Sigma\Lambda}_t$ of the SDE of the factors and the market price of volatility risk $\boldsymbol{\gamma}_{x,t}$. This is similar to the risk premium on the stock, $\sigma_t \gamma_{S,t}$, which equals the product of the volatility function σ_t of the stock return dS_t / S_t, and the market price of stock risk $\gamma_{S,t}$.

this set-up has not been imposed the absence of arbitrage strategies yet. Assuming the existence of a risk-free asset earning a return of r_f, Ross shows that in markets in which a large number of assets trades and that are free of arbitrage opportunities, for most assets it must hold that

$$\mathbb{E}_{\mathbb{P}}\left[r_i\right] = r_f + \mathbf{b}_i{}'\boldsymbol{\lambda}, \tag{5.7}$$

in which $\boldsymbol{\lambda}$ is a vector of *factor risk premia*. Hence, like in the CAPM, from (5.7) it is clear that idiosyncratic risk is not compensated for: $\text{var}_{\mathbb{P}}[e_i]$ does not enter (5.7). Relation (5.7) holds for all but possibly a small number of assets which may be mispriced; i.e. earn an expected return not commensurate with risk. For well-diversified portfolios the relation is always exact.

Although the APT does not specify the number or nature of the abstract factors, the factor risk premia can in theory be proxied for using the concept of *factor-mimicking portfolios*. Suppose a change in the value of portfolio j is only affected by a change in factor j, such that $b_{jk} = 0$ for all factors $k \neq j$. Rearranging from (5.7), factor j's risk premium is thus equal to $\lambda_j = (\mathbb{E}_{\mathbb{P}}[r_j] - r_f)/b_{jj}$; i.e. the risk premium on factor j's *mimicking portfolio* per unit of risk as measured by the factor sensitivity. (Note that a logical consequence is that if portfolios j and J both mimic factor j, then it must hold that $(\mathbb{E}_{\mathbb{P}}[r_j] - r_f)/b_{jj} = (\mathbb{E}_{\mathbb{P}}[r_J] - r_f)/b_{Jj}$, otherwise there is arbitrage possible. [36]) Identifying these factor-mimicking portfolios one by one results in asset i's (with $i \neq j$) expected return being rewritten in terms of the risk premia on the factor-mimicking portfolios as

$$\mathbb{E}_{\mathbb{P}}\left[r_i\right] = r_f + \sum_j \frac{b_{ij}}{b_{jj}}\left(\mathbb{E}_{\mathbb{P}}\left[r_j\right] - r_f\right). \tag{5.8}$$

Hence, in a sense, it is the factor-mimicking portfolios that essentially "span" the space of assets in this setting.

Comparison

So how does the 1-period APT setting compare to our continuous-time derivative pricing setting (and derivative pricing models in general)? First, regard the interval of infinitesimal length $[t, t+dt]$ as "one period". Second, in our model asset returns are driven by two "common factors". The first factor f_1 is the shock $\sigma_t dW_{S,t}$ in the stock return dS_t / S_t. The second factor f_2 is the unpredictable shock $\boldsymbol{\Sigma\Lambda}_t d\mathbf{W}_{x,t}$ in a stock-volatility change, governed by $d\mathbf{x}_t$. Notice that our model does not know "idiosyncratic risks e_i", i.e., asset-specific risk. Instead, all asset uncertainty is fully governed by the underlying common factors. The derivative betas (implicitly) measure the sensitivity of an asset's return to the common factors. These betas clearly serve as the "factor loadings" b_{i1} and b_{i2}.

Consider now the asset returns in our market. Both "factor loadings" of the bond return $dB_t / B_t = r_t dt$ are zero, whereas it's a_i-term from (5.6) equals $r_t dt$. For the reinvestment portfolio return we have $dS_t^r / S_t^r = (\mu_t + q_t)dt + \sigma_t dW_{S,t}$. Hence, its a_i-term from (5.6) equals $(\mu_t + q_t)dt$, it has a loading of 1 to the first factor $f_1 = \sigma_t dW_{S,t}$, and a zero-loading with respect to the second factor $f_2 = \boldsymbol{\Sigma\Lambda}_t d\mathbf{W}_{x,t}$.

[36] Notice the analogy with a continuous-time setting. For example, suppose two assets S_1 and S_2 are driven by the same risky $\mathbb{P}$-Brownian motion W and evolve according to $dS_{it} = \mu_i S_{it} dt + \sigma_i S_{it} dW_t$ for $i = 1, 2$. Imposing no-arbitrage in this case results in the equal Sharpe-ratio restriction $(\mu_1 - r)/\sigma_1 = (\mu_2 - r)/\sigma_2 := \gamma$ where γ is the market price of W – risk and r is the risk-free rate. This is basically an alternative way of stating that the relative prices of all tradable assets should be martingales under the *same* measure $\mathbb{Q}$ in the absence of arbitrage.

Consider finally the return on a derivative with price $F_t = F(t, S_t, \mathbf{x}_t)$. From Itô's lemma, $dF_t / F_t = (..)/F_t dt + \beta_{F,S,t}(\sigma_t dW_{S,t}) + \boldsymbol{\beta}_{F,x,t}{}'(\boldsymbol{\Sigma\Lambda}_t d\mathbf{W}_{x,t})$ under $\mathbb{P}$, where the expression for the dots is as in (3.1). The analogy with the APT relation (5.6) will be clear; notice once more that there is no derivative-specific risk e_i.

Considered in the interval $[t, t+dt]$, our model setting thus exactly fits into the 1-period APT-setting. Therefore, imposing the no-arbitrage condition in our market should lead to an expected-return expression similar to what the APT predicts; i.e., an expression like (5.7). Moreover, as there is no asset-specific risk e_i [37], we expect this expression to hold with equality for all assets. And indeed, this is precisely where we arrived at in the previous section in (5.4) and (5.5). The factor risk premia are given by $\sigma_t \gamma_{S,t}$ and $\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t}$. The analogy with the classical APT is evident. Hence indeed, we could have expected an expected return – beta relationship as (5.5) to hold in our market. It is really the condition of no-arbitrage that is driving this result.

Mimicking the volatility risk premium

Similar to the factor-mimicking portfolios of the APT, we can rewrite the unobserved volatility risk premium in terms of the risk premia on n traded derivatives G_i, $i = 1,..,n$, with prices $G_{it} = G_i(t, S_t, \mathbf{x}_t)$. Consider some derivative F with price $F_t = F(t, S_t, \mathbf{x}_t)$. Assume the smallest maturity among the G_i's to be at least as large as F's maturity. From (5.5), G_i's spot risk premium is given by

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dG_{it}}{G_{it}} - r_t dt \mid \mathcal{F}_t\right] = \left[\beta_{i,S,t}\left(\sigma_t \gamma_{S,t}\right) + \boldsymbol{\beta}_{i,x,t}{}'\left(\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t}\right)\right] dt, \qquad (5.9)$$

with $\beta_{i,S,t}$ and $\boldsymbol{\beta}_{i,x,t}$ the stock and volatility beta of G_i respectively. Stacking these n risk premia in an $(nx1)$ vector we obtain

$$\mathbb{E}_{\mathbb{P}}\left[d\mathbf{G}_t./\mathbf{G}_t - r_t \mathbf{1} dt \mid \mathcal{F}_t\right] = \left[\boldsymbol{\beta}_{\mathbf{G},S,t}\left(\sigma_t \gamma_{S,t}\right) + \boldsymbol{\beta}_{\mathbf{G},x,t}\left(\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t}\right)\right] dt, \qquad (5.10)$$

with $(nx1)\,\boldsymbol{\beta}_{\mathbf{G},S,t} \equiv (\beta_{1,S,t},..,\beta_{n,S,t})'$, $(nxn)\,\boldsymbol{\beta}_{\mathbf{G},x,t}{}' \equiv [\boldsymbol{\beta}_{1,x,t},..,\boldsymbol{\beta}_{n,x,t}]$, and where we use the notation $d\mathbf{G}_t./\mathbf{G}_t \equiv (dG_{1t}/G_{1t},..,dG_{nt}/G_{nt})'$ and $(nx1)\,\mathbf{1}$ is a vector of ones. From (5.4), $\sigma_t \gamma_{S,t} dt$ is the spot risk premium on the reinvestment portfolio S^r. The volatility risk premium can then be written as

$$\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t} = \boldsymbol{\beta}_{\mathbf{G},x,t}^{-1}\left(\mathbb{E}_{\mathbb{P}}\left[d\mathbf{G}_t./\mathbf{G}_t - r_t \mathbf{1} dt \mid \mathcal{F}_t\right] - \boldsymbol{\beta}_{\mathbf{G},S,t}\,\mathbb{E}_{\mathbb{P}}\left[\frac{dS_t^r}{S_t^r} - r_t dt \mid \mathcal{F}_t\right]\right) / dt, \qquad (5.11)$$

provided that the inverse of the matrix $(nxn)\,\boldsymbol{\beta}_{\mathbf{G},x,t}$ exists. The sensitivity vectors $\boldsymbol{\beta}_{i,x,t}$ (and hence the vegas of the G_i's) thus ought to be linear independent. For example, the set of G_i's is not allowed to contain derivatives that mature at the same date and which payoff is a linear combination of the payoffs of other G_i's. We thus need "*linear-independent*" derivatives. We encountered this condition earlier in section 4.2 on hedging. (Indeed, this is essentially a *rank condition* for *completeness* of the market.)

---

[37] This is clearly common to all derivative pricing models: The link between the underlying factors and the derivative is exact; there is *no room for "pricing error"*. Thinking about stock and option trading in practice, this is arguably both a *weak* point (as models are only reflections of reality) and a *strong* point (as we can quote an exact model-based price) of derivative pricing models at the same time. For parameter-estimation purposes it seems best to introduce some source of error between model-based derivative prices and observed real-world prices. This is exactly what we will do in the next chapter.

Having rewritten the unobserved volatility risk premium in terms of the risk premia on observed assets G_i and S^r, derivative F's spot return (5.5) becomes

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = r_t dt + \left(\beta_{F,S,t} - \boldsymbol{\beta}_{F,x,t}{}' \boldsymbol{\beta}_{\mathbf{G},x,t}^{-1} \boldsymbol{\beta}_{\mathbf{G},S,t}\right) \mathbb{E}_{\mathbb{P}}\left[\frac{dS_t^r}{S_t^r} - r_t dt \mid \mathcal{F}_t\right]$$
$$+ \boldsymbol{\beta}_{F,x,t}{}' \boldsymbol{\beta}_{\mathbf{G},x,t}^{-1} \mathbb{E}_{\mathbb{P}}\left[d\mathbf{G}_t./\mathbf{G}_t - r_t \mathbf{1} dt \mid \mathcal{F}_t\right]. \quad (5.12)$$

The analogy with the concept of factor-mimicking portfolios and equation (5.8) is apparent. Together the bond, reinvestment portfolio and derivatives G_i *span* the return on any arbitrary derivative.

5.4 Derivative returns and the hedging strategy

The expected spot return (5.5) on a general derivative F with price $F_t = F(t, S_t, \mathbf{x}_t)$ can alternatively be derived from its associated hedging strategy. Again, it is the arbitrage argument that fuels the derivation, and again it shows up in a somewhat different disguise, yielding additional insight. For simplicity we consider the 1-factor SV special case of our model.

As in section 4.3.1, we consider hedging F with a dynamic trading strategy in the bond B, the reinvestment portfolio S^r, and a longer-maturity derivative G. The hedging strategy $(\psi_t, \phi_t^r, \xi_t), t \in [0, T_F]$, is given by equations (4.9)-(4.11). As we can perfectly replicate F's payoff and as the strategy is self-financing, it holds that $F_t = \psi_t B_t + \phi_t^r S_t^r + \xi_t G_t$ by no-arbitrage, and $dF_t = \psi_t dB_t + \phi_t^r dS_t^r + \xi_t dG_t$. Dividing this SDE by F_t and rearranging, we obtain

$$\frac{dF_t}{F_t} = \left(\frac{\psi_t B_t}{F_t}\right)\frac{dB_t}{B_t} + \left(\frac{\phi_t^r S_t^r}{F_t}\right)\frac{dS_t^r}{S_t^r} + \left(\frac{\xi_t G_t}{F_t}\right)\frac{dG_t}{G_t}. \quad (5.13)$$

This equation tells us that the spot derivative return is a replicating-portfolio-weighted average of the spot returns on its composing instruments. Recalling that $dB_t / B_t = r_t dt$, rewriting in terms of excess returns, and using $\psi_t B_t + \phi_t^r S_t^r + \xi_t G_t = F_t$, we get

$$\frac{dF_t}{F_t} = r_t dt + \left(\frac{\phi_t^r S_t^r}{F_t}\right)\left[\frac{dS_t^r}{S_t^r} - r_t dt\right] + \left(\frac{\xi_t G_t}{F_t}\right)\left[\frac{dG_t}{G_t} - r_t dt\right]. \quad (5.14)$$

If we next take conditional expectations, the expected derivative return becomes

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = r_t dt + \left(\frac{\phi_t^r S_t^r}{F_t}\right)\mathbb{E}_{\mathbb{P}}\left[\frac{dS_t^r}{S_t^r} - r_t dt \mid \mathcal{F}_t\right] + \left(\frac{\xi_t G_t}{F_t}\right)\mathbb{E}_{\mathbb{P}}\left[\frac{dG_t}{G_t} - r_t dt \mid \mathcal{F}_t\right]. \quad (5.15)$$

Derivative F's spot risk premium is thus a replicating-portfolio-weighted average of the spot risk premia on the hedging instruments, being the reinvestment portfolio and the "*volatility-mimicking"* derivative G. This result is something that was not at all apparent from either (5.5) or (5.12). Taking the explicit hedging strategy (4.9)-(4.11) into account and the definitions of delta, vega and the stock and volatility betas, it requires some algebra to show that the portfolio weights in equation (5.15) are equal to $\phi_t^r S_t^r / F_t = \beta_{F,S,t} - \beta_{F,x,t}\beta_{G,S,t} / \beta_{G,x,t}$ and $\xi_t G_t / F_t = \beta_{F,x,t} / \beta_{G,x,t}$. Hence, equation (5.15) can be rewritten as the 1-factor version of (5.12), as it should be of course. The link is obvious.

5.5 Asset returns and the stochastic discount factor

It is interesting to go one step further and explore the relation between expected derivative returns and the stochastic discount factor (SDF). Its importance will become apparent in section 7, which proposes a theory for volatility-risk premia.

Consider the no-arbitrage pricing relation (3.10) that holds for all tradable assets,

$$M_t F_t = \mathbb{E}_{\mathbb{P}}\left[M_T F_T \mid \mathcal{F}_t\right]. \tag{5.16}$$

As the process $\{M_t F_t; t \geq 0\}$ is a 1-dimensional $(\mathbb{P},\{\mathcal{F}_t\})-$martingale, the *martingale representation theorem* tells us that

$$M_t F_t = M_0 F_0 + \int_0^t \boldsymbol{\xi}_u' d\mathbf{W}_u \Leftrightarrow d(M_t F_t) = \boldsymbol{\xi}_t' d\mathbf{W}_t , \qquad (\mathbb{P}) \tag{5.17}$$

for some $(n+1)-$dimensional $\mathcal{F}_t-$adapted process $\{\boldsymbol{\xi}_t; t \geq 0\}$. [38] It follows that

$$\mathbb{E}_{\mathbb{P}}\left[d(M_t F_t) \mid \mathcal{F}_t\right] = 0. \tag{5.18}$$

The $\mathbb{P}$-expected increment in the martingale given current information thus equals 0. It is essentially the infinitesimal analogue of the martingale property (5.16). [39] As both the derivative price and the stochastic discount factor follow Itô processes, integration by parts gives us

$$d(M_t F_t) = M_t dF_t + F_t dM_t + (dM_t)(dF_t). \tag{5.19}$$

Substitution in (5.18), dividing by $M_t F_t$ (which is allowed as no-arbitrage yields a strictly positive stochastic discount factor, and $F_t \neq 0$ as long as the derivative does not pay off zero in all states of the world), and rearranging yields

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = -\mathbb{E}_{\mathbb{P}}\left[\frac{dM_t}{M_t} \mid \mathcal{F}_t\right] - \mathbb{E}_{\mathbb{P}}\left[\frac{dM_t}{M_t}\frac{dF_t}{F_t} \mid \mathcal{F}_t\right]. \tag{5.20}$$

To further simplify, notice first that the latter expectation equals the conditional covariance between dM_t / M_t and dF_t / F_t, as both the conditional mean of dM_t / M_t and dF_t / F_t are of order dt, such that their product equals zero. Second, as (5.18) holds for all tradable assets, (5.20) holds for the bond as well. As the bond obeys $dB_t / B_t = r_t dt$ and does not covary with dM_t / M_t, it follows

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dM_t}{M_t} \mid \mathcal{F}_t\right] = -r_t dt . \tag{5.21}$$

Equation (5.20) can thus be rewritten as

[38] Recall that the MRT is not constructive: it only tells us that *some* process $\{\boldsymbol{\xi}_t; t \geq 0\}$ exists such that (5.17) holds, without giving guidance how to find it. In this case it is possible to find the process $\{\boldsymbol{\xi}_t; t \geq 0\}$ however, as we have explicit expressions for the SDF M_t and for dF_t.

[39] Notice the analogy with the risk-neutral arbitrage-free pricing relation $F_t / B_t = \mathbb{E}_{\mathbb{Q}}[F_T / B_T \mid \mathcal{F}_t]$, which is dictated by The Fundamental Theorem of Asset Pricing. Alternatively stated it thus says $0 = \mathbb{E}_{\mathbb{Q}}[d(B_t^{-1} F_t) \mid \mathcal{F}_t]$. As $F_t = F(t, S_t, \mathbf{x}_t)$ follows an Itô process, integration by parts yields $d(B_t^{-1} F_t) = B_t^{-1} dF_t - r_t B_t^{-1} F_t dt + (dB_t^{-1})(dF_t)$ with $(dB_t^{-1})(dF_t) = 0$. Substitution followed by rearranging yields $\mathbb{E}_{\mathbb{Q}}[dF_t / F_t \mid \mathcal{F}_t] = r_t dt$. Indeed, this is a quick way of showing that no-arbitrage requires the expected growth rate of all tradable assets under $\mathbb{Q}$ to equal the risk-free bond rate.

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = r_t dt - \text{cov}_{\mathbb{P}}\left[\frac{dM_t}{M_t}, \frac{dF_t}{F_t} \mid \mathcal{F}_t\right]. \tag{5.22}$$

This equation tells us that, apart from the risk-free rate as compensation for delay in expenditure, the expected derivative return is determined by the extent in which the derivative return covaries with the "return" on the SDF, dM_t / M_t. A negative (positive) correlation implies an expected derivative return above (below) the risk-free rate. Notice that no matter how large the "derivative-specific" risk $\text{var}_{\mathbb{P}}[dF_t / F_t \mid \mathcal{F}_t]$, it does not influence the derivative return. Similar results appear in different model settings and are well known in the literature (see e.g. Cochrane (2001)). The common underlying force that leads to this result is evidently the condition of no arbitrage. We refer to section 7 for further discussion.

Obviously, equation (5.22) must lead to expression (5.5), which also resulted from imposing the absence of arbitrage. From expression (3.11) for the Radon-Nikodym process $\{L_t; t \geq 0\}$, we obtain that $d\ln L_t = -\boldsymbol{\gamma}_t' d\mathbf{W}_t - \frac{1}{2}\boldsymbol{\gamma}_t'\boldsymbol{\gamma}_t dt$. By Itô's formula, $dL_t = -\boldsymbol{\gamma}_t' L_t d\mathbf{W}_t$. This illustrates the well-known fact that the Radon-Nikodym process is a $(\mathbb{P}, \{\mathcal{F}_t\})-$martingale. Recall that the SDF is defined by $M_t = B_t^{-1} L_t$ such that $dM_t = B_t^{-1} dL_t + L_t dB_t^{-1} = B_t^{-1} dL_t - r_t B_t^{-1} L_t dt$, and hence

$$dM_t / M_t = -r_t dt - \boldsymbol{\gamma}_t' d\mathbf{W}_t = -r_t dt - \gamma_{S,t} dW_{S,t} - \boldsymbol{\gamma}_{x,t}' d\mathbf{W}_{x,t}. \tag{5.23}$$

Notice the apparent roles of the market prices of time, stock and volatility risk, i.e., r_t, $\gamma_{S,t}$ and $\boldsymbol{\gamma}_{x,t}$ in the process for the stochastic discount factor. The expression for dF_t / F_t is given in (3.1), which is of the form $dF_t / F_t =$ $(..)dt + \beta_{F,S,t}\sigma_t dW_{S,t} + \boldsymbol{\beta}_{F,x,t}' \boldsymbol{\Sigma\Lambda}_t d\mathbf{W}_{x,t}$. We find for the covariance

$$\text{cov}_{\mathbb{P}}\left[\frac{dM_t}{M_t}, \frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = -\left[\beta_{F,S,t}\left(\sigma_t \gamma_{S,t}\right) + \boldsymbol{\beta}_{F,x,t}'\left(\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t}\right)\right] dt. \tag{5.24}$$

In the absence of arbitrage equation (5.22) holds. We obtain (5.5) once again:

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = \left[r_t + \beta_{F,S,t}\left(\sigma_t \gamma_{S,t}\right) + \boldsymbol{\beta}_{F,x,t}'\left(\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t}\right)\right] dt. \tag{5.25}$$

6. Volatility risk and compensation for volatility risk

In this section volatility risk and compensation for volatility risk are central. The issue whether investors are compensated for bearing (market) volatility risk in practice has recently attracted much attention in the finance literature. Standard asset pricing theory (the APT) states that if volatility risk is *systematic* –i.e. cannot be diversified away-, then investors should be compensated for it. As Bakshi and Kapadia (2003) point out in their first paragraph: "*However, a less than understood phenomenon is whether volatility risk is compensated, and whether this compensation is higher or lower than the risk-free rate.* [..]". Moreover, they end their article with "*Much more remains to be learned about how volatility risk is priced in financial markets*."

Section 6.1 reviews the assumptions made so far in the literature with regard to the market price of volatility risk. Before turning to the evidence, in section 6.2

we first consider a number of popular derivative strategies in which volatility risk plays a dominant role. We look at *straddles*, *variance swaps* and *delta hedging* a derivative. Then, in section 6.3, we provide a detailed overview of the empirical evidence on volatility-risk compensation. The evidence has only recently become available, and arrives from very different studies. Our empirical work to be discussed in later chapters further adds to this evidence.

6.1 Market-price-of-volatility-risk assumptions in the literature

To obtain unique derivative prices under SV formally requires an asset pricing model that integrates explicit investor attitude towards volatility risk. Aware of this, some of the earliest researchers on SV option pricing use general equilibrium arguments (based on Cox et al. (1985)) to be able to say more about the form of the market price of volatility risk. Examples are Wiggins (1987) and Chesney and Scott (1989). However, only for a few special (restrictive) types of investor preferences a closed-form expression for the price of volatility risk can be derived.

Given the complexity of such a model, others in the earliest stream simply assume a zero correlation between volatility and aggregate consumption. Volatility risk can then be diversified away, and hence should not be priced in equilibrium. Examples of this approach include Hull and White (1987), Johnson and Shanno (1987), and Scott (1987). Subsequent researchers (e.g. Melino and Turnbull (1990) and Stein and Stein (1991)) assume the market price of volatility risk to be constant. This can be justified if an investor derives logarithmic utility from consumption. Heston (1993) (implicitly) models the price of volatility risk as being proportional to the square root of the stock variance. His choice is motivated by a combination of Breeden's (1979) consumption-based asset pricing model, and the consumption process that emerges in the general equilibrium model of Cox et al. (1985). The resulting $\mathbb{Q}$-volatility process is still CIR.

The more recent literature has more or less side-stepped from such direct motivations. Researchers currently seem to favor tractable functional forms for the market price of volatility risk that depend on extra (free) parameters, which can be estimated from option prices. The price of volatility risk is then essentially only an *indirect reflection* of investor preferences. These functional forms are often chosen rather ad hoc (e.g., Jones (2003)). For instance, it is nowadays common to assume that the volatility process under the market and risk-neutral measures is of the same type. [40] Examples include Chernov and Ghysels (2000), Pan (2002), Jones (2003), and our model specification.

6.2 Investment strategies dominated by volatility risk

In this section we look at investment strategies in which volatility risk is of paramount importance. The idea is that if derivatives incorporate a volatility risk premium in practice, then its existence should be apparent from an investment strategy that has hedged all, but volatility risk. We consider *straddles*, *variance swaps* and *delta-hedging* a general derivative. These strategies are commonly employed in practice.

---

[40] This is also common practice in the term-structure-of-interest-rates literature for the short interest-rate specification under $\mathbb{P}$ and $\mathbb{Q}$; see e.g. Dai and Singleton (2000) and de Jong (2000).

6.2.1 Straddles: bets on volatility

A long *straddle* (Str) is a derivative asset consisting of a portfolio of a European call C and put P written on the same stock S, with equal strike K and remaining maturity $\tau_t = T - t$, where $T > t$ is the expiration date of the options. By no-arbitrage, the time-t straddle price, delta and vega equal

$$Str_t = C_t + P_t\,, \quad \Delta_{Str,t} = 2\Delta_{C,t} - \exp(-\bar{q}_t \tau_t)\,, \quad \mathcal{V}_{Str,t} = 2\mathcal{V}_{C,t}\,, \qquad (6.1)$$

with $\Delta_{C,t}$ and $\mathcal{V}_{C,t}$ the call delta and vega, and where we exploited the put-call parity (3.25). (Notice in particular that the vega of a call and put option coincide.)

Now suppose for the moment that we live in a world without stochastic volatility, such that $dS_t = \mu_t S_t dt + \sigma S_t dW_{S,t}$ - the world of Black-Scholes (BS). The BS call price is given by (3.23) and (3.24), but with $\bar{\sigma}_t$ replaced by the parameter σ. The BS call delta equals $\exp(-\bar{q}_t\tau_t)\Phi(d_{1t})$. The BS call vega –which is defined as $\partial BS_t / \partial\sigma$ - is given by $\exp(-\bar{q}_t\tau_t)S_t\sqrt{\tau_t}\,\phi(d_{1t})$, where $\phi(.)$ is the standard normal density. If the straddle is currently *at-the-money-forward* (ATMF) such that $F_t = K$, it holds that $d_{1t} = \frac{1}{2}\sigma\sqrt{\tau_t} \approx 0$, especially for short-maturity options. As $\Phi(0) = \frac{1}{2}$ and $\phi(.)$ reaches its maximum for $d_{1t} = 0$, it is clear that an ATMF-straddle position is virtually delta neutral, whereas its vega is maximal. In other words, an ATMF straddle is (instantaneously) insensitive to stock price fluctuations, whereas it is maximally perceptive to volatility changes. This analysis explains why ATMF straddles are typically being referred to as *bets on volatility* in practice, and allow investors *to trade volatility risk*. It is a clear example of a (spot) *delta-neutral positive-vega* strategy. An investor will buy such a straddle if he –different from other market participants- expects big swings in the stock price, and will write one if he expects tranquillity.

In our world with stochastic volatility, the straddle delta is given by $\exp(-\bar{q}_t\tau_t)\,\mathbb{E}_{\mathbb{Q}}[2\Phi(d_{1t}) - 1 \mid \mathcal{F}_t]$. If the straddle is currently ATMF, then it still holds that $d_{1t} = \frac{1}{2}\bar{\sigma}_t\sqrt{\tau_t} \approx 0$ by approximation, for whatever empirically reasonable value the random variable $\bar{\sigma}_t$ assumes, and in particular for short-maturity options. So an ATMF straddle is still delta neutral by approximation. Moreover, Hull (2003) mentions that the vega of an option calculated from a (1-factor) stochastic volatility model is very similar in magnitude to the Black-Scholes vega. [41] Hence, also in our world an ATMF straddle is particularly sensitive to volatility movements.

Given that its vega is maximal, its delta (and thus stock beta) is approximately zero, an ATMF-straddle position seems ultimately suited for enlarging our understanding of the volatility risk premium. From (5.5), the ATMF-straddle return is

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dStr_t}{Str_t} \mid \mathcal{F}_t\right] \approx \left[r_t + \boldsymbol{\beta}_{Str,x,t}{}'\left(\boldsymbol{\Sigma}\boldsymbol{\Lambda}_t^2\boldsymbol{\gamma}\right)\right]dt\,. \qquad (6.2)$$

The risk premium on such a straddle is virtually only determined by the volatility risk premium. Notice that in case volatility risk is not priced, the expected

[41] In our world, our vega definition implies $\mathcal{V}_{C,t} = \partial\,\mathbb{E}_{\mathbb{Q}}[BS(F_t, K, \tau_t, \bar{r}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t] / \partial\mathbf{x}_t = \mathbb{E}_{\mathbb{Q}}[\partial BS(F_t, K, \tau_t, \bar{r}_t, \bar{\sigma}_t^2) / \partial\mathbf{x}_t \mid \mathcal{F}_t]$ which –it seems- can only be obtained by simulation.

straddle return equals the risk-free rate. [42] (However, realize that the ATMF straddle is only instantaneously delta neutral, and that we are considering its *spot* return.) The straddle volatility beta is given by $\boldsymbol{\beta}_{Str,x,t} = 2\mathcal{V}_{C,t} / Str_t$. If $F_t = K$ (ATMF), then $P_t = C_t$, and hence $\boldsymbol{\beta}_{Str,x,t} = \mathcal{V}_{C,t} / C_t = \boldsymbol{\beta}_{C,x,t} = \boldsymbol{\beta}_{P,x,t}$. For a numerical impression of these betas we refer to the empirical analyses in later chapters.

Special case: 1-factor SV model
If we specialize to the 1-factor SV special case with $\delta_0 = 0$ and $\delta_1 = 1$ such that $\sigma_t^2 = x_t$, $dx_t = k(\theta - x_t)dt + \sigma\sqrt{\alpha + \beta x_t}dW_{x,t}$, the ATMF-straddle return becomes

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dStr_t}{Str_t} \mid \mathcal{F}_t\right] \approx \left[r_t + \beta_{Str,x,t}\,\sigma\gamma(\alpha + \beta x_t)\right]dt, \tag{6.3}$$

in which $\sigma\gamma(\alpha + \beta x_t)$ is the volatility risk premium. As the volatility beta of an option and hence a straddle is positive, we observe that the risk premium on the straddle is negative (resp. positive) if γ is negative (resp. positive). The parameter γ determines the market price of volatility risk, which in the 1-factor case is given by $\gamma_{x,t} = \gamma\sqrt{\alpha + \beta x_t}$.

6.2.2 Variance swaps: trades on realized variance

A *volatility derivative* that has recently gained much popularity in financial markets is the so-called *variance swap*, *VS*. This instrument is nowadays actively traded over the counter. The theoretical payoff of a long variance swap equals a fixed dollar-multiple A of the difference between the average stock variance, $\bar{\sigma}^2$, realized over the life $[0,T]$ of the contract, and a fixed *swap rate*, *SW*. That is, $VS_T = A[\bar{\sigma}^2 - SW]$. To determine the payoff at maturity in practice, $\bar{\sigma}^2$ is estimated by the average *realized variance*. [43] This volatility derivative is clearly path-dependent and, moreover, of the *Asian* type. A variance swap allows investors to trade volatility (or more precisely, *quadratic variation*) directly, without being exposed to stock price risk. Other than a straddle, a variance swap is constantly immune to stock price fluctuations.

A variance swap is a mutual agreement between two parties to essentially exchange volatility. Therefore, as there are no costs associated with entering such a contract (hence $VS_0 \equiv 0$), the risk-neutral valuation formula (3.9) yields for the variance-swap rate $SW = \mathbb{E}_{\mathbb{Q}}[\bar{\sigma}^2 \mid \mathcal{F}_0]$. From time 0 onwards, the arbitrage-free contract value at time t, VS_t, can next be determined from (3.9) again:

$$\begin{aligned} VS_t &= \exp(-\bar{r}_t\tau_t)\mathbb{E}_{\mathbb{Q}}\left[VS_T \mid \mathcal{F}_t\right] \\ &= A\exp(-\bar{r}_t\tau_t)\left\{\mathbb{E}_{\mathbb{Q}}\left[\bar{\sigma}^2 \mid \mathcal{F}_t\right] - SW\right\} \\ &= A\exp(-\bar{r}_t\tau_t)\left\{\frac{1}{T}\left(I_t + \mathbb{E}_{\mathbb{Q}}\left[\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t\right]\right) - SW\right\}; \quad I_t \equiv \int_0^t \sigma_s^2 ds \\ &= VS(t, I_t, \mathbf{x}_t), \end{aligned} \tag{6.4}$$

[42] This is moreover exactly what the conventional Black-Scholes model predicts. In the BS world the price of a derivative for which $F_T = f(S_T)$ is given by $F_t = F(t, S_t)$, as there is no volatility risk. Following a similar analysis as before, it can be shown that $\mathbb{E}_{\mathbb{P}}\left[dF_t / F_t \mid \mathcal{F}_t\right] = \left[r_t + \beta_{F,S,t}\left(\sigma\gamma_{S,t}\right)\right]dt$ with $\gamma_{S,t} = (\mu_t + q_t - r_t)/\sigma$. An ATMF straddle is thus expected to earn the risk-free rate instantaneously.

[43] This average realized stock return variance is computed as the annualized average of the squared (log) stock returns, in which the sampling frequency of the returns is specified in the contract.

in which $\{I_t; 0 \le t \le T\}$ represents the integrated stock-variance process. The conditional expectation appearing after the third equality is a function of $\mathbf{x}_t$ only (and not of its past), as $\{\mathbf{x}_t\}$ follows a Markov diffusion. [44] The variance-swap value is therefore a function of t, $\mathbf{x}_t$ *and* I_t : $VS_t = VS(t, I_t, \mathbf{x}_t)$. [45]

In appendix IIc we show that the expected spot variance-swap return is given by

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dVS_t}{VS_t} \mid \mathcal{F}_t\right] = \left[r_t + \boldsymbol{\beta}_{VS,x,t}{}'\left(\boldsymbol{\Sigma\Lambda}_t^2\boldsymbol{\gamma}\right)\right]dt\,. \qquad (6.5)$$

Because of its permanent delta-neutrality, the return on this exotic product is fully determined by the volatility risk premium.

Special case: 1-factor SV model
As an illustration, let us consider the 1-factor SV special case of our model with $\sigma_t^2 = x_t$ (i.e., $\delta_0 = 0, \delta = 1$). Under $\mathbb{Q}$, the factor dynamics read

$$dx_t = \tilde{k}(\tilde{\theta} - x_t)dt + \sigma\sqrt{\alpha + \beta x_t}\,d\tilde{W}_{x,t}\,, \qquad (\mathbb{Q}) \qquad (6.6)$$

in which the risk-neutral parameters are given by $\tilde{k} = k + \sigma\gamma\beta$, $\tilde{\theta} = (k\theta - \sigma\gamma\alpha)/\tilde{k}$. Using equation (9.17) from appendix B, the swap rate equals

$$SW = \mathbb{E}_{\mathbb{Q}}\left[\bar{\sigma}^2 \mid \mathcal{F}_0\right] = \mathbb{E}_{\mathbb{Q}}\left[\frac{1}{T}\int_0^T \sigma_s^2 ds \mid \mathcal{F}_0\right] = \tilde{\theta} + \frac{1 - \exp(-\tilde{k}T)}{\tilde{k}T}\left(x_0 - \tilde{\theta}\right). \qquad (6.7)$$

Using (6.4) and this (9.17) again, we obtain for the variance-swap value:

$$\begin{aligned} VS_t &= A\exp(-\bar{r}_t\tau_t)\left\{\frac{1}{T}\left(I_t + \mathbb{E}_{\mathbb{Q}}\left[\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t\right]\right) - SW\right\} \\ &= A\exp(-\bar{r}_t\tau_t)\left\{\frac{1}{T}\left(I_t + \tilde{\theta}\tau_t + \frac{1 - \exp(-\tilde{k}\tau_t)}{\tilde{k}}\left[x_t - \tilde{\theta}\right]\right) - SW\right\}. \end{aligned} \qquad (6.8)$$

From (6.5), the expected spot variance-swap return equals

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dVS_t}{VS_t} \mid \mathcal{F}_t\right] = \left[r_t + \beta_{VS,x,t}\,\sigma\gamma(\alpha + \beta x_t)\right]dt\,, \qquad (6.9)$$

in which the swap's volatility beta is given by

$$\beta_{VS,x,t} = \frac{1}{VS_t}\frac{\partial VS_t}{\partial x_t} = \frac{A\exp(-\bar{r}_t\tau_t)}{VS_t}\left(\frac{1 - \exp(-\tilde{k}\tau_t)}{\tilde{k}T}\right). \qquad (6.10)$$

A more physical interpretation of the market price of volatility risk can now be given as follows. The spot variance-swap risk premium is given by

[44] See appendix B, equation (9.17) for an illustration.
[45] Notice that if it were the case that $VS_t = VS(t, \mathbf{x}_t)$, then the analysis from section 3.1 could directly be used to obtain an expression for the spot variance-swap return as in equation (3.6) (but leaving out the terms involving S_t). However, it does not hold that $VS_t = VS(t, \mathbf{x}_t)$; instead, $VS_t = VS(t, I_t, \mathbf{x}_t)$. Additional analysis is therefore required.

$\mathbb{E}_{\mathbb{P}}[\frac{dVS_t}{VS_t} \mid \mathcal{F}_t] - r_t dt = \beta_{VS,x,t}\sigma\sqrt{\alpha + \beta x_t}\,\gamma_{x,t}dt$ with $\gamma_{x,t} = \gamma\sqrt{\alpha + \beta x_t}$. From appendix IIc, (c.6), it follows that $\mathrm{var}_{\mathbb{P}}[\frac{dVS_t}{VS_t} \mid \mathcal{F}_t] = \sigma^2(\alpha + \beta x_t)\beta^2_{VS,x,t}dt$. Dividing yields

$$\frac{\mathbb{E}_{\mathbb{P}}[\frac{dVS_t}{VS_t} \mid \mathcal{F}_t] - r_t dt}{\sqrt{\mathrm{var}_{\mathbb{P}}[\frac{dVS_t}{VS_t} \mid \mathcal{F}_t]}} = \gamma_{x,t}\sqrt{dt} \,. \tag{6.11}$$

The market price of volatility risk thus essentially represents the instantaneous Sharpe ratio of a variance swap. Moreover, any asset's risk premium can be expressed as a linear combination of the risk premia on the dividend-paying stock and a variance swap: These instruments essentially span the universe of assets in the 1-factor SV market. (Notice once more the APT-analogy.)

As the swap's volatility beta (6.10) is positive (as long as $\tilde{k} > 0$ and $VS_t > 0$), it follows that the risk premium on a variance swap is negative (resp. positive) if the market price of volatility risk (and hence γ) is negative (resp. positive). If volatility risk is not priced, then the variance swap is expected to earn the risk-free rate.

6.2.3 Delta hedging

As a final example of investment strategies that are dominated by volatility risk, we consider *delta hedging* a derivative. Consider, at time 0, a general derivative F with maturity T and time-t price given by $F_t = F(t, S_t, \mathbf{x}_t)$. Similar to Bakshi and Kapadia (2003) [46], we define the *delta-hedged gains* over $[0,T]$, $\Pi_{0,T}$, as the gain or loss on the delta-hedged derivative position (with immediate reinvestment of dividends [47]), and where the net investment earns the risk-free rate:

$$\begin{aligned}\Pi_{0,T} &\equiv F_T - F_0 - \int_0^T \exp(-\bar{q}_{0t}t)\Delta_{F,t}dS_t^r - \int_0^T r_t\left[F_t - \exp(-\bar{q}_{0t}t)\Delta_{F,t}S_t^r\right]dt \\ &= F_T - F_0 - \int_0^T \Delta_{F,t}dS_t - \int_0^T\left[r_tF_t - (r_t - q_t)\Delta_{F,t}S_t\right]dt \,. \qquad (\mathbb{P})\end{aligned} \tag{6.12}$$

The second equality uses $S_t^r = \exp(\bar{q}_{0t}t)S_t$ such that $dS_t^r = \exp(\bar{q}_{0t}t)[dS_t + q_tS_tdt]$. Combining (3.1) with the no-arbitrage restriction (3.4), and using $\mu_t - \gamma_{S,t}\sigma_t = r_t - q_t$ (see (2.8)), we obtain for the gain on derivative F over the interval $[0,T]$:

$$F_T - F_0 = \int_0^T\left[r_tF_t - (r_t - q_t)S_t\Delta_{F,t}\right]dt + \int_0^T \Delta_{F,t}dS_t + \int_0^T \mathcal{V}_{F,t}{}'d\mathbf{x}_t - \int_0^T \mathcal{V}_{F,t}{}'\left[\mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t) - \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t\boldsymbol{\gamma}_{x,t}\right]dt \qquad (\mathbb{P}) \tag{6.13}$$

Substituting the SDE (2.6) for $\mathbf{x}$ results in the delta-hedged gains being equal to:

$$\Pi_{0,T} = \int_0^T \mathcal{V}_{F,t}{}'\boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t} + \int_0^T \mathcal{V}_{F,t}{}'\boldsymbol{\Sigma}\boldsymbol{\Lambda}_t\boldsymbol{\gamma}_{x,t}dt \,. \qquad (\mathbb{P}) \tag{6.14}$$

---

[46] These authors consider the delta-hedged gains from both continuous and discrete-time delta-hedging a call option written on a *non-dividend-paying* stock in a 1-factor SV model, and test its implications on S&P500 option data. See section 6.3.

[47] Due to dividends, we need to delta-hedge with the tradable reinvestment portfolio. Recall from section 4.3.1 (or appendix IIa), that starting out with 1 stock at time 0 followed by continuous dividend-reinvestment yields $\exp(\bar{q}_{0t}t)$ stocks at time t, such that $S_t^r = \exp(\bar{q}_{0t}t)S_t$.

There are a number of things to learn from this expression. First, this dynamic trading strategy hedges all stock price risk. It is therefore delta neutral at all times. Second, notice that if the value of derivative F does not respond to volatility changes such that its vega is persistently zero, then the delta-hedged gains are zero. In that case the derivative price satisfies $F_t = F(t, S_t)$. Moreover, the derivative payoff can then perfectly be replicated by the bond-stock trading strategy just outlined, which is self-financing in this case. An obvious example of such a derivative is a forward contract. Third, if derivative F *is* perceptive to volatility movements such that $F_t = F(t, S_t, \mathbf{x}_t)$, then the delta-hedged gains are non-zero. This illustrates once again that a perfect hedge is not possible with a bond-stock strategy only; such a strategy results in a *tracking error*, the delta-hedged gains. [48] These derivatives are essentially why we label our market as incomplete. As explained in section 4.2, we need to add other derivatives to our set of hedge instruments for perfect replication. Fourth and finally, the expected delta-hedged gains are given by

$$\mathbb{E}_{\mathbb{P}}\left[\Pi_{0,T} \mid \mathcal{F}_0\right] = \int_0^T \mathbb{E}_{\mathbb{P}}\left[\mathcal{V}_{F,t}{}' \boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t} \mid \mathcal{F}_0\right] dt . \tag{6.15}$$

The mean gains are essentially determined by the product of the premium for volatility risk (i.e. $\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t}$ which equals $\boldsymbol{\Sigma\Lambda}_t^2 \boldsymbol{\gamma}$), and the amount of volatility risk inherent in the derivative, as measured by its vega. Notice that if volatility risk is not priced, then the delta-hedged derivative position is expected to break even. If the volatility risk premium is negative (resp. positive), and the vega of the derivative is positive (resp. negative), then this delta-hedging strategy is expected to underperform (resp. overperform) zero.

6.3 Empirical evidence on volatility-risk compensation

The previous section has highlighted the role volatility risk plays in various investment strategies. It has moreover indicated how investors should be compensated for this risk source if indeed (market) volatility risk is systematic. But is volatility risk systematic?

As Bakshi and Kapadia (2003) briefly point out in their introduction, evidence that market volatility risk is systematic can be motivated by three empirical findings. First, the *hedging motive*. Long option positions like calls and puts are hedges against large downward market movements. As volatility tends to increase as markets go down, such positions increase in value and thus provide a hedge. An interpretation is that investors are willing to pay a premium for this downside protection. The hedging motive is indicative of a *negative* volatility risk premium. Second, ATM Black-Scholes *implied* volatilities (which are typically considered market forecasts of future volatility) are consistently found to be larger than subsequent *realized* volatilities; see e.g. Jackwerth and Rubinstein (1996) and Christensen and Prabhala (1998). A possible explanation is that the market price of volatility risk is negative. Ceteris paribus, this leads to an increase in the drift of the volatility process under $\mathbb{Q}$ (as compared to $\mathbb{P}$), hence in option prices and thus BS implied volatilities. Third, empirically observed stock-index options are found to be *non-redundant* securities; see e.g. Buraschi and Jackwerth (2001) and Coval and Shumway (2001). They are redundant in the Black-Scholes world

[48] In this case, the strategy is also clearly not self-financing which should be apparent from reconsidering equation (6.12).

(or in "deterministic" volatility models in which volatility depends on the asset price but not on a separate Brownian motion), but not in a world with stochastic volatility, as we have seen in section 4.3.1. Index-option models omitting the economic impact of a volatility risk premium may therefore be inconsistent with observed option prices.

The empirical evidence on volatility-risk compensation seems to arrive from two broad directions. More indirect evidence comes from studies that integrate both time-series data on the underlying and an option series in an estimation strategy for SV option pricing models. Important examples are Chernov and Ghysels (2000), Pan (2002) and Jones (2003). Other evidence comes from studies that either test implications of the conventional BS model (or "deterministic" volatility models), or look at investment strategies in which volatility risk plays an important role. Examples include Coval and Shumway (2001), Buraschi and Jackworth (2001), Bakshi and Kapadia (2003), Driessen and Maenhout (2003) and Carr and Wu (2004).

No consensus on a volatility-risk-premium definition
Before we start discussing the evidence, it should be noted that there does not seem to exist a clear consensus on a definition for the volatility risk premium. Most authors refer to a divergence between the measures $\mathbb{P}$ to $\mathbb{Q}$. They mention that this is indicative of volatility risk being priced -and thus that investors require a premium for volatility risk-, but without explicitly defining this premium. Our definition (recall section 5.2) seems attractive in the sense that it has a clear interpretation in terms of a derivative's expected return; e.g., straddles and variance swaps. Moreover, having estimated "our" volatility risk premium, this premium may next be used to estimate returns on other, arbitrary derivatives.

Because of this lack of consensus, and as the evidence on volatility-risk compensation is so recent, we find a rather detailed discussion both interesting and useful. Moreover, it motivates our economic theory on volatility-risk compensation, as will be developed in section 7. Whenever possible we provide a link with the definition maintained by us. This allows for a comparison with our own empirical results to be discussed in later chapters.

6.3.1 Discussion of the evidence

For convenient reference below, consider the following 1-factor SV special case of our model with $\sigma_t^2 = x_t$ and $\alpha = 0,\ \beta = 1$. The factor dynamics become

$$dx_t = k\left(\theta - x_t\right)dt + \sigma\sqrt{x_t}dW_{x,t} \qquad (\mathbb{P}) \qquad (6.16)$$

$$\begin{aligned} dx_t &= \left[k(\theta - x_t) - \sigma\gamma x_t\right]dt + \sigma\sqrt{x_t}d\tilde{W}_{x,t} \qquad (\mathbb{Q}) \qquad (6.17) \\ &= (k + \sigma\gamma)\left[\frac{k\theta}{k + \sigma\gamma} - x_t\right]dt + \sigma\sqrt{x_t}d\tilde{W}_{x,t} \\ &= \tilde{k}\left(\tilde{\theta} - x_t\right)dt + \sigma\sqrt{x_t}d\tilde{W}_{x,t}\,. \end{aligned}$$

The risk-neutral speed-of-adjustment coefficient equals $\tilde{k} = k + \sigma\gamma$, whereas the risk-neutral mean of the variance process equals $\tilde{\theta} = k\theta / \tilde{k}$. The market price of volatility risk is given by $\gamma_{x,t} = \gamma\sqrt{x_t}$. The volatility risk premium equals $\sigma\gamma x_t$, and averages at $\sigma\gamma\theta$. This particular 1-factor SV model essentially coincides with the Heston (1993) model, but without leverage effect.

Chernov and Ghysels (2000)
Chernov and Ghysels (2000) extend the *efficient method of moments* (EMM, Gallant and Tauchen (1996)) to estimate the Heston (1993) model, using both daily return and option data on the S&P500 index over the period 1986 – 1994. Their main concern is the estimation method. (See section 2.1.1 of the next chapter for more details). They do not explicitly comment on the estimated market-price-of-volatility-risk process, nor mention the volatility risk premium.

Chernov and Ghysels find $k = 0.931$, $\theta = 0.0154$, $\sigma = 0.0615$, $\tilde{k} = 0.690$ and $\tilde{\theta} = 0.00956$. From these estimates we compute the market price of volatility risk-determining parameter γ as $\gamma = (\tilde{k} - k)/\sigma = -3.92$, which indicates a *negative* average volatility risk premium of $\sigma\gamma\theta =$ -0.37% per annum.

Coval and Shumway (2001)
Coval and Shumway (2001) are the first to focus on the theoretical and empirical nature of option returns over an option's lifetime, in the context of broader asset pricing theory than Black-Scholes or CAPM. Under the weak assumption of the existence of an SDF that prices all assets in the economy (which is implied by no-arbitrage; recall section 3.1), they prove two theorems that option returns should satisfy. Specifically, long calls are expected to earn a positive lifetime return above that of the underlying security, whereas long puts generate an expected lifetime return below the risk-free rate. Moreover, both call and put returns are increasing in the strike price. Coval and Shumway find that S&P100 and S&P500 index-option returns conform to these implications. They also find that empirical option returns are too low to be consistent with the BS and CAPM frameworks.

Recall from section 6.2.1 that the BS world predicts expected spot ATMF-straddle returns to equal the risk free rate, as these instruments are virtually delta neutral. Coval and Shumway find that market-neutral S&P500-index straddles earn returns that are significantly lower than the risk-free rate: -3% per week on average (period 1990-1995). Notice that, again, this corresponds again to a *negative* market volatility risk premium. They conclude that a natural interpretation for these results is that other risks than just market risk, and in particular stochastic volatility, is priced in the market.

Buraschi and Jackwerth (2001)
Buraschi and Jackwerth (2001) reach a similar conclusion by showing that S&P500-index options are non-redundant securities. They develop a statistical test for testing if options are needed for spanning of the pricing kernel. They reject the null hypothesis of redundancy (as implied by BS or "deterministic" volatility models), and conclude that this is indicative of additionally priced risk factors such as volatility, interest rate or jump risk.

Pan (2002)
Pan (2002) estimates the parameters of the Heston (1993) and Bates (2000) models [49] by her own developed *implied-state GMM* (IS-GMM) method. (We refer to section 2.1.1 of the next chapter for details on IS-GMM.) She uses weekly data on the S&P500 index (period 1989 – 1996) and a near-the-money short-dated option series for estimation.

Fitting the Heston model, Pan finds a significant *negative* volatility risk premium, although the model is rejected by the joint data. Remarkably, her results imply a

[49] The Bates (2000) model extends the Heston (1993) model to allow for jumps in stock prices.

risk-neutral volatility process that has both a negative mean and speed-of-adjustment coefficient, and is therefore explosive. [50] Specifically, Pan specifies the drift term of the Heston variance process $\{V_t\}$ under $\mathbb{Q}$ as $[\kappa_v(\bar{v} - V_t) + \eta^v V_t]dt$, and mentions that "*volatility risk is priced by the extra term $\eta^v V_t$, which is absent under $\mathbb{P}$.* [..]". In terms of our SV specification (6.16)-(6.17), Pan's specification implies a risk-neutral speed-of-adjustment coefficient of $\tilde{k} = \kappa_v - \eta^v$, which she estimates at –0.5. The corresponding risk-neutral mean equals $\tilde{\theta} = \kappa_v \bar{v} / (\kappa_v - \eta^v)$, and is estimated at –0.2. From (6.17), Pan's $-\eta^v V_t$ should be compared to our $\sigma\gamma x_t$ in our empirical work in later chapters.

Setting V_t to its long-run mean of $\bar{v}$ and using Pan's estimation results on the Heston model, implies an average volatility risk premium equal to $-\eta^v \bar{v}$, estimated at –(7.6)(0.0137) = -10.4% per annum for the S&P500 index over 1989-1996. The long-run variance is estimated at 0.0137, or 11.7% S&P500 volatility over 1989 - 1996. Notice that although Pan (2002) and Chernov and Ghysels (2000) both estimate the Heston model and virtually use the same dataset, their results dramatically differ on the implied magnitude of the volatility risk premium.

Jones (2003)

Jones (2003) also fits the Heston model to joint daily time series of S&P100-index returns and the VIX [51] for the period 1988-2000. He uses a Bayesian perspective. (A discussion of his estimation strategy is in section 2.1.1 of the next chapter.) Jones uses a form for the market price of volatility risk that ensures an identical type of volatility process under $\mathbb{P}$ and $\mathbb{Q}$. However, he does not explicitly comment on the volatility risk premium, let alone its magnitude implied by his estimation results. In particular, Jones models the drift term of the variance process under $\mathbb{P}$ as $[\alpha + \beta V_t]dt$. In our notation (6.16)-(6.17), this implies a long-run mean of $\theta = -\alpha / \beta$ and a speed-of-adjustment coefficient of $k = -\beta$. Under $\mathbb{Q}$ he models the drift as $[\alpha + \beta^* V_t]dt$, which implies a long-run mean of $\tilde{\theta} = -\alpha / \beta^*$ and a mean-reversion parameter of $\tilde{k} = -\beta^*$. Rewriting yields $[\alpha + \beta V_t + (\beta^* - \beta)V_t]dt$, which shows that the volatility risk premium is given by $-(\beta^* - \beta)V_t$.

Jones quotes his Heston-model estimates on a daily basis, assuming 264 days per annum. Replacing V_t by its long-run mean estimated at $-(2.41*10^{-6})/(-1.81*10^{-2})*264$ = 0.0351 per annum (implying a long-run S&P100 volatility level of 18.7% over 1988 – 2000), yields a *negative* average S&P100 volatility risk premium of $-264*(1.13*10^{-2}+1.81*10^{-2})*0.035$ = -27%. Jones's results on the Heston model further imply $\tilde{\theta} = -0.0563$ and $\tilde{k} = -2.98$, such that the volatility process is explosive under $\mathbb{Q}$.

Explosive risk-neutral volatility process?

Let us comment on the finding by both Pan (2002) and Jones (2003) of a risk-neutral volatility process (6.17) with negative $\tilde{k}$ and $\tilde{\theta}$. Imagine what this

[50] We concentrate on Pan's findings on the Heston model as this model is most comparable to ours. It should be noted however, that fitting the Bates model with a zero volatility risk premium but allowing for a jump risk premium, Pan finds a significant premium for jump risk. This model is not rejected by the data. Fitting the Bates model and allowing for both a volatility and jump risk premium, Pan finds that the jump risk premium dominates by far the volatility risk premium. We refer to sections 5.3.3 and 7.2.1 of respectively chapters IV and V for more comments on Pan's findings in light of our own results.

[51] The "market volatility index" VIX of the Chicago Board of Options Exchange used to represent an average of eight 30-days-to-maturity near-the-money BS implied volatilities from S&P100 index (OEX) options. In September 2003 the CBOE redefined the VIX; its (different) calculation is now based on S&P500 (SPX) index options. See http://www.cboe.com/micro/vix/faq.asp for details.

implies for out-of-sample call option pricing using this estimated process; recall formula (3.19). This requires simulated paths of $\{x_t\}$ under $\mathbb{Q}$, given an initial positive $\mathbb{P}$-value of x_t (i.e., the current stock variance $\sigma_t^2 = x_t$). A negative $\tilde{k}$ and $\tilde{\theta}$ imply $\mathbb{E}_{\mathbb{Q}}[dx_t \mid \mathcal{F}_t] = \tilde{k}(\tilde{\theta} - x_t)dt > 0$, such that the $\mathbb{Q}$-tendency of the simulated paths is to go upward, and ever faster increase. The volatility process under $\mathbb{Q}$ is thus explosive (and does not mean-revert), which will imply large, unrealistic option prices, especially when the option's maturity is large. (Another thing to worry about is, if the SDE has a properly defined solution at all.) A possible reason for this finding is the following. ATM BS implied volatilities (which are essentially "$\mathbb{Q}$-volatilities") are typically found to be larger than subsequent realized ("$\mathbb{P}$-") volatilities (e.g., Jackwerth and Rubinstein (1996) and Christensen and Prabhala (1998)). From (6.16)-(6.17), this translates into a negative γ (and thus a negative volatility risk premium, $\sigma \gamma x_t$)). This γ directly influences the risk-neutral parameter $\tilde{k}$, and hence $\tilde{\theta}$. If this difference between these $\mathbb{Q}$ and $\mathbb{P}$-volatilities in the data is too large, implying a large negative estimated γ, this results in a negative $\tilde{k}$ and thus $\tilde{\theta}$, and hence in an explosive risk-neutral volatility process. [52]

Bakshi and Kapadia (2003)
Bakshi and Kapadia (2003) consider the delta-hedged gains from both continuous-time and discrete-time delta-hedging a call option written on a non-dividend paying stock in a 1-factor SV model, where the net investment earns the risk-free rate. (See also section 6.2.3.) Testing their theoretical implications on daily S&P500 index-option data over the period 1988-1995, they find that the mean delta-hedged gains underperform zero, which is evidence of a *negative* market volatility risk premium (even after having accounted for jump fears).

Driessen and Maenhout (2003)
Driessen and Maenhout (2003) find evidence of a *negative* market volatility risk premium in US and UK markets, in the context of the standard asset allocation problem, extended with index options as an additional investment instrument. Given historical prices, they find that (non-) expected-utility investors act optimally when shorting puts and straddles, in order to exploit the premia for volatility and jump risk. Long put positions are never optimal. Given the well-known insurance motive for holding such positions in practice, this is a clear anomaly. In later chapters we consider the same FTSE100 index-option dataset as Driessen and Maenhout do. Their sample period covers March 1992 - Dec 2001. Of special interest to us is the summary statistic they provide on the average monthly empirical return of –13.1% (std.dev. 37%) on a short-maturity ATM-straddle strategy. Later we will contrast this empirical return with the expected return on such a strategy, implied by our model and estimation results.

Carr and Wu (2004)
Carr and Wu (2004) also find evidence of a *negative* volatility risk premium in US index markets. They use a "direct and robust" ([..]) method for quantifying this premium, based on the variance swap. Notice from section 6.2.2 that if volatility risk is not priced (such that the volatility processes under $\mathbb{P}$ and $\mathbb{Q}$ coincide), the variance swap rate equals $SW_{t,T} = \mathbb{E}_{\mathbb{P}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]$. (This assumes the contract is entered at time t and expires at T, such that its lifetime equals $\tau_t = T - t$.) Carr and Wu argue that the difference between $\mathbb{E}_{\mathbb{P}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]$ and $\mathbb{E}_{\mathbb{Q}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]$ therefore reflects the magnitude of the volatility risk premium. They next *define* the

[52] Indeed, the direct dependence between γ and $\tilde{k}$ (and thus $\tilde{\theta}$) is causing this. Duffee (2002) proposes the *essentially-affine* class of term-structure models to overcome this dependence.

volatility risk premium as such, i.e. by $\mathbb{E}_{\mathbb{P}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t] - \mathbb{E}_{\mathbb{Q}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]$. They show that it can be represented as the negative of the conditional $\mathbb{P}$-covariance between some general normalized pricing kernel and the average variance. They also show how to synthesize and approximate the swap rate, by a particular linear combination of option prices. By next comparing it to ex-post average realized variances, they find strong evidence of a negative volatility risk premium in S&P100, S&P500, DJIA, and NASDAQ100 stock index-markets over the period 1996-2003.

Let us compare Carr and Wu's definition of the volatility risk premium to ours in the 1-factor SV model (6.16)-(6.17). Carr and Wu's volatility risk premium is given by $\mathbb{E}_{\mathbb{P}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t] - \mathbb{E}_{\mathbb{Q}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]$, which (from appendix B, equation (9.17)) equals

$$\theta + \frac{1-\exp(-k\tau_t)}{k\tau_t}(x_t - \theta) - \left(\tilde{\theta} + \frac{1-\exp(-\tilde{k}\tau_t)}{\tilde{k}\tau_t}(x_t - \tilde{\theta})\right). \tag{6.18}$$

Their volatility risk premium averages at $\mathbb{E}_{\mathbb{P}}\left(\mathbb{E}_{\mathbb{P}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t] - \mathbb{E}_{\mathbb{Q}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]\right)$ which equals

$$\theta - \left(\tilde{\theta} + \frac{1-\exp(-\tilde{k}\tau_t)}{\tilde{k}\tau_t}(\theta - \tilde{\theta})\right) = \frac{\sigma\gamma\theta}{\tilde{k}}\left[1 - \frac{1-\exp(-\tilde{k}\tau_t)}{\tilde{k}\tau_t}\right]. \tag{6.19}$$

In contrast, our volatility risk premium is given by $\sigma\gamma x_t$, and averages at $\sigma\gamma\theta$. Although both definitions agree on the sign of the average volatility risk premium (as long as $\tilde{k} > 0$, the expression in [..] is in the interval [0,1]), they are clearly different and not directly comparable.

Summary
To summarize, recent empirical evidence supports the notion that market volatility risk is systematic. The evidence comes from various angles and very different studies. Investors seem indeed negatively be compensated for market volatility risk (even after having accounted for jump fears). Although a clear consensus on a definition does not exist, the associated volatility risk premium is found to be negative. (Hence, a long position in an ATM straddle yields a negative risk premium, a short position a positive risk premium.)

7. A theory on the negative volatility risk premium

This section provides an economic basis for the (likely) existence of a negative (market) volatility risk premium. To achieve this, we connect to both *consumption-based asset pricing theory* and the *permanent-income hypothesis* from macroeconomics. We show that there are theoretical indications that the volatility risk premium should be negative. The analysis in this section may be considered as a thorough theoretical justification of the *hedging motive* mentioned by Bakshi and Kapadia (2003). Given our analysis, we think the *consumption-insurance motive* is a somewhat more economically-justified name.

Our model as described in section 2 does not explicitly characterize investor preferences and investor behavior. What is the aim of an investor? How does he respond to volatility risk at which he is exposed? We only gave an indirect reflection of these issues in terms of an exogenously given market price of volatility risk. To use the model for pricing and hedging in practice, such a description is not necessary; we can do without. To derive an economic theory on the nature of volatility-risk compensation however, requires preference-based analysis. [53]

Section 7.1 characterizes an investor that considers investing in our market, and considers the problem he is facing. We derive a well-known formula that relates asset prices to investor's marginal utility of consumption. In section 7.2 we relate the investor's optimal consumption plan to derivative returns. The case in which the investor derives *power utility* from consumption serves as an illustration. Based on the permanent-income hypothesis, in section 7.3 we finally argue why the volatility risk premium should be negative.

7.1 The investor problem

Consider some infinitely-lived investor that enters our market at, say, time 0. At any point in time he has to decide how much to *consume* (i.e., the number of consumption goods to buy), how much to *save* for later (by investing in the cash bond), and how much to *invest* in the risky stock and its derivatives. The investor derives instantaneous utility $u(.)$ from consumption. Let $\rho \geq 0$ be his time-preference rate, i.e., the rate at which he discounts utility of future consumption. This parameter captures his impatience. If $\rho = 0$ current and future consumption are as valuable to him, whereas future consumption is less important if $\rho > 0$. We assume positive, but decreasing marginal utility, i.e. $u'(.) > 0$ and $u''(.) < 0$. Thus, the investor prefers more consumption to less, but the more he already consumes the less he values additional consumption. The concavity of the instantaneous utility function captures *risk-averse* behavior. We assume the investor's lifetime utility to be time-separable; i.e., it is given by $\int_0^\infty e^{-\rho t} u(c_t)dt$.

Let $w > 0$ be the investor's initial wealth, and let $\{c_t; t \geq 0\}$ be his nonnegative consumption-rate process of the consumption good. Let $\mathbf{P}_t$ denote the vector of time-t prices of all assets traded in the market, including the bond, stock and all derivatives. Let $\{\boldsymbol{\eta}_t; t \geq 0\}$ represent his dynamic trading strategy in these assets. The process $\{\boldsymbol{\eta}_t\}$ is $\{\mathcal{F}_t\}$-adapted, with $\boldsymbol{\eta}_t$ being a vector of equal length as $\mathbf{P}_t$.

Taking prices as given, the investor aims at maximizing his expected lifetime utility $U = U(\{c_t\},\{\boldsymbol{\eta}_t\})$, by choosing a consumption-rate process and a particular trading strategy that are budget-feasible. That is, he solves the intertemporal problem

$$\max_{\{c_t \geq 0\},\{\boldsymbol{\eta}_t\}} U(\{c_t\},\{\boldsymbol{\eta}_t\}) = \mathbb{E}_{\mathbb{P}}\left[\int_0^\infty e^{-\rho t} u(c_t)dt \mid \mathcal{F}_0\right] \tag{7.1}$$

s.t.

$$0 \leq \boldsymbol{\eta}_t{}'\mathbf{P}_t \leq w + \int_0^t \boldsymbol{\eta}_s{}' d\mathbf{P}_s - \int_0^t c_s ds, \quad t \geq 0. \tag{7.2}$$

The budget-feasibility restriction (7.2) says that, at all times, the value of his asset portfolio must be smaller than or equal to the sum of the investor's initial wealth and the gains from trading, minus total consumption so far. Since marginal utility is positive, the budget restriction will hold with equality in the optimum. Notice that this consumption-investment plan implies that no money is

[53] For more comprehensive treatments of consumption-based asset pricing and general equilibrium aspects, see Duffie (2001) and Cochrane (2001). Here we basically cover individual-investor optimality with prices being taken as given, but that suffices for our discussion.

added to it besides trading gains, or withdrawn from it, other than the amounts consumed. [54]

Duffie (2001) discusses (complicated) methods for solving similar classic consumption-investment problems. He gives numerous references that extend the problem in various directions. One solution method is the *stochastic-control* or *dynamic-programming* approach, which relies on the Hamilton-Jacobi-Bellman equation. Dynamic programming regards the problem at any point in time as maximizing the remaining expected utility. So in a sense, the investor updates his strategy as both time goes by and what state of the world has occurred.

Derivation of an explicit solution using stochastic-control methods is fortunately not necessary to arrive at the desired expression that relates asset prices to an investor's marginal utility of consumption. It is commonly known that, if the investor is optimizing, the investor equates *utility loss* with (time-discounted) *expected utility gain*. Let us illustrate this argument in our continuous-time derivative pricing setting.

Relating asset prices to marginal utilities
Suppose that $\{c_t; t \geq 0\}$ is the investor's *optimal* consumption process. (For simplicity, we refrain from discriminating between notation for *some* and the *optimal* consumption process.) Consider the following deviating experiment. Suppose the investor has arrived at time t. He decreases his optimal consumption in the time interval $[t, t+\Delta t]$ by going long, at time t, in an additional small number of $\zeta \neq 0$ units of some derivative F that has maturity T_F and price F_t. [55] Assume that he finances this purchase, costing ζF_t, by evenly decreasing his consumption over the interval. That is, at any time $s \in [t, t+\Delta t]$ he consumes $\zeta F_t / \Delta t$ less than is optimal for him. (In the limit for $\Delta t \to 0$, in which we are ultimately interested, this clearly does not matter.) Suppose next that the investor plans to increase his optimal consumption in some future time interval $[T, T+\Delta t]$ by selling the extra units of derivative F bought at time t, for the amount ζF_T (with the proceeds of this sale evenly consumed over the interval; but again in the limit this does not matter). In the intervals $[t+\Delta t, T]$ and $[T+\Delta t, \infty)$ he leaves his optimal consumption unaffected. As $\{c_t; t \geq 0\}$ is his expected-utility-maximizing consumption process, this deviating experiment must leave him with less or equal remaining expected utility. Therefore, given time-t information, it must hold that

$$\mathbb{E}_{\mathbb{P}}\left[\int_t^{t+\Delta t} e^{-\rho(s-t)}\, u\left(c_s - \frac{\zeta F_t}{\Delta t}\right) ds + \int_T^{T+\Delta t} e^{-\rho(s-t)}\, u\left(c_s + \frac{\zeta F_T}{\Delta t}\right) ds \mid \mathcal{F}_t\right] \tag{7.3}$$

$$\leq \mathbb{E}_{\mathbb{P}}\left[\int_t^{t+\Delta t} e^{-\rho(s-t)}\, u(c_s)\, ds + \int_T^{T+\Delta t} e^{-\rho(s-t)}\, u(c_s)\, ds \mid \mathcal{F}_t\right].$$

Taking everything to the left-hand side, dividing by ζ, and rewriting yields

---

[54] This is seen as follows. As the budget restriction holds with equality, taking differentials yields $d(\boldsymbol{\eta}_t{}'\mathbf{P}_t) = \boldsymbol{\eta}_t{}'d\mathbf{P}_t - c_t dt$ (as $d(w) = 0$), or $\boldsymbol{\eta}_t{}'d\mathbf{P}_t + \mathbf{P}_t{}'d\boldsymbol{\eta}_t + d\boldsymbol{\eta}_t{}'d\mathbf{P}_t = \boldsymbol{\eta}_t{}'d\mathbf{P}_t - c_t dt$ which simplifies to $\mathbf{P}_t{}'d\boldsymbol{\eta}_t + d\boldsymbol{\eta}_t{}'d\mathbf{P}_t = -c_t dt$. The left-hand side of this equality essentially represent portfolio *rebalancing costs*, which are absorbed by consumption.

[55] Derivative F is chosen without loss of generality (as long as $T_F > T$); i.e. he may alternatively have chosen to increase his investment in the reinvestment portfolio or the bond. The resulting equation to be derived below holds for all assets.

$$\mathbb{E}_{\mathbb{P}}\left[-\frac{F_t}{\Delta t}\int_t^{t+\Delta t} e^{-\rho(s-t)}\ \frac{u\left(c_s-\frac{\zeta F_t}{\Delta t}\right)-u\left(c_s\right)}{-\frac{\zeta F_t}{\Delta t}}ds+\frac{F_T}{\Delta t}\int_T^{T+\Delta t} e^{-\rho(s-t)}\ \frac{u\left(c_s+\frac{\zeta F_T}{\Delta t}\right)-u\left(c_s\right)}{\frac{\zeta F_T}{\Delta t}}ds \mid \mathcal{F}_t\right]\leq 0 \tag{7.4}$$

Taking the limit for the number of additionally purchased units of derivative F in the interval $[t, t+\Delta t]$ approaching zero, i.e. $\zeta \to 0$, we get

$$\mathbb{E}_{\mathbb{P}}\left[-\frac{F_t}{\Delta t}\int_t^{t+\Delta t} e^{-\rho(s-t)}\,u'\left(c_s\right)ds+\frac{F_T}{\Delta t}\int_T^{T+\Delta t} e^{-\rho(s-t)}\,u'\left(c_s\right)ds \mid \mathcal{F}_t\right]\leq 0\,. \tag{7.5}$$

Finally, letting the interval length go to zero, i.e. $\Delta t \to 0$, we obtain [56]

$$\mathbb{E}_{\mathbb{P}}\left[-F_t u'\left(c_t\right)+F_T\, e^{-\rho(T-t)}u'\left(c_T\right) \mid \mathcal{F}_t\right]\leq 0\,, \tag{7.6}$$

or

$$u'\left(c_t\right)\, F_t \geq \mathbb{E}_{\mathbb{P}}\left[e^{-\rho(T-t)}u'\left(c_T\right)\, F_T \mid \mathcal{F}_t\right]. \tag{7.7}$$

Clearly, this buy-sell strategy can also be reversed (as the investor can go short): The investor can increase his optimal consumption in the interval $[t, t+\Delta t]$ by short-selling ζ $F-$derivatives which yields him ζF_t extra for consumption. At time T he buys the ζ units back in the market for a total cost of ζF_T, which he finances with a reduced consumption in the interval $[T, T+\Delta t]$. It will be clear that in the limit equation (7.7) will hold, but with an $\leq$-sign. As both inequalities must hold at the same time, we arrive at the relation

$$u'\left(c_t\right)F_t = \mathbb{E}_{\mathbb{P}}\left[e^{-\rho(T-t)}u'(c_T)F_T \mid \mathcal{F}_t\right], \tag{7.8}$$

or $u'\left(c_t\right)=\mathbb{E}_{\mathbb{P}}[e^{-\rho(T-t)}(1+r_{F,tT})\,u'(c_T) \mid \mathcal{F}_t]$, with $r_{F,tT}=(F_T-F_t)/F_t$ the relative derivative return over $[t,T]$. This latter equation illustrates the following: If the investor is optimizing, the marginal utility *loss* of consuming a little less of the consumption good at time t and instead investing it in the asset, should equal the expected time-discounted marginal utility *benefit* of holding the extra assets until time T, then selling them, and using the proceeds to increase consumption then.

Rewriting (7.8) yields another commonly-known and useful interpretation. The time-t derivative price can be written as its expected discounted time-T value (and hence in particular, as its expected discounted time-T_F payoff) as follows:

$$F_t = \mathbb{E}_{\mathbb{P}}\left[e^{-\rho(T-t)}\,\frac{u'(c_T)}{u'(c_t)}\,F_T \mid \mathcal{F}_t\right]. \tag{7.9}$$

The time-discounted ratio of marginal utilities is known as the *marginal rate of (intertemporal consumption) substitution*. It clearly acts as a (stochastic) discount factor. It is the rate at which the investor is willing to substitute time-T consumption for time-t consumption. It basically measures his willingness (or incentive) to shift consumption through time by investing in the assets.

[56] Here we use $\lim_{\Delta x\to 0}\frac{1}{\Delta x}\int_x^{x+\Delta x} g(s)ds = \lim_{\Delta x\to 0}\frac{G(x+\Delta x)-G(x)}{\Delta x} = G'(x) = g(x)$.

Realize that this substitution rate depends on the uncertain state of the world (which is governed by the Brownian motion $\mathbf{W}$). As the main purpose of financial assets is arguably to provide an opportunity to shift consumption through time and states of the world, it makes sense that the marginal rate of substitution shows up in the asset-pricing equation (7.9).

Notice that we have looked at the problem of *some* investor to derive equation (7.9). [57] Equation (7.9) should hold for each optimizing investor (and for each tradable asset), regardless of his specific utility function, time-preference rate, budget-feasibility set, or optimal consumption-investment plan. We know from e.g. Duffie (2001) that, if the market were complete, the marginal rate of substitution of the *representative investor* (which exists in that case) can be used to price the assets. In an incomplete-markets setting, there may exist at least as many marginal rates of substitution that price assets according to (7.9) as there are investors. However, given that we have "fixed" the $\mathbb{Q}$-measure by having chosen a specific market price of volatility risk (and hence have essentially completed the market, as derivatives have a unique price), the marginal rate of substitution in (7.9) can be seen as the one of the representative investor.

7.2 Relating derivative returns to optimal consumption

Equation (7.9) relates the derivative price to the optimal consumption process $\{c_t\}$ of an investor. Equation (5.16) on the other hand, relates the derivative price to the stochastic-discount-factor process $\{M_t\}$: It states that in the absence of arbitrage, it must hold that $M_t F_t = \mathbb{E}_{\mathbb{P}}[M_T F_T \mid \mathcal{F}_t]$. Combining both equations shows that the SDF equals the time-discounted marginal utility of optimal investor consumption:

$$M_t = e^{-\rho t} u'(c_t). \tag{7.10}$$

Realize that in this way, we essentially link the no-arbitrage condition (which must obviously hold in equilibrium) with the investor's equilibrium behavior.

Equation (7.10) next allows us to relate the expected derivative return expressed in terms of the SDF (recall (5.22)) to investor consumption. As the optimal consumption process is an Itô process [58], Itô's lemma yields

$$\begin{aligned} dM_t &= -\rho e^{-\rho t} u'(c_t)dt + e^{-\rho t} u''(c_t)dc_t + \tfrac{1}{2} e^{-\rho t} u'''(c_t)(dc_t)^2 \\ &= -\rho M_t dt + M_t \frac{u''(c_t)}{u'(c_t)} dc_t + \tfrac{1}{2} M_t \frac{u'''(c_t)}{u'(c_t)} (dc_t)^2 . \end{aligned} \tag{7.11}$$

As $(dc_t)^2$ is of order dt, we obtain for dM_t / M_t an expression like

$$\frac{dM_t}{M_t} = (..)dt + c_t \frac{u''(c_t)}{u'(c_t)} \frac{dc_t}{c_t}, \tag{7.12}$$

such that the expected derivative return (5.22) may also be written as

[57] As Cochrane (2001, section 2.1) points out, (7.9) holds under very general circumstances in many different model settings. His discussion in section 2.2 on general equilibrium, and on either asset prices or consumption being the chicken or the egg, is interesting. For our purposes it all does not matter.
[58] Notice that $c_t = v(e^{\rho t} M_t)$ with $v(.)$ the inverse function of $u'(.)$. As $\{M_t\}$ follows the Itô process (5.23), $\{c_t\}$ follows an Itô process as well.

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = r_t dt - c_t \frac{u''(c_t)}{u'(c_t)} \text{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, \frac{dc_t}{c_t} \mid \mathcal{F}_t\right]. \tag{7.13}$$

This is a particularly illuminating relationship: [59] It connects the expected derivative return to the optimal consumption plan of an investor. Specifically, it specifies the required return the investor demands for bearing the risk inherent in the derivative. The factor $-c_t u''(c_t)/u'(c_t)$ is the investor's *coefficient of relative risk aversion* (CRRA). Notice that the CRRA is positive. Equation (7.13) implies that the larger (smaller) the covariance between the derivative return and the investor's consumption growth, the larger (smaller) the expected derivative return. In particular, if there is negative correlation between the derivative return and consumption growth, the expected derivative return is below the risk-free rate.

Optimal consumption in case of power utility
Before proceeding, let us derive the investor's optimal consumption-rate process in a special case, and confirm that (7.13) can indeed be used for pricing. Given its convenient properties, consider an investor that derives *power utility* from consumption. Specifically, the investor's instantaneous utility function is given by

$$u(c_t) = \frac{c_t^{1-\psi}}{1-\psi}; \ \psi > 0 . \tag{7.14}$$

Here, ψ is the investor's coefficient of relative consumption-risk aversion in different states of the world: his CRRA equals $-c_t u''(c_t)/u'(c_t) = \psi$ .The investor's willingness to shift consumption through time, i.e., his *elasticity of intertemporal substitution*, is given by $1/\psi$. [60] Notice that the risk-averter the investor, the less he is simultaneously willing to transfer consumption through time.

Recall from section 3.1 that the no-arbitrage condition defines the SDF as the bond-discounted Radon-Nikodym process: $M_t = B_t^{-1} L_t$. Hence, by (2.2) and (3.11),

$$M_t = B_0^{-1} \exp\left(-\int_0^t \left(r_u + \tfrac{1}{2}\mathbf{\gamma}_u{}'\mathbf{\gamma}_u\right)du - \int_0^t \mathbf{\gamma}_u{}' d\mathbf{W}_u\right). \tag{7.15}$$

On the other hand, equilibrium investor behavior defines the SDF as his time-discounted marginal utility of optimal consumption: $M_t = e^{-\rho t} u'(c_t)$. The investor's marginal utility process $\{u'(c_t); t \geq 0\}$ thus equals

$$u'(c_t) = B_0^{-1} \exp\left(-\int_0^t \left(r_u - \rho + \tfrac{1}{2}\mathbf{\gamma}_u{}'\mathbf{\gamma}_u\right)du - \int_0^t \mathbf{\gamma}_u{}' d\mathbf{W}_u\right). \tag{7.16}$$

Marginal power utility equals by $u'(c_t) = c_t^{-\psi}$. The investor's optimal consumption rate is thus given by

$$c_t = B_0^{\frac{1}{\psi}} \exp\left(\frac{1}{\psi}\int_0^t \left(r_u - \rho + \tfrac{1}{2}\mathbf{\gamma}_u{}'\mathbf{\gamma}_u\right)du + \frac{1}{\psi}\int_0^t \mathbf{\gamma}_u{}' d\mathbf{W}_u\right). \tag{7.17}$$

[59] We have essentially derived what is known as the *consumption capital asset pricing model* (CCAPM; Breeden (1979)) in our derivative-pricing setting.
[60] The fact that this elasticity equals the reciprocal of CRRA is an oft-cited drawback of power utility: it is too restrictive. *Epstein-Zin utility* breaks this 1-to-1 link, but is unfortunately much harder to work with.

Logarithmic consumption evolves according to

$$d\ln c_t = \frac{1}{\psi}\left[\left(r_t - \rho + \tfrac{1}{2}\boldsymbol{\gamma}_t'\boldsymbol{\gamma}_t\right)dt + \boldsymbol{\gamma}_t' d\mathbf{W}_t\right], \qquad (7.18)$$

which, after applying Itô's lemma to $\exp(\ln c_t)$, yields the SDE for optimal consumption,

$$\frac{dc_t}{c_t} = \frac{1}{\psi}\left[\left(r_t - \rho + \tfrac{1}{2}\boldsymbol{\gamma}_t'\boldsymbol{\gamma}_t\left[1+\frac{1}{\psi}\right]\right)dt + \boldsymbol{\gamma}_t' d\mathbf{W}_t\right]. \qquad (7.19)$$

Taking the derivative-return process (5.3) into account, the covariance between the derivative return and investor's consumption growth then equals

$$\begin{aligned}\mathrm{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, \frac{dc_t}{c_t} \mid \mathcal{F}_t\right] &= \mathrm{cov}_{\mathbb{P}}\left[\beta_{F,S,t}\left(\sigma_t dW_{S,t}\right) + \boldsymbol{\beta}_{F,x,t}'\left(\boldsymbol{\Sigma\Lambda}_t d\mathbf{W}_{x,t}\right), \frac{1}{\psi}\boldsymbol{\gamma}_t' d\mathbf{W}_t \mid \mathcal{F}_t\right] \\ &= \frac{1}{\psi}\left[\beta_{F,S,t}\left(\sigma_t \gamma_{S,t}\right)dt + \boldsymbol{\beta}_{F,x,t}'\left(\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t}\right)dt\right]. \qquad (7.20)\end{aligned}$$

The second equality follows from $\boldsymbol{\gamma}_t' d\mathbf{W}_t = \gamma_{S,t} dW_{S,t} + \boldsymbol{\gamma}_{x,t}' d\mathbf{W}_{x,t}$, the independence of W_S and $\mathbf{W}_x$, and some further manipulations. As the CRRA of power utility equals ψ, we finally obtain from the consumption-return formula (7.13) our by now familiar expression,

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = \left[r_t + \beta_{F,S,t}\left(\sigma_t \gamma_{S,t}\right) + \boldsymbol{\beta}_{F,x,t}'\left(\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t}\right)\right]dt. \qquad (7.21)$$

7.3 Consumption smoothing and the volatility risk premium

To better be able to interpret result (7.13), remember that the key idea of the *life-cycle* or *permanent-income hypothesis* from macroeconomics [61] is that individuals (economic agents) tend to *smooth* their *consumption* over time by borrowing and saving. The primary aim of saving is future consumption, whereas borrowing is primarily used for increasing consumption now.

The CRRA measures the individual's aversion to having different consumption opportunities in different states of the world. To achieve a smooth consumption pattern, it may therefore be argued that an investor values an asset more if it pays off more when consumption is low (and vice versa). Such an asset is less risky in terms of his consumption opportunities: it clearly provides *insurance*. This theory thus suggests him to be willing to pay a higher price for a derivative which return covaries less with his consumption growth (and vice versa). As high (resp. low) prices are associated with low (resp. high) expected returns, formula (7.13) makes perfect sense. Moreover, as consumption smoothing is the investor's primary purpose according to the permanent-income hypothesis, it is this covariance that matters for his required return, not the riskiness of the derivative itself (i.e., $\mathrm{var}_{\mathbb{P}}[dF_t / F_t \mid \mathcal{F}_t]$ does not show up in (7.13)). The analogy with the ordinary CAPM is apparent. Similar reasoning is known in the literature.

[61] Modigliani and Brumberg (1954), Friedman (1957). See Romer (2001) for a textbook treatment.

Volatility risk premium is negative on theoretical grounds
Given these appealing arguments, is it possible to derive a theoretical prediction on the sign of the volatility risk premium? First, combine (7.13) with (5.25) to get

$$-c_t \frac{u''(c_t)}{u'(c_t)} \text{cov}_{\mathbb{P}} \left[\frac{dF_t}{F_t}, \frac{dc_t}{c_t} \mid \mathcal{F}_t \right] = \left[\beta_{F,S,t} \left(\sigma_t \gamma_{S,t} \right) + \boldsymbol{\beta}_{F,x,t}' \left(\boldsymbol{\Sigma} \boldsymbol{\Lambda}_t \boldsymbol{\gamma}_{x,t} \right) \right] dt \,. \quad (7.22)$$

This equation relates the riskiness of a derivative -in terms of its impact on an investor's optimal consumption plan- to both the stock risk premium and the volatility risk premium.

To make our statement, it is natural to focus on assets that are especially prone to volatility risk. Arguably, the majority of investors allocates most of its wealth to long stock and bond positions (in practice). So, as long as markets are either fairly steady or go up, an investor's consumption stream will be rather stable. Now assume that a certain investor holds a delta-neutral positive-vega derivative G in his portfolio, like a long straddle or variance swap. Such an instrument does not (instantaneously) respond to stock price fluctuations, but *is* prone to volatility risk. Since $\beta_{G,S,t} = 0$ and $\boldsymbol{\beta}_{G,x,t} > \mathbf{0}$, it holds for G:

$$-c_t \frac{u''(c_t)}{u'(c_t)} \text{cov}_{\mathbb{P}} \left[\frac{dG_t}{G_t}, \frac{dc_t}{c_t} \mid \mathcal{F}_t \right] = \boldsymbol{\beta}_{G,x,t}' \left(\boldsymbol{\Sigma} \boldsymbol{\Lambda}_t \boldsymbol{\gamma}_{x,t} \right) dt \,. \quad (7.23)$$

Suppose then that the market suddenly crashes unexpectedly. Such a collapse typically goes hand in hand with decreased consumption: the investor's long stock and bond positions devaluate. However, as dropping markets are typically also accompanied by increasing volatility in practice (the leverage effect [62]), the value of positive-vega derivative G will typically rise. As such, the increase in G (partly) offsets the drop in value in the investor's stock and bond portfolio (and hence his consumption opportunities). Holding a delta-neutral positive-vega derivative G in his portfolio thus indeed helps the investor pursuing a smooth consumption pattern over time (as opposed to not holding such a derivative). Hence, such a derivative investment provides *consumption insurance* in bad times. So the correlation between $G's$ return and the investor's consumption is likely to be negative.

Looking back at formula (7.23), our theory thus predicts the volatility risk premium (and hence the market price of volatility risk) to be *negative*. [63] In particular, reconsidering (5.5), the expected return on delta-neutral positive-vega derivatives such as straddles and variance swaps is expected to be smaller than the risk-free rate.

The recent empirical evidence on market volatility-risk compensation discussed in section 6.3 suggests that financial markets practice conforms with this theory.

---

[62] Indeed, our reasoning here is explicitly based on the existence of a leverage effect, which we did not model. However, as we show in sections 3.4 and 3.5 of chapter VI on the Heston (1993) model (which does model the leverage effect), a similar formula as (7.23) results in case of leverage. Our reasoning here therefore theoretically still applies in case leverage is explicitly taken into account.

[63] A similar argument based on e.g. a long call option (which also rises in value if volatility rises) instead of a *delta-neutral* positive-vega derivative seems impossible to make: Reconsidering (7.22), dropping markets *may* induce a (temporarily) negative stock risk premium $\sigma_t \gamma_{S,t}$. Moreover, the value of a call option decreases if the underlying declines in value. Hence, basing the argument on a non-delta-neutral derivative, we cannot disentangle what the sign of the volatility risk premium should be theoretically.

8. Foreign exchange markets

So far the context of our model has been a stock market setting. This section explains how our multifactor SV model may also be used for derivative pricing and hedging in foreign exchange (FX) markets. (Apart from this section the remainder of this thesis focuses on the stock market setting of our model however.) Section 8.1 first discusses the Garman-Kohlhagen (1983), or Black-Scholes currency model. In section 8.2, the subjects are the currency-derivatives model of Bates (1996b), the evidence on correlation between exchange rate and volatility shocks, and the evidence on multiple SV factors in FX markets. Finally, in section 8.3 we state our model in a foreign-exchange market setting.

8.1 Garman-Kohlhagen (1983) model

Currency derivatives are contracts written on an underlying exchange rate; e.g., the euro/dollar exchange rate. Consider for example a Dutch company that does business in the US. It receives part of its profits in dollars. Exchanging these dollars for euros entails exchange-rate risk. The company may wish to partly hedge this risk by buying FX put options with strike K, giving the right to sell one dollar for K euros at the expiration date T of the contract. The company thus knows the minimum euro amount for which it can sell dollars for. If $\{S_t; t \geq 0\}$ represents the euro/dollar exchange rate [64], the put payoff is $\max\{0, K - S_T\}$.

A widely-used model for pricing currency derivatives is the Garman-Kohlhagen (1983) model. This model assumes that the exchange rate S follows a geometric Brownian motion as in (2.3), but with constant volatility σ. There are two assets: a domestic, domestic-currency-denominated cash bond B, which evolves as $dB_t = r_t B_t dt$, and a foreign, foreign-currency-denominated cash bond B^f, which evolves as $dB_t^f = r_t^f B_t^f dt$. [65]

In the Garman-Kohlhagen model, the time-t price C_t of a European FX call option C struck at K having maturity T is given by the conventional Black-Scholes call price formula (3.20), but with the average dividend yield $\bar{q}_t$ replaced by the average foreign risk-free interest rate, $\bar{r}_t^f$, over the remaining option's life $\tau_t = T - t$. The price of a FX forward contract is given by (3.21), but again, with $\bar{q}_t$ replaced by $\bar{r}_t^f$. The intuitive reason for these results is that the owner of foreign currency receives a "dividend yield q_t" equal to the foreign risk-free interest rate r_t^f when investing this currency in the foreign market.

In FX-markets banking practice, prices of currency options are quoted in terms of their associated implied volatilities. These implied volatilities are computed using the Garman-Kohlhagen model, stressing its importance.

8.2 The Bates (1996) model and empirical evidence

It is well known that, like stock prices, exchange rates exhibit stochastic volatility as well. [66] The Garman-Kohlhagen model is therefore misspecified. As mentioned

---

[64] It will soon become clear why we use the same notation as for a stock price.
[65] It should be noted that the original model assumes constant interest rates r and r^f. Here, we already partly extend the model towards our proposed extension to be discussed in section 8.3.
[66] See, e.g., Hull (2003), Franses and Van Dijk (2000), Bates (1996), Shephard (1996), Ghysels, Harvey and Renault (1996), Bollerslev, Engle and Nelson (1994), Bollerslev, Chou and Kroner (1992), Melino and Turnbull (1990), Bollerslev (1990), and numerous others.

by Carr and Wu (2004b), the jump-diffusion model of Bates (1996b) so far still represents the state of the art in the currency option pricing literature.

Bates's model assumes that the exchange rate follows the geometric Brownian motion-type SDE (2.3), but with a Heston (1993) volatility (i.e. CIR) process, extended with Merton (1976) jumps. The model allows for correlation between exchange-rate returns and volatility shocks (i.e., the "leverage effect").

Empirical evidence on correlation and multiple SV factors
Bates (1996b) performs estimation using a calibration-type, implicit parameter estimation strategy on a panel of Deutsche Mark call and put option prices over the period 1984 – 1991. His estimation results for his model without jumps imply a correlation of only 0.045 (std.dev. 0.002), whereas with jumps this correlation is estimated at 0.078 (std.dev. 0.003). Bates thus finds that the correlation between exchange-rate returns and volatility shocks is close to zero in value.

Other evidence that supports this is in, e.g., Chesney and Scott (1989), who estimate this correlation in a particular SV model at –0.0065 (std.dev. 0.76) for the dollar/Swiss-franc exchange rate (period 1979 – 1984). Melino and Turnbull (1990) estimate this correlation at –0.18 (std.dev. 0.09) for the Canadian-US dollar exchange rate (period 1975 – 1986). Hull (2003, p. 459) mentions that the typical pattern of FX implied volatilities over moneyness is a volatility *smile*, which is consistent with a zero correlation between FX returns and volatility shocks. Franses and Van Dijk (2000, p. 18) provide further evidence on the in general small magnitude of this correlation for eight exchange rates, for the period 1980 - 1998.

The evidence thus suggests that for exchange rates, the assumption of a zero correlation between FX returns and volatility shocks seems appropriate. Moreover, there seems to be no real reason why there should be a consistent positive or negative correlation in the first place: If the domestic/foreign currency exchange rate is inverted, one obtains the foreign/domestic rate, which is the series closely followed by traders and investors in the foreign country. This point is also mentioned by Franses and Van Dijk (2000).

And what about multiple SV factors in FX markets? Bates (1996b) concludes that *"The postulated 1-factor model […] does a poor job in capturing the evolution over time of implicit volatilities from multiple option maturities. There is therefore substantial scope for improvement from using multifactor rather than single-factor models of stochastic volatility."*

8.3 Our model in a FX-markets context

Given the evidence, our multifactor SV model may appear to be of particular use in derivative pricing contexts regarding foreign exchange markets. Although the assumption of no leverage effect may generally not be realistic in stock markets, the assumption of a zero correlation between exchange-rate returns and volatility shocks seems more realistic. So how can we translate our multifactor SV model into an FX-markets setting?

Assume that two basic assets are traded in the domestic market, a domestic cash bond B, and a foreign bond B^f. The price of the domestic bond evolves as $dB_t = r_t B_t dt$, where $\{r_t\}$ represents the (deterministic) domestic risk-free interest rate process. The foreign bond price evolves as $dB_t^f = r_t^f B_t^f dt$ in the foreign currency, where $\{r_t^f\}$ is the (deterministic) foreign risk-free interest rate process. The domestic/foreign currency exchange rate S follows the SDE (2.3), with SV specification (2.5)-(2.7). This specification assumes a zero correlation between exchange-rate returns and exchange-rate volatility shocks.

Obviously, neither the exchange-rate process $\{S_t\}$, nor the foreign-currency-denominated price process of the foreign bond $\{B_t^f\}$, represent price processes of assets that are tradable in the domestic market. The domestic-currency-denominated price of the foreign bond, which we denote by $\{S_t^r\}$, is given by $S_t^r = B_t^f S_t$.[67] Itô's formula yields the SDE

$$dS_t^r = (\mu_t + r_t^f) S_t^r dt + \sigma_t S_t^r dW_{S,t} . \qquad (\mathbb{P}) \qquad (8.1)$$

In contrast to $\{B_t^f\}$, the process $\{S_t^r\}$ represents the value of an asset (the foreign bond) that *is* tradable in the domestic market. The risk from holding the foreign bond in the domestic market is derived from exchange-rate fluctuations.

Comparing the SDE (8.1) with the SDE (2.4) reveals that, although the contexts of this FX-market setting and the stock market setting (as discussed in section 2) differ, the mathematical set-ups are essentially similar. The only difference is that in the stock market setting, the stock pays a dividend yield of q_t, whereas in the FX-market setting, the foreign bond pays a "dividend yield" of r_t^f. Therefore, all of the theoretical analysis in sections 3, 4, 5, and 6.2 on derivative pricing, hedging, asset returns, and volatility-risk dominated strategies, carries through, and can directly be applied to this FX-markets context in a straightforward manner. We only need to replace q_t with r_t^f.

For example, the market price of exchange-rate risk is given by $\gamma_{S,t} = (\mu_t + r_t^f - r_t)/\sigma_t$. Moreover, all FX derivatives can be priced using the risk-neutral or SDF valuation formulas (3.9) or (3.10). In (3.9), $\mathbb{Q}$ is now the measure under which the domestic-bond-discounted domestic-currency-denominated foreign-bond price, i.e. the process $\{B_t^{-1} S_t^r\}$, is a martingale. (In the stock market setting $\mathbb{Q}$ is the measure under which the discounted reinvestment portfolio $\{B_t^{-1} S_t^r\}$ is a martingale.) The exchange rate follows $dS_t = (r_t - r_t^f) S_t dt + \sigma_t S_t d\tilde{W}_{S,t}$ under $\mathbb{Q}$. As a final illustration, the price of a European FX call option having payoff $C_T = \max\{0, S_T - K\}$ is given by (3.17) and (3.19)-(3.24) with $\bar{q}_t$ replaced by $\bar{r}_t^f$.

9. From theory to practice

So far the discussion of our model has mostly been theoretical. We have shown that having available a full specification of a stock (or foreign exchange) market that is prone to stochastic volatility, given a specific pricing measure $\mathbb{Q}$, we know how to uniquely price and hedge derivative securities. We know what returns to expect. Moreover, we have acquired a lot of intuition for volatility risk; both in terms of investment strategies, and, economically, with regard to volatility-risk compensation. But what about using our model for practical purposes? How *do* we obtain such a full specification which includes numerical values of parameters,

[67] Again on purpose, we use the same notation as for the (tradable) reinvestment portfolio.

the specific $\mathbb{Q}$-measure, and the unobserved volatility? In other words, how can we numerically relate our theory to practice?

First of all, we ought to be willing to assume that observed real-world data is truly generated by our specific arbitrage-free model. For example, the market-observed term structure of interest rates can serve as information on the cash bond process. Information on the stock price and dividend yield processes can be obtained from observed stock and forward prices and their history. As volatility affects both stock and derivative prices, information on the stochastic volatility process ideally comes from both sources jointly. Given the latent character of volatility, this poses an additional challenge. And what about the specific $\mathbb{Q}$-measure that the market uses for pricing derivatives in practice? In other words, how can we numerically determine the empirical market price of volatility risk? In practice the pricing measure $\mathbb{Q}$ is not directly observable; it is only implicitly present in the prices of derivatives traded in the market. We therefore need a means to back out this measure from these prices, and as such "let the market speak" on what specific $\mathbb{Q}$ it uses for valuation.

Once both the model parameters have been estimated, and an estimate of the unobserved volatility-driving factors is available, we can in principle price and hedge any other derivative by means of, e.g., Monte Carlo simulation. Suppose for example, that we desire to price and hedge a new-to-be-issued exotic over-the-counter financial derivative in the 1-factor SV special case of our model. The risk-neutral valuation formula (3.9) can be used for pricing. To hedge this contract, we could follow some discretized version of the hedge strategy outlined in section 4.3. As continuous rebalancing of the hedge portfolio is not possible in practice due to transaction costs, such a strategy will not neutralize all risk. Not only due to the necessary discretization, but also due to model error and the unavoidable parameter uncertainty inherent to any estimation procedure. Such a hedging strategy requires an estimate of the latent volatility-driving factor x at each point in time the hedging portfolio is rebalanced: Prices and hence deltas and vegas depend on it. These price sensitivities can in principle only be obtained by Monte Carlo simulation, or one should be willing to rely on some approximation. The derivative instrument with which to hedge volatility shocks must be market-observable, and preferably highly liquid. For example, an index option may do the job if the aim is pricing and hedging derivatives written on a stock index.

The issue still left unspecified in this discussion however, is *how* to exactly extract the necessary information from real-world stock and option prices, that may serve to numerically specify our model, and *how* to obtain an estimate of the unobserved volatility. The next chapter is devoted to this problem.

Appendix

IIa. The bond-stock market is arbitrage-free: A proof

In section 2 we state that the bond-stock market is arbitrage free. Here we prove this claim by showing that the prices of the tradable assets (the cash bond and dividend-paying stock) expressed in terms of the value of the cash bond, are $(\mathbb{Q},\{\mathcal{F}_t\})-$martingales, as required by the First Fundamental Theorem of Asset Pricing.

First, and obviously, the relative price of the cash bond equals 1 for all t under all measures, as $B_t^{-1}B_t = 1\ \forall t$. It is therefore a martingale, irrespective of the specific probability measure and filtration. Second, the ex-dividend stock price process $\{S_t; t \geq 0\}$ does not represent the value of a tradable asset, because it does not account for dividend payments. The stock cannot be bought without its dividend payouts: The dividend payment of $q_t S_t dt$ in the time interval $[t, t+dt]$ is automatically received. Notice that this payment instantaneously buys an additional $q_t dt$ stocks. Consider then the following portfolio strategy. Suppose an investor buys a_0 (some positive number) stocks at time 0, and suppose he immediately reinvests all received dividends by buying additional units of stock. Doing so, at time t the investor has

$$a_t \equiv a_0 \exp\left(\int_0^t q_u du\right) \qquad \text{(a.1)}$$

stocks in his so-called *reinvestment portfolio*, which we denote by S^r. [68] The reinvestment portfolio is a tradable asset. Its time-t value equals $S_t^r = a_t S_t$. Notice that $S_t^r = f(t, S_t)$, a function of t and S_t. Using (2.3), Itô's lemma yields

$$dS_t^r = (\mu_t + q_t)S_t^r dt + \sigma_t S_t^r dW_{S,t}. \qquad (\mathbb{P}) \qquad \text{(a.2)}$$

Consider now the relative time-t value of the reinvestment portfolio, Z_t^r, defined as $Z_t^r \equiv B_t^{-1} S_t^r$. Integration by parts yields the SDE for $\{Z_t^r\}$ under $\mathbb{P}$, being

$$dZ_t^r = (\mu_t + q_t - r_t)Z_t^r dt + \sigma_t Z_t^r dW_{S,t}. \qquad (\mathbb{P}) \qquad \text{(a.3)}$$

Under our choice of $\mathbb{Q}$ for which

$$d\tilde{W}_{S,t} = dW_{S,t} + \gamma_{S,t} dt, \quad \gamma_{S,t} = \frac{\mu_t + q_t - r_t}{\sigma_t}, \qquad \text{(a.4)}$$

$\{Z_t^r\}$ follows

$$dZ_t^r = \sigma_t Z_t^r d\tilde{W}_{S,t}. \qquad (\mathbb{Q}) \qquad \text{(a.5)}$$

(Notice that the assumed specific form of the market price of volatility risk does not play a role here.) The process $\{Z_t^r; t \geq 0\}$ is driftless under $\mathbb{Q}$. As it holds that $\mathbb{E}_{\mathbb{Q}}(\exp[\frac{1}{2}\int_0^T \sigma_s^2 ds]) < \infty$ (which is clear from the results in section 10 of appendix B), $\{Z_t^r\}$ is a martingale under $\mathbb{Q}$ w.r.t. $\{\mathcal{F}_t\}$ (and not a *local* martingale; see e.g. Baxter and Rennie (1996, p. 79)). Therefore, the bond-stock market is indeed arbitrage-free.

---

[68] Notice the analogy with depositing an amount B_0 in the money market account at time 0. Continuous reinvestment of received interest-rate payments, results in an amount of $B_t = B_0 \exp\left(\int_0^t r_u du\right)$ at time t.

IIb. Derivative betas

In this part we derive the expressions for the stock and volatility betas of a general derivative F for which $F_t = F(t, S_t, \mathbf{x}_t)$, as given in section 5.1. Recall that we define the stock beta of derivative F as

$$\beta_{F,S,t} \equiv \text{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, \frac{dS_t}{S_t} \mid \mathcal{F}_t\right] / \text{var}_{\mathbb{P}}\left[\frac{dS_t}{S_t} \mid \mathcal{F}_t\right]. \tag{b.1}$$

From (2.3) and (3.6), the covariance between the spot derivative and stock returns equals

$$\begin{aligned}\text{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, \frac{dS_t}{S_t} \mid \mathcal{F}_t\right] &= \text{cov}_{\mathbb{P}}\left[\frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t}\sigma_t dW_{S,t} + \frac{1}{F_t}\frac{\partial F_t}{\partial \mathbf{x}_t{}'}\mathbf{\Sigma\Lambda}_t d\mathbf{W}_{x,t},\ \sigma_t dW_{S,t} \mid \mathcal{F}_t\right] \\ &= \sigma_t^2 \frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t}\text{var}_{\mathbb{P}}\left[dW_{S,t} \mid \mathcal{F}_t\right] = \sigma_t^2 \frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t} dt,\end{aligned} \tag{b.2}$$

in which the independence of the Brownian motions $\mathbf{W}_x$ and W_S is exploited, and in which we use $\text{var}_{\mathbb{P}}[dW_{S,t} \mid \mathcal{F}_t] = \text{var}_{\mathbb{P}}[dW_{S,t}] = dt$, which holds due to the independent-increments property of Brownian motion. The variance of the instantaneous stock return equals

$$\text{var}_{\mathbb{P}}\left[\frac{dS_t}{S_t} \mid \mathcal{F}_t\right] = \text{var}_{\mathbb{P}}\left[\mu_t dt + \sigma_t dW_{S,t} \mid \mathcal{F}_t\right] = \sigma_t^2 dt. \tag{b.3}$$

The time-t stock sensitivity of derivative F, as measured by its stock beta, then equals

$$\beta_{F,S,t} = \frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t} = \frac{S_t}{F_t}\Delta_{F,t}. \tag{b.4}$$

The volatility beta of derivative F is defined as

$$\boldsymbol{\beta}_{F,x,t} \equiv \left(\text{var}_{\mathbb{P}}\left[d\mathbf{x}_t \mid \mathcal{F}_t\right]\right)^{-1}\text{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, d\mathbf{x}_t \mid \mathcal{F}_t\right]. \tag{b.5}$$

By (2.6),

$$\begin{aligned}\text{var}_{\mathbb{P}}\left[d\mathbf{x}_t \mid \mathcal{F}_t\right] &= \text{var}_{\mathbb{P}}\left[\mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t)dt + \mathbf{\Sigma\Lambda}_t d\mathbf{W}_{x,t} \mid \mathcal{F}_t\right] \\ &= \mathbf{\Sigma\Lambda}_t\, \text{var}_{\mathbb{P}}\left[d\mathbf{W}_{x,t} \mid \mathcal{F}_t\right]\mathbf{\Lambda}_t\mathbf{\Sigma}' = \mathbf{\Sigma\Lambda}_t^2\mathbf{\Sigma}' dt,\end{aligned} \tag{b.6}$$

in which the third equality uses $\text{var}_{\mathbb{P}}[d\mathbf{W}_{x,t} \mid \mathcal{F}_t] = \text{var}_{\mathbb{P}}[d\mathbf{W}_{x,t}] = \mathbf{I}_n dt$, as the Brownian motions $W_1, .., W_n$ are independent. For the covariance between the spot derivative return and spot factor changes, we get by (2.6) and (3.6):

$$\begin{aligned}\text{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, d\mathbf{x}_t \mid \mathcal{F}_t\right] &= \text{cov}_{\mathbb{P}}\left[\frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t}\sigma_t dW_{S,t} + \frac{1}{F_t}\frac{\partial F_t}{\partial \mathbf{x}_t{}'}\mathbf{\Sigma\Lambda}_t d\mathbf{W}_{x,t},\ \mathbf{\Sigma\Lambda}_t d\mathbf{W}_{x,t} \mid \mathcal{F}_t\right] \\ &= \text{cov}_{\mathbb{P}}\left[\frac{1}{F_t}\frac{\partial F_t}{\partial \mathbf{x}_t{}'}\mathbf{\Sigma\Lambda}_t d\mathbf{W}_{x,t},\ \mathbf{\Sigma\Lambda}_t d\mathbf{W}_{x,t} \mid \mathcal{F}_t\right] \\ &= \text{cov}_{\mathbb{P}}\left[\mathbf{\Sigma\Lambda}_t d\mathbf{W}_{x,t},\ d\mathbf{W}_{x,t}{}'\mathbf{\Lambda}_t\mathbf{\Sigma}'\frac{1}{F_t}\frac{\partial F_t}{\partial \mathbf{x}_t} \mid \mathcal{F}_t\right] \\ &= \mathbf{\Sigma\Lambda}_t\mathbb{E}_{\mathbb{P}}\left[d\mathbf{W}_{x,t}d\mathbf{W}_{x,t}{}' \mid \mathcal{F}_t\right]\mathbf{\Lambda}_t\mathbf{\Sigma}'\frac{1}{F_t}\frac{\partial F_t}{\partial \mathbf{x}_t} = \mathbf{\Sigma\Lambda}_t^2\mathbf{\Sigma}'\frac{1}{F_t}\frac{\partial F_t}{\partial \mathbf{x}_t}dt.\end{aligned} \tag{b.7}$$

The second equality uses the independence of $\mathbf{W}_x$ and W_S, and the fourth equality uses $\mathbb{E}_{\mathbb{P}}[d\mathbf{W}_{x,t} \mid \mathcal{F}_t] = \mathbb{E}_{\mathbb{P}}[d\mathbf{W}_{x,t}] = \mathbf{0}$. [69] F's volatility beta then equals

$$\boldsymbol{\beta}_{F,x,t} = \frac{1}{F_t}\frac{\partial F_t}{\partial \mathbf{x}_t} = \frac{1}{F_t}\mathcal{V}_{F,t}. \tag{b.8}$$

IIc. Variance-swap analysis

As explained in the main text, the time-t value of a variance swap is given by $VS_t = VS(t, I_t, \mathbf{x}_t)$; recall (6.4). Here we aim at deriving the SDE the arbitrage-free variance-swap value follows under $\mathbb{P}$. We follow a similar analysis as in section 3.1. Itô's formula yields, under $\mathbb{P}$,

$$\begin{aligned} dVS_t &= \frac{\partial VS_t}{\partial t}dt + \frac{\partial VS_t}{\partial I_t}dI_t + \frac{\partial VS_t}{\partial \mathbf{x}_t{'}}d\mathbf{x}_t + \tfrac{1}{2}d\mathbf{x}_t{'}\frac{\partial^2 VS_t}{\partial \mathbf{x}_t \partial \mathbf{x}_t{'}}d\mathbf{x}_t \qquad (\mathbb{P}) \qquad \text{(c.1)} \\ &= \left[\frac{\partial VS_t}{\partial t} + \sigma_t^2\frac{\partial VS_t}{\partial I_t} + \frac{\partial VS_t}{\partial \mathbf{x}_t{'}}\mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t) + \tfrac{1}{2}\operatorname{tr}\left(\boldsymbol{\Lambda}_t\boldsymbol{\Sigma}'\frac{\partial^2 VS_t}{\partial \mathbf{x}_t \partial \mathbf{x}_t{'}}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_t\right)\right]dt + \frac{\partial VS_t}{\partial \mathbf{x}_t{'}}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t}, \end{aligned}$$

as $dI_t = \sigma_t^2 dt$ such that $(dI_t)(dt) = 0$, $(dI_t)^2 = 0$ and $(dI_t)(d\mathbf{x}_t) = \mathbf{0}$. This SDE does not include the no-arbitrage condition yet. Using $d\tilde{\mathbf{W}}_{x,t} = d\mathbf{W}_{x,t} + \boldsymbol{\gamma}_{x,t}$, it holds under $\mathbb{Q}$

$$\begin{aligned} dVS_t = &\left(\frac{\partial VS_t}{\partial t} + \sigma_t^2\frac{\partial VS_t}{\partial I_t} + \frac{\partial VS_t}{\partial \mathbf{x}_t{'}}\left[\mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t) - \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t\boldsymbol{\gamma}_{x,t}\right] + \tfrac{1}{2}\operatorname{tr}\left(\boldsymbol{\Lambda}_t\boldsymbol{\Sigma}'\frac{\partial^2 VS_t}{\partial \mathbf{x}_t \partial \mathbf{x}_t{'}}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_t\right)\right)dt \\ &+ \frac{\partial VS_t}{\partial \mathbf{x}_t{'}}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\tilde{\mathbf{W}}_{x,t}. \qquad (\mathbb{Q}) \qquad \text{(c.2)} \end{aligned}$$

The discounted variance-swap value follows under $\mathbb{Q}$,

$$\begin{aligned} d(B_t^{-1}VS_t) &= B_t^{-1}\left[dVS_t - r_t\, VS_t dt\right] \\ &= B_t^{-1}\left\{\left[(\ldots) - r_t\, VS_t\right]dt + \frac{\partial VS_t}{\partial \mathbf{x}_t{'}}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\tilde{\mathbf{W}}_{x,t}\right\}, \qquad (\mathbb{Q}) \qquad \text{(c.3)} \end{aligned}$$

in which the expression for the dots is given by the drift function in (c.2). Absence of arbitrage requires the drift of $\{B_t^{-1}VS_t\}$ to be zero under $\mathbb{Q}$. This leads to the following no-arbitrage restriction

$$r_t\, VS_t = \frac{\partial VS_t}{\partial t} + \sigma_t^2\frac{\partial VS_t}{\partial I_t} + \frac{\partial VS_t}{\partial \mathbf{x}_t{'}}\left[\mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t) - \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t\boldsymbol{\gamma}_{x,t}\right] + \tfrac{1}{2}\operatorname{tr}\left(\boldsymbol{\Lambda}_t\boldsymbol{\Sigma}'\frac{\partial^2 VS_t}{\partial \mathbf{x}_t \partial \mathbf{x}_t{'}}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_t\right). \tag{c.4}$$

As in section 3.1, this restriction implicitly defines a parabolic PDE that the pricing function $VS(t, I_t, \mathbf{x}_t)$ satisfies, with boundary condition $VS_T = A[\bar{\sigma}^2 - SW]$. From (c.2), the *arbitrage-free variance-swap value* thus follows under $\mathbb{Q}$,

[69] Notice that the covariance after the second equality is of the form $\operatorname{cov}[\mathbf{a}'\mathbf{A}\mathbf{y}, \mathbf{A}\mathbf{y}]$, with $\mathbf{a}$ and $\mathbf{A}$ a vector and matrix of constants respectively, and $\mathbf{y}$ a random vector with $\mathbb{E}[\mathbf{y}] = \mathbf{0}$. The third equality then follows from $\operatorname{cov}[\mathbf{a}'\mathbf{A}\mathbf{y}, \mathbf{A}\mathbf{y}] = \operatorname{cov}[\mathbf{A}\mathbf{y}, \mathbf{y}'\mathbf{A}'\mathbf{a}] = \mathbf{A}\mathbb{E}[\mathbf{y}\mathbf{y}']\mathbf{A}'\mathbf{a}$.

$$dVS_t = r_t VS_t dt + \frac{\partial VS_t}{\partial \mathbf{x}_t'} \mathbf{\Sigma \Lambda}_t d\tilde{\mathbf{W}}_{x,t} , \qquad (\mathbb{Q}) \qquad \text{(c.5)}$$

and under the market measure (using $d\tilde{\mathbf{W}}_{x,t} = d\mathbf{W}_{x,t} + \boldsymbol{\gamma}_{x,t}$),

$$dVS_t = \left[r_t + \left(\frac{1}{VS_t} \frac{\partial VS_t}{\partial \mathbf{x}_t'} \right) \mathbf{\Sigma \Lambda}_t \boldsymbol{\gamma}_{x,t} \right] VS_t dt + \frac{\partial VS_t}{\partial \mathbf{x}_t'} \mathbf{\Sigma \Lambda}_t d\mathbf{W}_{x,t} . \qquad (\mathbb{P}) \qquad \text{(c.6)}$$

Using equation (b.6), and following a similar analysis that leads to equation (b.7), shows that the volatility beta of the variance swap is given by $\boldsymbol{\beta}_{VS,x,t} = (1/VS_t)(\partial VS_t / \partial \mathbf{x}_t)$. The expected spot variance-swap return thus equals

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dVS_t}{VS_t} \mid \mathcal{F}_t \right] = \left[r_t + \boldsymbol{\beta}_{VS,x,t}'\left(\mathbf{\Sigma \Lambda}_t \boldsymbol{\gamma}_{x,t} \right) \right] dt = \left[r_t + \boldsymbol{\beta}_{VS,x,t}'\left(\mathbf{\Sigma \Lambda}_t^2 \boldsymbol{\gamma} \right) \right] dt . \qquad \text{(c.7)}$$

- III -

A State Space Approach to the Estimation of the Multifactor Affine Stochastic Volatility Derivative Pricing Model

1. Introduction

In this chapter we propose a method for estimating the multifactor affine SV derivative pricing model. Given a sample of stock and option data that is generated (or believed to be generated) by this model, we show how this sample may be used for parameter estimation and for unveiling the hidden volatilities underlying the data. We adopt a *state space* approach and base our technique on the *Kalman filter* and *smoother*.

The chapter proceeds as follows. Section 2 discusses current estimation methods of SV models, with a specific focus on recently developed methods that combine stock and option data for estimation. We motivate why we advocate a state space approach, and provide a review of the linear state space framework.

Section 3 derives a discrete-time linear state space representation from the multifactor affine SV derivative pricing model. We consider ways to extract information from stock returns, *realized volatilities* and call option prices. At several stages during the discussion we connect to the earlier literature and provide comparisons and explain differences. GARCH-type models and discrete-time SV models with associated state space estimation methods get specific attention. We also provide an overview of different strands of research towards a recent novelty in financial econometrics, *realized volatility*. We propose a fast method for extracting information from option data that circumvents Monte Carlo simulation during estimation. Preliminary insight into the performance of the method is provided. Several possible state space models eventually result; we explain the differences and argue why we prefer one model above the other.

Section 4 contains concluding remarks on estimation, parameter identification, and the inclusion of put options and multiple option series in the analysis. Some first remarks on more explicitly incorporating level-dependent volatility-of-volatility in the state space framework are given as well. An appendix completes this chapter.

2. Current estimation methods of SV models and the state space framework

This section gives an overview of current estimation methods of SV models, and summarizes the linear state space framework. The SV literature is huge and still expanding. The largest part of this literature deals with SV in pure stock price models, rather than option pricing models. As a result, typically only stock price data is used for estimation. Section 2.1 discusses some general issues and complications that arise in estimating SV models. We consider current approaches, with an emphasis on methods that combine stock and option data for estimating SV option pricing models, as these are most relevant for our purposes. These techniques have only recently been developed. We discuss drawbacks of these methods, and motivate why we advocate a state space approach. Section 2.2 reviews the linear state space framework, the Kalman filter and smoother, QML estimation, and diagnostic checking in state space models.

2.1 Current methods and motivating the state space approach

Recall the formula for the price of a European call option, as implied by our multifactor model: $C_t = \mathbb{E}_{\mathbb{Q}}[BS(S_t, K, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t]$. This price may be computed by Monte Carlo simulation, if the model parameters are known. In practice however, these parameters are unknown, and must therefore first be estimated from real-world data. Both stock and option prices contain complementary information in this respect. Stock prices contain information on the stock's real-world ($\mathbb{P}$) distribution. Option prices in contrast, accommodate information on the stock's risk-neutral ($\mathbb{Q}$) distribution, and in particular, on the market price of volatility risk, which is a necessary input for pricing derivatives.

Combining both sources of information for estimation is complicated however, for a number of reasons. As volatility is unobserved, no closed-form expression for the likelihood function exists (i.e., the *transition density* does not exist in closed-form). Therefore, the volatility essentially needs to be "integrated out" during estimation. Moreover, as option prices can only be obtained by simulation and/or numerical integration, computationally very demanding procedures result, which make a fast practical implementation infeasible (so it seems).

As a consequence, the earlier literature on estimating SV option pricing models has mainly focused on either the time-series information in stock prices, or the cross-section information in option prices, to estimate the diffusion parameters, but not both. [1] A drawback of using stock prices only is that the price of volatility

[1] Numerous estimation methods for SV models have been developed in the last decade, mainly in pure stock price settings. Early examples include various *(generalized) method of moments (GMM)*-based approaches, as employed in e.g. Scott (1987) and Taylor (1994). For many SV models however, a sufficient number of moments cannot be derived, such that a straightforward application of (G)MM is not always possible. This has led researchers as e.g. Wiggins (1987), Chesney and Scott (1989), Melino and Turnbull (1990), and notably Duffie and Singleton (1993), to simulate the unknown moments, and then apply GMM. This method is known as *simulated method of moments*. Shephard (1996) lists a large number of drawbacks of GMM for estimating SV models. Importantly, GMM does not deliver a volatility forecast, such that another technique is required for this. Moreover, it is commonly known that GMM can have poor finite-sample properties, especially when the moment conditions are strongly dependent. Improvements to GMM are discussed in Van der Sluis (1999), and include the *efficient method of moments* of Gallant and Tauchen (1996). Other estimation techniques for SV models include the *indirect inference procedure* of Gourieroux et al. (1993), and various *Markov Chain Monte Carlo methods* (e.g., Jacquier et al. (1994) and Kim et al. (1998)). Another branch of estimation methods that uses stock

risk cannot be estimated. A commonly employed estimation method that uses option data only, is (daily re-) *calibration*. [2] Calibration estimates the $\mathbb{Q}$-parameters by minimizing the sum of squared deviations between the theoretical and market-observed option prices (and thus requires simulation) of a cross-section of options observed on a particular day. Although popular, this technique disregards the time-series dimension of the model completely, and is therefore inherently inconsistent with the dynamic principles of the model. Moreover, calibrated estimates depend on the information available on that particular day only; past information is not included. This does not seem a wise strategy to pursue in practice. Suppose for example, that today is an "extraordinary" day, leading to calibrated "extraordinary" estimates that differ much from the estimates obtained from exploiting the information in a full time series, i.e., the "time-series" estimates. If the calibrated model is subsequently used for pricing derivatives, it may well be the case that the resulting price differs greatly from the price obtained from the time-series estimates, which seems more reliable. Possible losses therefore seem more easily incurred –and hedging less reliable- if calibration is used, rather than an estimation strategy that uses time-series data.

2.1.1 Combining stock and option data for estimation

Combining stock and option data jointly in a dynamically consistent manner is important for a number of reasons (besides those just mentioned). First, the real-world and risk-neutral stock price dynamics can be estimated simultaneously, and the market price of volatility risk can be isolated. Second, it is likely to result in more efficient parameter estimates as more information is used. This is important for more reliable pricing and hedging of exotic over-the-counter derivative products by a financial institution. Third, volatility forecasts may improve. [3] This is important for risk management purposes; e.g., the reliability of a measure as *Value-at-Risk* may be enhanced. If volatility is estimated from stock prices the estimate is based on historical data only. In contrast, *implied volatilities* obtained from option prices are generally considered as being market forecasts of future volatility in practice (see e.g. Hull (2003)). As such, combining both sources of volatility information may prove of additional value for predicting volatility.

In recent years, progress has been made in combining stock and option data for the purpose of parameter estimation in 1-factor SV option pricing models. We

---

prices only, explicitly recognizes that a latent process is driving the volatility, subsequently writes the model in *linear state space form*, and performs QML estimation. We refer to the main text and section 3.2.1 for a discussion. However, as discrete-time SV models can often be considered as *non-linear state space* models, other state space methods for reliable and efficient estimation have been developed; see e.g. Durbin and Koopman (1997, 2000, 2001). Danielsson (1994) adopts a *simulated maximum likelihood* (SML) estimation approach, and Danielsson (1998) extends this to the multivariate discrete-time SV case. Durham and Gallant (2002) also focus on a SML approach, but then for continuous-time SV models. Still, this is not a complete survey of all existing estimation methods. Earlier overviews include Taylor (1994), Shephard (1996) and Ghysels et al. (1996). See also Van der Sluis (1999), and the introduction of Bollerslev and Zhou (2002). Recent developments in SV models are discussed in e.g. Barndorff-Nielsen, Nicolato and Shephard (2002), Tauchen (2004) and Shephard (2005). See also Neil Shephard's homepage for numerous articles on both SV (and RV): http://www.nuff.ox.ac.uk/users/shephard/. Studies that compare the information in stock and option prices are summarized in Bates (1996). Bates (2003) and Garcia, Ghysels and Renault (2003) provide overviews of option pricing research.

[2] Bates (1996b, 2000), Bakshi et al. (1997), Duffie et al. (2000) among others. Calibration is also popular in banking practice.

[3] A recent survey on volatility forecasting is Poon and Granger (2003). See Poon (2005) for a practical guide on volatility forecasting.

briefly discuss the estimation methods of Chernov and Ghysels (2000), Pan (2002) and Jones (2003) below.

Chernov and Ghysels (2000)

Chernov and Ghysels (2000) extend the *efficient method of moments* (EMM; Gallant and Tauchen (1996)) to estimate the Heston (1993) model, using both daily return and short-maturity ATM option data on the S&P500 index. EMM matches the scores of the likelihood function of an auxiliary model via simulation. EMM is asymptotically as efficient as maximum likelihood, if the auxiliary model is a good approximation to the distribution of the data. Chernov and Ghysels use *reprojection* to filter the latent volatilities. They do not perform a Monte Carlo study to investigate the finite-sample performance of their estimation method.

Pan (2002)

Pan (2002) proposes the *implied-state GMM* (IS-GMM) approach to estimate the parameters of the Heston (1993) and Bates (2000) models. [4] She uses weekly data on the S&P500 index and a short-dated near-the-money option series for estimation. The basic idea of IS-GMM is first to back out a proxy for the unobserved volatility from an option price, by inverting the option-pricing formula implied by the model (for a given set of parameter values). Using this proxy, one can directly focus on the dynamics of the state variables, being the stock price and variance. As the affine structure of the log-stock price and variance in both models allows for a joint closed-form conditional moment-generating function of stock returns and variance, moment conditions can be derived. Replacing the true volatility with the proxy in these moment conditions, conventional GMM can next be applied. IS-GMM differs from GMM as one of the state variables (the stock variance) is parameter-dependent. Pan reports results of a small Monte Carlo study for one set of parameters that indicate that IS-GMM seems to work well.

Jones (2003)

Jones (2003) fits the Heston, CEV and 2GAM models [5] to joint daily time series of S&P100-index returns and the "market volatility index" VIX [6], using a *Bayesian perspective* and *Markov Chain Monte Carlo* (MCMC) methods. Jones's estimation strategy is briefly described as follows. To be able to extract information from option prices, Jones assumes a number of approximations. He first approximates the model-implied risk-neutral expected average variance over the life of the option (i.e., $\mathbb{E}_{\mathbb{Q}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]$ in our notation) by a tractable linear function of the current stock variance (σ_t^2). To implement this relation for estimation purposes, he next approximates $\mathbb{E}_{\mathbb{Q}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]$ by the observed Black-Scholes implied variance. To validate this approximation, Jones uses a Hull-White (1987)-type argument: If the stock and volatility process were independent, the option price would equal the $\mathbb{Q}$-expected Black-Scholes price, with the random $\bar{\sigma}_t^2$ as its argument. (Recall our option pricing formula.) Jones then argues that, since the BS formula is close to linear in its variance argument for short-term ATM options, approximating $\mathbb{E}_{\mathbb{Q}}[BS(\bar{\sigma}_t^2) \mid \mathcal{F}_t]$ by $BS(\mathbb{E}_{\mathbb{Q}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]$ may not result in large approximation errors, as the Jensen-inequality error will tend to be minor. Jones finally argues that the correlation between the stock and volatility processes is an

[4] The Bates (2000) model extends the Heston (1993) model to allow for jumps in stock prices.

[5] Jones (2003) develops the CEV (constant elasticity of variance) model, which is similar to the Heston model, but the square root in the volatility-of-volatility function is replaced by an arbitrary power that entails an extra free parameter. He also develops the 2GAM model. The cost of these generalizations is the loss of "closed-form" option pricing formulas.

[6] See http://www.cboe.com/micro/vix/faq.asp for details.

additional reason to invalidate the approximations, but says that the practical importance of correlation is unclear for ATM options. For a wide range of parameter values the validity of the approximation is verified, he mentions. To extract information from stock prices, Jones uses an Euler approximation of the stock price and variance processes. Jones does not perform a Monte Carlo study to investigate the performance of his estimation method. As will become apparent in section 3.4, the way Jones extracts information from option prices is related to our method. The (mathematical) foundations and justifications for our method are nonetheless more solid, given our specific derivative pricing model.

Drawbacks of current methods
Each of the estimation methods discussed above has its drawbacks. EMM for example, is not very transparent, not easy to implement, is simulation-based and computationally very demanding. Tauchen (2004) for example mentions "*EMM requires very long simulations to obtain accurate numerical integrations* [..]". EMM does not generate a direct volatility forecast either; another method, reprojection, is needed for this. Besides, although EMM has desirable asymptotic properties, accumulating evidence from empirical studies indicates that the finite-sample behavior of this estimation method can be very poor. Chernov and Ghysels (2000) mention that '*the precision of the estimates is very poor* [..]', and that their result is consistent with previous findings as in e.g. Gallant, Hsieh and Tauchen (1997). Duffee and Stanton (2001), who apply EMM to estimate models of the term structure of interest rates, find that '*EMM behaves extremely poorly in samples of the size and type usual in term structure estimation* [..]'. When the dimensionality of the problem increases (i.e. multiple SV factors or including more option series), we expect that problems with EMM are even more likely to occur.

Besides seeming computationally demanding, a major drawback of Pan's method is that IS-GMM assumes the option series from which to back out the volatility proxy, to be measured without error. Different selected options will therefore most likely give different values for the volatility proxy, due to model error. So what option to use? As a model is never a complete description of reality, allowing for measurement error is important. [7] The Bayesian, MCMC estimation method of Jones (2003) is apparently computationally very intensive as well: Jones mentions that 600 000 iterations (!) of his estimation algorithms are needed to estimate the parameters using the joint data.

2.1.2 Our approach and advantages of our approach

Given the disadvantages of each of the estimation methods considered above, our aim is to develop a different, QML-based estimation technique that overcomes most of these drawbacks. The method we suggest is based on a *state space* approach of the problem, and the *Kalman filter*. Kalman filter-based QML estimation approaches for SV stock price models (thus not option pricing models) have previously been considered by e.g., Ruiz (1994), Harvey et al. (1994), Harvey and Shephard (1996), and recently by Alizadeh et al. (2002). (We refer to section 3.2.1 for a discussion.) Moreover, state space methods are commonly employed in the literature on the term structure of interest rates, to estimate a wide range of different interest rate models. State space estimation has not been applied in SV option pricing contexts yet (at least, to the best of our knowledge), the main reason probably being that it is unclear how to use this approach in this setting. This is in sharp contrast to an affine interest-rate model setting for

[7] See Renault (1997) for a discussion on the importance of option pricing errors.

example, where it is a natural way to proceed. Indeed, the inspiration for our method (and multifactor SV model as well) was found in the articles on affine term structure models, introduced by Duffie and Kan (1996), and explored in e.g. Lund (1997), de Jong (2000), Dai and Singleton (2000) and Duffee and Stanton (2001).

Using a number of approximations, in section 3 we derive a discrete-time linear state space model (with time-varying coefficients) from our continuous-time multifactor affine SV derivative pricing model, that can be estimated by Kalman filter QML. The benefits of Kalman filter QML are its relative simplicity and consistency of estimates; a drawback is inefficiency of the estimates.

The advantages of our approach include the following. *First*, the state space framework treats the multiple, unobserved volatility-driving factors as latent objects. It is therefore naturally suited for incorporation of these hidden factors. *Second*, the smoothed evolution of the factors and volatility is a direct by-product of the estimation output. Volatility forecast can be generated at the same time. *Third*, the joint simultaneous inclusion of time series on the underlying stock and several option series is easily dealt with, due to the panel-data (i.e. time-series cross-section) character of state space models. *Fourth*, our method naturally allows for measurement error in all option series, which may be caused by e.g. market microstructure effects or differences in liquidity. *Fifth*, our method circumvents simulation of option prices during estimation. Convergence is rapidly achieved. Our method is fast. This makes it particularly useful for practical (banking) purposes. *Sixth*, in the literature on the term structure of interest rates, Kalman filter QML estimation methods have proven to be rather robust (Lund (1997), de Jong (2000) and Duffee and Stanton (2001)). These methods tend to perform well in finite samples, even though some of the different sub-methods deliver inconsistent estimates. The papers just mentioned contain Monte Carlo evidence showing the "inconsistency" (bias) to be small in finite samples.

Before deriving a state space representation of our model, we first review the linear state space framework in the next section. For more extensive discussions see, e.g., Harvey (1989), Hamilton (1994) and Durbin and Koopman (2001).

2.2 A review of the linear state space representation, the Kalman filter and smoother, and QML estimation

A linear state space representation

We consider the following linear state space representation of a dynamic system, which is particularly suited for our needs. The *observation* or *measurement equation* reads

$$\boldsymbol{y}_t = \boldsymbol{a}_t + \boldsymbol{H}_t{}'\boldsymbol{\xi}_t + \boldsymbol{w}_t\,; \qquad \boldsymbol{w}_t \sim (\mathbf{0}, \boldsymbol{R})\,; \qquad t = 1,..,T\,. \qquad (2.1)$$

Here, $(mx1)\boldsymbol{y}_t$ is a vector of observed variables at time t, where time is measured in some time unit. It is described in terms of a *state vector* $(rx1)\ \boldsymbol{\xi}_t$, which r state variables may be (partly) unobserved, and a vector white noise error term $(mx1)\ \boldsymbol{w}_t$. The *transition* or *state equation* describes the evolution of the state over time. The state is assumed to evolve as a vector autoregressive process of order 1, i.e. VAR(1). It reads

$$\boldsymbol{\xi}_{t+1} = \boldsymbol{d} + \boldsymbol{F}\boldsymbol{\xi}_t + \boldsymbol{v}_{t+1}\,; \qquad \boldsymbol{v}_{t+1} \sim (\mathbf{0},\ \boldsymbol{Q})\,, \qquad (2.2)$$

in which $\boldsymbol{v}_{t+1}$ represents the time $t+1$ state innovation. The innovation series $\{\boldsymbol{v}_{t+1}\}$ behaves like white noise. The system matrices $(mx1)\,\boldsymbol{a}_t$ and $(mxr)\,\boldsymbol{H}_t{}'$, $(mxm)\,\boldsymbol{R}$, $(rx1)\boldsymbol{d}$, $(rxr)\,\boldsymbol{F}$, $(rxr)\boldsymbol{Q}$ contain parameters. The parameters in $\boldsymbol{a}_t$ and $\boldsymbol{H}_t{}'$ are allowed to change over time, although only in a deterministic way. The disturbance series $\{\boldsymbol{w}_t\}$ and $\{\boldsymbol{v}_{t+1}\}$ are mutually uncorrelated at all time points:

$$\text{cov}\left[\boldsymbol{w}_t, \boldsymbol{v}_{s+1}\right] = \mathbb{E}\left[\boldsymbol{w}_t \boldsymbol{v}_{s+1}{}'\right] = \mathbf{0} \quad \forall t, s = 1,..,T. \tag{2.3}$$

Finally, it is assumed that the initial state $\boldsymbol{\xi}_1$ is uncorrelated with the series $\{\boldsymbol{v}_{t+1}\}$, such that $\mathbb{E}[\boldsymbol{v}_{t+1}\boldsymbol{\xi}_1{}'] = \mathbf{0}$. Given these assumptions, it follows that the state innovation $\boldsymbol{v}_{t+1}$ is uncorrelated with lagged values of the state. That is $\mathbb{E}[\boldsymbol{v}_{t+1}\boldsymbol{\xi}_{t+1-s}{}'] = \mathbf{0};\ s = 1,..,t$.

The Kalman filter and smoother

The representation above can be analyzed by the *Kalman filter* and *smoother*, assuming that the numerical values of $\boldsymbol{a}_t$, $\boldsymbol{H}_t$, $\boldsymbol{R}$, $\boldsymbol{d}$, $\boldsymbol{F}$ and $\boldsymbol{Q}$ are known. [8] In particular, we are interested in the properties of the unobserved state variables in $\boldsymbol{\xi}_t$. If both the initial state $\boldsymbol{\xi}_1$ and the series $\{\boldsymbol{w}_t, \boldsymbol{v}_{t+1}\}$ are multivariate Gaussian [9], the Kalman filter may be motivated as a recursive algorithm for calculating the conditional mean and variance of the state at time $t+1$, given the information set $\boldsymbol{Y}_t \equiv (\boldsymbol{y}_1{}',..,\boldsymbol{y}_t{}')'$ at time t, i.e.

$$\boldsymbol{\xi}_{t+1|t} \equiv \mathbb{E}\left[\boldsymbol{\xi}_{t+1} \mid \boldsymbol{Y}_t\right], \qquad \boldsymbol{P}_{t+1|t} \equiv \text{var}\left[\boldsymbol{\xi}_{t+1} \mid \boldsymbol{Y}_t\right], \tag{2.4}$$

for $t = 2,3,...,T$ in succession, given some initial $\boldsymbol{\xi}_{1|0}$ and $\boldsymbol{P}_{1|0}$. In the Gaussian case it thus holds that $\boldsymbol{\xi}_{t+1} \mid \boldsymbol{Y}_t \sim \mathcal{N}(\boldsymbol{\xi}_{t+1|t}, \boldsymbol{P}_{t+1|t})$. $\boldsymbol{\xi}_{t+1|t}$ is labeled the *filtered state* and $\boldsymbol{P}_{t+1|t}$ the *filtered state variance matrix*. Provided that the eigenvalues of $\boldsymbol{F}$ are inside the unit circle so that the process for $\boldsymbol{\xi}_t$ is covariance-stationary, the Kalman filter can be started with the unconditional mean and variance of $\boldsymbol{\xi}_1$,

$$\begin{aligned} \boldsymbol{\xi}_{1|0} &= \mathbb{E}\left[\boldsymbol{\xi}_1\right] = (\mathbf{I}_r - \boldsymbol{F})^{-1}\boldsymbol{d} \\ \text{vec}\left(\boldsymbol{P}_{1|0}\right) &= \text{vec}\left(\text{var}\left[\boldsymbol{\xi}_1\right]\right) = \left[\mathbf{I}_{r^2} - (\boldsymbol{F} \otimes \boldsymbol{F})\right]^{-1} \text{vec}(\boldsymbol{Q}). \end{aligned} \tag{2.5}$$

We then iterate on

$$\begin{aligned} \boldsymbol{e}_t &= \boldsymbol{y}_t - \boldsymbol{a}_t - \boldsymbol{H}_t{}'\boldsymbol{\xi}_{t|t-1}, & \boldsymbol{E}_t &= \boldsymbol{H}_t{}'\boldsymbol{P}_{t|t-1}\boldsymbol{H}_t + \boldsymbol{R}, \\ \boldsymbol{K}_t &= \boldsymbol{F}\boldsymbol{P}_{t|t-1}\boldsymbol{H}_t\boldsymbol{E}_t^{-1}, & \boldsymbol{L}_t &= \boldsymbol{F} - \boldsymbol{K}_t\boldsymbol{H}_t{}', \\ \boldsymbol{\xi}_{t+1|t} &= \boldsymbol{d} + \boldsymbol{F}\boldsymbol{\xi}_{t|t-1} + \boldsymbol{K}_t\boldsymbol{e}_t, & \boldsymbol{P}_{t+1|t} &= \boldsymbol{F}\boldsymbol{P}_{t|t-1}\boldsymbol{L}_t{}' + \boldsymbol{Q}. \qquad t = 1,..,T. \end{aligned} \tag{2.6}$$

Here $(mx1)\,\boldsymbol{e}_t = \boldsymbol{y}_t - \mathbb{E}[\boldsymbol{y}_t \mid \boldsymbol{Y}_{t-1}]$ is the *prediction error* or *innovation* in $\boldsymbol{y}_t$ given $\boldsymbol{Y}_{t-1}$, and $(mxm)\,\boldsymbol{E}_t = \text{var}[\boldsymbol{e}_t]$ is the *prediction error variance*. The (rxm) matrix $\boldsymbol{K}_t$ is known as the time-t *gain matrix*. The equations for $\boldsymbol{\xi}_{t+1|t}$ and $\boldsymbol{P}_{t+1|t}$ are known as the *prediction equations*. Notice that the current filtered state is essentially computed as a weighted sum of the most recently filtered state, and the forecast error in the observed variables.

---

[8] The Kalman filter originated in the engineering literature, in the path-breaking work of Kalman (1960).

[9] In that case subsets of variables, given other subsets of variables, are all normally distributed; e.g., $\boldsymbol{\xi}_{t+1} \mid \boldsymbol{Y}$ is Gaussian as well.

The Kalman filter can be used for producing a forecast (or estimate) of the future state, given *current* data. In contrast, the *Kalman smoother* is particularly suited for obtaining an estimate of the state, given *all* observations. That is, for the *smoothed state* $\boldsymbol{\xi}_{t|T} \equiv \mathbb{E}[\boldsymbol{\xi}_t \mid \boldsymbol{Y}_T]$, and the *smoothed state variance matrix* $\boldsymbol{P}_{t|T} \equiv \text{var}[\boldsymbol{\xi}_t \mid \boldsymbol{Y}_T]$; $t = 1,..,T$. They can be obtained from a backwards recursion which uses the output from the Kalman filter. The recursion starts with $(rx1)\boldsymbol{r}_T = \mathbf{0}$, $(rxr)\boldsymbol{N}_T = \mathbf{0}$, and subsequently reads

$$\begin{aligned} \boldsymbol{r}_{t-1} &= \boldsymbol{H}_t \boldsymbol{E}_t^{-1} \boldsymbol{e}_t + \boldsymbol{L}_t{}' \boldsymbol{r}_t, & \boldsymbol{N}_{t-1} &= \boldsymbol{H}_t \boldsymbol{E}_t^{-1} \boldsymbol{H}_t{}' + \boldsymbol{L}_t{}' \boldsymbol{N}_t \boldsymbol{L}_t, \\ \boldsymbol{\xi}_{t|T} &= \boldsymbol{\xi}_{t|t-1} + \boldsymbol{P}_{t|t-1} \boldsymbol{r}_{t-1}, & \boldsymbol{P}_{t|T} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1} \boldsymbol{N}_{t-1} \boldsymbol{P}_{t|t-1}, \end{aligned} \tag{2.7}$$

for $t = T,..,1$.

In the Gaussian case, the filtered state $\boldsymbol{\xi}_{t+1|t}$ is the *minimum mean squared error (MMSE) estimator* of the state $\boldsymbol{\xi}_{t+1}$ given time-t information, whereas the smoothed state $\boldsymbol{\xi}_{t|T}$ is the MMSE estimator of the state $\boldsymbol{\xi}_t$ given the full information set $\boldsymbol{Y}_T$. The Kalman filter and smoother recursions then provide us with conditional means and variances of the state vector.

When the normality assumption does not hold this is no longer true. However, as discussed in e.g. Harvey (1989) and Hamilton (1994), in the non-Gaussian case the same Kalman filter and smoother recursions as above can still be used to compute *linear least squares forecasts* of $\boldsymbol{\xi}_t$ given $\boldsymbol{Y}_{t-1}$ (filtering), or $\boldsymbol{Y}_T$ (smoothing), together with the mean squared error matrices of their respective estimation errors. The Kalman filter then produces the *linear projection* of $\boldsymbol{\xi}_{t+1}$ on $\boldsymbol{Y}_t$ and a constant, as denoted by $\hat{\boldsymbol{\xi}}_{t+1|t} = \hat{\mathbb{E}}[\boldsymbol{\xi}_{t+1} \mid \boldsymbol{Y}_t]$ in Hamilton (1994), together with MSE matrix $\hat{\boldsymbol{P}}_{t+1|t} \equiv \mathbb{E}[(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t})']$ of the associated estimation error. The Kalman smoother then produces the linear projection $\hat{\boldsymbol{\xi}}_{t|T}$ with associated MSE matrix $\hat{\boldsymbol{P}}_{t|T}$ of the estimation error. In the non-Gaussian case, the state estimators $\hat{\boldsymbol{\xi}}_{t+1|t}$ (the filtered state) and $\hat{\boldsymbol{\xi}}_{t|T}$ (the smoothed state) are still optimal, in the sense that they minimize the MSE within the class of all *linear* estimators of the state vector. [10]

Estimation

In practice, the numerical values of the parameters appearing in $\boldsymbol{a}_t$, $\boldsymbol{H}_t$, $\boldsymbol{R}$, $\boldsymbol{d}$, $\boldsymbol{F}$ and $\boldsymbol{Q}$ are typically unknown, and must be estimated from the observed data $\boldsymbol{Y}_T$ at hand. If the initial state $\boldsymbol{\xi}_1$ and the series $\{\boldsymbol{w}_t, \boldsymbol{v}_{t+1}\}$ are multivariate Gaussian, then the distribution of $\boldsymbol{y}_t$ conditional on $\boldsymbol{Y}_{t-1}$ is Gaussian,

$$\boldsymbol{y}_t \mid \boldsymbol{Y}_{t-1} \sim \mathcal{N}\left[\boldsymbol{a}_t + \boldsymbol{H}_t{}' \boldsymbol{\xi}_{t|t-1},\, \boldsymbol{E}_t\right], \tag{2.9}$$

from which the sample loglikelihood is constructed in a straightforward way. The loglikelihood can subsequently be maximized with respect to the unknown parameters, invoking the Kalman filter recursion. Kalman filter ML estimation yields consistent, asymptotically normal, and (asymptotically) efficient estimates. It should be noted however, that in the absence of sufficient restrictions on the parameters, not all parameters might be identified.

[10] The MMSE estimator is still the conditional expectation (as always), but is now generally a non-linear function of the data $\boldsymbol{Y}$ and much more tedious to compute; in general by means of simulation. Recall that in case of Gaussianity $\boldsymbol{\xi}_{t+1|t}$ and $\boldsymbol{\xi}_{t|T}$ have minimal MSE of *all* estimators of the state, both linear and non-linear.

In the non-Gaussian case, the same loglikelihood (with $\boldsymbol{\xi}_{t|t-1}$ replaced by $\hat{\boldsymbol{\xi}}_{t|t-1}$ and $\boldsymbol{P}_{t|t-1}$ by $\hat{\boldsymbol{P}}_{t|t-1}$) may be maximized. This *quasi-maximum likelihood* (QML) estimation procedure still yields consistent and asymptotically normal (but inefficient) estimates of the parameters, provided that some mild regularity conditions are satisfied (see Watson (1989)). [11] Robust QML standard errors can be computed with the method of White (1982); see e.g. Hamilton (1994, p.145).

Diagnostic checking

Once the parameters have been estimated, correct econometric practice is to subject the estimated model to a number of diagnostic checks. Given the estimation output, the *standardized prediction errors* or *innovations* may be computed. These are defined as $\boldsymbol{E}_t^{-\frac{1}{2}}\boldsymbol{e}_t$, and equal

$$\left(\boldsymbol{H}_t{}'\boldsymbol{P}_{t|t-1}\boldsymbol{H}_t+\boldsymbol{R}\right)^{-\frac{1}{2}}\left(\boldsymbol{y}_t-\mathbb{E}\left[\boldsymbol{y}_t \mid \boldsymbol{Y}_{t-1}\right]\right)=\left(\boldsymbol{H}_t{}'\boldsymbol{P}_{t|t-1}\boldsymbol{H}_t+\boldsymbol{R}\right)^{-\frac{1}{2}}\left(\boldsymbol{y}_t-\boldsymbol{a}_t-\boldsymbol{H}_t{}'\boldsymbol{\xi}_{t|t-1}\right) \tag{2.10}$$

If the model is well specified, the standardized innovations should form a mean-zero, homoskedastic, serially uncorrelated series (hence white noise), with a variance equal to 1.

The so-called *smoothed disturbances* or *auxiliary residuals* of the state space model yield additional information with respect to model misspecification. They represent the best (i.e., MMSE) estimates of the disturbances given the data. The smoothed disturbances associated with the measurement equation are given by

$$\mathbb{E}[\boldsymbol{w}_t \mid \boldsymbol{Y}_T]=\boldsymbol{y}_t-\boldsymbol{a}_t-\boldsymbol{H}_t{}'\boldsymbol{\xi}_{t|T}, \tag{2.11}$$

whereas for the transition equation, they are given by

$$\mathbb{E}[\boldsymbol{v}_{t+1} \mid \boldsymbol{Y}_T]=\boldsymbol{\xi}_{t+1|T}-\boldsymbol{d}-\boldsymbol{F}\boldsymbol{\xi}_{t|T}. \tag{2.12}$$

Durbin and Koopman (2001) point out that the auxiliary residuals are autocorrelated (in theory), but can be useful for detecting outliers and structural breaks, as they are the best estimators of the error terms given the data.

3. Deriving a state space representation

This section derives a linear state space representation from the multifactor affine SV derivative pricing model, which can be used for parameter and volatility estimation. Section 3.1 derives an exact discrete-time equivalent of the SDE for the latent factors. Section 3.2 explains how to extract information from daily stock returns. For illustrative purposes, we link to the earlier literature on estimating SV models using the Kalman filter. We also compare the 1-factor SV special case of the multifactor model with the GARCH(1,1) model for stock returns. Section 3.3 designs a strategy for extracting information from high-frequency intraday stock price data, via *realized volatilities*. Section 3.4 considers information extraction from a time series of call option prices. Section 3.5 comments on two possible resulting state space models, of which one is eventually chosen to work with in the remainder.

[11] Durbin and Koopman (2001) provide an extensive treatment and simulation-based estimation methodology for non-Gaussian state space models in which the distributions of the measurement and transition equation disturbances are known. Unfortunately, these distributions are unknown in all cases we will encounter, such that we resort to QML estimation.

In the coming sections, we need the following SDEs and expressions that appear in the multifactor affine SV option pricing model (see sections 2 and 3.2 of the previous chapter):

Stock price: $$dS_t = \mu_t S_t dt + \sigma_t S_t dW_{S,t} \qquad (\mathbb{P}) \qquad (3.1)$$

Stock variance: $$\sigma_t^2 = \delta_0 + \boldsymbol{\delta}' \mathbf{x}_t \qquad (3.2)$$

Latent factors: $$d\mathbf{x}_t = \mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t)dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t} \qquad (\mathbb{P}) \qquad (3.3)$$

$$d\mathbf{x}_t = \tilde{\mathbf{K}}(\tilde{\boldsymbol{\theta}} - \mathbf{x}_t)dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\tilde{\mathbf{W}}_{x,t} \qquad (\mathbb{Q}) \qquad (3.4)$$

Call price: $$C_t = \mathbb{E}_{\mathbb{Q}}\left[BS(S_t, K, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t\right], \qquad (3.5)$$

with the risk-neutral parameters given by $\tilde{\mathbf{K}} = \mathbf{K}_d + \boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathcal{B}'$ and $\tilde{\boldsymbol{\theta}} = \tilde{\mathbf{K}}^{-1}(\mathbf{K}_d\boldsymbol{\theta} - \boldsymbol{\Sigma}\boldsymbol{\Gamma}\boldsymbol{\alpha})$, and where $\boldsymbol{\Lambda}_t = \text{diag}[\sqrt{\alpha_1 + \boldsymbol{\beta}_1' \mathbf{x}_t}, .., \sqrt{\alpha_n + \boldsymbol{\beta}_n' \mathbf{x}_t}]$.

Given this *continuous-time* model, our object is to translate the model into the *discrete-time* state space representation from section 2.2, which can be used for estimation purposes. We assume the empirical data to consist of T *daily* observations of stock returns, realized volatilities, and/or a (panel of) call option prices. We measure time t in trading years, with 260 days per annum. The timing of the data points is denoted by $t = \Delta t, 2\Delta t, .., T\Delta t$, with $\Delta t = 1/260$.

3.1 Tackling the SDE of the volatility-driving factors

The unobserved volatility-driving factors $\mathbf{x}$ will form (part of) the state of the state space model (2.1)-(2.2). To obtain a discrete-time equation from their SDE $d\mathbf{x}_t = \mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t)dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t}$ under $\mathbb{P}$, one could consider a (non-exact) Euler discretization. Fortunately, an exact discretization exists however, as we will now show. We first express $\mathbf{x}_{t+\Delta t}$ in terms of $\mathbf{x}_t$. To achieve this, consider the transformation

$$\mathbf{y}_t \equiv \exp[\mathbf{K}_d t]\mathbf{x}_t, \qquad (3.6)$$

where we define the exponent of a diagonal matrix as follows. If (nxn) $\mathbf{A}_d = \text{diag}[a_1, ..., a_n]$ then $\exp[\mathbf{A}_d] \equiv \text{diag}[\exp(a_1), ..., \exp(a_n)]$. Itô gives

$$d\mathbf{y}_t = \mathbf{K}_d \exp[\mathbf{K}_d t]\mathbf{x}_t dt + \exp[\mathbf{K}_d t]\, d\mathbf{x}_t = \exp(\mathbf{K}_d t)[\mathbf{K}_d\boldsymbol{\theta} dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t}]. \qquad (3.7)$$

Integrating this SDE over the interval $[t, t+\Delta t]$ yields

$$\mathbf{y}_{t+\Delta t} = \mathbf{y}_t + \int_t^{t+\Delta t} \exp[\mathbf{K}_d u]\mathbf{K}_d\boldsymbol{\theta} du + \int_t^{t+\Delta t} \exp[\mathbf{K}_d u]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u}. \qquad (3.8)$$

Transforming back to $\mathbf{x}_{t+\Delta t}$, by premultiplying with $\exp[-\mathbf{K}_d(t+\Delta t)]$, yields

$$(3.9)$$

$$\begin{aligned} \mathbf{x}_{t+\Delta t} &= \exp[-\mathbf{K}_d \Delta t]\mathbf{x}_t + \int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t+\Delta t - u)]\mathbf{K}_d\boldsymbol{\theta} du + \int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t+\Delta t - u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} \\ &= \exp[-\mathbf{K}_d \Delta t]\mathbf{x}_t + (\mathbf{I}_n - \exp[-\mathbf{K}_d \Delta t])\boldsymbol{\theta} + \int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t+\Delta t - u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u}. \end{aligned}$$

Defining the factors in deviation from their mean, $\mathbf{x}_t^*$, as

$$\mathbf{x}_t^* \equiv \mathbf{x}_t - \boldsymbol{\theta} \ \forall t, \tag{3.10}$$

and

$$\mathbf{u}_{t+\Delta t} \equiv \int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t+\Delta t-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u}, \qquad (\mathbb{P}) \tag{3.11}$$

this equation can be simplified towards

$$\mathbf{x}_{t+\Delta t}^* = \exp[-\mathbf{K}_d \Delta t]\mathbf{x}_t^* + \mathbf{u}_{t+\Delta t}, \quad t = \Delta t, .., T\Delta t. \qquad (\mathbb{P}) \tag{3.12}$$

This equation already has the basic form we are looking for, if we interpret the random vector $\mathbf{u}_{t+\Delta t}$ as an error term. The state vector $\boldsymbol{\xi}_t$ in the state space model (2.1)-(2.2) will (partly) be formed by the latent factors in deviation from their mean, $\mathbf{x}_t^* = \mathbf{x}_t - \boldsymbol{\theta}$. What remains to be checked, are the statistical properties of the series $\{\mathbf{u}_{t+\Delta t}\}$, $t = \Delta t, ..., T\Delta t$. The state space representation (2.1)-(2.2) requires this series to be white noise. In section 6 of appendix B we prove that this is indeed the case. Specifically, we find

$$\mathbf{u}_{t+\Delta t} \sim \left(\mathbf{0}, \mathbf{G}(\Delta t) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right) \qquad (\mathbb{P}) \tag{3.13}$$

with

$$\mathbf{M}_d \equiv \text{diag}\left[\alpha_1 + \boldsymbol{\beta}_1'\boldsymbol{\theta}, .., \alpha_n + \boldsymbol{\beta}_n'\boldsymbol{\theta}\right], \tag{3.14}$$

$$(nxn)\ \mathbf{G}(\Delta t) \text{ with } [\mathbf{G}(\Delta t)]_{ij} = \frac{1-\exp[-(k_i+k_j)\Delta t]}{k_i+k_j},$$

(see also appendix A for these definitions), and in which $\odot$ denotes the Hadamard product; i.e., element-by-element multiplication. As $\mathbb{E}_{\mathbb{P}}[\mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t] = \mathbf{0}$, it moreover follows that $\mathbf{u}_{t+\Delta t}$ is uncorrelated with the past $\{\mathbf{x}_t, \mathbf{x}_{t-\Delta t}, ..\}$, i.e. $\{\mathbf{u}_{t+\Delta t}\}$ represents a true innovation series. The errors in $\mathbf{u}_{t+\Delta t}$ are interpreted as the unpredictable shocks to the volatility factors from day t to day $t+\Delta t$.

The continuous-time process (3.3) thus implies that the factors evolve as a vector autoregressive process of order 1 (VAR(1)) in discrete time. Given that $\sigma_t^2 = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t = \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} + \boldsymbol{\delta}'\mathbf{x}_t^*$, it is clear that the stock variance is essentially modeled as a sum of n AR(1) processes, which may be correlated.

3.2 Extracting information from stock returns

Consider the SDE (3.1) for the stock price. An exact discrete-time discretization does not exist, not even for the log stock price (due to stochastic volatility). An Euler discretization over the interval $[t, t+\Delta t]$ yields [12]

$$r_{t+\Delta t} = \mu_t \Delta t + \sigma_t \eta_{t+\Delta t}, \qquad (\mathbb{P}) \tag{3.15}$$

in which the stock return $r_{t+\Delta t}$ and the Brownian increment $\eta_{t+\Delta t}$ are defined as

[12] This Euler discretization is not exact. Nevertheless, for simplicity, we write an equality sign (=) instead of an approximately-equal sign ($\approx$).

$$r_{t+\Delta t} \equiv \frac{S_{t+\Delta t} - S_t}{S_t}, \qquad \eta_{t+\Delta t} \equiv W_{S,t+\Delta t} - W_{S,t} \sim \text{i.i.d. } \mathcal{N}(0, \Delta t). \qquad (3.16)$$

Given time-t information, collected in the filtration $\mathcal{F}_t$, the returns are approximately normally distributed: [13]

$$r_{t+\Delta t} \mid \mathcal{F}_t \approx \mathcal{N}(\mu_t \Delta t, \sigma_t^2 \Delta t). \qquad (\mathbb{P}) \qquad (3.17)$$

The process $\{\sigma_t^2\}$ represents the (per-annum) *conditional variance* process of the stock returns. As $\{\sigma_t^2\}$ varies stochastically over time, periods of high and low volatility, featured by large and small stock returns (i.e., *volatility clustering*) are possible. The *unconditional variance* of the stock returns is constant, and given by $\mathbb{E}_{\mathbb{P}}[\sigma_t^2] = \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta}$, per annum. [14] We are ultimately interested in the volatility process. As volatility clustering is associated with temporal dependence in second-order central moments, our focus is on squared returns in deviation from their mean, instead of the raw returns themselves, as these are not very informative on this. Consider then the equation [15]

$$\frac{1}{\Delta t}\left(r_{t+\Delta t} - \mu_t \Delta t\right)^2 = \mathbb{E}_{\mathbb{P}}\left[\frac{1}{\Delta t}\left(r_{t+\Delta t} - \mu_t \Delta t\right)^2 \mid \mathcal{F}_t\right] + \omega_{t+\Delta t}, \qquad (\mathbb{P}) \qquad (3.18)$$

in which, by construction, the error term $\omega_{t+\Delta t}$ has (conditional) mean zero. The conditional expectation in (3.18) equals $\mathbb{E}_{\mathbb{P}}[\sigma_t^2 \eta_{t+\Delta t}^2 \mid \mathcal{F}_t] / \Delta t = \sigma_t^2\, \mathbb{E}_{\mathbb{P}}[\eta_{t+\Delta t}^2 \mid \mathcal{F}_t] / \Delta t$ $= \sigma_t^2$, in which the final equality uses the independent-increments property of Brownian motion. Substitution yields

$$\frac{1}{\Delta t}\left(r_{t+\Delta t} - \mu_t \Delta t\right)^2 = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t + \omega_{t+\Delta t}, \qquad \omega_{t+\Delta t} \sim (0, \sigma_\omega^2). \qquad (\mathbb{P}) \qquad (3.19)$$

This equation is linear in the latent factors. The state space representation (2.1)-(2.2) requires the error series $\{\omega_t\}$ to be white noise. We confirm this in the appendix where we also derive an explicit expression for σ_ω^2, assuming the Euler discretization to be exact. As this is not the case however, we treat σ_ω as a free parameter in the estimations. We moreover show the errors $\{\omega_t\}$ to be uncorrelated with the errors $\{\mathbf{u}_t\}$ in (3.12) at all points in time.

Equation (3.19) serves to extract information from a series of stock returns. If the conditional mean-return process $\{\mu_t\}$ is assumed to be constant, such that $\mu_t = \mu$, the sample average of returns may be substituted in (3.19) to make the left-hand side "observable". If not, a possible empirical strategy is to take *prewhitened* returns in (3.19). We refer to the empirical analysis in section 2.1 of chapter IV for a motivation and discussion.

---

[13] As the Euler discretization is not exact, this only holds by approximation. For $\mu_t = \mu$ and treating the Euler discretization as exact, it can moreover be shown that the returns are serially uncorrelated. In section 12 of appendix B, we show however that this only holds true by close approximation, due to SV.

[14] For the *instantaneous* stock return, it holds that $dS_t / S_t \mid \mathcal{F}_t \sim \mathcal{N}(\mu_t dt, \sigma_t^2 dt)$ exactly. It follows that $\text{var}[dS_t / S_t] = \mathbb{E}(\text{var}[dS_t / S_t \mid \mathcal{F}_t]) + \text{var}(\mathbb{E}[dS_t / S_t \mid \mathcal{F}_t]) = \mathbb{E}[\sigma_t^2]dt + \text{var}[\mu_t]dt^2$ $= \mathbb{E}[\sigma_t^2]dt$. Hence, the variation in the $\mathcal{F}_t$ – adapted conditional mean-return process $\{\mu_t\}$ has no "first-order effect" on the unconditional variance of the spot return. In a similar way, treating the Euler discretization as exact, we obtain $\text{var}[r_{t+\Delta t}] = \mathbb{E}[\sigma_t^2]\Delta t + \text{var}[\mu_t]\Delta t^2$. For $\Delta t = 1/260$, the variation in $\{\mu_t\}$ seems negligible.

[15] This equation boils down to $\frac{1}{\Delta t}\left(r_{t+\Delta t} - \mathbb{E}_{\mathbb{P}}[r_{t+\Delta t} \mid \mathcal{F}_t]\right)^2 = \frac{1}{\Delta t}\text{var}_{\mathbb{P}}[r_{t+\Delta t} \mid \mathcal{F}_t] + \omega_{t+\Delta t}$ (treating the Euler discretization as exact). Notice that the conditional expectation of the left-hand side equals $\frac{1}{\Delta t}\text{var}_{\mathbb{P}}[r_{t+\Delta t} \mid \mathcal{F}_t]$.

3.2.1 ARSV(1) model and relation with the earlier literature

Let us relate our state space way of extracting information from returns with the ways others have previously done this in the literature, in the discrete-time *ARSV(1) model*. This model, introduced by Taylor (1986), is the most popular discrete-time SV model for stock returns. [16] Volatility is assumed to be driven by a latent Gaussian AR(1) process $\{x_t\}$. Daily returns are generated according to

$$r_t = \mu + \sigma_t \eta_t , \qquad \eta_t \sim \text{i.i.d } \mathcal{N}(0,1) \quad (\mathbb{P}) \qquad (3.20)$$
$$\sigma_t^2 = \exp(x_t), \quad x_{t+1} = \phi_0 + \phi_1 x_t + u_{t+1}, \quad u_t \sim \text{i.i.d } \mathcal{N}(0, \sigma_u^2),$$

with η_t and u_s independent $\forall t,s$. [17] An alternative formulation of the variance process is $\ln \sigma_{t+1}^2 = \phi_0 + \phi_1 \ln \sigma_t^2 + u_{t+1}$. Notice that the returns are conditionally Gaussian: $r_t \mid \sigma_t \sim \mathcal{N}(\mu, \sigma_t^2)$. As mean daily returns are close to zero, the literature commonly assumes $\mu = 0$. In order to derive a possible state space representation of (3.20), it follows from $r_t = \sigma_t \eta_t$ that $\ln r_t^2 = \ln \sigma_t^2 + \ln \eta_t^2$ $= x_t + \ln \eta_t^2$. As $\eta_t^2 \sim \chi^2(1)$, it holds that $\mathbb{E}[\ln \eta_t^2] = -1.27$ and $\text{var}[\ln \eta_t^2] = \pi^2/2$. Defining the mean-zero random variable $\omega_t \equiv \ln \eta_t^2 - \mathbb{E}[\ln \eta_t^2]$, the state space representation becomes

$$\ln r_t^2 + 1.27 = x_t + \omega_t, \quad x_{t+1} = \phi_0 + \phi_1 x_t + u_{t+1}. \qquad (3.21)$$

Evidently, the distribution of the error term ω_t is not Gaussian. If one is willing to assume that the distribution of ω_t may be approximated by $\mathcal{N}(0, \pi^2/2)$, the parameters can still be estimated by QML. This method for estimating the ARSV(1) model is due to Ruiz (1994) and Harvey, Ruiz and Shephard (1994). (See also Harvey and Shephard (1996)). Although QML generates consistent estimates, it is likely to have poor finite-sample properties in this case, as the real distribution of ω_t is far from being Gaussian: It has a much longer (resp. shorter) left (resp. right) tail, and is thus not even symmetric.

To improve on QML, Kim, Shephard and Chib (1998) propose to approximate the distribution of ω_t by a mixture of normals, the number of normal distributions depending on the desired level of accuracy. Sandmann and Koopman (1998) also improve on QML, by decomposing the likelihood function into a linear Gaussian part constructed by the Kalman filter, and a remainder term, which expectation is evaluated by simulation. As the likelihood function can be approximated arbitrarily closely, their method is efficient.

In the 1-factor SV special case of our multifactor model, the latent factor x also follows an AR(1) process (recall (3.12)). The similarities and differences between (3.19) and (3.21) will be clear. It may well be the case that the finite-sample performance of our QML method for extracting information from squared returns will turn out to be poor.

3.2.2 Relation with the GARCH(1,1) model for stock returns

The way in which equation (3.19) extracts information from stock returns is related to the way the GARCH(1,1) model deals with this, if there is only one

[16] Some of the more recent references include Sandmann and Koopman (1998), Mahieu and Schotman (1998), Kim, Shephard and Chib (1998), Liesenfeld and Jung (2000), Carnero et al. (2001).
[17] The ARSV(1) model thus assumes no leverage effect. The *asymmetric* ARSV(1) model of Harvey and Shephard (1996) allows for negative correlation between η_t and u_{t+1} (notice the timing).

factor driving the volatility. Since the GARCH(1,1) model is probably the most widely applied model for describing empirical daily stock returns, a brief discussion of some of the similarities and differences is in place.

1-factor SV model with Heston (1993) volatility

Let us first focus on the 1-factor SV special case of the multifactor model with $\sigma_t^2 = x_t$ and $dx_t = k(\theta - x_t)dt + \sigma\sqrt{x_t}dW_{x,t}$ under $\mathbb{P}$; i.e. Heston (1993) volatility (without leverage effect). (Properties of this SV process used below can be found in appendix B, section 14.) Recall from (3.17) that returns are approximately conditionally Gaussian, $r_{t+\Delta t} \mid \mathcal{F}_t \approx \mathcal{N}(\mu_t\Delta t, \sigma_t^2\Delta t)$. Using (3.12), the stock variance at day $t + \Delta t$ is related to the stock variance of day t by

$$\sigma_{t+\Delta t}^2 = \theta\left[1 - \exp(-k\Delta t)\right] + \exp(-k\Delta t)\sigma_t^2 + u_{t+\Delta t} \qquad (\mathbb{P}) \qquad (3.22)$$

with

$$u_{t+\Delta t} \sim \left(0;\ \sigma^2\theta\left[\frac{1-\exp(-2k\Delta t)}{2k}\right]\right) \qquad (3.23)$$

$$u_{t+\Delta t} \mid \mathcal{F}_t \sim \left(0;\ \sigma^2\theta\left[\frac{1-\exp(-2k\Delta t)}{2k}\right] + \sigma^2\, x_t^*\left[\frac{\exp(-k\Delta t)-\exp(-2k\Delta t))}{k}\right]\right).$$

The volatility shock $u_{t+\Delta t}$ is unconditionally homoskedastic. Conditionally, it is heteroskedastic, with a variance depending on the current volatility-determining state x_t^*, which illustrates the fact that the volatility-of-volatility is level-dependent. For $k = 0$ we have $\sigma_{t+\Delta t}^2 = \sigma_t^2 + u_{t+\Delta t}$, such that the stock variance follows a *random walk* in discrete time. In that case, shocks have an everlasting, *persistent* impact on future volatility. For $k > 0$, the volatility is *mean reverting*. The *volatility persistence* is measured by $\exp(-k\Delta t)$, on a daily basis. The closer k is to 0, the more the volatility behaves as a random walk, the closer the persistence is to 1. The larger k, the faster the volatility reverts back to its mean θ. The *half-life* (in days) of a shock in the stock variance is given by $\ln 2 / k\Delta t$. Using (3.10) and substituting (3.12) in (3.19), followed by rewriting, we find

$$\frac{1}{\Delta t}\left(r_{t+\Delta t} - \mu_t\Delta t\right)^2 = \theta\left[1-\exp(-k\Delta t)\right] + \exp(-k\Delta t)\frac{1}{\Delta t}\left(r_t - \mu_{t-\Delta t}\Delta t\right)^2 + \omega_{t+\Delta t} - \exp(-k\Delta t)\omega_t + u_t. \qquad (3.24)$$

This equation shows that the squared returns are essentially modeled as an autoregressive moving-average (ARMA(1,1)) process, "plus" white noise.

GARCH(1,1) volatility

Consider next the Gaussian GARCH(1,1) model for daily returns, introduced by Bollerslev (1986): [18]

$$r_t = \mu + \sigma_t\eta_t\,, \quad \sigma_t^2 = \phi_0 + \phi_1 e_{t-1}^2 + \phi_2\sigma_{t-1}^2\,, \quad \eta_t \sim \text{i.i.d. } \mathcal{N}(0,1)\,, \quad (\mathbb{P}) \qquad (3.25)$$

in which $e_t \equiv \sigma_t\eta_t = r_t - \mu = r_t - \mathbb{E}[r_t \mid \mathcal{F}_{t-1}]$ and $\phi_0 > 0;\ \phi_1, \phi_2 \geq 0$. For $\phi_1 + \phi_2 < 1$, the GARCH conditional variance process is stationary, and has unconditional variance $\phi_0/(1-\phi_1-\phi_2)$. As in our model and the ARSV(1) model, returns are

[18] There is a bewildering literature on GARCH-type models, both on the theoretical and applied side, starting with the seminal papers of Engle (1982) and Bollerslev (1986). Since then, literally hundreds papers have appeared on the subject. Overviews can be found in Bollerslev et al. (1992), Bera and Higgins (1993), Bollerslev et al. (1994), Engle (1995), and Franses and van Dijk (2000) among others. A recent, up-to-date survey of research on multivariate GARCH-type models is Bauwens et al. (2003).

conditionally Gaussian: $r_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(\mu, \sigma_t^2)$. Introducing the error term $\varsigma_t \equiv e_t^2 - \mathbb{E}[e_t^2 \mid \mathcal{F}_{t-1}] = e_t^2 - \sigma_t^2$, the variance process can be written as

$$\sigma_t^2 = \phi_0 + (\phi_1 + \phi_2)\sigma_{t-1}^2 + \phi_1\varsigma_{t-1}, \tag{3.26}$$

with

$$\varsigma_t \sim \left(0, 2\mathbb{E}[\sigma_t^4]\right), \qquad \varsigma_t \mid \mathcal{F}_{t-1} \sim \left(0, 2\sigma_t^4\right). \tag{3.27}$$

In the GARCH model, the volatility persistence is measured by $\phi_1 + \phi_2$, and the half-life of a shock by $-\ln 2 / \ln(\phi_1 + \phi_2)$. If $\phi_1 + \phi_2 = 1$, the *integrated* GARCH (IGARCH; Engle and Bollerslev (1986)) model results, with the variance showing random walk behavior. [19] The standard GARCH model does not model the leverage effect. [20] Notice from (3.27) that the GARCH volatility-of-volatility is level-dependent. If we substitute the expression $\sigma_t^2 = e_t^2 - \varsigma_t$ into the GARCH conditional variance process (3.25) and perform some rewriting, we obtain that $e_t^2 = \phi_0 + (\phi_1 + \phi_2)e_{t-1}^2 + \varsigma_t - \phi_2\varsigma_{t-1}$, or

$$\left(r_t - \mu\right)^2 = \phi_0 + (\phi_1 + \phi_2)\left(r_{t-1} - \mu\right)^2 + \varsigma_t - \phi_2\varsigma_{t-1}. \tag{3.28}$$

This expression reveals that GARCH(1,1) models the squared returns as an ARMA(1,1) process.

Notice the similarity between (3.15), (3.17), (3.22)-(3.24) on one side, and (3.25)-(3.28) on the other side. As (3.24) and (3.28) illustrate, the way the 1-factor SV model and the GARCH(1,1) model model squared returns seems similar. There is one crucial difference however. In contrast to (3.28), (3.24) includes the extra white-noise error term u_t, due to stochastic volatility. Moreover, although (3.22) and (3.26) look the same, the error $u_{t+\Delta t}$ represents a true *innovation* in the variance process (3.22), whereas the error ς_{t-1} in (3.26) is determined from the *history* of the GARCH process. This observation touches the core of the difference between GARCH-type and SV models, and explains why GARCH models are so easy to estimate (by ML), as opposed to SV models. As the time-t GARCH conditional variance is a deterministic function of the available information at time $t-1$, true surprises in the time-t variance do not occur. In contrast, our model -and SV models in general- do allow for such unpredictable surprises, and that is precisely the reason for the additional error term u_t showing up in (3.24) as opposed to (3.28). [21]

---

[19] If we further restrict $\phi_0 = 0$, the *exponentially weighted moving average* (EWMA) model for volatility results, $\sigma_t^2 = (1-\phi_2)e_{t-1}^2 + \phi_2\sigma_{t-1}^2$, which can be rewritten as $\sigma_t^2 = (1-\phi_2)\sum_{i=0}^{\infty}\phi_2^i e_{t-1-i}^2$. J.P. Morgan's well-known *RiskMetrics system* uses this model for calculating *Value-at-Risk* measures, assuming the decay-factor ϕ_2 to equal 0.94.

[20] The following extensions of the GARCH model model the leverage effect in different ways: *Exponential* GARCH (EGARCH; Nelson (1991)), *Threshold* ARCH (TARCH) also known as GJR-GARCH (Rabemananjara and Zokoian (1993) and Glosten et al. (1993)), *Quadratic* GARCH (QGARCH; Sentana (1995)), *Volatility Switching* GARCH (VS-GARCH; Fornari and Mele (1997)), and *Smooth Transition* GARCH (ST-GARCH; Gonzalez (1998)).

[21] It should be noted that despite this difference, there also exists an intimate connection between GARCH-type and SV models. Nelson (1990) shows that e.g. the GARCH(1,1) model can be seen as a discrete-time approximation of a certain continuous-time SV model. More specific, the GARCH(1,1) model converges weakly to an SV diffusion the finer the (intraday) time grid on which the GARCH model is sampled.

3.3 Extracting information from realized volatilities

The squared return equation (3.19) can be used for information extraction from stock prices, if one has access to daily data. What if one has access to *high-frequency intraday* data? Can we simply replace the squared returns by *realized volatilities* as we may intuitively expect? This section lays out the theoretical foundation for including high-frequency data in the state space analysis.

Section 3.3.1 explains the concept of realized volatility (RV). Section 3.3.2 gives an overview of RV research. Section 3.3.3 motivates the measurement equation for RV (given in section 3.3.4), by considering a particular equation satisfied by the average variance over a day. Section 3.3.5 takes a closer look at the measurement equation for RV, and revisits the squared return equation (3.19).

3.3.1 Realized volatility

For $\Delta t = 1/260$, the interval $[t, t+\Delta t]$ corresponds to one trading day. Suppose that we divide this interval into I intraday intervals of equal length $\delta t \equiv \Delta t / I$. Suppose further that we observe the stock price at the intraday time points $t,\ t+\delta t, .., t+I\delta t = t+\Delta t$. Define the so-called *realized variance* $\sigma^2_{RV,t+\Delta t}$ over day $[t, t+\Delta t]$, as the annualized sum of squared intraday logreturns,

$$\sigma^2_{RV,t+\Delta t} \equiv \frac{1}{\Delta t}\sum_{i=1}^{I}\left(\ln\frac{S_{t+i\delta t}}{S_{t+(i-1)\delta t}}\right)^2 . \qquad (\mathbb{P}) \qquad (3.29)$$

[22] Its square-root is labeled the *realized volatility* over that day. From the theory of *quadratic variation* [23], the realized variance converges in probability to the average variance over that day, if sampling becomes continuous:

$$\begin{aligned}\operatorname*{plim}_{I\to\infty} \sigma^2_{RV,t+\Delta t} &= \frac{1}{\Delta t}\operatorname*{plim}_{I\to\infty}\sum_{i=1}^{I}\left(\ln\frac{S_{t+i\delta t}}{S_{t+(i-1)\delta t}}\right)^2 \qquad (\mathbb{P}) \qquad (3.30)\\ &= \frac{1}{\Delta t}\left([\ln S]_{t+\Delta t} - [\ln S]_t\right) = \frac{1}{\Delta t}\int_t^{t+\Delta t}\left(d\ln S_u\right)^2 = \frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_u^2 du .\end{aligned}$$

---

[22] Our choice for *equidistant* intraday time points is a notational convenience, and is neither essential for the definition of realized variance, nor for the derivation of the measurement equation.

[23] From Itô's lemma, the log stock price evolves as $d\ln S_t = (\mu_t - \frac{1}{2}\sigma_t^2)dt + \sigma_t dW_{S,t}$ under $\mathbb{P}$. The *quadratic variation* $[\ln S]_T$ of the logprice process over the interval $[0,T]$ is defined as the mean-square limit of the sum of squared logprice increments, for any sequence of partitions $0 = t_0 < t_1 < .. < t_N = T$ of $[0,T]$, for which $\lim_{N\to\infty}\max_i (t_i - t_{i-1}) = 0$. That is, $[\ln S]_T$ is defined such that

$$\lim_{N\to\infty}\mathbb{E}_{\mathbb{P}}\left[\left(\sum_{i=1}^{N}\left(\ln S_{t_i} - \ln S_{t_{i-1}}\right)^2 - [\ln S]_T\right)^2\right] = 0$$

holds. In other words, the sum of squared logreturns converges in mean-square (and hence in probability) to the quadratic variation of $\{\ln S_t\}$ over $[0,T]$, the more frequently we sample the logprice process. As $\{\ln S_t\}$ is a continuous *semimartingale* (i.e., $\ln S_t$ is the sum of a continuous martingale, $M_t \equiv \int_0^t \sigma_u dW_{S,u}$, and a process of bounded variation, $A_t = \int_0^t (\mu_u - \frac{1}{2}\sigma_u^2)du$, the latter which has zero quadratic variation), its quadratic variation is only determined by the martingale part, and equals $[\ln S]_T = \int_0^T (d\ln S_u)^2 = \int_0^T \sigma_u^2 du$. Recall that for Brownian motion W, $[W]_T = T$.

Here, $[\ln S]_u$ represents the quadratic variation of the logprice process $\{\ln S_t; t \geq 0\}$ over the interval $[0,u]$. Result (3.30) essentially means that, given an *observed* (i.e., realized) stock price path over a certain day, the implied *latent* average variance over that day, can arbitrarily well be approximated by computing the sum of squared intraday returns from that path; i.e., the realized variance. The approximation gets better for larger I, hence the more often the stock price is sampled during that day. The realized variance (3.29) is thus an estimator of the (ex-post) average variance.

3.3.2 Overview of research on RV

This section provides an overview of the research on realized volatility. [24] Research on RV is fairly novel, but is rapidly expanding at an increasing rate. After a brief discussion of the early papers, rather than attempting to be complete, we will focus on current, leading main strands of research towards RV.

Early papers and properties of empirical RV

The work of notably Andersen and Bollerslev (1998) started a whole new area of research on the use of high-frequency data for measuring and forecasting financial volatility. At the time numerous studies suggested that ARCH and SV models provide poor volatility forecast, despite the good fit of these models. Evaluation of these forecasts was typically done by regressing daily squared returns on e.g. GARCH volatility forecasts, leading to R-squareds of only around 3%. Andersen and Bollerslev showed however, that when evaluating these GARCH forecasts against realized volatilities, R-squareds of close to 50% result. As such, they concluded that squared returns are typically noisy volatility estimators, and that GARCH models provide much better volatility forecasts than was commonly suggested at that time.

A series of papers, notably Ebens (1999), Andersen, Bollerslev, Diebold and Ebens (2001a,b) and Andersen, Bollerslev, Diebold and Labys (2001a,b) next investigated the empirical distribution of realized volatilities (computed using 5-minute returns). Studying 60 stocks, 1 stock index and 2 exchange rates, these authors find the following. First, daily logreturns standardized by realized volatilities are approximately Gaussian. (This is in sharp contrast to the remaining excess kurtosis typically found in the standardized returns of an estimated GARCH-type or SV model.) Second, the distribution of realized variances is right skewed and leptokurtic. Third, logarithmic realized volatilities are approximately Gaussian. Fourth, the dynamics of realized logarithmic volatilities are well approximated by a fractionally integrated long-memory (though still stationary) process.

Direct modeling of RV for the purpose of volatility forecasting

From the previous section it is clear that, if one believes that empirical stock prices are truly generated by an SV model in the *continuous semimartingale* class (hence no jumps), then RV provides a non-parametric, or *model-free* way of measuring *ex-post* volatility (see (3.29)-(3.30)). No full specification of a specific SV model is required for this, which makes RV a robust, ex-post volatility measure. Volatility essentially becomes *"observable"* (disregarding measurement error, and errors due to the fact that I is finite). This sharply contrasts to

[24] In line with the literature, we will use the terms *realized variance* and *realized volatility* interchangeably, except there where the distinction is important.

measuring volatility using GARCH-type and SV models, which treat volatility as latent, and which make specific distributional assumptions regarding volatility.

This observation has led researchers to directly model (log) realized volatilities using (non)linear (V)AR(FI)MA-type models, and investigate the forecasting performance of these models. (Accurate volatility forecasts are of obvious practical importance in all sorts of financial decision-making problems as e.g., asset allocation decisions.) The current conclusion of the papers in this area is that direct modeling of RV using ARMA-type models yields much better volatility forecasts than traditional GARCH and SV models. See e.g., Andersen, Bollerslev, Diebold and Labys (2003), Martens et al. (2003) and Koopman et al. (2005).

Do not sample too frequently (5–30 minutes)
Equation (3.30) demonstrates that the more frequent the stock price is sampled during a day, the better RV approximates the ex-post volatility over that day. In the limit for $I \to \infty$, RV is free of measurement error. In practice unlimited sampling is not possible however, given that transactions do not take place continuously. Moreover, given that very-high-frequency data tends to be contaminated by *market-microstructure noise*, as a result of e.g. infrequent trading, bid-ask bounce, and price discreteness, the sampling frequency should not be too large. Andersen and Bollerslev (1998), and Andersen, Bollerslev, Diebold and Ebens (2001a,b) argue that five-minute sampling should be the lowest frequency. The typical frequency used in the literature is in the range of 5 to 30 minutes.

RV is still a noisy estimator of quadratic variation
Barndorff-Nielsen and Shephard (2001a) show that RV is normally distributed asymptotically, and thus has a variance. A simulation experiment by these authors shows that even with 2.5-minute sampling, the estimation error in the RV estimate is still rather large. So although RV means an obvious improvement to squared daily returns, ex-post volatility not truly becomes observable.

Barndorff-Nielsen and Shephard have developed several asymptotic theories for RV-type measures recently. For a fairly general SV model, Barndorff-Nielsen and Shephard (2002, 2004b) establish a central limit theorem (CLT) for RV as $I \to \infty$. Barndorff-Nielsen and Shephard (2003, 2004a) show that a CLT also applies to measures based on powers of intraday returns (*realized power variation*) and products of powers of absolute returns (e.g. *bipower variation*). Barndorff-Nielsen and Shephard (2005) provide a joint asymptotic distribution theory for RV and bipower variation, and show how this distribution can be used for testing for jumps in speculative prices (see also below).

Effects of market-microstructure noise on RV
Another branch of research is currently studying the impact of market-microstructure noise on RV, including Aït-Sahalia et al. (2005a, 2005b), Zhang et al. (2002), Zhang (2004), Bandi and Russel (2004), Hansen and Lunde (2004a, 2004b) and Barndorff-Nielsen et al. (2004). These papers propose alternative estimators of quadratic variation that are robust to microstructure noise.

Consider the papers of Aït-Sahalia et al. (2005a) and Zhang et al. (2002) for example. The idea of these papers is as follows. The observed, high-frequency transaction price is modeled as the sum of an unobserved, efficient price plus some noise component, due to adverse microstructure effects. The efficient price follows a continuous diffusion with SV (no jumps), and the interest is in estimating the (ex-post) volatility of the efficient price. In the absence of noise, asymptotic theory suggests to sample the stock price intradaily as often as possible (recall (3.30)). In the presence of simple i.i.d. noise, Aït-Sahalia et al. (2005a) and Zhang et al. (2002) show that computing RV using e.g. second-by-second data, results in an estimate of the variance of the noise term, and thus not of the realized quadratic variation.

Their analysis confirms that RV should not be calculated by sampling too frequently. Sampling every five minutes only if tick-by-tick data is available, means however that a lot of information is thrown away. From an intuitive statistical point of view, there should be gains from exploiting all data. These authors propose to estimate ex-post volatility by *two-scales realized volatility*. By computing RV at two (or more) different frequencies [25], averaging the results, and taking some linear combination, results in a consistent estimator of the quadratic variation, which has a much smaller variance than RV. Aït-Sahalia et al. (2005b) deal with an extension, the case of dependent microstructure noise.

Effects of jumps on RV

The discussion so far has not included the effects of jumps on realized volatility. The realized variance $\sigma^2_{RV,t+\Delta t}$ converges to the (average) quadratic variation over day $[t, t+\Delta t]$ -which equals the average variance $\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_u^2 du$-, only if stock prices are generated by a semimartingale process with SV, but without jumps (as in our model). Now, suppose that the data-generating process (DGP) contains jumps in prices (but no microstructure noise). [26] What are the effects of jumps on the asymptotic properties of the RV measure?

If the DGP contains rare jumps, then RV is an estimator of $\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_u^2 du$ *plus* a certain jump contribution (see e.g., Barndorff-Nielsen and Shephard (2004a)). Hence, whenever a jump has occurred in the interval $[t, t+\Delta t]$, $\sigma^2_{RV,t+\Delta t}$ does not consistently estimate $\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_u^2 du$. Barndorff-Nielsen and Shephard (2004a) introduce a different, RV-related measure coined *bipower variation* (BV), which is a consistent estimator of $\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_u^2 du$. BV is thus robust to jumps in the DGP. They moreover show that for a process with no jumps, RV is a slightly better estimator of $\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_u^2 du$ then BV. Barndorff-Nielsen and Shephard (2005) propose a test for jumps in asset prices, based on the difference between RV and BV. Huang and Tauchen (2004) propose another test, based on the relative RV - BV difference.

Using RV for estimating parametric SV models

Direct modeling of realized volatilities using ARMA-type models is useful for volatility forecasting; no specific parametric SV model is required for this, as we have seen. Nonetheless, Barndorff-Nielsen and Shephard (2002) and Bollerslev and Zhou (2002) consider ways to use the information in realized volatilities to estimate continuous-time SV stock price models. Bollerslev and Zhou (2002) match the sample moments of realized volatility to the population moments of integrated volatility for the Heston model, in a GMM-type estimation procedure.

[25] This means: Compute a first RV by sampling, e.g., every 5 minutes, using the 1, 6, 11,… minute returns. Then compute a second RV using the 2, 7, 12,… minute returns, etcetera.

[26] The modeling of jumps in prices seems particularly important for high-frequency data analysis. E.g., Andersen, Bollerslev and Diebold (2003) and Huang and Tauchen (2004) provide empirical evidence of jumps in high-frequency financial prices.

The state space approach of Barndorff-Nielsen and Shephard (2002) is discussed in section 3.3.4, after we have explained our own approach.

Although in a pure stock price setting the estimation of SV models is perhaps not of so much use anymore, the use of RV for estimating option pricing models seems particularly useful: Eventually, the primary purpose of such a model in practice is to price and hedge exotic options. As RV is a less noisy estimator of ex-post volatility than the squared return, we expect efficiency gains from exploiting the information in high-frequency data.

3.3.3 An equation for the average variance

To derive a possible measurement equation for RV, we focus on the distributional properties of its probability limit, the average variance. The moments of the integrated variance are derived in sections 9 and 10 of appendix B. Consider the equation

$$\begin{aligned}\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_s^2 ds &= \mathbb{E}_{\mathbb{P}}\left[\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_s^2 ds \mid \mathcal{F}_t\right] + \varpi_{RV,t+\Delta t} \qquad (\mathbb{P}) \qquad (3.31)\\ &= \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} + \boldsymbol{\delta}'\frac{\mathbf{D}(\Delta t)}{\Delta t}\mathbf{x}_t^* + \varpi_{RV,t+\Delta t}\\ &= \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} + \boldsymbol{\delta}'\left(\mathbf{K}_d\Delta t\right)^{-1}\left(\mathbf{I}_n - \exp[-\mathbf{K}_d\Delta t]\right)\mathbf{x}_t^* + \varpi_{RV,t+\Delta t},\end{aligned}$$

in which the noise term $\varpi_{RV,t+\Delta t}$ has (conditional) mean zero by construction. In words, we decompose the average variance over day $[t, t+\Delta t]$ into its optimal forecast given time-t information [27], plus a random error term $\varpi_{RV,t+\Delta t}$, the estimation error. This equation is linear in the state variables $\mathbf{x}_t^* = \mathbf{x}_t - \boldsymbol{\theta}$. If this equation -in some slightly adjusted version- is to be incorporated into the linear state space representation (2.1)-(2.2), the error series $\{\varpi_{RV,t+\Delta t}\}$ ought to be white noise. Let us therefore check if this is indeed the case.

The white noise property of $\{\varpi_{RV,t+\Delta t}\}$
In a straightforward way, it follows that both the conditional and unconditional mean of $\varpi_{RV,t+\Delta t}$ are zero under $\mathbb{P}$: $\mathbb{E}_{\mathbb{P}}[\varpi_{RV,t+\Delta t} \mid \mathcal{F}_t] = 0$ and $\mathbb{E}_{\mathbb{P}}[\varpi_{RV,t+\Delta t}] = 0$. From appendix B, section 9, the conditional variance of $\varpi_{RV,t+\Delta t}$ equals

$$\begin{aligned}\mathrm{var}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t} \mid \mathcal{F}_t\right] &= \mathrm{var}_{\mathbb{P}}\left[\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_s^2 ds \mid \mathcal{F}_t\right] \qquad (3.32)\\ &= \frac{1}{\Delta^2 t}\boldsymbol{\delta}'\left[\mathbf{N}(\Delta t)\odot\boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' + \sum_{i=1}^n\left(\left\{\left[\boldsymbol{\beta}_i\odot\mathbf{x}_t^*\right]'\otimes\mathbf{I}_n\right\}\mathcal{I}(\Delta t)\right)\odot\boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'\right]\boldsymbol{\delta}.\end{aligned}$$

(We refer to appendix A for the (complicated) definitions of the matrices $\mathbf{N}(\Delta t)$ and $\mathcal{I}(\Delta t)$.) The unconditional variance of $\varpi_{RV,t+\Delta t}$ follows then as

$$\mathrm{var}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}\right] = \mathbb{E}_{\mathbb{P}}\left[\mathrm{var}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t} \mid \mathcal{F}_t\right]\right] = \frac{1}{\Delta^2 t}\boldsymbol{\delta}'\left[\mathbf{N}(\Delta t)\odot\boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right]\boldsymbol{\delta}, \quad (3.33)$$

[27] That is, its minimum MSE forecast or "estimator" given time-t information, which is its conditional expectation. Technically, the estimation error is perpendicular or *orthogonal* to the information set $\mathcal{F}_t$.

which reveals that the series is homoskedastic. The autocovariance of order $p = 1, 2, ...$ equals

$$\text{cov}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}, \varpi_{RV,t-(p-1)\Delta t}\right] = \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}}\left(\varpi_{RV,t+\Delta t}\,\varpi_{RV,t-(p-1)\Delta t} \mid \mathcal{F}_t\right)\right] \qquad (3.34)$$
$$= \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}}\left(\varpi_{RV,t+\Delta t} \mid \mathcal{F}_t\right)\varpi_{RV,t-(p-1)\Delta t}\right] = 0\,,$$

such that the series is uncorrelated. Hence, the error series $\{\varpi_{RV,t+\Delta t}\}$ is indeed a white noise series.

Investigating the correlation between $\{\varpi_{RV,t+\Delta t}\}$ and $\{\mathbf{u}_{t+\Delta t}\}$

Appendix IIIb examines the correlation between the disturbances $\{\varpi_{RV,t+\Delta t}\}$ and the volatility-factor shocks $\{\mathbf{u}_{t+\Delta t}\}$. Although their non-contemporaneous correlation equals zero (for $p = 1, 2, ..$, $\text{corr}_{\mathbb{P}}[\varpi_{RV,t+\Delta t}, \mathbf{u}_{t-(p-1)\Delta t}] = \mathbf{0}$), these errors are contemporaneously correlated with covariance

$$\text{cov}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}, \mathbf{u}_{t+\Delta t}\right] = \left(\left[\mathbf{D}(\Delta t)\mathbf{1}\mathbf{1}' - \mathbf{G}(\Delta t)\right]\mathbf{K}_d^{-1} \odot \mathbf{\Sigma}\mathbf{M}_d\mathbf{\Sigma}'\right)\frac{\boldsymbol{\delta}}{\Delta t}\,, \qquad (3.35)$$

and correlation vector

$$\text{corr}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}, \mathbf{u}_{t+\Delta t}\right] = \left(\frac{1}{\Delta^2 t}\boldsymbol{\delta}'\left[\mathbf{N}(\Delta t) \odot \mathbf{\Sigma}\mathbf{M}_d\mathbf{\Sigma}'\right]\boldsymbol{\delta}\right)^{-\frac{1}{2}} \left[\mathbf{I}_n \odot \mathbf{G}(\Delta t) \odot \mathbf{\Sigma}\mathbf{M}_d\mathbf{\Sigma}'\right]^{-\frac{1}{2}}$$
$$*\left(\left[\mathbf{D}(\Delta t)\mathbf{1}\mathbf{1}' - \mathbf{G}(\Delta t)\right]\mathbf{K}_d^{-1} \odot \mathbf{\Sigma}\mathbf{M}_d\mathbf{\Sigma}'\right)\frac{\boldsymbol{\delta}}{\Delta t}\,. \qquad (3.36)$$

These expressions are for the general multifactor SV model. To gain insight into the magnitude of this contemporaneous correlation, consider the 1-factor affine SV special case with $\sigma_t^2 = x_t$ and

$$dx_t = k(\theta - x_t)dt + \sigma\sqrt{\alpha + \beta x_t}\,dW_{x,t}\,. \qquad (\mathbb{P}) \qquad (3.37)$$

Invoking the matrix definitions stated in appendix A, we find for the covariance (3.35) between $\varpi_{RV,t+\Delta t}$ and $u_{t+\Delta t}$

$$\text{cov}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}, u_{t+\Delta t}\right] = \frac{\sigma^2(\alpha + \beta\theta)}{k}\left(\frac{1 - \exp[-k\Delta t]}{k\Delta t} - \frac{1 - \exp[-2k\Delta t]}{2k\Delta t}\right), \qquad (3.38)$$

whereas the variance of $\varpi_{RV,t+\Delta t}$ in (3.33) becomes

$$\text{var}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}\right] = \frac{\sigma^2(\alpha + \beta\theta)}{k^2\Delta t}\left[1 - 2\left(\frac{1 - \exp[-k\Delta t]}{k\Delta t}\right) + \frac{1 - \exp[-2k\Delta t]}{2k\Delta t}\right]. \qquad (3.39)$$

As the volatility shock has a variance of (see (3.13))

$$\text{var}_{\mathbb{P}}\left[u_{t+\Delta t}\right] = \sigma^2(\alpha + \beta\theta)\frac{1 - \exp[-2k\Delta t]}{2k}\,, \qquad (3.40)$$

we find for the contemporaneous correlation between $\{\varpi_{RV,t+\Delta t}\}$ and $\{u_{t+\Delta t}\}$

$$\text{corr}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}, u_{t+\Delta t}\right] = \frac{\frac{1-\exp[-k\Delta t]}{k\Delta t} - \frac{1-\exp[-2k\Delta t]}{2k\Delta t}}{\sqrt{\left[1 - 2\left(\frac{1-\exp[-k\Delta t]}{k\Delta t}\right) + \frac{1-\exp[-2k\Delta t]}{2k\Delta t}\right]\left[\frac{1-\exp[-2k\Delta t]}{2k\Delta t}\right]}}\,. \qquad (3.41)$$

This correlation depends on the mean-reversion parameter k only. The three volatility-of-volatility parameters σ, α, β are clearly just scale factors. The mean stock variance θ does not play a role either, which seems intuitively understandable, as both $\varpi_{RV,t+\Delta t}$ and $u_{t+\Delta t}$ measure deviations between random variables and their conditional expectations. Notice that the correlation has the same magnitude for the 1-factor affine, CIR ($\alpha = 0$, $\beta = 1$), and Ornstein-Uhlenbeck ($\alpha = 1, \beta = 0$) SV specifications.

Figure 3.1 shows how this correlation varies for $k \in [0.10; 25]$, which corresponds to a volatility persistence ($\exp[-k\Delta t]$) in the range of 0.90832 (for $k = 25$) and 0.99962 ($k = 0.10$). The graph seems virtually linear. The magnitude of the contemporaneous correlation is substantial, around 0.86 .

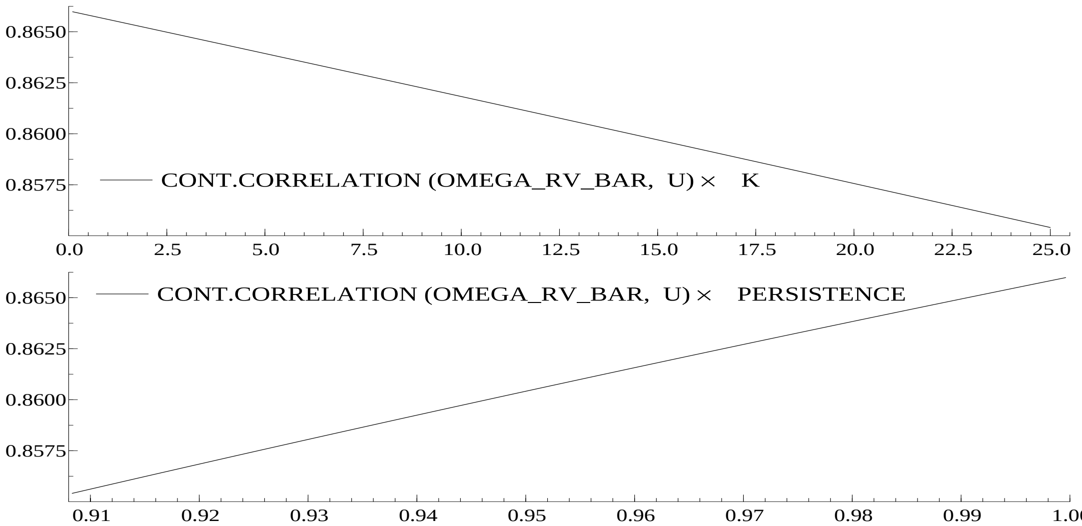


Figure 3.1: Contemp. correlation between $\{\varpi_{RV,t+\Delta t}\}$ and $\{u_{t+\Delta t}\}$ in the 1-factor affine/ CIR/OU SV model for $k \in [0.10; 25]$, i.e. a vol persistence in range [0.90832; 0.99962].

3.3.4 A measurement equation for realized volatilities

Summarizing from the previous section, the following exact relation thus holds:

$$\begin{aligned}\operatorname*{plim}_{I\to\infty} \sigma^2_{RV,t+\Delta t} &= \mathbb{E}_{\mathbb{P}}\left[\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_s^2 ds \mid \mathcal{F}_t\right] + \varpi_{RV,t+\Delta t} \qquad (\mathbb{P}) \qquad (3.42)\\ &= \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} + \boldsymbol{\delta}'\left(\mathbf{K}_d\Delta t\right)^{-1}\left(\mathbf{I}_n - \exp[-\mathbf{K}_d\Delta t]\right)\mathbf{x}_t^* + \varpi_{RV,t+\Delta t}\,.\end{aligned}$$

The error series $\{\varpi_{RV,t+\Delta t}\}$ is white noise with variance given in (3.33). $\{\varpi_{RV,t+\Delta t}\}$ is non-contemporaneously uncorrelated with the volatility-factor shocks $\{\mathbf{u}_{t+\Delta t}\}$, but contemporaneously correlated, with correlation vector given in (3.36).

Given these results, we propose the following measurement equation for extracting parameter and volatility information from a series of observed realized variances:

$$(\mathbb{P}) \qquad (3.43)$$

$$\sigma^2_{RV,t+\Delta t} = \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} + \boldsymbol{\delta}'\left(\mathbf{K}_d\Delta t\right)^{-1}\left(\mathbf{I}_n - \exp[-\mathbf{K}_d\Delta t]\right)\mathbf{x}_t^* + \omega_{RV,t+\Delta t}\,,$$

with

$$\omega_{RV,t+\Delta t} \sim (0,\ \sigma^2_{\omega,RV}),$$
$$\text{corr}_{\mathbb{P}}\left[\omega_{RV,t+\Delta t}, \mathbf{u}_{t+\Delta t}\right] = \mathbf{c}, \quad \text{corr}_{\mathbb{P}}\left[\omega_{RV,t+\Delta t}, \mathbf{u}_{t-(p-1)\Delta t}\right] = \mathbf{0},\ p = 1,2,..,$$

for some real-valued $\sigma^2_{\omega,RV}$ and $(nx1)\mathbf{c}$, and $t = 0, \Delta t,..,(T-1)\Delta t$. This measurement equation is not fully exact: As we take I finite, the realized variance $\sigma^2_{RV,t+\Delta t}$ only approximates its probability limit, the average variance $\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_u^2 du$ over the interval $[t, t+\Delta t]$. Nonetheless, given that $\{\varpi_{RV,t+\Delta t}\}$ is white noise, and given its non-contemporaneous zero-correlation but contemporaneous nonzero-correlation with $\{\mathbf{u}_{t+\Delta t}\}$, it seems reasonable to assume these properties to hold for the error series $\{\omega_{RV,t+\Delta t}\}$ as well. The next section elaborates further on the mean-zero assumption for $\omega_{RV,t+\Delta t}$. When estimating the resulting state space model using (3.43), we choose to treat both $\sigma^2_{\omega,RV}$ and $\mathbf{c}$ as free (unrestricted) parameters. Hence, we neither impose the restriction that $\sigma^2_{\omega,RV}$ equals the variance of $\varpi_{RV,t+\Delta t}$, nor that $\mathbf{c}$ equals $\text{corr}_{\mathbb{P}}[\varpi_{RV,t+\Delta t}, \mathbf{u}_{t+\Delta t}]$, because the measurement equation is not fully exact. Treating $\sigma^2_{\omega,RV}$ and $\mathbf{c}$ as free parameters, circumvents the possible bias incurred from imposing wrong restrictions.

With regard to the correlation for example, simulation results for $I = 48$ for the 1-factor OU and CIR SV models (to be discussed in coming chapters), show that $\text{corr}(\omega_{RV,t+\Delta t}, u_{t+\Delta t})$ is much smaller than the approximate value of 0.86 found for $\text{corr}(\varpi_{RV,t+\Delta t}, u_{t+\Delta t})$ above. We find simulated values in the range [0.09, 0.33]. Evidently, the larger I, the closer $\text{corr}(\omega_{RV,t+\Delta t}, u_{t+\Delta t})$ approaches 0.86 (although this goes very slowly; we need to increase I till in the thousands.) In the other extreme case, i.e. for $I = 1$ (such that the realized variance reduces to the squared daily logreturn), this correlation is close to zero. [28]

Comparing the RV equation (3.43) with the squared return equation (3.19), there is one difference besides the contemporaneous correlation: the diagonal matrix $(\mathbf{K}_d\Delta t)^{-1}(\mathbf{I}_n - \exp[-\mathbf{K}_d\Delta t])$ multiplying the vector $\mathbf{x}_t^*$ in (3.43). This difference is caused by the fact that we used an Euler discretization to arrive at (3.19), whereas for (3.43), we used the theory of quadratic variation to first obtain an exact relation, and then perform an approximation. Nonetheless, as $(\mathbf{K}_d\Delta t)^{-1}(\mathbf{I}_n - \exp[-\mathbf{K}_d\Delta t])$ approximately equals the identity matrix $\mathbf{I}_n$ since $\lim_{a\to 0}(1-\exp[-a])/a = 1$, both equations are essentially the same. The earlier mentioned intuition of replacing squared returns with realized variances to obtain a measurement equation for RV is thus approximately right. However, although the contemporaneous correlation in case of squared returns seems negligible, this is not true when realized variances are considered.

Barndorff-Nielsen and Shephard (2002)
Before proceeding, let us discuss the state space approach of Barndorff-Nielsen and Shephard (2002) to estimating multifactor SV stock price models using RV. In particular, Barndorff-Nielsen and Shephard (2001b, 2002) model the stock variance σ_t^2 as a sum (or *superposition*) of n unobserved, independent

[28] Recall the measurement equation for the squared returns (3.19), obtained from a non-exact Euler discretization. Treating this discretization as exact, there is a zero correlation between $\omega_{t+\Delta t}$ and $\mathbf{u}_{t+\Delta t}$. The simulation results for $I = 1$ justify that the exact correlation is indeed close to zero.

pure jump non-Gaussian OU processes, which are driven by positive-increment *Lévy processes*: [29]

$$\sigma_t^2 = \mathbf{1}'\mathbf{x}_t\,, \qquad dx_{it} = -k_i x_{it}dt + d\mathcal{L}_{i,k_i t}\,, \qquad i = 1,..,n\,, \qquad (\mathbb{P}) \qquad (3.44)$$

Here $\{\mathcal{L}_{i,t}; t \geq 0\}$ are independent Lévy processes [30] with nonnegative increments, with k_i the memory (persistence) parameter. Barndorff-Nielsen and Shephard (2002) decompose the realized variance $\sigma_{RV,t+\Delta t}^2$ over day $[t, t+\Delta t]$ as

$$\sigma_{RV,t+\Delta t}^2 = \frac{1}{\Delta t}\int_t^{t+\Delta t} \sigma_u^2 du + w_{t+\Delta t} = \frac{1}{\Delta t}\mathbf{1}'\int_t^{t+\Delta t} \mathbf{x}_u du + w_{t+\Delta t}\,;\; w_{t+\Delta t} \sim (0, \sigma_w^2)\,, \quad (3.45)$$

with $\mathbb{E}_{\mathbb{P}}[w_{t+\Delta t} \mid \frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_u^2 du] = 0$, and call $w_{t+\Delta t}$ the realized volatility error. They show that $\{w_{t+\Delta t}\}$ is a white noise series. The key feature is now that each $\int_t^{t+\Delta t} x_{iu} du$ admits an ARMA(1,1) representation. The sum of independent ARMA(1,1) processes can be cast into linear state space form (see e.g., Durbin and Koopman (2001)). The resulting unobserved ARMA components model can next be estimated by QML, as the error distributions involved are not known. The primary purpose of this is model-based volatility forecasting, which is particularly easy with the Kalman filter.

Differences with our approach

Our approach to extracting information from RV differs in several respects from Barndorff-Nielsen and Shephard (2002). Assume for simplicity that $\delta_0 = 0$, $\boldsymbol{\delta} = \mathbf{1}$. First, our approach allows for correlated latent factors, whereas their approach requires independent factors. Second, in our model the driving processes are Brownian motions, whereas their model allows for more general Lévy processes, which imply that the volatility jumps from time to time. As such, their model can produce fatter-tailed stock return distributions. Third, our approach is based on a decomposition of the average variance $\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_u^2 du$ over $[t, t+\Delta t]$ into its optimal forecast given time-t information (see 3.31) plus an error term, after which we apply an approximation; recall (3.42)-(3.43). In contrast, decomposition (3.45) of Barndorff-Nielsen and Shephard involves the full path of the latent factors $\mathbf{x}$ over the interval $[t, t+\Delta t]$. More specific, our approach links $\sigma_{RV,t+\Delta t}^2$ to $\mathbf{x}_t$, which follows a VAR(1) process in discrete time. Barndorff-Nielsen and Shephard link $\sigma_{RV,t+\Delta t}^2$ to the integral $\frac{1}{\Delta t}\int_t^{t+\Delta t}\mathbf{x}_u du$, of which each component follows an ARMA(1,1) process. Notice that in case of Brownian motion-driven, persistent volatility, the difference between $\mathbf{x}_t$ and $\frac{1}{\Delta t}\int_t^{t+\Delta t}\mathbf{x}_u du$ may intuitively not be that large, as volatility then typically only changes slowly.

3.3.5 A closer look at the measurement equation for RV

The realized variance $\sigma_{RV,t+\Delta t}^2$ measures the average variance $\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_u^2 du$ (i.e. the quadratic variation of the process $\{\ln S_t\}$) with error. Comparing the exact relation (3.42) to the measurement equation (3.43), reveals that this error is implicitly subsumed in the disturbance term $\omega_{RV,t+\Delta t}$. We assumed $\omega_{RV,t+\Delta t}$ to have mean zero, as we intuitively expect the mean approximation error to be

[29] (NB: We use a notation close to our own, which facilitates a comparison.) See also Nicolato and Venardos (2002). Koopman et al. (2005) provide an empirical illustration in which they apply the state space estimation method of this model, as proposed by Barndorff-Nielsen and Shephard (2002).

[30] A *Lévy process* is a process with stationary and independent increments, and can be thought of as a random walk in continuous time. Prominent examples are Brownian motion and the Poisson process. Brownian motion is the only Lévy process that has continuous sample paths. See e.g. Protter (1990).

close to zero (and it gets closer to zero, the larger I). However, it will be clear that a substantial deviation from this assumption is likely to lead to bias in the state space estimates. Let us therefore better justify the mean-zero assumption for the disturbance term $\omega_{RV,t+\Delta t}$ by examining its true mean. The error $\omega_{RV,t+\Delta t}$ can be written as

$$\omega_{RV,t+\Delta t} = \sigma^2_{RV,t+\Delta t} - \frac{1}{\Delta t}\int_t^{t+\Delta t} \sigma_s^2 ds + \varpi_{RV,t+\Delta t}, \qquad (3.46)$$

and has expectation $\mathbb{E}_{\mathbb{P}}[\omega_{RV,t+\Delta t}] = \mathbb{E}_{\mathbb{P}}[\sigma^2_{RV,t+\Delta t}] - (\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})$. If the mean of the realized variance coincides with the mean $\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta}$ of the average variance, then the mean of $\omega_{RV,t+\Delta t}$ equals zero. Evidently, this holds in the limit case for $I \to \infty$, hence $\delta t \to 0$. For finite I however, which is typically the case when studying empirical data, the mean of $\omega_{RV,t+\Delta t}$ equals

$$\mathbb{E}_{\mathbb{P}}\left[\omega_{RV,t+\Delta t}\right] = \left[\mu - \tfrac{1}{2}(\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\right]^2 \delta t + \tfrac{1}{4}\boldsymbol{\delta}'\left[\left(\frac{\mathbf{N}(\delta t)}{\delta t} + \frac{\mathbf{D}(\delta t)}{\delta t}\mathbf{1}\mathbf{1}'\mathbf{D}(\delta t)\odot \mathbf{J}\right)\odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right]\boldsymbol{\delta}. \qquad (3.47)$$

Here we have used expression (c.9) in the appendix for $\mathbb{E}_{\mathbb{P}}[\sigma^2_{RV,t+\Delta t}]$. (For simplicity that appendix assumes a constant stock price drift of $\mu_t = \mu\ \forall t$.) As it holds that $\lim_{\delta t\to 0}\mathbf{N}(\delta t)/\delta t = \mathbf{0}$, $\lim_{\delta t\to 0}\mathbf{D}(\delta t)/\delta t = \mathbf{I}_n$ and $\lim_{\delta t\to 0}\mathbf{D}(\delta t) = \mathbf{0}$ (see appendix A), we expect the mean of $\omega_{RV,t+\Delta t}$ to be close to 0 for small δt.

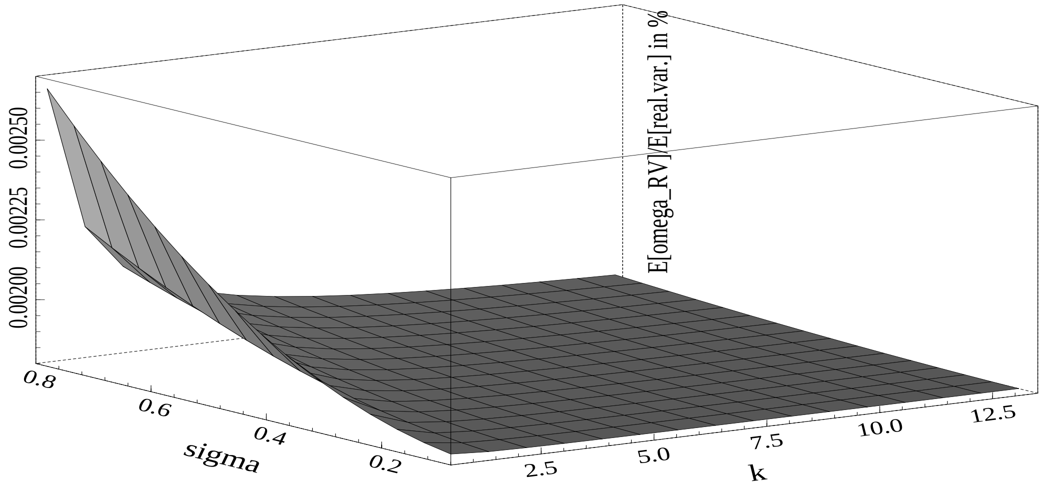
To get more insight into its magnitude, consider the 1-factor SV specification (3.37). In that case

$$\mathbb{E}_{\mathbb{P}}\left[\omega_{RV,t+\Delta t}\right] = \left[(\mu - \tfrac{1}{2}\theta)^2 + \frac{\sigma^2(\alpha+\beta\theta)}{8k}\left(\frac{1-\exp[-k\,\delta t]}{k\,\delta t}\right)^2\right]\delta t + \frac{\sigma^2(\alpha+\beta\theta)}{4k^2}\left[1 - 2\left(\frac{1-\exp[-k\,\delta t]}{k\,\delta t}\right) + \frac{1-\exp[-2k\,\delta t]}{2k\,\delta t}\right]. \qquad (3.48)$$

We choose I=48, meaning that we sample the stock price every 10 minutes during a trading day, which lasts for 8 hours. This implies $\delta t = \Delta t / I \approx 0.8 * 10^{-5}$. Concerning parameter values, we restrict $\alpha = 0, \beta = 1$ such that CIR volatility results. We take μ=0.0825; i.e., an annual mean rate of stock return of 8.25%. We plot the relative mean of $\omega_{RV,t+\Delta t}$ as a function of $k \in [0.75, 13.3]$ (implying a persistence range of $[0.95, 0.9971]$) and $\sigma \in [0.10, 0.80]$, for two values of θ. We take $\theta = 0.0225$ and $\theta = 0.0625$, which correspond to an unconditional stock return volatility of (approximately) 15% and 25% respectively. All parameter values are empirically plausible, as will become apparent in later chapters.

The 3-dimensional graphs in figure 3.2 plot the relative magnitude (in %) of the true mean of $\omega_{RV,t+\Delta t}$, as a fraction of the mean realized variance; i.e. the quantity $\mathbb{E}_{\mathbb{P}}[\omega_{RV,t+\Delta t}]/\mathbb{E}_{\mathbb{P}}[\sigma^2_{RV,t+\Delta t}]$*100%. This quantity seems most suitable to examine for justifying our assumption that $\mathbb{E}_{\mathbb{P}}[\omega_{RV,t+\Delta t}] = 0$ in the measurement equation (3.43). Figure 3.2 plots the relative magnitude of the true mean of $\omega_{RV,t+\Delta t}$ as a function of k and σ, for $\theta = 0.0225$ and $\theta = 0.0625$ respectively. We conclude that the magnitude of $\mathbb{E}_{\mathbb{P}}[\omega_{RV,t+\Delta t}]$ is indeed very close to zero for empirically plausible parameter values. Assuming its value to equal zero in the measurement equation (3.43) is not expected to result in any major bias in subsequent state space parameter estimates.

$\theta = 0.0225$:



$\theta = 0.0625$:

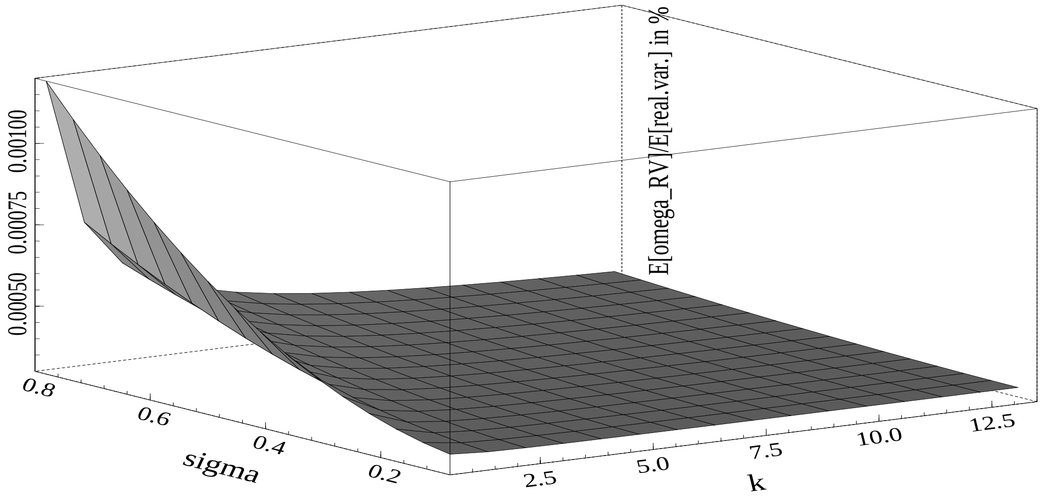


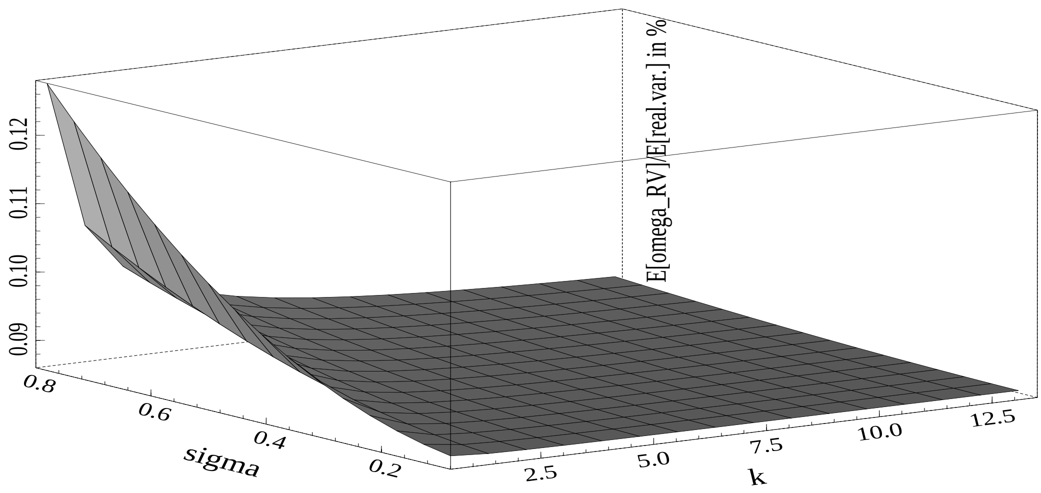
Figure 3.2: Relative magnitude (in %) of the true mean of $\omega_{RV,t+\Delta t}$ in case $I = 48$, as a fraction of the mean realized variance, for $\theta = 0.0225$ and $\theta = 0.0625$.

The measurement equation for the squared returns revisited

Recall the mean-zero assumption for the disturbance $\omega_{t+\Delta t}$ in the measurement equation (3.19) for the squared returns. This equation was obtained from a non-exact Euler discretization, which discretization error is implicitly subsumed in $\omega_{t+\Delta t}$. By assuming $\mathbb{E}_{\mathbb{P}}[\omega_{t+\Delta t}] = 0$, we implicitly assume the mean discretization error to be zero. Again, if this assumption is not reasonable, bias may result in subsequent state space estimates. Given the analysis in this section, we are now able to better justify the reasonability of this assumption.

Notice first that for $I = 1$, the realized variance reduces to the (annualized) squared daily logreturn, which virtually equals the relative return, as $\ln(S_{t+\Delta t} / S_t) \approx (S_{t+\Delta t} - S_t) / S_t = r_{t+\Delta t}$ by a first-order Taylor series expansion. Figure 3.3 plots the relative magnitude (in %) of the true mean of $\omega_{RV,t+\Delta t}$ in case $I = 1$ (i.e. essentially of $\omega_{t+\Delta t}$), as a fraction of the mean annualized squared logreturn, for the same parameter values as before.

$\theta = 0.0225$:



$\theta = 0.0625$:

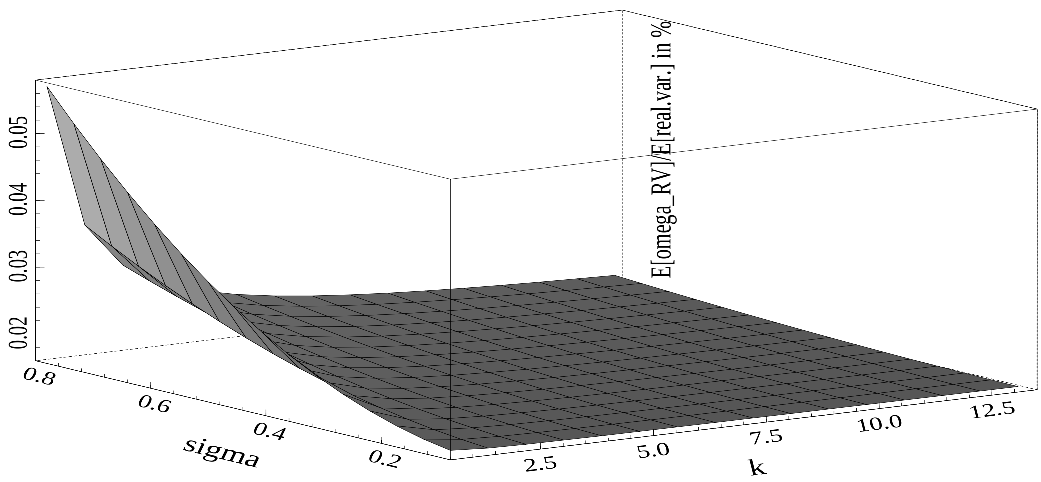


Figure 3.3: Relative magnitude (in %) of the true mean of $\omega_{RV,t+\Delta t}$ in case $I = 1$, as a fraction of the mean annualized squared logreturn, for $\theta = 0.0225$ and $\theta = 0.0625$.

Although the relative magnitude is -as expected- larger than in case $I = 48$, it is still close to zero. Assuming $\mathbb{E}_{\mathbb{P}}[\omega_{t+\Delta t}] = 0$ in the measurement equation for the squared returns, seems therefore reasonable, and is not expected to affect subsequent state space estimates in any major way. Further investigation by Monte Carlo simulation is nevertheless required to examine if our expectations hold indeed true; we refer to the next two chapters for details.

3.4 Extracting information from option prices

This section outlines a strategy for extracting information from a time series of option prices that circumvents simulation during estimation. Section 3.4.1 linearizes the Black-Scholes pricing function in such a way that a theoretical relationship is obtained that is linear in the latent volatility-driving factors. Section 3.4.2 explains how to use this relationship for estimation purposes in practice. Section 3.4.3 provides preliminary insight in the quality of the linearization and explains how a Monte Carlo study can add to this insight. Section 3.4.4 discusses a different way of linearizing the BS pricing function. Section 3.4.5 considers higher-order approximations and proposes a measure to compare the relative merits of using a parabolic versus a linear approximation in a simulation environment.

3.4.1 Linearizing the call price formula

Reconsider the price of a European call option, $C_t = \mathbb{E}_{\mathbb{Q}}[BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t]$, as implied by the multifactor SV model. [31] It is not clear how to translate this expression towards an equation that is linear in the latent factors, which can serve as a measurement equation for information extraction from option prices. An exact way does not exist. Below we propose an approximate method based on linearizing the Black-Scholes pricing function.

As a prelude, recall the affine stock variance specification $\sigma_t^2 = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t$, with the factors evolving under the risk-neutral measure as $d\mathbf{x}_t = \tilde{\mathbf{K}}(\tilde{\boldsymbol{\theta}} - \mathbf{x}_t)dt + \boldsymbol{\Sigma\Lambda}_t d\tilde{\mathbf{W}}_{x,t}$, in which $\tilde{\mathbf{K}} = \mathbf{K}_d + \boldsymbol{\Sigma\Gamma}\mathcal{B}'$ and $\tilde{\boldsymbol{\theta}} = \tilde{\mathbf{K}}^{-1}(\mathbf{K}_d\boldsymbol{\theta} - \boldsymbol{\Sigma\Gamma\alpha})$. From the results of Duffie and Kan (1996), the risk-neutral expectation of the exponent of the integrated variance is an exponential-affine function of the latent factors, [32]

$$\mathbb{E}_{\mathbb{Q}}\left[\exp\left(\int_t^{T_t} \sigma_u^2 du\right) \mid \mathcal{F}_t\right] = \mathbb{E}_{\mathbb{Q}}\left[\exp\left(\tau_t \bar{\sigma}_t^2\right) \mid \mathcal{F}_t\right] = \exp\left[A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\mathbf{x}_t\right], \quad (3.49)$$

in which $A_1(.)$ is a deterministic function of the time to maturity $\tau_t = T_t - t$, and $\mathbf{B}_1(.)$ is a deterministic $(nx1)$ vector function of τ_t. These functions satisfy the following system of Ricatti ordinary differential equations (ODEs),

$$\frac{dA_1(\tau_t)}{d\tau_t} = \tilde{\boldsymbol{\theta}}'\tilde{\mathbf{K}}'\mathbf{B}_1(\tau_t) + \tfrac{1}{2}\sum_{i=1}^{n}\left[\boldsymbol{\Sigma}'\mathbf{B}_1(\tau_t)\right]_i^2 \alpha_i + \delta_0 \quad (3.50)$$

$$\frac{d\mathbf{B}_1(\tau_t)}{d\tau_t} = -\tilde{\mathbf{K}}'\mathbf{B}_1(\tau_t) + \tfrac{1}{2}\sum_{i=1}^{n}\left[\boldsymbol{\Sigma}'\mathbf{B}_1(\tau_t)\right]_i^2 \boldsymbol{\beta}_i + \boldsymbol{\delta},$$

with boundary conditions $A_1(0) = 0$ and $\mathbf{B}_1(0) = \mathbf{0}$. Notice that $A_1(.)$ and $\mathbf{B}_1(.)$ depend on the risk-neutral parameters $\tilde{\mathbf{K}}$ and $\tilde{\boldsymbol{\theta}}$, and thus on the market price of volatility risk-determining parameters $\boldsymbol{\gamma}$, which appear in $\boldsymbol{\Gamma} = \text{diag}[\gamma_1, .., \gamma_n]$. In the Gaussian, multifactor Ornstein-Uhlenbeck SV special case for which $d\mathbf{x}_t = \tilde{\mathbf{K}}(\tilde{\boldsymbol{\theta}} - \mathbf{x}_t)dt + \boldsymbol{\Sigma} d\tilde{\mathbf{W}}_{x,t}$ and $\tilde{\mathbf{K}} = \mathbf{K}_d$ and $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \mathbf{K}_d^{-1}\boldsymbol{\Sigma\Gamma 1}$, these functions can

[31] From now on we add time-subscripts to the strike price K and maturity date T (hence K_t, T_t and thus $\tau_t = T_t - t$) of a call option, to explicitly denote that we may be dealing with different option contracts for each $t = \Delta t, 2\Delta t, .., T\Delta t$.

[32] Duffie and Kan (1996) model the short interest rate by $r_t = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t$, with the latent factors $\mathbf{x}$ following the same SDE under $\mathbb{Q}$ as above. Given this set-up, they show that the time-t price $P(t,T)$ of a zero coupon bond maturing at time T is given by

$$P(t,T) = \mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_t^T r_u du\right) \mid \mathcal{F}_t\right] = \exp[A(\tau) + \mathbf{B}(\tau)'\mathbf{x}_t],$$

in which $\tau \equiv T - t$ and $A(.)$ and $\mathbf{B}(.)$ satisfy a similar -but not exactly equal- system of ODEs as in (3.50). The ODE system (3.50) is obtained from a reparametrization. Indeed, the inspiration for our multifactor SV model arose when reading their and related papers in the literature on the term structure of interest rates.

explicitly be solved for. For non-Gaussian models, the system can numerically be solved relatively easily with Euler's or the Runge-Kutta method for example.

To derive a measurement equation for option prices, we aim at exploiting result (3.50). Reconsidering the call price formula, $C_t = \mathbb{E}_{\mathbb{Q}}[BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t]$, notice that, given time-t information, all arguments of the Black-Scholes pricing function are known constants, except for the average variance $\bar{\sigma}_t^2 = \frac{1}{\tau_t}\int_t^{T_t} \sigma_u^2 du$ over the remaining life τ_t of the option contract, which is random. Now, $\bar{\sigma}_t^2$ may be considered a function of the exponent of the integrated variance, Y_t:

$$\bar{\sigma}_t^2 = \bar{\sigma}_t^2(Y_t) = \frac{1}{\tau_t}\ln Y_t, \qquad Y_t \equiv \exp\left(\int_t^{T_t} \sigma_u^2 du\right) = \exp\left(\tau_t \bar{\sigma}_t^2\right). \tag{3.51}$$

Similarly, the Black-Scholes pricing function $BS(.)$ may be considered as a function $g(.)$ of Y_t,

$$g(Y_t) \equiv BS\left(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2(Y_t)\right). \tag{3.52}$$

The function $g(.)$ thus represents the BS pricing function rewritten in terms of the exponent of the integrated variance Y_t. For the crux of the argument, consider a first-order Taylor series expansion of the function $g(.)$ in some point $Y_t = b_t$,

$$g(Y_t) = g(b_t) + (Y_t - b_t)g'(b_t) + HOT_t, \tag{3.53}$$

in which HOT stands for higher-order terms. The arbitrage-free theoretical call value can then be written as

$$\begin{aligned} C_t &= \mathbb{E}_{\mathbb{Q}}\left[BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t\right] \qquad (3.54) \\ &= \mathbb{E}_{\mathbb{Q}}\left[g(Y_t) \mid \mathcal{F}_t\right] \\ &= \mathbb{E}_{\mathbb{Q}}\left[g(b_t) + (Y_t - b_t)g'(b_t) + HOT_t \mid \mathcal{F}_t\right] \\ &= g(b_t) - b_t g'(b_t) + \mathbb{E}_{\mathbb{Q}}\left[Y_t \mid \mathcal{F}_t\right] g'(b_t) + \mathbb{E}_{\mathbb{Q}}\left[HOT_t \mid \mathcal{F}_t\right]. \end{aligned}$$

Rewriting gives

$$\frac{C_t - g(b_t) + b_t g'(b_t) - \mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]}{g'(b_t)} = \mathbb{E}_{\mathbb{Q}}\left[Y_t \mid \mathcal{F}_t\right] = \exp\left[A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\mathbf{x}_t\right]. \tag{3.55}$$

Taking logarithms yields

$$\ln\left(\frac{C_t - g(b_t) + b_t g'(b_t) - \mathbb{E}_{\mathbb{Q}}\left[HOT_t \mid \mathcal{F}_t\right]}{g'(b_t)}\right) = A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\mathbf{x}_t. \tag{3.56}$$

This equation is linear in the latent factors $\mathbf{x}_t$.

3.4.2 Using the theoretical relationship in practice

Equation (3.56) is an exact theoretical relationship that linearly relates the call price to the latent factors. Some issues naturally arise if (3.56) is to be implemented for the practical purposes of parameter estimation and volatility extraction. Around which point b_t^* to linearize? What about the quality of the linearization? Is the function $g(.)$ sufficiently well behaved to be adequately approximated by a linear function?

Regarding the point of linearization, b_t^*, we argue as follows. Consider the conventional Black-Scholes world with time-varying, but *deterministic* volatility (see section 3.2 of chapter II). The call price equals $C_t^{det} = BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2)$ in that world, such that the Black-Scholes implied variance [33] $\sigma_{implied,t}^2$ -which is defined implicitly as $C_t^{det} \equiv BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \sigma_{implied,t}^2)$ - equals the average variance over the remaining option's life in that setting: $\sigma_{implied,t}^2 = \bar{\sigma}_t^2$. Consider now our time-varying, *random* volatility world, where we view the Black-Scholes pricing function as a function of $Y_t = \exp(\tau_t \bar{\sigma}_t^2)$. An arguably reasonable point of linearization may thus be $Y_t = b_t^* \equiv \exp(\tau_t \, \sigma_{implied,t}^2)$, where $\sigma_{implied,t}^2$ is the BS implied variance associated with the call price as implied by the multifactor SV model. That is, $\sigma_{implied,t}^2$ is defined implicitly from the equality $C_t = \mathbb{E}_{\mathbb{Q}}[BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t] \equiv BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \sigma_{implied,t}^2)$. We perform this linearization for each $t = \Delta t, .., T\Delta t$. For this specific choice of b_t it holds that

$$g(b_t^*) = BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2(b_t^*)) = BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \sigma_{implied,t}^2) = C_t . \quad (3.57)$$

Neglecting higher-order terms, it then follows from (3.56) that

$$\sigma_{implied,t}^2 \approx \frac{A_1(\tau_t)}{\tau_t} + \frac{\mathbf{B}_1(\tau_t)'}{\tau_t} \mathbf{x}_t , \qquad (\mathbb{P}) \qquad (3.58)$$

as $\ln b_t^* = \tau_t \, \sigma_{implied,t}^2$. [34] To implement this approximate relationship in practice, we introduce noise in the form of an additive white noise error term $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$. We add a constant term μ_v as well, reflecting that the error incurred from linearizing may not equal zero on average (see also below). We then obtain

$$\sigma_{implied,t}^2 = \mu_v + \frac{A_1(\tau_t)}{\tau_t} + \frac{\mathbf{B}_1(\tau_t)'}{\tau_t} \mathbf{x}_t + \varepsilon_t , \qquad \varepsilon_t \sim (0, \sigma_\varepsilon^2). \qquad (\mathbb{P}) \qquad (3.59)$$

This equation will serve as the measurement equation for extracting information from option prices.

Both ε_t and μ_v serve to more or less compensate for the performed simplifications. Besides, the introduction of noise is well motivated by the fact that a model is never a complete description of reality. It moreover allows for measurement error in empirically observed option prices. Regarding the assumed statistical properties for the error series $\{\varepsilon_t\}$, these are admittedly rather ad hoc. In theory, the error incurred from linearizing is not random, but is instead some fixed quantity. Assuming $\{\varepsilon_t\}$ to be white noise is therefore mostly for

[33] In general, the (model-implied) Black-Scholes implied variance is defined as that variance, for which the model-implied call price is equal to the theoretical Black-Scholes call price. Different models thus yield different BS implied variances. In practice, the BS implied variance is determined from choosing the variance parameter σ^2 in the BS formula in such a way, that the theoretical BS call (or put) price equals the observed market price of the call (or put).

[34] The implied *yield to maturity* $y(t,T) \equiv -\ln P(t,T)/(T-t)$ on a zero coupon bond in affine models of the term structure of interest rates (Duffie and Kan (1996); see the earlier footnote) satisfies the exact equation

$$-y(t,T) = \frac{A(\tau)}{\tau} + \frac{\mathbf{B}(\tau)'}{\tau} \mathbf{x}_t ,$$

with $\tau = T - t$ the bond's maturity. For estimation purposes, a random error term is typically added in the literature; see e.g. de Jong (2000). Notice the clear analogy with our approach.

convenience. We will additionally assume that $\{\varepsilon_t\}$ is uncorrelated with the other error series that appear in the state space model, i.e., $\{\omega_t\}$, $\{\omega_{RV,t}\}$ and $\{\mathbf{u}_t\}$.

Moneyness does not play a role
One thing to notice about measurement equation (3.59) is that an option's moneyness (i.e., K_t / S_t) does not directly play a role in it: The functions $A_1(.)$ and $\mathbf{B}_1(.)$ depend on maturity only. Equation (3.59) can therefore not fully explain the *volatility smile* across same-maturity different-moneyness options. This is of minor importance if the aim is parameter estimation based on one option time series, characterized by options having approximately the same moneyness at each point in time (e.g., a near-the-money option series). Neither is it of importance if more than one option series is considered, in which each series has approximately the same moneyness (e.g., two near-the-money option series). If such is not the case however, and the aim is information extraction from more than one different-moneyness option series, then higher-order approximations of $g(.)$ may prove useful. See section 3.4.5 for more details.

It should be noted however that it seems wisest to focus on (close-to) ATM option series to extract information from. The reason is twofold. First, near-the-money options are typically most liquid, and hence their prices seem most reliable. (See e.g. the empirical evidence in Cont and Fonseca (2002)). Second, ATM options have largest vega, and are thus most sensitive to volatility fluctuations. As such, these options are expected to contain the most valuable volatility information.

Another rationale for the option equation: Jensen's inequality
Another rationale that leads to equation (3.59) is the following. The BS implied variance $\sigma^2_{implied,t}$ is defined implicitly as $C_t = BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \sigma^2_{implied,t})$, or, abbreviated, $C_t = BS(\sigma^2_{implied,t})$. As the BS pricing function is monotonically increasing in its variance argument, we may write $\sigma^2_{implied,t} = BS^{-1}(C_t)$, and hence

$$\sigma^2_{implied,t} = BS^{-1}\left(\mathbb{E}_{\mathbb{Q}}\left[BS\left(\bar{\sigma}^2_t(Y_t)\right) \mid \mathcal{F}_t\right]\right), \tag{3.60}$$

in which we have abbreviated the call price $C_t = \mathbb{E}_{\mathbb{Q}}[BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}^2_t(Y_t) \mid \mathcal{F}_t]$ to $C_t = \mathbb{E}_{\mathbb{Q}}[BS(\bar{\sigma}^2_t(Y_t)) \mid \mathcal{F}_t]$, again for simplicity. Simulation shows that for ATM and near-the-money options, the $BS(.)$ function is concave (though still rather linear) in its variance argument, irrespective of maturity. Similar pictures as in figure 3.4 result. (For farther in-the-money and farther out-of-the-money options, convex graphs result.) From Jensen's inequality, it thus holds that $\mathbb{E}_{\mathbb{Q}}[BS(\bar{\sigma}^2_t(Y_t)) \mid \mathcal{F}_t] \leq BS(\mathbb{E}_{\mathbb{Q}}[\bar{\sigma}^2_t(Y_t) \mid \mathcal{F}_t])$, and hence

$$\sigma^2_{implied,t} \leq \mathbb{E}_{\mathbb{Q}}\left[\bar{\sigma}^2_t(Y_t) \mid \mathcal{F}_t\right]. \tag{3.61}$$

As $\bar{\sigma}^2_t(Y_t) = \frac{1}{\tau_t}\ln Y_t$ is concave in Y_t, applying Jensen's inequality once again leads to

$$\sigma^2_{implied,t} \leq \frac{1}{\tau_t}\ln\left(\mathbb{E}_{\mathbb{Q}}\left[Y_t \mid \mathcal{F}_t\right]\right) = \frac{A_1(\tau_t)}{\tau_t} + \frac{\mathbf{B}_1(\tau_t)'}{\tau_t}\mathbf{x}_t, \tag{3.62}$$

from which approximation (3.58), and measurement equation (3.59) may subsequently follow. Notice that this derivation shows that μ_ν will be negative for near-the-money options.

3.4.3 Preliminary insight in the quality of the linearization

Let us consider an illustrative example to get some insight into the quality of the linearization. The BS pricing function rewritten in terms of $Y_t = \exp(\tau_t \bar{\sigma}_t^2)$ equals

$$g(Y_t) = S_t \exp(-\bar{q}_t \tau_t)\,\Phi\left[d_{1t}(Y_t)\right] - K_t \exp(-\bar{r}_t \tau_t)\,\Phi\left[d_{2t}(Y_t)\right], \tag{3.63}$$

with

$$d_{1t}(Y_t) = \frac{\ln\frac{S_t}{K_t} + (\bar{r}_t - \bar{q}_t)\tau_t + \frac{1}{2}\ln Y_t}{\sqrt{\ln Y_t}}; \;\; d_{2t}(Y_t) = d_{1t}(Y_t) - \sqrt{\ln Y_t}\,. \tag{3.64}$$

Suppose that the current stock price equals $S_t = 100$. The stock pays an average dividend yield of $\bar{q}_t = 3.5\%$ per annum over the option's life, whereas the average risk-free interest rate equals $\bar{r}_t = 6.5\%$. Consider an ATM call option with strike $K_t = 100$, maturity $\tau_t = 0.50$ years, and price $C_t = 7.63$. These data imply that the associated BS implied volatility of the call option equals 25%. We thus linearize the function $g(.)$ around the point $b_t^* = \exp(\tau_t \sigma_{implied,t}^2) = 1.032$. Recall that $Y_t = \exp(\tau_t \bar{\sigma}_t^2)$. The range of values for Y_t for which we plot $g(.)$ corresponds to an average volatility over the option's life in the range $\bar{\sigma}_t \in [5\%, 40\%]$. Figure 3.4 plots $g(.)$ and its linear approximation in $Y_t = b_t^*$ as functions of Y_t.

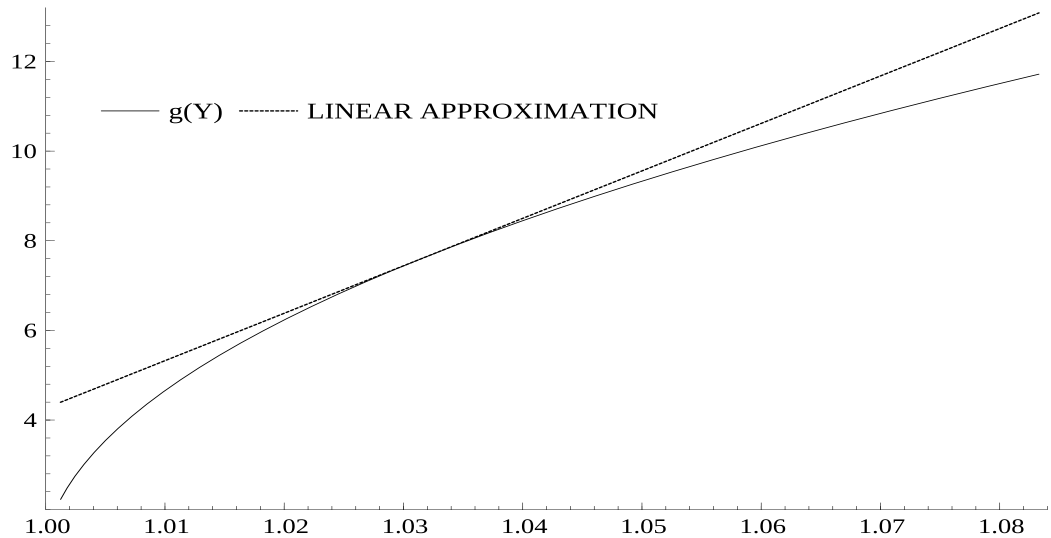


Figure 3.4: Function $g(.)$ and its linear approximation in $Y_t = b_t^*$.

The pricing formula $C_t = \mathbb{E}_{\mathbb{Q}}[BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2(Y_t) \mid \mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}}[g(Y_t) \mid \mathcal{F}_t]$ essentially implies that the call price equals a probability-weighted average of $g(.)$-values, each of which corresponds to a value that Y_t can assume. Intuitively, we expect the most "probability weight" to be put on values of Y_t around the point of linearization b_t^*, where the error is smallest. The further away from the linearization point, the greater the errors, but the less weight these points are expected to have in the overall call value.

Evidently, the current example is just one example. A ceteris-paribus analysis in which we only vary the option's maturity between one month and one year reveals however, that similar pictures result. Similar concave pictures also result when we look at near-the-money options, i.e. $S_t \in [90, 110]$.

Gaining additional insight by Monte Carlo simulation
This preliminary, rather qualitative analysis suggests that a linearization may perform reasonably well. A Monte Carlo study should nonetheless further clarify if this is quantitatively the case too. In particular, the magnitude of the error incurred due to linearizing $g(.)$ is of importance. Relatedly, the size of the "parameter" μ_v in (3.59) is also of significance.

Reconsidering (3.54), the quantity $\mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]$ represents the error incurred when the call price at time t is approximated by a value obtained via a linear approximation of the function $g(.)$ in b_t. Our choice of linearization point $b_t^* = \exp(\tau_t \sigma^2_{implied,t})$ implies that

$$\mathbb{E}_{\mathbb{Q}}\left[HOT_t \mid \mathcal{F}_t\right] = g'(b_t^*)\left(b_t^* - \mathbb{E}_{\mathbb{Q}}\left[Y_t \mid \mathcal{F}_t\right]\right). \tag{3.65}$$

For convenience, we will label this error the *linearization error*. To compute it, the derivative $g'(.)$ is needed. Since $g(Y_t) = BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2(Y_t))$, it follows that

$$g'(Y_t) = \frac{\partial BS}{\partial \bar{\sigma}_t}\frac{\partial \bar{\sigma}_t}{\partial \bar{\sigma}_t^2}\frac{\partial \bar{\sigma}_t^2(Y_t)}{\partial Y_t} = \frac{S_t \exp(-\bar{q}_t \tau_t)\phi\left(d_{1t}(Y_t)\right)}{2Y_t\sqrt{\ln Y_t}}, \tag{3.66}$$

as $\partial BS / \partial \bar{\sigma}_t = \exp(-\bar{q}_t \tau_t) S_t \sqrt{\tau_t}\phi(d_{1t})$, the conventional Black-Scholes vega. Given a simulated time-series dataset, the linearization error can be computed for each $t = \Delta t, .., T\Delta t$. Summary statistics of (the absolute value of) the relative linearization error over the sample (i.e. $\mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]/C_t$) provide quantitative insight in how well the linearization performs.

We want to stress however, that the interpretation of these statistics ought to be taken with some care, given measurement equation (3.59) that will be used for estimation. Recall that one of the reasons for introducing the constant term μ_v and the error ε_t is to more or less compensate for the negligence of the linearization error. Hence, we do not completely neglect this error. What we do neglect however, is the remaining model structure inherent in this error.

Estimating "parameter" μ_v
Can we somehow approximate the introduced constant term μ_v? First of all, notice that from a theoretical point of view, μ_v is not really a parameter. Some additional analysis helps clarifying this. As $g(b_t^*) = C_t$, it follows from (3.56) that

$$\ln\left(b_t^* - \frac{\mathbb{E}_{\mathbb{Q}}\left[HOT_t \mid \mathcal{F}_t\right]}{g'(b_t^*)}\right) = A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\mathbf{x}_t. \tag{3.67}$$

Using an infinite Taylor series expansion [35], the left-hand side can be written as

$$\ln\left(b_t^* - \frac{\mathbb{E}_{\mathbb{Q}}\left[HOT_t \mid \mathcal{F}_t\right]}{g'(b_t^*)}\right) = \ln b_t^* + \sum_{i=1}^{\infty}\frac{(-1)^{2i-1}}{i}\left(\frac{\mathbb{E}_{\mathbb{Q}}\left[HOT_t \mid \mathcal{F}_t\right]}{b_t^* g'(b_t^*)}\right)^i. \tag{3.68}$$

As $\ln b_t^* = \tau_t \sigma^2_{implied,t}$, we obtain from (3.67)

[35] Recall that an infinite Taylor series expansion of an infinitely differentiable function $f(.)$ in a point $x + \Delta x$ is given by $f(x+\Delta x) = f(x) + \sum_{k=1}^{\infty}\frac{1}{k!}f^{(k)}(x)\,(\Delta x)^k$. If $f(x) = \ln(x)$ then $f^{(n)}(x) = (-1)^{n-1}(n-1)!\,x^{-n}, n = 1,2,..$ such that $\ln(x+\Delta x) = \ln x + \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}x^{-k}(\Delta x)^k$.

$$\sigma^2_{implied,t} = \frac{A_1(\tau_t)}{\tau_t} + \frac{\mathbf{B}_1(\tau_t)'}{\tau_t}\mathbf{x}_t + \frac{1}{\tau_t}\sum_{i=1}^{\infty}\frac{(-1)^{2i}}{i}\left(\frac{\mathbb{E}_{\mathbb{Q}}\left[HOT_t \mid \mathcal{F}_t\right]}{b_t^* \, g'(b_t^*)}\right)^i. \qquad (3.69)$$

Given time-t information, the summation term in (3.69) is a non-random, fixed quantity. (Notice that it depends implicitly on the option's moneyness, the latent factors, maturity and the BS implied variance itself in a highly non-linear way.) In essence, our estimation method thus boils down to assuming that this deterministic summation term can be approximated by the sum of a constant μ_v, and a random (white noise) error term ε_t.

As argued before, introducing randomness is a sound argument for using our model and method in practice. [36] In a simulation environment however, where model and reality coincide, assuming that the fixed summation term can be replaced by a constant μ_v and a random ε_t seems somewhat strange. Nonetheless, we may still *pretend* if this is possible, and examine the performance of our estimation method by simulation.

An "estimate" of the "parameter" μ_v is then obtained as follows. Our method assumes the summation term in (3.69) to equal the sum $\mu_v + \varepsilon_t$. Given the assumption that $\varepsilon_t \sim (0, \sigma^2_\varepsilon)$ such that $\frac{1}{T}\sum_{t=\Delta t}^{T\Delta t} \varepsilon_t \approx 0$, an estimate of μ_v can then be obtained from averaging the summation terms for $t = \Delta t, .., T\Delta t$ over the simulated sample.

In coming chapters we discuss the results of several Monte Carlo studies. The remarks made here are important to keep in mind.

3.4.4 Linearizing around the integrated variance instead

We linearized the BS pricing function around the exponent of the integrated variance $Y_t = \exp(\tau_t \bar{\sigma}_t^2)$ to be able to invoke analytical result (3.50). If we linearize around the integrated variance $\int_t^{T_t} \sigma_u^2 du = \tau_t \bar{\sigma}_t^2$ instead and follow a similar analysis and reasoning, we arrive at the approximate relationship

$$\sigma^2_{implied,t} \approx \frac{1}{\tau_t}\mathbb{E}_{\mathbb{Q}}\left[\int_t^{T_t} \sigma_u^2 du \mid \mathcal{F}_t\right] = \mathbb{E}_{\mathbb{Q}}\left[\bar{\sigma}_t^2 \mid \mathcal{F}_t\right], \qquad (3.70)$$

after which noise could again be introduced, and so forth. As $\sigma_t^2 = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t$ and as the latent factors follow a Markov diffusion, the expectation in (3.70) depends on the current factor value $\mathbf{x}_t$ only in principle (and not on its past). Deriving an explicit analytical expression for this expectation in terms of $\mathbf{x}_t$ in the general multifactor case seems difficult however, as the mean-reversion matrix $\tilde{\mathbf{K}}$ under $\mathbb{Q}$ is not diagonal. Section 9 of appendix B derives an expression for this conditional expectation in case the mean-reversion matrix is diagonal (which, under $\mathbb{Q}$, is only the case in the multifactor OU SV case). That derivation builds on the results of section 2 in appendix B, which explicitly exploit the diagonality of the mean-reversion matrix. Section 10 of appendix B derives the conditional moment-generating function (cMGF) of the integrated variance for general, non-diagonal mean-reversion matrices, by generalizing the results of Duffie and Kan (1996). As computing the cMGF involves solving a system of ODEs numerically in the general case, the cMGF (which, recall, is unique) does not seem to easily

[36] In the end, it remains a choice between misspecified models anyway in practice.

allow for computing an explicit expression for the expectation in (3.70) in terms of $\mathbf{x}_t$ either.

We therefore opt for linearizing around $Y_t = \exp(\tau_t \bar{\sigma}_t^2)$ instead of $\frac{1}{\tau_t}\int_t^{T_t} \sigma_u^2 du = \tau_t \bar{\sigma}_t^2$, allowing us to invoke the convenient result (3.50). Moreover, regarding the Black-Scholes pricing function as a function of Y_t allows for an easy consideration of higher-order approximations of the call price: As section 10 of appendix B shows, it namely holds that $\mathbb{E}_{\mathbb{Q}}[Y_t^{\mathrm{m}} \mid \mathcal{F}_t] = \exp[A_{\mathrm{m}}(\tau_t) + \mathbf{B}_{\mathrm{m}}(\tau_t)' \mathbf{x}_t]$, $\mathrm{m} = 1, 2, ...$

Will it matter much?
One may wonder if it might make a large difference, linearizing either around $Y_t = \exp(\tau_t \bar{\sigma}_t^2)$ or $\tau_t \bar{\sigma}_t^2$ (assuming one has succeeded in deriving an explicit expression for $\mathbb{E}_{\mathbb{Q}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]$ in terms of $\mathbf{x}_t$). A second motivation for an options measurement equation based on (3.70) follows again from Jensen's inequality, see (3.61). To arrive at (3.62) which, recall, may be seen as a second motivation for measurement equation (3.59), involves an extra application of Jensen's inequality. Linearizing around $Y_t = \exp(\tau_t \bar{\sigma}_t^2)$ instead of $\tau_t \bar{\sigma}_t^2$ thus involves an additional "convexity effect". Nonetheless, intuitively we do not expect this effect to be large however, as both the linear function $f_1(x) = x$ and the exponential function $f_2(x) = \exp(x)$ have equal slope 1 in $x = 0$, and as $\tau_t \bar{\sigma}_t^2$ typically assumes values close to zero. (E.g., suppose the option's maturity is 2 months and the average volatility is 20%. Then $\tau_t \bar{\sigma}_t^2$ =0.0067 and Y_t = 1.0067.) Figure 3.5 seems to confirm these thoughts. For three different option maturities, i.e. 1.5 months, 6 months and 1 year, we have plotted $Y_t = \exp(\tau_t \bar{\sigma}_t^2)$ against $\tau_t \bar{\sigma}_t^2$ for $\bar{\sigma}_t \in [5\%, 70\%]$. The graphs seem virtually linear. As such, intuitively, we do not expect it to matter much if either around the exponent of the integrated variance is linearized, or around the integrated variance itself (assuming an explicit expression is available for $\mathbb{E}_{\mathbb{Q}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]$ in terms of $\mathbf{x}_t$).

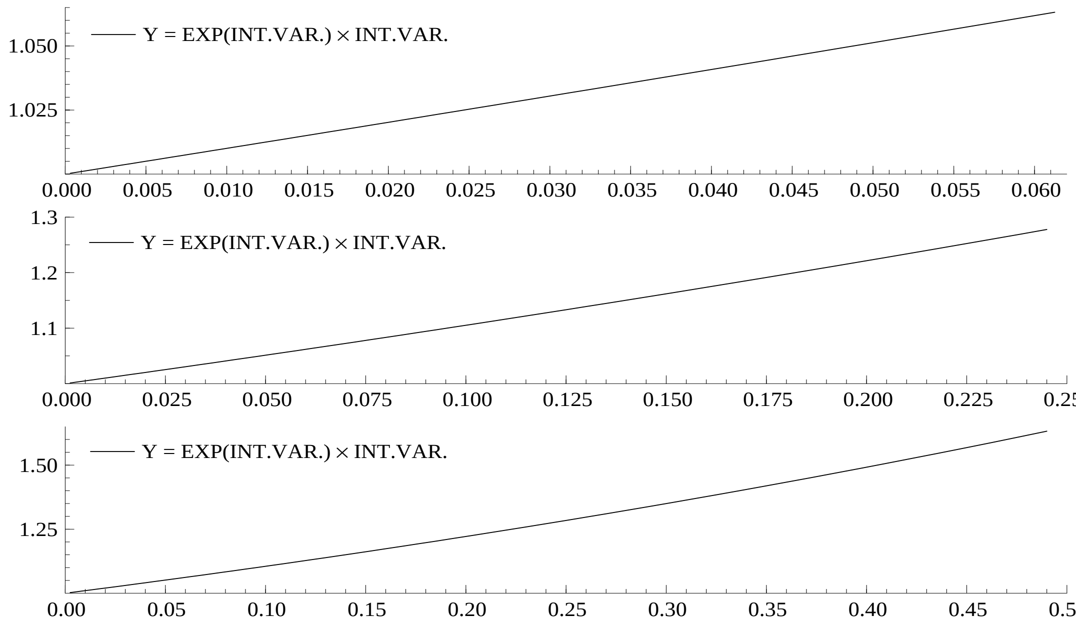


Figure 3.5: Plots of $Y_t = \exp(\tau_t \bar{\sigma}_t^2)$ against $\tau_t \bar{\sigma}_t^2$ for maturities $\tau_t = 0.125$ (1.5 month; upper plot), $\tau_t = 0.50$ (6 months) and $\tau_1 = 1$ (1 year), for $\bar{\sigma}_t \in [5\%, 70\%]$.

In independent, concurrent work, Jones (2003) essentially extracts information from option prices using equation (3.70), in which he *approximates* the conditional expectation $\mathbb{E}_{\mathbb{Q}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]$ by a tractable linear function of σ_t^2 and moreover adds noise. Jones considers estimation of the 1-factor SV models of Heston (1993), and his own CEV and 2GAM models. We refer to section 2.1 for more details.

Interpreting Black-Scholes implied volatilities
Equation (3.70) yields a convenient interpretation of Black-Scholes implied volatilities. A BS implied volatility may be interpreted (by approximation) as the average future stock volatility over the remaining option's life, under the *risk-neutral* measure. [37] In practice, BS implied volatilities are generally considered as being market forecasts of future stock volatility (under measure $\mathbb{P}$); see e.g. Hull (2003). We observe that this latter interpretation is actually only correct if volatility risk is not priced, as in that case the volatility processes under $\mathbb{P}$ and $\mathbb{Q}$ coincide. This is an important observation.

3.4.5 Higher-order approximations

If it turns out that a linearization performs poorly, or if improved accuracy is desired, higher-order approximations of the function $g(.)$ may be considered, which include curvature. These may also be considered if the aim is information extraction from different-moneyness option series, though this has its drawbacks, as explained earlier.

Consider for example a parabolic approximation of the function $g(.)$ in the point $Y_t = b_t^* = \exp(\tau_t \sigma_{implied,t}^2)$:

$$g(Y_t) = q_0(b_t^*) + q_1(b_t^*)Y_t + q_2(b_t^*)Y_t^2 + HOT2_t\,, \tag{3.71}$$

with

$$\begin{aligned} q_0(b_t^*) &\equiv g(b_t^*) - b_t^* g'(b_t^*) + \tfrac{1}{2}(b_t^*)^2 g''(b_t^*) \\ q_1(b_t^*) &\equiv g'(b_t^*) - b_t^* g''(b_t^*) \\ q_2(b_t^*) &\equiv \tfrac{1}{2} g''(b_t^*)\,, \end{aligned} \tag{3.72}$$

and in which $HOT2$ stands for the higher-order terms. An expression for $g'(.)$ is given in (3.66). The second-order derivative of $g(.)$ equals

$$\begin{aligned} g''(Y_t) &= \frac{S_t \exp(-\bar{q}_t \tau_t)}{2Y_t\sqrt{\ln Y_t}}\left[\frac{d\phi[d_{1t}(Y_t)]}{dd_{1t}(Y_t)}\frac{dd_{1t}(Y_t)}{dY_t} - \phi\left[d_{1t}(Y_t)\right]\left(\frac{1}{Y_t} + \frac{1}{2Y_t \ln Y_t}\right)\right] \\ &= -g'(Y_t)\left[d_{1t}(Y_t)\frac{dd_{1t}(Y_t)}{dY_t} + \frac{1}{Y_t} + \frac{1}{2Y_t \ln Y_t}\right] \end{aligned} \tag{3.73}$$

with

$$\frac{dd_{1t}(Y_t)}{dY_t} = \frac{\ln Y_t - 2[\ln \frac{S_t}{K_t} + (\bar{r}_t - \bar{q}_t)\tau_t]}{4Y_t(\ln Y_t)^{\frac{3}{2}}}\,. \tag{3.74}$$

(Here we used $d\phi[d_{1t}(Y_t)]/d(d_{1t}(Y_t)) = -d_{1t}(Y_t)\phi[d_{1t}(Y_t)]$.) To obtain some preliminary insight into the quality of a quadratic approximation, and its relation to a linear approximation, consider the earlier discussed example (see also figure 3.4).

[37] Assuming that real-world option prices obey the Hull and White (1987) model, Ghysels, Harvey and Renault (1996) arrive at the following approximation regarding the BS implied volatility of ATM options: $\sigma_{implied,t} \approx \mathbb{E}_{\mathbb{Q}}[\bar{\sigma}_t \mid \mathcal{F}_t]$. Their approximation is based on the fact that the Gaussian cumulative distribution function is roughly linear around zero.

Figure 3.6 plots the function $g(.)$, its tangent and its parabolic approximation in $Y_t = b_t^*$.

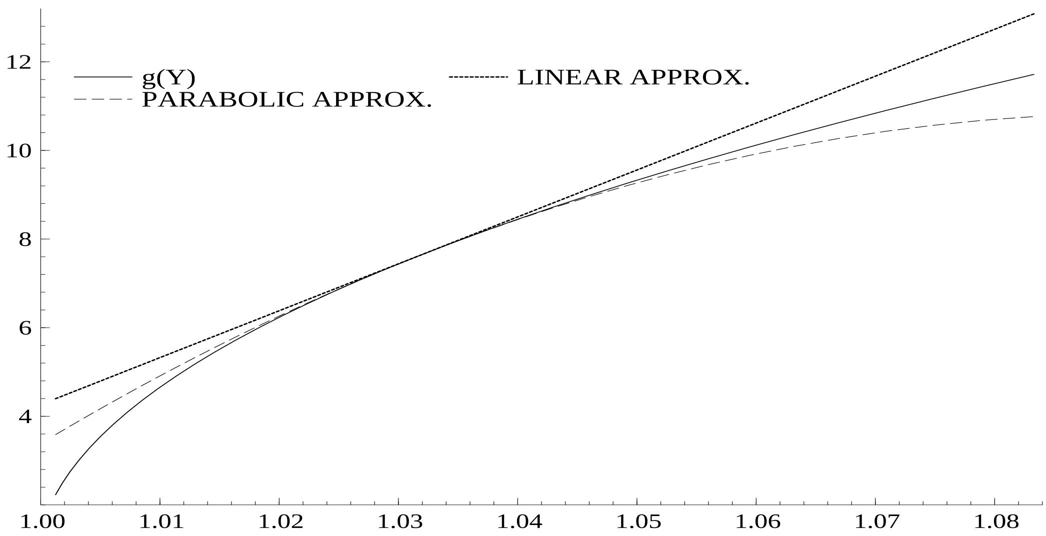


Figure 3.6: Function $g(.)$ and its linear and parabolic approximations (see also figure 3.4)

Although the linear approximation is consistently larger than $g(.)$, the parabolic approximation exceeds $g(.)$ for $Y_t < b_t^*$ and is smaller for $Y_t > b_t^*$. As it should, it picks up part of the curvature of $g(.)$ around b_t^*, and performs better in the neighborhood of the expansion point. Nevertheless, the differences do not seem very large in that area.

To derive a possible measurement equation based on a parabolic approximation, using (3.71), the call price $C_t = \mathbb{E}_{\mathbb{Q}}[g(Y_t) \mid \mathcal{F}_t]$ can be written as

$$\begin{aligned} C_t &= q_0(b_t^*) + q_1(b_t^*)\,\mathbb{E}_{\mathbb{Q}}\left[Y_t \mid \mathcal{F}_t\right] + q_2(b_t^*)\,\mathbb{E}_{\mathbb{Q}}\left[Y_t^2 \mid \mathcal{F}_t\right] + \mathbb{E}_{\mathbb{Q}}\left[HOT2_t \mid \mathcal{F}_t\right] \\ &= q_0(b_t^*) + q_1(b_t^*)\exp\left[A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\mathbf{x}_t\right] \\ &\qquad + q_2(b_t^*)\exp\left[A_2(\tau_t) + \mathbf{B}_2(\tau_t)'\mathbf{x}_t\right] + \mathbb{E}_{\mathbb{Q}}\left[HOT2_t \mid \mathcal{F}_t\right], \end{aligned} \tag{3.75}$$

where the second equality follows from the cMGF of the integrated variance; see section 10 of appendix B for details. The deterministic functions $A_{\mathrm{m}}(.)$ and $(nx1)$ $\mathbf{B}_{\mathrm{m}}(.)$ satisfy the following system of Ricatti ODEs for $\mathrm{m} = 1, 2$:

$$\begin{aligned} \frac{dA_{\mathrm{m}}(\tau_t)}{d\tau_t} &= \tilde{\boldsymbol{\theta}}'\tilde{\mathbf{K}}'\mathbf{B}_{\mathrm{m}}(\tau_t) + \tfrac{1}{2}\sum_{i=1}^{n}\left[\boldsymbol{\Sigma}'\mathbf{B}_{\mathrm{m}}(\tau_t)\right]_i^2 \alpha_i + \mathrm{m}\delta_0\,; \quad A_{\mathrm{m}}(0) = 0 \\ \frac{d\mathbf{B}_{\mathrm{m}}(\tau_t)}{d\tau_t} &= -\tilde{\mathbf{K}}'\mathbf{B}_{\mathrm{m}}(\tau_t) + \tfrac{1}{2}\sum_{i=1}^{n}\left[\boldsymbol{\Sigma}'\mathbf{B}_{\mathrm{m}}(\tau_t)\right]_i^2 \boldsymbol{\beta}_i + \mathrm{m}\boldsymbol{\delta}\,; \quad \mathbf{B}_{\mathrm{m}}(0) = \mathbf{0}, \end{aligned} \tag{3.76}$$

which can be solved for numerically. In the multifactor OU SV special case analytical expressions exist for these functions.

Equation (3.75) cannot further be rewritten in terms of an expression that is linear in the latent factors $\mathbf{x}_t$. Nor can it be rewritten in terms of an equation that involves $\sigma^2_{implied,t}$ on the left-hand side. The same holds if cubic or even higher-order approximations are considered. This is in clear contrast with the linear approximation. If we neglect the higher-order terms in $\mathbb{E}_{\mathbb{Q}}[HOT2_t \mid \mathcal{F}_t]$ and introduce noise as before, a measurement equation results that is *non-linear* in the state variables $\mathbf{x}_t^* = \mathbf{x}_t - \boldsymbol{\theta}$.

This non-linearity ensures that the conventional Kalman filter techniques cannot be applied straight away. Nonetheless, as suggested by Harvey (1989), a practical solution is to first linearize the measurement equation around a suitable estimate of the state, and then proceed with the linearized model as usual, as the linearized measurement equation fits again in the linear state space framework. Harvey (1989) proposes to linearize around the filtered state, $\mathbf{x}_{t|t-1}^*$. Durbin and Koopman (2001) also propose to linearize the measurement equation, but advocate a more sophisticated method: Their idea is to match the conditional modes of the state variables given the data, of the original non-linear and the linear approximating models. Evidently, the performance of these approaches depends on the degree of non-linearity, and on the accuracy of e.g. the filtered state $\mathbf{x}_{t|t-1}^*$ as an estimator of the true underlying state $\mathbf{x}_t^*$.

Relative merits of a parabolic versus a linear approximation
Both approaches of Harvey (1989) and Durbin and Koopman (2001) require the necessary additional computations. Moreover, the complexity of the computations involved increases as well. We may therefore wonder if considering a parabolic approximation is really worth the extra effort. Stated differently, do we lose a considerable amount of information by using a linear rather than a parabolic approximation, or may a linearization be considered sufficient? It would particularly be convenient if we had some measure that provides an indication of the *relative merits* of using a parabolic instead of a linear approximation, without a priori having to implement the associated non-linear state space model.

Reconsidering (3.75), we label $\mathbb{E}_{\mathbb{Q}}[HOT2_t \mid \mathcal{F}_t]$ the *parabolic error*. It represents the error incurred from approximating the call price by a value obtained via a parabolic approximation of the function $g(.)$. Some algebra shows that the parabolic error can be written in terms of the linearization error as

$$\mathbb{E}_{\mathbb{Q}}\left[HOT2_t \mid \mathcal{F}_t\right] = \mathbb{E}_{\mathbb{Q}}\left[HOT_t \mid \mathcal{F}_t\right] - \tfrac{1}{2} g''(b_t^*)\left(b_t^{*2} - 2b_t^* \,\mathbb{E}_{\mathbb{Q}}\left[Y_t \mid \mathcal{F}_t\right] + \mathbb{E}_{\mathbb{Q}}\left[Y_t^2 \mid \mathcal{F}_t\right]\right). \tag{3.77}$$

The parabolic error thus establishes some correction to the linearization error. Its magnitude can be investigated by simulation. Given a simulated time-series dataset for $t = \Delta t, .., T\Delta t$, the summary statistics of (the absolute value of) the relative parabolic error, $\mathbb{E}_{\mathbb{Q}}[HOT2_t \mid \mathcal{F}_t]/C_t$, provide information on how well the quadratic approximation performs.

To investigate the relative merits of using a parabolic rather than a linear approximation, the difference between the mean absolute relative linearization and parabolic errors seems a useful measure:

$$\text{Relative merits} = \left(\frac{1}{T}\sum_{t=\Delta t}^{T\Delta t} \frac{\left|\mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]\right| - \left|\mathbb{E}_{\mathbb{Q}}[HOT2_t \mid \mathcal{F}_t]\right|}{C_t}\right) * 100\% . \tag{3.78}$$

We expect this measure to be positive. If this *relative merits measure* is close to zero, it may not be considered worthwhile to use a parabolic approximation (if the single aim is improved accuracy), given the additional computations and increased complexity of implementation. However, if the aim is fitting the model to different-moneyness option series (i.e. to the volatility smile over time), a parabolic approximation should effectively be taken into account. The Monte Carlo studies in coming chapters report the magnitude of the relative merits measure based on simulated data.

3.5 Possible state space models

Given the discrete-time equations derived from the continuous-time multifactor affine SV option pricing model, several possible state space models can be considered for estimation. Section 3.5.1 first collects these equations together. Next, each of the sections 3.5.2 and 3.5.3 discusses two possible state space models in case either return – option, or RV – option data is considered for estimation. We explain their differences, and discuss why we favor one model above the other.

3.5.1 Refresher

As a refresher, recall the general linear state space representation from section 2.2. The measurement and transition equations are given by

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{a}_t + \boldsymbol{H}_t{}'\boldsymbol{\xi}_t + \boldsymbol{w}_t, \qquad && \boldsymbol{w}_t \sim (\mathbf{0}, \boldsymbol{R}), \\ \boldsymbol{\xi}_{t+\Delta t} &= \boldsymbol{d} + \boldsymbol{F}\boldsymbol{\xi}_t + \boldsymbol{v}_{t+\Delta t}, \qquad && \boldsymbol{v}_{t+\Delta t} \sim (\mathbf{0}, \boldsymbol{Q}), \qquad t = \Delta t, .., T\Delta t, \end{aligned} \tag{3.79}$$

with the white noise error series $\{\boldsymbol{w}_t\}$ and $\{\boldsymbol{v}_{t+\Delta t}\}$ being mutually uncorrelated at all points in time. For future reference, we will refer to this representation as the *unconditional* state space model.

In the foregoing sections, we derived the following discrete-time equations from the continuous-time multifactor SV model (recall that $\mathbf{x}_t^* = \mathbf{x}_t - \boldsymbol{\theta}$ denotes the factors in deviation from their mean):

Latent factors:

$$\mathbf{x}_{t+\Delta t}^* = \exp[-\mathbf{K}_d \Delta t]\mathbf{x}_t^* + \mathbf{u}_{t+\Delta t}, \qquad \mathbf{u}_{t+\Delta t} \sim \left(\mathbf{0},\ \mathbf{G}(\Delta t) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right) \tag{3.80}$$

Squared returns:

$$\begin{aligned} \frac{1}{\Delta t}\left(r_{t+\Delta t} - \hat{\mu}_t \Delta t\right)^2 &= \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t + \omega_{t+\Delta t} \\ &= \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} + \boldsymbol{\delta}'\mathbf{x}_t^* + \omega_{t+\Delta t}, \qquad \omega_{t+\Delta t} \sim (0, \sigma_\omega^2) \end{aligned} \tag{3.81}$$

Realized volatilities:

$$\sigma_{RV,t+\Delta t}^2 = \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} + \boldsymbol{\delta}'\left(\mathbf{K}_d \Delta t\right)^{-1}\left(\mathbf{I}_n - \exp[-\mathbf{K}_d \Delta t]\right)\mathbf{x}_t^* + \omega_{RV,t+\Delta t}, \tag{3.82}$$
$$\omega_{RV,t+\Delta t} \sim (0, \sigma_{\omega,RV}^2)$$

BS implied volatilities (call options):

$$\begin{aligned} \sigma_{implied,t}^2 &= \mu_v + \frac{A_1(\tau_t)}{\tau_t} + \frac{\mathbf{B}_1(\tau_t)'}{\tau_t}\mathbf{x}_t + \varepsilon_t \\ &= \mu_v + \frac{1}{\tau_t}\left[A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\boldsymbol{\theta}\right] + \frac{\mathbf{B}_1(\tau_t)'}{\tau_t}\mathbf{x}_t^* + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2). \end{aligned} \tag{3.83}$$

Regarding the co-dependence between the various disturbance series the following. The series $\{\mathbf{u}_t\}$ and $\{\omega_t\}$ are mutually uncorrelated (based on an Euler discretization of the stock price SDE). The errors $\{\mathbf{u}_t\}$ and $\{\omega_{RV,t}\}$ are uncorrelated at different points in time, but are contemporaneously correlated with correlation vector $\text{corr}_{\mathbb{P}}[\omega_{RV,t},\mathbf{u}_t]=\mathbf{c}$. By assumption, the option error series $\{\varepsilon_t\}$ is uncorrelated with all other error series.

3.5.2 System matrices for return - option data

Suppose the aim is extracting information from a series of daily stock returns and a time series of daily call option prices. Equations (3.80), (3.81) and (3.83) can be cast into the unconditional linear state space representation (3.79) in two different ways, each characterized by its own system matrices. A logical first set up is defining the system matrices as:

$$\boldsymbol{y}_t=\begin{bmatrix}\frac{1}{\Delta t}\left(r_{t+\Delta t}-\hat{\mu}_t\Delta t\right)^2\\ \sigma^2_{implied,t}\end{bmatrix}\qquad \boldsymbol{a}_t=\begin{bmatrix}\delta_0+\boldsymbol{\delta}'\boldsymbol{\theta}\\ \mu_V+\frac{1}{\tau_t}\left[A_1(\tau_t)+\mathbf{B}_1(\tau_t)'\boldsymbol{\theta}\right]\end{bmatrix}\qquad \boldsymbol{H}_t'=\begin{bmatrix}\boldsymbol{\delta}'\\ \frac{1}{\tau_t}\mathbf{B}_1(\tau_t)'\end{bmatrix}$$

$$\boldsymbol{\xi}_t=\mathbf{x}_t^*\qquad \boldsymbol{w}_t=\begin{bmatrix}\omega_{t+\Delta t}\\ \varepsilon_t\end{bmatrix}\qquad \boldsymbol{R}=\begin{bmatrix}\sigma^2_\omega & 0\\ 0 & \sigma^2_\varepsilon\end{bmatrix}\qquad \boldsymbol{d}=\mathbf{0}\tag{3.84}$$

$$\boldsymbol{F}=\exp\left[-\mathbf{K}_d\Delta t\right]\qquad \boldsymbol{v}_{t+\Delta t}=\mathbf{u}_{t+\Delta t}\qquad \boldsymbol{Q}=\mathbf{G}(\Delta t)\odot\boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'.$$

In this set up the state variables $\boldsymbol{\xi}_t$ are formed by the n unobserved volatility-driving factors (in deviation from their mean), $\mathbf{x}_t^*$. Representation (3.84) is fine if the single aim is parameter estimation. If a second aim is forecasting of future volatilities, this set up is of no use however. Kalman filter QML estimation of (3.84) implies that the BS implied variance observed at time t, but the squared return observed at time $t+\Delta t$ are used to predict the latent factors at time $t+\Delta t$. To be more specific, the Kalman filter produces the linear projection $\hat{\mathbb{E}}[\boldsymbol{\xi}_{t+\Delta t}\,|\,\boldsymbol{Y}_t]=\hat{\mathbb{E}}[\boldsymbol{\xi}_{t+\Delta t}\,|\,\boldsymbol{y}_t,\boldsymbol{y}_{t-\Delta t},..]$, which for (3.84) boils down to $\hat{\mathbb{E}}[\mathbf{x}^*_{t+\Delta t}\,|\,\frac{1}{\Delta t}\left(r_{t+\Delta t}-\hat{\mu}_t\Delta t\right)^2,\sigma^2_{implied,t},...]$. Hence, set up (3.84) uses time-$t+\Delta t$ information to predict a time-$t+\Delta t$ quantity. This is obviously not what we mean by true forecasting.

As a second possible set up, representation (3.85) circumvents this by using only time-t information in $\boldsymbol{y}_t=[\frac{1}{\Delta t}(r_t-\hat{\mu}_{t-\Delta t}\Delta t)^2,\sigma^2_{implied,t}]'$, to predict the state variables at time $t+\Delta t$. As a result, the state in (3.85) differs from the state in (3.84). The state is now defined as $\boldsymbol{\xi}_t=[\frac{1}{\Delta t}(r_t-\hat{\mu}_{t-\Delta t}\Delta t)^2,\ \mathbf{x}_t^*{}']'$, and contains besides the latent factors also the squared returns. A second consequence is that the first equation of the measurement equation $\boldsymbol{y}_t=\boldsymbol{a}_t+\boldsymbol{H}_t'\boldsymbol{\xi}_t+\boldsymbol{w}_t$ in (3.85) is now a trivial identity.

$$\boldsymbol{y}_t=\begin{bmatrix}\frac{1}{\Delta t}\left(r_t-\hat{\mu}_{t-\Delta t}\Delta t\right)^2\\ \sigma^2_{implied,t}\end{bmatrix}\qquad \boldsymbol{a}_t=\begin{bmatrix}0\\ \mu_V+\frac{1}{\tau_t}\left[A_1(\tau_t)+\mathbf{B}_1(\tau_t)'\boldsymbol{\theta}\right]\end{bmatrix}\qquad \boldsymbol{H}_t'=\begin{bmatrix}1 & \mathbf{0}'\\ 0 & \frac{1}{\tau_t}\mathbf{B}_1(\tau_t)'\end{bmatrix}$$

$$\boldsymbol{\xi}_t=\begin{bmatrix}\frac{1}{\Delta t}\left(r_t-\hat{\mu}_{t-\Delta t}\Delta t\right)^2\\ \mathbf{x}_t^*\end{bmatrix}\qquad \boldsymbol{w}_t=\begin{bmatrix}0\\ \varepsilon_t\end{bmatrix}\qquad \boldsymbol{R}=\begin{bmatrix}0 & 0\\ 0 & \sigma^2_\varepsilon\end{bmatrix}\qquad \boldsymbol{d}=\begin{bmatrix}\delta_0+\boldsymbol{\delta}'\boldsymbol{\theta}\\ \mathbf{0}\end{bmatrix}$$

$$\boldsymbol{F} = \begin{bmatrix} 0 & \boldsymbol{\delta}' \\ \mathbf{0} & \exp[-\mathbf{K}_d \Delta t] \end{bmatrix} \quad \boldsymbol{v}_{t+\Delta t} = \begin{bmatrix} \omega_{t+\Delta t} \\ \mathbf{u}_{t+\Delta t} \end{bmatrix} \quad \boldsymbol{Q} = \begin{bmatrix} \sigma_\omega^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{G}(\Delta t) \odot \boldsymbol{\Sigma} \mathbf{M}_d \boldsymbol{\Sigma}' \end{bmatrix}. \tag{3.85}$$

Given that set up (3.85) allows for parameter estimation and volatility forecasting at the same time, and is therefore arguably most relevant in practice, we will use representation (3.85) in our Monte Carlo study and empirical work in coming chapters.

3.5.3 System matrices for RV - option data

Suppose now that the aim is extracting information from a time series of realized volatilities and a time series of daily call option prices. Similar as in the return – option case, an intuitively logical first way of casting equations (3.80), (3.82) and (3.83) into the unconditional state space representation (3.79) seems perhaps:

$$\boldsymbol{y}_t = \begin{bmatrix} \sigma_{RV,t+\Delta t}^2 \\ \sigma_{implied,t}^2 \end{bmatrix} \qquad \boldsymbol{a}_t = \begin{bmatrix} \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} \\ \mu_v + \frac{1}{\tau_t}[A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\boldsymbol{\theta}] \end{bmatrix} \tag{3.86}$$

$$\boldsymbol{H}_t{}' = \begin{bmatrix} \boldsymbol{\delta}'(\mathbf{K}_d \Delta t)^{-1}(\mathbf{I}_n - \exp[-\mathbf{K}_d \Delta t]) \\ \frac{1}{\tau_t}\mathbf{B}_1(\tau_t)' \end{bmatrix} \qquad \boldsymbol{\xi}_t = \mathbf{x}_t^* \qquad \boldsymbol{w}_t = \begin{bmatrix} \omega_{RV,t+\Delta t} \\ \varepsilon_t \end{bmatrix}$$

$$\boldsymbol{R} = \begin{bmatrix} \sigma_{\omega,RV}^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix} \qquad \boldsymbol{d} = \mathbf{0} \qquad \boldsymbol{F} = \exp[-\mathbf{K}_d \Delta t]$$

$$\boldsymbol{v}_{t+\Delta t} = \mathbf{u}_{t+\Delta t} \qquad \boldsymbol{Q} = \mathbf{G}(\Delta t) \odot \boldsymbol{\Sigma}\mathbf{M}_d \boldsymbol{\Sigma}'.$$

Besides the fact that this set up does not allow for true volatility forecasting, there is one important caveat to take into account: Set up (3.86) does not exactly fit into framework (3.79), as the disturbances $\{\omega_{RV,t}\}$ and $\{\mathbf{u}_t\}$ are contemporaneously correlated. The error series $\{\boldsymbol{w}_t\}$ and $\{\boldsymbol{v}_{t+\Delta t}\}$ are therefore correlated as well.

The Kalman filter and smoother equations given in section 2.2 are based on the assumption of uncorrelated $\{\boldsymbol{w}_t\}$ and $\{\boldsymbol{v}_{t+\Delta t}\}$. If this assumption is violated, these equations do no longer apply. In case of contemporaneous correlation, they may nevertheless be adjusted to take this correlation into account; see e.g., Harvey (1989) for details. [38] Alternatively, Koopman (1993) and Koopman, Shephard and Doornik (1999) for example, prefer to start off with a general state space representation different from (3.79), in which there is an overall disturbance vector for the measurement and transition equation errors, and to work with selection matrices. In that case the correlation shows up as correlation between errors within that single disturbance vector. As a result, there is no need to adjust the Kalman filter equations that hold in that setting, as the possibility of contemporaneous correlation has already been accounted for from the outset.

[38] It should be noted that the original versions of the Kalman filter that appeared in the engineering literature (e.g. Kalman (1960)) already assumed correlated measurement and transition equation errors. See also the book of Anderson and Moore (1979).

In the remainder we will work with a set up different from (3.86), as it does not allow for real volatility prediction. The following representation of the system matrices bypasses this:

$$\boldsymbol{y}_t = \begin{bmatrix} \sigma^2_{RV,t} \\ \sigma^2_{implied,t} \end{bmatrix} \qquad \boldsymbol{a}_t = \begin{bmatrix} 0 \\ \mu_v + \frac{1}{\tau_t}[A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\boldsymbol{\theta}] \end{bmatrix} \qquad (3.87)$$

$$\boldsymbol{H}_t' = \begin{bmatrix} 1 & \mathbf{0}' \\ 0 & \frac{1}{\tau_t}\mathbf{B}_1(\tau_t)' \end{bmatrix} \qquad \boldsymbol{\xi}_t = \begin{bmatrix} \sigma^2_{RV,t} \\ \mathbf{x}^*_t \end{bmatrix} \qquad \boldsymbol{w}_t = \begin{bmatrix} 0 \\ \varepsilon_t \end{bmatrix} \qquad \boldsymbol{R} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma^2_\varepsilon \end{bmatrix}$$

$$\boldsymbol{d} = \begin{bmatrix} \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} \\ \mathbf{0} \end{bmatrix} \qquad \boldsymbol{F} = \begin{bmatrix} 0 & \boldsymbol{\delta}'(\mathbf{K}_d\Delta t)^{-1}(\mathbf{I}_n - \exp[-\mathbf{K}_d\Delta t]) \\ \mathbf{0} & \exp[-\mathbf{K}_d\Delta t] \end{bmatrix} \qquad \boldsymbol{v}_{t+\Delta t} = \begin{bmatrix} \omega_{RV,t+\Delta t} \\ \mathbf{u}_{t+\Delta t} \end{bmatrix}$$

$$\boldsymbol{Q} = \begin{bmatrix} \sigma^2_{\omega,RV} & \sigma_{\omega,RV}\mathbf{c}'(\mathbf{I}_n \odot \mathbf{G}(\Delta t) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}')^{\frac{1}{2}} \\ \sigma_{\omega,RV}(\mathbf{I}_n \odot \mathbf{G}(\Delta t) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}')^{\frac{1}{2}}\mathbf{c} & \mathbf{G}(\Delta t) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' \end{bmatrix}.$$

Similar as in (3.85), a trivial identity for the realized variance $\sigma^2_{RV,t}$ appears in the measurement equation of the resulting state space model. The equation for $\sigma^2_{RV,t}$ is incorporated in the transition equation. Notice that the correlation between $\omega_{RV,t+\Delta t}$ and $\mathbf{u}_{t+\Delta t}$ is subsumed in the transition equation disturbance $\boldsymbol{v}_{t+\Delta t}$. Set up (3.87) does fit in the unconditional linear state space framework (3.79), since the error series $\{\boldsymbol{w}_t\}$ and $\{\boldsymbol{v}_{t+\Delta t}\}$ are again uncorrelated at all points in time, as required. As such, (3.87) can be analyzed by the regular Kalman filter and smoother equations given in section 2.2.

The Kalman filter generates the linear projection of the state vector $\boldsymbol{\xi}_{t+\Delta t}$ at time $t+\Delta t$ on the observed data $\boldsymbol{Y}_t$ until time t, as usual. In this case the Kalman filter thus produces a forecast of both the realized variance and the latent factors at time $t+\Delta t$, given the observed series of realized and BS implied variances up to and including time t. Forecasts of future stock volatilities can next be obtained from the formula $\sigma^2_t = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t = \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} + \boldsymbol{\delta}'\mathbf{x}^*_t$.

4. Concluding remarks

For the unconditional state space models considered in section 3.5, the full distributions of the error series $\{\boldsymbol{w}_t\}$ and $\{\boldsymbol{v}_{t+\Delta t}\}$ are not known. As explained in section 2.2, we therefore estimate the parameters by Kalman filter QML. Apart from potential distortions as a result of the various approximations that we carried out to arrive at a linear state space representation, this procedure yields consistent and asymptotically normal estimates (though not efficient).

The state space formulations in section 3.5 suppose that only one call option series is analyzed besides the stock price series. Generalizing the formulations to incorporate more than one call option series simultaneously, is more or less obvious. It may then be assumed that the error terms belonging to each call option series have their own variance, and perhaps be cross-sectionally contemporaneously correlated, but uncorrelated over time. Incorporating put option series in the analysis is easily dealt with by exploiting the put-call parity.

Hence, first compute the prices of the corresponding artificial call options by $C_t = P_t + \exp(-\bar{r}_t \tau_t)(F_{t,T} - K_t)$, and then perform state space estimation as usual.

Prior to estimation, there is one caveat to take into account however. In the absence of sufficient parametric restrictions, not all of the parameters may be identified. The identification problem is contingent on both the specific SV process that is considered, and the type of data that is used for estimation; that is, only return, RV or option data, or combinations of these. For each specific SV case that we will consider in subsequent chapters, we take up the identification problem separately.

The unconditional state space representation (3.79) (and hence the set ups (3.84)-(3.87)) takes the unconditional distribution of the latent factors $\mathbf{x}$ into account only. Although the multifactor SV model allows for level-dependent volatility-of-volatility, this feature is not fully exploited in the state space formulations above. The fact that the volatility-of-volatility is level-dependent is clear, in discrete time, from the *conditional* variance of the volatility-factor shocks $\mathbf{u}_{t+\Delta t}$. This implies that when confronting the unconditional state space model with empirical data characterized by level-dependent volatility-of-volatility, that not full advantage is being taken of the structure in the data. A loss of information is likely to result, which in turn may imply less efficient estimates. Evidently, in that case there is just more information in the data than the unconditional state space model can pick up.

The unconditional state space model is nevertheless still useful to investigate for the following reasons. First, it may serve as a convenient benchmark or reference point when considering more complicated models. Second, in case the volatility is driven by a multifactor OU SV process, which implies that the volatility-of-volatility is level-*in*dependent, the conditional and unconditional distribution of the volatility-factor shocks $\mathbf{u}_{t+\Delta t}$ coincide, and hence the filter is exact (apart from the various performed approximations). Considering the OU SV case allows us to fully focus on the consequences regarding estimation bias, resulting from the diverse approximations that we carried out to arrive at a linear state space model.

In chapter V we propose a different state space formulation, labeled the *conditional* state space model, which is more explicitly designed to take level-dependent volatility-of-volatility into account. We then resort to the *Extended Kalman filter* for QML estimation. But first, in the next chapter, we study the performance of the unconditional state space model in the OU SV case.

Appendix

IIIa. Proof of $\{\omega_t\}$ being white noise and uncorrelated with $\{\mathbf{u}_{t+\Delta t}\}$

This appendix proves that the series $\{\omega_t\}$ is a white noise series. We also show that $\{\omega_t\}$ is uncorrelated with the daily volatility-factor shocks $\{\mathbf{u}_{t+\Delta t}\}$. For simplicity we treat the Euler discretization of the stock price SDE as exact in the derivations below; i.e., $r_{t+\Delta t} = \mu_t \Delta t + \sigma_t \eta_{t+\Delta t}$ with $\eta_{t+\Delta t} = W_{S,t+\Delta t} - W_{S,t}$.

From (3.18), $\omega_{t+\Delta t}$ can be written as

$$\omega_{t+\Delta t} = \frac{1}{\Delta t}\left(r_{t+\Delta t} - \mu_t \Delta t\right)^2 - \sigma_t^2 = \sigma_t^2\left(\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1\right). \tag{a.1}$$

From decomposition (3.18), by construction $\{\omega_t\}$ has conditional –and hence unconditional- mean zero under $\mathbb{P}$: $\mathbb{E}_{\mathbb{P}}[\omega_{t+\Delta t} \mid \mathcal{F}_t] = 0 = \mathbb{E}_{\mathbb{P}}[\omega_{t+\Delta t}]$. The variance of $\omega_{t+\Delta t}$ equals

$$\text{var}_{\mathbb{P}}\left[\omega_{t+\Delta t}\right] = \mathbb{E}_{\mathbb{P}}\left[\sigma_t^4\left(\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1\right)^2\right] = \mathbb{E}_{\mathbb{P}}\left[\sigma_t^4\right]\mathbb{E}_{\mathbb{P}}\left[\left(\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1\right)^2\right], \tag{a.2}$$

in which the latter equality exploits the independence between σ_t and $\eta_{t+\Delta t}$. Now,

$$\mathbb{E}_{\mathbb{P}}\left[\left(\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1\right)^2\right] = \mathbb{E}_{\mathbb{P}}\left[\frac{\eta_{t+\Delta t}^4}{\Delta^2 t} - 2\frac{\eta_{t+\Delta t}^2}{\Delta t} + 1\right] = \frac{3\Delta^2 t}{\Delta^2 t} - 2\frac{\Delta t}{\Delta t} + 1 = 2, \tag{a.3}$$

as $\mathbb{E}_{\mathbb{P}}[\eta_{t+\Delta t}^4] = 3\Delta^2 t$. Also,

$$\mathbb{E}_{\mathbb{P}}\left[\sigma_t^4\right] = \mathbb{E}_{\mathbb{P}}\left[\left(\delta_0 + \boldsymbol{\delta}'\mathbf{x}_t\right)^2\right] = \delta_0^2 + 2\delta_0\boldsymbol{\delta}'\boldsymbol{\theta} + \boldsymbol{\delta}'\left(\text{var}_{\mathbb{P}}\left[\mathbf{x}_t\right] + \boldsymbol{\theta}\boldsymbol{\theta}'\right)\boldsymbol{\delta}. \tag{a.4}$$

Section 5 of appendix B shows that the unconditional variance of the factors equals

$$\text{var}_{\mathbb{P}}\left[\mathbf{x}_t\right] = \mathbf{J} \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}', \tag{a.5}$$

in which $\mathbf{J}$ is an (nxn) matrix having its ij – element equal to $[\mathbf{J}]_{ij} = 1/(k_i + k_j)$ (see also appendix A). The symbol $\odot$ represents the Hadamard product; i.e., element-by-element multiplication. The variance of $\omega_{t+\Delta t}$ then equals

$$\text{var}_{\mathbb{P}}\left[\omega_{t+\Delta t}\right] = 2\delta_0^2 + 4\delta_0\boldsymbol{\delta}'\boldsymbol{\theta} + 2\boldsymbol{\delta}'\left(\mathbf{J} \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' + \boldsymbol{\theta}\boldsymbol{\theta}'\right)\boldsymbol{\delta}. \tag{a.6}$$

The series $\{\omega_t\}$ is thus homoskedastic. Consider next the autocovariance of order $p = 1, 2, \ldots$,

$$\begin{aligned}\text{cov}_{\mathbb{P}}\left[\omega_{t+\Delta t}, \omega_{t-(p-1)\Delta t}\right] &= \mathbb{E}_{\mathbb{P}}\left[\sigma_t^2\left(\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1\right)\sigma_{t-p\Delta t}^2\left(\frac{\eta_{t-(p-1)\Delta t}^2}{\Delta t} - 1\right)\right] \\ &= \mathbb{E}_{\mathbb{P}}\left[\sigma_t^2\sigma_{t-p\Delta t}^2\right]\mathbb{E}_{\mathbb{P}}\left[\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1\right]\mathbb{E}_{\mathbb{P}}\left[\frac{\eta_{t-(p-1)\Delta t}^2}{\Delta t} - 1\right] = 0.\end{aligned} \tag{a.7}$$

Here we exploited the fact that the variance process $\{\sigma_t^2\}$ is independent of $\{\eta_t\}$ and that $\eta_{t+\Delta t} \sim$ i.i.d. $\mathcal{N}(0, \Delta t)$ under $\mathbb{P}$. The series $\{\omega_t\}$ is thus serially uncorrelated. To conclude, we find $\{\omega_t\}$ to be a white noise series.

Regarding the correlation between the error series $\{\omega_t\}$ and the daily volatility-factor shocks $\{\mathbf{u}_{t+\Delta t}\}$, we find for their covariance, $\forall t, s$:

$$\begin{aligned}\text{cov}_{\mathbb{P}}\left[\omega_{t+\Delta t}, \mathbf{u}_{s+\Delta t}\right] &= \mathbb{E}_{\mathbb{P}}\left[\sigma_t^2\left(\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1\right)\mathbf{u}_{s+\Delta t}\right] \\ &= \mathbb{E}_{\mathbb{P}}\left[\sigma_t^2\mathbf{u}_{s+\Delta t}\right]\mathbb{E}_{\mathbb{P}}\left[\frac{\eta_{t+\Delta t}^2}{\Delta t} - 1\right] = \mathbf{0},\end{aligned} \tag{a.8}$$

such that these series are uncorrelated at all points in time. The argument is as follows. Both the error term $\mathbf{u}_{s+\Delta t}$ and the stock variance σ_t^2 are only determined from the Brownian motion $\mathbf{W}_x$, and not from the Brownian motion driving the stock price, W_S, which determines $\eta_{t+\Delta t}$ completely. Therefore, the product $\sigma_t^2\mathbf{u}_{s+\Delta t}$ is independent of $\eta_{t+\Delta t}$ as $\mathbf{W}_x$ and W_S are independent, from which the result follows.

IIIb. Investigating the correlation between $\{\varpi_{RV,t+\Delta t}\}$ and $\{\mathbf{u}_{t+\Delta t}\}$

This appendix proves that the series $\{\varpi_{RV,t+\Delta t}\}$ is non-contemporaneously uncorrelated, but contemporaneously correlated, with the series $\{\mathbf{u}_{t+\Delta t}\}$ of daily volatility-factor shocks.

Non-contemporaneous correlation

Concerning the non-contemporaneous correlation between the error series $\{\varpi_{RV,t+\Delta t}\}$ and $\{\mathbf{u}_{t+\Delta t}\}$, their covariance of order $p = 1,2,..$ equals

$$\begin{aligned}\mathrm{cov}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}, \mathbf{u}_{t-(p-1)\Delta t}\right] &= \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}}\left(\varpi_{RV,t+\Delta t}\, \mathbf{u}_{t-(p-1)\Delta t} \mid \mathcal{F}_t\right)\right] \qquad \text{(b.1)}\\ &= \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}}\left(\varpi_{RV,t+\Delta t} \mid \mathcal{F}_t\right)\mathbf{u}_{t-(p-1)\Delta t}\right] = \mathbf{0}\,,\end{aligned}$$

such that the errors at different points in time are mutually uncorrelated.

Contemporaneous correlation

To derive their contemporaneous correlation, $\mathrm{corr}_{\mathbb{P}}[\varpi_{RV,t+\Delta t}, \mathbf{u}_{t+\Delta t}]$, recall first from (3.11) - (3.12) that the factors in deviation from their mean evolve in discrete time under $\mathbb{P}$ as $\mathbf{x}^*_{t+\Delta t} = \exp[-\mathbf{K}_d\Delta t]\mathbf{x}^*_t + \mathbf{u}_{t+\Delta t}$ with

$$\mathbf{u}_{t+\Delta t} = \int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t+\Delta t-u)]\mathbf{\Sigma\Lambda}_u d\mathbf{W}_{x,u} = \int_t^{t+\Delta t} \mathbf{C}_d(u,t+\Delta t)\mathbf{\Sigma\Lambda}_u d\mathbf{W}_{x,u}\,. \qquad (\mathbb{P}) \qquad \text{(b.2)}$$

(See also appendix A for a definition of the matrix function $\mathbf{C}_d(.,.)$.) Notice next that, by appendix B, (9.5) and (9.17), the error $\varpi_{RV,t+\Delta t}$ can be written as

$$\begin{aligned}\varpi_{RV,t+\Delta t} &= \frac{1}{\Delta t}\int_t^{t+\Delta t} \sigma_s^2 ds - \mathbb{E}_{\mathbb{P}}\left[\frac{1}{\Delta t}\int_t^{t+\Delta t} \sigma_s^2 ds \mid \mathcal{F}_t\right] \qquad (\mathbb{P}) \qquad \text{(b.3)}\\ &= \frac{\boldsymbol{\delta}'}{\Delta t}\int_t^{t+\Delta t} \mathbf{u}_{t,s} ds = \frac{\boldsymbol{\delta}'}{\Delta t}\int_t^{t+\Delta t} \mathbf{D}(u,t+\Delta t)\mathbf{\Sigma\Lambda}_u d\mathbf{W}_{x,u}\,,\end{aligned}$$

in which the expression for the integrated disturbance term $\int_t^{t+\Delta t}\mathbf{u}_{t,s}ds$ in the last equality follows from appendix B, (9.6). The first steps in deriving the covariance between $\varpi_{RV,t+\Delta t}$ and $\mathbf{u}_{t+\Delta t}$ then become [39]

$$\begin{aligned}&\mathrm{cov}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}, \mathbf{u}_{t+\Delta t}\right] \qquad \text{(b.4)}\\ &= \mathrm{cov}_{\mathbb{P}}\left[\frac{\boldsymbol{\delta}'}{\Delta t}\int_t^{t+\Delta t} \mathbf{D}(u,t+\Delta t)\mathbf{\Sigma\Lambda}_u d\mathbf{W}_{x,u}, \int_t^{t+\Delta t} \mathbf{C}_d(u,t+\Delta t)\mathbf{\Sigma\Lambda}_u d\mathbf{W}_{x,u}\right]\end{aligned}$$

[39] The second equality in this derivation follows from the fact that it can be shown that for two arbitrary random vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ and vector of constants $\mathbf{a}$ (of equal length as $\mathbf{v}_1$), it holds that the column vector of covariances $\mathrm{cov}[\mathbf{a}'\mathbf{v}_1, \mathbf{v}_2]$ can be written as $\mathrm{cov}[\mathbf{a}'\mathbf{v}_1, \mathbf{v}_2] = \mathrm{cov}[\mathbf{v}_2, \mathbf{v}_1'\mathbf{a}] = \mathbb{E}[\mathbf{v}_2\mathbf{v}_1'\mathbf{a}] - \mathbb{E}[\mathbf{v}_2]\mathbb{E}[\mathbf{v}_1'\mathbf{a}] = \mathrm{cov}[\mathbf{v}_2, \mathbf{v}_1]\mathbf{a}$. Notice that it furthermore holds that $\mathrm{cov}[\mathbf{v}_1, \mathbf{v}_2] = (\mathrm{cov}[\mathbf{v}_2, \mathbf{v}_1])'$.

$$= \mathrm{cov}_{\mathbb{P}}\left[\int_t^{t+\Delta t} \mathbf{C}_d(u,t+\Delta t)\mathbf{\Sigma\Lambda}_u d\mathbf{W}_{x,u}, \int_t^{t+\Delta t} \mathbf{D}(u,t+\Delta t)\mathbf{\Sigma\Lambda}_u d\mathbf{W}_{x,u}\right]\frac{\boldsymbol{\delta}}{\Delta t}$$

$$= \mathbb{E}_{\mathbb{P}}\left[\left(\int_t^{t+\Delta t} \mathbf{C}_d(u,t+\Delta t)\mathbf{\Sigma\Lambda}_u d\mathbf{W}_{x,u}\right)\left(\int_t^{t+\Delta t} \mathbf{D}(u,t+\Delta t)\mathbf{\Sigma\Lambda}_u d\mathbf{W}_{x,u}\right)'\right]\frac{\boldsymbol{\delta}}{\Delta t}.$$

In order to further simplify, think about the two Itô integrals in the latter expression as (limits of) discrete sums of weighted Brownian increments, each with respect to the same Brownian motion. When multiplying these sums with each other, cross-terms disappear, as non-overlapping Brownian increments are independent, such that their product equals zero. [40] We then obtain

$$\begin{aligned} &\mathrm{cov}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}, \mathbf{u}_{t+\Delta t}\right] \qquad\qquad (b.5)\\ &= \mathbb{E}_{\mathbb{P}}\left[\int_t^{t+\Delta t} \mathbf{C}_d(u,t+\Delta t)\mathbf{\Sigma\Lambda}_u d\mathbf{W}_{x,u} d\mathbf{W}_{x,u}'\mathbf{\Lambda}_u\mathbf{\Sigma}'\mathbf{D}(u,t+\Delta t)\right]\frac{\boldsymbol{\delta}}{\Delta t}\\ &= \mathbb{E}_{\mathbb{P}}\left[\int_t^{t+\Delta t} \mathbf{C}_d(u,t+\Delta t)\mathbf{\Sigma\Lambda}_u^2\mathbf{\Sigma}'\mathbf{D}(u,t+\Delta t)du\right]\frac{\boldsymbol{\delta}}{\Delta t}\\ &= \left(\int_t^{t+\Delta t} \mathbf{C}_d(u,t+\Delta t)\mathbf{\Sigma M}_d\mathbf{\Sigma}'\mathbf{D}(u,t+\Delta t)du\right)\frac{\boldsymbol{\delta}}{\Delta t}\\ &= \left[\left(\int_t^{t+\Delta t} \mathbf{C}_d(u,t+\Delta t)\mathbf{11}'\mathbf{D}(u,t+\Delta t)du\right)\odot\mathbf{\Sigma M}_d\mathbf{\Sigma}'\right]\frac{\boldsymbol{\delta}}{\Delta t}. \end{aligned}$$

The second equality uses $d\mathbf{W}_{x,u}d\mathbf{W}_{x,u}' = \mathbf{I}_n du$, which holds due to the fact that the individual Brownian motions in the vector standard Brownian motion $\mathbf{W}_x$ are independent. The third equality uses $\mathbb{E}_{\mathbb{P}}[\mathbf{\Lambda}_t^2] = \mathrm{diag}[\alpha_1 + \boldsymbol{\beta}_1'\boldsymbol{\theta},..,\alpha_n + \boldsymbol{\beta}_n'\boldsymbol{\theta}] = \mathbf{M}_d$, see appendix A. The final equality invokes the lemma stated in appendix A, such that $\mathbf{C}_d(u,t+\Delta t)\mathbf{\Sigma M}_d\mathbf{\Sigma}'\mathbf{D}(u,t+\Delta t) = \mathbf{C}_d(u,t+\Delta t)\mathbf{11}'\mathbf{D}(u,t+\Delta t)\odot\mathbf{\Sigma M}_d\mathbf{\Sigma}'$. Before proceeding, let us tackle the integral in the latter expression first:

$$\begin{aligned} &\int_t^{t+\Delta t} \mathbf{C}_d(u,t+\Delta t)\mathbf{11}'\mathbf{D}(u,t+\Delta t)du \qquad\qquad (b.6)\\ &= \int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t+\Delta t-u)]\mathbf{11}'(\mathbf{I}_n - \exp[-\mathbf{K}_d(t+\Delta t-u)])\mathbf{K}_d^{-1}du\\ &= \left(\int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t+\Delta t-u)]du\right)\mathbf{11}'\mathbf{K}_d^{-1}\\ &\quad - \left(\int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t+\Delta t-u)]\mathbf{11}'\exp[-\mathbf{K}_d(t+\Delta t-u)]du\right)\mathbf{K}_d^{-1}\\ &= \left[\left(\int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t+\Delta t-u)]du\right)\mathbf{11}' - \mathbf{G}(t,t+\Delta t)\right]\mathbf{K}_d^{-1}, \end{aligned}$$

[40] Think also about the Itô isometry to compute the variance of an Itô integral. Alternatively, one may generalize the results in section 4.4 of Etheridge (2002) to the multivariate case, to see that this first simplification indeed holds.

with $\mathbf{G}(t, t+\Delta t) = \mathbf{G}(0, \Delta t) = \mathbf{G}(\Delta t)$ (see appendix A again). In the first equality we used $\mathbf{D}(u, t+\Delta t) = \mathbf{K}_d^{-1}(\mathbf{I}_n - \exp[-\mathbf{K}_d(t+\Delta t - u)]) = (\mathbf{I}_n - \exp[-\mathbf{K}_d(t+\Delta t-u)])\mathbf{K}_d^{-1}$, which is allowed as all matrices involved are diagonal. By changing the integration variable from u to $v \equiv t + \Delta t - u$, such that $dv = -du$, the integral in the latter expression becomes

$$\int_t^{t+\Delta t} \exp[-\mathbf{K}_d(t+\Delta t - u)]du = -\int_{\Delta t}^{0} \exp[-\mathbf{K}_d v]dv = \int_0^{\Delta t} \exp[-\mathbf{K}_d v]dv = \mathbf{D}(\Delta t). \quad \text{(b.7)}$$

Collecting together, we find for the $(nx1)$ covariance vector between $\varpi_{RV,t+\Delta t}$ and $\mathbf{u}_{t+\Delta t}$:

$$\mathrm{cov}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}, \mathbf{u}_{t+\Delta t}\right] = \left(\left[\mathbf{D}(\Delta t)\mathbf{1}\mathbf{1}' - \mathbf{G}(\Delta t)\right]\mathbf{K}_d^{-1} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right)\frac{\boldsymbol{\delta}}{\Delta t}. \quad \text{(b.8)}$$

Using (3.13) and (3.33), the $(nx1)$ contemporaneous correlation vector between the series $\{\varpi_{RV,t+\Delta t}\}$ and $\{\mathbf{u}_{t+\Delta t}\}$ then finally follows as

$$\begin{aligned}
&\mathrm{corr}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}, \mathbf{u}_{t+\Delta t}\right] \quad \text{(b.9)}\\
&= \left(\mathrm{var}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}\right]\right)^{-\frac{1}{2}}\left(\mathrm{var}_{\mathbb{P}}\left[\mathbf{u}_{t+\Delta t}\right] \odot \mathbf{I}_n\right)^{-\frac{1}{2}} \mathrm{cov}_{\mathbb{P}}\left[\varpi_{RV,t+\Delta t}, \mathbf{u}_{t+\Delta t}\right]\\
&= \left(\frac{1}{\Delta^2 t}\boldsymbol{\delta}'\left[\mathbf{N}(\Delta t) \odot \mathbf{\Sigma M}_d \mathbf{\Sigma}'\right]\boldsymbol{\delta}\right)^{-\frac{1}{2}} \left[\mathbf{I}_n \odot \mathbf{G}(\Delta t) \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right]^{-\frac{1}{2}}\\
&\qquad * \left(\left[\mathbf{D}(\Delta t)\mathbf{1}\mathbf{1}' - \mathbf{G}(\Delta t)\right]\mathbf{K}_d^{-1} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right)\frac{\boldsymbol{\delta}}{\Delta t}.
\end{aligned}$$

The contemporaneous correlation is not equal to zero.

IIIc. The expected value of the realized variance $\sigma^2_{RV,t+\Delta t}$

This appendix calculates the mean of the realized variance $\sigma^2_{RV,t+\Delta t}$ if a finite number of intraday time points I is considered, in the general multifactor SV model. For simplicity we assume the mean rate of stock return to be constant: $\mu_t = \mu \ \forall t$.

Recall the definition of the realized variance over $[t, t+\Delta t]$:

$$\sigma^2_{RV,t+\Delta t} \equiv \frac{1}{\Delta t}\sum_{i=1}^{I}\left(\ln\frac{S_{t+i\delta t}}{S_{t+(i-1)\delta t}}\right)^2. \quad (\mathbb{P}) \quad \text{(c.1)}$$

To compute its expected value, we first focus on the logreturn $R_{t+\delta t}$ over the interval $[t, t+\delta t]$ of length $\delta t = \Delta t / I$. From Itô's lemma, the log stock price follows the SDE $d\ln S_t = (\mu - \frac{1}{2}\sigma_t^2)dt + \sigma_t dW_{S,t}$ under $\mathbb{P}$, such that

$$R_{t+\delta t} \equiv \ln\frac{S_{t+\delta t}}{S_t} = \mu\,\delta t - \tfrac{1}{2}\int_t^{t+\delta t}\sigma_u^2 du + \int_t^{t+\delta t}\sigma_u dW_{S,u}. \quad (\mathbb{P}) \quad \text{(c.2)}$$

The squared logreturn over $[t, t+\delta t]$ can be written as

$$R^2_{t+\delta t} = \mu^2\delta t^2 - \mu\,\delta t\left(\int_t^{t+\delta t}\sigma_u^2 du\right) + 2\mu\,\delta t\left(\int_t^{t+\delta t}\sigma_u dW_{S,u}\right) \quad (\mathbb{P}) \quad \text{(c.3)}$$

$$+\tfrac{1}{4}\left(\int_t^{t+\delta t}\sigma_u^2 du\right)^2-\left(\int_t^{t+\delta t}\sigma_u^2 du\right)\left(\int_t^{t+\delta t}\sigma_u dW_{S,u}\right)+\left(\int_t^{t+\delta t}\sigma_u dW_{S,u}\right)^2.$$

We are interested in its expected value. We first compute the expected value of the five integrals occurring in (c.3). The integrated variance has an expected value of

$$\mathbb{E}_{\mathbb{P}}\left[\int_t^{t+\delta t}\sigma_u^2 du\right]=\int_t^{t+\delta t}\mathbb{E}_{\mathbb{P}}[\sigma_u^2]du=(\delta_0+\boldsymbol{\delta}'\boldsymbol{\theta})\delta t\,, \qquad (c.4)$$

whereas the expectation of the second (i.e., the Itô) integral is simply 0. The expected value of the squared integrated variance equals

$$\mathbb{E}_{\mathbb{P}}\left[\left(\int_t^{t+\delta t}\sigma_u^2 du\right)^2\right]=\mathrm{var}_{\mathbb{P}}\left[\int_t^{t+\delta t}\sigma_u^2 du\right]+\left(\mathbb{E}_{\mathbb{P}}\left[\int_t^{t+\delta t}\sigma_u^2 du\right]\right)^2 \qquad (c.5)$$
$$=\boldsymbol{\delta}'\left[\left(\mathbf{N}(\delta t)+\mathbf{D}(\delta t)\mathbf{1}\mathbf{1}'\mathbf{D}(\delta t)\odot\mathbf{J}\right)\odot\boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right]\boldsymbol{\delta}+(\delta_0+\boldsymbol{\delta}'\boldsymbol{\theta})^2\delta t^2.$$

The second equality follows from the moments of the integrated variance; see section 9 in appendix B. By conditioning on the $\sigma-$field generated by the Brownian motion driving the stock variance ($\mathbf{W}_x$) over the interval $[t,t+\delta t]$, it follows that

$$\mathbb{E}_{\mathbb{P}}\left[\left(\int_t^{t+\delta t}\sigma_u^2 du\right)\left(\int_t^{t+\delta t}\sigma_u dW_{S,u}\right)\right]=0\,. \qquad (c.6)$$

Finally, from the Itô isometry, the mean of the fifth integral in (c.3) equals

$$\mathbb{E}_{\mathbb{P}}\left[\left(\int_t^{t+\delta t}\sigma_u dW_{S,u}\right)^2\right]=\int_t^{t+\delta t}\mathbb{E}_{\mathbb{P}}\left[\sigma_u^2\right]du=(\delta_0+\boldsymbol{\delta}'\boldsymbol{\theta})\delta t\,. \qquad (c.7)$$

Collecting the intermediate results and rearranging yields the expected value of the squared logreturn over $[t,t+\delta t]$:

$$\mathbb{E}_{\mathbb{P}}\left[R_{t+\delta t}^2\right]=(\delta_0+\boldsymbol{\delta}'\boldsymbol{\theta})\delta t+\left[\mu-\tfrac{1}{2}(\delta_0+\boldsymbol{\delta}'\boldsymbol{\theta})\right]^2\delta t^2 \qquad (c.8)$$
$$+\tfrac{1}{4}\boldsymbol{\delta}'\left[\left(\mathbf{N}(\delta t)+\mathbf{D}(\delta t)\mathbf{1}\mathbf{1}'\mathbf{D}(\delta t)\odot\mathbf{J}\right)\odot\boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right]\boldsymbol{\delta}\,.$$

The expected value of the realized variance then becomes

$$\mathbb{E}_{\mathbb{P}}\left[\sigma_{RV,t+\Delta t}^2\right]=\mathbb{E}_{\mathbb{P}}\left[\frac{1}{\Delta t}\sum_{i=1}^{I}\left(\ln\frac{S_{t+i\delta t}}{S_{t+(i-1)\delta t}}\right)^2\right]=\frac{I}{\Delta t}\mathbb{E}_{\mathbb{P}}\left[R_{t+\delta t}^2\right] \qquad (c.9)$$
$$=\delta_0+\boldsymbol{\delta}'\boldsymbol{\theta}+\left[\mu-\tfrac{1}{2}(\delta_0+\boldsymbol{\delta}'\boldsymbol{\theta})\right]^2\delta t$$
$$+\tfrac{1}{4}\boldsymbol{\delta}'\left[\left(\frac{\mathbf{N}(\delta t)}{\delta t}+\frac{\mathbf{D}(\delta t)}{\delta t}\mathbf{1}\mathbf{1}'\mathbf{D}(\delta t)\odot\mathbf{J}\right)\odot\boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right]\boldsymbol{\delta}\,,$$

in which the second equality uses $I/\Delta t=1/\delta t$.

Notice that the realized variance $\sigma^2_{RV,t+\Delta t}$ is a biased estimator of the mean stock return variance $\mathbb{E}_{\mathbb{P}}[\sigma^2_t] = \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta}$. Hence, it is also a biased estimator of the mean average variance $\mathbb{E}_{\mathbb{P}}[\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma^2_u du]$, which equals $\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta}$ as well. However, as it can be shown that $\lim_{\delta t\to 0}\mathbf{N}(\delta t)/\delta t = \mathbf{0}$, $\lim_{\delta t\to 0}\mathbf{D}(\delta t)/\delta t = \mathbf{I}_n$ and $\lim_{\delta t\to 0}\mathbf{D}(\delta t) = \mathbf{0}$ (see appendix A), it is clear that this bias disappears for $\delta t \to 0$ (i.e., $I \to \infty$):

$$\lim_{\delta t\to 0}\mathbb{E}_{\mathbb{P}}\left[\sigma^2_{RV,t+\Delta t}\right] = \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} = \mathbb{E}_{\mathbb{P}}\left[\sigma^2_t\right] = \mathbb{E}_{\mathbb{P}}\left[\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma^2_u du\right]. \qquad \text{(c.10)}$$

This is conform our intuitions. Hence, the mean realized variance only coincides with the mean average variance if $I \to \infty$ (or $\delta t \to 0$).

A related way of clarifying the same issue is as follows. Recall the stochastic convergence of the random variable $\sigma^2_{RV,t+\Delta t}$ towards the random average variance over $[t, t+\Delta t]$,

$$\operatorname*{plim}_{I\to\infty}\sigma^2_{RV,t+\Delta t} = \frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma^2_u du\,. \qquad (\mathbb{P}) \qquad \text{(c.11)}$$

Taking expectations yields

$$\mathbb{E}_{\mathbb{P}}\left[\operatorname*{plim}_{I\to\infty}\sigma^2_{RV,t+\Delta t}\right] = \mathbb{E}_{\mathbb{P}}\left[\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma^2_u du\right] = \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta}\,, \qquad \text{(c.12)}$$

which clarifies that in the limit for I approaching infinity (or δt approaching 0), the realized variance is an unbiased estimator of $\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta}$. In our earlier explanation we first took expectations and then limits, whereas here it is the other way round. The operators may thus be interchanged here.

- IV -

Monte Carlo and Empirical Results for Ornstein-Uhlenbeck SV: Examining UK Financial Markets

1. Introduction

In this chapter we consider a special case of the multifactor affine SV option pricing model, in which the volatility is driven by one or more Ornstein-Uhlenbeck (OU) processes. This implies that the volatility-of-volatility is constant, i.e., the volatility-of-volatility does not depend on the current volatility level itself.

We first examine the performance of the unconditional state space model as a method for parameter estimation and volatility filtering by Monte Carlo simulation, for 1-factor OU SV. We then confront the OU model to real-world financial markets data. We study the major stock index from the United Kingdom, the *FTSE100 index*, and the European option contract traded on that index. The data consists of daily index returns, and three *at-the-money* call option series of various maturities: a *short*, *medium* and *long-maturity* series.

An obvious drawback of the Gaussian OU assumption is that the stock variance σ_t^2 can become negative. Although fitting the model to empirical data is not expected to pose problems in this respect, in a simulation environment in which multiple paths of $\{\sigma_t^2\}$ are generated, sample paths may well turn negative.

Although the possibility of sampling negative stock variances is theoretically inconsistent, there are a number of reasons why we think investigating the OU SV case is still of value. First, the OU assumption yields large analytical tractability. [1] In our multifactor SV estimations considered below in which multiple option series are included, this tractability is very convenient. Second, suppose we find that the volatility is driven by more than one hidden factor. If all factors are modeled as OU factors, diagnostic checking allows us to examine which factors feature level-dependent volatility-of-volatility and which do not.

Last and important, as the unconditional and conditional distributions of the volatility-factor shocks $\{\mathbf{u}_t\}$ coincide in the OU case, the associated *unconditional* linear state space model is excellently suited for estimation. Apart from the various approximations that we carried out to arrive at linear measurement

[1] Similarly, in the interest-rate literature the Gaussian OU assumption has often been used as a model for the short interest rate, see e.g. Vasicek (1977), Langetieg (1980), the Ho and Lee (1986) model and de Jong (2000). Stein and Stein (1991) and Scott (1987) assume a 1-factor OU process for stock volatility, not the least because of its analytical tractability. More recently, Bakshi and Kapadia (2003) also assume an OU process for the stock volatility in their proposition 2.

equations, Kalman filter QML generates *consistent* estimates in this case. So if any large estimation bias is found, this seems more attributable to the performed approximations, rather than the state space estimator as such, given its asymptotic properties. For SV processes featuring level-dependent volatility-of-volatility, the *conditional* state space model to be discussed in the next chapter is more suited for estimation. However, this model will be estimated by *Extended* Kalman filter QML, which is a priori known to generate *inconsistent* estimates. So, whenever we are to find a large bias in the estimates there, it is less clear if this is either caused by Extended Kalman filter QML, or the various approximations that we carried out to arrive at linear measurement equations, or both. The OU SV case thus serves as a convenient benchmark or reference point.

The remainder of this chapter proceeds as follows. Section 2 discusses and explores the FTSE100-index data. Special attention is devoted to the complex empirical dynamics of the at-the-money *volatility term structure* over time.

Section 3 performs a Monte Carlo study towards the 1-factor OU SV model. We simulate option data which characteristics largely match the observed, short-maturity FTSE100-index option data. To prevent stock variance paths from turning negative during simulations, the parameters are chosen in such a way that chances are virtually negligible that this will occur. We study the quality of the various approximations that we carried out to arrive at linear measurement equations, the assumptions regarding the error terms of the state space model, and the performance of our estimation method. We compare this performance for several types of data used for estimation. Although RV data is not considered in the empirical analysis of the FTSE100-index market (we lack this data), we do consider 10-minute realized volatilities in the Monte Carlo study. It appears that using squared returns only for estimation performs worst. Although RV data performs much better, option data is even more informative. The combination of RV and option data performs best.

Section 4 assumes the FTSE100-index data to have been generated by the 1-factor OU SV option pricing model. We initially estimate the model using only squared returns, only short-maturity option data, and a combination of these. We interpret the estimation results, compare the SV with GARCH volatilities and the "observed" BS implied volatilities, and consider compensation for FTSE100-index volatility risk. We find a volatility risk premium of −15% per annum, which implies an expected short-maturity ATM FTSE100-index straddle return of −174% per annum. It appears that the 1-factor SV model overprices the longer-dated options out of sample. Additional evidence pointing towards insufficient capability of describing the observed volatility term structure movements next comes from estimating the 1-factor OU model using all data jointly.

Section 5 therefore extends to multiple OU SV factors. Although two SV factors yield a considerable improvement in fit, three factors are needed to obtain an adequate description of the volatility dynamics observed in the joint data. A first interpretation of the factors concerns their role in the volatility evolution, i.e., long-term, medium-term and short-term volatility trends. A second interpretation concerns their role in the shape dynamics (i.e., level, slope and convexity) of the volatility term structure over time. Each factor impacts differently on the prices of options of different maturity. The long-term volatility-trend factor impacts on all options similarly; the short-term trend factor virtually influences short-maturity options only. We interpret the risk premia associated with each of the factors, by considering expected straddle returns. Diagnostic checks reveal that the 3-factor OU SV model can be improved upon in several ways, one of which concerns the modeling of level-dependent volatility-of-volatility.

Section 6 summarizes the main results of this chapter on OU stochastic volatility. An appendix concludes.

2. FTSE100-index data: An explorative analysis

This section introduces the FTSE100-index data that provides the basis for the empirical analysis in this chapter and the next. Section 2.1 details the data collection and construction. An explorative analysis of the data motivates why we will not use all data for estimation. Section 2.2 examines the shape and complex dynamics of the at-the-money *volatility term structure* observed in the sample.

2.1 Data collection and construction

We examine daily data on the most important stock index of the United Kingdom, the FTSE100 index, and the European call option contract traded on that index. The FTSE100 index tracks changes in the value of a hypothetical portfolio of 100 major U.K. shares, listed on the *London Stock Exchange*. [2] The original dataset covers 9 years of data for the period 4 Jan 1993 till 28 Dec 2001.

Stock-index and index-option data
The source of the index data is *DataStream*, which records the daily closing prices of the index. Trading at the London Stock Exchange ends at 4.30 pm. We use the closing prices to compute daily index returns. The left panel of figure 2.1 plots the FTSE100 index and its daily returns.

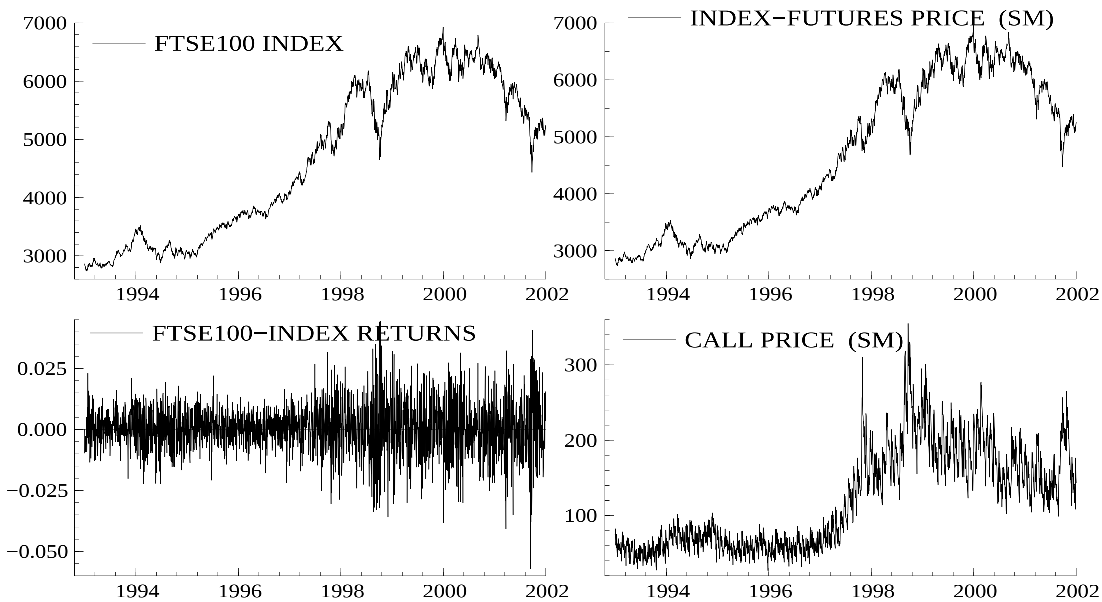


Figure 2.1: Left: FTSE100 index and its daily returns. Right: Index-futures prices and call prices associated with the SM option series.

[2] Stock dividends are not included in the computation of the index, and hence are not reinvested. The index merely tracks the capital gains and losses of the hypothetical portfolio. In terms of our model, $\{S_t\}$ thus represents the FTSE100 index (and not $\{S_t^r\}$, which is the reinvestment portfolio).

Volatility clustering is clearly present. Table 2.1 reports summary statistics for the daily returns for the period Oct 1997 – Dec 2001. (As motivated below, this is the sample we will eventually use for estimation.) The return distribution is slightly negatively skewed and has somewhat fatter tails than the normal distribution: Its skewness equals –0.14, its kurtosis 3.67.

The source of the index option data is the *London International Financial Futures and Options Exchange* (LIFFE). [3] The dataset consists of daily closing prices on a wide range of different call and put options, for a total of 902,445 observations. Specifically, for each observation the dataset contains the date, call/put flag, strike price, option price, expiry month, open interest, volume, and the daily settlement price in the FTSE100-index *futures contract* that has the same maturity as the option contract. The daily settlement time of the futures contract is at 4.30pm; option trading ends at 4.30pm as well. As we select *at-the-money-forward* (ATMF [4]) options only, which are typically most liquid, we expect possible non-synchronicity between daily settlement times to be least a problem. For simplicity, we assume *non-synchronicity biases* between index-futures prices, the FTSE100 index, and the index option prices to be negligible.

With regard to the available option contracts, on any given trading day European index options are traded with expiry months March, June, September, December, and additional months such that the 3 nearest calendar months are always available for trading. The options expire on the *third Friday* of the expiration month. On each day, there are 6 different-maturity contracts traded. This implies that the *maximum maturity* of all contracts ranges in between 9 and 12 months on any given day. [5] For each maturity, a wide range of options trades that differ in strike prices.

Selecting a SM, MM and LM option series

We select three close-to ATMF call option series from the database: a *short-maturity* (SM), a *medium-maturity* (MM), and a *long-maturity* (LM) series. As ATM options are typically most actively traded, and have maximal vega, we expect the prices of these options to contain the most valuable information. The maturity of the SM series ranges from 20 to 44 trading days, and averages at 1.4 months. For the MM series, the maturity range is 4.5 - 7 months (average 6 months). For the LM series this is 9 – 12 months (average 10.5 months). The average *moneyness* (F/K) of each of these series is 1.000, and has a small standard deviation. [6] Table 2.1 reports summary statistics on the maturity and moneyness for the SM, MM and LM option series.

---

[3] We thank Joost Driessen from the Finance Department of the University of Amsterdam for the option data. See http://www.liffe.com for a complete description of the data.

[4] "Forward-"moneyness is defined here as F/K, with F the futures price in the futures contract that has the same expiry date as the option, which has strike K. ATMF means that $F = K$; i.e. $F/K = 1$.

[5] To understand this, consider e.g. a day in March just before expiration. Contracts that mature in March, April, May, June, September and December trade on that day, so that the maximum maturity of all traded contracts equals 9 months. Consider next the first day in March after expiration. The contracts that mature in April, May, June, September and December still trade. LIFFE now issues new contracts that mature in next year's March, so that the maximum maturity instantly changes from 9 to 12 months.

[6] We extracted the option series from the database as follows. Consider e.g. the SM call series. Having separated the calls from the puts, for each day, we selected the calls with a maturity of at least (but at the same time closest to) 20 trading days. From these calls, on each day, we next selected that call which is closest ATMF by minimizing $|\ln(F_{ti}/K_i)|$, where K_i is the strike price of call i and F_{ti} is the current futures price of the index-futures contract that has the same maturity as call i. (Following Hull (2003), we assume forward and futures prices to be equal.) We minimize $|\ln(F_{it}/K_i)|$ rather than $|F_{it}/K_i - 1|$ because ("forward"-)moneyness may be defined as either F_{it}/K_i or K_i/F_{it}. Now

Figure 2.1 plots the call prices of the SM series; figure 2.2 shows its maturity and moneyness over time. Although ATM, the average call price is substantially higher from Oct 1997 onwards. The reason is the increased FTSE100-index volatility from Oct 1997 onwards, which is apparent from the graph of the index returns. Figure 2.1 also shows the futures price associated with SM option series. Not surprisingly, it closely tracks the FTSE100 index.

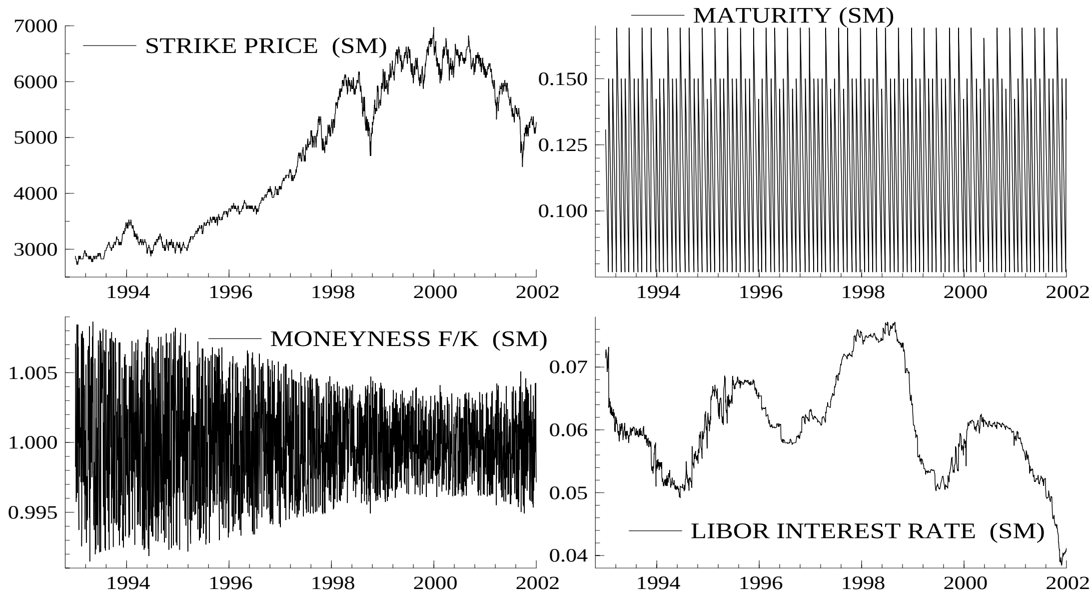


Figure 2.2: Series associated with the SM option series: strike price, maturity (in years), moneyness (F / K), and LIBOR risk-free interest rate (per annum).

Computing the SM, MM and LM BS implied volatility series

To compute the BS implied volatilities associated with each of the SM, MM and LM option series, we use the BS formula expressed in terms of the forward price. Daily estimates of the average dividend yield $\bar{q}_t$ are therefore not needed. [7] Specifically, we solve for $\sigma^2_{implied,it}$ from $C_{it} = BS(F_{t,T_{it}}, K_{it}, \tau_{it}, \bar{r}_{it}, \sigma^2_{implied,it})$ numerically, for $i = SM, MM, LM$ and $t = \Delta t,..,T\Delta t$. This requires for each option i on each day t an estimate of the average risk-free interest rate $\bar{r}_{it}$ over the remaining option's life. We estimate $\bar{r}_{it}$ by performing linear interpolation between the two nearest (in terms of maturity) continuously compounded *London Interbank Offer Rates* (LIBOR) on that day, which are typically considered risk free. The daily LIBOR rates for terms of 1 month, 2 months up to 12 months are taken from *DataStream*. For our sample period, these fluctuate between 3.8% and 7.7%. Figure 2.2 shows the interest rate series associated with the SM option series. It averages at 6.1%. (The MM and LM series also average at 6.1%.)

---

searching for an option for which $| F_{it} / K_i - 1 |$ is closest to 0, may yield another option than searching for an option for which $| K_i / F_{it} - 1 |$ is closest to 0. Minimizing $| \ln(F_{ti} / K_i) |$ circumvents this, which is equal to minimizing $| \ln(K_i / F_{it}) |$. This eventually left us with the SM-ATMF call option series, which counts 2258 daily observations, and for which the maturity range turned out to be 20 – 44 trading days. We extracted the MM-ATMF and LM-ATMF option series in similar ways; both contain 2258 observations as well. We carefully ensured that the characteristic series (i.e. price, strike, maturity, LIBOR interest rate, FTSE100 index, futures price etc.) associated with the 3 option series contain as many observations, and that all observation dates are fully synchronous. (Data on 20 trading days are missing in the original database. Furthermore, the data for 1-10-1997 is incomplete and the data for 28-5-1998 contains obvious errors, such that we have discarded these dates.)

[7] Using the data and forward price formula $F_{t,T} = S_t \exp[(\bar{r}_t - \bar{q}_t)(T - t)]$, we estimate the average continuously compounded FTSE100-index dividend yield at 3.53% per annum (for Jan 93 – Dec 01).

Squared returns $\frac{1}{\Delta t}(r_{t+\Delta t} - \hat{\mu}_t \Delta t)^2$ used in state space estimations

According to our model, the FTSE100 index evolves according to the SDE $dS_t = \mu_t S_t dt + \sigma_t S_t dW_{S,t}$. We did not explicitly specify the drift process $\mu_t = \mu(t, S_t, \mathbf{x}_t)$ in our theoretical discussion. Our estimation method nonetheless requires a prior estimate of μ_t at each point in time; recall the squared return equation $\frac{1}{\Delta t}(r_{t+\Delta t} - \hat{\mu}_t \Delta t)^2 = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t + \omega_{t+\Delta t}$. How can we obtain such an estimate? From an Euler discretization, it holds by approximation that $r_{t+\Delta t} \mid \mathcal{F}_t \sim \mathcal{N}(\mu_t \Delta t, \sigma_t^2 \Delta t)$. $\mu_t \Delta t$ thus basically represents the conditional mean daily index return, for which we require an estimate. Estimating an AR(2) model for the index returns results in residuals that are not significantly autocorrelated. [8] We therefore propose to estimate the conditional mean $\mu_t \Delta t$ by the fitted returns from this AR(2) regression. [9] The (uncorrelated) residuals of this regression are known as the *prewhitened* returns. [10] The squared prewhitened returns thus form the input to the squared return equation.

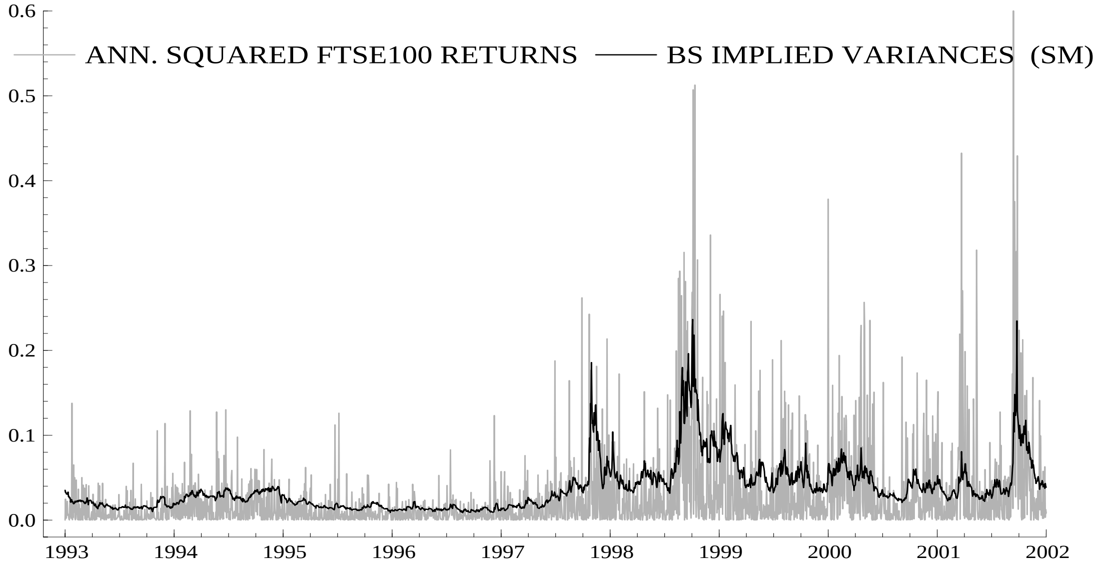


Figure 2.3: Annualized squared daily (prewhitened) FTSE100-index returns, and the Black-Scholes implied variance series associated with the SM option series.

Analysis of the squared returns and BS implied volatilities

Figure 2.3 shows the squared returns $\frac{1}{\Delta t}(r_{t+\Delta t} - \hat{\mu}_t \Delta t)^2$ and the SM BS implied variances $\sigma^2_{implied,t}$ in one graph. A clear increase in the average volatility level and the volatility-of-volatility is visible after the third quarter of 1997. The impact on global financial markets of the *Asian crisis* is apparent (fall 1997). The near-

[8] Our model implies that non-overlapping returns are (virtually) uncorrelated (see section 12 of appendix B). The fact that we find the empirical index returns to be significantly autocorrelated is generally believed to be caused by *non-synchronous trading effects* associated with the individual stocks that together compose the FTSE100 index; see e.g. Campbell et al. (1997).

[9] Doing so, our estimation procedure thus essentially boils down to two-step estimation and volatility extraction of the parameters of the model $r_{t+\Delta t} \mid \mathcal{F}_t \sim \mathcal{N}(\phi_0 + \phi_1 r_t + \phi_2 r_{t-\Delta t}, \sigma_t^2 \Delta t)$ in which σ_t is the annualized stochastic volatility. In the first step we estimate the conditional mean parameters ϕ_0, ϕ_1, ϕ_2; in the second step we estimate the parameters governing the stochastic volatility and moreover extract the latent volatility series. Notice the clear analogy with an AR(2)-GARCH(1,1) model for daily stock returns. But recall the important difference: In our model we have true stochastic volatility, whereas in the GARCH model today's volatility is a deterministic function of yesterday's information set. See also section 3.2.2 of chapter III.

[10] E.g. Andersen et al. (2002) also prefilter the data prior to estimation to get rid of the autocorrelation.

collapse of hedge fund *Long-Term Capital Management*, the influence of the *Soviet Union crisis* and the uncertainty about the *European Monetary Union* (i.e., the introduction of the euro) is clearly visible too (fall 1998). The aftermath of *September 11*, 2001 and the *war on terrorism* are also obvious. [11] These events have led to increased fluctuations in financial markets in recent years.

Plots of the SM, MM, and LM BS implied volatility series (figure 2.4) confirm that the volatility seems to evolve in two very different "regimes" during 1993 – 2002. To a priori avoid model misspecification as much as possible, we opt to use data for the period 6 Oct 1997 – 28 Dec 2001 only in all our estimations, for a total of 1058 observations. For the period till Oct 1997 for example, an estimate of the unconditional stock volatility (computed as the average of the annualized squared returns) equals 11.4%, after Oct 1997 it equals 20.3%. If volatility is driven by a 1-factor OU process (for which the unconditional volatility equals $\sqrt{\theta}$), this would imply that θ has more than tripled after Oct 1997. Using all data would lead to obvious biases in this case.

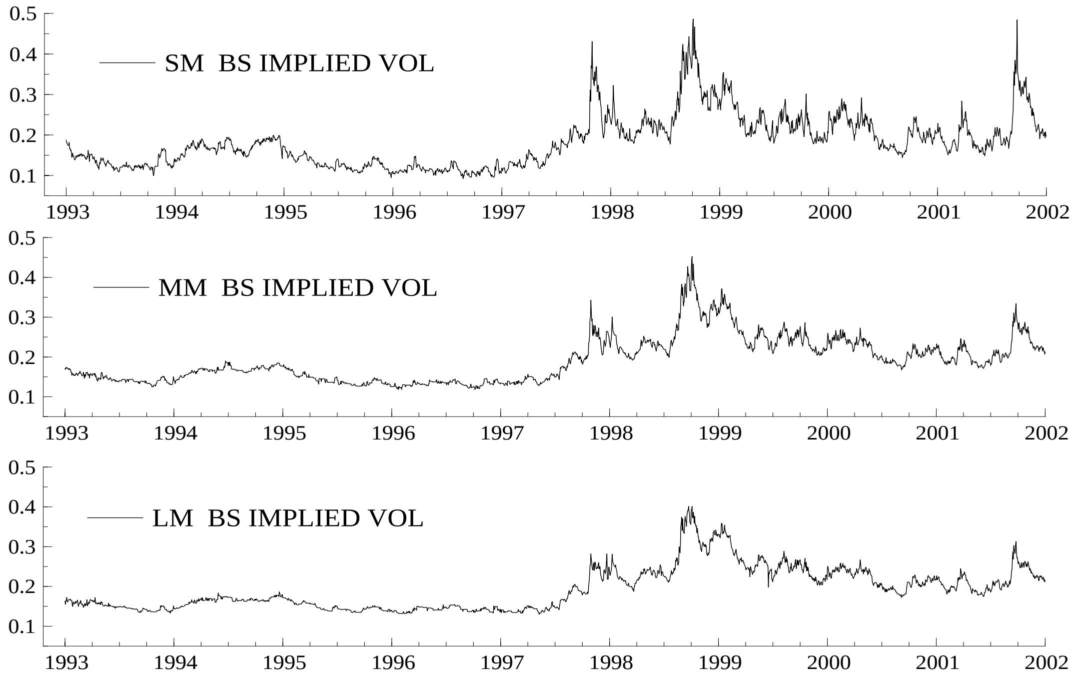


Figure 2.4: The Black-Scholes implied volatility series associated with the short-maturity, medium-maturity and long-maturity call option series.

Negative market price of volatility risk?
Summary statistics for Oct 1997–Dec 2001 are in table 2.1. The average BS implied volatility of 23.70% is substantially larger than the just-mentioned 20.3% unconditional stock volatility. Moreover, the SM BS implied volatility series fluctuates most, the LM series least. As our SV model implies that $\sigma^2_{implied,t} \approx \mathbb{E}_{\mathbb{Q}}[\bar{\sigma}^2_t \mid \mathcal{F}_t]$ (see (3.70) in the previous chapter), the first data characteristic may be interpreted as a preliminary indication that measures $\mathbb{P}$ and $\mathbb{Q}$ diverge, with the market price of volatility risk being *negative*. The second characteristic is understood if there is *mean reversion* in the volatility.

[11] The peak of 0.86 in the annualized squared FTSE100 return at September 11, 2001 has been cut off at 0.6 to enhance the visual quality of the graph.

Table 2.1: Summary statistics FTSE100-index data for the period Oct 1997 – Dec 2001

	Returns	*Sq.returns*	*SM series:*	*MM series:*	*LM series:*
			BS impl.vol.	*BS impl.vol.*	*BS impl.vol.*
Mean	0.0000	0.0413	23.25%	23.93%	23.89%
Median	0.0003	0.0151	21.84%	22.96%	23.22%
Std.deviation	0.0126	0.0676	5.73%	4.68%	4.35%
Skewness	-0.14	4.23	1.33	1.36	1.28
Kurtosis	3.67	37.73	4.94	5.18	4.68
			Moneyness	*Moneyness*	*Moneyness*
Mean			1.000	1.000	1.000
Std.deviation			0.002	0.005	0.005
			Maturity	*Maturity*	*Maturity*
Mean			0.116	0.500	0.872
Std.deviation			0.025	0.074	0.072
Minimum			0.077	0.377	0.750
Maximum			0.169	0.638	1.015

The table reports summary statistics for the daily (prewhitened) FTSE100-index returns, the annualized squared (prewhitened) index returns $\frac{1}{\Delta t}(r_{t+\Delta t} - \hat{\mu}_t \Delta t)^2$, and the SM, MM, and LM index option series, for 6 Oct 1997 – 28 Dec 2001. Moneyness is defined here as the ratio of futures price and strike price (F / K). Maturity is reported in years.

2.2 Analyzing the volatility term structure

In real-world financial markets, BS implied volatilities of options written on the same underlying typically differ across strikes and maturity (see e.g. Hull (2003)). If the Black-Scholes model were true, the *implied volatility surface* should be flat at any moment in time: each moneyness-maturity combination should reveal the same BS implied volatility. Instead, for a given maturity, one typically observes a smile or skew-shaped pattern across implied volatilities, known as the *volatility smile* or *skew*. This pattern tends to be less pronounced the longer the maturity of the options. For given moneyness, one observes a variety of shapes of the so-called *volatility term structure,* which plots the implied volatilities against term-to-maturity of the options. The implied volatility surface serves an important function in practice: Traders use it as a sophisticated interpolation tool for pricing *vanilla* options consistently with the market. [12] Moreover, traders typically *quote* an *option price* in terms of its *BS implied volatility*.

We focus on the relation between BS implied volatilities and maturity. Our interest is in the *shape* and *dynamics* of the ATM volatility term structure (VTS) over time. Given the three ATM option series, the VTS of ATM options can have four possible shapes on each of the 1058 days in our sample. It may either be *upward sloping*, exhibit a *hump shape*, be *downward sloping*, or display an *inverted hump shape*. Table 2.2 reports the shape frequencies (in percents), and assigns each possible shape a number for later reference. The ATM VTS is upward sloping on almost 50% of the days; the inverted hump shape occurs seldom.

[12] That is, European call and put options. *Exotic* derivatives cannot be priced with the surface, unless the exotic can be decomposed into a combination of vanilla options. European call and put options written on the same underlying with the same strike and maturity, have identical implied volatility (this follows from the put-call parity). As such the surface can be used to price both call and put options.

Table 2.2: Shape frequencies volatility term structure ATMF options (Oct 1997 – Dec 2001)

Shape of vol. term structure	*1: upward sloping*	*2: hump shape*	*3: downward sloping*	*4: inverted hump shape*
Frequency	48.7%	27.5%	18.3%	5.5%

VTS changes shape often

As option prices change through time, so does the shape of the implied volatility surface. To explore the daily dynamics in the ATM volatility term structure, figure 2.5 provides two matrices. The first matrix reports the number of VTS *shape transitions*, as a percentage of the total number of transitions, 1057. For example, of all transitions, 9.4% entailed a shape change from upward sloping to hump shape from one day to the next, and in 14.4% of the cases the shape remained downward sloping. The second matrix reports the empirical shape transition "probabilities", with each row adding to 100%. Given that the VTS has a particular shape today, it is most likely that it has equal shape tomorrow.

Shape transition frequencies (in %)

t \ t+1	***1***	***2***	***3***	***4***
1	37.2	9.4	0.9	1.2
2	9.9	14.9	1.7	1.0
3	0.4	2.1	14.4	1.5
4	1.2	1.2	1.3	1.7

Shape transition "probabilities" (in %)

t \ t+1	***1***	***2***	***3***	***4***
1	76.4	19.3	1.8	2.5
2	36.0	54.0	6.2	3.8
3	2.1	11.3	78.4	8.2
4	22.4	22.4	24.1	31.1

Figure 2.5: Day-to-day dynamics in the shape of the ATMF volatility term structure (Oct 97 – Dec 01). Left: transition frequencies in % of total number of transitions. Right: empirical transition "probability" matrix (in %). Shape codes are reported in table 2.2.

Shifts in the VTS: 34% is non-parallel

We define a *parallel shift* in the VTS as one in which the SM, MM and LM BS implied volatilities either jointly go up from one day to the next, or go down, but not necessarily by the same amount. Table 2.3 shows that parallel shifts occur 66% of the time; 34% of the shifts are non-parallel.

Table 2.3: Daily shifts in the ATMF volatility term structure (Oct 97 – Dec 01)

	Upward	*Downward*	*Total*
Parallel shifts	32%	34%	66%
Non-parallel shifts			34%
		Total:	100%

Dynamics of level, slope and curvature of the VTS

Additional evidence showing that the empirical VTS evolves in complex ways is in figure 2.6, which plots the evolution of the *level*, *slope* and *curvature* (or *convexity*) of the VTS over time. We define its general level as the average BS implied volatility: $level_t \equiv (\sigma_{implied,SM,t} + \sigma_{implied,MM,t} + \sigma_{implied,LM,t})/3$. We define its slope as the difference between the LM and SM implied volatility, divided by the maturity difference: $slope_t \equiv (\sigma_{implied,LM,t} - \sigma_{implied,SM,t})/(\tau_{LM,t} - \tau_{SM,t})$. We define its curvature as $curvature_t \equiv a_t\,\sigma_{implied,LM,t} + b_t\,\sigma_{implied,MM,t} + c_t\,\sigma_{implied,SM,t}$, where $a_t = 2/[h_{1t}(h_{1t} + h_{2t})]$, $c_t = 2/[h_{2t}(h_{1t} + h_{2t})]$, $b_t = -a_t - c_t$, $h_{1t} = \tau_{LM,t} - \tau_{MM,t}$ and $h_{2t} = \tau_{MM,t} - \tau_{SM,t}$. As explained in the appendix, this convexity measure essentially represents a numerical approximation of the second-order derivative of a smooth function in a certain point, computed using three coordinate pairs.

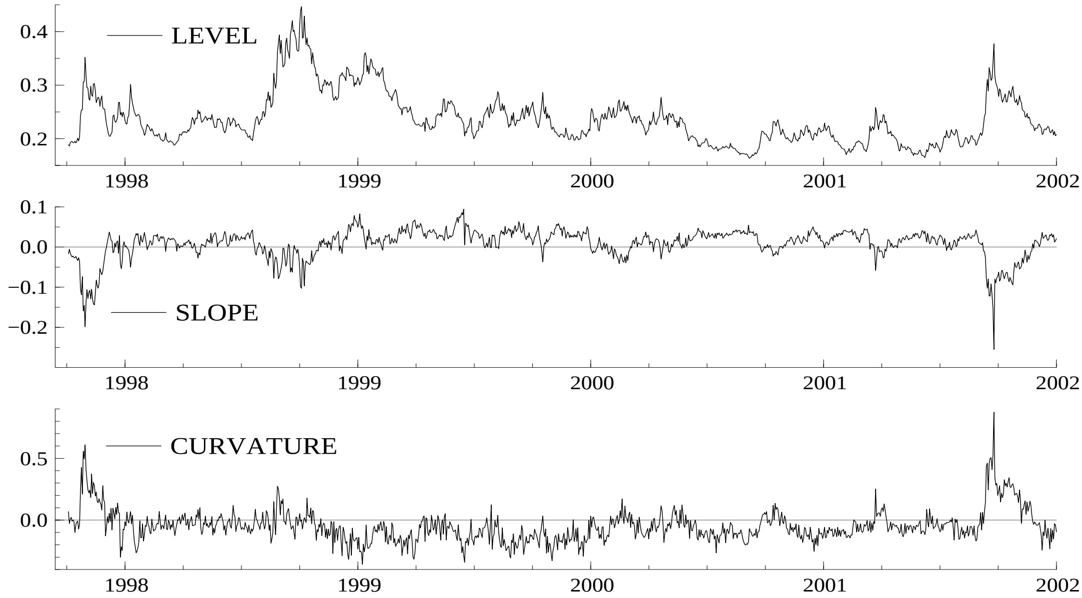


Figure 2.6: Evolution of level, slope and curvature of the volatility term structure.

Figure 2.6 shows that when the general volatility level is high (low), the slope is typically negative (positive). This suggests volatility mean-reversion, and illustrates why traders attach forecasting power to BS implied volatilities. The slope numbers reveal that the VTS can be steep: a slope of e.g. −0.10 implies a BS implied volatility difference of roughly −7.5% between the LM and SM series.

Is one SV factor enough?
It is clear that the empirical VTS evolves over time in complicated manners. A realistic model must be capable of reproducing these complex dynamics. Evidently, the question is if one SV factor is sufficient in this respect. Recalling that our model implies that

$$\sigma^2_{implied,t} \approx \frac{A_1(\tau_t)}{\tau_t} + \frac{\mathbf{B}_1(\tau_t)'}{\tau_t}\mathbf{x}_t\ ,\ d\mathbf{x}_t = \mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t)dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t}\ , \qquad (\mathbb{P}) \qquad (2.1)$$

reveals that, by approximation, the VTS is driven by an affine function of the underlying hidden volatility-driving factors $\mathbf{x}$. Equation (2.1) is indicative of why 1-factor SV option pricing models are found to be misspecified in practice. [13] If (2.1) held with equality, in the 1-factor SV case, all BS implied variances of a given moneyness would be perfectly correlated. [14] That is, 1-factor SV could only describe *parallel* shifts in the VTS in that case: A volatility shock would either simultaneously raise all implied volatilities, or lower them. We find 34% of all

[13] Meddahi (2002), Chernov et al. (2003) and Tauchen (2004) argue that one of the reasons for the poor fit of 1-factor SV diffusion models is that this one factor cannot simultaneously fit both the fat tails of the return distribution and the volatility persistence. Introducing a second volatility factor (or adding jumps) breaks this link. Eraker et al. (2003) argue that as the extracted volatility from 1-factor SV models (without jumps) is found to be so persistent, it cannot change rapidly enough during periods of market stress. A second volatility factor allows for faster changing volatility.
[14] Notice the analogy with affine models of the term structure of interest rates (Duffie and Kan (1996)). In those models, yields of arbitrary maturity are an exact affine function of the factors. This implies that 1-factor models can only generate parallel shifts in the *yield curve*, as all yields are perfectly correlated in that case. This is not realistic in practice. Moreover, it is by now commonly known that 1-factor models cannot describe –in a satisfactory way- the rich number of yield-curve shapes typically observed.

shifts in the empirical FTSE100 VTS to be non-parallel. Although (2.1) does not hold with equality, it seems unlikely that 1-factor SV is able to describe the rich dynamics observed in the joint FTSE100-index data sufficiently well.

3. A Monte Carlo study to the 1-factor OU SV model

This section performs a Monte Carlo study towards the 1-factor OU SV option pricing model. Section 3.1 states its main equations, and considers some implied properties of the volatility process. Section 3.2 outlines the simulation strategy, and discusses the assumptions underlying the simulated squared return, RV and option data. These assumptions are largely based on the information in the FTSE100-index data. Section 3.3 discusses the state space model associated with 1-factor OU SV, and considers parameter identification. Section 3.4 discusses the simulation results.

3.1 The 1-factor OU SV option pricing model

If there is one OU factor driving the volatility, the main ingredients of the multifactor model read (see section 2 and 3.2 of chapter I for more details): [15]

Stock price: $$dS_t = \mu_t S_t dt + \sigma_t S_t dW_{S,t} \qquad (\mathbb{P}) \qquad (3.1)$$

Stock variance: $$\sigma_t^2 = x_t \qquad (3.2)$$

Latent factor: $$dx_t = k(\theta - x_t)dt + \sigma dW_{x,t} \qquad (\mathbb{P}) \qquad (3.3)$$

$$dx_t = \tilde{k}(\tilde{\theta} - x_t)dt + \sigma d\tilde{W}_{x,t} \qquad (\mathbb{Q}) \qquad (3.4)$$

Risk-neutral parameters: $$\tilde{k} = k, \quad \tilde{\theta} = \theta - \sigma\gamma / k \qquad (3.5)$$

Market price of vol. risk: $$\gamma_{x,t} = \gamma \qquad (3.6)$$

Volatility risk premium: $$\sigma\gamma \qquad (3.7)$$

Call price (strike K_t, mat. τ_t) $$C_t = \mathbb{E}_{\mathbb{Q}}\left[BS(F_{t,T}, K_t, \tau_t, \bar{r}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t\right] \qquad (3.8)$$

$$BS(F_{t,T}, K_t, \tau_t, \bar{r}_t, \bar{\sigma}_t^2) = \exp(-\bar{r}_t\tau_t)\left[F_{t,T}\Phi(d_{1t}) - K_t\,\Phi(d_{2t})\right]$$

$$d_{1t} = \frac{\ln\frac{F_{t,T}}{K_t} + \frac{1}{2}\bar{\sigma}_t^2\tau_t}{\bar{\sigma}_t\sqrt{\tau_t}}; \quad d_{2t} = d_{1t} - \bar{\sigma}_t\sqrt{\tau_t}$$

Forward price (maturity τ_t): $$F_{t,T} = S_t \exp\left[(\bar{r}_t - \bar{q}_t)\tau_t\right]. \qquad (3.9)$$

The model is fully determined by four parameters, θ, k, σ and γ with $\theta, k, \sigma > 0$ and $\gamma \in \mathbb{R}$. Recall that $\{\sigma_t^2\}$ represents the per-annum *conditional* variance process of the instantaneous stock return, dS_t / S_t. The unconditional stock return variance equals $\mathbb{E}_{\mathbb{P}}[\sigma_t^2] = \mathbb{E}_{\mathbb{P}}[x_t] = \theta$ per annum, such that the *unconditional stock volatility* equals $\sqrt{\theta}$. From the properties of the OU process (see appendix B, section 13), the *invariant distribution* of the stock variance equals $\sigma_t^2 \sim \mathcal{N}(\theta, \sigma^2 / 2k)$. The *volatility-of-the-variance* thus equals $\sqrt{(\sigma^2 / 2k)}$. The mean-reversion parameters k and $\tilde{k}$ coincide under $\mathbb{P}$ and $\mathbb{Q}$. The *volatility persistence* is measured by $\exp[-k\Delta t]$ with $\Delta t = 1/260$. The *half-life* (measured

[15] This special case is obtained from imposing $n = 1$, $\alpha = 1$ and $\beta = 0$ in the general multifactor model. For identification reasons, we moreover a priori impose $\delta_0 = 0$, $\delta = 1$ such that $\sigma_t^2 = x_t$.

in days) of a volatility shock is given by $\ln 2/(k\Delta t)$. The *market price of volatility risk* equals γ, whereas the *volatility risk premium* is given by $\sigma\gamma$. Both are constant over time in the OU case.

3.2 Simulating from the 1-factor OU SV option pricing model

3.2.1 Assumptions

As usual, we assume 260 trading days per annum and denote the timing of the daily data points by $t = \Delta t, 2\Delta t, .., T\Delta t$, with $\Delta t = 1/260$. We simulate for a total of $T = 1058$ daily observations, corresponding to the UK data used in our empirical work. We assume $S_0 = 5300$, $\mu_t = \mu = 8.25\%$, $q_t = q = 3.5\%$ and $r_t = r = 6\%$. (These values correspond to the information in the FTSE100-index data for the period 6 Oct 1997 – 28 Dec 2001.)

We simulate time series of squared returns, realized volatilities and short-maturity (close-to) ATMF call options. The maturity and moneyness of the option series exactly match those of the empirical short-maturity ATMF FTSE100-index option series, for each t. The average option maturity is 1.4 months. (See section 2.1 for details.) This permits a consistent comparison between the real-world and simulated call data.

What can we do to prevent stock variance paths from turning negative during simulations? The parameters with which to simulate data with, should be chosen in such a way that chances are negligible that this will occur: θ and k must be chosen large enough, σ small enough, and γ enough negative. [16] At the same time however, we want their values to be as "realistic" as possible. This tradeoff results in the following choice:

$\theta = 0.16$	(unconditional stock volatility of 40%)
$k = 5.25$	(persistence in daily stock variance of 0.98)
$\sigma = 0.10$	(volatility-of-volatility parameter)
$\gamma = -1$	(market price of volatility risk).

The implied risk-neutral parameters equal $\tilde{k} = 5.25$ and $\tilde{\theta} = 0.179$, and the volatility risk premium –10%. (See table 4.1 for further motivation of this choice.)

To keep the amount of computations to a reasonable level, we simulate a total of 100 datasets for this set of parameter values.

3.2.2 Set up of the Monte Carlo experiment

Below we outline (in chronological order) the different steps in the Monte Carlo experiment, to obtain one simulated dataset from the 1-factor OU SV option pricing model. Repeating these steps a 100 times yields the 100 datasets.

Step 1: Simulate a stock price and volatility path under $\mathbb{P}$; compute squared returns and realized volatilities.

The first step consists of simulating a sample path of $\{S_t\}$ under $\mathbb{P}$ over the interval $[0, T\Delta t]$, given $S_0 = 5300$; see (3.1). As the stock price is governed by independently evolving stochastic volatility, a simulated $\mathbb{P}$-path of $\{\sigma_t^2 = x_t\}$ is

[16] An alternative "solution" seems perhaps a *reflecting barrier* at $\sigma_t^2 = x_t = 0$. However, if this barrier is hit too often, sample paths result that can no longer be considered as (approximate) OU paths anymore. Distorted simulated data results, which does not provide a reliable basis for investigating the performance of our estimation method.

needed first. We simulate both $\{S_t\}$ and $\{x_t\}$ on $I = 48$ *intraday* time points. As a trading day lasts for approximately 8 hours in practice, this means that we sample a value every *10 minutes*. (For a discussion on the intraday sampling frequency for RV, see section 3.3.2 of the previous chapter.) The discretization step (δt) of the continuous-time processes then equals $\delta t = \Delta t / I = 8.0 * 10^{-5}$.

Rather than sampling x_0 from its invariant distribution (which does not preclude negative values), we assume the initial value of $\{x_t\}$ to equal its invariant mean, $x_0 = \theta$. From equations (6.1), (6.2) and (13.29) in appendix B, for the 1-factor OU SV process it holds under $\mathbb{P}$ that

$$x_{t+\delta t} = \theta + \exp[-k\delta t](x_t - \theta) + u_{t,t+\delta t}\ ,\ u_{t,t+\delta t} \sim \mathcal{N}\left[0, \sigma^2\left(\frac{1-\exp[-2k\delta t]}{2k}\right)\right] \quad (3.10)$$

This result leads to the following scheme for obtaining the desired sample path:

$$\begin{aligned} x_{[0]} &= \theta \qquad (\mathbb{P}) \qquad (3.11) \\ x_{[(i+1)\delta t]} &= \theta + \exp[-k\delta t](x_{[i\delta t]} - \theta) + \sqrt{\sigma^2\left(\frac{1-\exp[-2k\delta t]}{2k}\right)}\epsilon_i\ , \end{aligned}$$

in which $\epsilon_i \sim$ *i.i.d.* $\mathcal{N}(0,1)$; $i = 0,..,TI-1$. The end-of-day factor values are given by $x_{[0]}, x_{[\Delta t]} = x_{[I\delta t]},.., x_{[T\Delta t]} = x_{[TI\delta t]}$. In contrast to a conventional Euler approximation scheme of the factor SDE (3.3), this discrete-time process has the advantage that at each (discrete) point in time, it has exactly the same distribution as the continuous-time process it approximates. As such, (3.11) is preferable to a naïve Euler approximation scheme of (3.3), for which this only holds by approximation (as long as $\delta t \neq 0$). Although sample paths of (3.11) may turn negative, this did not happen for the 100 datasets that we simulated.

A sample path of $\{S_t\}$ under $\mathbb{P}$ is next obtained as follows. First, recall that due to SV, an exact solution of the stock price SDE does not exist. We therefore need to resort to discrete-time processes that approximate the distribution of $\{S_t\}$ at each point in time; e.g. an Euler approximation. Considering the SDE of the log stock price, $d\ln S_t = (\mu - \frac{1}{2}\sigma_t^2)dt + \sigma_t dW_{S,t}$, adding the increments over $[t, t+\delta t]$, and then transforming back to S_t yields

$$S_{t+\delta t} = S_t \exp\left[\mu\,\delta t - \tfrac{1}{2}\int_t^{t+\delta t} \sigma_u^2 du + \int_t^{t+\delta t} \sigma_u dW_{S,u}\right]. \qquad (\mathbb{P}) \qquad (3.12)$$

Approximating the Riemann and Itô integrals in this expression by the product of their respective integrands observed at the begin point of the interval, and the increment in the integrator over $[t, t+\delta t]$, yields the following approximation scheme of $\{S_t\}$ on the interval $[0; T\Delta t]$: [17]

$$\begin{aligned} S_{[0]} &= S_0 \qquad (\mathbb{P}) \qquad (3.13) \\ S_{[(i+1)\delta t]} &= S_{[i\delta t]} \exp\left[(\mu - \tfrac{1}{2}x_{[i\delta t]})\delta t + \sqrt{x_{[i\delta t]}}\sqrt{\delta t}\,\eta_i\right], \end{aligned}$$

[17] Notice that the scheme (3.13) means an improvement to a conventional Euler scheme of the stock price SDE obtained from $S_{t+\delta t} - S_t = \int_t^{t+\delta t} \mu S_u du + \int_t^{t+\delta t} \sigma_u S_u dW_{S,u}$: As the integrand of the Itô integral in (3.12) will tend to vary less over $[t, t+\delta t]$ than the one in the latter equation, the approximation in (3.13) will generally be better; especially when S_t is large.

in which $\eta_i \sim i.i.d.\mathcal{N}(0,1)$, $i = 0,..,TI-1$. From the end-of-day stock prices $S_{[0]}$, $S_{[\Delta t]} = S_{[I\delta t]},.., S_{[T\Delta t]} = S_{[TI\delta t]}$, the simulated annualized squared returns in deviation from their mean, $(r_{[j\Delta t]} - \mu\Delta t)^2 / \Delta t;\ j = 1,..,T$, can next be computed. Here, the daily stock return is given by $r_{[j\Delta t]} = (S_{[j\Delta t]} - S_{[(j-1)\Delta t]}) / S_{[(j-1)\Delta t]}$. Given the simulated intraday path of $\{S_t\}$, the associated daily 10-minute realized variance over day $[t, t+\Delta t]$ is calculated as

$$\sigma^2_{RV,[t+\Delta t]} = \frac{1}{\Delta t}\sum_{i=1}^{I}\left(\ln\frac{S_{[t+i\delta t]}}{S_{[t+(i-1)\delta t]}}\right)^2, \qquad (\mathbb{P}) \qquad (3.14)$$

and so for all $t = 0, \Delta t,.., (T-1)\Delta t$.

Step 2: Simulate forward prices and option strike prices
The forward prices for delivery on the option maturity dates are first computed by $F_{[t,T]} = S_{[t]}\exp[(r-q)\tau_t];\ t = \Delta t,.., T\Delta t$. Here, $\{\tau_t\}$ is the maturity series of the empirical SM FTSE100-index option series. We next simulate the option strike prices $\{K_t\}$ such that the moneyness of the simulated and observed option series exactly coincide for all t. [18] This results in a simulated SM close-to ATMF call option series, which maturity and moneyness exactly match those of the empirical SM FTSE100-index option series.

Step 3: Simulate call prices
Given r, q, the simulated $x_{[t]}, S_{[t]}, F_{[t,T]}, K_t$, and τ_t, we next simulate the theoretical call price C_t for each $t = \Delta t, 2\Delta t,.., T\Delta t$. Given time-$t$ information, the only random quantity in the pricing formula $C_t = \mathbb{E}_{\mathbb{Q}}[BS(F_{t,T}, K_t, \tau_t, r, \bar{\sigma}_t^2) \mid \mathcal{F}_t]$ is the average variance $\bar{\sigma}_t^2 = \frac{1}{\tau_t}\int_t^{T_t}\sigma_u^2 du$ over the remaining option's life $\tau_t = T_t - t$. Hence, to compute C_t, we essentially only need to simulate a large number of $\bar{\sigma}_t^2$'s under measure $\mathbb{Q}$, given the initial stock variance $\sigma_t^2 = x_t$ at time t under measure $\mathbb{P}$ (!). From appendix B, section 13.1, the conditional distribution of $\bar{\sigma}_t^2$ under $\mathbb{Q}$ is Gaussian,

$$\bar{\sigma}_t^2 \mid \mathcal{F}_t \overset{\mathbb{Q}}{\sim} \mathcal{N}\left[\tilde{\theta} + \frac{1-\exp[-\tilde{k}\tau_t]}{\tilde{k}\tau_t}\left(x_t - \tilde{\theta}\right);\ \left(\frac{\sigma}{\tau_t\tilde{k}}\right)^2\left[\tau_t - 2\left(\frac{1-\exp[-\tilde{k}\tau_t]}{\tilde{k}}\right) + \frac{1-\exp[-2\tilde{k}\tau_t]}{2\tilde{k}}\right]\right] \qquad (3.15)$$

which makes simulating $\bar{\sigma}_t^2$'s fairly easy. (3.15) requires as input for x_t its simulated value under $\mathbb{P}$ obtained in step 1.

As its distribution is Gaussian, it is possible to sample a negative $\bar{\sigma}_t^2$. This will mainly occur if x_t is close to zero under $\mathbb{P}$; i.e., if the current stock volatility is close to zero. This is obviously something that we want to circumvent as much as possible. Our parameter choice for k, θ, σ, γ aids in this respect, but does not fully preclude negative sampled $\bar{\sigma}_t^2$'s. To deal with this, we choose to cut off both the negative tail of the distribution of $\bar{\sigma}_t^2 \mid \mathcal{F}_t$, and that part of the positive tail

[18] This is done as follows. Recall that, in order to extract a SM ATMF call series from the FTSE100-index database, we minimized the quantity $m_t = |\ln(F_{t,T} / K_t)|$. To obtain simulated strike prices from this m_t series (given the simulated forward prices), we consider two possible cases to correctly deal with the absolute value function $|..|$. If the *observed* $F_{t,T}$ is larger than or equal to the *observed* K_t, the *simulated* K_t is computed as $K_t = F_{[t,T]}\exp(-m_t)$, with $F_{[t,T]}$ the *simulated* forward price. Instead, the simulated K_t is computed as $K_t = F_{[t,T]}\exp(m_t)$. (The resulting simulated strike prices are not necessarily integer numbers.)

that has the same probability as the negative tail. [19] This procedure seems more reasonable than just discarding the negative tail, as both the mean and symmetry of the sampling distribution remain in tact in this way (although the variance changes of course). Regarding the 100 datasets that we simulated, it only happened for 1 dataset that 0.000045% of the average variances initially drawn, fell into the a priori excluded tails; for the 99 other datasets this never happened. This indicates that the simulated data is (virtually) not distorted by this specific procedure. (The 0.000045% is based on 2000 simulated $\bar{\sigma}_t^2$'s for each t.)

Having simulated N average variances at time t, which we denote by $\bar{\sigma}_{[t]}^2(i)$; $i = 1,..,N$, we proceed as follows. Each of these $\bar{\sigma}_{[t]}^2(i)$'s corresponds to a simulated Black-Scholes call price value, $BS(F_{[t,T]}, K_t, \tau_t, r, \bar{\sigma}_{[t]}^2(i))$. By the weak law of large numbers, the average of these values converges in probability to the true call price at time t: [20]

$$\frac{1}{N}\sum_{i=1}^{N} BS(F_{[t,T]}, K_t, \tau_t, r, \bar{\sigma}_{[t]}^2(i)) \xrightarrow{p} C_t . \tag{3.16}$$

Given this result, we approximate the call price C_t at day t by the left-hand side of (3.16), and denote this simulated value by $C_{[t]}$. Obviously, the approximation improves for larger N, but comes at the cost of increased computing time. We choose $N = 2000$ in our simulations. This is trade-off between computational burden and a sufficiently small Monte Carlo standard error of the simulated call price. (The average Monte Carlo standard error as a percentage of the call price equals 0.11% (std.dev. 0.02%) only over all 100*1058 simulated call prices.) Repeating this procedure for each $t = \Delta t,..,T\Delta t$, yields the desired simulated call price series $\{C_{[\Delta t]}, C_{[2\Delta t]},.., C_{[T\Delta t]}\}$.

Step 4: Obtain the Black-Scholes implied variance series
Given the call price series $\{C_{[t]}; t = \Delta t,..,T\Delta t\}$, we obtain the associated Black-Scholes implied variance series $\{\sigma_{implied,[t]}^2;\ t = \Delta t,..,T\Delta t\}$ from numerically solving for $\sigma_{implied,[t]}^2$, from the equation $C_{[t]} = BS(F_{[t,T]}, K_t, \tau_t, r, \sigma_{implied,[t]}^2)$. As the Black-Scholes function is monotonically increasing in its variance argument, this results in a unique $\sigma_{implied,[t]}^2$ for each t.

3.3 State space model

Depending on the specific data used for estimation, the state space model associated with the 1-factor OU SV option pricing model has several forms. For $t = \Delta t,...,T\Delta t$, each model consists of a selection of the following components: [21]

Return data:

$$\frac{1}{\Delta t}(r_t - \hat{\mu}_{t-\Delta t}\Delta t)^2 = \theta + x_{t-\Delta t}^* + \omega_t , \qquad \omega_t \sim (0, \sigma_\omega^2) \tag{3.17}$$

---

[19] More specific: Write $\bar{\sigma}_t^2 = \mu(x_t, \tau_t) + \sigma(\tau_t)\varepsilon_t$ where $\varepsilon_t \sim i.i.d.\mathcal{N}(0,1)$ and $\mu(x_t, \tau_t)$ and $\sigma^2(\tau_t)$ are as in (3.15). If $\varepsilon_t < -\mu(x_t, \tau_t)/\sigma(\tau_t)$ we draw a negative $\bar{\sigma}_t^2$, which we want to prevent. Therefore, we exclude this tail and do not allow for sampling ε_t's in that area. We also discard ε_t's for which $\varepsilon_t > \mu(x_t, \tau_t)/\sigma(\tau_t)$ (these would lead to average variances in the far positive tail).

[20] That is, to be precise, if we are willing to assume that $F_{[t,T]} = F_{t,T}$ and $\bar{\sigma}_{[t]}^2 = \bar{\sigma}_t^2$ for the moment.

[21] For the general multifactor model, these equations and the resulting system matrices of the unconditional state space model are given in section 3.5 of the previous chapter. Recall that we choose representations (3.85) and (3.87) for estimation. In the 1-factor OU case, the functions $A_1(.)$ and $B_1(.)$ have explicit expressions; see (13.38) in appendix B, although the expressions there need to be tailored to the risk-neutral measure $\mathbb{Q}$. The option equation (3.19) has been rewritten in terms of the $\mathbb{P}$-parameters k and θ by substituting $\tilde{k} = k$ and $\tilde{\theta} = \theta - \sigma\gamma / k$.

RV data:

$$\sigma^2_{RV,t} = \theta + \frac{1-\exp(-k\Delta t)}{k\Delta t} x^*_{t-\Delta t} + \omega_{RV,t}\,, \quad \omega_{RV,t} \sim (0,\ \sigma^2_{\omega,RV}) \quad (3.18)$$

Option data:

$$\sigma^2_{implied,t} = \mu_v + \theta + \gamma\,\frac{\sigma}{k}\left(\frac{1-\exp(-k\tau_t)}{k\tau_t} - 1\right) + \frac{\sigma^2}{2k^2}\left(1 - 2\left[\frac{1-\exp(-k\tau_t)}{k\tau_t}\right] + \frac{1-\exp(-2k\tau_t)}{2k\tau_t}\right)$$
$$+ \frac{1-\exp(-k\tau_t)}{k\tau_t}\, x^*_t + \varepsilon_t\,, \qquad \varepsilon_t \sim (0, \sigma^2_\varepsilon) \qquad (3.19)$$

Latent factor:

$$x^*_t = \exp(-k\Delta t)\, x^*_{t-\Delta t} + u_t\,, \qquad u_t \sim \left(0, \sigma^2\,\frac{1-\exp(-2k\Delta t)}{2k}\right). \quad (3.20)$$

The error series $\{\omega_t\}$, $\{\omega_{RV,t}\}$ and $\{u_t\}$ are white noise and mutually uncorrelated, except that $\{\omega_{RV,t}\}$ and $\{u_t\}$ are contemporaneously correlated with correlation $\text{corr}[\omega_{RV,t}, u_t] = c$. The option error series $\{\varepsilon_t\}$ is assumed white noise, and mutually uncorrelated with the other errors.

Parameter identification

Depending on the data used for estimation, not all parameters can either be estimated or may be identified. For example, if only return or RV data is used, the market price of volatility risk γ cannot be estimated. In those cases, the problem becomes one of estimating the parameters of an SV stock price model, given by equations (3.1)-(3.3).

The appendix to this chapter considers the identification problem in detail. Here we summarize the results. In case only *return data* is analyzed, all four parameters θ, k, σ_ω and σ that can potentially be estimated, are identified. In case only *RV data* is used, parameters θ, k, $\sigma_{\omega,RV}$, σ, c can potentially be estimated. Although θ and k can be identified, $\sigma_{\omega,RV}$, σ and c cannot separately be identified. As soon as one of these parameters is restricted, both others can be identified. An incorrect restriction may lead to bias in the other estimates. When working with empirical RV data in practice, it seems most natural to restrict $c = 0$, given its small magnitude in the simulations (see below) and our specific interest in the volatility-of-volatility parameter σ.

In case only *option data* is analyzed, potential parameters to be estimated are μ_v, θ, γ, σ, k and σ_ε, of which k, σ and σ_ε can be identified. Importantly, information on the other three parameters μ_v, θ, γ can only be derived from the mean BS implied variance over the sample. Parameters μ_v and θ cannot separately be identified; only their sum is identifiable. As the simulation results below show that μ_v is close to zero, it is most natural to restrict $\mu_v = 0$. The market price of volatility risk γ is then still left to be identified. The term multiplying γ in (3.19) knows time variation through the option maturity τ_t, which varies over t. This variation will nonetheless be rather limited if one analyzes one e.g. short-maturity option series. Therefore, although theoretically identified, the empirical identification of γ may appear weak. [22] Data information on γ may moreover "intermingle" with information on the sum $\mu_v + \theta$. [23]

[22] Theoretically, we expect γ to be more easily identifiable empirically, if the maturity-variation in the option series is increased; i.e. by including a second, longer-maturity option series for estimation. The

In case a combination of *return–option data* is analyzed, all potential parameters θ, k, σ, γ, μ_v, σ_ω, σ_ε can be identified. A similar story for γ holds nonetheless. If a combination of *RV-option data* is used for estimation, all potentially to-be-estimated parameters θ, k, σ, γ, μ_v, $\sigma_{\omega,RV}$, σ_ε, c can be identified. But again, possible weak empirical identification of γ remains in place.

3.4 Simulation results for 1-factor OU SV

This section discusses the simulation results for the 1-factor OU SV option pricing model. Before investigating the performance of the estimation method, we first pay attention to the various approximations and assumptions made, to arrive at a linear state space model.

Section 3.4.1 examines the quality of the linear and parabolic call price approximations, the magnitude of the parameter μ_v, and the relative merits measure. Section 3.4.2 examines the statistical properties assumed for the error terms of the state space model. Section 3.4.3 investigates the performance of the estimation method, based on different types of data. Central is the issue what type of data yields the most reliable estimation results.

3.4.1 Quality of the call price approximations

Linearization and parabolic errors

Recall the call price formula $C_t = \mathbb{E}_{\mathbb{Q}}[BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t]$, as implied by our multifactor model. In section 3.4 of the previous chapter, we linearized the function $g(Y_t) = BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2(Y_t))$ with $Y_t = \exp(\tau_t \bar{\sigma}_t^2)$ around the point $Y_t = b_t^* = \exp(\tau_t \sigma_{implied,t}^2)$, to arrive at the rewritten call price formula

$$C_t = g(b_t^*) - b_t^* g'(b_t^*) + \mathbb{E}_{\mathbb{Q}}\left[Y_t \mid \mathcal{F}_t\right] g'(b_t^*) + \mathbb{E}_{\mathbb{Q}}\left[HOT_t \mid \mathcal{F}_t\right], \tag{3.21}$$

in which $\mathbb{E}_{\mathbb{Q}}[Y_t \mid \mathcal{F}_t] = \exp[A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\mathbf{x}_t]$. To subsequently arrive at an equation that is linear in the unobserved factors $\mathbf{x}_t$, we proposed to neglect the *linearization error* $\mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]$. In a similar way, considering a parabolic approximation of the function $g(.)$, we obtained the following rewritten call price:

$$C_t = q_0(b_t^*) + q_1(b_t^*)\mathbb{E}_{\mathbb{Q}}\left[Y_t \mid \mathcal{F}_t\right] + q_2(b_t^*)\mathbb{E}_{\mathbb{Q}}\left[Y_t^2 \mid \mathcal{F}_t\right] + \mathbb{E}_{\mathbb{Q}}\left[HOT2_t \mid \mathcal{F}_t\right], \tag{3.22}$$

and dubbed the quantity $\mathbb{E}_{\mathbb{Q}}[HOT2_t \mid \mathcal{F}_t]$ the *parabolic error*.

Section 3.4.5 of the previous chapter provided some preliminary, graphical insight into the quality of the linear and parabolic approximations of the function $g(.)$. The simulated data of the 1-factor OU model allows for a more quantitative examination of the magnitude of the linear and parabolic errors.

Table 3.1 contains summary statistics of the (absolute value of the) relative linearization and parabolic errors, over the full sample of 100*1058 observations. Regarding the magnitude of the simulated call option prices C_t over the full

empirical results in section 4 show however, that the 1-factor SV model is heavily misspecified for the joint data. Thus,although theoretically an appealing strategy, in practice it is not due to misspecification.

[23] By this we mean that if the option maturity τ_t were constant over the sample, none of the parameters μ_v, θ and γ would be separately identified. Hence, in case of little variation in τ_t, clear empirical identification (i.e. based on data) may still appear to be hard to establish.

sample: The average call price equals 318 (std.dev. 213). (The mean simulated stock price S_t over the sample equals 5850 (std.dev. 3749)).

Table 3.1: Summary statistics (absolute value of) relative linearization and parabolic errors

	$\mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]/C_t$	$\mid \mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]/C_t \mid$	$\mathbb{E}_{\mathbb{Q}}[HOT2_t \mid \mathcal{F}_t]/C_t$	$\mid \mathbb{E}_{\mathbb{Q}}[HOT2_t \mid \mathcal{F}_t]/C_t \mid$
Mean	-0.13%	0.14%	-0.0013%	0.087%
Std.dev.	0.12%	0.10%	0.11%	0.068%
Minimum	-1.49%	0.00%	-0.67%	0.00%
Maximum	0.44%	1.49%	0.57%	0.67%

The errors are small in general, suggesting that the linear and parabolic approximations provide accurate approximations of the true call price. The linearization error is typically negative, indicating that a linearization typically over-estimates the true call price. The parabolic errors more evenly spread around zero. Figure 3.1 shows the relative errors for a typical simulated dataset. The errors seem rather randomly distributed.

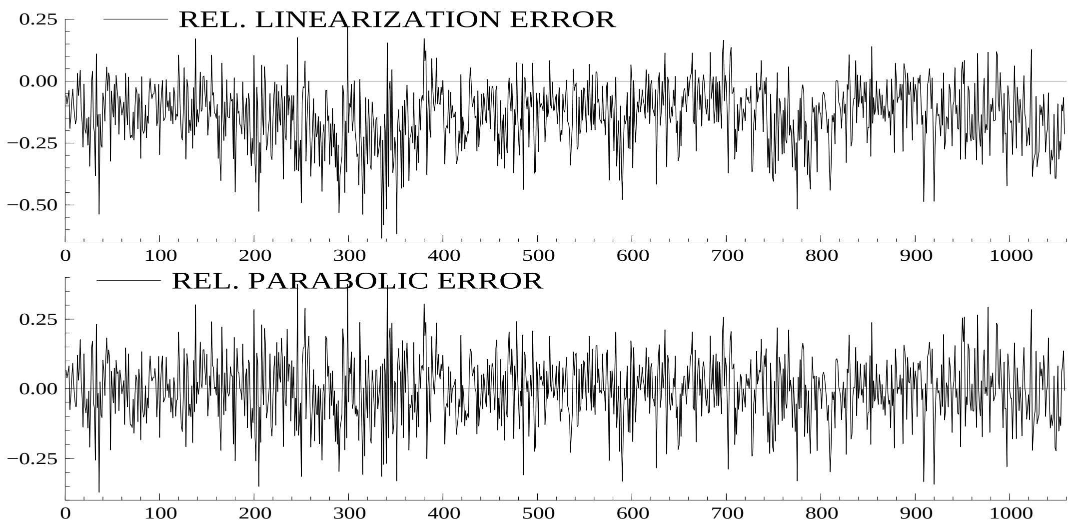


Figure 3.1: Relative linearization errors $\mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]/C_t * 100\%$, and relative parabolic errors $\mathbb{E}_{\mathbb{Q}}[HOT2_t \mid \mathcal{F}_t]/C_t * 100\%$, for a typical simulated dataset.

"Parameter" μ_ν

From Jensen's inequality, we showed in section 3.4.2 of the previous chapter that $\sigma^2_{implied,t} \leq \frac{1}{\tau_t}[A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\mathbf{x}_t]$. This result provided a second rationale for the options measurement equation as $\sigma^2_{implied,t} = \mu_\nu + \frac{1}{\tau_t}[A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\mathbf{x}_t] + \varepsilon_t$, with $\varepsilon_t \sim (0, \sigma^2_\varepsilon)$. We outlined how to estimate the "parameter" μ_ν based on a simulated sample of T observations.

Summary statistics of the 100 estimates of μ_ν are in table 3.2. As expected, μ_ν is negative for all datasets. Its magnitude is very close to zero. Its relative magnitude (i.e. μ_ν as percentage of the average $\sigma^2_{implied,t}$ over the sample) is also small, on average only –0.25%. This suggests that restricting $\mu_\nu = 0$ prior to estimation may not result in any major bias in the other estimates, though may possibly lead to an improved empirical identification of the market price of volatility risk, γ. (Recall the identification discussion in section 3.3.)

Table 3.2: Summary statistics of the 100 μ_v-estimates and relative merits measures

	μ_v	μ_v *as % of mean BS impl.var.*	*Relative merits measure*
Mean	-0.00041	-0.25%	0.056%
Std.dev.	$2.1*10^{-5}$	0.023%	0.0084%
Minimum	-0.00047	-0.32%	0.040%
Maximum	-0.00036	-0.20%	0.093%

Relative merits of using a parabolic vs. a linear approximation

In section 3.4.5 of the previous chapter, we introduced the *relative merits measure* as an indicative measure to investigate if a parabolic (rather than a linear) approximation is really worth the additional complexity in the computations. We defined it as $\frac{1}{T}\sum_t^T \left[| \mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]/C_t | - | \mathbb{E}_{\mathbb{Q}}[HOT2_t \mid \mathcal{F}_t]/C_t | \right]$ *100%, i.e. as the percentual difference between the mean absolute relative linearization and parabolic errors.

Table 3.2 shows that the average of the 100 relative merits measures equals 0.056% only. This suggests that the improved accuracy obtained from using a parabolic rather than a linear approximation may be expected to be limited [24], and hence does not seem worth the extra effort. (In practice, simple is better.)

3.4.2 Disturbance terms of the state space model

The validity of the Kalman filter equations depends on the white noise assumption for the different disturbance terms that appear in the state space model; i.e. $\{u_t\}$, $\{\omega_{RV,t}\}$, $\{\omega_t\}$ and $\{\varepsilon_t\}$. We were able to prove that the volatility-factor shock series $\{u_t\}$ is indeed white noise. But what about the other error terms?

Error series $\{\omega_{RV,t}\}$ (RV equation)

Recall section 3.3 on RV of the previous chapter. In deriving the equation for the average variance as $\frac{1}{\Delta t}\int_t^{t+\Delta t} \sigma_u^2 du = \mathbb{E}_{\mathbb{P}}[\frac{1}{\Delta t}\int_t^{t+\Delta t} \sigma_u^2 du \mid \mathcal{F}_t] + \varpi_{RV,t+\Delta t}$ over day $[t, t+\Delta t]$, we proved that the series $\{\varpi_{RV,t}\}$ is white noise. Based on this equation, we next proposed a measurement equation for RV in which we assumed its associated error series $\{\omega_{RV,t}\}$ to be white noise: $\omega_{RV,t} \sim (0, \sigma^2_{\omega,RV})$. Although reasonable, this remains an assumption, as $\sigma^2_{RV,t+\Delta t}$ measures $\frac{1}{\Delta t}\int_t^{t+\Delta t} \sigma_u^2 du$ with error. The simulated data allows us to examine this assumption.

The simulated data confirm that considering the series $\{\omega_{RV,t}\}$ as white noise is indeed reasonable: The average mean of $\{\omega_{RV,t}\}$ over all 100 datasets equals $-1.6*10^{-5}$, and the series are neither significantly autocorrelated nor heteroskedastic unconditionally. [25]

Recall that we proved the series $\{\varpi_{RV,t}\}$ and $\{u_t\}$ to be contemporaneously correlated only, with approximate correlation coefficient 0.86. Based on this result, we allow for contemporaneous correlation between $\{\omega_{RV,t}\}$ and $\{u_t\}$ in the state space model. The simulated data supports this assumption: The average contemporaneous correlation coefficient between $\{\omega_{RV,t}\}$ and $\{u_t\}$ over the 100

[24] If we only use one e.g. ATMF option series for estimation, or multiple series that have approximately the same moneyness (e.g. 3 ATMF series). Incorporating different-moneyness option series requires a parabolic approximation as explained in section 3.4.5 of the previous chapter.

[25] We checked this for several simulated datasets. The empirical analysis was done in EViews. To save space we prefer to summarize our findings verbally.

datasets equals 0.086 (with a standard deviation of 0.031, a maximum of 0.17, and a minimum of 0.016). As the number of intraday sampling points I equals 48 only (and not infinite), the simulated correlations are much smaller than 0.86.

Error series $\{\omega_t\}$ (Squared return equation)
Treating the Euler discretization of the stock price SDE as exact, we proved that $\{\omega_t\}$ is white noise, and is uncorrelated with $\{u_t\}$. However, as the discretization is not fully exact, these are actually merely assumptions.

We find that the mean of the simulated $\{\omega_t\}$ series averaged over the 100 datasets equals $1.2*10^{-4}$, which is close to zero. The non-autocorrelatedness and homoskedasticity assumptions are generally supported by the simulated data as well. The white noise assumption for $\{\omega_t\}$ seems therefore reasonable (though its distribution is largely asymmetric [26]). Cross-correlograms between $\{\omega_t\}$ and $\{u_t\}$ show that these series are indeed generally non-contemporaneously uncorrelated. The average contemporaneous correlation coefficient between the hundred $\{\omega_t\}$ and $\{u_t\}$ series equals 0.011 (and hence centers close to zero), with a standard deviation of 0.033, a minimum of –0.084 and a maximum of 0.083. Assuming $\{\omega_t\}$ and $\{u_t\}$ to be uncorrelated seems therefore reasonable.

Option error series $\{\varepsilon_t\}$
The reason for assuming the option error series $\{\varepsilon_t\}$ to be white noise is mainly for convenience (see section 3.4.2 of the previous chapter). Recall that in theory, our method for extracting information from option prices essentially boils down to replacing a deterministic summation term with the random sum $\mu_v + \varepsilon_t$ (recall (3.69)). This is clearly strange if we believe our theoretical model to exactly hold in practice. It is not strange however if we acknowledge that a model is never a complete description of reality.

In a simulation environment (where theory and reality coincide), we can compute the "random" error series $\{\varepsilon_t\}$, and examine if it at least resembles white noise. The average mean of the 100 $\{\varepsilon_t\}$-series over all datasets equals $1.0*10^{-20}$, whereas the average standard deviation equals 0.00036. Graphs of the $\{\varepsilon_t\}$-series show that the homoskedasticity assumption seems reasonable. The series appears to be slightly autocorrelated. The first five autocorrelation coefficients, averaged over all datasets, equal 0.034, 0.034, 0.031, 0.025, 0.020 (with standard deviations 0.037, 0.040, 0.033, 0.033, 0.037).

And what about the assumed zero-correlation between $\{\varepsilon_t\}$ and the other errors $\{\omega_t\}$, $\{\omega_{RV,t}\}$ and $\{u_t\}$? Inspection of the cross-correlograms reveals that this assumption seems reasonable. Specifically, we find for the various cross-correlation coefficients (averaged over all 100 datasets): $\text{corr}(\varepsilon_t, \omega_t) = 0.00085$, $\text{corr}(\varepsilon_t, \omega_{t-\Delta t}) = 0.00188$, $\text{corr}(\varepsilon_t, \omega_{t-2\Delta t}) = -0.00664$, $\text{corr}(\varepsilon_t, \omega_{RV,t}) = 0.00324$, $\text{corr}(\varepsilon_t, \omega_{RV,t-\Delta t}) = -0.00151$, $\text{corr}(\varepsilon_t, \omega_{RV,t-2\Delta t}) = -0.00275$, $\text{corr}(\varepsilon_t, u_t) = 0.03132$, $\text{corr}(\varepsilon_t, u_{t-\Delta t}) = 2.75 * 10^{-6}$ and $\text{corr}(\varepsilon_t, u_{t-2\Delta t}) = 0.00061$.

As an illustration, figure 3.2 shows the error series $\{\omega_t\}$, $\{\omega_{RV,t}\}$ and $\{\varepsilon_t\}$ for a typical dataset (i.e., the same dataset as used in figure 3.1). The error series $\{\omega_{RV,t}\}$ and $\{\varepsilon_t\}$ seem to largely resemble white noise. The series $\{\omega_t\}$

[26] The distribution of $\{\omega_t\}$ is skewed to the right as we extract information from *squared* returns; see also figure 3.2. Hence, although $\{\omega_t\}$ may be close-to white noise, as its distribution is not symmetric (and hence deviates a lot from the Gaussian distribution), we a priori may perhaps expect the *quasi* maximum likelihood procedure to perform poorly. (See also section 3.2.1 of the previous chapter.)

associated with the squared return equation may look heteroskedastic. This is deceptive however: Running a regression of ω_t on a constant and its lag results in an insignificant coefficient, and a White test does not reject homoskedasticity of the residuals.

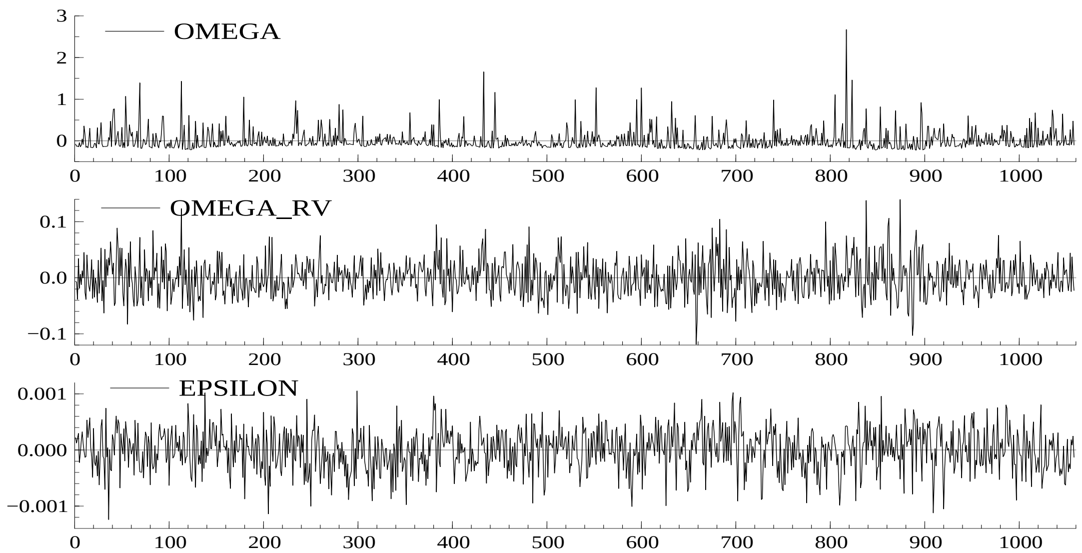


Figure 3.2: The simulated error terms $\{\omega_t\}$, $\{\omega_{RV,t}\}$ and $\{\varepsilon_t\}$ for a typical dataset.

We conclude that the properties of the errors $\{\omega_t\}$, $\{\omega_{RV,t}\}$ and $\{u_t\}$ closely correspond to the assumptions underlying the linear state space framework, and on which the Kalman filter equations are built. For the option error series $\{\varepsilon_t\}$ this holds by approximation only, due to its small positive autocorrelation. This may possibly lead to estimation bias when estimating the state space model. The next section investigates this further.

3.4.3 Performance of the state space estimation method

Our goal here is to explore the performance of our Kalman filter QML estimation method for the 1-factor OU SV option pricing model. If we are willing to ignore the small autocorrelation in the $\{\varepsilon_t\}$-series, the state space model derived from the OU SV option pricing model fits exactly into the (unconditional) linear state space framework. We know that Kalman filter QML yields consistent and asymptotically normal estimates in that framework. Hence, apart from examining its finite-sample behavior in our specific context, why should we bother about investigating its performance, if we already know that the method yields consistent estimates?

The main reason why this is still important has, of course, mostly to do with *how* we derived the measurement equations for the squared returns, RV and the options. All three equations were obtained from approximations, with mainly the option equation obtained in a far from trivial way. The issue thus boils down to the question if the approximations perform reasonably well.

Are the estimates obtained from our method close to their true values? Are the stock volatilities extracted by the Kalman smoother close to the underlying true stock volatilities? What data yields the best results in these respects?

For parameter set $\theta = 0.16$, $k = 5.25$, $\sigma = 0.10$, $\gamma = -1$, we estimate the unconditional state space model (3.17)-(3.20) using the 100 simulated datasets from the 1-factor OU model. We consider estimation based on the following types of data: only squared return data, only RV data, only (short-maturity ATMF) option data, both return and option data, and the combination RV – option data. (We refer to section 3.3 for a discussion on identification.) In the latter two cases, we also consider restricting $\mu_v = 0$ prior to estimation: As the simulations show that μ_v is close to zero, restricting $\mu_v = 0$ may possibly lead to an improved empirical identification of the market price of volatility risk γ, as explained earlier.

Table 3.3: Estimation results **1-factor OU SV** based on 100 simulated datasets.

	Sq.return data	*RV data ($c = 0$)*	*Option data ($\mu_v = 0$)*	*Sq.return & option data*	*Sq.return & option data ($\mu_v = 0$)*	*RV & option data*	*RV & option data ($\mu_v = 0$)*
$\theta = 0.16$	**0.160**	**0.160**	**0.159**	**0.160**	**0.159**	**0.160**	**0.159**
	(0.012)	(0.009)	(0.008)	(0.011)	(0.009)	(0.009)	(0.009)
Bias	-3e-4	-4e-4	-7e-4	-3e-4	-7e-4	-5e-4	-6e-4
MSE	1e-4	8e-5	8e-5	1e-4	8e-5	8e-5	8e-5
$k = 5.25$	**13.74**	**6.74**	**5.42**	**5.41**	**5.41**	**5.30**	**5.29**
	(18.28)	(2.59)	(0.79)	(0.75)	(0.75)	(0.49)	(0.48)
Bias	8.49	1.49	0.17	0.16	0.16	0.05	0.04
MSE	406	8.95	0.64	0.59	0.59	0.24	0.23
$\sigma = 0.10$	**0.149**	**0.103**	**0.100**	**0.100**	**0.100**	**0.099**	**0.099**
	(0.139)	(0.013)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Bias	0.049	0.003	-2e-4	-2e-4	-2e-4	-8e-4	-8e-4
MSE	0.022	2e-4	2e-5	2e-5	2e-5	1e-5	1e-5
$\gamma = -1$	-	**-**	**-1.00**	**-1.00**	**-1.00**	**-1.00**	**-0.97**
			(0.55)	(0.55)	(0.55)	(0.54)	(0.51)
Bias	-	-	0.00	0.00	0.00	0.00	0.03
MSE	-	-	0.30	0.30	0.30	0.30	0.26
$\mu_v \approx -4e\text{-}4$	**-**	**-**	**0**	**-3.7e-4**	**0**	**-2.0e-4**	**0**
				(0.0067)		(0.0017)	
$c \approx 0.086$	**-**	**0**	**-**	**-**	**-**	**0.085**	**0.085**
						(0.031)	(0.031)
RMSPE	**8.38**	**3.30**	**0.60**	**1.92**	**0.58**	**0.41**	**0.35**
in %	(1.31)	(0.26)	(0.31)	(1.33)	(0.29)	(0.15)	(0.12)
MAPE	**6.54**	**2.59**	**0.47**	**1.81**	**0.46**	**0.34**	**0.28**
in %	(0.91)	(0.18)	(0.23)	(1.31)	(0.22)	(0.15)	(0.10)

The Kalman filter QML estimation results are summarized over the 100 simulated datasets jointly. The state space model (3.17)-(3.20) was estimated using different types of data. The average estimates of θ, k, σ, γ are reported (std.dev. in parentheses), with associated sample bias and MSE. The average estimates of μ_v, c (std.dev. in parentheses) ought to be compared with their average simulated values of $\mu_v \approx -4e\text{-}4$ and $c \approx 0.086$. The average RMSPE and MAPE for comparing the smoothed and true volatilities are also reported (std.dev. in parentheses).

Table 3.3 reports the main estimation results, averaged over all datasets. The average estimates of θ, k, σ, γ are reported in boldface (standard deviations in parentheses), together with the sample bias and mean squared errors (MSE). As the exact values of μ_v and $c = \text{corr}[u_t, \omega_{RV,t}]$ are not known, these may be contrasted to their average simulated values. The *root mean squared percent error* (RMSPE) and *mean absolute percent error* (MAPE) for comparing the smoothed stock volatilities (obtained from the Kalman smoother at the optimum) with the true underlying volatilities are given in the bottom part. [27]

Squared returns perform worst, RV – option data best

Using *squared returns* only for estimation performs worst. Although θ (unconditional stock volatility) is estimated well, the estimates of the mean-reversion (k) and volatility-of-volatility (σ) parameters are heavily biased and have large standard deviations. The smoothed and true volatilities deviate most in this case. The results confirm why squared returns are typically considered noisy estimators of the stock variance (see e.g., Andersen and Bollerslev (1998)).

As expected, using 10-minute *RV data* instead, leads to a substantial improvement in bias, MSE and volatility evaluation criteria. [28] Figure 3.3 illustrates the fact that squared returns contain far less precise information than RV. It plots the squared returns $(r_t - \mu\Delta t)^2 / \Delta t$, the realized variances $\sigma^2_{RV,t}$ and the true underlying stock variances $\sigma^2_t = x_t$ in one graph, for a typical dataset.

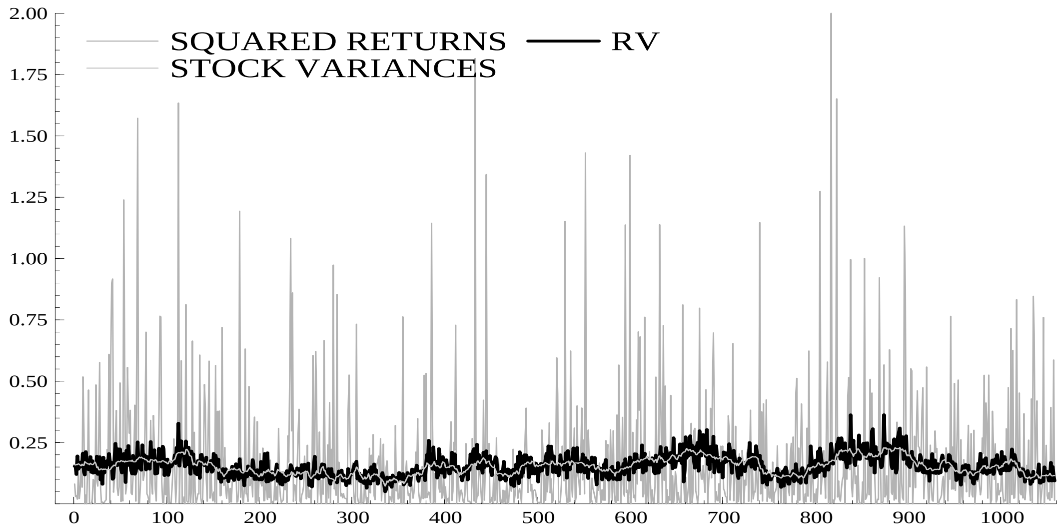


Figure 3.3: Squared returns $(r_t - \mu\Delta t)^2 / \Delta t$, realized variances $\sigma^2_{RV,t}$, and true underlying stock variances $\sigma^2_t = x_t$ for a typical dataset.

In turn, *option data* outperforms RV on all criteria. The bias and MSE of the estimates are small, and the RMSPE and MAPE indicate that the smoothed and true volatilities are close. Option data is clearly very informative. Figure 3.4 visualizes the dramatic increase in volatility information, obtained from subsequently using squared returns, RV, and options data for estimation. Option data contains the strongest signal on the hidden underlying volatilities.

Combining *return and option data* performs somewhat better than only option data; at least if μ_v is restricted to zero. Remarkably, although restricting $\mu_v = 0$ does not impact on estimation bias and MSE (which is understood given its small

[27] To remind: $\text{RMSPE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\sigma_t - \hat{\sigma}_{t|T}}{\sigma_t}\right)^2} * 100\%$, $\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\sigma_t - \hat{\sigma}_{t|T}}{\sigma_t}\right| * 100\%$.

[28] Using only RV data, but imposing the restriction that $c = \text{corr}[u_t, \omega_{RV,t}]$ equals its dataset-specific simulated value, leads to virtually the same results. The evident reason is that the simulated c's are generally close to 0 (i.e. they average at 0.086).

magnitude), the RMSPE and MAPE are much lower than if μ_V is not restricted. This result suggests that restricting $\mu_V = 0$ is preferable. [29]

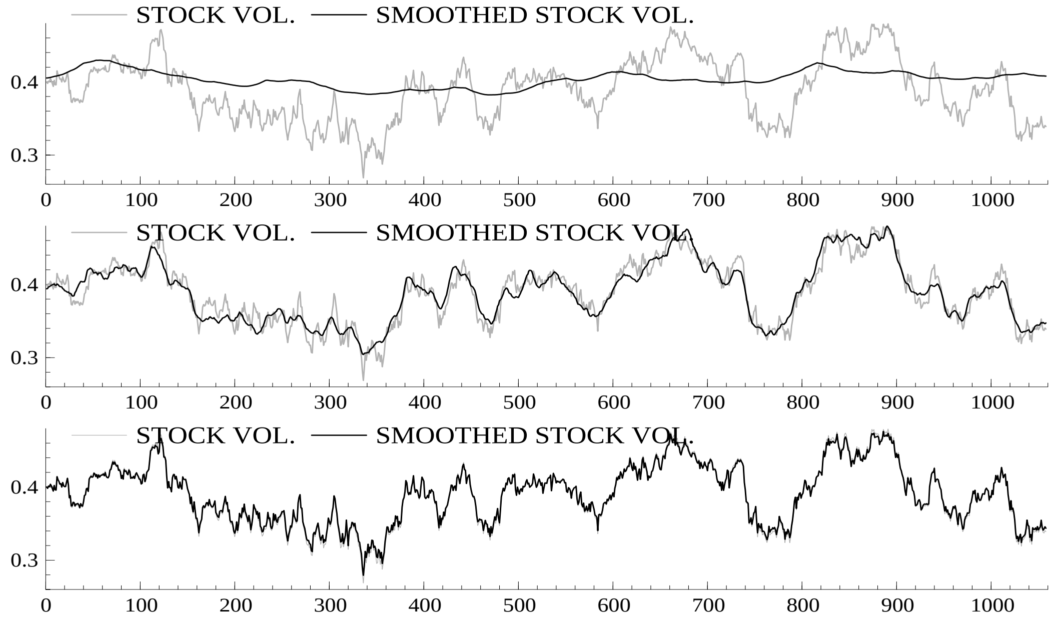


Figure 3.4: Smoothed and true underlying stock volatilities, in case only squared return data (upper), only RV data (middle), or only option data (lower) is used for estimation.

The best results are obtained from using both *RV and option data* for estimation. The bias is very small for all estimates, it performs best on the MSE criterion, and the smoothed and underlying true volatility series are particularly close. As expected, there are *efficiency gains* from using RV-option data as opposed to e.g. return-option data. As in the return-option case, the results suggest that restricting $\mu_V = 0$ prior to estimation is preferable. This combination of data contains the most precise information. As such, if this data is available in practice its use is to be advocated for obtaining the most reliable estimates.

Market price of volatility risk γ

Although there is no estimation bias associated with the market price of volatility risk (γ), the precision of the estimates is poor. This is irrespective of the data used for estimation. Precise information on γ is thus difficult to distill. Restricting $\mu_V = 0$ does not help in this respect. [30]

We conclude the following from this Monte Carlo study towards the 1-factor OU SV option pricing model. Although we have considered one (rather realistic) parameter set only, the results suggest that our QML estimation strategy performs well for this model if option data is included. This is partly attributed to the volatility factor being Gaussian in the OU SV case, resulting in QML estimation being "close to" ML estimation. Apparently, our option price approximation works very well in this case. (In the next chapter, in which modeling of level-dependent volatility-of-volatility is taken into account, we perform a simulation study for the (non-Gaussian) 1-factor CIR SV model, for *five* different sets of parameters. There, we do not have the problem of sampling negative stock variances.)

[29] The only modest improvement obtained from using both return and option data, as opposed to using option data only, is attributed to the inherent noisiness of squared returns.

[30] See footnote 22 for remarks about including a second, longer-maturity option series for estimation.

4. FTSE100-index data: Results for 1-factor OU SV

This section presents estimation results for the 1-factor OU SV option pricing model based on the FTSE100-index data, for the period Oct 1997 – Dec 2001. Initially, in section 4.1, we estimate the model using three types of data: only squared returns, only the SM option data, and a combination of these. We interpret the results and contrast them to GARCH estimates. Section 4.2 compares the SV with GARCH volatilities and the BS implied volatilities, and considers the in-sample fit of the state space model. Section 4.3 considers the compensation for FTSE100-index volatility risk that is implicitly present in the data. Section 4.4 considers diagnostic checking. Special attention is given to the out-of-sample overpricing of the longer-maturity options. In section 4.5, we estimate the 1-factor OU model using the four time series of squared returns and ATM options jointly. The model appears considerably dynamically misspecified.

4.1 Parameter estimates

We estimate the 1-factor OU SV model using three types of data: only squared return data, only SM option data, and both squared return – SM option data. Given the Monte Carlo evidence reported in the previous section, we restrict $\mu_v = 0$ prior to estimation. [31] We also estimate a Gaussian GARCH(1,1) model for the daily (prewhitened) FTSE100-index returns. [32] As both our model and the GARCH model do not allow for the leverage effect, this permits a consistent comparison. The GARCH(1,1) conditional stock variance follows $\sigma^2_{t+\Delta t} = \varphi_0 + \varphi_1(r_t - \mu\Delta t)^2 + \varphi_2\sigma^2_t$. We find

$$\sigma^2_{t+\Delta t} = \underset{(4.8e-6)}{9.9e-0.6} + \underset{(0.028)}{0.086}\,(r_t - \hat{\mu}\Delta t)^2 + \underset{(0.05)}{0.85}\,\sigma^2_t\,, \tag{4.1}$$

with robust standard errors in parentheses. For the GARCH model, the volatility persistence is measured by $\varphi_1 + \varphi_2$, the half-life of a volatility shock by $-\ln 2/\ln(\varphi_1 + \varphi_2)$, and the per-annum unconditional stock volatility may be approximated by $\sqrt{\varphi_0/[(1-\varphi_1-\varphi_2)\Delta t]}$. To obtain a rough estimate of the per-annum volatility-of–the-variance, we compute the standard deviation of the estimated daily GARCH variance series, after multiplication by 260.

Table 4.1 reports the estimation results, with robust White (1982) QML standard errors in parentheses. If only return or return–option data is used, the unconditional stock volatility ($\sqrt{\theta}$) is estimated rather close to its *method-of-moments estimate* of $\sqrt{\hat{\theta}_{SqR}}$ = 20.3%, computed as the root sample average of the squared returns. GARCH implies an identical value. If only option data is used, θ is seemingly estimated too large. Restricting θ to $\hat{\theta}_{SqR}$ prior to estimation, leads to virtually the same results, except for the market price of volatility risk γ, which is then estimated negative. (Notice the very marginal reduction in the quasi-loglikelihood.) This illustrates our remarks made in section

[31] Restricting μ_v = 0 leads to better results in the Monte Carlo study. Moreover, leaving μ_v unrestricted in case of return-option data, results in an estimate of 0.0174 with a standard error of 0.0030. From the Monte Carlo study and Jensen's inequality, we know that μ_v ought to be very close to zero but negative, if the data were truly generated by the 1-factor OU model. However, we a priori believe that such is not the case. The misspecification partly translates itself in this estimate of μ_v: As this parameter appears as a constant term in just one place in the state space model (the option equation), it easily picks up such misspecification. Restricting $\mu_v = 0$ yields more plausible results.

[32] Notice that this essentially boils down to 2-step estimation of an AR(2)-GARCH(1,1) model for the *non*-prewhitened returns. Note the analogy with our state space method; see footnote 9.

3.3 on the difficult empirical identification of θ and γ, if only option data is used that features little maturity variation. In that case, information on θ and γ seems to intermingle and hard to distinguish from each other. Section 4.3 interprets the magnitude of the market price of volatility risk.

Table 4.1: Estimation results **1-factor OU SV** (FTSE100-index data; Oct 1997 – Dec 2001)

	Sq.return data	*SM option data*	*SM option data* $\theta = \hat{\theta}_{SqR}$	*Sq.return & SM option data*	*GARCH*
θ	**0.0409**	**0.0572**	**0.0413**	**0.0450**	
	(0.0052)	(0.0136)		(0.0317)	
k	**16.3**	**7.12**	**7.00**	**1.85**	
	(6.93)	(5.05)	(5.00)	(1.78)	
σ	**0.164**	**0.211**	**0.210**	**0.151**	
	(0.049)	(0.056)	(0.056)	(0.022)	
γ	-	**0.128**	**-0.437**	**-0.967**	
		(0.541)	(0.176)	(0.527)	
σ_ω	**0.0612**	-	-	**0.0657**	
	(0.0056)			(0.0062)	
σ_ε	-	**0.0029**	**0.0029**	**0.0038**	
		(0.0018)	(0.0018)	(0.0018)	
Vol. returns	20.2%	23.9%	20.3%	21.2%	20.3%
Vol-of-var.	0.0287	0.0558	0.0560	0.0785	0.0175
Persistence	0.939	0.973	0.973	0.993	0.937
Half-life (days)	11	25	26	98	11
Std.dev. u_t	0.0099	0.0129	0.0128	0.0093	
Loglikelihood	1394	3408	3407	4761	3165

Parameter estimates are in boldface, with robust White (1982) QML standard errors in parentheses. We estimated the state space model (3.17)-(3.20) using only return data, only SM option data, only SM option data under the restriction $\theta = \hat{\theta}_{SqR} = 0.0413$, and both return - SM option data. Also reported are the unconditional stock volatility $\sqrt{\theta}$, the volatility-of-the-variance $\sqrt{\sigma^2/2k}$, the volatility persistence $\exp[-k\Delta t]$, the half-life (in days) of a volatility shock $\ln 2/k\Delta t$, the QMLE of the standard deviation of the daily volatility shock u_t, and the maximized quasi-loglikelihood value. The last column shows comparable quantities for the estimated Gaussian GARCH(1,1) model (4.1).

The volatility persistence ($\exp[-k\Delta t]$) is estimated smallest (or the mean-reversion quickest) in case only return data is analyzed. The persistence of 0.939 corresponds to the value implied by the GARCH model. This is not surprising given that both the 1-factor SV and GARCH(1,1) models can be rewritten as ARMA(1,1) models for the squared returns. (See section 3.2.2 of the previous chapter). Using only option data, the estimated persistence of 0.973 is larger, but not as large as the 0.993 implied by the combination return-option data. A persistence of 1 would imply random walk behavior for the stock volatility. Assuming (for the moment) that the 1-factor OU model is correctly specified, and taking the Monte Carlo results into account, finding an estimated persistence from the (noisy) squared return data that is so much smaller makes sense.

4.2 Volatilities and in-sample fit

Consider the estimation results obtained from using return and SM option data jointly. Figure 4.1 shows the smoothed stock volatilities ($\sigma_{t|T} := \sqrt{x^*_{t|T} + \theta}$) obtained from the Kalman smoother at the optimum, the annualized GARCH volatilities obtained from (4.1), and the "observed" SM BS implied volatilities. The smoothed volatilities will be referred to as the *SV volatilities* below. The GARCH volatilities were annualized by multiplying the daily volatilities with $\sqrt{260}$.

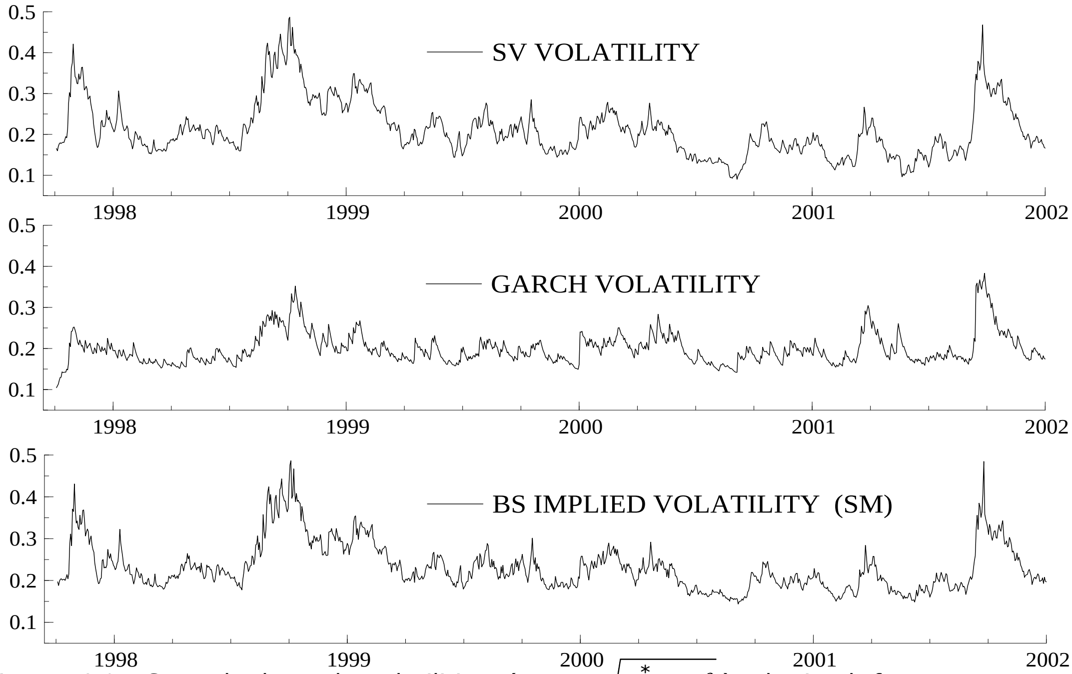


Figure 4.1: Smoothed stock volatilities ($\sigma_{t|T} = \sqrt{x^*_{t|T} + \theta}$) obtained from state space estimation using **return and SM option data**, annualized GARCH volatilities, and the Black-Scholes implied volatilities associated with the SM option series.

GARCH versus SV volatilities

The pattern in the three volatility series is roughly similar, although the GARCH series looks somewhat *smoother* and does not peak as much as the other series. Moreover, in contrast to the SV volatilities, the GARCH volatilities seem to obey some *lower barrier*, which is virtually never exceeded. The SV volatilities pick up *fast-changing patterns* in the data better than GARCH does.

The smoother behavior of GARCH may be caused by the fact that the impact of a time-t shock in the conditional mean return (i.e., $r_t - \mu\Delta t$) on the time-$t + \Delta t$ GARCH variance is given by φ_1 times the *squared* shock $(r_t - \mu\Delta t)^2$, which is typically small. The SV model in contrast, allows for an autonomous shock $u_{t+\Delta t}$ that fully impacts on the SV conditional variance of time $t + \Delta t$: $\sigma^2_{t+\Delta t} = \theta[1 - \exp(-k\Delta t)] + \exp(-k\Delta t)\sigma^2_t + u_{t+\Delta t}$. This also explains why SV tracks sudden changes in the data better than GARCH does. The virtual lower barrier in the GARCH series is probably due to the fact that shocks $\phi_1(r_t - \mu\Delta t)^2$ to the GARCH variance are always positive, whereas shocks $u_{t+\Delta t}$ to the SV variance can also be negative.

BS implied volatilities and the market price of volatility risk

Figure 4.2 plots the differences between the three volatility series (notice the scale). The observed BS implied volatilities are generally larger than the SV volatilities. Hence, the market consistently seems to expect a future volatility larger than the current, "true" volatility; the BS implied volatilities seem to be biased upward. They are also generally larger than the GARCH volatilities, though this is less so in the second part of the sample. The difference between the SV and GARCH volatilities fluctuates more around zero (but can be substantial).

These observations are understood as follows. For ATM options, our model implies $\sigma^2_{implied,t} \approx \mathbb{E}_{\mathbb{Q}}[\bar{\sigma}^2_t \mid \mathcal{F}_t]$. If volatility risk is not priced, then $\sigma^2_{implied,t} \approx \mathbb{E}_{\mathbb{P}}[\bar{\sigma}^2_t \mid \mathcal{F}_t]$. We therefore attribute the fact that the BS implied volatilities are found to be larger than the "true" (i.e., SV $\mathbb{P}$-) volatilities to the negative market price of volatility risk (γ), as this implies a risk-neutral mean volatility larger than the mean $\mathbb{P}$-volatility. This also explains why the difference between the SV and GARCH (which are both $\mathbb{P}$-) volatilities centers more near zero. The lower barrier of GARCH explains why the BS implied volatilities are less large than GARCH in the second part of the sample. Finally, observe that the common practical interpretation of BS implied volatilities being forecasts of future ($\mathbb{P}$) volatility (see e.g. Hull (2003)) should clearly be taken with care if volatility risk is priced.

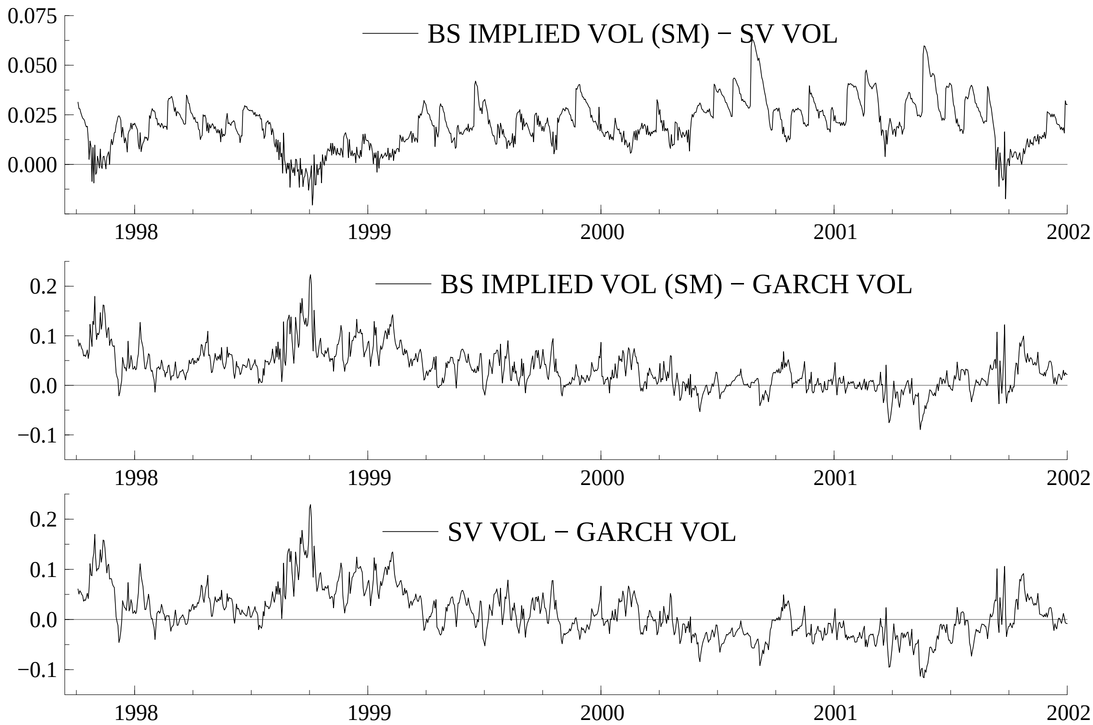


Figure 4.2: Differences between: SM BS implied and SV volatilities (upper graph), SM BS implied and GARCH volatilities (middle), and SV and GARCH volatilities (lower).

GARCH vs. SV volatilities obtained from squared returns

The SV volatilities obtained from state space estimation using return data only, seem more directly comparable with the GARCH volatilities, as both are based on information in stock returns only. Figure 4.3 plots these volatilities and their difference. Although both series track each other rather well, the GARCH volatilities *lag behind* the SV volatilities. Moreover, the volatility-of-volatility is for GARCH less (see table 4.1).

This confirms that GARCH does not as quickly respond to news as SV does, as GARCH does not allow for unpredictable news, but instead models today's volatility as a deterministic function of yesterday's information, in a recursive manner. Comparing figure 4.1 with 4.3, reveals that the SV volatilities obtained from the return data only, evolve much *smoother* than those extracted from the joint return – SM option data. This is as expected, given the less precise information in squared returns.

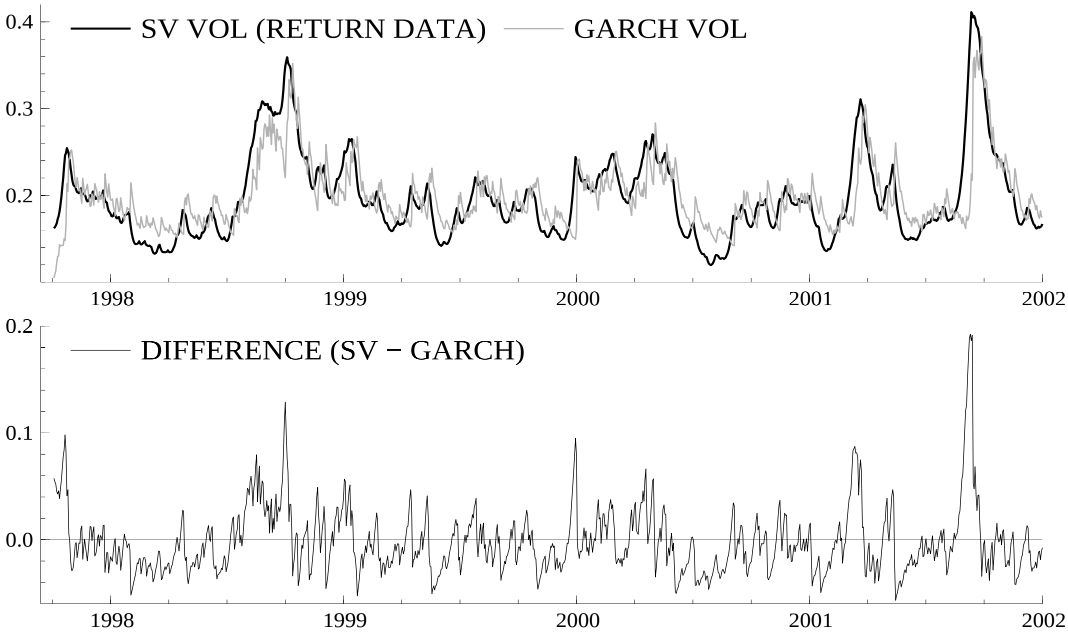


Figure 4.3: The SV volatilities obtained from estimating the state space model using **return data only**, the annualized GARCH volatilities, and their difference.

Relation between shape VTS and current volatility

Table 2.2 reported the shape frequencies of the volatility term structure observed in the sample. But when does a particular shape occur most often? Figure 4.4 plots the distribution (i.e. *boxplots*) of the SV volatilities (obtained from return – option data) for each of the four shapes the VTS can have on each day. An *upward* (resp. downward) *sloping* VTS is generally associated with a *low* (high) current stock volatility. This observation supports *volatility mean reversion* and the notion that traders attach forecasting value to BS implied volatilities.

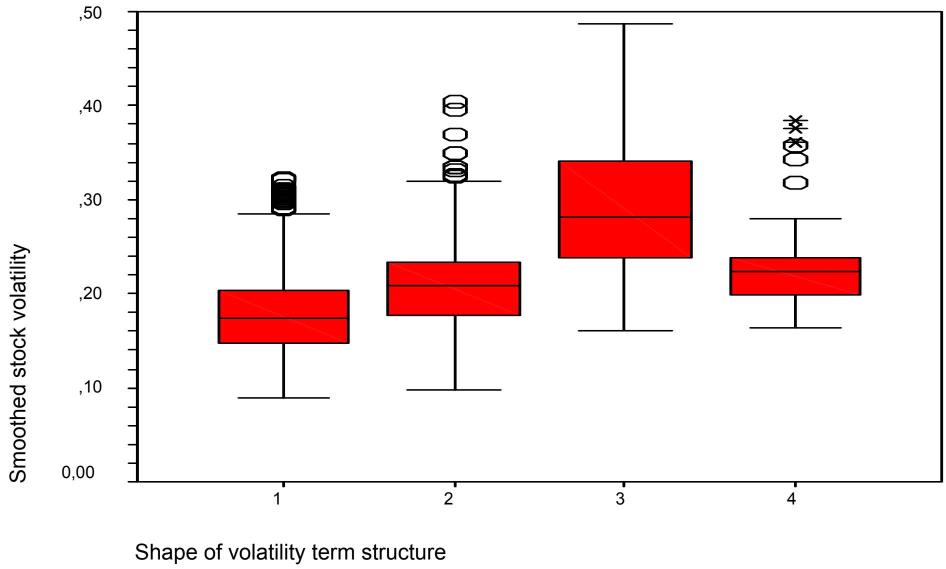


Figure 4.4: Distribution (boxplots) of the smoothed stock volatility (obtained from return - SM option data) for each of the four VTS shapes (1 = upward sloping, 2 = hump shape, 3 = downward sloping, 4 = inverted hump shape).

In sample fit
Consider again the estimated state space model based on the combination return - SM option data. Figure 4.5 shows the extent in which the model fits the data. The upper graph displays the squared returns and their inherent signal, the smoothed stock variances $\sigma^2_{t|T} = x^*_{t|T} + \theta$ (which equal the fitted squared returns). The lower graph shows the observed SM BS implied variances and their fitted (smoothed) counterparts, computed from equation (3.19) as $\sigma^2_{implied,t} - \varepsilon_t$.

The model seems to fit the data well. Regressing the squared returns on a constant and the fitted squared returns yields an R-squared of 0.11. (This rather low value is attributed to the noisiness of the squared returns.) The observed and fitted BS implied variances are virtually indistinguishable in the graph. A similar regression yields an R^2 of 0.99634, which is very close to the perfect-fit value 1.

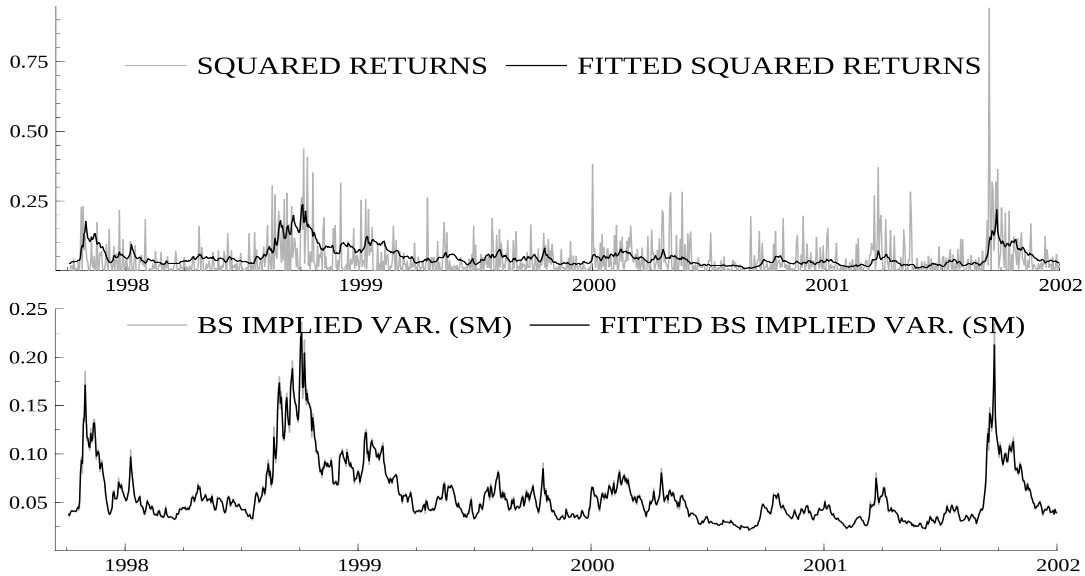


Figure 4.5: Observed and fitted squared returns and SM BS implied variances.

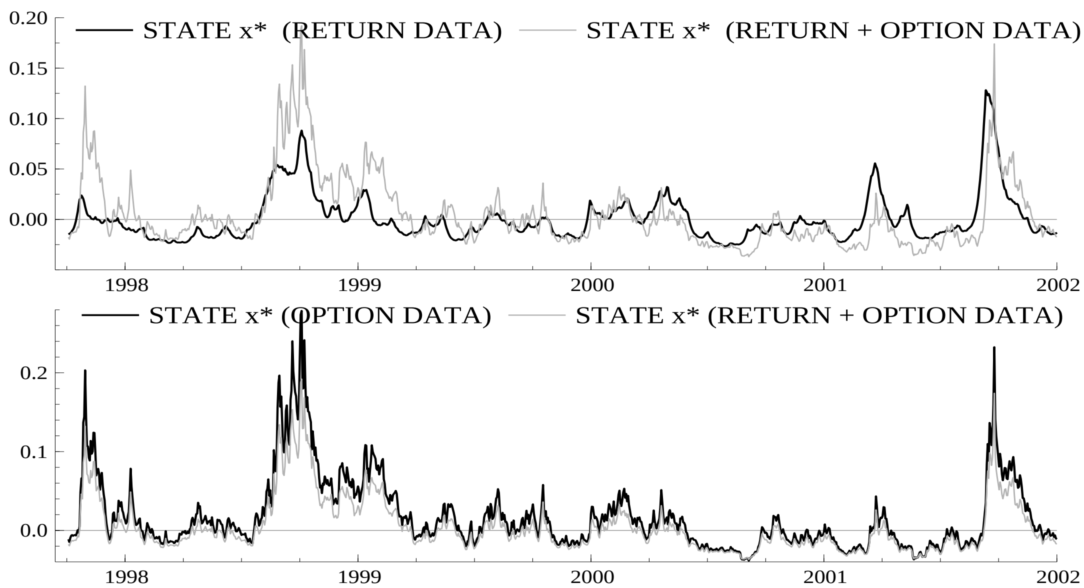


Figure 4.6: Upper plot: Smoothed states $x^*_{t|T} = x_{t|T} - \theta$ obtained from using the return data only, and the joint data. Lower: Smoothed states obtained from the SM option data (under the restriction that $\theta = \hat{\theta}_{SqR}$), and the joint data.

Option data dominates
Comparing the different smoothed states $x^*_{t|T} = x_{t|T} - \theta$ obtained from state space estimation using the three types of data reveals that the option data *dominates* the estimation results. The upper plot in figure 4.6 shows the state based on return data only, together with the state obtained from the joint data. The states deviate much from time to time. The lower graph plots the state extracted from the SM option data, together with the joint-data state. These states track each other much better. As was also apparent from the Monte Carlo results, the information in the option data dominates the estimation results when both return and option data are used for estimation. Again, squared returns contain far less precise information, and much less pronounced structure than option data.

4.3 Compensation for volatility risk and straddle returns

What do the estimation results imply for the FTSE100-index volatility risk premium? Consider a general, path-independent European-style derivative F. From chapter II, section 5.2, in the 1-factor OU SV case, the spot return an investor is expected to earn on this derivative equals

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = \left[r_t + \beta_{F,S,t}\left(\sigma_t \gamma_{S,t}\right) + \beta_{F,x,t}\left(\sigma\gamma\right)\right], \tag{4.2}$$

in which $\beta_{F,S}$ and $\beta_{F,x}$ are the derivative's stock and volatility beta respectively, $\sigma_t \gamma_{S,t}$ is the stock risk premium, and $\sigma\gamma$ is the volatility risk premium.

The joint return - SM option data yields an estimate of -0.967 for the market price of volatility risk γ. As γ differs significantly from zero, this suggests that investing in FTSE100-index derivatives yields a (negative) premium for bearing the volatility risk in the index. Market volatility risk thus seems *systematic*. The volatility risk premium is *negative*, which agrees with the consumption-insurance or hedging motive. Investors thus *pay* for the uncertainty in the FTSE100-index volatility. Our findings are in line with the empirical evidence that has recently appeared in the literature; see section 6.3.1 of chapter II.

Volatility risk premium: -15% per annum
Our results imply a volatility risk premium of $\sigma\gamma$ =(0.151)(-0.967) = -14.6% per annum for the FTSE100 index over 1997-2001. The long-run FTSE100-index volatility level equals 21%. To recall: Pan's (2002) results on the Heston model imply an average volatility risk premium of -10.4% for the S&P500 index over 1989-1996, with an associated long-run S&P500 volatility of 11.7%. Jones's (2003) Heston-model results imply an average S&P100 volatility risk premium of -27% over 1988 - 2000, with associated long-run S&P100 volatility of 18.7%.

Straddle return: -174% per annum
What do our results imply for the expected return to be earned on a short-maturity ATM FTSE100-index straddle? Such a straddle is virtually spot delta neutral (and hence $\beta_{Str,S,t} = (S_t / Str_t)\Delta_{Str,t} \approx 0$), such that its expected return equals

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dStr_t}{Str_t} \mid \mathcal{F}_t\right] \approx \left[r_t + \beta_{Str,x,t}\left(\sigma\gamma\right)\right] dt \tag{4.3}$$

by approximation, in the 1-factor OU SV case. To compute this return, an approximation of the *straddle volatility beta* $\beta_{Str,x,t} = (1/Str_t)\mathcal{V}_{Str,t}$ is needed.

Using the SM ATM call option series and its associated data on $\{C_t, \bar{r}_t, \tau_t, K_t, F_{t,T}\}$, we exploit the put-call parity $P_t = C_t + \exp(-\bar{r}_t \tau_t)(K_t - F_{t,T})$ to first construct a virtual short-maturity ATM *straddle price* series, by $Str_t = C_t + P_t$. As $\sigma_t^2 = x_t$, the *straddle vega* equals

$$\mathcal{V}_{Str,t} = \frac{\partial Str_t}{\partial x_t} = \frac{\partial Str_t}{\partial \sigma_t^2} = 2\mathcal{V}_{C,t} = 2\frac{\partial C_t}{\partial \sigma_t^2} = 2\frac{\partial C_t}{\partial \sigma_t}\frac{\partial \sigma_t}{\partial \sigma_t^2} = \frac{1}{\sigma_t}\frac{\partial C_t}{\partial \sigma_t}. \tag{4.4}$$

Our model implies that $\partial C_t / \partial \sigma_t$ can only be obtained by simulation. However, as Hull (2003, p.318) points out, the vega (which Hull defines as $\partial C_t / \partial \sigma_t$) of an SV model is very similar in magnitude to the Black-Scholes vega (which is defined as $\partial BS_t / \partial \sigma$). As such, we approximate $\partial C_t / \partial \sigma_t$ in (4.4) by the BS vega,

$$\frac{\partial C_t}{\partial \sigma_t} \approx \exp(-\bar{r}_t \tau_t) F_{t,T} \sqrt{\tau_t}\, \phi(d_{1t}), \tag{4.5}$$

in which we substitute the smoothed volatility $\sigma_{t|T}$ (obtained from the Kalman smoother) in the expression for d_{1t}. [33] We next compute $\beta_{Str,x,t}$ as

$$\beta_{Str,x,t} = \frac{\mathcal{V}_{Str,t}}{Str_t} \approx \frac{\exp(-\bar{r}_t \tau_t) F_{t,T} \sqrt{\tau_t}\, \phi(d_{1t})}{\sigma_{t|T}\, Str_t}. \tag{4.6}$$

By subsequently computing $r_t + \beta_{Str,x,t}(\sigma\gamma)$ for each t, a time series of (approximated) expected straddle returns (per annum) is obtained, which we average.

This procedure implies an average expected SM ATM FTSE100-index straddle return of −174% per annum. On a monthly (resp. weekly) basis, this return equals −14.5% (−3.34%). (We multiply the annual statistic by 1/12, resp. 1/52).

This result is in close correspondence to what other researchers have found, regarding *empirical* (i.e., realized) straddle returns. Driessen and Maenhout (2003) use the same FTSE100-index dataset as we do, though their results concern March 1992 - Dec 2001. They find an average monthly empirical return of −13.1% on short-maturity ATM FTSE100-index straddles. Coval and Shumway (2001) find an average weekly empirical return of −3% on delta-neutral *S&P500-index* straddles (period 1990-1995).

Magnitude of straddle price, vega, and volatility beta

To get an idea of the magnitude of the other quantities: The average *straddle price* equals 366, the average *straddle vega* ($\mathcal{V}_{Str,t} = \partial Str_t / \partial \sigma_t^2$) equals 4161, and the average *straddle volatility beta* equals 12.3. The average $\partial Str_t / \partial \sigma_t$ equals 1592, which implies that a 1-percent increase in the stock volatility raises the straddle price with approximately 1592. As the average straddle price is 366, this confirms that ATM straddles are true *bets on volatility*. [34]

The message is clear: Although in terms of expected returns, it seems particularly valuable to write short-maturity at-the-money index straddles in practice, there is also considerable risk involved with this investment strategy.

[33] Bakshi and Kapadia (2003) follow a comparable practice with GARCH volatilities.

[34] See also section 6.2.1 of chapter II. Some more numbers: The average SM *call price* equals 182 for the sample Oct 97 – Dec 2001, the average SM *call delta* (approximated by the Black-Scholes delta) equals 0.51. The average *Black-Scholes (call and put) vega* ($\partial BS_t / \partial \sigma_t$) equals 796. The average *call stock beta* equals 17.4, and the average *call volatility beta* equals 12.4.

4.4 Diagnostic checking

This section considers diagnostic checking. We focus on the model estimated using the joint return – SM option data. (See section 2.2 of the previous chapter for commonly applied diagnostic checks in linear state space models.)

Figure 4.7 shows the *standardized prediction errors* or *innovations* in the squared returns and BS implied variances. A well-specified model requires these innovations to be white noise, with unit variance. Their standard deviations equal 0.993 and 1.000 respectively. Both series are clearly *heteroskedastic*. As evidenced by table 4.2, both series are significantly *autocorrelated* as well. E.g., the p-value associated with the Ljung-Box Q-statistic for testing the null hypothesis of no autocorrelation up to order 5 equals 0.000 for both series. The autocorrelation may be interpreted as a preliminary indication of neglected dynamics; i.e. the 1-factor SV assumption is possibly not realistic.

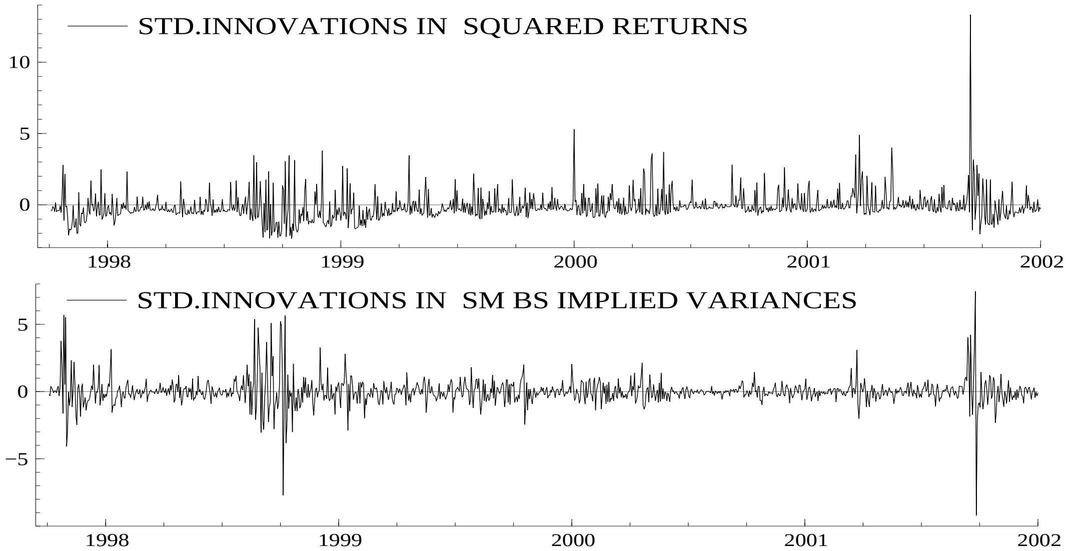


Figure 4.7: Standardized innovations in the squared returns and SM BS implied variances.

Table 4.2: Autocorrelations in the standardized innovations

Std. innovations in squared returns:				*Std.innovations in SM BS implied variances:*			
Order	AC	Q-Stat	Prob	Order	AC	Q-Stat	Prob
1	0.08	6.84	0.01	1	0.04	1.58	0.21
2	0.06	10.2	0.01	2	-0.15	25.9	0.00
3	0.13	28.9	0.00	3	-0.11	38.1	0.00
4	0.11	42.6	0.00	4	0.04	40.1	0.00
5	0.05	45.2	0.00	5	0.01	40.1	0.00

The table reports the autocorrelation coefficients (AC) up to order 5 and the Ljung-Box Q-statistics (Q-stat) for testing the null of no autocorrelation up to a certain order with associated p-values (Prob).

The *smoothed disturbances* of the state space model yield additional information regarding misspecification. They represent the best estimates of the errors given the data. Figure 4.8 shows the smoothed errors $\{\omega_t\}$ and $\{\varepsilon_t\}$ associated with the squared return and option equations. There is clear evidence of *conditional heteroskedasticity*. (The asymmetric pattern in the $\{\omega_t\}$-series is a logical consequence of the course of the squared returns over time; recall figure 2.3.)

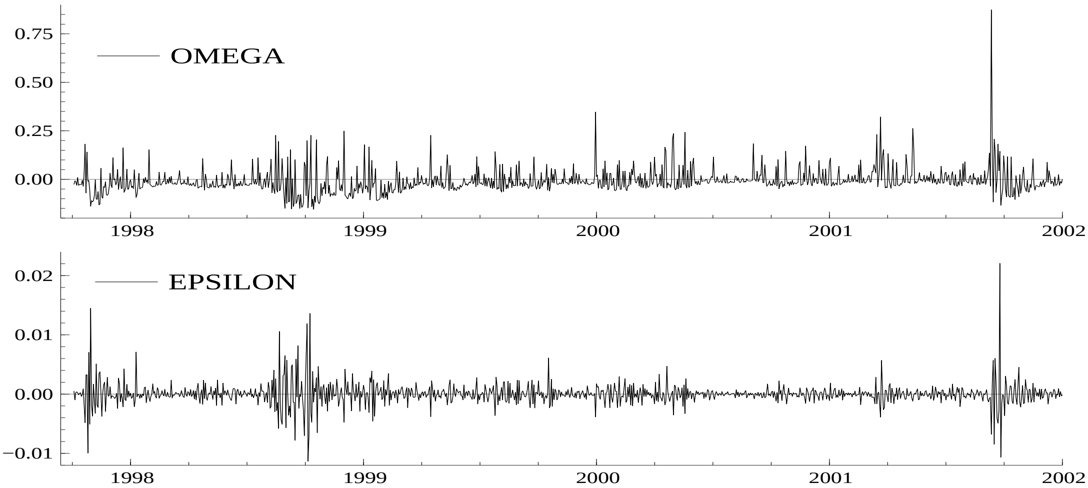


Figure 4.8: Smoothed $\{\omega_t\}$-series and $\{\varepsilon_t\}$-series.

Level-dependent volatility-of-volatility

Figure 4.9 shows the smoothed state $\{x_t^* = x_t - \theta\}$ and the daily volatility shocks $\{u_t\}$; recall that $\sigma_{t+\Delta t}^2 = \theta[1-\exp(-k\Delta t)] + \exp(-k\Delta t)\sigma_t^2 + u_{t+\Delta t}$. When the level of x_t^* is high (resp. low), the volatility shock u_t is generally large (small) in absolute value. As the state determines the stock variance ($\sigma_t^2 = x_t^* + \theta$), this means that volatility shocks are larger (smaller) if the current stock variance is large (small). This is clear evidence of *level-dependent volatility-of-volatility*. The 1-factor OU SV assumption does not allow for this effect, and is thus additionally misspecified in this sense.

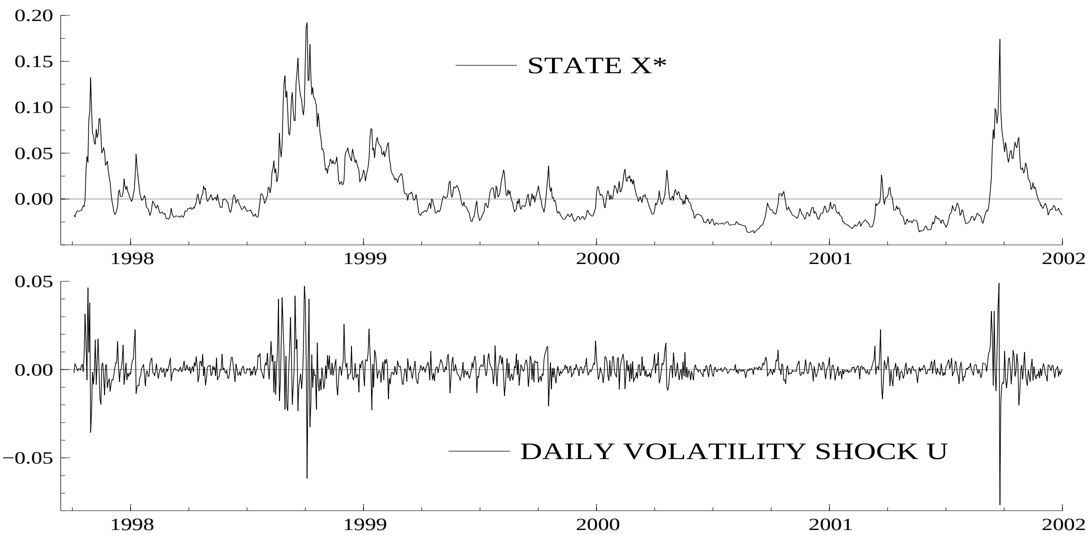


Figure 4.9: Smoothed state $\{x_t^*\}$, and daily volatility shocks $\{u_t\}$.

Option pricing errors and fit of the VTS

To what extent is the 1-factor OU SV model capable of fitting the observed option prices, and in particular, those of the longer-maturity options out of sample? As traders quote option prices in terms of BS implied volatilities, we express the *option pricing errors* in terms of deviations between "observed" and fitted Black-Scholes implied volatilities.

Based on the options measurement equation (for the general multifactor affine SV derivative pricing model), we compute fitted SM, MM and LM BS implied volatilities according to

$$\hat{\sigma}_{implied,it} = \sqrt{\tfrac{1}{\tau_{it}}\left[A_1(\tau_{it}) + \mathbf{B}_1(\tau_{it})'\boldsymbol{\theta}\right] + \tfrac{1}{\tau_{it}}\mathbf{B}_1(\tau_{it})'\mathbf{x}_t^*}\ , \qquad (4.7)$$

for $t = \Delta t,..,T\Delta t$ and $i = SM, MM, LM$, in which we substitute the parameter estimates and smoothed state $\mathbf{x}_t^*$. [35] The *option pricing errors* (expressed in percents BS implied volatility) are then given by

$$\text{error}_{it} = \sigma_{implied,it} - \hat{\sigma}_{implied,it}\ . \qquad (4.8)$$

The left panel of figure 4.10 shows the observed and fitted SM, MM and LM BS implied volatilities in one graph; the right panel shows the pricing errors.

1-factor SV: Out-of-sample overpricing of longer-dated options

The 1-factor OU SV model generally overfits the longer-maturity BS implied volatilities out of sample. Hence, the predicted option prices are typically larger than the observed market prices. The *overpricing* seems to get worse the longer the option maturity. Table 4.3 reports the mean and standard deviation of the three option pricing error series. The overpricing can be substantial.

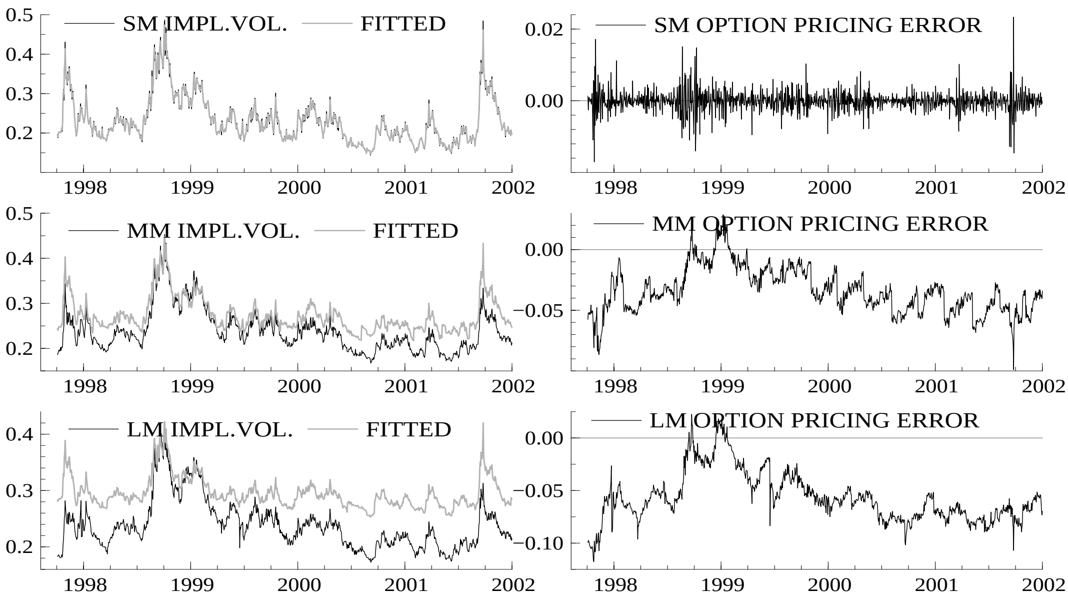


Figure 4.10: Fit of the VTS: *In-sample* fit of the SM BS implied volatility series and *out-of-sample* fit of the MM and LM series. Left: "observed" and fitted BS implied volatilities for each maturity. Right: their difference, the option pricing errors.

Table 4.3: Summary statistics option pricing errors (1-factor OU, return – SM option data)

	Error in SM impl.vol.	*Error in MM impl.vol.*	*Error in LM impl.vol.*
Mean	-0.0069%	-3.38%	-5.66%
Std.deviation	0.31%	1.89%	2.38%

[35] We avoid computationally intensive Monte-Carlo simulations here, to first obtain the model-implied call prices, and then transform them to BS implied volatilities. We use our easy-to-use approximation instead, for which the Monte Carlo results have shown that it is rather accurate.

Chernov and Ghysels (2000) reach a similar conclusion (irrespective of moneyness) when assessing the out-of-sample pricing performance of the Heston (1993) model. Pan (2002) also reports the severe out-of-sample overpricing of long-dated near-the-money options by the Heston model. She attributes this to the fact that her estimation results imply an *explosive* risk-neutral volatility process. Our results imply a *stationary* volatility process under $\mathbb{Q}$ (i.e., $\tilde{k} = 1.85$ and $\tilde{\theta} = 0.124$), as do the results of Chernov and Ghysels. Hence, the overpricing seems to occur irrespective of a stationary or explosive $\mathbb{Q}$-volatility process.

4.5 1-factor OU SV: Estimation results using return, SM, MM and LM option data jointly

The previous section presented evidence of misspecification of the 1-factor OU SV model estimated with return - SM option data. Before moving to a richer model structure, this section presents further evidence by providing estimation results based on the joint return, SM, MM and LM option data. As we will use all data in subsequent multifactor SV estimations, a comparison with these results later yields insight in the contribution of each additional SV factor.

The state space model now consists of the squared return equation (3.17), the equation for the OU factor (3.20), and three equations like (3.19), one for each option series. As the SM, MM and LM call option series are all ATMF, it seems reasonable to assume an equal parameter μ_V for each series. (Recall that the maturity effect is explicitly incorporated in the options measurement equation.) Given the Monte Carlo evidence, we restrict μ_V equal to zero (but see also footnote 36). We allow for *contemporaneous correlation* between the three option error series $\{\varepsilon_{it}\}$, but assume a zero-correlation at other points in time. *Cross-sectional heteroskedasticity* is allowed for as well; i.e., each option error series may have its own variance.

Table 4.4: Results **1-factor OU SV** using return, SM, MM and LM option data jointly

θ	**0.0475** (0.0408)		corr($\varepsilon_{SM,t}, \varepsilon_{MM,t}$)	**0.950** (0.0122)
k	**0.321** (0.0714)		corr($\varepsilon_{SM,t}, \varepsilon_{LM,t}$)	**0.895** (0.0482)
σ	**0.0639** (0.0063)		corr($\varepsilon_{MM,t}, \varepsilon_{LM,t}$)	**0.781** (0.0770)
γ	**0.00389** (0.206)		Loglikelihood	13,135
σ_ω	**0.0704** (0.0057)		Vol. Returns	21.8%
$\sigma_{\varepsilon,SM}$	**0.0193** (0.00124)		Vol-of-var.	0.0799
$\sigma_{\varepsilon,MM}$	**0.00728** (0.00089)		Persistence Half-life	0.9988 2.2 (years)
$\sigma_{\varepsilon,LM}$	**0.00394** (0.00071)		Std.dev. u_t	0.0040

Parameter estimates are in boldface; robust White (1982) QML standard errors are in parentheses. The state space model was estimated using all data (under the restriction that $\mu_V = 0$ for all option series), allowing for contemporaneous correlation between the three $\{\varepsilon_{it}\}$-series. Also reported are: the unconditional volatility of returns $\sqrt{\theta}$, the volatility-of-the-variance $\sqrt{\sigma^2/2k}$, the volatility persistence $\exp[-k\Delta t]$, the half-life of a volatility shock $\ln 2/k\Delta t$, and the QMLE of the std.dev. of the daily volatility shock u_t.

Table 4.4 reports the estimation results for Oct 1997 – Dec 2001. [36] Comparing the estimates to those obtained from using return – SM option data, the following stands out. The volatility persistence is estimated even larger, and particularly close to the random walk value of 1. The market price of volatility risk is estimated near zero. The contemporaneous correlation between the three option error series $\{\varepsilon_t\}$ is substantial (which seems to suggest misspecification). However, as the model is heavily misspecified, these estimates should not be taken too seriously.

A first indication of the model's incapability of adequately describing the full set of empirical data is that, if the model were correctly specified, roughly the same estimates for θ, k, σ, γ, σ_ω and $\sigma_{\varepsilon,SM}$ should have resulted as when using only return - SM option data. This is obviously not the case.

Figure 4.11 shows the in-sample fit of the model, i.e. the observed and fitted (smoothed) data. The LM option series is fitted best, which is also apparent from the estimate of $\sigma_{\varepsilon,LM}$, which is smallest among the option error standard deviations.

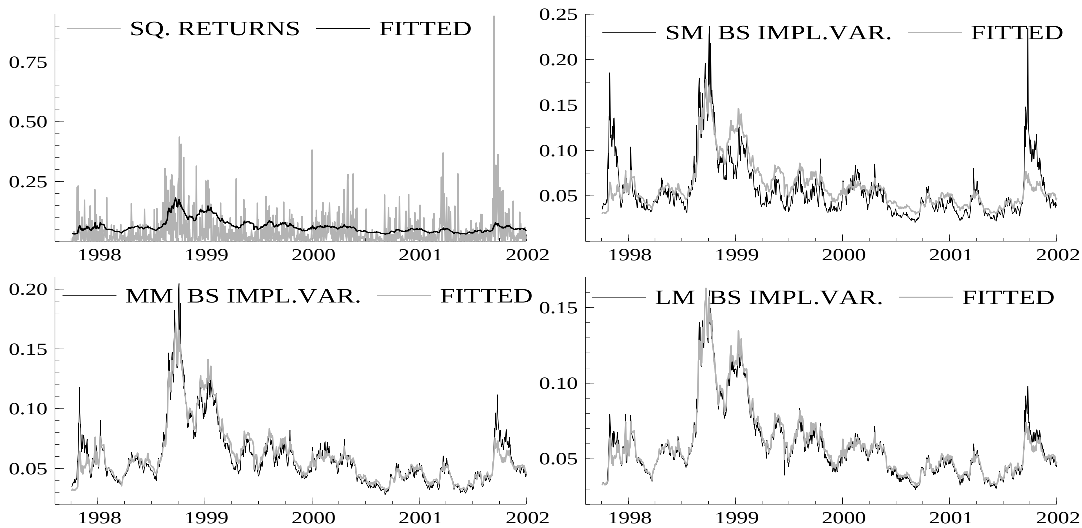


Figure 4.11: In-sample fit (**1-factor OU SV**): Observed and fitted data.

Figure 4.12 plots the standardized innovations in the squared returns, SM, MM and LM BS implied variances, table 4.5 provides summary statistics. A well-specified model requires these innovations to have mean zero, be serially and *mutually* uncorrelated, and be homoskedastic with unit variance. The misspecification is obvious. In particular, there is considerable autocorrelation in each series. This indicates that one SV factor is insufficient to describe the volatility dynamics present in the joint data sufficiently well.

[36] Leaving μ_v unrestricted yields the following results: $\mu_v = 0.021$, $\theta = 0.027$, $k = 0.318$, $\sigma = 0.0640$, $\gamma = 0.0252$, similar values for the correlations, and a quasi-loglikelihood of 13,181. As in the return - SM case, μ_v is estimated *positive* (and large), though on theoretical grounds it ought to be *negative* (and small; if the model were correctly specified). As before, μ_v seems to pick up part of the misspecification (recall footnote 31). Note that although k and σ are estimated similar in magnitude as when imposing $\mu_v = 0$, θ and γ are most affected by leaving μ_v free (which is as expected; recall the identification discussion). θ assumes a less plausible value in the latter case.

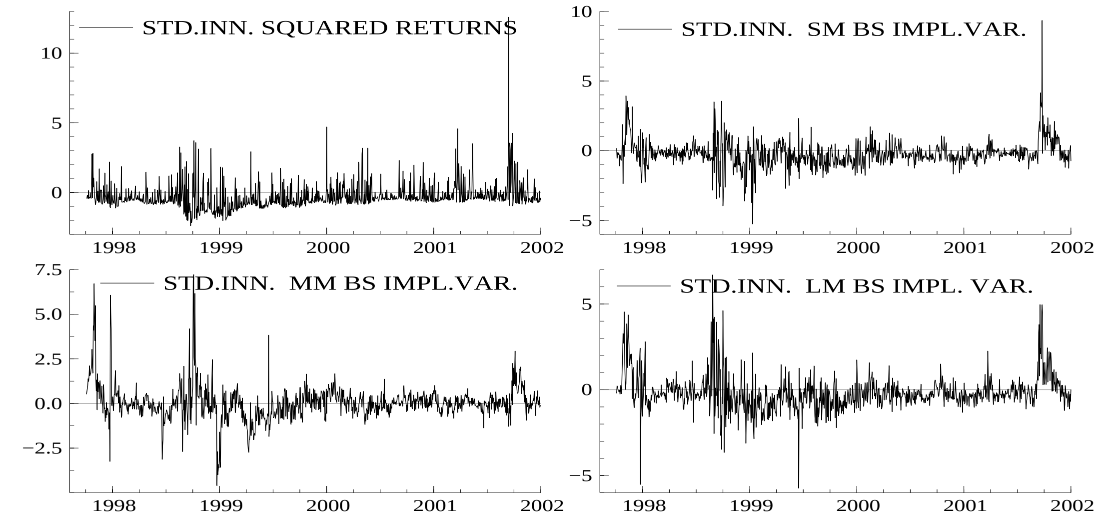


Figure 4.12: Standardized innovations (**1-factor OU SV**).

Table 4.5: Summary statistics standardized innovations (**1-factor OU SV**)

	Std.inn. sq. return	*Std.inn. SM series*	*Std.inn. MM series*	*Std.inn. LM series*
Mean	-0.28	-0.21	0.05	-0.17
Std.deviation	0.96	0.95	0.99	1.03
AC(1)	0.14	0.40	0.62	0.43
AC(2)	0.12	0.35	0.53	0.37
AC(3)	0.18	0.30	0.44	0.26
AC(4)	0.17	0.25	0.42	0.34
AC(5)	0.11	0.28	0.37	0.33
Cont.correlation matrix	1.00	0.09	0.05	0.34
		1.00	-0.02	-0.04
			1.00	0.03
				1.00

The table reports the mean, standard deviation, autocorrelation coefficients (AC) up to order 5, and the contemporaneous correlation matrix of the standardized innovations.

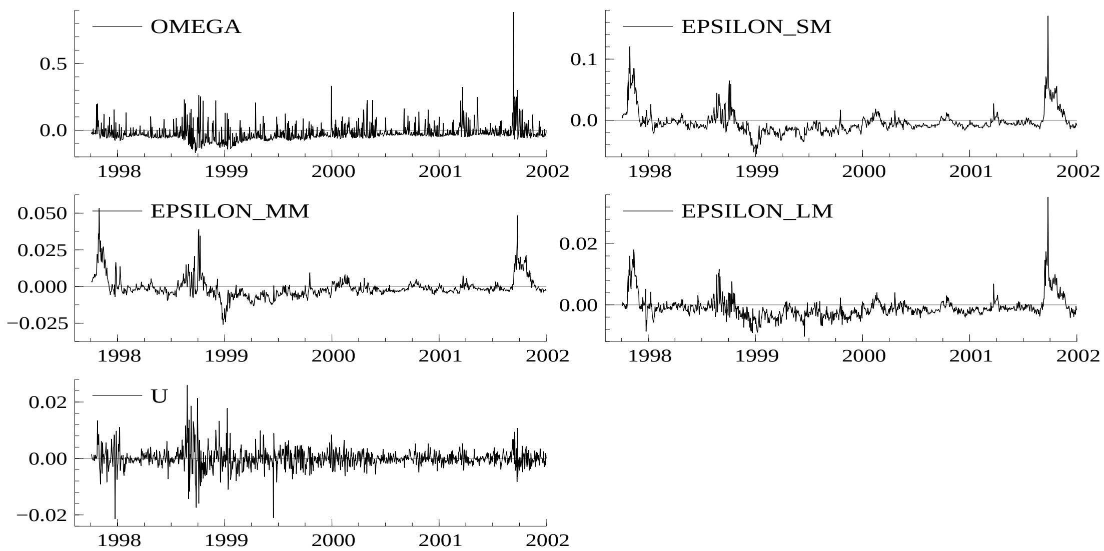


Figure 4.13: Smoothed disturbances of the state space model (**1-factor OU SV**).

Figure 4.13 shows the smoothed disturbances of the state space model. The misspecification is most apparent from the graphs of the three option error series, which ought to look like white noise approximately. They clearly do not.

In sample fit of the volatility term structure
How does the 1-factor OU SV model fit the ATMF volatility term structure in sample? [37] Following a similar practice as before, we compute fitted SM, MM and LM BS implied volatilities by (4.7), and option pricing errors by (4.8). Figure 4.14 shows the observed and fitted series, and the errors. The errors can be substantial. Table 4.6 reports their means and standard deviations (also for the absolute errors). On average, the 1-factor OU model predicts somewhat higher prices than those observed in the market.

The *short end* of the volatility term structure is generally fitted worst, and the *long end* best. We get back to this observation later in section 5.3.4.

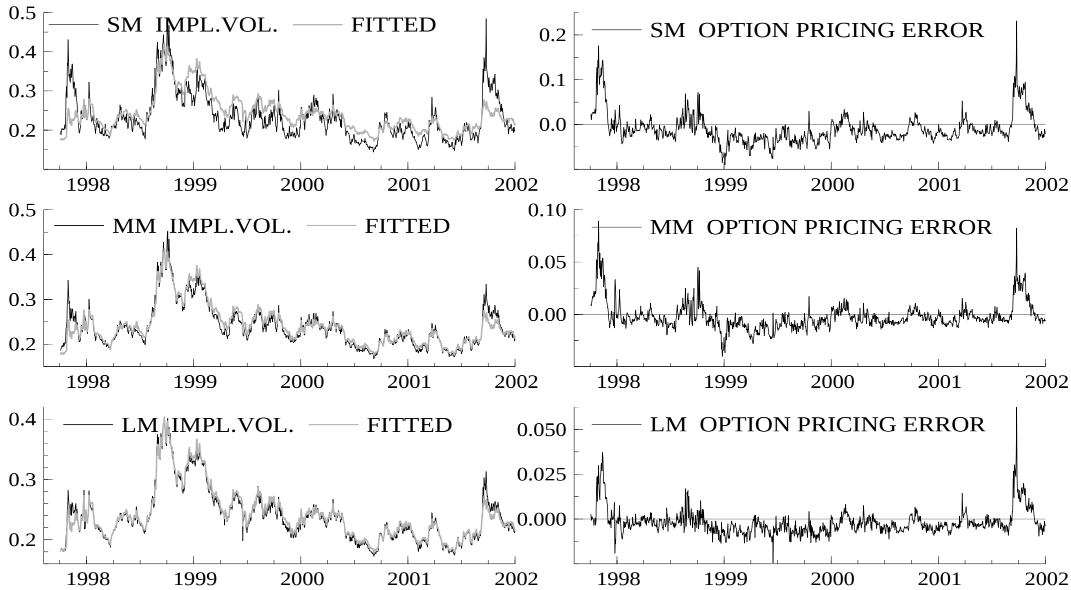


Figure 4.14: Fit (in-sample) of the VTS (**1-factor OU SV**). Left: observed and fitted BS implied volatilities for each maturity. Right: their difference, the option pricing errors.

Table 4.6: Summary statistics (absolute) option pricing errors (**1-factor OU SV**)

	Error SM	*Error MM*	*Error LM*	*|Error SM|*	*|Error MM|*	*|Error LM|*
Mean	-1.01%	-0.22%	-0.20%	2.62%	0.93%	0.53%
Std.dev.	3.33%	1.31%	0.70%	2.30%	0.95%	0.51%

We expect that fitting the Heston (1993) model to a similar dataset incorporating more than one option series would lead to similar conclusions as found here. To the best of our knowledge this has not been pursued so far in the literature. We pursue this extension in section 5.2 of the next chapter.

We conclude by noting that the findings in this section are similar in spirit to what de Jong (2000) finds in his empirical analysis of the *US bond market*. Fitting

[37] Obviously, figures 4.11 and 4.13 already provide a clue, but then in terms of less easy-to-interpret BS implied variances, instead of implied volatilities.

various 1-factor *affine* models of the *term structure of interest rates* (Duffie and Kan (1996)) to monthly US zero-coupon yields of four different maturities, de Jong concludes that these models are considerably misspecified. In particular, they fail to give a good fit of the yield curve at the short end, and do not describe the yield curve dynamics very well.

5. FTSE100-index data: Results for multifactor OU SV

The main defect of the 1-factor OU SV model seems its incapability of capturing most of the observed VTS movements. This section therefore extends to multiple OU SV factors. Section 5.1 discusses the associated state space model. Not all parameters can be identified such that certain restrictions need to be imposed prior to estimation. Section 5.2 presents estimation results for 2-factor OU SV. Two factors still appear insufficient. Section 5.3 extends to three OU factors, which appear to provide an adequate description of the dynamics present in the joint data. We interpret the factors in two ways, explain how each impacts on the prices of options and consider their associated risk premia. Specification tests next reveal how the 3-factor OU SV model can still be improved upon.

5.1 The model in case of multifactor OU SV

In case of multifactor OU SV, the SV specification reads $\sigma_t^2 = \mathbf{1}'\mathbf{x}_t$ (we impose $\delta_0 = 0$, $\boldsymbol{\delta} = \mathbf{1}$ for identification reasons), with

$$d\mathbf{x}_t = \mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t)dt + \boldsymbol{\Sigma} d\mathbf{W}_{x,t}, \qquad (\mathbb{P}) \qquad (5.1)$$

$$d\mathbf{x}_t = \tilde{\mathbf{K}}(\tilde{\boldsymbol{\theta}} - \mathbf{x}_t)dt + \boldsymbol{\Sigma} d\tilde{\mathbf{W}}_{x,t}, \quad \tilde{\mathbf{K}} = \mathbf{K}_d, \quad \tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \mathbf{K}_d^{-1}\boldsymbol{\Sigma\Gamma}\mathbf{1}, \qquad (\mathbb{Q}) \qquad (5.2)$$

in which $\boldsymbol{\Gamma} = \text{diag}[\gamma_1,..,\gamma_n]$ contains the market prices risk associated with the SV factors. The stock variance is driven by a sum of n OU factors, which are independent if $(nxn)\,\boldsymbol{\Sigma}$ is diagonal. We refer to section 13 of appendix B for the implied statistical properties of the factors and volatility.

State space model
Using all data for estimation, extracting information from the stock returns occurs via

$$\frac{1}{\Delta t}(r_t - \hat{\mu}_{t-\Delta t}\Delta t)^2 = \mathbf{1}'\boldsymbol{\theta} + \mathbf{1}'\mathbf{x}^*_{t-\Delta t} + \omega_t, \quad \omega_t \sim (0, \sigma^2_\omega), \qquad (5.3)$$

and from the option series via

$$\sigma^2_{implied,it} = \mu_v + \frac{1}{\tau_{it}}\left[A_1(\tau_{it}) + \mathbf{B}_1(\tau_{it})'\boldsymbol{\theta}\right] + \frac{\mathbf{B}_1(\tau_{it})'}{\tau_{it}}\mathbf{x}^*_t + \varepsilon_{it},\ \varepsilon_{it} \sim (0, \sigma^2_{\varepsilon i}), \quad (5.4)$$

for $t = \Delta t,..,T\Delta t$ and $i = SM, MM, LM$. The $n-$factor OU SV assumption yields the closed-form expressions (13.15)-(13.16) in appendix B for the functions $(1x1)\,A_1(.)$ and $(nx1)\,\mathbf{B}_1(.)$. These functions depend on the risk-neutral parameters. Rewritten in terms of the $\mathbb{P}$-parameters, they become

$$A_1(\tau_t) = \mathbf{1}'\left[\tau_t\,\mathbf{I}_n - \mathbf{D}(\tau_t)\right]\left(\boldsymbol{\theta} - \mathbf{K}_d^{-1}\boldsymbol{\Sigma\Gamma}\mathbf{1}\right) + \tfrac{1}{2}\mathbf{1}'\left[\mathbf{N}(\tau_t) \odot \boldsymbol{\Sigma\Sigma}'\right]\mathbf{1} \qquad (5.5)$$
$$\mathbf{B}_1(\tau_t) = \mathbf{D}(\tau_t)\mathbf{1},$$

in which $\mathbf{D}(\tau_t) = \text{diag}[d_1(\tau_t),..,d_n(\tau_t)]$ with $d_i(\tau_t) = [1-\exp(-k_i\tau_t)]/k_i$; $i = 1,..,n$, and the matrix $(nxn)\mathbf{N}(\tau_t)$ has $ij-$th element equal to

$$N_{ij}(\tau_t) = \tfrac{1}{k_i k_j}\left[\tau_t - d_i(\tau_t) - d_j(\tau_t) + \tfrac{1-\exp[-(k_i+k_j)\tau_t]}{k_i+k_j}\right]. \qquad (5.6)$$

Using these results, the options measurement equation can be rewritten as

$$\sigma^2_{implied,it} = \mu_v + \mathbf{1}'\boldsymbol{\theta} + \mathbf{1}'\left(\frac{\mathbf{D}(\tau_{it})}{\tau_{it}} - \mathbf{I}_n\right)\mathbf{K}_d^{-1}\boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathbf{1} + \frac{1}{2\tau_{it}}\mathbf{1}'\left[\mathbf{N}(\tau_{it}) \odot \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right]\mathbf{1}$$
$$+ \mathbf{1}'\frac{\mathbf{D}(\tau_{it})}{\tau_{it}}\mathbf{x}_t^* + \varepsilon_{it}; \qquad \varepsilon_{it} \sim (0, \sigma^2_{\varepsilon i}). \qquad (5.7)$$

As before, we take μ_v the same for each option series as they are all ATMF. Given the Monte-Carlo evidence, we next set μ_v equal to 0. We allow for possible contemporaneous correlation between the option error series $\{\varepsilon_{SM,t}\}$, $\{\varepsilon_{MM,t}\}$ and $\{\varepsilon_{LM,t}\}$, and allow for cross-sectional heteroskedasticity. In discrete time, the factors evolve as

$$\mathbf{x}^*_{t+\Delta t} = \exp[-\mathbf{K}_d\Delta t]\mathbf{x}^*_t + \mathbf{u}_{t+\Delta t}, \; \mathbf{u}_{t+\Delta t} \sim \mathcal{N}[\mathbf{0}; \mathbf{G}(\Delta t) \odot \boldsymbol{\Sigma}\boldsymbol{\Sigma}'] \sim \mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t, \quad (5.8)$$

in which $[\mathbf{G}(\Delta t)]_{ij} = (1-\exp[-(k_i+k_j)\Delta t])/(k_i+k_j)$. The distribution of the *daily volatility-factor shocks* $\{\mathbf{u}_t\}$ is both conditionally and unconditionally Gaussian, as level-dependent volatility-of-volatility is not modeled by $n-$factor OU SV. The system matrices of the state space model are given in the appendix.

Parameter identification
Not all parameters of the n-factor OU SV model are identified. The parameters $\theta_1,..,\theta_n$ cannot separately be identified; only their sum is identifiable. We therefore choose to leave θ_1 unrestricted, and restrict $\theta_2 = .. = \theta_n = 0$ prior to estimation. The parameters σ_{ij}; $i,j = 1,..,n$ appearing in matrix $\boldsymbol{\Sigma}$ cannot all be identified either. However, if $\boldsymbol{\Sigma}$ is *diagonal* (i.e., *independent* factors), or if $\boldsymbol{\Sigma}$ is *lower* (or upper) *diagonal* (which means *correlated* factors), then all parameters can be identified. Both cases are considered in the estimations below.

5.2 Estimation results for 2-factor OU SV

This section presents estimation results for 2-factor OU SV. As dynamic misspecification still remains, we keep the discussion short. We consider two sets of a priori imposed restrictions to deal with the identification problem. *Restrictions (a)* considers independent factors, with zero mutual option error correlation. *Restrictions (b)* allows for factor and contemporaneous option error correlation.

Table 5.1 reports the results. Allowing for two OU factors instead of one raises the quasi-loglikelihood from 13,135 to more than 14,300. Allowing for correlated rather than independent factors does not seem to improve the fit dramatically. The factor correlation is estimated at 0.166 and seems significant (as σ_{21} differs significantly from zero). The option error series $\{\varepsilon_{it}\}$ are not significantly contemporaneously correlated. The unconditional stock return volatility, $\sqrt{\mathbf{1}'\boldsymbol{\theta}}$, is estimated near the 20.3% obtained from averaging the squared returns. The two volatility factors differ in their characteristics. The first factor has similar properties as the factor distilled from the joint data in the 1-factor OU case. Its persistence $\exp(-k_1\Delta t)$ is close to the random walk value of 1, with a half-life $\ln 2/k_1\Delta t$ of about 1.45 years. Shocks in this factor die out slowly. The second factor shows much quicker mean reversion with a half-life of about a month. The market prices of factor risk do not differ significantly from zero (neither jointly).

Table 5.1: Estimation results **2-factor OU SV model** using all data

	Restrictions (a)	*Restrictions (b)*		*Restrictions (a)*	*Restrictions (b)*
θ_1	**0.0431** (0.0342)	**0.0408** (0.0386)	corr($\varepsilon_{SM,t}, \varepsilon_{MM,t}$)	0	**0.548** (0.342)
k_1	**0.516** (0.080)	**0.445** (0.076)	corr($\varepsilon_{SM,t}, \varepsilon_{LM,t}$)	0	**-0.089** (0.548)
k_2	**6.72** (0.731)	**8.71** (1.37)	corr($\varepsilon_{MM,t}, \varepsilon_{LM,t}$)	0	**-0.516** (0.658)
σ_{11}	**0.0602** (0.0046)	**0.0669** (0.0058)	Vol. returns	20.8%	20.2%
σ_{12}	0	0	Vol-of-var.	0.0804	0.0858
σ_{21}	0	**0.0612** (0.0216)	Std.dev. x_{1t}	0.059	0.071
σ_{22}	**0.199** (0.018)	**0.146** (0.035)	Std.dev. x_{2t}	0.054	0.038
γ_1	**0.116** (0.258)	**0.090** (0.241)	corr(x_{1t}, x_{2t})	0	0.166
γ_2	**-0.330** (0.467)	**-0.661** (0.595)	Persistence x_1	0.9980	0.9983
σ_ω	**0.0671** (0.0057)	**0.0672** (0.0056)	Persistence x_2	0.9745	0.9671
$\sigma_{\varepsilon,SM}$	**1.59*10^{-5}** (2.10*10^{-5})	**0.00306** (0.00158)	Half-life x_1	1.3 years	1.6 years
$\sigma_{\varepsilon,MM}$	**0.00201** (0.00016)	**0.00205** (0.00069)	Half-life x_2	27 days	21 days
$\sigma_{\varepsilon,LM}$	**0.00206** (0.00027)	**0.00165** (0.00038)	Std.dev. u_{1t}	0.0037	0.0041
Loglik.	14,305	14,363	Std.dev. u_{2t}	0.0122	0.0097

The table reports (restricted) parameter estimates (in boldface) with robust White (1982) QML standard errors in parentheses, resulting from estimating the state space model associated with the **2-factor OU SV** assumption using the combination of return, SM, MM and LM option data, under two sets of restrictions (see main text), together with the QMLEs of some other quantities of interest. For identification reasons, we restrict $\theta_2 = 0$ and $\sigma_{12} = 0$ such that $\mathbf{\Sigma}$ is lower diagonal.

Fit of the volatility term structure

For further investigation and diagnostic checking, we concentrate on the least restricted model (Restrictions (b)).

Figure 5.1 shows the observed and fitted SM, MM and LM Black-Scholes implied volatilities, and the corresponding option pricing errors. Comparing this figure to figure 4.14 (1-factor OU) reveals that although the error graphs have "improved", they still do not seem to be randomly distributed over time.

Table 5.2 reports the mean and standard deviation of the (absolute value of the) option pricing errors. Comparing this table to table 4.6 (1-factor OU), reveals a substantial increase in fit obtained by allowing for two volatility factors. The biggest improvement is obtained for the SM option series. The pricing errors concentrate around zero now, irrespective of maturity. The average fit is similar for all series, though the LM series is still fitted somewhat best.

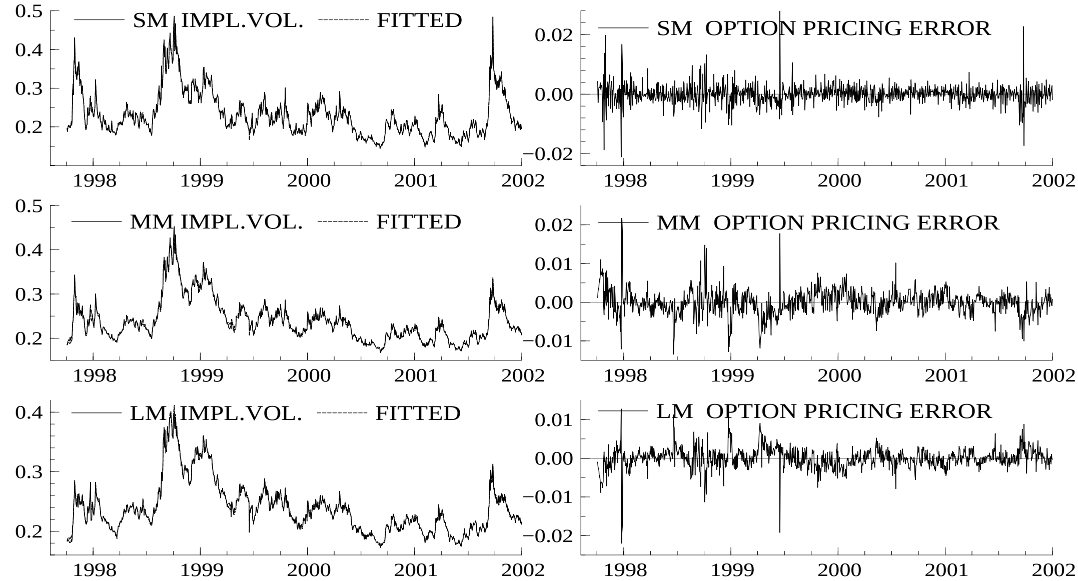


Figure 5.1: Fit of the VTS (**2-factor OU SV, Restrictions (b)**). Left: observed and fitted BS implied volatilities for each maturity. Right: their difference, the option pricing errors.

Table 5.2: Summary stat. (absolute) option pricing errors (**2-factor OU SV, Restr.(b)**)

	Error SM	*Error MM*	*Error LM*	*|Error SM|*	*|Error MM|*	*|Error LM|*
Mean	0.00%	0.00%	0.00%	0.23%	0.24%	0.19%
Std.dev.	0.34%	0.33%	0.28%	0.25%	0.24%	0.20%

Specification tests

Figure 5.2 shows the smoothed factors (in deviation from their mean) $x_i^* = x_i - \theta_i$, and the daily volatility-factor shocks u_1 and u_2. Level-dependent volatility-of-volatility is present in both series, a feature not accounted for.

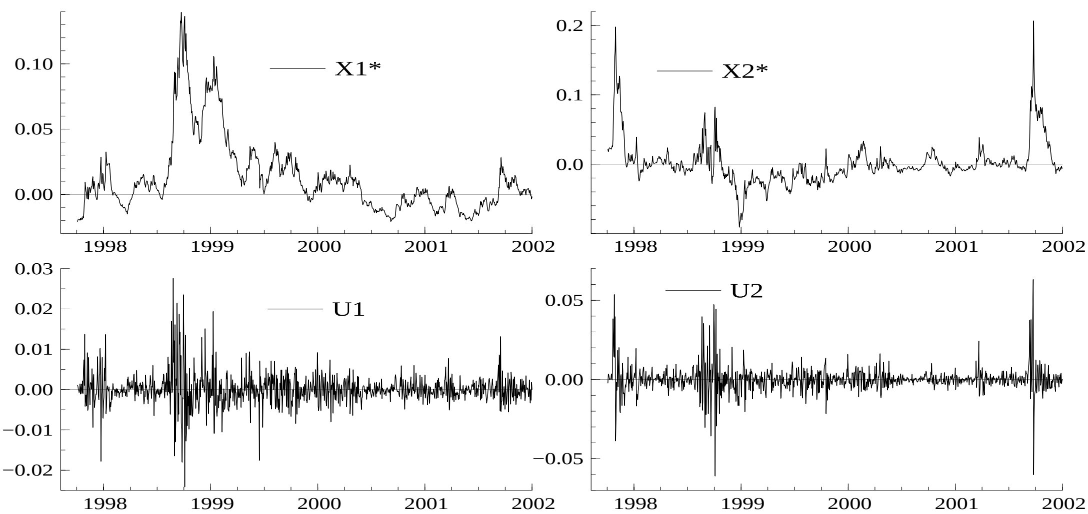


Figure 5.2: Smoothed factors in deviation from their mean (x_{1t}^* and x_{2t}^*), and smoothed state equation errors u_{1t} and u_{2t} (**2-factor OU SV, Restrictions (b)**).

Table 5.3 provides summary statistics for the standardized innovations. Recalling table 4.5 (1-factor OU), the 2-factor assumption evidently provides a much better description of the data. Especially prominent is the dramatic reduction in autocorrelation in three of the four series. The dynamics in the MM option series are however still not well captured by this model. [38] Figure 5.3 shows the smoothed disturbances. A comparison with figure 4.13 confirms the improved data description. But again, these pictures seem to indicate a lack of dynamics.

Table 5.3: Summary statistics standardized innovations (**2-factor OU SV, Restr. (b)**)

	Std.inn. sq. return	*Std.inn. SM series*	*Std.inn. MM series*	*Std.inn. LM series*
Mean	-0.19	-0.01	-0.01	0.01
Std.deviation	0.98	1.00	1.01	0.99
AC(1)	0.09	0.08	0.40	0.05
AC(2)	0.06	-0.02	0.23	0.02
AC(3)	0.14	-0.03	0.20	-0.10
AC(4)	0.11	-0.02	0.19	0.07
AC(5)	0.05	0.01	0.21	0.02
Cont.correlation matrix	1.00	0.07	0.00	0.22
		1.00	0.00	0.02
			1.00	0.02
				1.00

The table reports the mean, standard deviation, the autocorrelation coefficients (AC) up to order 5, and the contemporaneous correlation matrix of the standardized innovations.

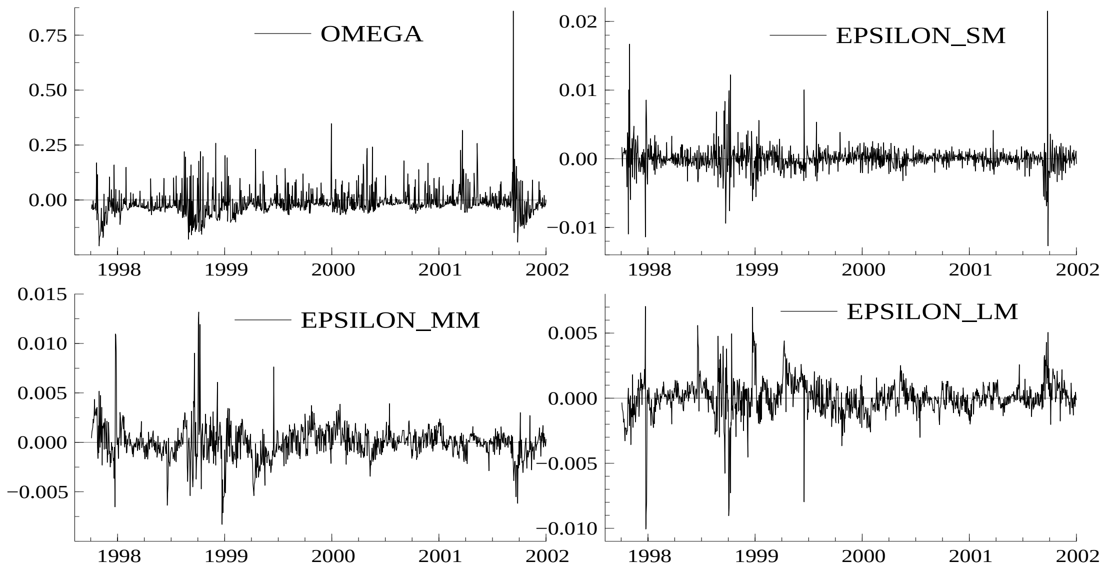


Figure 5.3: Smoothed disturbances of state space model (**2-factor OU SV, Restr.(b)**).

5.3 Estimation results for 3-factor OU SV

This section extends to three OU factors. As not all parameters can be identified we consider two sets of a priori imposed restrictions, which are similar as in the 2-factor case. *Restrictions (a)* assumes independent factors and zero correlation between the option errors $\{\varepsilon_{it}\}$. *Restrictions (b)* assumes correlated factors and contemporaneous option error correlation.

[38] Based on the Ljung-Box Q-statistic, we reject the null hypothesis of zero autocorrelation up to order 5 at 5% significance for each innovation series, except for the SM series.

Table 5.4 presents the results. Allowing for three instead of two OU factors increases the quasi-loglikelihood from more than 14,300 to more than 14,500. Permitting for factor and option error correlation does not seem to further raise the quasi-loglikelihood very much. It does however introduce much additional uncertainty in most estimates, as evidenced by the larger standard errors (relative to the estimates). There are obvious opportunity costs associated with this larger flexibility. We feel most confident with the results of Restrictions (a), the independent-factors case, such that in the remainder we focus on those. [39]

Table 5.4: Estimation results **3-factor OU SV model** using all data

	Restrictions (a)	*Restrictions (b)*		*Restrictions (a)*	*Restrictions (b)*
θ_1	**0.0392** (0.0417)	**0.0412** (0.0990)	corr($\varepsilon_{SM,t}, \varepsilon_{MM,t}$)	0	**0.59** (0.47)
k_1	**0.023** (0.029)	**0.167** (0.388)	corr($\varepsilon_{SM,t}, \varepsilon_{LM,t}$)	0	**0.14** (0.50)
k_2	**3.32** (0.29)	**1.22** (3.40)	corr($\varepsilon_{MM,t}, \varepsilon_{LM,t}$)	0	**-0.16** (0.84)
k_3	**15.7** (2.5)	**13.6** (2.98)	Vol. returns	19.81%	20.30%
σ_{11}	**0.0402** (0.0036)	**0.171** (0.626)	Std.dev. x_{1t}	0.189	0.295
σ_{21}	0	**-0.266** (0.679)	Std.dev. x_{2t}	0.057	0.181
σ_{22}	**0.147** (0.013)	**0.097** (0.132)	Std.dev. x_{3t}	0.031	0.041
σ_{31}	0	**0.129** (0.149)	corr(x_{1t}, x_{2t})	0	-0.61
σ_{32}	0	**0.047** (0.185)	corr(x_{1t}, x_{3t})	0	0.13
σ_{33}	**0.176** (0.035)	**0.161** (0.037)	corr(x_{2t}, x_{3t})	0	-0.27
γ_1	**-0.034** (0.177)	**-0.165** (0.231)	Persistence x_1	0.9999	0.9994
γ_2	**0.266** (0.481)	**0.243** (0.475)	Persistence x_2	0.9873	0.9953
γ_3	**-1.47** (0.77)	**-1.29** (0.73)	Persistence x_3	0.9413	0.9490
$\sigma_{\varepsilon,SM}$	**0.00096** (0.00090)	**0.00274** (0.00151)	Half-life x_1	31 years	4.2 years
$\sigma_{\varepsilon,MM}$	**0.00140** (0.00013)	**0.00129** (0.00095)	Half-life x_2	54 days	147 days
$\sigma_{\varepsilon,LM}$	**0.00182** (0.00024)	**0.00137** (0.00043)	Half-life x_3	11 days	13 days
σ_{ω}	**0.0669** (0.0059)	**0.0673** (0.0059)	Std.dev. u_{1t}	0.0025	0.0106
Loglik.h.	14,500	14,560	Std.dev. u_{2t}	0.0090	0.0175
Vol-of-var.	0.2000	0.2355	Std.dev. u_{3t}	0.0106	0.0128

The table reports (restricted) parameter estimates (boldface) with robust White (1982) QML standard errors in parentheses, resulting from estimating the state space model associated with the **3-factor OU SV** assumption using the combination of return, SM, MM and LM option data under two sets of restrictions (see main text), together with the QMLEs of some other quantities of interest. For identification reasons, we restrict $\theta_2 = \theta_3 = 0$ and $\sigma_{12} = \sigma_{13} = \sigma_{23} = 0$ such that $\mathbf{\Sigma}$ is lower diag.

[39] The outcome and interpretation of the diagnostic checks only marginally differ for both cases. Moreover, the evidence in Cont and Fonseca (2002) seems to support this choice further.

5.3.1 Decomposing stock volatility: long-memory, medium-term and short-term volatility trends

The upper graph of figure 5.4 plots the smoothed stock volatilities obtained from the 3-factor OU model. The middle graph plots this series again, but together with the volatilities obtained from the 1-factor OU model (estimated using return - SM option data (recall figure 4.1)). The lower graph shows that their difference can be substantial. A close inspection reveals that mainly in times of sudden rises in the volatility (fall 1997 and September 11, 2001), the 3-factor volatilities respond quicker than the 1-factor volatilities, which makes sense intuitively.

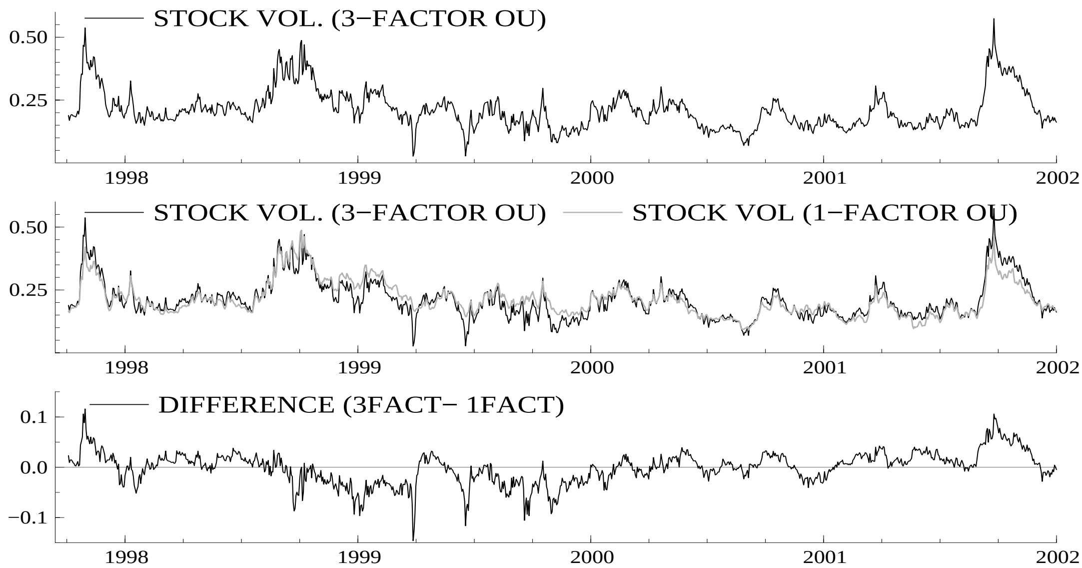


Figure 5.4: Upper graph: Stock volatilities obtained from 3-factor OU SV, Restrictions (a). The middle graph plots these again, together with the volatilities obtained from 1-factor OU estimated using return – SM option data. Their difference is plotted in the lower graph.

The 3-factor model decomposes the hidden volatility evolution into three different dynamic components. Table 5.4 shows that the volatility factors distilled from the data differ greatly in their features. Factor x_1 is extremely persistent: it has very long memory and behaves almost like a random walk. Shocks to this factor have close-to permanent effects. (This explains why the standard deviation of x_1 is much larger than for the other factors.) Factor x_2 reverts much faster to its mean, with a half-life of about 2.5 months. Factor x_3 is quickest mean reverting. It takes about 11 days for a shock in this factor to lose half its impact.

The estimates indicate that x_1 may be interpreted as determining the *long-term volatility trend*. The long-term trend averages at 20%. Factor x_2 seems more associated with *medium-term volatility movements*; factor x_3 determines the very *short-term volatility fluctuations*. Both average at zero (due to the imposed identification restrictions).

Figure 5.5 plots the smoothed factors $\{x_{it}^*\}$ and the daily shocks to these factors, $\{u_{it}\}$. Note the scales on the vertical axes, and in particular the graph of $\{u_{3t}\}$. Factor x_3 is clearly responsible for large volatility changes in short time periods. This is confirmed by the fact that its instantaneous volatility, σ_{33}, is largest, and the fact that the shocks u_{3t} to this factor have largest variance.

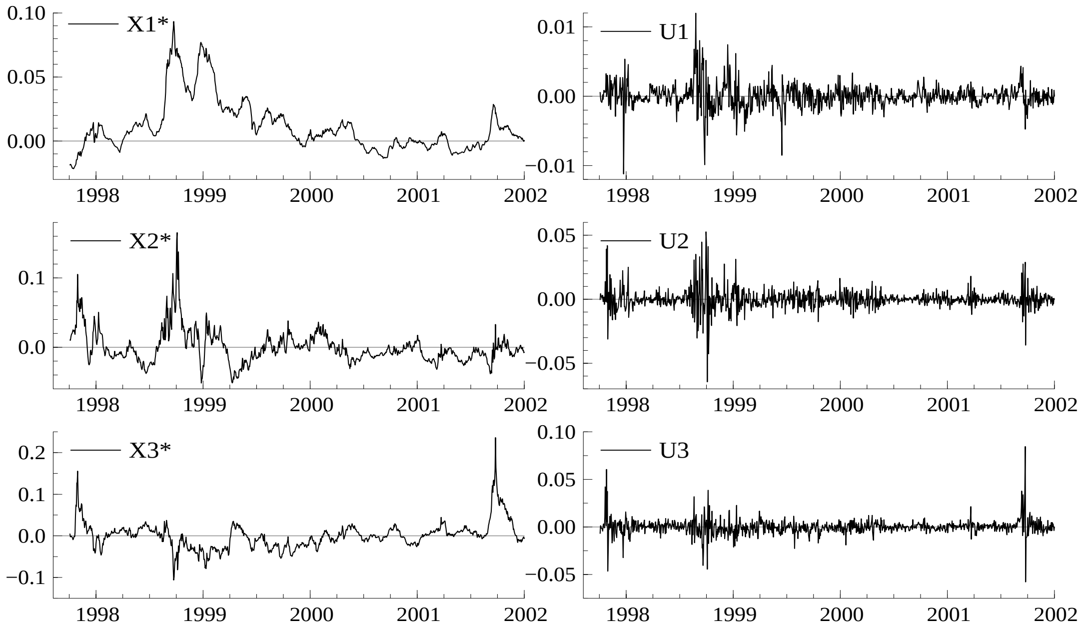


Figure 5.5: Smoothed factors in deviation from their mean x^*_{1t}, x^*_{2t} and x^*_{3t}, and daily volatility-factor shocks u_{1t}, u_{2t} and u_{3t} (**3-factor OU SV, Restrictions (a)**).

5.3.2 Another interpretation: Level, slope and curvature factors

A second interpretation of the factors concerns the impact of each factor on the prices of options, i.e., on the shape and dynamics of the volatility term structure.

Reconsidering the options measurement equation in the general multifactor model, i.e. $\sigma^2_{implied,t} = \mu_v + \frac{1}{\tau_t}\left[A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\boldsymbol{\theta}\right] + \frac{1}{\tau_t}\mathbf{B}_1(\tau_t)'\mathbf{x}^*_t + \varepsilon_t$, reveals that the functions $B_{1j}(.)/(.); j = 1,..,n$ may be interpreted as *reaction coefficients*, or *volatility-factor loadings*. Each of these coefficients measures the instantaneous ceteris-paribus response of the BS implied variance to a change in one of the factors: $\partial\sigma^2_{implied,t} / \partial x_{jt} = B_{1j}(\tau_t)/\tau_t$. These reaction coefficients depend on the option maturity. Options of different maturity -and hence the VTS- respond differently to changes in each of the volatility factors. To examine how a certain shock affects each option, it is natural to take into account that the factors have different variances. As $\mathrm{var}_{\mathbb{P}}[\mathbf{x}_t] = \mathbf{J} \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'$, the direct impact of a one-standard-deviation shock in each of the factors on the BS implied variance of an option with maturity τ_t is given by the vector ($nx1$) $(\mathbf{I}_n \odot \mathbf{J} \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}')^{1/2}\,\mathbf{B}_1(\tau_t)/\tau_t$. In the OU case, the factor loadings are given by $B_{1j}(\tau_t)/\tau_t = (1-\exp[-k_j\tau_t])/k_j\tau_t$ for $\tau_t \geq 0$. If the OU factors are independent, then $\mathrm{var}_{\mathbb{P}}[x_{jt}] = \sigma^2_{jj}/2k_j$.

Figure 5.6 shows the reaction coefficients as a function of maturity, and the instant responses of BS implied variances of all maturities to 1-standard-deviation shocks in the factors, as implied by our estimation results. These pictures yield clear insight in how the VTS is affected by such shocks. [40]

[40] Equal-maturity equal-strike call and put options have identical BS implied volatility. Moreover, the vega of such a call and put coincide, such that they respond equally to volatility changes. (This follows from the put-call parity.) Figure 5.6 thus plots the instant response of the BS implied variance (read: call or put option price) of both call and put option to factor shocks, as a function of maturity.

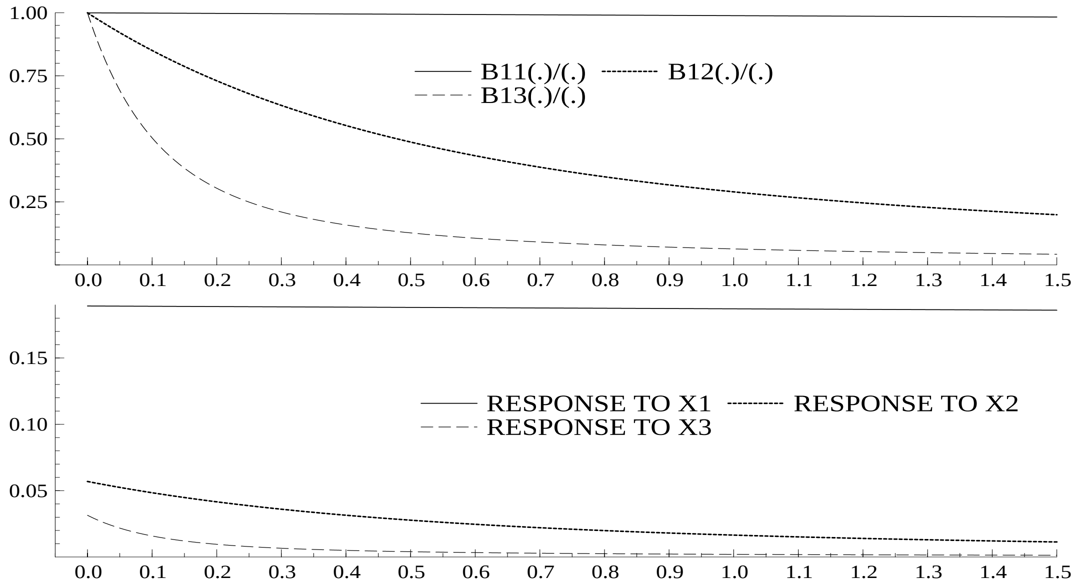


Figure 5.6. Upper plot: Reaction coefficients $B_{1j}(.)/(.); j=1,2,3$ as a function of option maturity (in years). Lower plot: Instantaneous responses of the BS implied variances to 1-standard-deviation shocks in each of the factors, as function of maturity. (**3-factor OU**)

Impact of factors on option prices and the VTS
A shock to factor x_1 has *similar impact* on all options, irrespective of maturity. Ceteris paribus, this factor seems to cause *parallel shifts* in which all implied volatilities increase with approximately the same amount. We attribute this to its persistent character, with inherent shocks that die out very slowly. As such, it mainly influences the general level of the volatility term structure. Factor x_1 may be interpreted as a *VTS level factor*.

A shock to factor x_2 affects options of all maturities as well, but by different amounts. Short-maturity option prices respond most, and the *response gradually diminishes* the longer the maturity of the option.

Shocks to factor x_3 have biggest impact on the implied volatilities (and hence prices) of short-maturity options. Its effect *quickly wears off* as the option maturity increases. Shocks to x_3 do not virtually seem to affect options with a maturity longer than a half year. As this factor is so fast mean-reverting, this makes sense: The longer the maturity, the more often x_3 tends to revert back to its mean, the closer its average value $\bar{x}_{3t}$ over the option's lifetime will be to its mean, the less a current shock has an impact on this mean. As this average value is a major determinant of the option price (i.e. $C_t = \mathbb{E}_{\mathbb{Q}}[BS(S_t, K_t, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t]$ with $\bar{\sigma}_t^2 = \bar{x}_{1t} + \bar{x}_{2t} + \bar{x}_{3t}$), it is clear that shocks to x_3 have virtually no effect on sufficiently long-dated options. Shocks to x_3 tend to "average out".

VTS level, slope and curvature factors
Given this analysis, and given that x_1 seems mainly associated with parallel shifts in the level of the VTS, it makes intuitively sense that factors x_2 and x_3 will largely be responsible for (or associated with) dynamic changes in the *slope* and *curvature* of the VTS over time (recall section 2.2).

The left panel of figure 5.7 plots the dynamics of the level, slope and convexity of the VTS through time (see also figure 2.6). The right panel shows the smoothed factors x_1^*, $-x_2^*$ and x_3^*. (Given the plot of the instant responses in figure 5.6, we expect a negative relation between x_2 and the slope of the VTS.) Especially the correspondences between x_1^* and *level*, and x_3^* and *curvature* are striking.

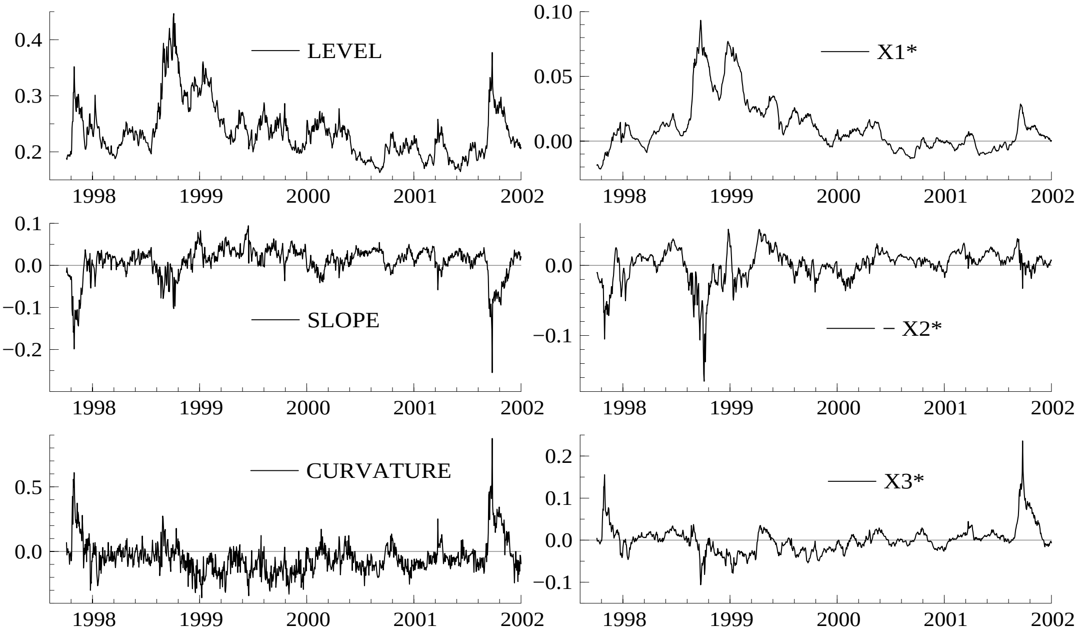


Figure 5.7. Left: Evolution of level, slope and curvature of the ATMF volatility term structure through time. Right: Smoothed x_1^*, $-x_2^*$ and x_3^*. (**3-factor OU SV**)

Factor x_1 may indeed be interpreted as mainly a VTS level factor: The correlation between x_1 and the general level of the VTS equals 0.86. Given that the impact of a shock to x_2 only gradually becomes less the longer the option maturity, it is not surprising that x_{2t} and $level_t$ are also rather correlated (correlation 0.68). In contrast, the correlation between x_{3t} and $level_t$ is only –0.08. Factor x_2 tracks changes in the slope of the VTS rather well. As expected, the slope is negatively correlated with x_2, with coefficient –0.60. In contrast, the correlation between x_{1t} and $slope_t$ is only –0.02. Not surprisingly, the correlation between x_{3t} and our slope measure is also rather large, -0.62. However, factor x_3 even more closely follows the movements of the curvature in the VTS. The correlation between $curvature_t$ and x_{3t} is substantial, 0.82, whereas the correlation between $curvature_t$ and x_{1t} (resp. x_{2t}) is only –0.07 (resp. 0.19).

We conclude the following. The persistent, long-memory factor x_1 is mainly responsible for the level dynamics of the VTS. The dynamics of factor x_2 are largely associated with changes in the slope. Dynamic changes in the convexity of the VTS are mainly driven by the very fast mean-reverting factor x_3. [41]

[41] Similar results are found in empirical implementations of affine models of the term structure of interest rates (Duffie and Kan (1996)). The term structure of interest rates plots the yield on default-free zero-coupon bonds as a function of the maturity of the bonds. Implementations of affine models show that more than one short interest rate-driving factor is necessary to obtain a good fit of empirical zero-coupon bond data. Three (correlated) factors seem needed; see e.g. de Jong (2000), Dai and Singleton (2000) and Andersen and Lund (1997). Though the interest rate setting is very different, the factors have similar interpretations as in our stock option pricing setting under SV.

5.3.3 Risk premia, straddle returns and consumption smoothing

Each volatility factor x_j has an associated market price of risk γ_j. Table 5.4 reports the estimates; notice the large standard errors. None of the γ_j's differs significantly from zero, neither do they jointly. [42] Prominent however is that γ_3 is "close-to significant". [43] Given the Monte Carlo evidence in section 3, the estimation imprecision is not so surprising. Pan (2002) reports similar difficulties in obtaining precise estimates of the market prices of volatility and jump risk. As such, some care seems in place when (numerically) interpreting the results.

Compensation for volatility risk
Compensation for volatility risk decomposes in three parts now. For the 3-independent-factors OU SV model, the spot return an investor is expected to earn when investing in an arbitrary path-independent European-style derivative F equals

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = \left[r_t + \beta_{F,S,t}\left(\sigma_t\gamma_{S,t}\right) + \boldsymbol{\beta}_{F,x,t}{}'(\boldsymbol{\Sigma\gamma})\right]dt = \left[r_t + \beta_{F,S,t}\left(\sigma_t\gamma_{S,t}\right) + \sum_{j=1}^{3}\beta_{F,j,t}\left(\sigma_{jj}\gamma_j\right)\right]dt \tag{5.9}$$

in which the risk premium associated with factor x_j is given by $\sigma_{jj}\gamma_j$. (See section 5.2 of chapter II.) The derivative price sensitivity towards factor j is given by $\beta_{F,j,t} = \text{cov}_{\mathbb{P}}[dF_t / F_t, dx_{jt} \mid \mathcal{F}_t] / \text{var}_{\mathbb{P}}[dx_{jt} \mid \mathcal{F}_t] = (1/F_t)(\partial F_t / \partial x_{jt})$, i.e. by *volatility-factor beta* j.

The estimation results show that each factor is priced differently in the FTSE100-index market. Table 5.5 reports the *factor risk premia* (per annum). The risk-premium associated with long-memory factor x_1 is negative, but equals -0.14% only. This does not seem large at all. However, we know from figure 5.6 that option prices are most sensitive to this factor, irrespective of maturity. This may possibly result in large betas associated with x_1, and may hence still yield a considerable expected-return component. The risk premium on factor x_2 is estimated at 3.9%. We lack an intuition on why it is estimated positive. The very fast mean-reverting factor x_3 knows a substantial risk premium of –26%.

SM, MM and LM at-the-money FTSE100-index straddle returns
To better be able to interpret the magnitude of these risk premia, consider the expected return on an ATM straddle:

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dStr_t}{Str_t} \mid \mathcal{F}_t\right] \approx \left[r_t + \sum_{j=1}^{3}\beta_{Str,j,t}\left(\sigma_{jj}\gamma_j\right)\right]dt\,. \tag{5.10}$$

To compute this expectation requires the *straddle factor betas* $\beta_{Str,j}$, which can only be obtained by simulation in principle. These betas may nonetheless be approximated as follows. The straddle price equals $Str_t = C_t + P_t$, in which the call price is given by $C_t = \mathbb{E}_{\mathbb{Q}}[BS(F_{t,T}, K_t, \tau_t, \bar{r}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t] = C(t, S_t, \mathbf{x}_t)$, i.e., a function $C(.)$ of time, stock price S_t, and $\mathbf{x}_t$ only (and not of its past; recall chapter II). From the put-call parity $P_t = C_t + \exp(-\bar{r}_t\tau_t)(K_t - F_{t,T})$, it next follows that $\partial C_t / \partial x_{jt} = \partial P_t / \partial x_{jt}$. Moreover, the model-implied Black-Scholes implied variance

---

[42] The Wald test statistic for testing their joint significance equals 6.0, which is smaller than the 95% (resp. 90%) critical value of 7.82 (resp. 6.25) of its asymptotic chi-squared(3) distribution.
[43] Indeed, this is an abuse of language, though the meaning should be clear.

is defined implicitly as $C_t = BS(F_{t,T}, K_t, \tau_t, \bar{r}_t, \sigma^2_{implied,t})$, such that C_t may also be considered as a function of $\sigma^2_{implied,t}$ (or, equivalently, $\sigma^2_{implied,t}$ may directly be considered a function of $\mathbf{x}_t$, as $\sigma^2_{implied,t} = BS^{-1}(C_t) = BS^{-1}(C(t, S_t, \mathbf{x}_t))$). These observations allow the straddle factor beta to be written as [44]

$$\beta_{Str,j,t} = \frac{1}{Str_t}\frac{\partial Str_t}{\partial x_{jt}} = \frac{1}{Str_t}2\frac{\partial C_t}{\partial x_{jt}} = \frac{1}{Str_t}2\frac{\partial C_t}{\partial \sigma^2_{implied,t}}\frac{\partial \sigma^2_{implied,t}}{\partial x_{jt}}, \qquad (5.11)$$

with (5.12)

$$\frac{\partial C_t}{\partial \sigma^2_{implied,t}} = \frac{\partial BS(.)}{\partial \sigma^2_{implied,t}} = \frac{\partial BS(.)}{\partial \sigma_{implied,t}}\frac{\partial \sigma_{implied,t}}{\partial \sigma^2_{implied,t}} = \frac{1}{2\sigma_{implied,t}}\frac{\partial BS(.)}{\partial \sigma_{implied,t}} = \frac{\exp(-\bar{r}_t\tau_t)F_{t,T}\sqrt{\tau_t}\,\phi(d_{1t})}{2\sigma_{implied,t}},$$

in which we used the expression for the conventional BS vega (which needs to be evaluated in the BS implied volatility). From our options measurement equation,

$$\frac{\partial \sigma^2_{implied,t}}{\partial x_{jt}} \approx \frac{B_{1j}(\tau_t)}{\tau_t} = \frac{1-\exp(-k_j\tau_t)}{k_j\tau_t}, \qquad (5.13)$$

in which the equality follows from the OU SV assumption. The ATM straddle factor betas can thus be approximated by:

$$\beta_{Str,j,t} \approx \frac{1}{Str_t}\left(\frac{\exp(-\bar{r}_t\tau_t)F_{t,T}\sqrt{\tau_t}\,\phi(d_{1t})}{\sigma_{implied,t}}\right)\frac{1-\exp(-k_j\tau_t)}{k_j\tau_t}; \quad j = 1,2,3. \qquad (5.14)$$

Given the three FTSE100-index option series with associated data, we can construct a *virtual*, near-the-money *SM, MM and LM straddle price series* in a similar way as before. Using our estimation results, nine time series of straddle factor betas (for each maturity three), can next be computed using approximation (5.14). This subsequently allows us to compute three time series of expected SM, MM, and LM ATM straddle returns using (5.10).

Table 5.5 reports the so-constructed volatility-factor betas and the expected straddle returns for each maturity, averaged over the $T = 1058$ daily observations. As expected, the pattern in the betas confirms figure 5.6. The SM straddle has largest exposure to the fast mean-reverting factor x_3 and this exposure quickly wears off the longer the straddle maturity. The exposure towards the persistent factor is similar for all maturities. Note that table 5.5 yields insight in the magnitude of the different straddle return components.

Fitting the 1-factor OU SV model using squared return – SM option data only, we found an average expected SM straddle return of –174% per annum, which closely agrees with empirical straddle returns found in the literature. Now we find an expected return of –90% "only". This may be attributed to the larger parameter imprecision. Remarkably, the implied expected MM and LM straddle

[44] Although perhaps tempting at first sight, the following approximation is *not* correct (as there are $n \geq 2$ factors):

$$\frac{\partial C_t}{\partial \mathbf{x}_t} \neq \frac{\partial \hat{C}_t}{\partial \sigma^2_t}\frac{\partial \sigma^2_t}{\partial \mathbf{x}_t} = \frac{\partial \hat{C}_t}{\partial \sigma^2_t}\mathbf{1} = \frac{1}{2\sigma_t}\left(\frac{\partial \hat{C}_t}{\partial \sigma_t}\right)\mathbf{1} \approx \frac{1}{2\sigma_t}\left(BS\,vega\right)\mathbf{1},$$

as it does *not* hold that C_t can be (re)written as a function of $\sigma^2_t = \mathbf{1}'\mathbf{x}_t$ directly; i.e., $C_t = C(t, S_t, \mathbf{x}_t) \neq \hat{C}(t, S_t, \sigma^2_t)$. This is immediately intuitively clear as well, since this latter approximation would imply that the call price has equal sensitivity $\partial C_t / \partial x_{jt}$ towards all factors, whereas from figure 5.6 we know that that is not true. (Recall the earlier footnote in chapter II.)

returns are much smaller in absolute value, but again, the magnitudes ought to be taken with care.

Table 5.5: Implied expected SM, MM and LM ATM straddle returns (3-factor OU, Restr.(a))

	Risk premium per annum ($\sigma_{jj}\gamma_j$):	**SM** ATM straddle *Volatility-factor beta:*	**MM** ATM straddle *Volatility-factor beta:*	**LM** ATM straddle *Volatility-factor beta:*
Factor x_1	**-0.14%**	**10.7** (4.3)	**9.5** (3.1)	**9.4** (2.9)
Factor x_2	**3.91%**	**8.9** (3.6)	**4.7** (1.5)	**3.1** (1.0)
Factor x_3	**-26%**	**5.0** (2.1)	**1.2** (0.4)	**0.7** (0.2)
	Av. exp. return: (per annum)	**-90%**	**-9.0%**	**-1.1%**

The table reports the implied factor risk premia, the averaged expected SM, MM and LM ATM straddle returns, and the averaged straddle factor betas (std.dev. in parentheses).

Long-term and short-term consumption smoothing

The risk premium associated with factor x_3 is much more negative than for factor x_1.[45] In particular, SM straddles seem to earn a much more negative return than LM straddles. Apart from the fact that SM straddles seem riskier, is there a possible additional economic explanation for this large difference?

First, recall our consumption-insurance argument as a theoretical justification for why we expect a *negative* volatility risk premium in practice: To smooth consumption over time, investors are willing to pay a premium for assets that pay off in times of increased volatility, as, arguably, such times are typically perceived as bad states of the world. Second, we have decomposed volatility risk into three components, the first associated with long-term volatility trends (driven by x_1), the second with medium-term (x_2), and the third with short-term trends (x_3).

Consider then a typical investor whose portfolio consists mainly of long stock and bond positions. According to the permanent-income hypothesis, this investor aims at consumption smoothing. As explained in chapter II, to achieve his goal he may well want to invest in delta-neutral positive-vega derivatives like ATMF straddles, which pay off if volatility increases.

Let us suppose that the investor is primarily interested in *long-term* (LT) consumption smoothing, but does not worry too much about *short-term* (ST) consumption fluctuations. For him the most important component of volatility risk is the LT volatility trend, x_1. Figure 5.6 reveals that this LT-investor may choose either a SM or a LM straddle to achieve his goal: Factor x_1 has similar impact on all option prices and hence straddles, irrespective of maturity.

Consider next an investor who wants to smooth his consumption no matter what. This investor cares about ST consumption fluctuations *as well*. For him all volatility-risk components are important. From figure 5.6, to achieve his goal he

[45] One may perhaps expect the factor that most negatively correlates with the FTSE100-index returns (i.e., the one that has "biggest leverage") to have the largest negative premium. This is not the case however. Based on the smoothed series, we find: $\mathrm{corr}[d\sigma_t^2, \frac{dS_t}{S_t}] = -0.66$ (leverage effect), and $\mathrm{corr}[dx_{1t}, \frac{dS_t}{S_t}] = -0.47$, $\mathrm{corr}[dx_{2t}, \frac{dS_t}{S_t}] = -0.69$, $\mathrm{corr}[dx_{3t}, \frac{dS_t}{S_t}] = -0.38$.

had better choose a SM straddle. A LM straddle is much less suitable, as a LM straddle responds to ST volatility fluctuations to a much lesser extent. A LM straddle is therefore expected not to be as valuable to him as a SM straddle.

Although the LT-investor has many straddles to choose from, the ST-investor has not. It is thus not unreasonable to think that the ST-investor is willing to pay a higher price for a SM straddle, whereas the LT-investor picks whatever straddle is cheapest. As high straddle prices mean a low expected return, this may perhaps additionally explain (besides its larger risk) why we find a far more negative risk premium associated with the ST volatility trend than for the LT trend.

We want to stress however that this reasoning is not meant to be conclusive at all for why x_3 seems to have a far more negative risk premium associated with it than x_1 has. Instead it merely represents some initial thoughts on this finding. (For example, data on the number of ST and LT investors and on the amount of SM and LM straddles that are traded in the market (i.e. demand - supply data) is lacking.) Obviously, more remains to be learned about how investors are exactly compensated for the different components of volatility risk. We leave this for future research. Note that the finding of negative expected ATM straddle returns agrees with negative (total) volatility-risk compensation however.

Short-term volatility risk versus jump risk in stock prices
Our results may be interpreted as being more or less in line with Pan's (2002) results, although the estimated models clearly differ. Recall that the Bates (2000) model (with 1 SV factor) adds price jumps to the Heston (1993) model, such that there are two risk factors (besides stock price risk): SV and jumps. When fitting this model to time series on stock returns and SM options, Pan finds that the jump risk premium dominates by far the volatility risk premium.

With a bit of imagination Pan's findings are essentially similar to what we find, as factor x_3 mainly accommodates fast changes in brief periods of time. What is different however, is that the *nature* of the risk factors is fundamentally different, i.e., jumps in *price* shocks versus "jumps" in *volatility*. [46] So at first sight, it may seem that either risk factor picks up the patterns in the data. However, this is not to say that one may be indifferent between *either* including price jumps *or* an extra volatility factor. Pan mentions that it is likely that multiple SV factors are needed to obtain a better description of the VTS. Our results do precisely confirm this, something which cannot be "solved" by including price jumps only it seems. As the diagnostic checks have shown, we really need additional, extended dynamics. (We refer to section 7.2.1 of the next chapter for more discussion.)

5.3.4 Fit of the volatility term structure

Consider next the fit of the VTS achieved by allowing for three OU factors. Figure 5.8 shows the observed and fitted BS implied volatilities, and the errors. In contrast to figures 4.14 and 5.1 for the 1-factor and 2-factor OU models, there does not seem any systematic pattern left in the error series. The improvement in fit is also apparent from table 5.6. The pricing errors concentrate around zero and are typically small. The average absolute errors seem small as well. The SM option series is now fitted best.

[46] Obviously, this should not be taken literally; x_3 has continuous sample paths. It will nonetheless be clear what we mean.

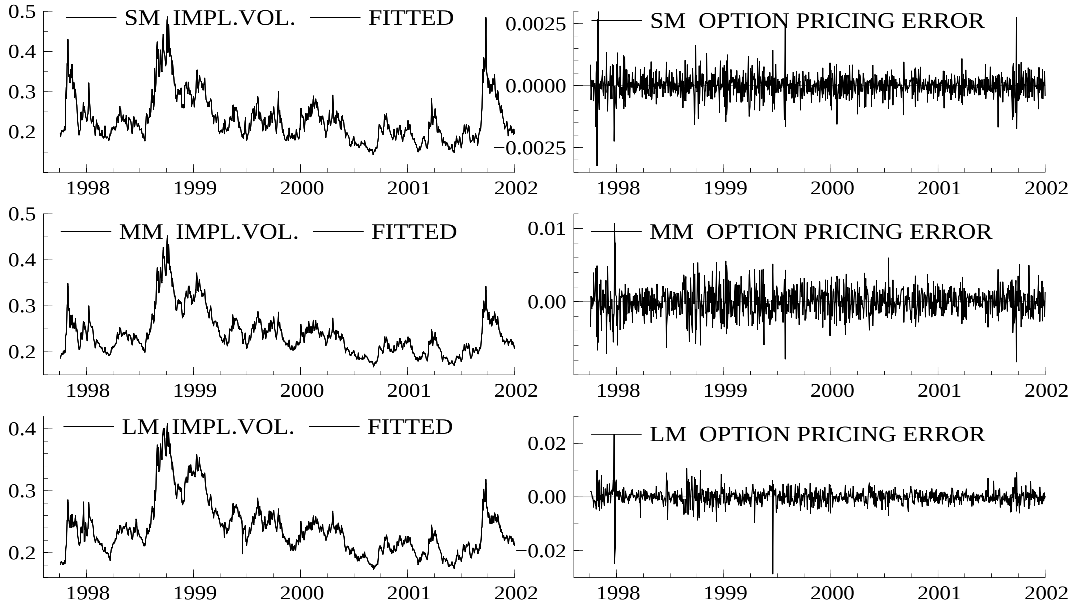


Figure 5.8: Fit of the VTS (**3-factor OU SV**, Restrictions (a)). Left: observed and fitted BS implied volatilities for each maturity. Right: their difference, the option pricing errors.

Table 5.6: Summary stat. (absolute) option pricing errors (**3-factor OU SV, Restr. (a)**)

	Error SM	*Error MM*	*Error LM*	*|Error SM|*	*|Error MM|*	*|Error LM|*
Mean	0.00%	0.00%	0.00%	0.03%	0.15%	0.19%
Std.dev.	0.05%	0.19%	0.29%	0.03%	0.13%	0.22%

A logical pattern has emerged

If we subsequently compare the absolute errors in table 4.6 (1-factor OU) with those in table 5.2 (2-factor OU) and table 5.6 (3-factor OU), the observed pattern is logically understood, given our factors interpretation and analysis so far.

1-factor SV

Fitting the 1-factor SV model to the joint data, resulted in distillation of the overall volatility trend from the data; i.e., essentially the long-memory or VTS-level factor. Due to its high persistence, it cannot respond fast enough to sudden changes in short time periods, which (given that figure 5.6 shows that it is especially SM options that are prone to these changes) resulted in the SM series fitted worst. Not surprisingly, the LM series was fitted best: Intuitively, short-term fluctuations tend to “average out” the further one looks into the future.

2-factor SV

Moving from one to two SV factors, yielded the biggest improvement in fit for the MM and especially the SM option series: The second factor in the 2-factor SV model is much faster mean-reverting, and therefore more easily picks up sudden changes. The increase in fit of the LM series was rather modest, which may again be explained by the averaging-out effect.

3-factor SV

Going from two to three SV factors, resulted in distillation of an additional, very fast mean-reverting factor. It will now be clear why we observed the biggest improvement for the SM series. Taking the averaging-out effect into account once more, it also makes sense why the fit of the LM series remained virtually constant when going from two to three factors.

5.3.5 Diagnostic checking

As opposed to the 2-factor model, is the 3-factor OU SV model dynamically well specified? If yes, in what respects can the model be improved upon?

Level-dependent volatility-of-volatility
Figure 5.9 shows the standardized innovations. Conditional heteroskedasticity is still present in all series, with biggest fluctuations in times of largest volatility (see figure 5.4). This may possibly be attributed to the fact that OU SV does not model level-dependent volatility-of-volatility. Other evidence that supports the presence of volatility feedback comes from figure 5.5: Fluctuations in the daily shocks $\{u_{jt}\}$ to the volatility-driving factors are generally largest when the (absolute) level of the factors is largest. If the model were correctly specified these shocks ought to (more or less) resemble white noise instead. Figure 5.5 also suggests that for each of the factors, volatility feedback needs to be modeled.

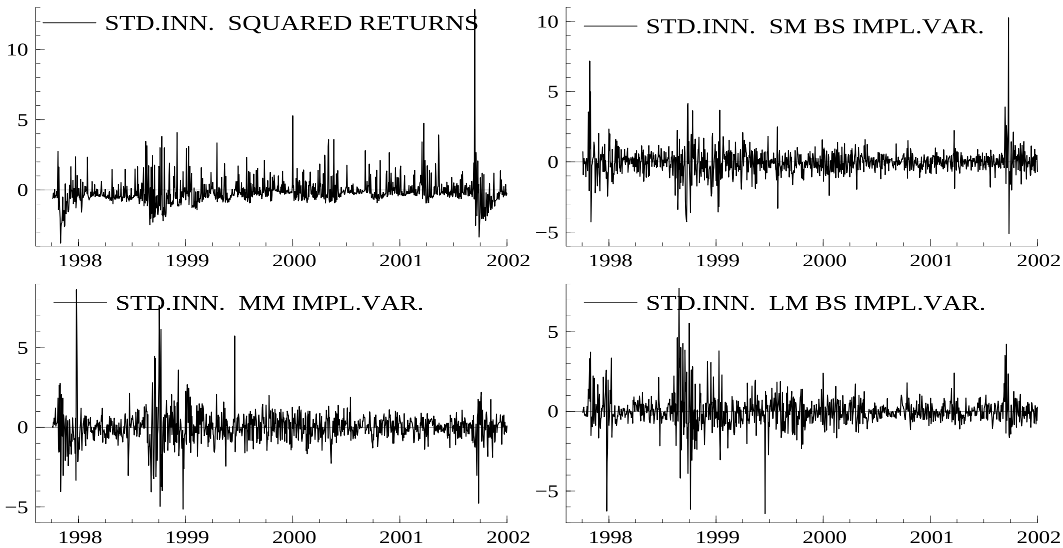


Figure 5.9: Standardized innovations (**3-factor OU SV, Restrictions (a)**).

Table 5.7: Summary statistics standardized innovations (**3-factor OU SV, Restr. (a)**)

	Std.inn. sq. return	*Std.inn. SM series*	*Std.inn. MM series*	*Std.inn. LM series*
Mean	-0.13	-0.01	0.00	0.00
Std.deviation	1.00	0.97	1.04	0.99
AC(1)	0.10	0.04	0.13	0.09
AC(2)	0.06	-0.01	-0.02	0.02
AC(3)	0.14	-0.03	-0.04	-0.10
AC(4)	0.11	-0.04	-0.04	0.06
AC(5)	0.05	-0.01	0.03	0.02
Cont.correlation matrix	1.00	0.09	0.02	0.19
		1.00	-0.06	0.10
			1.00	-0.03
				1.00

The table reports the mean, standard deviation, autocorrelation coefficients (AC) up to order 5, and the contemporaneous correlation matrix of the standardized innovations.

Table 5.7 reports summary statistics for the standardized innovations. Allowing for three factors removes most autocorrelation. Note the substantial decrease in autocorrelation in the MM series, as compared to the 2-factor case (table 5.3). Although most autocorrelations are near zero now, the Ljung-Box Q-test still rejects the null hypothesis of no serial correlation up to order 5 for all but the SM series, with respective p-values of 0.00, 0.40, 0.00 and 0.00. Remember however, that this test assumes homoskedasticity of the underlying series, and this is obviously violated.

Figure 5.10 draws the smoothed disturbances. They more closely resemble white noise than in the 1 and 2-factor OU cases. Still, also in these pictures, some effect of neglected level-dependent volatility-of-volatility seems present.

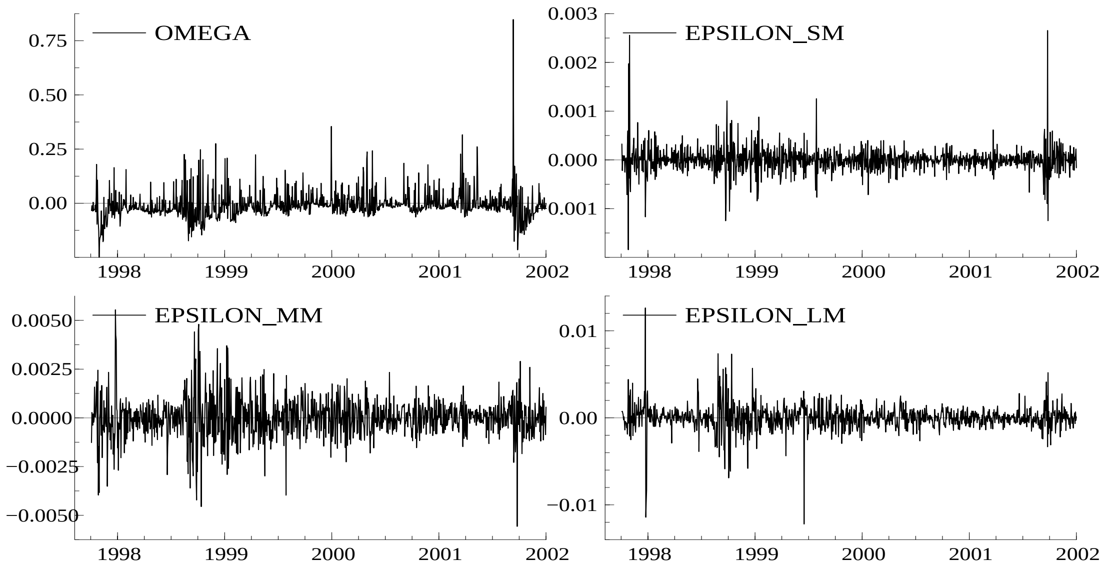


Figure 5.10: Smoothed disturbances of state space model (**3-factor OU SV, Restr.(a)**).

Leverage effect

One issue left largely unaddressed in this chapter, is the possible presence of a leverage effect in the FTSE100-index data. The multifactor affine SV derivative pricing model does not describe this effect: stock price and volatility changes occur independently from each other. Looking back at the evolution of the FTSE100 index over time (figure 2.1) and taking the course of the smoothed volatilities in figure 5.4 into account, it seems that when markets go down, volatility goes up. The correlation between the FTSE100-index returns and smoothed daily stock-variance changes equals –0.66. The leverage effect seems thus indeed present.

However, it seems unlikely that this misspecification invalidates most of our findings, given that our analysis is based on ATM options only. In chapter VI on the Heston (1993) model, we show by simulation that the VTS of ATM options is hardly affected by the leverage effect. (In contrast, leverage does impact much on the prices of in and out-of-the-money options.)

6. Summary

This chapter considers a special case of the multifactor affine SV derivative pricing model: OU SV. In the OU case, the volatility is driven by one or multiple Gaussian OU processes, such that the volatility-of-volatility is level-independent.

OU SV yields large analytical tractability. Moreover, as the conditional and unconditional moments of the daily volatility-factor shocks $\{\mathbf{u}_t\}$ coincide, the unconditional state space model is excellently suited for estimation.

Monte Carlo study based on the 1-factor OU SV model
For the 1-factor OU SV model, we conduct a Monte Carlo study for one set of (rather realistic) parameter values. We simulate time series of daily returns, 10-minute realized volatilities (i.e., 48 intraday points) and short-maturity at-the-money options. Our main findings are summarized as follows.

Regarding the quality of the linear and parabolic call price approximations that we carried out to arrive at a linear options measurement equation, the following. The linearization error is small and negative. The parabolic error is somewhat smaller, but centers around zero. Both approximations seem to provide accurate approximations of the true call price. The relative merits measure is very small. This suggests that using a parabolic approximation for improved accuracy does not outweigh the associated, additional implementation cost of increased computational complexity (at least, if one or multiple option series are analyzed characterized by approximately the same moneyness). "Parameter" μ_v (which we introduced in the options measurement equation to allow for a non-zero average approximation error) is negative, but very close to zero.

Regarding the white noise assumption for the errors $\{\omega_t\}$, $\{\omega_{RV,t}\}$ and $\{\varepsilon_t\}$ associated with the squared return, RV, and options measurement equations respectively, the simulated data supports the reasonability of these assumptions. On average, the errors $\{\varepsilon_t\}$ appear to be slightly positively autocorrelated however. As such, one of the assumptions on which the Kalman filter equations are built is strictly speaking violated.

We estimate the state space model associated with the 1-factor OU SV model using five different types of simulated data: only squared return data, only RV data, only (short-maturity ATM) option data, both return - option data, and both RV and option data. Investigating the performance of our state space method in recovering the true parameters and volatility paths underlying the simulated data reveals the following.

Squared return data performs worst: The bias and MSEs are largest, mainly with regard to the mean-reversion and volatility-of-volatility parameters. Moreover, the smoothed and true underlying volatilities deviate most. The results confirm why squared returns are generally considered as noisy estimators of the stock variance: They do not contain very precise information. Consistent with the literature, QML performs poorly in this case. Using RV data instead, yields a substantial improvement in bias, MSE and volatility evaluation criteria.

In turn, option data alone outperforms RV on all criteria: both the bias and MSEs are small, and the smoothed and true volatilities are close. Option data is clearly very informative.

Combining squared return and option data performs somewhat better than option data alone. Imposing the restriction $\mu_V = 0$ prior to estimation is preferable in this case: Although the parameter estimates remain virtually the same, it leads to better volatility filtering.

Combining RV and option data provides the best results: The estimation bias is small for all parameters, it performs best on the MSE criterion (i.e. yields the most efficient estimates), and the smoothed volatility series is particularly close to the underlying true volatility series. The combination RV – option data contains the most precise information, and is therefore preferable for estimation. As in the squared return–option case, restricting μ_V to 0 leads to somewhat better results.

The results further show that, although the estimation bias in the market price of volatility risk is small, the precision of the estimates is poor, irrespective of the data used for estimation. (This underwrites empirical findings in the literature.) Information on this parameter in either type of data seems particularly weak.

Empirical results OU SV based on the FTSE100-index data
We next confront the OU model to empirical FTSE100 stock-index and option data. The data consists of time series of daily index returns, a SM, MM and LM at-the-money index call option series.

Explorative data analysis shows that the level and fluctuations in the index volatility have substantially increased from the fall of 1997 onwards. Augmented market uncertainty due to e.g. the Asian crisis and the war on terrorism are held accountable for this. We therefore only use the data for Oct 97–Dec 2001 in the further analysis. Examination of the ATM VTS reveals its different shapes, and its complex evolution over time. 34% of its shifts is non-parallel. A realistic model must obviously be capable of reproducing these observed, complicated dynamics. As, by approximation, a 1-factor SV model implies parallel shifts in the VTS only, one SV factor is likely to be insufficient in this respect.

Initially, we assume there to be only one OU factor driving the volatility. Comparing the results obtained from using squared return data only with GARCH results, shows that although the volatility persistence (0.94; half-life 11 days) and unconditional volatility (20%) are estimated similar, the GARCH volatilities do not fluctuate as much, and do not respond as fast to news as the SV volatilities.

Using return - SM option data jointly, the persistence is estimated much larger, at 0.993 (half-life 98 days). The option data dominates the estimation results, being much more informative than the return data. The observed BS implied volatilities are typically larger than the “true” underlying hidden volatilities. We attribute this to the negative market price of volatility risk. This illustrates that the common market interpretation of BS implied volatilities being forecasts of future volatility should be taken with care if volatility risk is priced.

Investing in FTSE100-index derivatives yields a negative premium for volatility risk of −15% per annum. In line with our theory, investors thus pay for consumption insurance, which can be achieved by investing in delta-neutral positive-vega derivatives. Investing in SM ATM straddles yields an expected return of −174% per annum. This closely agrees with average empirical SM ATM index-straddle returns that other researchers have found. Hence, although writing straddles is expected to yield a large return, we show that this investment strategy entails considerable risk, and is a true bet on volatility.

The misspecification of the 1-factor OU model estimated using return - SM option data seems rather modest at first sight. The main deficiency seems negligence of volatility feedback (apart from the leverage effect). However, the estimated model leads to substantial out-of-sample overpricing of the longer-dated options.

When the model is estimated using all (return, SM, MM and LM option) data jointly, the most important misspecification of the 1-factor OU model appears to be insufficient dynamics. The VTS is fitted worst at the short end and best at the long end.

We next investigate the contribution of each additional SV factor to the fit of the data. Although two OU factors yield a considerable increase in fit, dynamic misspecification still remains. The fit of the three option series is now similar.

Extending to three OU factors appears sufficient to satisfactorily describe the dynamics observed in the four joint time series. Allowing for correlation between the factors seems to not much further increase the fit, though does introduce a lot of additional parameter uncertainty. We therefore choose to analyze the independent-factors results further.

Comparing the smoothed stock volatilities obtained from the 1-factor model (estimated using return – SM option data) to those obtained from the 3-factor model, reveals that especially in times of sudden increases in financial markets uncertainty, the latter ones respond quicker. This makes sense: Decomposing the volatility in different dynamic components allows the volatility to respond quicker.

The three OU factors distilled from the data differ greatly in their characteristics. Factor x_1 is extremely persistent. It has a daily persistence of 0.9999, with shocks that have close-to permanent effects on its future evolution; they seem to last for years. Factor x_2 is much quicker mean reverting (persistence 0.9873) with shocks that have a half-life of about 2.5 months. Factor x_3 is very fast mean-reverting (persistence 0.9413). Shocks to x_3 lose half their impact in 11 days, and have moreover largest variance. This factor clearly governs large volatility changes in relatively brief periods of time.

We interpret the factors in two ways. The first interpretation is that x_1, x_2 and x_3 determine the long, medium and short-term volatility trends respectively. As such, they influence the prices of options of different maturities in different ways. Factor x_1 impacts on all options similarly, irrespective of maturity. Factor x_2 affects all options as well, but gradually by smaller amounts the longer the option's lifetime. Short-maturity options are especially prone to shocks in factor x_3. The impact of x_3 quickly diminishes as the option maturity increases. As x_3 is so rapidly mean reverting, its shocks tend to average out over a sufficiently long time span, resulting in an only marginal impact on long-dated options.

Our second interpretation of the factors concerns their impact on the shape and dynamics of the volatility term structure. x_1 is mainly responsible for changes in the general level of the VTS. x_2 is largely associated with changes in the slope. x_3 is surprisingly closely related to dynamic changes in the convexity of the VTS.

Consistent with the literature, and not surprising given the Monte Carlo results, the market prices of risk associated with the SV factors are estimated imprecisely. Only x_3 seems to have a price of risk significantly different from zero. We do not have an intuition for why the price of risk of x_2 is estimated positive; both others are negative. The risk premium associated with fast mean-reverting factor x_3 is much more negative (-26% per annum) than the premium for the long-memory factor x_1 (-0.14%). The results imply that SM ATM straddles earn a much more negative return than LM ATM straddles on average. This agrees with the fact that investing in SM straddles seems riskier. (The differences are substantial however.)

The fit of the VTS obtained by allowing for three OU factors is good. The option pricing errors concentrate around zero, are typically small, and do not reveal any systematic pattern left. The SM series is fitted somewhat best. Comparing, in succession, the pricing errors from the 1, 2 and 3-factor models yields a pattern that is logically understood, given our previous analysis.

Specification tests of the 3-factor OU SV model reveal that the dynamics in the data are satisfactorily captured. We also find evidence of the presence of the leverage effect and level-dependent volatility-of-volatility. The multifactor OU SV model does not model both data features.

It seems unlikely however, that these misspecifications invalidate most of our findings. Given that our analysis is based on ATM options only, we expect the effects of negligence of the leverage effect to be modest. (In chapter VI on the Heston (1993) model, we show by simulation that the ATM VTS is hardly affected by the leverage effect.) Level-dependent volatility-of-volatility mainly concerns conditional heteroskedasticity. As such, it is less associated with the time-series dimension (the dynamics) of the volatility.

The next chapter allows for volatility feedback in the SV-driving processes. We then resort to a different state space model (the *conditional* state space model) and estimation method: Extended Kalman filter QML.

Appendix

IVa. Parameter identification in the OU SV case

This appendix considers parameter identification in the 1-factor OU SV special case. The identification problem is contingent on the data used for estimation.

Using only RV data for estimation

Using only RV data for estimation, the state space model reduces to

$$\sigma^2_{RV,t+\Delta t} = \theta + \frac{1-\exp(-k\Delta t)}{k\Delta t} x_t^* + \omega_{RV,t+\Delta t}, \quad \omega_{RV,t+\Delta t} \sim (0, \sigma^2_{\omega,RV}), \qquad \text{(a.1)}$$
$$x^*_{t+\Delta t} = \exp(-k\Delta t) x_t^* + u_{t+\Delta t}, \qquad u_{t+\Delta t} \sim \left(0,\ \sigma^2 \frac{1-\exp(-2k\Delta t)}{2k}\right),$$

for $t = 0,\ \Delta t, .., (T-1)\Delta t$, in which $\{\omega_{RV,t+\Delta t}\}$ and $\{u_{t+\Delta t}\}$ are non-contemporaneously uncorrelated, but contemporaneously correlated with $\text{corr}_{\mathbb{P}}[\omega_{RV,t+\Delta t}, u_{t+\Delta t}] = c$. To examine parameter identification, the Kalman filter recursions could explicitly be worked out in principle. This is a tedious exercise however, even for this simple model.

We therefore attempt another strategy. We focus on the method-of-moments estimators (MMEs) of the parameters. We construct the MMEs only from the mean and autocovariance function of the observed RV series $\{\sigma^2_{RV,t+\Delta t}\}$; we do not take higher-order moments into account. Why not? Intuitively, it is only the mean and autocovariance function of the observed data from which Kalman filter (Q)ML extracts information, and bases its estimates on. To better justify this intuition, the following. $\{u_{t+\Delta t}\}$ is a (discrete-time) Gaussian process in the 1-factor OU case. Now, suppose that $\{\omega_{RV,t+\Delta t}\}$ were Gaussian as well (though we know it is not). Then the whole system would be Gaussian, and the RV series would follow a Gaussian process. Now, a Gaussian process is completely specified by its expectation and covariance function; all other moments are fixed constant numbers which do not depend on parameters. Hence, a sound estimation strategy (and thus certainly Kalman filter ML, as for given parameter values, the Kalman filter delivers the MMSE estimator of the state vector in the Gaussian case) should only extract information from the sample mean and covariance function of the observed data, and not from other (higher-order) moments. Therefore, to be able to "transfer" the identification problem encountered here with Kalman filter QML towards a similar (but easier manageable) identification problem for MM estimation, it seems that we should only focus on the mean and autocovariance function of the observed RV series, and not on its higher moments.

Having motivated our strategy, let us examine what it brings us. First, for simplicity introduce the parameter $\rho \equiv \text{cov}_{\mathbb{P}}[\omega_{RV,t+\Delta t}, u_{t+\Delta t}] = c\,\sigma_{\omega,RV}\sigma\sqrt{[1-\exp(-2k\Delta t)]/2k}$. From (a.1),

$$\mathbb{E}_{\mathbb{P}}\left[\sigma^2_{RV,t+\Delta t}\right] = \theta, \tag{a.2}$$

such that θ can essentially be identified from the sample mean of the RV data. The variance of the realized variance, γ_0, equals

$$\gamma_0 \equiv \text{var}_{\mathbb{P}}\left[\sigma^2_{RV,t+\Delta t}\right] = \frac{\sigma^2}{2k}\left(\frac{1-\exp[-k\Delta t]}{k\Delta t}\right)^2 + \sigma^2_{\omega,RV}, \tag{a.3}$$

in which we used $\text{var}_{\mathbb{P}}[x^*_t] = \text{var}[x_t] = \sigma^2/2k$, see appendix B, section 13.1. The first-order autocovariance $\gamma_{\Delta t}$ equals

$$\begin{aligned}
\gamma_{\Delta t} &\equiv \text{cov}_{\mathbb{P}}\left[\sigma^2_{RV,t+\Delta t}, \sigma^2_{RV,t}\right] \\
&= \text{cov}_{\mathbb{P}}\left[\tfrac{1-\exp[-k\Delta t]}{k\Delta t}x^*_t + \omega_{RV,t+\Delta t}, \tfrac{1-\exp[-k\Delta t]}{k\Delta t}x^*_{t-\Delta t} + \omega_{RV,t}\right] \\
&= \left(\tfrac{1-\exp[-k\Delta t]}{k\Delta t}\right)^2 \text{cov}_{\mathbb{P}}\left[x^*_t, x^*_{t-\Delta t}\right] + \tfrac{1-\exp[-k\Delta t]}{k\Delta t}\text{cov}_{\mathbb{P}}\left[x^*_t, \omega_{RV,t}\right] \\
&= \left(\tfrac{1-\exp[-k\Delta t]}{k\Delta t}\right)^2 \text{cov}_{\mathbb{P}}\left[x^*_t, x^*_{t-\Delta t}\right] + \tfrac{1-\exp[-k\Delta t]}{k\Delta t}\text{cov}_{\mathbb{P}}\left[\exp[-k\Delta t]x^*_{t-\Delta t} + u_t, \omega_{RV,t}\right] \\
&= \left(\tfrac{1-\exp[-k\Delta t]}{k\Delta t}\right)^2 \text{cov}_{\mathbb{P}}\left[x^*_t, x^*_{t-\Delta t}\right] + \tfrac{1-\exp[-k\Delta t]}{k\Delta t}\rho \\
&= \frac{\sigma^2}{2k}\exp[-k\Delta t]\left(\frac{1-\exp[-k\Delta t]}{k\Delta t}\right)^2 + \frac{1-\exp[-k\Delta t]}{k\Delta t}\rho,
\end{aligned} \tag{a.4}$$

where the final equality uses $\text{cov}_{\mathbb{P}}[x^*_t, x^*_{t-\Delta t}] = \sigma^2/(2k)\exp[-k\Delta t]$; see appendix B, (7.1). The second-order autocovariance $\gamma_{2\Delta t}$ equals

$$\begin{aligned}
\gamma_{2\Delta t} &\equiv \text{cov}_{\mathbb{P}}\left[\sigma^2_{RV,t+\Delta t}, \sigma^2_{RV,t-\Delta t}\right] \\
&= \text{cov}_{\mathbb{P}}\left[\tfrac{1-\exp[-k\Delta t]}{k\Delta t}x^*_t + \omega_{RV,t+\Delta t}, \tfrac{1-\exp[-k\Delta t]}{k\Delta t}x^*_{t-2\Delta t} + \omega_{RV,t-\Delta t}\right]
\end{aligned} \tag{a.5}$$

$$= \left(\tfrac{1-\exp[-k\Delta t]}{k\Delta t}\right)^2 \mathrm{cov}_{\mathbb{P}}\left[x_t^*, x_{t-2\Delta t}^*\right] + \tfrac{1-\exp[-k\Delta t]}{k\Delta t}\mathrm{cov}_{\mathbb{P}}\left[x_t^*, \omega_{RV,t-\Delta t}\right].$$

By (7.1) in appendix B,

$$\mathrm{cov}_{\mathbb{P}}\left[x_t^*, x_{t-2\Delta t}^*\right] = \exp[-k\Delta t]\,\mathrm{cov}_{\mathbb{P}}\left[x_t^*, x_{t-\Delta t}^*\right], \tag{a.6}$$

and

$$\begin{aligned}\mathrm{cov}_{\mathbb{P}}\left[x_t^*, \omega_{RV,t-\Delta t}\right] &= \mathrm{cov}_{\mathbb{P}}\left[\exp[-k\Delta t]x_{t-\Delta t}^* + u_t,\ \omega_{RV,t-\Delta t}\right] \\ &= \exp[-k\Delta t]\,\mathrm{cov}_{\mathbb{P}}\left[x_{t-\Delta t}^*,\ \omega_{RV,t-\Delta t}\right] \\ &= \exp[-k\Delta t]\,\mathrm{cov}_{\mathbb{P}}\left[\exp[-k\Delta t]x_{t-2\Delta t}^* + u_{t-\Delta t},\ \omega_{RV,t-\Delta t}\right] \\ &= \exp[-k\Delta t]\,\rho,\end{aligned} \tag{a.7}$$

such that

$$\begin{aligned}\gamma_{2\Delta t} &= \left(\tfrac{1-\exp[-k\Delta t]}{k\Delta t}\right)^2 \exp[-k\Delta t]\,\mathrm{cov}_{\mathbb{P}}\left[x_t^*, x_{t-\Delta t}^*\right] + \tfrac{1-\exp[-k\Delta t]}{k\Delta t}\exp[-k\Delta t]\,\rho \\ &= \exp[-k\Delta t]\left[\left(\tfrac{1-\exp[-k\Delta t]}{k\Delta t}\right)^2 \mathrm{cov}_{\mathbb{P}}\left[x_t^*, x_{t-\Delta t}^*\right] + \tfrac{1-\exp[-k\Delta t]}{k\Delta t}\,\rho\right] \\ &= \exp[-k\Delta t]\,\gamma_{\Delta t}.\end{aligned} \tag{a.8}$$

A similar analysis shows that $\gamma_{3\Delta t} = \exp[-k\Delta t]\gamma_{2\Delta t}$, and so forth. For general $p = 2, 3, ..$, the autocovariance of order p thus equals

$$\begin{aligned}\gamma_{p\Delta t} &= \exp[-k\Delta t]\,\gamma_{(p-1)\Delta t} \\ &= \exp[-(p-1)k\Delta t]\,\gamma_{\Delta t};\quad p = 2, 3, ..\end{aligned} \tag{a.9}$$

(Note that the autocorrelation function of $\{\sigma_{RV,t+\Delta t}^2\}$ declines geometrically towards zero, which is as expected, since the RV series is covariance-stationary.)

We now return to the identification problem. As θ has already been identified, this leaves $\sigma, k, \rho, \sigma_{\omega,RV}$ to still be identified. The sample autocovariance function computed from the RV data yields estimates of $\gamma_0, \gamma_{\Delta t}, \gamma_{2\Delta t}, \gamma_{3\Delta t}, ..$. Parameter k can next be identified from e.g. $\gamma_{3\Delta t}$ and $\gamma_{2\Delta t}$. As $\gamma_{3\Delta t} = \exp[-k\Delta t]\gamma_{2\Delta t}$, $k = -\ln(\gamma_{3\Delta t} / \gamma_{2\Delta t}) / \Delta t$. [47] This yields three parameters $\{\sigma, \rho, \sigma_{\omega,RV}\}$ left to identify, but only two quantities to still be able to extract information from: Since $\gamma_0 = \gamma_0(\sigma, \sigma_{\omega,RV}, k)$ and $\gamma_{\Delta t} = \gamma_{\Delta t}(\sigma, \rho, k)$ it follows that $\sigma, \sigma_{\omega,RV}$ and ρ (and hence parameter c) are not separately identified. As soon as one of these parameters is restricted to a certain value, the other two parameters can be identified. (Keep in mind that an incorrect restriction will probably lead to estimation bias in both other parameters.

Using other types of data for estimation

Consider next the state space model estimated using *only squared return data*. The model reduces to equations (3.17) and (3.20). The parameters that can potentially be estimated are θ, k, σ_ω and σ. The identification problem is similar as in case of RV data. However, as $\mathrm{corr}[\omega_t, u_t] = 0$, all four parameters can be identified.

Consider next using *only option data* for estimation. The state space model reduces to (3.19)-(3.20). Potential parameters to be estimated are μ_v, θ, γ, σ, k and σ_ε. Regarding k, σ and σ_ε, a close comparison reveals that the problem is similar as in the

[47] Notice that higher-order sample autocovariances provide additional information on k but not directly on the other parameters. Therefore, rather than basing the estimate of k on just the sample analogues of $\gamma_{3\Delta t}$ and $\gamma_{2\Delta t}$, a more efficient GMM estimate of k can in principle be obtained by taking these higher-order sample autocovariances into account.

RV case with $c = 0$, such that these can be identified. Information on μ_v, θ, γ can only be derived from the mean BS implied variance over the sample. As such, the parameters μ_v and θ cannot separately be identified; only their sum is identifiable. Hence, either of these parameters ought to be restricted. γ is theoretically identified as the option maturity τ_t differs over t.

Consider next using the *combination return-option data* for estimation. The state space model reduces to equations (3.17), (3.19) and (3.20) and has parameters θ, k, σ, γ, μ_v, σ_ω, σ_ε. From the analysis on using either only squared return or option data, it follows that all parameters can be identified.

Consider finally using *both RV – option data* for estimation. The state space model reducing to (3.18)-(3.20), and has parameters θ, k, σ, γ, μ_v, $\sigma_{\omega,RV}$, σ_ε, c. From the analysis on using either only RV or option data, some thinking reveals that in this case all parameters can be identified. Specifically, the option data provides information on σ, after which $\sigma_{\omega,RV}$ and c can also be identified from the RV data. As θ can be identified from the RV data, μ_v is now identifiable as well.

IVb. Curvature measure of the volatility term structure

This appendix provides a rationale for the *convexity* or *curvature measure* of the volatility term structure, as discussed in section 2.2.

The VTS plots the BS implied volatilities as a function of maturity. On each day we observe three BS implied volatilities, a SM, MM and LM Black Scholes implied volatility. The way these volatilities are located with respect to each other determines the shape and hence the curvature of the VTS. (Think about three points on the graph of a smooth function.)

Consider then an at least twice differentiable function $f(.)$. The aim is to obtain an approximation of the curvature of this function in a certain point y, given three coordinate-pairs $[y+h_1, f(y+h_1)]$, $[y, f(y)]$ and $[y-h_2, f(y-h_2)]$, with $h_1, h_2 > 0$. The second-order derivative of $f(.)$ evaluated in the point y, i.e. $f''(y)$, determines this curvature. As such, our goal is to determine the coefficients a, b and c such that

$$f''(y) \approx a\,f(y+h_1) + b\,f(y) + c\,f(y-h_2)\,. \tag{b.1}$$

Second-order Taylor series approximations lead to

$$f(y+h_1) \approx f(y) + h_1 f'(y) + \tfrac{1}{2}h_1^2 f''(y) \tag{b.2}$$

$$f(y-h_2) \approx f(y) - h_2 f'(y) + \tfrac{1}{2}h_2^2 f''(y)\,. \tag{b.3}$$

Substitution into (b.1), followed by rewriting gives

$$f''(y) \approx (a+b+c)f(y) + (ah_1 - ch_2)f'(y) + \tfrac{1}{2}(ah_1^2 + ch_2^2)f''(y)\,. \tag{b.4}$$

This leads to the following restrictions

$$a+b+c \equiv 0, \quad ah_1 - ch_2 \equiv 0, \quad \tfrac{1}{2}(ah_1^2 + ch_2^2) \equiv 1\,, \tag{b.5}$$

which yield

$$a = \frac{2}{h_1(h_1 + h_2)}, \quad c = \frac{2}{h_2(h_1 + h_2)}, \quad b = -a - c. \tag{b.6}$$

These coefficients together with (b.1) determine the approximate curvature in the point $[y, f(y)]$. Notice in particular that if $f(y + h_1) = f(y) = f(y - h_2)$ (i.e. the function is flat), the curvature is zero, as it ought to be. Note moreover that if $h_1 = h_2 = h$, the convexity approximation reduces to the well-known formula

$$f''(y) \approx \frac{f(y+h) - 2f(y) + f(y-h)}{h^2}. \tag{b.7}$$

In our application, the role of function $f(.)$ is played by the BS implied volatilities considered as a function of maturity, i.e., by the volatility term structure. The maturity of the MM series $\tau_{MM,t}$ is associated with y, h_1 equals the maturity difference between the LM and MM series, and h_2 equals the maturity difference between MM and SM series. As these differences are not equal (and vary over time as well) we approximate the curvature of the VTS by equations (b.1) and (b.6) leading to our curvature measure as defined as $curvature_t \equiv a_t\, \sigma_{implied,LM,t} + b_t\, \sigma_{implied,MM,t} + c_t\, \sigma_{implied,SM,t}$ in the main text in which $a_t = 2/[h_{1t}(h_{1t} + h_{2t})]$, $c_t = 2/[h_{2t}(h_{1t} + h_{2t})]$, $b_t = -a_t - c_t$, $h_{1t} = \tau_{LM,t} - \tau_{MM,t}$ and $h_{2t} = \tau_{MM,t} - \tau_{SM,t}$. This curvature measure thus represents a numerical approximation of the second-order derivative of a smooth function in a certain point, computed using three coordinate pairs.

IVc. System matrices state space model for multifactor OU SV

The state space model associated with the multifactor affine OU SV derivative pricing model, in which both the squared returns and the three option series are jointly used for estimation, reads

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{a}_t + \boldsymbol{H}_t{}'\boldsymbol{\xi}_t + \boldsymbol{w}_t & \boldsymbol{w}_t &\sim (\mathbf{0}, \boldsymbol{R}) \quad t = \Delta t, \ldots, T\Delta t \\ \boldsymbol{\xi}_{t+\Delta t} &= \boldsymbol{d} + \boldsymbol{F}\boldsymbol{\xi}_t + \boldsymbol{v}_{t+\Delta t} & \boldsymbol{v}_{t+\Delta t} &\sim (\mathbf{0}, \boldsymbol{Q}), \end{aligned} \tag{c.1}$$

in which the error series $\{\boldsymbol{w}_t\}$ and $\{\boldsymbol{v}_{t+\Delta t}\}$ are both serially and mutually uncorrelated at all points in time, and in which the system matrices read [48]

$$(4x1)\,\boldsymbol{y}_t \equiv \begin{bmatrix} \frac{1}{\Delta t}(r_t - \hat{\mu}_{t-\Delta t}\Delta t)^2 \\ \sigma^2_{implied,1t} \\ \sigma^2_{implied,2t} \\ \sigma^2_{implied,3t} \end{bmatrix}, \qquad (4x1)\,\boldsymbol{a}_t \equiv \begin{bmatrix} 0 \\ \mu_v + \frac{1}{\tau_{1t}}[A_1(\tau_{1t}) + \mathbf{B}_1(\tau_{1t})'\boldsymbol{\theta}] \\ \mu_v + \frac{1}{\tau_{2t}}[A_1(\tau_{2t}) + \mathbf{B}_1(\tau_{2t})'\boldsymbol{\theta}] \\ \mu_v + \frac{1}{\tau_{3t}}[A_1(\tau_{3t}) + \mathbf{B}_1(\tau_{3t})'\boldsymbol{\theta}] \end{bmatrix} \tag{c.2}$$

$$(4x\,n+1)\,\boldsymbol{H}_t{}' \equiv \begin{bmatrix} 1 & \mathbf{0}' \\ 0 & \frac{1}{\tau_{1t}}\mathbf{B}_1(\tau_{1t})' \\ 0 & \frac{1}{\tau_{2t}}\mathbf{B}_1(\tau_{2t})' \\ 0 & \frac{1}{\tau_{3t}}\mathbf{B}_1(\tau_{3t})' \end{bmatrix}, \qquad (n+1\,x1)\,\boldsymbol{\xi}_t \equiv \begin{bmatrix} \frac{1}{\Delta t}(r_t - \hat{\mu}_{t-\Delta t}\Delta t)^2 \\ \mathbf{x}_t^* \end{bmatrix},$$

[48] For notational convenience we index the option series with 1, 2 and 3 instead of SM, MM, and LM.

$$(4x1)\,\boldsymbol{w}_t \equiv \begin{bmatrix} 0 \\ \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}, \qquad (4x4)\,\boldsymbol{R} \equiv \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma_{\varepsilon1}^2 & \rho_{12}\sigma_{\varepsilon1}\sigma_{\varepsilon2} & \rho_{13}\sigma_{\varepsilon1}\sigma_{\varepsilon3} \\ 0 & \rho_{12}\sigma_{\varepsilon1}\sigma_{\varepsilon2} & \sigma_{\varepsilon2}^2 & \rho_{23}\sigma_{\varepsilon2}\sigma_{\varepsilon3} \\ 0 & \rho_{13}\sigma_{\varepsilon1}\sigma_{\varepsilon3} & \rho_{23}\sigma_{\varepsilon2}\sigma_{\varepsilon3} & \sigma_{\varepsilon3}^2 \end{bmatrix},$$

$$(n+1\,x1)\,\boldsymbol{d} \equiv \begin{bmatrix} \mathbf{1}'\boldsymbol{\theta} \\ \mathbf{0} \end{bmatrix}, \qquad (n+1\,x\,n+1)\,\boldsymbol{F} \equiv \begin{bmatrix} 0 & \mathbf{1}' \\ \mathbf{0} & \exp[-\mathbf{K}_d\Delta t] \end{bmatrix},$$

$$(n+1\,x1)\,\boldsymbol{v}_{t+\Delta t} \equiv \begin{bmatrix} \omega_{t+\Delta t} \\ \mathbf{u}_{t+\Delta t} \end{bmatrix}, \qquad (n+1\,x\,n+1)\,\boldsymbol{Q} \equiv \begin{bmatrix} \sigma_\omega^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{G}(\Delta t)\odot\boldsymbol{\Sigma\Sigma}' \end{bmatrix}.$$

Matrix $(nxn)\,\mathbf{G}(\Delta t)$ has $ij-$th element $[\mathbf{G}(\Delta t)]_{ij} = (1-\exp[-(k_i+k_j)\Delta t])/(k_i+k_j)$. Notice that we adopt a state space form that is suitable for volatility forecasting (with $\delta_0 = 0$, $\boldsymbol{\delta} = \mathbf{1}$ imposed), as explained in section 3.5.2 of chapter III. We estimate the state space model by Kalman filter QML.

- V-

Exploiting
Level-Dependent Volatility-of-Volatility:
An Extended Kalman Filter Approach to the Estimation of the Multifactor Affine SV Derivative Pricing Model

1. Introduction

The estimation results presented in the previous chapter show that an empirical shortcoming of the OU SV specification is its inability to describe *level-dependent volatility-of-volatility*, also known as *volatility feedback*. Although OU SV is able to describe volatility clustering, it can do that to some extent only. OU SV cannot describe the periods of heavy clustering observed in the FTSE100-index data sufficiently well. We expect improved results to be obtained by allowing the volatility-of-volatility to change over time.

Data that features volatility feedback is characterized by volatility shocks which depend on the current level of the volatility. As such, more efficient use of information is made if the *conditional* distribution of the volatility-factor shocks $\{\mathbf{u}_t\}$ is explicitly taken into account in the estimation procedure. The unconditional state space model so far worked with is based on the unconditional distribution of $\{\mathbf{u}_t\}$ only. Although in the OU SV case the unconditional and conditional distributions of the shocks coincide, this does not hold for the general case. Therefore, in this chapter we extend our estimation method to better deal with volatility feedback-featuring SV processes. We present a somewhat different state space model, labeled the *conditional state space model*, which can be estimated by quasi maximum likelihood based on the *Extended Kalman filter*.

The outline is as follows. Section 2 briefly reviews the model and highlights the main traits of the multifactor affine SV process in discrete time. Section 3 explains how to incorporate the conditional distribution of the volatility-factor shocks $\{\mathbf{u}_t\}$ in the state space framework. We discuss the estimation method, comment on the inconsistency of QML based on the Extended Kalman filter, and provide the system matrices of the conditional state space model.

Section 4 reports the results of a Monte Carlo study towards the 1-factor CIR SV option pricing model, which is the Heston (1993) model without leverage effect. Central is the issue what data yields the best estimation results. As in the 1-factor

OU SV case, using squared returns only performs worst, the combinations of squared return and option data, and RV – option data best.

Section 5 confronts the 1-factor CIR SV model and associated estimation method to the FTSE100 stock-index and option data. As for the 1-factor OU SV model, we initially estimate the model using either only squared returns, or SM option data, or both. We contrast the results to the 1-factor OU results, and focus on the main differences. We consider compensation for FTSE100-index volatility risk, straddle returns, and perform a specification analysis. We show that modeling volatility feedback by 1-factor CIR SV is still not sufficient to capture the occasional periods of fast-changing volatility present in the data. Moreover, the model overprices the longer-maturity options out of sample. Estimating the 1-factor CIR SV model using all data shows that the model lacks sufficient volatility dynamics.

Section 6 presents estimation results for various 2-factor affine SV specifications. Estimating a specification with 1 OU and 1 affine SV factor reveals that the OU assumption is misspecified. The first factor also features volatility feedback. This contrasts to the findings of Chernov et al. (2003). Subsequently estimating a specification with 1 CIR and 1 affine SV factor yields a considerable improvement in fit. However, the model still lacks sufficient volatility dynamics.

Section 7 extends to 3-factor affine SV specifications. A specification with 1 OU and 2 affine SV factors appears misspecified: The long-memory, OU factor is characterized by level-dependent volatility as well. The most suitable model for the joint FTSE100 stock-index and option data within the affine class appears to be a 3-factor SV model with 1 CIR and 2 affine independent volatility factors. Not surprisingly, the interpretations of the factors are similar as in the 3-factor OU SV case, covered in the previous chapter. Nonetheless, extensive diagnostic checks reveal that the 3-factor affine SV derivative pricing model can still be improved upon in several ways. Section 8 summarizes. An appendix concludes this chapter.

2. Recap of the model

Let us review the main ingredients of the multifactor affine SV derivative pricing model. (See section 2 of chapter II for a full discussion.) The characteristic model equations are [1]

Stock price: $$dS_t = \mu_t S_t dt + \sigma_t S_t dW_{S,t} \qquad (\mathbb{P}) \qquad (2.1)$$

Stock variance: $$\sigma_t^2 = \mathbf{1}'\mathbf{x}_t, \quad d\sigma_t^2 = \mathbf{1}'d\mathbf{x}_t \qquad (2.2)$$

Latent factors: $$d\mathbf{x}_t = \mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t)dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t} \qquad (\mathbb{P}) \qquad (2.3)$$

$$d\mathbf{x}_t = \tilde{\mathbf{K}}(\tilde{\boldsymbol{\theta}} - \mathbf{x}_t)dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\tilde{\mathbf{W}}_{x,t} \qquad (\mathbb{Q}) \qquad (2.4)$$

$$\boldsymbol{\Lambda}_t = \text{diag}\left[\sqrt{\alpha_1 + \boldsymbol{\beta}_1'\mathbf{x}_t}, .., \sqrt{\alpha_n + \boldsymbol{\beta}_n'\mathbf{x}_t}\right] \qquad (2.5)$$

No leverage effect: $\mathbf{W} = (W_S, \mathbf{W}_x')'$ is a standard $\mathbb{P}$-Brownian motion

Risk-neutral parameters: $$\tilde{\mathbf{K}} = \mathbf{K}_d + \boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathcal{B}', \; \tilde{\boldsymbol{\theta}} = \tilde{\mathbf{K}}^{-1}(\mathbf{K}_d\boldsymbol{\theta} - \boldsymbol{\Sigma}\boldsymbol{\Gamma}\boldsymbol{\alpha}) \qquad (2.6)$$

Market price of vol risk: $$\boldsymbol{\gamma}_{x,t} = \boldsymbol{\Lambda}_t\boldsymbol{\gamma} \qquad (2.7)$$

[1] To simplify discussion, we already impose the identification restrictions $\delta_0 = 0$, $\boldsymbol{\delta} = \mathbf{1}$ in the stock variance equation $\sigma_t^2 = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t$ here.

Call option price: $$C_t = \mathbb{E}_{\mathbb{Q}}\left[BS(F_{t,T}, K_t, \tau_t, \bar{r}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t\right] \tag{2.8}$$

Forward price: $$F_{t,T} = S_t \exp\left[(\bar{r}_t - \bar{q}_t)\tau_t\right]. \tag{2.9}$$

Properties of the SV process

Given that we deal with our most general affine SV specification in this chapter, let us highlight some of the main discrete-time features of the stock variance process. Appendix B gives a more complete description. The unconditional mean and variance of the stock variance equal

$$\mathbb{E}_{\mathbb{P}}\left[\sigma_t^2\right] = \mathbf{1}'\boldsymbol{\theta}, \qquad \text{var}_{\mathbb{P}}\left[\sigma_t^2\right] = \mathbf{1}'\left[\mathbf{J} \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right]\mathbf{1}, \tag{2.10}$$

in which $(nxn)\mathbf{J}$ with $[\mathbf{J}]_{ij} = 1/(k_i + k_j)$ and $\mathbf{M}_d = \text{diag}[\alpha_1 + \boldsymbol{\beta}_1'\boldsymbol{\theta}, .., \alpha_n + \boldsymbol{\beta}_n'\boldsymbol{\theta}]$. (See also appendix A for the various matrix definitions). Suppose we live at day t. The best forecast of tomorrow's stock variance (i.e., at day $t + \Delta t$) given today's information equals

$$\mathbb{E}_{\mathbb{P}}\left[\sigma_{t+\Delta t}^2 \mid \mathcal{F}_t\right] = \mathbf{1}'\boldsymbol{\theta} + \mathbf{1}'\exp[-\mathbf{K}_d\Delta t]\,\mathbf{x}_t^*. \tag{2.11}$$

Equation (2.11) implies that the best forecast of the stock variance in the indefinite future equals $\mathbf{1}'\boldsymbol{\theta}$, the mean stock variance. This is a consequence of stationarity of the stock variance process. The conditional variance of tomorrow's stock variance equals

$$\text{var}_{\mathbb{P}}\left[\sigma_{t+\Delta t}^2 \mid \mathcal{F}_t\right] = \mathbf{1}'\left[\mathbf{G}(\Delta t) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' + \sum_{i=1}^{n}\left(\left\{\left[\boldsymbol{\beta}_i \odot \mathbf{x}_t^*\right]' \otimes \mathbf{I}_n\right\}\mathcal{H}(\Delta t)\right) \odot \boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'\right]\mathbf{1}, \tag{2.12}$$

in which matrix $(nxn)\mathbf{G}(\Delta t)$ has $[\mathbf{G}(\Delta t)]_{ij} = (1 - \exp[-(k_i + k_j)\Delta t])/(k_i + k_j)$, and $(n^2xn)\,\mathcal{H}(\Delta t) = [\mathcal{H}_1(\Delta t)', .., \mathcal{H}_n(\Delta t)']'$, in which the pq-th element of the individual matrix (nxn) $\mathcal{H}_j(\Delta t)$, $j = 1, .., n$ equals

$$\left[\mathcal{H}_j(\Delta t)\right]_{pq} = \frac{\exp[-k_j\Delta t] - \exp[-(k_p + k_q)\Delta t]}{k_p + k_q - k_j}. \tag{2.13}$$

The vector $(nx1)\,\boldsymbol{\sigma}_i$ in (2.12) represents the i-th column of $\boldsymbol{\Sigma}$ ($i = 1, .., n$), such that $\boldsymbol{\Sigma}$ is partitioned as $\boldsymbol{\Sigma} = [\boldsymbol{\sigma}_1, .., \boldsymbol{\sigma}_n]$.

The stock variance process is homoskedastic unconditionally. Conditionally it is heteroskedastic, except in the OU case for which $\boldsymbol{\beta}_i = \mathbf{0}\ \forall i$. Under OU SV periods of low and high volatility are possible, but the volatility-of-volatility is constant. In the general SV specification the volatility-of-volatility depends on the level of the volatility factors (see (2.12)). This typically allows for more extreme volatility clustering than under OU SV.

3. The conditional state space model and Extended Kalman filter QML

This section discusses the way in which we extract information from stock and option prices. We explain how we explicitly actuate level-dependent volatility-of-volatility in the state space framework, and discuss the estimation method, *Extended Kalman filter QML*. We also provide the system matrices of the conditional state space model.

3.1 Extracting information from stock and option data

The measurement equations for the squared returns, realized volatilities and BS implied volatilities are the same as before. To repeat, for $t = \Delta t, .., T\Delta t$, we extract information from stock returns via

$$\frac{1}{\Delta t}(r_t - \hat{\mu}_{t-\Delta t}\Delta t)^2 = \mathbf{1}'\boldsymbol{\theta} + \mathbf{1}'\mathbf{x}^*_{t-\Delta t} + \omega_t, \quad \omega_t \sim (0, \sigma^2_{\omega}), \tag{3.1}$$

and from realized volatilities via

$$\sigma^2_{RV,t} = \mathbf{1}'\boldsymbol{\theta} + \mathbf{1}'(\mathbf{K}_d\Delta t)^{-1}\left(\mathbf{I}_n - \exp[-\mathbf{K}_d\Delta t]\right)\mathbf{x}^*_{t-\Delta t} + \omega_{RV,t}, \; \omega_{RV,t} \sim (0, \sigma^2_{\omega,RV}). \tag{3.2}$$

From the option series $i = SM, MM, LM$, we extract information via

$$\sigma^2_{implied,it} = \mu_v + \frac{1}{\tau_{it}}\left[A_1(\tau_{it}) + \mathbf{B}_1(\tau_{it})'\boldsymbol{\theta}\right] + \frac{\mathbf{B}_1(\tau_{it})'}{\tau_{it}}\mathbf{x}^*_t + \varepsilon_{it}, \; \varepsilon_{it} \sim (0, \sigma^2_{\varepsilon i}). \tag{3.3}$$

Other than for OU SV, the general SV specification does not yield closed-form expressions for the maturity functions $(1x1)\, A_1(.)$ and $(nx1)\, \mathbf{B}_1(.)$, which depend on the risk-neutral parameters $\tilde{\mathbf{K}}$ and $\tilde{\boldsymbol{\theta}}$. [2] The system of Ricatti ODEs that they satisfy,

$$\frac{dA_1(\tau)}{d\tau} = \tilde{\boldsymbol{\theta}}'\tilde{\mathbf{K}}'\mathbf{B}_1(\tau) + \tfrac{1}{2}\sum_{i=1}^{n}\left[\boldsymbol{\Sigma}'\mathbf{B}_1(\tau)\right]_i^2 \alpha_i \tag{3.4}$$

$$\frac{d\mathbf{B}_1(\tau)}{d\tau} = -\tilde{\mathbf{K}}'\mathbf{B}_1(\tau) + \tfrac{1}{2}\sum_{i=1}^{n}\left[\boldsymbol{\Sigma}'\mathbf{B}_1(\tau)\right]_i^2 \boldsymbol{\beta}_i + \mathbf{1},$$

can be solved numerically using the boundary conditions $A_1(0) = 0$ and $\mathbf{B}_1(0) = \mathbf{0}$. We use Euler's method, which replaces differentials with differences. [3] As before, we take μ_v the same for each option series as they are all ATMF. Given the Monte Carlo evidence discussed in section 4 (and in chapter IV), we next set μ_v equal to zero. Cross-sectional heteroskedasticity is modeled by allowing the option error series $\{\varepsilon_{SM,t}\}$, $\{\varepsilon_{MM,t}\}$, $\{\varepsilon_{LM,t}\}$ to have their own variance. We may also allow for possible non-zero contemporaneous correlation between these series. We assume $\{\varepsilon_{it}\}$ to be uncorrelated with the other error series that appear in the state space model, i.e. with $\{\omega_t\}, \{\omega_{RV,t}\}$ and $\{\mathbf{u}_t\}$.

Exploiting level-dependent volatility-of-volatility

In discrete time, the hidden factors in deviation from their mean evolve as

$$\mathbf{x}^*_{t+\Delta t} = \exp[-\mathbf{K}_d\Delta t]\mathbf{x}^*_t + \mathbf{u}_{t+\Delta t}, \tag{3.5}$$

in which $(nx1)\, \mathbf{u}_{t+\Delta t}$ is interpreted as the vector of unpredictable shocks to the volatility factors from day t to day $t+\Delta t$. Both the unconditional and conditional distribution of $\mathbf{u}_{t+\Delta t}$ is non-Gaussian, except in the OU case. Unconditionally, the volatility-factor shocks are white noise with

$$\mathbf{u}_{t+\Delta t} \sim \left(\mathbf{0},\, \mathbf{G}(\Delta t) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right), \tag{3.6}$$

[2] This even holds for the 1-factor CIR SV case. We refer to section 14 of appendix B for a discussion.

[3] To be more specific, for each iteration of the loglikelihood optimization we evaluate these functions on the interval $[0, \max_{i,t} \tau_{it}]$ using a grid of 5000 intermediate points.

whereas their conditional distribution satisfies (see appendix B, section 6)

$$\mathbb{E}_{\mathbb{P}}\left[\mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t\right] = \mathbf{0}\,, \tag{3.7}$$
$$\text{var}_{\mathbb{P}}\left[\mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t\right] = \mathbf{G}(\Delta t) \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}' + \sum_{i=1}^{n}\left(\left\{\left[\boldsymbol{\beta}_i \odot \mathbf{x}_t^*\right]' \otimes \mathbf{I}_n\right\}\mathcal{H}(\Delta t)\right) \odot \boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'.$$

Moreover, recall the assumptions $\text{cov}_{\mathbb{P}}[\mathbf{u}_t, \omega_s] = \mathbf{0}\ \forall t,s$, $\text{cov}_{\mathbb{P}}[\mathbf{u}_t, \omega_{RV,s}] = \mathbf{0}\ \forall t \neq s$ and $\text{corr}_{\mathbb{P}}[\mathbf{u}_t, \omega_{RV,t}] = \mathbf{c}$.

The unconditional state space models stated in section 3.5 of chapter III and so far worked with, are based on the unconditional moments of $\mathbf{u}_{t+\Delta t}$. As such, they ignore the fact that the magnitude of tomorrow's volatility-factor shocks is linked to the level of the factors today, which is clear from (3.7). Observed periods of volatility clustering yield a rich source of information on the conditional dynamics of the volatility process. Hence, more efficient use of information is made if the conditional SV dynamics are explicitly acknowledged in the estimation method.

3.2 Conditional state space model and Extended Kalman filter

So how can we incorporate the conditional SV dynamics? In more general terms, how can one exploit conditional distributions in the linear state space framework? Recall first the general linear (unconditional) state space representation discussed in section 2.2 of chapter III. For $t = \Delta t, .., T\Delta t$ it reads

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{a}_t + \boldsymbol{H}_t'\boldsymbol{\xi}_t + \boldsymbol{w}_t\,, & \boldsymbol{w}_t &\sim (\mathbf{0}, \boldsymbol{R})\,, \\ \boldsymbol{\xi}_{t+\Delta t} &= \boldsymbol{d} + \boldsymbol{F}\boldsymbol{\xi}_t + \boldsymbol{v}_{t+\Delta t}\,, & \boldsymbol{v}_{t+\Delta t} &\sim (\mathbf{0}, \boldsymbol{Q})\,, \end{aligned} \tag{3.8}$$

in which $(mx1)\,\boldsymbol{y}_t$ contains the observable variables, $(rx1)\,\boldsymbol{\xi}_t$ is the (partly) unobserved state vector, and in which the white noise error series $\{\boldsymbol{w}_t\}$ and $\{\boldsymbol{v}_{t+\Delta t}\}$ are uncorrelated at all points in time. This representation can be analyzed by the Kalman filter, which yields the linear projection of the state vector on the data. Estimation is by Kalman filter QML. No further knowledge on the conditional distributions of either $\{\boldsymbol{w}_t\}$ or $\{\boldsymbol{v}_{t+\Delta t}\}$ is required for this.

For our specific purposes however, we would like to incorporate and exploit in this framework, the following additional information on the conditional distribution of the state innovation $\boldsymbol{v}_{t+\Delta t}$,

$$\boldsymbol{v}_{t+\Delta t} \mid \mathcal{F}_t \sim (\mathbf{0}, \boldsymbol{Q}_t)\,, \tag{3.9}$$

in which $\boldsymbol{Q}_t = \boldsymbol{Q}_t(\boldsymbol{\xi}_t)$ is state dependent. (Notice that $\boldsymbol{Q} = \mathbb{E}[\boldsymbol{Q}_t]$, i.e., the unconditional innovation variance is its expected conditional variance.) We will refer to the dynamic system (3.8)-(3.9) as the *conditional state space model*. In order to achieve this, write

$$\boldsymbol{v}_{t+\Delta t} \equiv \boldsymbol{Q}_t^{\frac{1}{2}}\,\boldsymbol{\zeta}_{t+\Delta t},\ (rx1)\,\boldsymbol{\zeta}_{t+\Delta t} \mid \mathcal{F}_t \sim (\mathbf{0}, \mathbf{I}_r) \sim \boldsymbol{\zeta}_{t+\Delta t}\,. \tag{3.10}$$

Now notice that if $\boldsymbol{Q}_t$ were not state dependent but instead deterministic, we would effectively be back in the original framework (3.8) with $\boldsymbol{Q} = \boldsymbol{Q}_t$; that is, the unconditional and conditional variance would coincide in that case.

Extended Kalman filter
However, $\boldsymbol{Q}_t$ *is* state dependent. As Harvey (1989, p.161) argues, obtaining an optimal filter for this type of model is not possible in general. Nevertheless, one may resort to an approximate filter known as the *Extended Kalman filter*. In our conditional volatility dynamics setting, the Extended Kalman filter essentially amounts to replacing the matrix $\boldsymbol{Q}$ with the matrix $\boldsymbol{Q}_t$ at time t in the Kalman filter and smoother recursions (2.6) and (2.7), given in section 2.2 of chapter III. [4] The state $\boldsymbol{\xi}_t$ that appears in $\boldsymbol{Q}_t = \boldsymbol{Q}_t(\boldsymbol{\xi}_t)$ is estimated by the *contemporaneously filtered state* $\boldsymbol{\xi}_{t|t}$. In the Gaussian state space framework, $\boldsymbol{\xi}_{t|t}$ is defined as $\boldsymbol{\xi}_{t|t} \equiv \mathbb{E}[\boldsymbol{\xi}_t \mid \boldsymbol{Y}_t]$ and has associated contemporaneously filtered state variance matrix $\boldsymbol{P}_{t|t} \equiv \text{var}[\boldsymbol{\xi}_t \mid \boldsymbol{Y}_t]$. Unlike the regular filtered state $\boldsymbol{\xi}_{t|t-1}$, which uses information until time $t-1$ to predict the state at time t, the contemporaneously filtered state $\boldsymbol{\xi}_{t|t}$ uses information up to and including time t. Harvey (1989, p.162) mentions that the quality of the Extended Kalman filter depends on the accuracy of $\boldsymbol{\xi}_{t|t}$ as an estimator of the state $\boldsymbol{\xi}_t$.

$\boldsymbol{\xi}_{t|t}$, $\boldsymbol{P}_{t|t}$ and $\boldsymbol{Q}_t$ can be computed by extending the Kalman filter recursion (2.6) in section 2.2 of chapter III with the following equations:

$$\begin{aligned} \boldsymbol{\xi}_{t|t} &= \boldsymbol{\xi}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t\boldsymbol{E}_t^{-1}\boldsymbol{e}_t \\ \boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t\,\boldsymbol{E}_t^{-1}\,\boldsymbol{H}_t{}'\boldsymbol{P}_{t|t-1} \\ \boldsymbol{Q}_t &= \boldsymbol{Q}_t(\boldsymbol{\xi}_{t|t}). \end{aligned} \tag{3.11}$$

These equations ought to be computed just prior to the prediction equations (which calculate $\boldsymbol{\xi}_{t+1|t}$ and $\boldsymbol{P}_{t+1|t}$). The equations for $\boldsymbol{\xi}_{t|t}$ and $\boldsymbol{P}_{t|t}$ are known as the *updating equations.*

Estimation
The conditional state space model can be analyzed by the Extended Kalman filter if the model parameters are known. It generates an estimate of the latent state vector given the data. If the parameters are unknown, they may be estimated by quasi maximum likelihood as before, but by using the Extended Kalman filter equations. We will refer to this method as *Extended Kalman filter QML*.

Inconsistency of Extended Kalman filter QML
Recall that in linear Gaussian state space models, the Kalman filter ML estimator is consistent, asymptotically normal and (asymptotically) efficient. In linear non-Gaussian state space models [5] such as (3.8), the Kalman filter QML estimator still yields consistent and asymptotically normal estimates (under sufficient regularity; see e.g. Hamilton (1994)). The estimates are not efficient however. The Extended Kalman filter QML estimator of the linear non-Gaussian conditional state space model (3.8)-(3.9) is not even consistent. The reason is that the state $\boldsymbol{\xi}_t$ in $\boldsymbol{Q}_t(\boldsymbol{\xi}_t)$ is not known exactly, but is estimated by $\boldsymbol{\xi}_{t|t}$, and is therefore measured with error.

The obvious issue is how large the inconsistency of the Extended Kalman filter QML estimator is in our specific setting. Exact results seem impossible to derive. As mentioned in chapter III, to estimate affine models of the term structure of interest rates, Extended Kalman filter QML has frequently been advocated as a favorable estimation method; see e.g., Lund (1997), de Jong (2000) and Duffee

[4] The unconditional variance matrix $\boldsymbol{Q}$ is however still needed to initialize the Kalman filter recursion.
[5] See Durbin and Koopman (2001) for numerous examples and an extensive discussion of these models.

and Stanton (2001). The Monte Carlo results reported in de Jong (2000) confirm the results of Lund (1997), who suggests that for the typical parameter values found in estimates of affine interest rate models, the bias in this estimator is small. Duffee and Stanton (2001) report similar evidence.

Although we are not dealing with affine interest rate models here, our affine SV setting and estimation method show obvious similarities. There is one important difference however: Unlike in the interest rate setting, in our SV setting we had to perform several approximations to arrive at a linear state space model. This motivates the investigation of the bias induced by the approximations, by performing a Monte Carlo study towards the 1-factor CIR SV model in section 4.

3.3 System matrices of the conditional state space model

The set up of the conditional state space model (3.8)-(3.9) is given in general terms. The measurement equations (3.1)-(3.3) used for information extraction and the equations for the discrete-time volatility-factors evolution (3.5)-(3.7) can be cast in this framework. The resulting system matrices of the conditional state space model are identical to those of the unconditional model; we refer to (3.85) and (3.87) in section 3.5 of chapter III for details. The difference lies in the introduction of an additional system matrix, $\boldsymbol{Q}_t = \text{var}_{\mathbb{P}}[\boldsymbol{\upsilon}_{t+\Delta t} \mid \mathcal{F}_t]$.

If squared returns are included for estimation

In case squared returns or the combination squared return - option data is used for estimation, $\boldsymbol{Q}_t = \text{var}_{\mathbb{P}}[(\omega_{t+\Delta t}, \mathbf{u}_{t+\Delta t}')' \mid \mathcal{F}_t]$. The expression for $\text{var}_{\mathbb{P}}[\mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t]$ is given in (3.7). Based on an Euler discretization of the stock price SDE, it follows that $\text{var}_{\mathbb{P}}[\omega_{t+\Delta t} \mid \mathcal{F}_t] = 2\sigma_t^4 = 2(\delta_0 + \boldsymbol{\delta}'\mathbf{x}_t)^2$ (and hence $\text{var}_{\mathbb{P}}[\omega_t] = 2\mathbb{E}_{\mathbb{P}}[\sigma_t^4]$), which is *non-linear* in $\mathbf{x}_t$. As this is not an exact result however, we choose to leave $\text{var}_{\mathbb{P}}[\omega_{t+\Delta t} \mid \mathcal{F}_t]$ unrestricted and take it equal to σ_ω^2 for simplicity, which (using a similar argument) we took to be equal to $\text{var}_{\mathbb{P}}[\omega_t]$. Hence, we assume that $\text{var}_{\mathbb{P}}[\omega_{t+\Delta t} \mid \mathcal{F}_t] = \sigma_\omega^2 = \text{var}_{\mathbb{P}}[\omega_{t+\Delta t}]$. Given that $\omega_{t+\Delta t}$ represents measurement error in the squared returns which are noisy, we expect this simplification to not have any major impact, though it evidently simplifies matters greatly. Based on an Euler discretization, $\text{cov}_{\mathbb{P}}[\omega_{t+\Delta t}, \mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t] = \mathbf{0}$. The Monte Carlo evidence reported in section 4.3.2 shows that this is a reasonable assumption. The matrix $\boldsymbol{Q}_t$ then becomes

$$\boldsymbol{Q}_t = \begin{bmatrix} \sigma_\omega^2 & \mathbf{0}' \\ \mathbf{0} & \text{var}_{\mathbb{P}}\left[\mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t\right] \end{bmatrix}. \tag{3.12}$$

The appendix states the system matrices in case the squared returns, the SM, MM and LM ATMF option series are jointly used for estimation.

If RV data is included for estimation

In case RV data or the combination RV - option data is used for estimation, $\boldsymbol{Q}_t = \text{var}_{\mathbb{P}}[(\omega_{RV,t+\Delta t}, \mathbf{u}_{t+\Delta t}')' \mid \mathcal{F}_t]$, with $\text{var}_{\mathbb{P}}[\mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t]$ given in (3.7). As the measurement equation for realized volatility is not exact, we make the following (dramatically) simplifying assumptions, $\text{var}_{\mathbb{P}}[\omega_{RV,t+\Delta t} \mid \mathcal{F}_t] = \sigma_{\omega,RV}^2 = \text{var}_{\mathbb{P}}[\omega_{RV,t+\Delta t}]$ and $\text{corr}_{\mathbb{P}}[\omega_{RV,t+\Delta t}, \mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t] = \mathbf{c} = \text{corr}_{\mathbb{P}}[\omega_{RV,t+\Delta t}, \mathbf{u}_{t+\Delta t}]$, such that

$$\boldsymbol{Q}_t = \begin{bmatrix} \sigma_{\omega,RV}^2 & \sigma_{\omega,RV}\mathbf{c}'\left(\mathbf{I}_n \odot \text{var}_{\mathbb{P}}\left[\mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t\right]\right)^{\frac{1}{2}} \\ \sigma_{\omega,RV}\left(\mathbf{I}_n \odot \text{var}_{\mathbb{P}}\left[\mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t\right]\right)^{\frac{1}{2}}\mathbf{c} & \text{var}_{\mathbb{P}}\left[\mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t\right] \end{bmatrix}. \tag{3.13}$$

(Note that although the conditional correlation is assumed constant, the conditional contemporaneous covariance between $\{\omega_{RV,t}\}$ and $\{\mathbf{u}_t\}$ is time varying.)

4. A Monte Carlo study to the 1-factor CIR SV model

This section performs a Monte Carlo study towards the 1-factor CIR SV option pricing model. Section 4.1 considers the 1-factor CIR SV specification and discusses some implied properties of the volatility process. Section 4.2 discusses the assumptions underlying the simulated data, and outlines the simulation strategy. Section 4.3 considers the quality of the linear and parabolic call price approximations, and investigates the assumptions underlying the error terms of the state space model. Section 4.4 examines the performance of the estimation method based on (combinations of) squared return, RV and option data.

4.1 1-factor CIR SV specification

In the 1-factor CIR SV case, the SV specification (2.2)-(2.6) reduces to [6]

$$\begin{aligned} \sigma_t^2 &= x_t && (4.1) \\ dx_t &= k(\theta - x_t)dt + \sigma\sqrt{x_t}dW_{x,t} && (\mathbb{P}) \\ dx_t &= \tilde{k}(\tilde{\theta} - x_t)dt + \sigma\sqrt{x_t}d\tilde{W}_{x,t}, && (\mathbb{Q}) \end{aligned}$$

in which the risk-neutral parameters are given by $\tilde{k} = k + \sigma\gamma$ and $\tilde{\theta} = k\theta/(k + \sigma\gamma)$. In contrast to the OU case, the market price of volatility risk (2.7) is now time-varying and given by $\gamma_{x,t} = \gamma\sqrt{x_t}$. The volatility risk premium equals $\sigma\gamma x_t$ and averages at $\sigma\gamma\theta$. The SV option pricing model is fully determined by four parameters only: θ, k, σ and γ.

The 1-factor CIR SV specification precludes negative stock variances if $k, \theta, \sigma_0^2 > 0$. As Brownian sample paths are continuous, so are the paths $\{\sigma_t^2\}$ follows. The stock variance therefore first has to become zero for it to possibly become negative. But if it becomes zero, then the diffusion component of the SDE drops out, and hence there is temporarily no stochastic variation in the stock variance. As both the mean θ of the process, and the mean-reversion parameter k are positive, the process will then naturally be pulled away from zero towards θ. As such, the stock variance will never become negative. If the Feller condition $2k\theta \geq \sigma^2$ holds (Feller (1951)), the upward drift in the CIR factor is sufficiently large for the factor to never exactly reach zero.

The main properties of the CIR SV process are the following (see also section 14 of appendix B). The mean stock variance is given by $\mathbb{E}_{\mathbb{P}}[\sigma_t^2] = \mathbb{E}_{\mathbb{P}}[x_t] = \theta$, such that the unconditional volatility of returns equals $\sqrt{\theta}$ (per annum, as we measure time in years). Moreover, $\text{var}_{\mathbb{P}}[\sigma_t^2] = \text{var}_{\mathbb{P}}[x_t] = \sigma^2\theta/2k$. The (unconditional) distribution of $\{\sigma_t^2\}$ is non-Gaussian, and has fatter tails than the normal distribution. Parameter k governs the speed at which the stock variance reverts back to its mean θ. Using $\sigma_t^2 = x_t = x_t^* + \theta$ and from (3.5)

[6] The *square root* process was first imported in finance by Cox et al. (1985), who proposed it as a model for the short interest rate. Heston (1993) was the first to adopt this process as a model for stock volatility. The 1-factor CIR SV model is obtained from imposing $n = 1$, $\alpha = 0$ and $\beta = 1$ in the general multifactor SV model.

$x_t^* = \exp(-k\Delta t)x_{t-\Delta t}^* + u_t$, the stock variance at day $t+\Delta t$ is related to the stock variance of day t via

$$\sigma_{t+\Delta t}^2 = \theta\left[1-\exp(-k\Delta t)\right] + \exp(-k\Delta t)\,\sigma_t^2 + u_{t+\Delta t}\,. \qquad (\mathbb{P}) \qquad (4.2)$$

If $k=0$ then $\sigma_{t+\Delta t}^2 = \sigma_t^2 + u_{t+\Delta t}$, such that the stock variance follows a random walk in discrete time, with shocks that persist for ever. The volatility persistence is measured by $\exp(-k\Delta t)$ on a daily basis. The half-life (in days) of a shock in the stock variance is given by $\ln 2 / k\Delta t$. In equation (4.2), $u_{t+\Delta t}$ represents the unpredictable shock in the stock variance from day t to day $t+\Delta t$. In the CIR SV case it holds that [7]

$$u_{t+\Delta t} \sim \left(0;\ \sigma^2\theta\left[\frac{1-\exp(-2k\Delta t)}{2k}\right]\right) \qquad (\mathbb{P}) \qquad (4.3)$$

$$u_{t+\Delta t} \mid \mathcal{F}_t \sim \left(0;\ \sigma^2\theta\left[\frac{1-\exp(-2k\Delta t)}{2k}\right] + \sigma^2\, x_t^*\left[\frac{\exp(-k\Delta t)-\exp(-2k\Delta t))}{k}\right]\right).$$

The volatility shock $u_{t+\Delta t}$ is thus unconditionally homoskedastic, but conditionally heteroskedastic, with a variance that depends on the current volatility-determining state x_t^*. The volatility-of-volatility is indeed level-dependent.

4.2 Simulating from the 1-factor CIR SV option pricing model

Assumptions

As usual, we assume 260 trading days per annum and denote the timing of the daily data points by $t = \Delta t, 2\Delta t, .., T\Delta t$, with $\Delta t = 1/260$. We simulate for a total of $T = 1058$ daily observations, corresponding to the UK data used in our empirical work. As before, we assume $S_0 = 5300$, $\mu_t = \mu = 8.25\%$, $q_t = q = 3.5\%$ and $r_t = r = 6\%$. (These values correspond to the information in the FTSE100-index data for the period 6 Oct 1997 – 28 Dec 2001, and were also used in the 1-factor OU SV Monte Carlo study.)

We simulate time series of squared returns, 10-minute realized volatilities (i.e. 48 intraday time points per trading day of eight hours) and short-maturity (close-to) ATMF call options. The maturity and moneyness of the simulated option series exactly match those of the empirical short-maturity ATMF FTSE100-index option series, for each t. The average option maturity is 1.4 months. (See section 2.1 of chapter IV for details.) Besides permitting a consistent comparison between the real-world and simulated call data, in practice, short-maturity ATMF calls are typically most liquid, are expected to suffer least from non-synchronicity biases, have maximal vega, and hence contain the most valuable volatility information. Recall that we advocate the use of such a series if our estimation method is to be used in practice. (Again, this is similar as in the OU Monte Carlo study.)

A decision needs to be made on the parameter values with which to simulate data from the 1-factor CIR SV model. We consider five "realistic" parameter sets, numbered I – V, given in table 4.1. The values are largely based on empirical estimates of the Heston (1993) model found in the literature. (The 1-factor CIR SV model coincides with the Heston (1993) model, except that it does not model

[7] In contrast, in the 1-factor OU SV case: $u_{t+\Delta t} \sim u_{t+\Delta t} \mid \mathcal{F}_t \sim \left(0;\ \sigma^2\left[\frac{1-\exp(-2k\Delta t)}{2k}\right]\right)$.

the leverage effect.) We refer to section 5.1 of the next chapter for a more complete motivation of these values.

Table 4.1: Parameter sets

Parameter set:	*I*	*II*	*III*	*IV*	*V*
θ	0.04	0.04	0.04	0.04	0.02
k	1.50	2.61	5.25	7	7
σ	0.20	0.30	0.45	0.60	0.30
γ	-4	-5	-6	-7	-20
Volatility of returns	20%	20%	20%	20%	14.14%
Persistence	0.9942	0.9900	0.9800	0.9734	0.9734
Std.dev. x_t	0.0231	0.0263	0.0278	0.0321	0.0113
$\tilde{k}$	0.700	1.11	2.55	2.80	1.00
$\tilde{\theta}$	0.0857	0.0941	0.0824	0.100	0.140
Av.vol. risk premium	-3.2%	-6%	-10.8%	-16.8%	-12%

The table reports five parameter sets with implied quantities that will be used for simulating data from the 1-factor CIR SV model (which assumes no leverage effect).

Set IV closely corresponds to the empirical estimates we find in section 5.1 for the 1-factor CIR SV model. We choose θ (mean stock variance) the same for parameter sets I - IV. We intuitively expect the bias of its estimate to be only weakly dependent on its magnitude. This allows us to focus (when moving from set I to IV) on the effects on bias of a gradual increase in mean-reversion (k), accompanied by increasing volatility-of-volatility (σ), which is often observed in practice (e.g., Jones (2003)). Set V largely corresponds to Pan's (2002) findings.

Set up of the Monte Carlo experiment

To save space, details on the set up of the Monte Carlo experiment are given in section 5.1.2 of the next chapter, which outlines how to obtain a simulated dataset of the Heston (1993) model. As imposing the restriction $\rho = 0$ (no leverage effect) in the Heston (1993) model results in the 1-factor CIR SV model, the steps to follow are readily obtained. We therefore refrain from duplication here.

To keep the amount of computations to a reasonable level, we simulate a total of 150 datasets for each set of parameter values I - V. (To simulate 150 datasets takes 41 hours with the program Ox on a Pentium-III computer. Again, see section 5.1.2 of the next chapter for additional discussion.)

4.3 Simulation results for 1-factor CIR SV: Call price approximations and error terms state space model

Section 4.3.1 examines the quality of the linear and parabolic call price approximations, the magnitude of μ_v and the relative merits measure. Section 4.3.2 examines the statistical properties assumed for the error terms of the state space model. We emphasize what a possible consequence is of the correlation found in the option error series regarding diagnostic checking in practice.

4.3.1 Quality of the call price approximations

Linearization and parabolic errors

From section 3.4 of chapter III, recall the linearized call option price formula: $C_t = g(b_t^*) - b_t^* g'(b_t^*) + \mathbb{E}_{\mathbb{Q}}\left[Y_t \mid \mathcal{F}_t\right] g'(b_t^*) + \mathbb{E}_{\mathbb{Q}}\left[HOT_t \mid \mathcal{F}_t\right]$. To arrive at a linear options measurement equation, we proposed to neglect the linearization error

$\mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]$. We labeled $\mathbb{E}_{\mathbb{Q}}[HOT2_t \mid \mathcal{F}_t]$ the parabolic error. This error appears in expression $C_t = q_0(b_t^*) + q_1(b_t^*)\mathbb{E}_{\mathbb{Q}}\left[Y_t \mid \mathcal{F}_t\right] + q_2(b_t^*)\mathbb{E}_{\mathbb{Q}}\left[Y_t^2 \mid \mathcal{F}_t\right] + \mathbb{E}_{\mathbb{Q}}\left[HOT2_t \mid \mathcal{F}_t\right]$, being a parabolic approximation of the call price.

Table 4.2 provides the means of the (absolute value of the) relative linearization and parabolic errors with standard deviations in parentheses, over the full sample of 150*1058 observations, for each parameter set. To interpret the percentages in perspective: The average simulated call price equals 163 (std.dev. 68), 164 (std.dev. 71), 169 (std.dev. 71), 173 (std.dev. 76) and 132 (std.dev. 44) for parameter sets I, II, III, IV and V respectively. [8]

Table 4.2: Summary statistics (absolute value of) relative linearization and parabolic errors

	$\mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]/C_t$	$\lvert \mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]/C_t \rvert$	$\mathbb{E}_{\mathbb{Q}}[HOT2_t \mid \mathcal{F}_t]/C_t$	$\lvert \mathbb{E}_{\mathbb{Q}}[HOT2_t \mid \mathcal{F}_t]/C_t \rvert$
I	-0.55% (0.40%)	0.55% (0.40%)	0.02% (0.23%)	0.18% (0.15%)
II	-1.13% (0.70%)	1.13% (0.70%)	0.07% (0.34%)	0.26% (0.22%)
III	-1.89% (0.86%)	1.89% (0.86%)	0.21% (0.45%)	0.38% (0.32%)
IV	-2.94% (1.17%)	2.94% (1.17%)	0.53% (0.64%)	0.64% (0.52%)
V	-1.42% (0.59%)	1.42% (0.59%)	0.12% (0.37%)	0.30% (0.24%)

The table reports the mean and standard deviation (in parentheses) of the (absolute value of the) relative linearization and parabolic errors, for each parameter set I - V.

The errors seem rather small in general. (Of course, it is difficult to define what exactly is "small" in this context. The ultimate test is eventually how our state space estimation method performs.) The linearization error is typically negative, the parabolic error positive. A linearization thus typically overestimates the true call price, a parabolic approximation generally underestimates the call price. The pattern in the errors seems to indicate that the errors become larger the less the volatility persistence, and the more the volatility fluctuates (recall table 4.1).

"Parameter" μ_v

From Jensen's inequality, we showed in section 3.4.2 of chapter III that $\sigma^2_{implied,t} \le \frac{1}{\tau_t}[A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\mathbf{x}_t]$. This result provided a second rationale for the options measurement equation as $\sigma^2_{implied,t} = \mu_v + \frac{1}{\tau_t}[A_1(\tau_t) + \mathbf{B}_1(\tau_t)'\mathbf{x}_t] + \varepsilon_t$, with $\varepsilon_t \sim (0, \sigma^2_\varepsilon)$. We outlined how to estimate "parameter" μ_v based on a simulated sample of T observations.

Table 4.3: Summary statistics of the 150 μ_v-estimates and relative merits measures

	μ_v		*μ_v as % of mean BS impl.var.*		*Relative merits measure*	
	Mean	(std.dev.)	Mean	(std.dev.)	Mean	(std.dev.)
I.	-0.00035	(1.1e-5)	-0.89%	(0.22%)	0.37%	(0.12%)
II.	-0.00073	(2.6e-5)	-1.78%	(0.37%)	0.87%	(0.19%)
III.	-0.00138	(5.3e-5)	-3.16%	(0.40%)	1.51%	(0.16%)
IV.	-0.00226	(9.5e-5)	-4.91%	(0.56%)	2.31%	(0.17%)
V.	-0.00066	(1.8e-5)	-2.54%	(0.22%)	1.11%	(0.09%)

[8] The average Monte Carlo standard error of the simulated call price as a percentage of the call price equals 0.23% (std.dev. 0.07%), 0.32% (std.dev. 0.09%), 0.43% (std.dev. 0.09%), 0.53% (std.dev. 0.11%) and 0.37% (0.06%) for sets I–V respectively. Moreover, the mean simulated stock price S_t equals 6134 (std.dev. 1841), 6135 (std.dev. 1822), 6143 (std.dev. 1832), 6159 (std.dev. 1850) and 6197 (std.dev. 1381) for sets I-V respectively.

Summary statistics of the 150 estimates of μ_v for each parameter set I – V are in table 4.3. As expected, μ_v is negative for all datasets. (We checked this for all datasets). Its magnitude is close to zero. Its relative magnitude (i.e. μ_v as a percentage of the average $\sigma^2_{implied,t}$ over the sample) seems also rather modest.

Relative merits of using a parabolic vs. a linear approximation
In section 3.4.5 of chapter III, we introduced the relative merits measure as an indicative measure to examine if a parabolic (rather than a linear) approximation is really worth the additional complexity in the computations. We defined it as $\frac{1}{T}\sum_t^T\left[|\,\mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]/C_t\,| - |\,\mathbb{E}_{\mathbb{Q}}[HOT2_t \mid \mathcal{F}_t]/C_t\,|\right]$ *100%, i.e. as the percentual difference between the mean absolute relative linearization and parabolic errors.

As table 4.3 shows, the average of the 150 relative merits measures for each of the parameter sets I-V does not seem particularly large, except perhaps for parameter set IV. Although this indicates that it may be worthwhile to consider a parabolic approximation, in the remainder we focus on the simpler, linear approximation, and leave the parabolic approximation for future research.

4.3.2 Disturbance terms of the state space model

The validity of the Kalman filter equations depends on the white noise assumption for the different disturbance terms that appear in the state space model, i.e. $\{u_t\}$, $\{\omega_{RV,t}\}$, $\{\omega_t\}$ and $\{\varepsilon_t\}$. We were able to prove that the volatility-factor shock series $\{u_t\}$ is indeed white noise. But what about the other error terms?

Error series $\{\omega_{RV,t}\}$ (RV equation)
Recall section 3.3 of chapter III on RV. In deriving the equation for the average variance as $\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_u^2du = \mathbb{E}_{\mathbb{P}}[\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_u^2du \mid \mathcal{F}_t] + \varpi_{RV,t+\Delta t}$ over day $[t, t+\Delta t]$, we proved that the series $\{\varpi_{RV,t}\}$ is white noise. Based on this equation, we next proposed a measurement equation for RV in which we assumed its associated error series $\{\omega_{RV,t}\}$ to be white noise: $\omega_{RV,t} \sim (0, \sigma^2_{\omega,RV})$. Although reasonable, this remains an assumption, as $\sigma^2_{RV,t+\Delta t}$ measures $\frac{1}{\Delta t}\int_t^{t+\Delta t}\sigma_u^2du$ with error. The simulated data allows us to examine this assumption.

The simulated data confirm that considering $\{\omega_{RV,t}\}$ as white noise is indeed reasonable. Table 4.4 reports the mean of $\omega_{RV,t}$ over all 150 datasets for each set I-V. The means are virtually equal to zero. The $\{\omega_{RV,t}\}$-series appear neither significantly autocorrelated, nor heteroskedastic unconditionally in general.[9]

Table 4.4: Statistics associated with error series $\{\omega_{RV,t}\}$ (RV equation)

	$Mean\{\omega_{RV,t}\}$	corr($\omega_{RV,t}, u_t$)	corr($\omega_{RV,t}, u_{t-\Delta t}$)	corr($\omega_{RV,t}, u_{t-2\Delta t}$)
I.	1.9e-6	0.14 (0.04)	0.00 (0.03)	0.00 (0.04)
II.	1.5e-5	0.19 (0.04)	0.00 (0.04)	0.00 (0.04)
III.	1.9e-5	0.27 (0.04)	0.00 (0.04)	0.00 (0.04)
IV.	2.0e-5	0.33 (0.05)	0.00 (0.05)	0.00 (0.04)
V.	4.6e-6	0.26 (0.04)	0.00 (0.04)	0.00 (0.04)

The table reports the mean of $\{\omega_{RV,t}\}$, the contemporaneous and "lagged" correlations between $\{\omega_{RV,t}\}$ and $\{u_t\}$, averaged over all 150 datasets (std.dev. in (.)), for each parameter set I - V.

We proved the series $\{\varpi_{RV,t}\}$ and $\{u_t\}$ to be contemporaneously correlated only, with approximate correlation coefficient 0.86. Based on this result, we allow for

[9] We checked this in EViews for several datasets. To save space we summarize our findings verbally.

contemporaneous correlation between $\{\omega_{RV,t}\}$ and $\{u_t\}$ in the state space model. As table 4.4 shows, the simulated data supports this assumption. The average contemporaneous correlations are much smaller than 0.86, because the number of intraday sampling points I equals 48 only (and not infinity).

Error series $\{\omega_t\}$ (Squared return equation)
Treating the Euler discretization of the stock price SDE as exact, we proved that $\{\omega_t\}$ is white noise, and is uncorrelated with $\{u_t\}$. However, as the discretization is not fully exact, these are actually merely assumptions.

The simulated data confirm that considering $\{\omega_t\}$ as white noise is indeed reasonable. Table 4.5 reports the mean of ω_t over all 150 datasets, for each parameter set I-V. The means are close to zero. The non-autocorrelatedness and homoskedasticity assumptions appear generally supported by the simulated data as well. Table 4.5 also reports the contemporaneous, and first and second-order lagged correlation coefficients between $\{\omega_t\}$ and $\{u_t\}$, averaged over all 150 datasets, for each set I-V. The contemporaneous correlation coefficients center near zero; the lagged correlations at zero. Assuming $\{\omega_t\}$ and $\{u_t\}$ to be uncorrelated seems reasonable.

Table 4.5: Statistics associated with error series $\{\omega_t\}$ (Squared return equation)

	Mean $\{\omega_t\}$	corr(ω_t, u_t)	corr($\omega_t, u_{t-\Delta t}$)	corr($\omega_t, u_{t-2\Delta t}$)
I.	-0.00026	0.02 (0.03)	0.00 (0.03)	0.00 (0.04)
II.	-0.00010	0.03 (0.04)	0.00 (0.04)	0.00 (0.04)
III.	-0.00010	0.04 (0.04)	0.00 (0.04)	0.00 (0.04)
IV.	-0.00004	0.05 (0.04)	0.00 (0.04)	0.00 (0.04)
V.	-0.00011	0.04 (0.03)	0.00 (0.04)	0.00 (0.04)

The table reports the mean of $\{\omega_t\}$, the contemporaneous and lagged correlations between $\{\omega_t\}$ and $\{u_t\}$, averaged over all 150 datasets (std.dev. in parentheses), for each parameter set I - V.

Option error series $\{\varepsilon_t\}$
The reason for assuming the option error series $\{\varepsilon_t\}$ to be white noise is mainly for convenience (see section 3.4.2 of chapter III). In theory, our method for extracting information from option prices essentially boils down to replacing a deterministic summation term with the random sum $\mu_v + \varepsilon_t$ (recall (3.69) in chapter III). This is obviously strange if we believe our theoretical model to exactly hold in practice. It is not strange however if we acknowledge that a model is never a complete description of reality. In a simulation environment (in which theory and reality coincide), we can compute the theoretical "random" error series $\{\varepsilon_t\}$, and examine if it at least resembles white noise.

As table 4.7 shows, the mean of $\{\varepsilon_t\}$ equals zero. Plots of the $\{\varepsilon_t\}$-series reveal that the homoskedasticity assumption seems reasonable. The series appear to be autocorrelated however, as evidenced by table 4.6. This observation is important for the following reason. Suppose reality would exactly coincide with the 1-factor CIR SV option pricing model, and suppose we apply our state space method for extracting information from option prices based on a linearization (assuming one SV factor). We know that the theoretical (deterministic) option error series $\{\varepsilon_t\}$ is "autocorrelated", whereas the state space model assumes this series to be white noise. The issue is now, if this autocorrelation translates into remaining autocorrelation in the standardized innovations of the *estimated* state space model, or not.

If it does, it will be clear that the ACF of the standardized innovations forms an inappropriate guide to judge if one SV factor is enough to describe the dynamics in the data. (For example, suppose we find the ACF of the standardized options innovations to be 0.20, 0.15, 0.10,… This would lead us to falsely conclude that one SV factor is not enough, whereas we know the data was generated by a 1-factor SV model.) If no autocorrelation is found, the ACF of the standardized options innovations of the estimated state space model can be used as usual, to examine if the model is dynamically well specified.

Table 4.6: Autocorrelation coefficients in the option error series $\{\varepsilon_t\}$

	AC(1)	AC(2)	AC(3)	AC(4)	AC(5)
I.	0.12 (0.04)	0.08 (0.03)	0.05 (0.03)	0.03 (0.03)	0.00 (0.03)
II.	0.21 (0.05)	0.15 (0.04)	0.10 (0.04)	0.07 (0.04)	0.03 (0.03)
III.	0.32 (0.05)	0.25 (0.05)	0.20 (0.05)	0.15 (0.04)	0.11 (0.04)
IV.	0.45 (0.05)	0.38 (0.05)	0.32 (0.05)	0.27 (0.05)	0.22 (0.05)
V.	0.26 (0.03)	0.20 (0.04)	0.14 (0.03)	0.10 (0.03)	0.05 (0.03)

The table reports the first five autocorrelation coefficients (AC) of the $\{\varepsilon_t\}$-series, averaged over all 150 datasets (std.dev. in parentheses), for each parameter set I–V.

Therefore, in the next section (which deals with the performance of the estimation method), we will also report the ACF of the standardized options innovations of the estimated state space model, to examine if the "autocorrelation" in the option error series $\{\varepsilon_t\}$ found here results in autocorrelation in the standardized innovations.

And what about the assumed zero-correlation between $\{\varepsilon_t\}$ and the other errors $\{\omega_t\}$, $\{\omega_{RV,t}\}$ and $\{u_t\}$? Table 4.7 shows the reasonability of these assumptions.

Table 4.7: Statistics associated with option error series $\{\varepsilon_t\}$ (Options equation)

	Mean $\{\varepsilon_t\}$	corr(ε_t, ω_t)	corr($\varepsilon_t, \omega_{t-\Delta t}$)	corr($\varepsilon_t, \omega_{t-2\Delta t}$)	corr($\varepsilon_t, \omega_{RV,t}$)
I.	1.4e-20	0.01 (0.04)	0.00 (0.03)	0.00 (0.04)	0.00 (0.03)
II.	2.4e-20	0.00 (0.04)	0.00 (0.04)	0.00 (0.04)	-0.01 (0.04)
III.	5.9e-20	0.00 (0.04)	0.00 (0.04)	0.00 (0.04)	-0.02 (0.04)
IV.	6.6e-20	0.00 (0.04)	0.00 (0.04)	0.00 (0.04)	-0.04 (0.04)
V.	1.5e-20	0.01 (0.04)	0.00 (0.04)	0.00 (0.04)	-0.02 (0.03)

	corr($\varepsilon_t, \omega_{RV,t-\Delta t}$)	corr($\varepsilon_t, \omega_{RV,t-2\Delta t}$)	corr(ε_t, u_t)	corr($\varepsilon_t, u_{t-\Delta t}$)	corr($\varepsilon_t, u_{t-2\Delta t}$)
I.	0.00 (0.04)	0.01 (0.04)	-0.01 (0.03)	0.00 (0.03)	0.00 (0.03)
II.	0.00 (0.04)	0.00 (0.04)	-0.03 (0.03)	0.00 (0.04)	0.00 (0.03)
III.	0.00 (0.04)	0.00 (0.04)	-0.07 (0.03)	0.00 (0.04)	0.00 (0.03)
IV.	0.00 (0.04)	0.00 (0.04)	-0.10 (0.03)	0.00 (0.04)	0.00 (0.03)
V.	0.00 (0.04)	0.01 (0.04)	-0.06 (0.03)	0.00 (0.03)	0.01 (0.03)

Besides the mean of $\{\varepsilon_t\}$, the table reports the contemporaneous and lagged correlations between $\{\varepsilon_t\}$ and $\{\omega_t\}$, $\{\omega_{RV,t}\}$ and $\{u_t\}$, averaged over all 150 datasets (std.dev. in parentheses) for each parameter set I–V.

For illustration, figure 4.1 shows the error series $\{\varepsilon_t\}$, $\{\omega_t\}$, $\{\omega_{RV,t}\}$ and the stock volatility series $\{\sigma_t\}$ for a typical simulated dataset, for parameter set III. As expected, the effect of level-dependent volatility-of-volatility is visible in the graphs of $\{\omega_t\}$ and $\{\omega_{RV,t}\}$ through conditional heteroskedasticity. (The series are unconditionally homoskedastic however.)

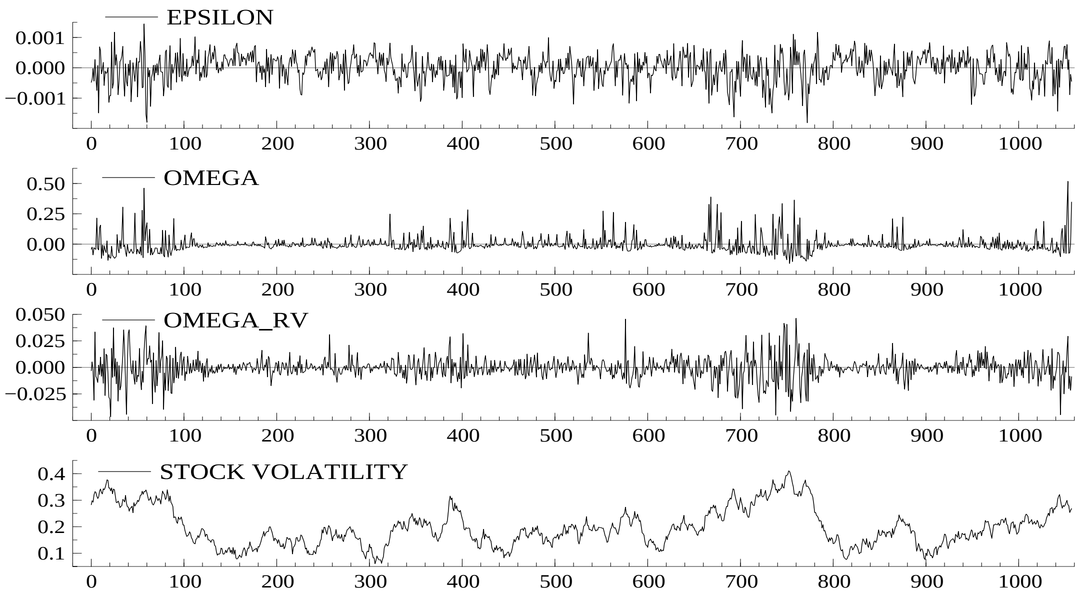


Figure 4.1: Simulated error terms $\{\varepsilon_t\}$, $\{\omega_t\}$, $\{\omega_{RV,t}\}$ and stock volatility series $\{\sigma_t\}$ for a typical simulated dataset, for parameter set III.

4.4 Simulation results for 1-factor CIR SV: Performance of the state space estimation method

This section examines the performance of the estimation method. How close are the estimates obtained by Extended Kalman filter QML to their true values? Are the smoothed and true underlying volatility series close? What data yields the best results in this respect? As motivated in the previous section, we pay separate attention to the standardized options innovations.

There are several causes of likely bias in the estimates. First, although (Q)ML yields consistent estimates, (Q)ML typically exhibits finite-sample bias. Bias is therefore likely to be introduced by assuming the time series in each simulated dataset to consist of $T = 1058$ daily observations only (corresponding to the FTSE100-index data). This finite-sample bias could in principle be reduced by increasing T. Second, Extended Kalman filter (Q)ML generates inconsistent estimates, and may therefore lead to additional finite-sample bias. An obvious third source of bias is introduced by the various approximations that we carried out to arrive at a linear state space model (i.e., the return, RV and options measurement equations). A fourth source is formed by using approximating discrete-time processes to simulate from the continuous-time processes. The issue of course is, how large the cumulative bias in the estimates is. Is this bias parameter dependent? Does it depend on the specific data used for estimation?

4.4.1 Results using squared returns for estimation

Table 4.8 reports the main estimation results if only squared return data is used for estimation. Parameter γ cannot be estimated using return data only.

The bias in the estimates is large. The volatility evaluation criteria RMSPE (root mean squared percent error) and MAPE (mean absolute percent error) for comparing the smoothed with the true underlying volatility series indicate a poor extraction from the data. As the standard deviations indicate, the estimates of mean-reversion and volatility-of-volatility parameters k and σ fluctuate wildly.

Table 4.8: Estimation results 1-factor CIR SV based on **squared return data**

	I	*II*	*III*	*IV*	*V*
θ	**0.039**	**0.035**	**0.031**	**0.022**	**0.018**
	(0.012)	(0.014)	(0.016)	(0.018)	(0.005)
Bias	-0.001	-0.005	-0.009	-0.018	-0.002
MSE	1.4e-4	2.2e-4	3.3e-4	6.6e-4	3.3e-5
k	**6.8**	**17.8**	**32.8**	**64.9**	**27.4**
	(17.5)	(41.9)	(68)	(90)	(45.3)
Bias	5.3	15.2	27.6	57.9	20.4
MSE	336	1989	5383	11484	2468
σ	**0.41**	**1.26**	**2.98**	**6.56**	**0.95**
	(1.22)	(3.16)	(9.98)	(13.9)	(1.72)
Bias	0.21	0.96	2.53	5.96	0.65
MSE	1.5	10.9	106	228	3.4
RMSPE	**14.4**	**22.3**	**30.2**	**43.3**	**23.2**
%	(5.1)	(10.3)	(10.2)	(12.4)	(6.5)
MAPE	**10.6**	**15.8**	**20.9**	**29.5**	**16.9**
%	(3.8)	(8.6)	(9.5)	(12.3)	(5.8)

For each parameter set I-V, the estimation results are summarized over the 150 simulated datasets jointly. The conditional state space model was estimated using **squared return data** only. The average estimates of θ, k, σ are reported (std.dev. in parentheses), with associated sample bias and MSE. The average RMSPE and MAPE, for comparing the smoothed with the true volatilities, are also reported (std.dev. in parentheses).

Further investigation reveals what precisely is going on. Figure 4.2 plots the estimates of θ, k, σ for parameter set III, for which the true values are $\theta = 0.04$, $k = 5.25$ and $\sigma = 0.45$. It often occurs that θ is estimated close to zero, which is generally accompanied by very large estimates of k and σ. The other parameter sets show similar results, and they become worse when subsequently going from set I to IV (as also evidenced by increasing bias and standard deviations).

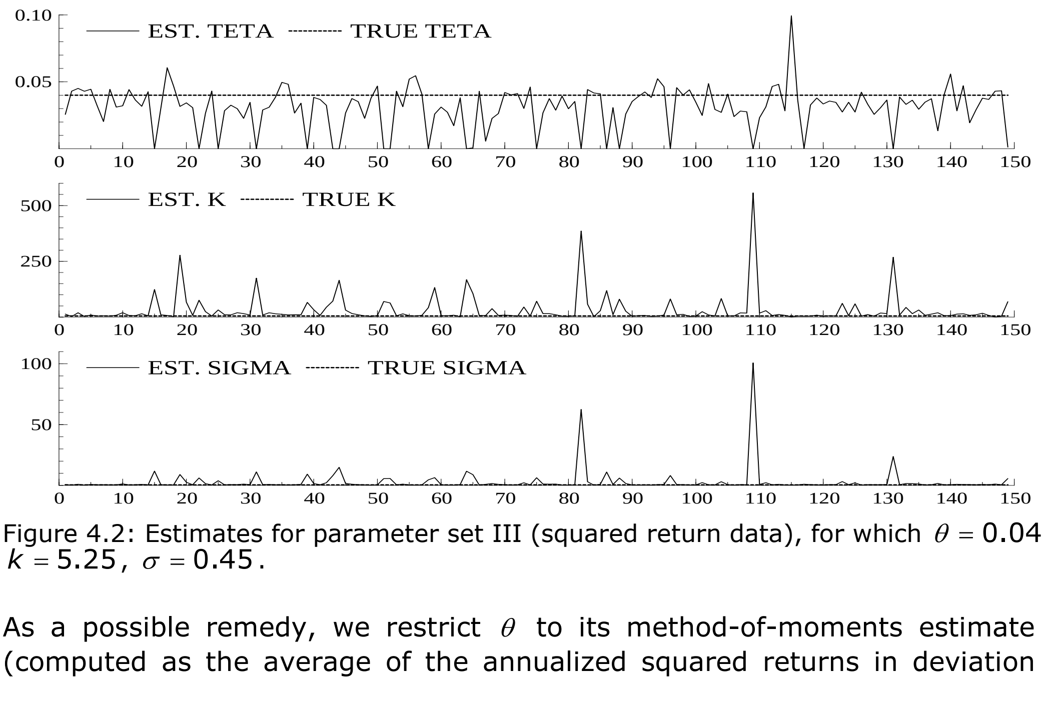


Figure 4.2: Estimates for parameter set III (squared return data), for which $\theta = 0.04$, $k = 5.25$, $\sigma = 0.45$.

As a possible remedy, we restrict θ to its method-of-moments estimate $\hat{\theta}_{SqR}$ (computed as the average of the annualized squared returns in deviation from their mean) prior to estimation. Table 4.9 reports the restricted estimation results. Although imposing $\theta = \hat{\theta}_{SqR}$ leads to a large bias reduction in the k and σ-estimates, the bias is still very large. Moreover, the volatility evaluation criteria have barely improved. Figure 4.3 plots the estimates of θ, k and σ for set III. As compared to figure 4.2, although extreme estimates of k and σ occur less frequently and are generally less large, they are still quite common.

Table 4.9: Estimation results using **squared return data**, under the restriction $\theta = \hat{\theta}_{SqR}$

	I	*II*	*III*	*IV*	*V*
θ	**0.040**	**0.041**	**0.040**	**0.040**	**0.020**
	(0.012)	(0.011)	(0.009)	(0.009)	(0.003)
Bias	0.000	0.001	0.000	0.000	0.000
MSE	1.4e-4	1.3e-4	7.6e-5	7.8e-5	9.4e-6
k	**5.08**	**10.8**	**20.9**	**26.8**	**24.6**
	(8.36)	(29.9)	(38.7)	(42.7)	(42.2)
Bias	3.6	8.2	15.6	19.8	17.6
MSE	83	961	1738	2211	2086
σ	**0.30**	**0.64**	**1.17**	**1.73**	**0.69**
	(0.24)	(0.99)	(1.35)	(1.61)	(0.86)
Bias	0.10	0.34	0.72	1.13	0.39
MSE	0.07	1.10	2.34	3.86	0.88
RMSPE	**14.2**	**20.7**	**28.2**	**40.9**	**22.6**
%	(3.5)	(6.6)	(7.3)	(11.0)	(5.0)
MAPE	**10.3**	**14.2**	**18.7**	**24.1**	**16.2**
%	(1.9)	(4.0)	(5.0)	(5.4)	(4.0)

See table 4.8 for further explanatory legend.

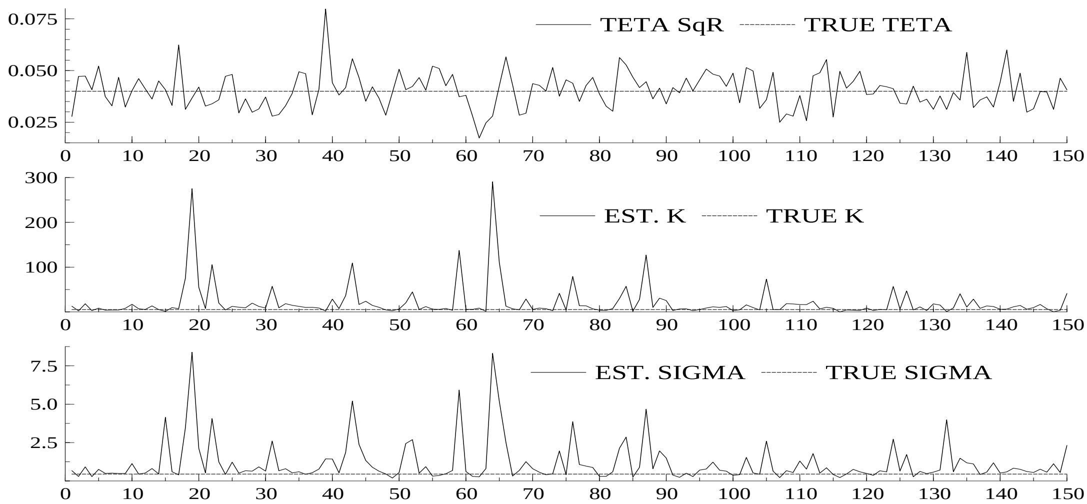


Figure 4.3: Restricted estimates for parameter set III (squared return data), for which $\theta = 0.04$, $k = 5.25$, $\sigma = 0.45$. From top to bottom: $\hat{\theta}_{SqR}$ and k, σ-estimates.

The simulation results confirm that squared returns are generally noisy estimators of the stock variance. It typically appears hard to discriminate between what should be interpreted as true noise, and what as a true signal in the data regarding the data-generating process. Moreover, the results confirm the expectation that QML performs poorly in case only squared returns are used for estimation, given that the distribution of $\{\omega_t\}$ is asymmetric and heavily skewed to the right, and hence departs much from the symmetric, Gaussian distribution. (See section 3.2.1 of chapter III for similar findings in the literature.)

We conclude that the use of squared returns only for QML estimation is generally not to be advocated, not even if we restrict θ to its moment estimate prior to estimation. We refer to section 3.2.1 of chapter III for SV estimation methods based on squared returns that are likely to improve on our method, such as the Monte Carlo maximum likelihood procedure of Sandmann and Koopman (1998).

4.4.2 Results using realized volatilities for estimation

Table 4.10 reports estimation results if 10-minute realized volatility data is used for estimation. Parameter γ cannot be estimated in this case. The correlation parameter $c = \text{corr}_{\mathbb{P}}[\omega_{t+\Delta t}, u_{t+\Delta t}]$ is not identified; we restrict its value to zero.

Table 4.10: Estimation results 1-factor CIR SV based on **RV data**

	I	*II*	*III*	*IV*	*V*
θ	**0.035**	**0.029**	**0.022**	**0.018**	**0.016**
	(0.014)	(0.015)	(0.011)	(0.010)	(0.004)
Bias	-0.005	-0.011	-0.018	-0.022	-0.004
MSE	2.1e-4	3.4e-4	4.2e-4	5.8e-4	2.7e-5
k	**6.02**	**8.96**	**16.9**	**15.4**	**16.5**
	(7.67)	(5.69)	(5.70)	(5.54)	(6.34)
Bias	4.5	6.4	11.6	8.4	9.5
MSE	79	73	168	101	130
σ	**0.374**	**0.622**	**1.01**	**1.15**	**0.536**
	(0.290)	(0.297)	(0.20)	(0.11)	(0.166)
Bias	0.17	0.32	0.56	0.55	0.24
MSE	0.11	0.19	0.36	0.32	0.08
RMSPE	**5.8**	**8.0**	**11.5**	**14.1**	**8.8**
in %	(2.4)	(2.5)	(1.8)	(1.3)	(2.0)
MAPE	**4.5**	**6.2**	**8.9**	**10.6**	**6.9**
in %	(1.9)	(2.0)	(1.4)	(0.7)	(1.6)

For each parameter set I-V, the estimation results are summarized over the 150 simulated datasets jointly. The conditional state space model was estimated using 10-minute **RV data** only. The average estimates of θ, k and σ are reported (std.dev. in parentheses), with associated sample bias and MSE. The average RMSPE and MAPE, for comparing the smoothed and true volatilities, are also reported (std.dev. in parentheses).

Although using RV data means a substantial reduction in bias and MSE with regard to the k and σ-estimates as opposed to using squared returns, the bias and MSEs are still large. The bias in θ is larger than if squared returns are used. The volatility is much better filtered out in the RV case however.

Also in this case it sometimes happens that θ is estimated near zero, though less frequently than in the squared return case. Plots of the estimates of k and σ show less extreme peaks, and display more constant variation. Figure 4.4 graphs the 150 estimates of θ, k and σ for parameter set III, for which the true values are $\theta = 0.04$, $k = 5.25$, and $\sigma = 0.45$.

Table 4.11 reports estimation results based on RV data only, but with θ restricted to its RV-based moment estimate $\hat{\theta}_{RV}$ for each dataset, computed as the mean realized variance. [10] Imposing $\theta = \hat{\theta}_{RV}$ leads to a large bias and MSE reduction in the estimates of θ and k, but only modestly affects the estimates of σ. Surprisingly, it has virtually no effect on the volatility evaluation criteria.

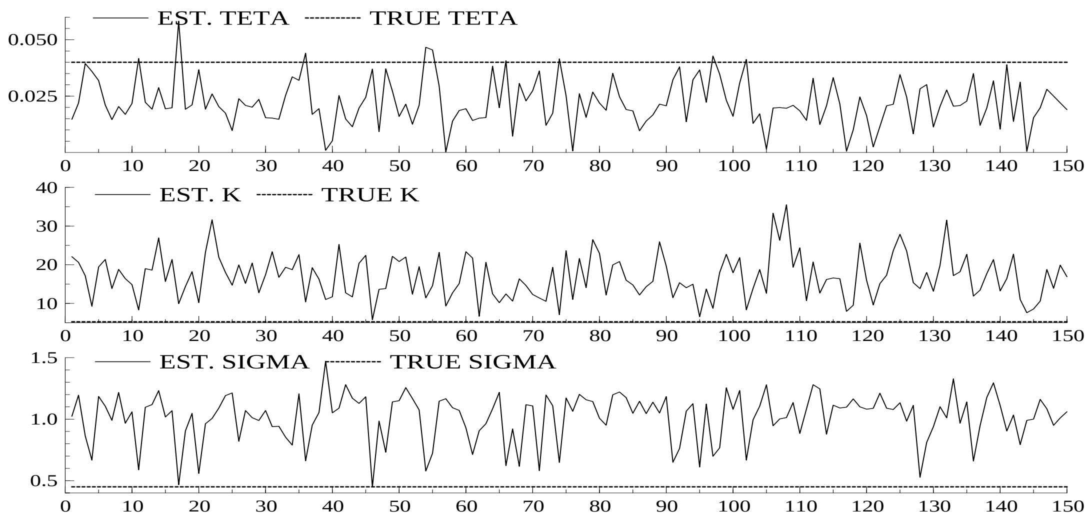


Figure 4.4: Estimates for parameter set III (RV data), for which $\theta = 0.04$, $k = 5.25$, $\sigma = 0.45$. From top to bottom: θ, k, σ -estimates.

Table 4.11: Estimation results based on **RV data** , under the restriction $\theta = \hat{\theta}_{RV}$

	I	*II*	*III*	*IV*	*V*
θ	**0.041**	**0.041**	**0.040**	**0.040**	**0.020**
	(0.012)	(0.011)	(0.008)	(0.009)	(0.003)
Bias	0.001	0.001	0.000	0.000	0.000
MSE	1.4e-4	1.2e-4	7.1e-5	7.3e-5	9.1e-6
k	**4.82**	**5.56**	**10.3**	**8.30**	**12.8**
	(8.28)	(3.30)	(5.3)	(5.56)	(4.2)
Bias	3.3	3.0	5.1	1.3	5.8
MSE	80	20	54	33	51
σ	**0.362**	**0.584**	**0.958**	**1.11**	**0.501**
	(0.273)	(0.270)	(0.189)	(0.102)	(0.142)
Bias	0.16	0.28	0.51	0.51	0.20
MSE	0.10	0.15	0.29	0.27	0.06
RMSPE	**5.8**	**7.8**	**11.3**	**14.1**	**8.5**
in %	(2.3)	(2.4)	(1.9)	(1.3)	(1.8)
MAPE	**4.5**	**6.0**	**8.8**	**10.6**	**6.7**
In %	(1.9)	(1.9)	(1.5)	(0.7)	(1.5)

See table 4.10 for further explanatory legend.

The latter finding is attributed to the following. The stock variance can be written as $\sigma_t^2 = x_t = x_t^* + \theta$, in which $x_t^* = x_t - \theta$ forms part of the unobserved state of the state space model, which needs to be estimated along with θ. The mean stock variance equals $\mathbb{E}_{\mathbb{P}}[\sigma_t^2] = \mathbb{E}_{\mathbb{P}}[x_t] = \mathbb{E}_{\mathbb{P}}[x_t^*] + \theta = 0 + \theta = \theta$. Table 4.10 shows that θ is estimated too low on average for all parameter sets, which seems to imply that the mean stock variance is consistently estimated too low. However, inspecting

[10] As shown in appendix IIIc, the realized variance $\sigma^2_{RV,t+\Delta t}$ is a biased estimator of θ if one samples at a finite number of intraday points I (as we do here, as $I = 48$). Table 4.11 shows that the bias is nevertheless small (for these parameter sets).

the mean of the smoothed $\{x_t^*\}$-series reveals that it deviates from zero, in such a way that the sum of the θ-estimate and the mean of the smoothed $\{x_t^*\}$ is close to θ. Consider e.g. parameter set III. The average estimate of θ equals 0.022, and the average mean of the smoothed $\{x_t^*\}$ equals 0.0188 (std.dev. 0.012), which add up to 0.0408, which is close to $\theta = 0.04$. Restricting $\theta = \hat{\theta}_{RV}$ yields an average estimate of 0.0404 for θ, whereas now the average mean smoothed $\{x_t^*\}$ equals 0.000, adding up to 0.0404, which is again close to $\theta = 0.04$. Apparently, in the RV case it is difficult for the estimation procedure to discriminate between the magnitude of θ and the mean level of $\{x_t^*\}$, although the average volatility level is fitted adequately. (Further investigation reveals that this is also the case in the squared return case, though to a lesser extent.)

For illustration, figure 4.5 plots the restricted estimates of θ, k and σ for parameter set III. Compared to figure 4.4, the estimates of θ have typically shifted upward, those of k downward, whereas those of σ have largely remained in place. Plots for the other parameter sets reveal the same pattern.

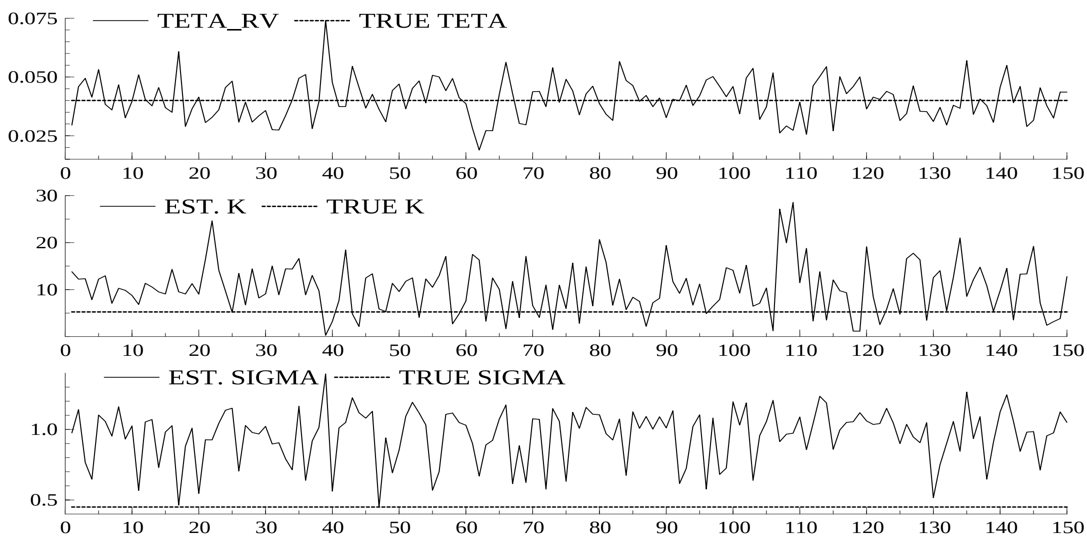


Figure 4.5: Restricted estimates for parameter set III (RV data), for which $\theta = 0.04$, $k = 5.25$, $\sigma = 0.45$. From top to bottom: $\hat{\theta}_{RV}$ and k, σ-estimates.

4.4.3 Results using option data for estimation

Table 4.12 reports the estimation results if only short-maturity at-the-money option data is used for estimation. Once option data is included for estimation, all four parameters θ, k, σ, γ of the 1-factor CIR SV model can be estimated. We restrict "parameter" μ_v to zero prior to estimation. [11]

[11] Leaving μ_v unrestricted, the BFGS optimization routine often keeps on iterating for a long time: Although the gradients associated with the other parameters become 0 quickly, it often cannot get the μ_v-gradient sufficiently close to 0. Instead this gradient keeps on "switching" between values very close to 0. (This also occurs if return–option or RV–option data is used.) Hence, for μ_v, there does not seem to be a clear, one direction that points to an improvement in the loglikelihood when leaving μ_v unrestricted. We attribute this to the fact that μ_v is not a true parameter (in a simulation environment): It essentially absorbs the linearization error, and this error differs for each option at each point in time. (Recall (3.69) in chapter III.) Restricting $\mu_v = 0$ does not result in any problems: the optimum is quickly reached. Given that μ_v is close to 0 in the simulations, and given the earlier reported (Monte Carlo) evidence in the OU chapter, restricting μ_v =0 may even be the better "choice".

The use of option data is clearly preferable to either using only squared return or RV data. Option data is much more informative on all parameters: both the bias and MSE of the estimates are dramatically less. Moreover, the volatility evaluation criteria reveal much better volatility extraction.

Table 4.12: Results 1-factor CIR SV based on (short-maturity ATMF) **option data**

	I	*II*	*III*	*IV*	*V*
θ	**0.040**	**0.040**	**0.040**	**0.039**	**0.020**
	(0.011)	(0.011)	(0.008)	(0.008)	0.003)
Bias	0.000	0.000	0.000	-0.001	0.000
MSE	1.1e-4	1.2e-4	6.0e-5	6.0e-5	8.6e-6
k	**1.73**	**2.86**	**5.49**	**7.19**	**7.24**
	(0.56)	(0.88)	(1.09)	(1.34)	(1.15)
Bias	0.23	0.25	0.24	0.19	0.24
MSE	0.37	0.83	1.25	1.82	1.39
σ	**0.202**	**0.304**	**0.461**	**0.616**	**0.306**
	(0.007)	(0.011)	(0.016)	(0.021)	(0.014)
Bias	0.002	0.004	0.011	0.016	0.006
MSE	5.4e-5	1.2e-4	3.8e-4	7.0e-4	2.5e-4
γ	**-3.9**	**-4.7**	**-5.4**	**-6.3**	**-19.0**
	(2.3)	(2.6)	(2.6)	(2.5)	(3.8)
Bias	0.1	0.3	0.6	0.7	1.0
MSE	5.3	7.0	7.2	6.7	15.2
RMSPE	**1.1**	**2.0**	**3.6**	**6.2**	**2.4**
in %	(0.6)	(1.2)	(1.8)	(2.5)	(1.2)
MAPE	**0.8**	**1.3**	**2.1**	**3.2**	**1.7**
in %	(0.4)	(0.7)	(1.1)	(1.5)	(0.8)

For each parameter set I-V, the estimation results are summarized over the 150 simulated datasets jointly. The conditional state space model was estimated using **option data** only under the restriction $\mu_v = 0$. The average estimates of θ, k, σ, γ are reported (std.dev. in parentheses), with associated sample bias and MSE. The average RMSPE and MAPE, for comparing the smoothed and true volatilities, are also reported (std.dev. in parentheses).

The bias in the estimates of θ and σ is small for all parameter sets, and their small standard deviations show that the estimates are generally close to their true values. The estimates of the mean-reversion and market price of volatility risk-determining parameters k and γ are biased upward. The relative bias of k decreases when moving from very persistent, calmly fluctuating volatility (set I), to less persistent, more heavily fluctuating volatility (set IV, which is close to our empirical estimates), whereas for γ it remains more constant.

The large standard deviations of the γ-estimates indicate that this price of risk parameter is hard to pin down precisely. We encountered this earlier in the OU Monte Carlo study and our empirical work. Moreover, this is a common finding in the literature; see e.g. Pan (2002) and the interest rate literature.

Standardized innovations

Table 4.13 reports the autocorrelation coefficients of the standardized innovations in the BS implied variances of the estimated state space model, averaged over all datasets (standard deviations in parentheses), for each parameter set I-V. Evidently, the theoretical autocorrelation found in the option error series $\{\varepsilon_t\}$ (recall table 4.6) does not translate into remaining autocorrelation in the standardized options innovations of the estimated state space model. As such, the ACF of these innovations can be used as usual for examining if the model is dynamically well specified in practice.

Table 4.13: Autocorrelation coefficients in the standardized innovations

	I	*II*	*III*	*IV*	*V*
AC(1)	0.01 (0.02)	0.01 (0.02)	0.00 (0.02)	0.00 (0.03)	0.01 (0.02)
AC(2)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)
AC(3)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)
AC(4)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)
AC(5)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)

The table reports the autocorrelation coefficients (AC) up to order 5 of the standardized innovation series associated with the options measurement equation, averaged over the 150 simulated datasets (std.dev. in parentheses), for each parameter set I - V.

4.4.4 Results using squared return and option data for estimation

Table 4.14 reports the estimation results if both squared return and short-maturity ATM option data is used for estimation, with μ_v restricted to zero.

This combination of data yields better results than the use of option data alone. The bias and MSE in the k, σ and γ-estimates has decreased further, and the volatility evaluation criteria have (modestly) improved. There are some efficiency gains to be obtained from including squared return in addition to option data.

Table 4.14: Estimation results 1-factor CIR SV based on **squared return - option data**

	I	*II*	*III*	*IV*	*V*
θ	**0.039**	**0.040**	**0.040**	**0.039**	**0.020**
	(0.010)	(0.011)	(0.008)	(0.008)	(0.003)
Bias	-0.001	0.000	0.000	-0.001	0.000
MSE	1.0e-4	1.2e-4	6.3e-5	6.2e-5	8.4e-6
k	**1.70**	**2.82**	**5.40**	**7.10**	**7.11**
	(0.55)	(0.85)	(1.09)	(1.33)	(1.14)
Bias	0.20	0.21	0.15	0.10	0.11
MSE	0.34	0.77	1.21	1.77	1.32
σ	**0.201**	**0.303**	**0.458**	**0.614**	**0.304**
	(0.007)	(0.011)	(0.017)	(0.024)	(0.015)
Bias	0.001	0.003	0.008	0.014	0.004
MSE	5.4e-5	1.3e-4	3.7e-4	7.8e-4	2.4e-4
γ	**-4.0**	**-4.8**	**-5.4**	**-6.3**	**-19.2**
	(2.3)	(2.6)	(2.6)	(2.5)	(3.7)
Bias	0.0	0.2	0.6	0.7	0.8
MSE	5.2	6.8	7.0	6.7	14.0
RMSPE	**1.1**	**2.0**	**3.4**	**6.1**	**2.3**
in %	(0.6)	(1.1)	(1.7)	(2.6)	(1.1)
MAPE	**0.8**	**1.2**	**2.0**	**3.1**	**1.6**
in %	(0.4)	(0.7)	(1.0)	(1.3)	(0.7)

For each parameter set I-V, the estimation results are summarized over the 150 simulated datasets jointly. The conditional state space model was estimated using **return - option data** with $\mu_v = 0$ imposed. The average estimates of θ, k, σ and γ are reported (std.dev. in (.)), with associated sample bias and MSE. The average RMSPE and MAPE, for comparing the smoothed with the true volatilities, are also reported (std.dev. in (.)).

As an illustration, figure 4.6 plots the distributions of the estimates for parameter set III, for which $\theta = 0.04$, $k = 5.25$, $\sigma = 0.45$ and $\gamma = -6$. Based on the 150 estimates of each parameter, the histogram is plotted together with the estimated (continuous) density. A normal density with the same mean and variance has been superimposed. The graphs provide an indication of the finite-sample distribution of the state space estimates.

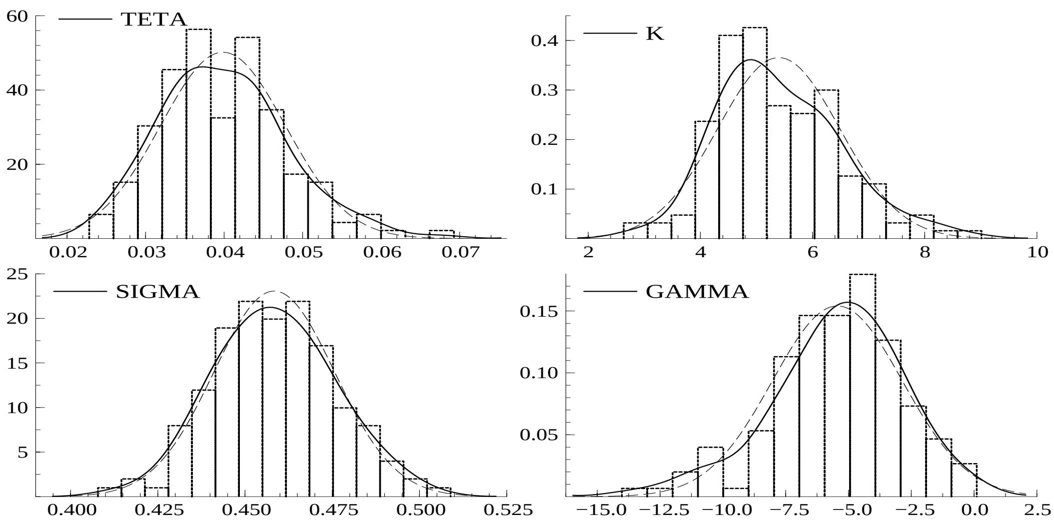


Figure 4.6: Histogram, estimated density, and associated normal density (with same mean and variance) of the 150 estimates (computed using **return - option data**) of θ, k, σ and γ for parameter set III. True values: $\theta = 0.04$, $k = 5.25$, $\sigma = 0.45$ and $\gamma = -6$.

Standardized innovations and diagnostic checking

Table 4.15 reports autocorrelation coefficients of the standardized innovations in the BS implied variances (upper part) and squared returns (lower part) of the estimated state space model, averaged over the 150 datasets (std.dev. in parentheses), for each parameter set I-V. As in the option data-only case, the theoretical autocorrelation found in the option error series $\{\varepsilon_t\}$ (table 4.6) does not translate into remaining autocorrelation in the standardized innovations of the estimated state space model. The ACF of these innovations can thus be used as usual for investigating if the model is dynamically well specified in practice.

Table 4.15: Autocorrelation coefficients in the standardized innovations

		I	*II*	*III*	*IV*	*V*
Option:	AC(1)	0.01 (0.02)	0.01 (0.02)	0.00 (0.03)	0.00 (0.03)	0.01 (0.02)
	AC(2)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)
	AC(3)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)
Sq.return:	AC(1)	0.00 (0.03)	0.00 (0.04)	0.00 (0.05)	0.00 (0.05)	0.00 (0.04)
	AC(2)	0.00 (0.04)	0.00 (0.04)	0.00 (0.04)	0.00 (0.05)	0.00 (0.05)
	AC(3)	0.00 (0.04)	0.00 (0.04)	0.00 (0.05)	0.00 (0.05)	0.00 (0.04)

The table reports the autocorrelation coefficients (AC) up to order 3 of the std. innovations associated with the options (upper part) and squared return (lower part) measurement equations, averaged over the 150 simulated datasets (std.dev. in (.)), for each set I - V.

What does a typical picture of these standardized innovations look like? And important, what do we learn from such a picture for our empirical analysis regarding diagnostic checking? Figure 4.7 plots the standardized innovations associated with the estimated state space model for a typical dataset, for parameter set III. This dataset was also used for illustration in figure 4.1. The standardized options innovations look homoskedastic. In contrast, the innovations in the squared returns display conditional heteroskedasticity, as a result of volatility feedback: [12] Compare their graph with the stock volatility graph in figure 4.1. Pictures for the other datasets (and parameter sets) look similar.

If the model is correctly specified, the standardized innovations in the option series thus ought to look homoskedastic, both unconditionally *and* conditionally. This suggests the following regarding diagnostic checking: If clear evidence of conditional heteroskedasticity is found in the graphs of the option innovations in our empirical FTSE100-index analysis, then the level-dependent volatility-of-volatility present in the data is not adequately (or sufficiently) captured by the model that was estimated. This turns out to be an important observation later.

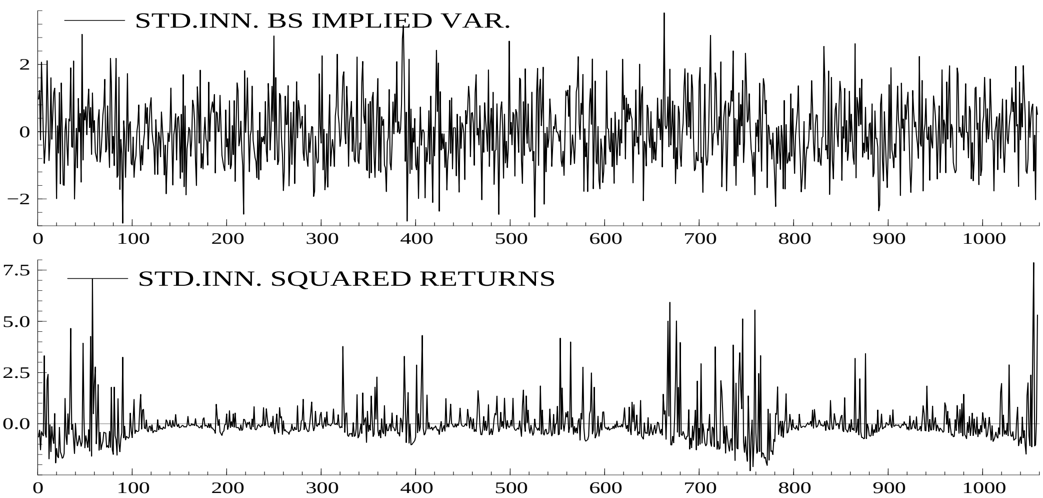


Figure 4.7: Standardized innovations of the estimated state space model for a typical simulated dataset, for parameter set III. (This dataset is also used in figure 4.1.)

4.4.5 Results using RV and option data for estimation

Table 4.16 reports estimation results based on using both RV and short-maturity ATM option data for estimation, with μ_V restricted to zero.

θ is accurately estimated, and with similar precision as in the squared return – option case. The bias in the mean-reversion parameter k is now close to zero for parameter set I and II, whereas for sets III - V the bias is somewhat larger than in the squared return - option case, though still seemingly small. Although for the other data types we found consequent over-estimation of k on average, note that for parameter sets II - V the bias is now negative, implying under-estimation. Parameter σ is very accurately estimated, with negligible bias and a small MSE. According to the MSE criterion, parameter γ is more precisely estimated than when using return–option data, although the bias is for sets II - V somewhat larger. The combination RV–option data yields the most favorable volatility evaluation criteria. Their magnitudes indicate that the state space model performs well in extracting the underlying latent volatility series.

There are clear efficiency gains associated with using RV and option data jointly for estimation. Overall speaking, we conclude that the combination of RV and option data yields the best estimation results.

[12] This is attributed to our assumption $\text{var}_{\mathbb{P}}[\omega_{t+\Delta t} \mid \mathcal{F}_t] = \sigma_\omega^2 = \text{var}_{\mathbb{P}}[\omega_{t+\Delta t}]$, which was motivated by, first, the non-exact Euler discretization, and second, that it simplifies implementation greatly.

For illustration, figure 4.8 plots the histogram, estimated density and associated normal density (with same mean and variance) of the estimates for parameter set III, for which $\theta = 0.04$, $k = 5.25$, $\sigma = 0.45$ and $\gamma = -6$.

Table 4.16: Estimation results 1-factor CIR SV based on **RV - option data**

	I	*II*	*III*	*IV*	*V*
θ	**0.039**	**0.040**	**0.040**	**0.039**	**0.020**
	(0.011)	(0.011)	(0.009)	(0.008)	(0.003)
Bias	-0.001	0.000	0.000	-0.001	0.000
MSE	1.1e-4	1.3e-4	7.5e-5	7.3e-5	8.0e-6
k	**1.58**	**2.60**	**5.07**	**6.83**	**6.75**
	(0.46)	(0.70)	(1.05)	(1.38)	(1.00)
Bias	0.08	-0.01	-0.18	-0.17	-0.25
MSE	0.22	0.50	1.14	1.93	1.07
σ	**0.200**	**0.299**	**0.450**	**0.601**	**0.299**
	(0.006)	(0.009)	(0.013)	(0.019)	(0.010)
Bias	0.000	-0.001	0.000	0.001	-0.001
MSE	4.0e-5	8.1e-5	1.8e-4	3.6e-4	1.0e-4
γ	**-4.0**	**-4.6**	**-5.1**	**-6.2**	**-18.7**
	(2.1)	(2.4)	(2.3)	(2.3)	(3.2)
Bias	0.0	0.4	0.9	0.8	1.3
MSE	4.6	5.6	6.0	6.1	11.7
RMSPE	**0.7**	**1.1**	**1.9**	**4.1**	**1.2**
in %	(0.4)	(0.6)	(1.0)	(2.9)	(0.4)
MAPE	**0.5**	**0.7**	**1.0**	**1.6**	**0.8**
in %	(0.2)	(0.3)	(0.3)	(0.5)	(0.2)

For each parameter set I-V, the estimation results are summarized over the 150 simulated datasets jointly. The conditional state space model was estimated using **RV and option data**. The average estimates of θ, k, σ, γ, ρ are reported (std.dev. in parentheses), with associated sample bias and MSE. The average RMSPE and MAPE, for comparing the smoothed with the true volatilities, are also reported (std.dev. in parentheses).

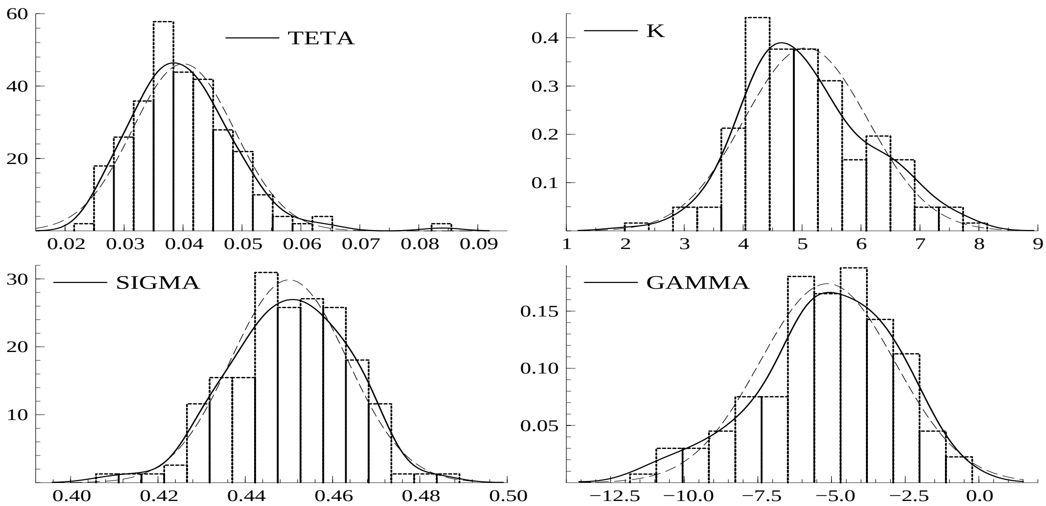


Figure 4.8: Histogram, estimated density, and associated normal density (with same mean and variance) of the 150 estimates (computed using **RV-option data**) of θ, k, σ and γ for parameter set III. True values: $\theta = 0.04$, $k = 5.25$, $\sigma = 0.45$ and $\gamma = -6$.

Standardized innovations and diagnostic checking

Table 4.17 reports autocorrelation coefficients of the standardized innovations in the BS implied variances and realized variances of the estimated state space model, averaged over the 150 datasets (std.dev. in parentheses). As in the option and return-option cases, there is no remaining autocorrelation left. A similar account as before thus applies.

Table 4.17: Autocorrelation coefficients in the standardized innovations

		I	*II*	*III*	*IV*	*V*
Option:	AC(1)	0.01 (0.02)	0.01 (0.02)	0.00 (0.03)	0.00 (0.03)	0.00 (0.02)
	AC(2)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)
	AC(3)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)
RV:	AC(1)	0.01 (0.04)	0.00 (0.04)	0.00 (0.05)	0.00 (0.05)	0.00 (0.04)
	AC(2)	0.00 (0.04)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
	AC(3)	0.00 (0.04)	0.00 (0.04)	0.00 (0.05)	0.00 (0.05)	0.00 (0.04)

The table reports the autocorrelation coefficients (AC) up to order 3 of the std. innovation series associated with the options (upper part) and RV (lower part) measurement equations, averaged over the 150 simulated datasets (std.dev. in (.)), for each set I - V.

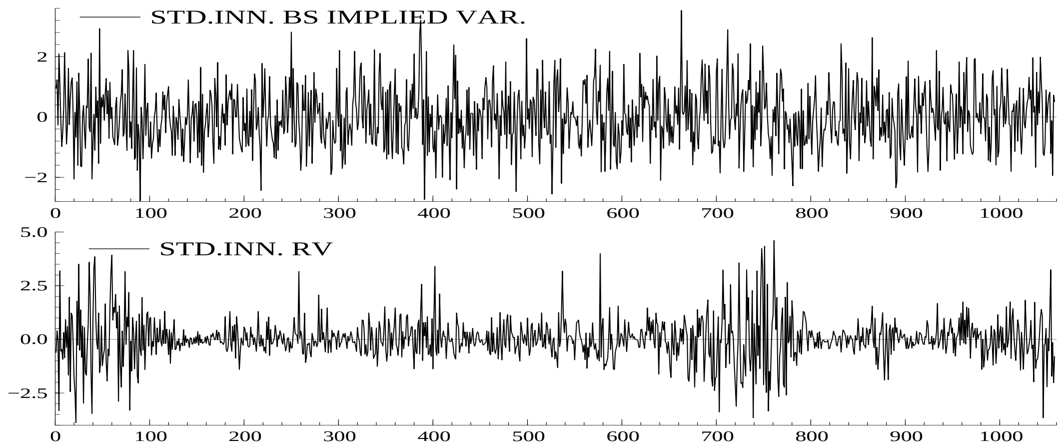


Figure 4.9: Standardized innovations of the estimated state space model for a typical simulated dataset, for parameter set III. (This dataset is also used in figures 4.1 and 4.7.)

Figure 4.9 plots the standardized innovations associated with the estimated state space model for a typical dataset, for parameter set III. This dataset was also used for illustration in figures 4.1 and 4.7. The standardized options innovations look homoskedastic, whereas the RV innovations display conditional heteroskedasticity, due to level-dependent volatility-of-volatility. [13] Compare their graph again with the stock volatility graph in figure 4.1. Pictures for the other datasets (and parameter sets) look similar.

It will be clear that a similar account as in the return - option case regarding diagnostic checking thus applies here in the RV – option case as well.

5. FTSE100-index data: Results for 1-factor CIR SV

This section presents estimation results for the 1-factor CIR SV option pricing model (the Heston (1993) model without leverage effect) based on the FTSE100-index data. Section 5.1 reports results based on either only squared return data or SM option data, and both. Section 5.2 reports estimation results when the returns and the SM, MM and LM option series are jointly used for estimation.

[13] We attribute this to our simplifying assumption of $\text{var}_{\mathbb{P}}[\omega_{RV,t+\Delta t} \mid \mathcal{F}_t] = \sigma^2_{RV,\omega} = \text{var}_{\mathbb{P}}[\omega_{RV,t+\Delta t}]$ which was motivated by the non-exact measurement equation for RV.

5.1 Results based on only return data, SM option data, and both

Table 5.1 presents estimation results for the 1-factor CIR SV model. Table 5.1 may be compared to table 4.1 in the last chapter for the 1-factor OU SV model, except for the estimates of σ and γ (which cannot be directly compared as the model differs). Given the extensive comments on the OU results (sections 4.1-4.4 of the previous chapter), and given that the implied CIR results are largely comparable, we keep the discussion short and focus on the main differences.

Table 5.1: Estimation results **1-factor CIR SV** (FTSE100-index data; Oct 1997–Dec 2001)

	Sq.return data	*SM option data*	*SM option data* $\theta = \hat{\theta}_{SqR}$	*Sq.return & SM option data*	*GARCH*
θ	**0.0318**	**0.0571**	**0.0413**	**0.0441**	
	(0.0118)	(0.0131)		(0.0111)	
k	**143**	**8.39**	**11.4**	**7.86**	
	(170)	(2.22)	(1.06)	(1.95)	
σ	**4.69**	**0.805**	**0.805**	**0.689**	
	(6.40)	(0.059)	(0.054)	(0.0367)	
γ	-	**0.910**	**-3.30**	**-6.91**	
		(2.72)	(1.11)	(3.20)	
σ_ω	**0.0453**	-	-	**0.0663**	
	(0.0261)			(0.0063)	
σ_ε	-	**0.00075**	**0.00072**	**1.15e-5**	
		(0.00031)	(0.00025)	(1.84e-5)	
Vol. returns	17.8%	23.9%	20.3%	21.0%	20.3%
Vol-of-var.	0.0495	0.0470	0.0343	0.0365	0.0175
Persistence	0.577	0.968	0.957	0.970	0.937
Half-life (days)	1.3	21	16	23	11
Std.dev. u_t	0.0404	0.0117	0.0099	0.0088	
Loglikelihood	1406	3823	3821	5162	3165

Parameter estimates are in boldface, with robust White (1982) QML standard errors in parentheses. We estimated the conditional state space model (3.8)-(3.9) using only return data, only SM option data, only SM option data under the restriction $\theta = \hat{\theta}_{SqR} = 0.0413$, and both return - SM option data. Also reported are the unconditional stock volatility $\sqrt{\theta}$, the volatility-of-the-variance $\sqrt{\sigma^2\theta/2k}$, the volatility persistence $\exp[-k\Delta t]$, the half-life (in days) of a volatility shock $\ln 2/k\Delta t$, the QMLE of the standard deviation of the daily volatility shock u_t, and the maximized quasi-loglikelihood value. The last column shows comparable quantities for the Gaussian GARCH(1,1) model (4.1) in the last chapter.

Notice first that although the OU and CIR SV specifications are nonnested (the affine specification nests both however), much larger quasi-loglikelihoods are achieved in the CIR case. For example, based on return-option data, the quasi-loglikelihood in the OU case equals 4761, here it equals 5162. This is indicative of the improved data description by CIR SV.

The results based on the squared returns only confirm the simulation results of section 4.4.1. The estimated k and σ are very large and have large standard errors, indicating that the signal in the data is hard to distinguish from the noise. In the OU case, we found less extreme estimates however.

Using SM option data only results in a too large θ-estimate it seems. The method-of-moment estimate of θ equals $\hat{\theta}_{SqR}$ = 0.0413, and implies an estimate of the mean stock volatility of 20.3%. Restricting θ to $\hat{\theta}_{SqR}$ does virtually not reduce the loglikelihood and gives similar results, except for the market price of volatility risk parameter γ, which is now estimated negative. We encountered this before in the OU case, and is illustrative of the seemingly difficult empirical identification of θ and γ if only option data is used for estimation.

5.1.1 Results based on the combination return - SM option data

We now focus on the results obtained from using both return and SM option data for estimation. The stock variance process obeys the Feller condition, and the mean stock volatility is estimated near its moment estimate of 20.3%. The upper-left graph of figure 5.1 shows the smoothed stock volatilities in the 1-factor CIR and OU cases in one graph; note the close correspondence.

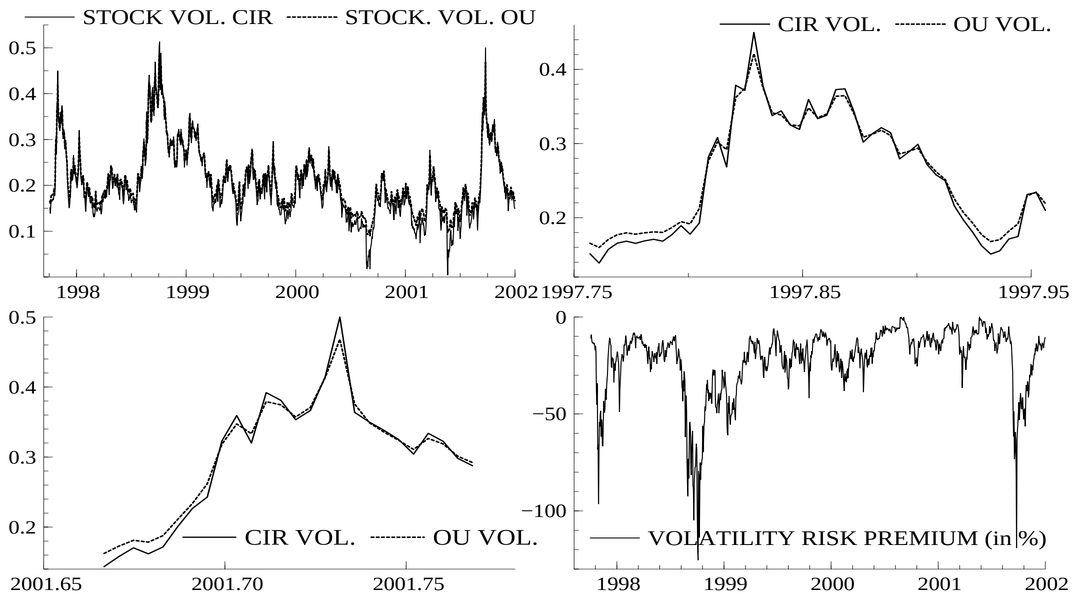


Figure 5.1: Upper-left: Stock volatilities in 1-factor CIR and OU cases (based on **return-SM option data**). Upper-right and lower-left: Zooming in: Volatilities around the time of the beginning Asian crisis and September 11, 2001. Lower-right: Volatility risk premium.

CIR SV appears less persistent than OU SV

The volatility persistence appears smaller in the CIR case (0.97) than in the comparable OU case (0.993). A possible explanation is the following. If the 1-factor OU SV model is fitted to data which features periods of sudden large volatility changes, then it is likely that the estimated volatility path will start rising (or falling) too early, in an attempt to capture or fit these "extremes" as well as possible. This is because OU volatility-of-volatility is constant and as such, the volatility cannot change fast enough. This seems to result in a somewhat deceptively high estimated persistence, not because the volatility is truly that persistent, but because the OU SV model is actually misspecified in that case. As under CIR SV the volatility can change more rapidly due to volatility feedback (and is therefore a better data descriptor), it thus makes sense that the CIR results indicate a lower volatility persistence.

Zooming in on the stock volatility graph in figure 5.1 in two periods of sudden increased turbulence (i.e. the beginning Asian crisis (upper-right graph of figure 5.1) and the period around September 11, 2001 (lower-left)) seems to confirm this reasoning to some extent.

A similar discussion as in section 4.2 of the previous chapter on the comparison of the 1-factor OU SV volatilities with the GARCH and BS implied volatilities, also applies here in 1-factor CIR SV case. Moreover, the option data dominates the estimation results if both squared return and option data are used for estimation, and an upward (resp. downward) sloping VTS is generally associated with a low (high) stock volatility. (We do not report specific results to save space.)

Compensation for FTSE100-index volatility risk
In the 1-factor CIR SV case, the expected spot return an investor faces when investing in a general, path-independent European-style derivative F equals

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = \left[r_t + \beta_{F,S,t}\left(\sigma_t \gamma_{S,t}\right) + \beta_{F,x,t}\left(\sigma \gamma x_t\right)\right], \tag{5.1}$$

in which $\beta_{F,S}$ and $\beta_{F,x}$ are the derivative's stock and volatility beta respectively, $\sigma_t \gamma_{S,t}$ is the stock risk premium, and $\sigma \gamma x_t$ is the volatility risk premium. In contrast to the OU case, the market price of volatility risk is time-varying and is given by $\gamma_{x,t} = \gamma \sqrt{x_t}$, leading to a time-varying volatility risk premium.

Parameter γ is estimated negative at –6.91 and is significant, indicating that market volatility risk is systematic and yields a negative premium. The larger the volatility, the more an investor is willing to pay for delta-neutral positive-vega derivatives, which yield him consumption insurance. As figure 5.1 shows, the magnitude of the volatility risk premium can be substantial. The volatility risk premium is largest negative around the times of the Asian crisis, the near-collapse of LTCM and September 11, 2001.

Volatility risk premium: -21% per annum on average
The CIR results imply an average volatility risk premium of $\sigma \gamma \theta$ = (0.689) (-6.91)(0.0441) = -21% per annum for the FTSE100 index over 1997-2001. This premium is estimated more negative than in the OU case, for which we found a volatility risk premium of –14.6%. (See section 6.3 of chapter II for more evidence on volatility risk compensation.)

Straddle return: -195% per annum
Following a similar strategy as in the OU case (see section 4.3 of the previous chapter), the CIR results imply an average expected short-maturity ATMF straddle return [14] of –3.75% on a weekly basis (OU: -3.34%). On a monthly, respectively annual basis this number equals –16.2% and -195% (OU: -14.5%, resp. –174%). The larger the volatility the more negative the expected straddle return is. Remember that Driessen and Maenhout (2003) find an average monthly empirical return of –13.1%, and Coval and Shumway (2001) an average weekly return of –3% on such straddles. These straddle numbers are close, which is quite remarkable given that both the datasets and time periods differ. In terms of expected returns, writing ATM index straddles seemingly proves very valuable in practice. However, straddles involve considerable risk.

---

[14] In the CIR SV case, the mean spot short-maturity ATMF straddle return is given by $\mathbb{E}_{\mathbb{P}}[dStr_t \,/\, Str_t \mid \mathcal{F}_t] \approx [r_t + \beta_{Str,x,t}\left(\sigma \gamma x_t\right)]dt$. The average (virtually constructed) straddle price equals 366. For comparison with the OU results: The average straddle vega (recall our definition $\mathcal{V}_{Str,t} = \partial Str_t \,/\, \partial \sigma_t^2$) now equals 4842 (OU: 4161), and the average straddle volatility beta equals 14.5 (OU: 12.3). The average Black-Scholes vega now equals 795 (recall that the BS vega is defined as $\partial BS_t \,/\, \partial \sigma_t$); in the OU case it was 796.

5.1.2 Diagnostic checking

We now turn to specification testing of the 1-factor CIR SV model estimated with return - SM option data, and compare the results to the 1-factor OU test results reported in section 4.4 of the last chapter.

Figure 5.2 shows the smoothed disturbances of the state space model. Compared to the OU case, the option error series $\{\varepsilon_t\}$ looks much more homoskedastic. (As the scale indicates, the option series is fitted very well.)

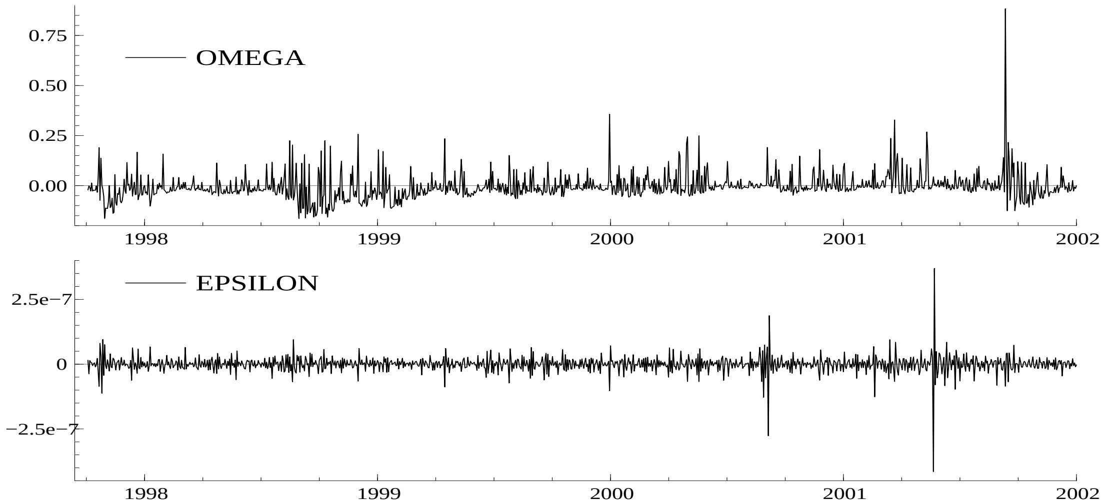


Figure 5.2: Smoothed $\{\omega_t\}$ and $\{\varepsilon_t\}$-series.

Figure 5.3 shows the standardized innovations, which ought to be unit-variance white noise if the model is correctly specified. Their respective means equal −0.08 and 0.01, their variances equal 1.00. As evidenced by table 5.2, although the autocorrelations in the innovation series are small, they are significant, as in the OU case. As expected, modeling level-dependent volatility-of-volatility does not mitigate the autocorrelation.

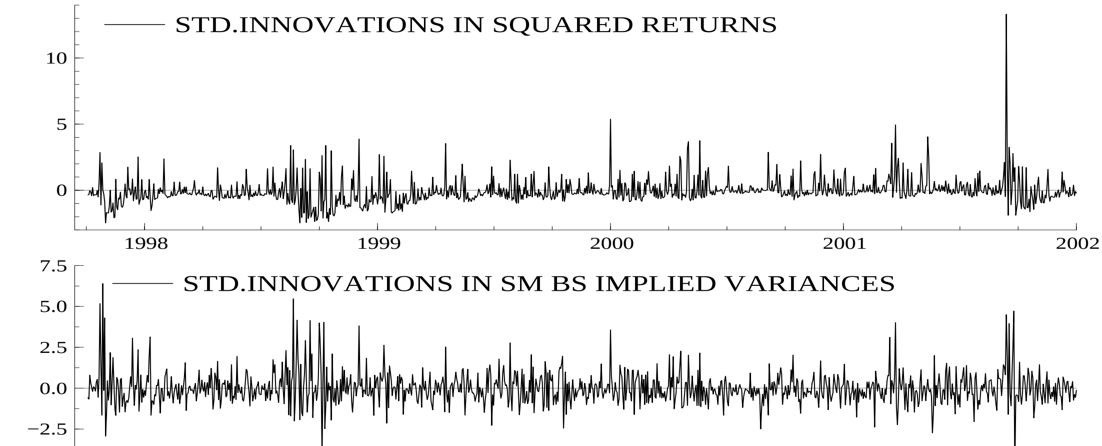


Figure 5.3: Standardized innovations in the squared returns and BS implied variances.

As compared to figure 4.7 in the last chapter for 1-factor OU SV, the result of explicit modeling of volatility feedback is clearly visible in the lower graph of figure 5.3. The standardized options innovations look more (conditionally) homoskedastic than in the OU case.

Table 5.2: Autocorrelations in the standardized innovations

Std. innovations in squared returns				*Std. innovations in SM BS implied variances*			
Order	AC	Q-Stat	Prob	Order	AC	Q-Stat	Prob
1	0.09	9.16	0.002	1	-0.01	0.14	0.71
2	0.07	13.7	0.001	2	-0.07	5.89	0.05
3	0.15	36.8	0.000	3	-0.05	8.41	0.04
4	0.12	52.8	0.000	4	0.05	10.9	0.03
5	0.06	56.8	0.000	5	0.02	11.5	0.04

The table reports the autocorrelation coefficients (AC) up to order 5 and the Ljung-Box Q-statistics (Q-stat) for testing the null of no autocorrelation up to a certain order with associated p-values (Prob).

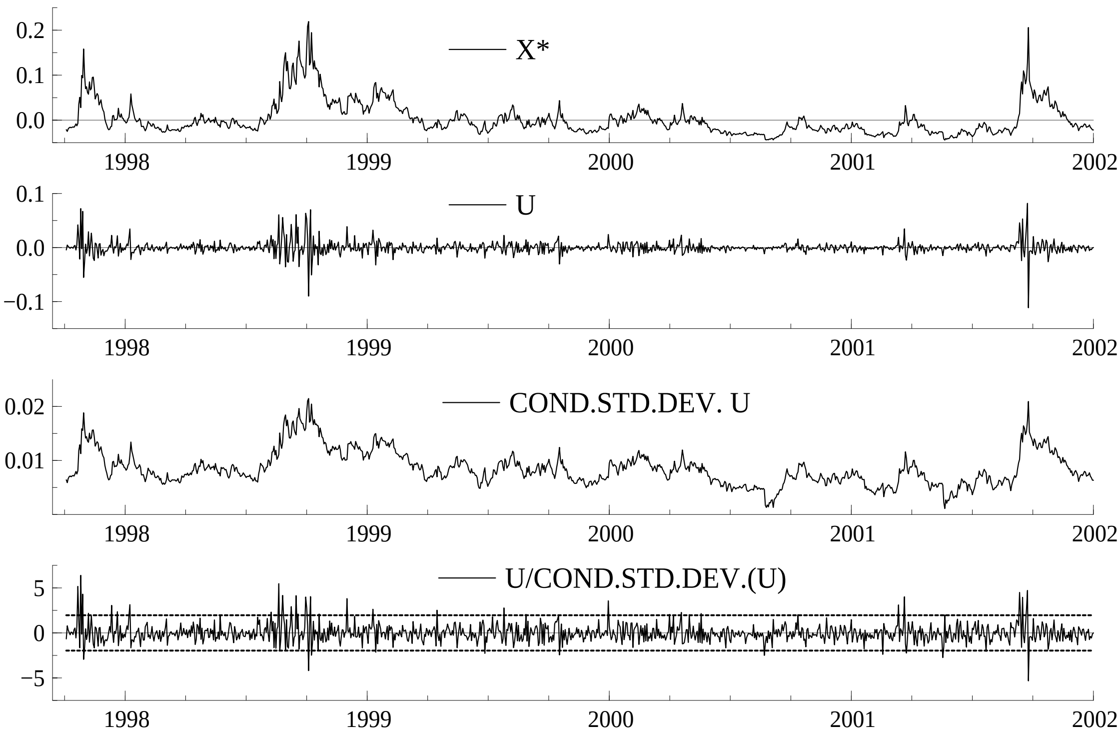


Figure 5.4: Smoothed x_t^*, smoothed daily volatility shock u_t, conditional standard deviation of u_t, and standardized volatility shock $u_t / \sqrt{\text{var}[u_t \mid \mathcal{F}_{t-\Delta t}]}$.

1-factor CIR SV still does not allow for sufficiently fast changing volatility
Nonetheless, there still seems to be conditional heteroskedasticity left in the standardized options innovations. This suggests that 1-factor CIR SV cannot sufficiently adequately describe the volatility feedback present in the data. To investigate this further, consider figure 5.4. The upper plot draws the (smoothed) volatility factor in deviation from its mean (the state $x_t^* = x_t - \theta$), the second plot draws the daily volatility shocks u_t, and the third plot draws the conditional standard deviation of these shocks (see (4.3)). If the model is correctly specified, u_t divided by its conditional standard deviation ought to be approximately

standard normally distributed. [15] The lower plot in figure 5.4 shows the standardized shocks $u_{t+\Delta t} / \sqrt{\text{var}_{\mathbb{P}}[u_{t+\Delta t} \mid \mathcal{F}_t]}$ together with the lines −1.96 and 1.96, which are exceeded on 4.6% of the days. Their skewness equals 1.13, their kurtosis 9.1. The associated Jarque-Bera normality test statistic equals 1861, which is larger than the 95%-critical value of 5.99 of its asymptotic chi-squared(2) distribution under the null. Normality is evidently rejected. [16]

Importantly, the graph of the standardized shocks shows that days on which the line 1.96 is exceeded (which mean large positive volatility shocks) appear in clusters. This suggests that 1-factor CIR volatility does not respond fast enough to rapidly increasing market stress. This point is also made by Chernov et al. (2003), Eraker et al. (2003), Jones (2003) and Tauchen (2004). Tauchen (2004) argues that the problem with the affine specification is that the square-root volatility function (i.e. $\sigma_t = \sqrt{x_t}$) is concave. As such, sudden large realizations of the factors (which are common in times of market stress) have a diminished impact on the stock volatility itself, thereby limiting the ability of the affine SV model to adequately describe such periods.

1-factor SV: Out-of-sample overpricing of longer-dated options

As shown in the last chapter, the 1-factor OU SV model generally overprices the MM and LM ATM options out of sample. Given similar findings on the Heston (1993) model reported in Chernov and Ghysels (2000) and Pan (2002), the 1-factor CIR SV model (which is the Heston model without leverage [17]) is not expected to resolve this overpricing. Table 5.3 and figure 5.5 confirm this.

Figure 5.5 shows the “observed” and fitted SM, MM and LM option prices (quoted in terms of BS implied volatilities) and their differences, the option pricing errors, which summary statistics are given in table 5.3. [18] Although the SM option series is virtually fitted perfectly in sample, the out-of-sample overfitting of the longer end of the volatility term structure can be considerable.

Table 5.3: Summary statistics option pricing errors (1-factor CIR, return – SM option data)

	Error in SM impl.vol.	*Error in MM impl.vol.*	*Error in LM impl.vol.*
Mean	0.000%	-4.02%	-6.08%
Std.deviation	0.000%	2.18%	2.74%

[15] The volatility shock can be written as $u_{t+\Delta t} = \int_t^{t+\Delta t} \exp[-k(t+\Delta t - s)]\sigma\sqrt{x_s}dW_{x,s}$ (see appendix B section 6). Although the exact conditional distribution of $u_{t+\Delta t}$ is unknown, for small Δt the integrand of the Itô integral is approximately constant, such that $u_{t+\Delta t}$ is approximately Gaussian, with mean and variance given in (4.3).

[16] Some caution is in place with this procedure of detecting misspecification. First, to compute the standardized volatility shocks we use the smoothed u_t and x_t , and estimates of the parameters, and not their *true* values. This seems the best we can do. Second, $u_{t+\Delta t} / \sqrt{\text{var}_{\mathbb{P}}[u_{t+\Delta t} \mid \mathcal{F}_t]}$ is only *approximately* normally distributed. Hence, it is not clear if small deviations from normality should either be attributed to misspecification, or to the approximate normality, and as such, cannot be considered a sign of misspecification. Here the deviation from normality seems obvious however.

[17] The correlation between the daily FTSE100-index returns and daily smoothed stock variance changes equals –0.67 (OU: -0.66), which indicates the presence of the leverage effect.

[18] Recall that (in terms of the general multifactor SV model) we compute fitted BS implied volatilities by $\hat{\sigma}_{implied,i,t} = \sqrt{\mu_V + \frac{1}{\tau_{i,t}}\left[A_1(\tau_{i,t}) + \mathbf{B}_1(\tau_{i,t})'\boldsymbol{\theta}\right] + \frac{1}{\tau_{i,t}}\mathbf{B}_1(\tau_{i,t})'\mathbf{x}_t^*}$ for $t = \Delta t, .., T\Delta t$, $i = SM, MM, LM$, in which we substitute the parameter estimates and smoothed $\{\mathbf{x}_t^*\}$ -series. The option pricing errors are computed as $\text{error}_{i,t} = \sigma_{implied,i,t} - \hat{\sigma}_{implied,i,t}$.

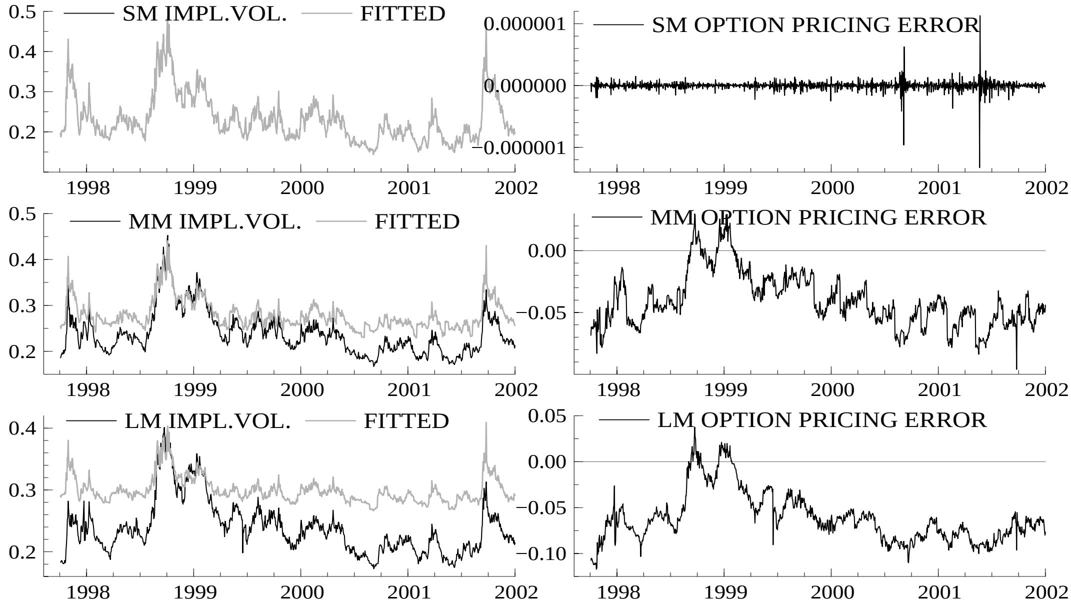


Figure 5.5: Fit of the VTS: In-sample fit of the SM BS implied volatility series and out-of-sample fit of the MM and LM series. Left: "observed" and fitted BS implied volatilities for each maturity. Right: their difference, the option pricing errors.

5.2 1-factor CIR SV: Estimation results using return, SM, MM and LM option data jointly

The out-of-sample overpricing of the longer end of the VTS by the 1-factor CIR SV model is evidence of misspecification. This section reports further evidence that indicates the specific type of misspecification: insufficient volatility dynamics.

Table 5.4: Results **1-factor CIR SV** using return, SM, MM and LM option data jointly

θ	**0.0314** (0.0192)	corr($\varepsilon_{SM,t}$, $\varepsilon_{MM,t}$)	**0.957** (0.014)
k	**0.782** (0.464)	corr($\varepsilon_{SM,t}$, $\varepsilon_{LM,t}$)	**0.691** (0.198)
σ	**0.319** (0.019)	corr($\varepsilon_{MM,t}$, $\varepsilon_{LM,t}$)	**0.218** (0.349)
γ	**-0.931** (1.48)	Loglikelihood	13,324
σ_{ω}	**0.0703** (0.0057)	Vol. Returns	17.7%
$\sigma_{\varepsilon,SM}$	**0.0169** (0.0012)	Vol-of-var.	0.0451
$\sigma_{\varepsilon,MM}$	**0.0047** (0.00076)	Persistence	0.9970
$\sigma_{\varepsilon,LM}$	**0.0020** (0.00044)	Half-life	230
		Std.dev. u_t	0.0035

Parameter estimates are in boldface; robust White (1982) QML standard errors are in parentheses. The conditional state space model was estimated using all data (with $\mu_v = 0$ restricted for all option series), allowing for contemporaneous correlation between the three $\{\varepsilon_{it}\}$-series. Also ~~reported~~ are: the unconditional volatility of returns $\sqrt{\theta}$, the volatility-of-the-variance $\sqrt{\sigma^2\theta/2k}$, the volatility persistence $\exp(-k\Delta t)$, the half-life of a volatility shock $\ln 2/k\Delta t$, and the QMLE of the std.dev. of the daily volatility shock u_t.

Table 5.4 reports the 1-factor CIR SV estimation results in case the four time series are jointly used for estimation. (See section 4.5 in the previous chapter for the comparable 1-factor OU results.) The volatility persistence is estimated smaller than in the OU case. The mean volatility is estimated too low it seems. The large positive correlations between the option errors $\varepsilon_{SM,t}$, $\varepsilon_{MM,t}$ and $\varepsilon_{LM,t}$ suggest that the model lacks sufficient structure. As in the 1-factor OU case, the standardized innovations are heavily (and significantly) autocorrelated, see table 5.5. This naturally motivates the investigation of extended volatility dynamics. Figure 5.6 shows the standardized innovations, figure 5.7 the smoothed disturbances. These series ought to look like white noise approximately if the specification is adequate. This is clearly not the case.

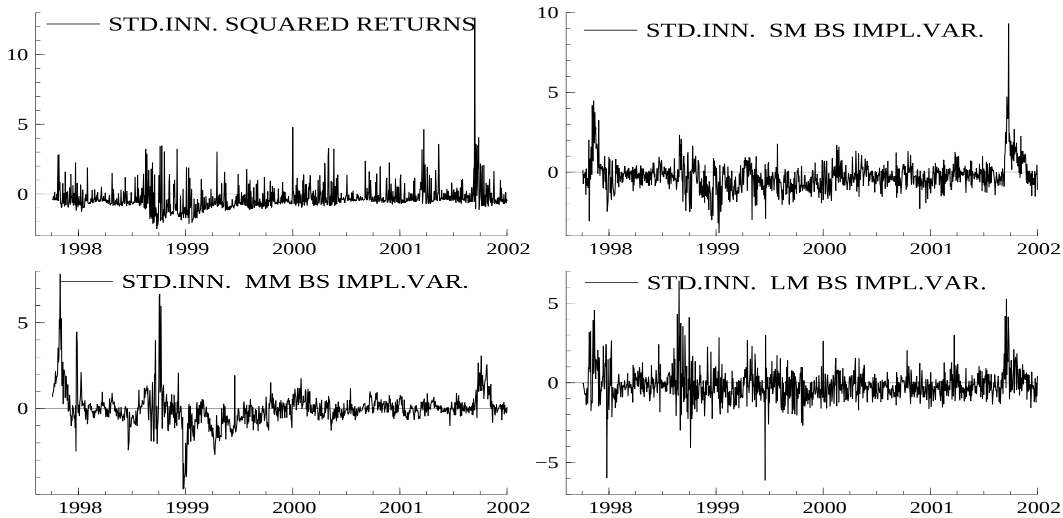


Figure 5.6: Standardized innovations (**1-factor CIR SV**).

Table 5.5: Summary statistics standardized innovations (**1-factor CIR SV**)

	Std.inn. sq. return	*Std.inn. SM series*	*Std.inn. MM series*	*Std.inn. LM series*
Mean	-0.28	-0.26	0.01	-0.14
Std.deviation	0.96	0.95	1.01	1.01
AC(1)	0.13	0.45	0.71	0.11
AC(2)	0.11	0.46	0.60	0.17
AC(3)	0.18	0.47	0.55	0.07
AC(4)	0.16	0.36	0.50	0.16
AC(5)	0.10	0.39	0.48	0.14
Cont.correlation matrix	1.00	0.10	0.15	0.30
		1.00	-0.03	-0.02
			1.00	0.09
				1.00

The table reports the mean, standard deviation, autocorrelation coefficients (AC) up to order 5, and the contemporaneous correlation matrix of the standardized innovations.

In-sample fit of the volatility term structure

Figure 5.8 shows to what extent the 1-factor CIR SV model fits the at-the-money VTS in sample. Figure 5.8 shows the observed and fitted BS implied volatilities of each maturity, and their difference. The pricing errors can be considerable and are not randomly distributed over time. The errors are largest around the times of the Asian crises (fall 1997), the near collapse of LTCM (fall 1998) and September

11, 2001. These periods were all characterized by sudden commotion in financial markets. The 1-factor SV model does not adequately describe these periods.

Table 5.6 reports the mean and standard deviation of the (absolute) pricing errors for each option series. The SM option series is fitted worst, the LM series best. The estimated 1-factor CIR SV model typically predicts somewhat higher option prices than quoted in the market. These findings are similar as in the comparable 1-factor OU case. The fit has (only modestly) improved however.

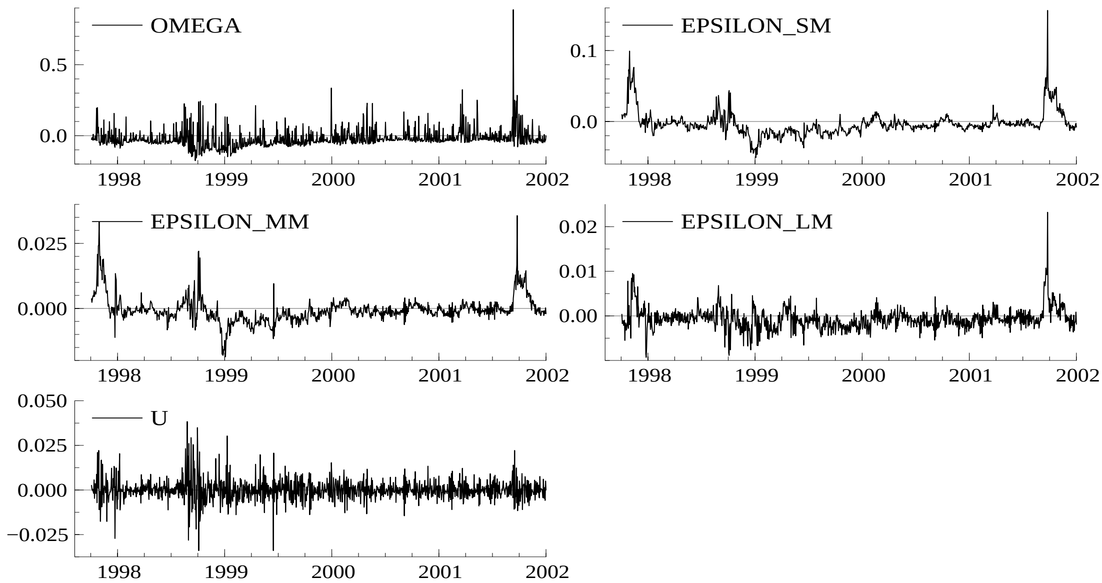


Figure 5.7: Smoothed disturbances of the state space model (**1-factor CIR SV**).

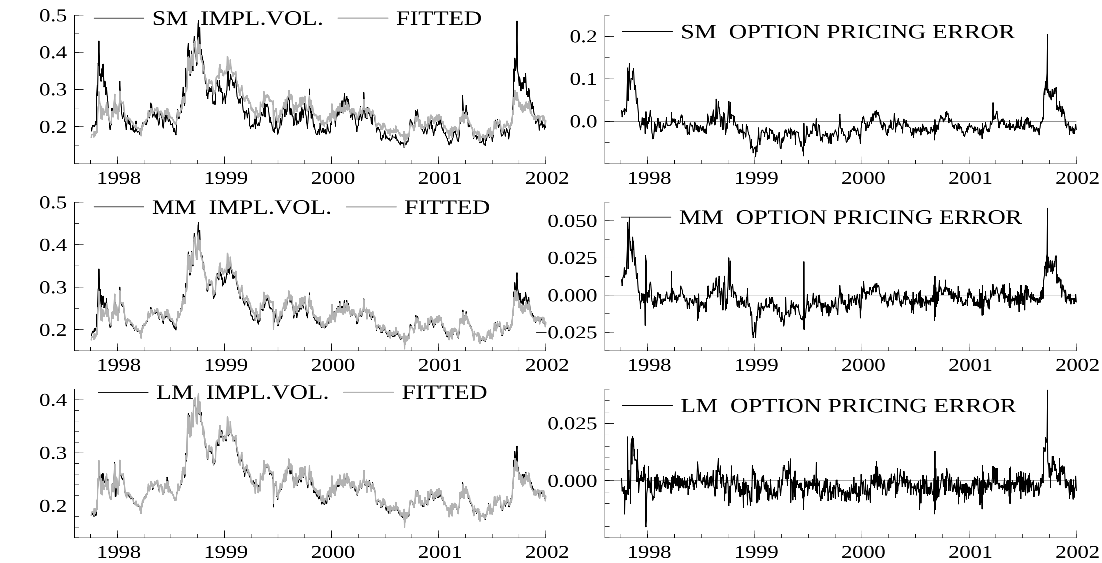


Figure 5.8: Fit (in-sample) of the VTS (**1-factor CIR SV**). Left: observed and fitted BS implied volatilities for each maturity. Right: their difference, the option pricing errors.

Table 5.6: Summary statistics (absolute) option pricing errors (**1-factor CIR SV**)

	Error SM	*Error MM*	*Error LM*	*|Error SM|*	*|Error MM|*	*|Error LM|*
Mean	-0.88%	-0.13%	-0.14%	2.27%	0.65%	0.37%
Std.dev.	2.92%	0.92%	0.48%	2.03%	0.67%	0.33%

6. FTSE100-index data: Results for 2-factor affine SV

Given the evidence of the 1-factor CIR SV model not being able to describe the rich dynamics observed in the joint data, we next extend to two volatility factors. Section 6.1 presents estimation results in case 1 OU and 1 affine factor are driving the volatility. Section 6.2 considers the case of 1 CIR and 1 affine factor. We covered the 2-factor OU case in section 5.2 of the previous chapter.

6.1 Results 2-factor SV model with 1 OU and 1 affine factor

In their analysis of daily 1953-1999 DJIA stock index return data, Chernov et al. (2003) find that two volatility factors are sufficient to adequately describe the data generating process. The first factor is found to be very persistent and not featuring volatility feedback. The second factor appears quickly mean-reverting and does feature volatility feedback.

Given this evidence, as a first extension, let us start by assuming that the stock variance is generated by the following 2-factor SV specification: [19]

$$\sigma_t^2 = x_{1t} + x_{2t}, \quad d\sigma_t^2 = dx_{1t} + dx_{2t} \tag{6.1}$$

with

$$dx_{1t} = k_1(\theta_1 - x_{1t})dt + \sigma_{11}dW_{1t} \qquad (\mathbb{P})$$

$$dx_{2t} = -k_2 x_{2t}dt + \sigma_{21}dW_{1t} + \sqrt{\alpha_2 + \beta_{22}x_{2t}}\,dW_{2t}\,. \qquad (\mathbb{P})$$

The OU factor x_1 reverts back to its mean θ_1 at speed k_1, and has constant volatility parameter σ_{11}. The affine factor x_2 reverts back to its mean $\theta_2 = 0$ at speed k_2. The volatility of x_2 is determined by both the shock in x_1 and an autonomous shock, which volatility depends on the current level of x_2. [20] If $\sigma_{21} = 0$, the factors are independent, otherwise they are correlated with correlation matrix given in (5.8) in appendix B. If $\sigma_{21} > 0$, ceteris paribus, a positive (respectively negative) shock in x_1 results in a positive (negative) shock in x_2.

Table 6.1 reports estimation results for model (6.1) in three cases. *Case a* assumes independent factors ($\sigma_{21} = 0$) and zero correlation between the option

[19] This version is obtained by imposing $\theta_2 = 0, \sigma_{12} = 0, \sigma_{22} = 1, \beta_{11} = \beta_{12} = \beta_{21} = 0, \alpha_1 = 1$ in the general multifactor affine SV model (2.2)-(2.3). Specification (6.1) is close in line with the estimated 2-factor OU specification (see last chapter). In the OU case we had to impose $\theta_2 = 0$ for identification reasons, in the affine case this seems not necessary. It appears however that θ_1 and θ_2 are individually difficult to estimate: We tried this for the 1 OU + 1 affine factor case, the 2-factor CIR case, and the 1 CIR + 1 affine factor case. First, in all these cases the estimated sum $\theta_1 + \theta_2 = \mathbb{E}_{\mathbb{P}}[\sigma_t^2]$ (the mean stock variance) was not very close to its moment-estimate of $\hat{\theta}_{SqR} = 0.0413$, and either θ_1 or θ_2 was estimated near 0. Moreover, mean(smoothed x_{1t}^*) and mean(smoothed x_{2t}^*) were not close to 0. Second, $\hat{\theta}_1$, $\hat{\theta}_2$, mean(smoothed x_{1t}^*) and mean(smoothed x_{2t}^*) nonetheless added up to approximately 0.054 all the time, indicating a seemingly somewhat too high mean stock volatility of 23%. A possible reason for the first finding is that there is too much "freedom" in the model when both θ_1 and θ_2 are left free; i.e. a clear signal is lacking. (Each of the four terms that determines the mean stock volatility in $\mathbb{E}_{\mathbb{P}}[\sigma_t^2] = \theta_1 + \theta_2 + \mathbb{E}_{\mathbb{P}}[x_{1t}^*] + \mathbb{E}_{\mathbb{P}}[x_{2t}^*] = \theta_1 + \theta_2$ needs to be estimated or filtered out of the data.) Restricting θ_2 to 0 seems to help in this respect. The second finding may possibly be attributed to misspecification of the 2-factor SV models.

[20] We have restricted $\sigma_{22} = 1$ for identification reasons: Suppose that x is a general affine factor such that it follows $dx_t = k(\theta - x_t)dt + \sigma\sqrt{\alpha + \beta x_t}\,dW_{x,t}$. This SDE can be rewritten as $dx_t = k(\theta - x_t)dt + \bar{\sigma}\sqrt{\bar{\alpha} + \bar{\beta}x_t}\,dW_{x,t}$ with $\bar{\sigma} \equiv \sigma / c, \bar{\alpha} \equiv c^2\alpha,\ \bar{\beta} \equiv c^2\beta$, for arbitrary $c > 0$, which is exactly the same process. As such, α, β, σ cannot separately be identified. We choose to restrict $\sigma = 1$.

errors $\{\varepsilon_{it}\}; i = SM, MM, LM$ at all times. *Case b* assumes correlated factors, but maintains the assumption of zero option error correlation. *Case c* assumes correlated factors, and allows for contemporaneous option error correlation.

The estimates are rather similar in all cases. The maximized quasi-loglikelihood in case (a) equals 14,475. In the comparable 2-factor OU case (Restrictions (a)) it equaled 14,305. Assuming yet one volatility feedback-featuring factor substantially increases the loglikelihood. [21] Subsequently allowing for factor dependence (case (b)), and option error correlation (case (c)) seems to only modestly further improve the quasi-loglikelihood as compared to case (a).

Table 6.1: Estimation results 2-factor SV model (6.1) using all data **(1 OU + 1 affine factor)**

	Case a	*Case b*	*Case c*		*Case a*	*Case b*	*Case c*
θ_1	**0.0397** (0.0365)	**0.0472** (0.0398)	**0.0385** (0.0382)	corr($\varepsilon_{SM}, \varepsilon_{MM}$)	0	0	**0.149** (0.704)
k_1	**0.356** (0.073)	**0.273** (0.068)	**0.304** (0.070)	corr($\varepsilon_{SM}, \varepsilon_{LM}$)	0	0	**-0.887** (0.545)
k_2	**7.82** (1.41)	**8.62** (1.74)	**9.13** (1.63)	corr($\varepsilon_{MM}, \varepsilon_{LM}$)	0	0	**-0.605** (0.397)
σ_{11}	**0.0565** (0.0045)	**0.0518** (0.0041)	**0.0596** (0.0050)	Vol. returns	19.9%	21.7%	19.6%
σ_{21}	0	**0.0515** (0.0081)	**0.0396** (0.0111)	Vol-of-var.	0.0794	0.0843	0.0878
α_2	**0.0285** (0.0058)	**0.0246** (0.0060)	**0.0236** (0.0054)	Std.dev. x_{1t}	0.067	0.070	0.076
β_{22}	**0.381** (0.058)	**0.454** (0.071)	**0.368** (0.062)	Std.dev. x_{2t}	0.043	0.040	0.037
γ_1	**0.134** (0.210)	**0.171** (0.190)	**0.122** (0.183)	corr(x_{1t}, x_{2t})	0	0.107	0.088
γ_2	**-2.86** (3.09)	**-2.83** (3.58)	**-4.11** (3.65)	Persist. x_1	0.9986	0.9990	0.9988
σ_ω	**0.0672** (0.0057)	**0.0673** (0.0057)	**0.0673** (0.0057)	Persist. x_2	0.9704	0.9674	0.9655
$\sigma_{\varepsilon,SM}$	**1.47e-5** (1.60e-5)	**1.68e-5** (1.64e-5)	**8.33e-4** (4.01e-4)	Half-life x_1	1.9y	2.5y	2.3y
$\sigma_{\varepsilon,MM}$	**0.00201** (0.00018)	**0.00208** (0.00015)	**0.00182** (0.00037)	Half-life x_2	23	21	20
$\sigma_{\varepsilon,LM}$	**0.00206** (0.00027)	**0.00209** (0.00024)	**0.00169** (0.00029)	Std.dev. u_{1t}	0.0035	0.0032	0.0037
Loglik.	14,475	14,509	14,519	Std.dev. u_{2t}	0.0103	0.0101	0.0097

The table reports (restricted) parameter estimates (in boldface) with robust White (1982) QML standard errors in parentheses, resulting from estimating the conditional state space model associated with the 2-factor SV assumption (6.1) using the combination of return, SM, MM and LM option data for three cases (see main text), together with the QMLEs of some other quantities of interest. Half-life is either reported in days or in years (if extension "y" is after the number). The 2-factor SV model (6.1) assumes **1 OU + 1 affine factor**.

The large standard errors of the θ_1-estimates suggest that the mean stock variance is hard to pin down precisely. An intuitive reason is that θ_1 represents the mean of the very persistent factor x_1, for which it takes long before it has reverted back to its mean. Shocks to x_1 have a half-life of around 2.3 years. In our sample of 4 years this factor does not very often "pass" through its mean, which may be the reason for the imprecise estimate. We encountered this imprecision before in the 2 and 3-factor OU cases as well.

[21] At least, it seems substantial. A likelihood-ratio test cannot be performed, as we are dealing with quasi loglikelihoods here.

Two of the option error correlations are estimated negative (which seems counterintuitive), but they are insignificant. Moreover, given that allowing for this extra freedom only modestly increases the loglikelihood when moving from case (b) to (c), this suggests that we may as well restrict these correlations to zero. Parameters γ_1 and γ_2 determine the market prices of volatility-factor risk. In none of the cases these parameters are significant, neither are they jointly. [22]

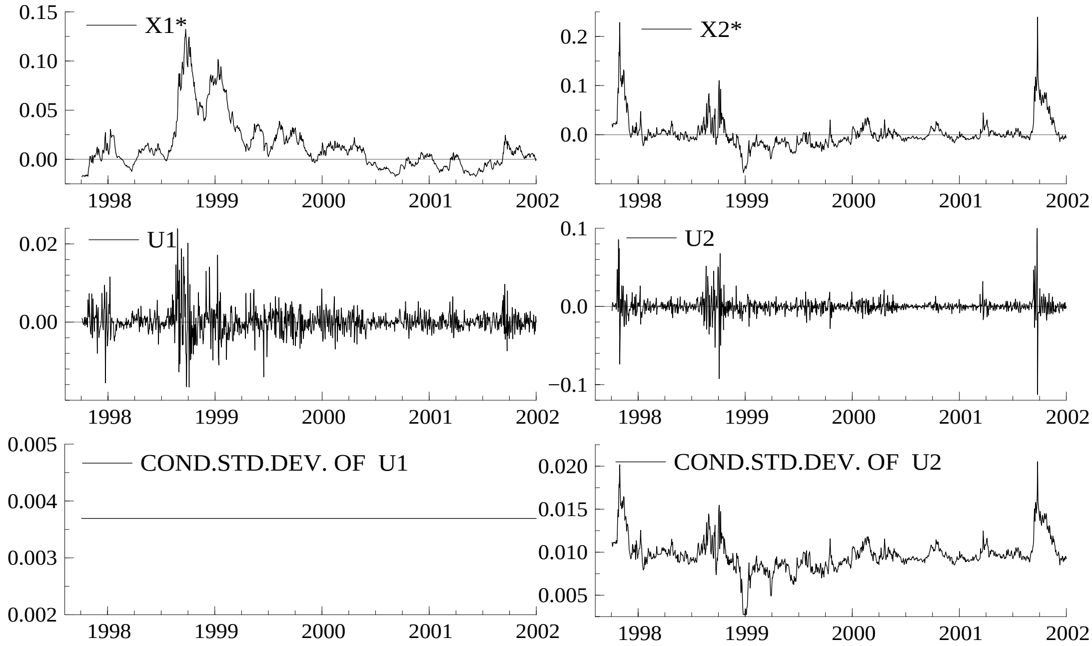


Figure 6.1: Smoothed x_{1t}^*, x_{2t}^*, u_{1t} and u_{2t}. Lower panel: conditional standard deviation of u_{1t} and u_{2t}. (2-factor SV model (6.1) with **1 OU + 1 affine factor**, **case c**)

Although the z-statistics associated with σ_{21} indicate that the persistent and fast mean-reverting factor are significantly correlated, their correlation is low, around 0.10 (2-factor OU: 0.16). The z-statistics of β_{22} indicate that the OU assumption for x_2 is strongly rejected, which is as expected.

The standard errors ought be taken with care however, as model (6.1) appears misspecified in several respects (apart from leverage). A lack of sufficient volatility dynamics appears to be the main misspecification. [23] In addition, consider figure 6.1. The upper panel shows the smoothed factors x_{1t}^* and x_{2t}^* in the least restricted case, case (c). Their course looks similar as in the 2-factor OU case. The middle and lower panel show the daily unpredictable factor shocks u_{1t} and u_{2t}, and their conditional standard deviations (see formula (3.7)). As is clear from the graphs of x_{1t}^* and u_{1t}, the persistent factor x_1 does feature volatility feedback as well (though to lesser extent as the fast mean-reverting factor). This contrasts with the findings of Chernov et al. (2003), who find the persistent factor to not feature feedback; recall the discussion at the beginning of this section.

[22] The respective Wald test statistics of 1.44, 1.59 and 1.75 are smaller than the 95%-critical value of 5.99 of their asymptotic chi-squared(2) distribution under the null hypothesis that $\gamma_1 = \gamma_2 = 0$.

[23] To save space we do not formally report on results of misspecification tests for this model. We will do so for the 2-factor SV model with 1 CIR and 1 affine factor in the next section.

6.2 Results 2-factor SV model with 1 CIR and 1 affine factor

Given that the persistent factor is also found to feature level-dependent volatility, this section reports estimation results in case 1 CIR and 1 affine factor are driving the volatility. We assume the following SV specification: [24]

$$\sigma_t^2 = x_{1t} + x_{2t}, \quad d\sigma_t^2 = dx_{1t} + dx_{2t} \tag{6.2}$$

with

$$dx_{1t} = k_1(\theta_1 - x_{1t})dt + \sigma_{11}\sqrt{x_{1t}}dW_{1t} \qquad (\mathbb{P})$$

$$dx_{2t} = -k_2 x_{2t}dt + \sigma_{21}\sqrt{x_{1t}}dW_{1t} + \sqrt{\alpha_2 + \beta_{22}x_{2t}}dW_{2t}\,. \qquad (\mathbb{P})$$

CIR factor x_1 reverts to its mean θ_1 at speed k_1, with a volatility that depends on its current level. Affine factor x_2 reverts back to its mean $\theta_2 = 0$ at speed k_2, with a volatility determined by the shock in x_1 and an autonomous, level-dependent shock. If $\sigma_{21} = 0$, the factors are independent, else they are correlated.[25]

Table 6.2: Estimation results 2-factor SV model (6.2) using all data **(1 CIR + 1 affine factor)**

	Case a	*Case b*	*Case c*		*Case a*	*Case b*	*Case c*
θ_1	**0.0413**	**0.0413**	**0.0413**	corr($\varepsilon_{SM}, \varepsilon_{MM}$)	0	0	**-0.111** (0.245)
k_1	**0.120** (0.053)	**0.103** (0.051)	**0.0867** (0.0471)	corr($\varepsilon_{SM}, \varepsilon_{LM}$)	0	0	**0.507** (0.215)
k_2	**12.1** (1.34)	**13.2** (1.36)	**13.5** (1.33)	corr($\varepsilon_{MM}, \varepsilon_{LM}$)	0	0	**-0.473** (0.439)
σ_{11}	**0.322** (0.028)	**0.306** (0.023)	**0.329** (0.021)	Vol. returns	20.3%	20.3%	20.3%
σ_{21}	0	**0.252** (0.064)	**0.173** (0.086)	Vol-of-var.	0.1368	0.1415	0.1646
α_2	**0.0210** (0.0027)	**0.0170** (0.0029)	**0.0216** (0.0041)	Std.dev. x_{1t}	0.134	0.137	0.161
β_{22}	**0.365** (0.059)	**0.311** (0.066)	**0.349** (0.065)	Std.dev. x_{2t}	0.029	0.027	0.029
γ_1	**1.15** (0.122)	**1.11** (0.115)	**1.03** (0.112)	corr(x_{1t}, x_{2t})	0	0.064	0.037
γ_2	**-11.5** (1.73)	**-15.3** (2.92)	**-13.0** (2.40)	Persist. x_1	0.9995	0.9996	0.9997
σ_ω	**0.0673** (0.0057)	**0.0674** (0.0056)	**0.0675** (0.0057)	Persist. x_2	0.9547	0.9506	0.9495
$\sigma_{\varepsilon,SM}$	**8.20e-6** (2.30e-5)	**7.32e-7** (8.68e-7)	**6.31e-7** (2.15e-7)	Half-life x_1	5.8y	6.7y	8.0y
$\sigma_{\varepsilon,MM}$	**0.00160** (0.00022)	**0.00172** (0.00016)	**0.00153** (0.00013)	Half-life x_2	15	14	13
$\sigma_{\varepsilon,LM}$	**0.00240** (0.00035)	**0.00235** (0.00028)	**0.00208** (0.00025)	Std.dev. u_{1t}	0.0041	0.0039	0.0042
Loglik.	14,648	14,682	14,717	Std.dev. u_{2t}	0.0088	0.0085	0.0091

The table reports (restricted) parameter estimates (in boldface) with robust White (1982) QML standard errors in parentheses, resulting from estimating the conditional state space model associated with the 2-factor SV specification (6.2) using the combination of return, SM, MM and LM option data for three cases (see main text), together with the QMLEs of some other quantities of interest. Half-life is either reported in days or in years (if extension "y" is after the number). 2-factor SV model (6.2) assumes **1 CIR + 1 affine factor**.

[24] Imposing $\theta_2 = 0, \sigma_{12} = 0, \sigma_{22} = 1, \beta_{11} = 1, \beta_{12} = \beta_{21} = 0, \alpha_1 = 0$ in the general multifactor affine SV model (2.2)-(2.3), yields specification (6.2).

[25] As in the 2-factor OU and 1 OU + 1 affine SV factor cases, we restrict $\theta_2 = \sigma_{12} = 0$, for reasons discussed in earlier footnotes.

Table 6.2 presents results for three cases. *Case a* assumes independent factors and zero option error correlation at all times. *Case b* assumes correlated factors, but maintains the assumption of zero option error correlation. *Case c* assumes correlated factors, and allows for contemporaneous option error correlation. In all cases we restrict θ_1 to its moment estimate of $\hat{\theta}_{SqR} = 0.0413$.

Leaving θ_1 free results in estimates of θ_1 close to zero at about 0.005 in all cases, with averaged smoothed x_{1t}^* and x_{2t}^* deviating too much from 0. (Recall that we encountered similar findings in the Monte Carlo study (see sections 4.4.1 and 4.4.2).) It is unclear why this happens. The high parameter non-linearity of the state space model, combined with the large imprecision found in the θ_1-estimates in the 2-factor OU, 3-factor OU and 1 OU + 1 affine factor cases (which indicates the "instability" of this estimate) may possibly be held responsible for this. As such, we advocate restricting $\theta_1 = \hat{\theta}_{SqR} = 0.0413$ from now on. [26] (We do not anticipate any major econometric problems regarding e.g. dramatically increased inconsistency as a result of imposing this restriction, though it may impact on the standard errors.)

The estimates are rather similar in the three cases. Case (a) yields a maximized quasi-loglikelihood of 14,648. This seems a considerable increase to the value 14,475 found in the comparable 1 OU + 1 affine factor case (a). As encountered before, modeling factor dependence (case (b)) and option error correlation (case (c)) in succession, does not seem to further raise the loglikelihood much.

In comparison to the 1 OU + 1 affine SV factor cases, the persistence of factor x_1 and the speed of mean reversion of x_2 both further increase. Their correlation seems significant but is again small, around 0.05 only. (The standard errors are distorted however, as the model appears misspecified.) The option error correlations do not seem significant. Remarkably, γ_1 and γ_2 (which determine the market prices of the factors) now seem highly significant. (This may possibly be attributed to restricting $\theta = \hat{\theta}_{SqR}$.) As in the previous 2-factor SV estimations, the market price of risk of x_1 is estimated positive, of x_2 negative. The SM option series is fitted nearly perfectly, as evidenced by the small $\sigma_{\varepsilon,SM}$-estimate. The estimates indicate that the Feller condition $2k_1\theta_1 \geq \sigma_{11}^2$ associated with CIR factor x_1 is violated. If the Feller condition holds, the upward drift in the CIR factor is sufficiently large for the factor to never exactly assume the value zero. The mean reversion of x_1 is apparently too slow for the Feller condition to hold.

For further examination and diagnostic checking, we concentrate on the least restricted case, case (c). Figure 6.2 shows the smoothed factors x_{1t}^* and x_{2t}^*. Although their course looks similar as in the previously discussed 2-factor SV estimations, the mean of x_{2t}^* (resp. x_{1t}^*) equals 0.021 (–0.0083) and hence still deviates rather much from zero. The middle and lower panel show the daily factor shocks u_{1t} and u_{2t}, and their time-varying conditional standard deviations.

Fit of the volatility term structure

Figure 6.3 shows the observed and fitted SM, MM and LM Black-Scholes implied volatilities, and their difference, the option pricing errors. The 2-factor SV assumption is a clear improvement over the 1-factor CIR SV assumption (see figure 5.8), which is also evidenced by table 6.3 (as compared to table 5.6). The

[26] We also estimated cases (a), (b) and (c) for the SV specification $\sigma_t^2 = \delta_0 + x_{1t} + x_{2t}$ with x_1 and x_2 both affine factors for which $\theta_1 = \theta_2 = 0$. This did not lead to much better results: δ_0 was estimated around 0.025 and the means of the smoothed x_{1t}^* and x_{2t}^* still deviated too much from 0.

largest improvement in fit is for the SM and MM option series, with the SM option series now fitted virtually perfectly. The pricing error graphs still seem to indicate unmodeled dynamic structure however.

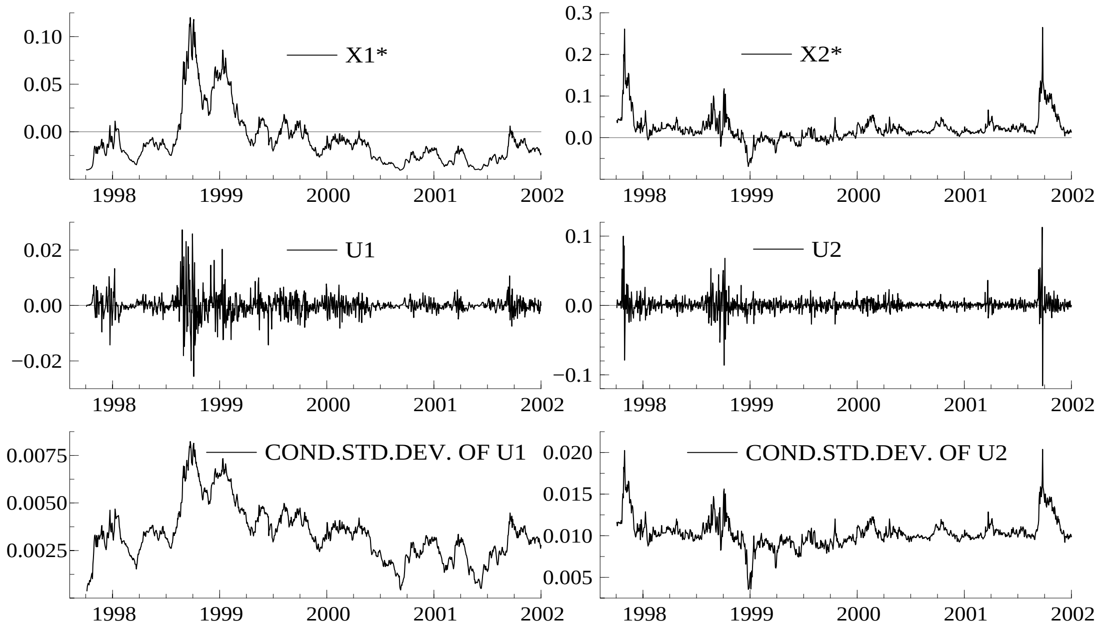


Figure 6.2: Smoothed x_{1t}^*, x_{2t}^*, u_{1t} and u_{2t}. Lower panel: conditional standard deviation of u_{1t} and u_{2t}. (2-factor model (6.2) with **1 CIR + 1 affine SV factor**, **case c**)

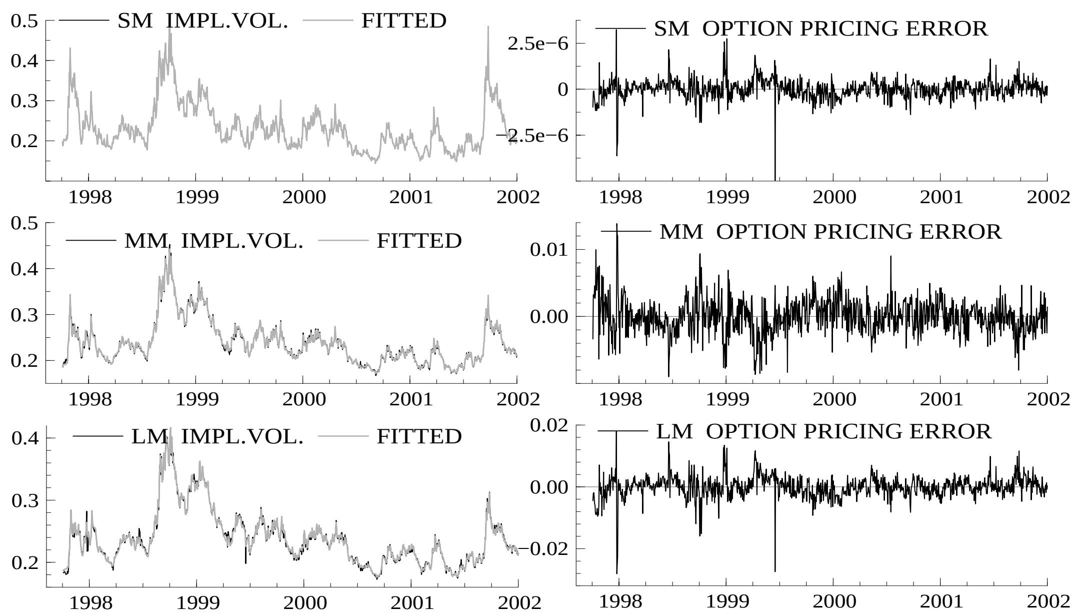


Figure 6.3: Fit of the VTS (**1 CIR + 1 affine SV factor, case c**). Left: observed and fitted BS implied volatilities for each maturity. Right: their difference, the pricing errors.

Table 6.3: Statistics (abs.) option pricing errors (**1 CIR + 1 affine SV factor, case c**)

	Error SM	*Error MM*	*Error LM*	*|Error SM|*	*|Error MM|*	*|Error LM|*
Mean	0.00%	-0.01%	0.00%	0.00%	0.20%	0.25%
Std.dev.	0.00%	0.26%	0.37%	0.00%	0.17%	0.27%

Specification tests
Figure 6.4 shows the standardized innovations, summary statistics are in table 6.4. The improved data description by the 2-factor SV specification with 1 CIR and 1 affine factor as compared to 1 CIR SV factor is obvious (recall figure 5.6 and table 5.5). The innovations more closely resemble white noise. There still seems to be unmodeled conditional heteroskedasticity left however. Although the reduction in autocorrelation is apparent, the dynamics in especially the MM option series are still not very well described by this 2-factor SV model. [27]

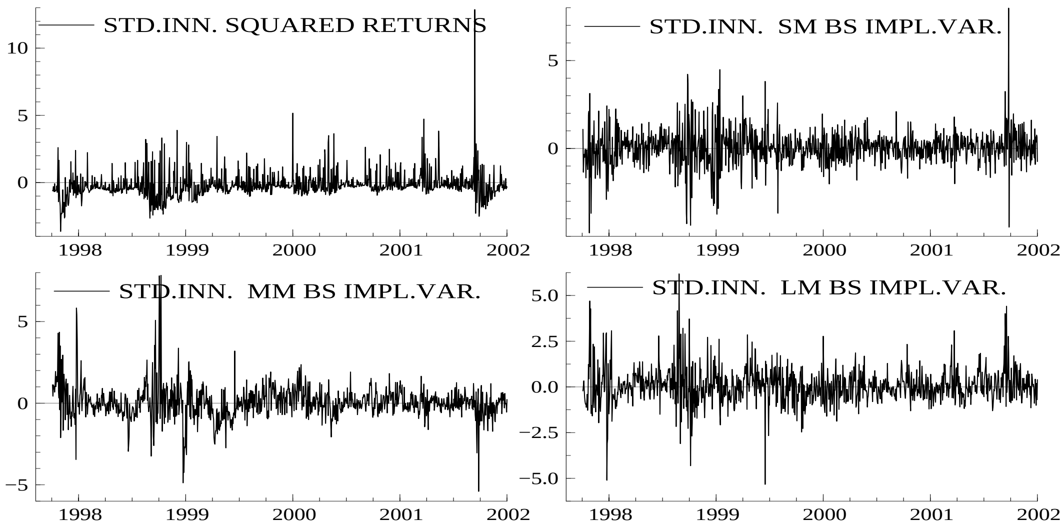


Figure 6.4: Standardized innovations (**1 CIR + 1 affine SV factor, case c**)

Table 6.4: Summary statistics std. innovations (**1 CIR + 1 affine SV factor, case c**)

	Std.inn. sq. return	*Std.inn. SM series*	*Std.inn. MM series*	*Std.inn. LM series*
Mean	-0.19	0.06	0.06	0.06
Std.deviation	0.98	1.00	1.04	0.96
AC(1)	0.09	-0.05	0.35	0.16
AC(2)	0.06	-0.07	0.23	0.07
AC(3)	0.14	-0.01	0.21	-0.02
AC(4)	0.11	-0.03	0.17	0.08
AC(5)	0.05	0.06	0.20	0.04
Cont.correlation	1.00	0.07	0.03	0.24
matrix		1.00	0.01	-0.01
			1.00	0.05
				1.00

The table reports the mean, standard deviation, autocorrelation coefficients (AC) up to order 5, and the contemporaneous correlation matrix of the standardized innovations.

Figure 6.5 shows the smoothed disturbances. A comparison with figure 5.7 for 1-factor CIR SV confirms the much improved data description. But again, these pictures seem to indicate a lack of sufficient volatility dynamics.

[27] Based on the Ljung-Box Q-statistic we still reject the null of zero autocorrelation up to order 5 at 5% significance for each innovation series, with respective p-values of 0.00, 0.03, 0.00, 0.00.

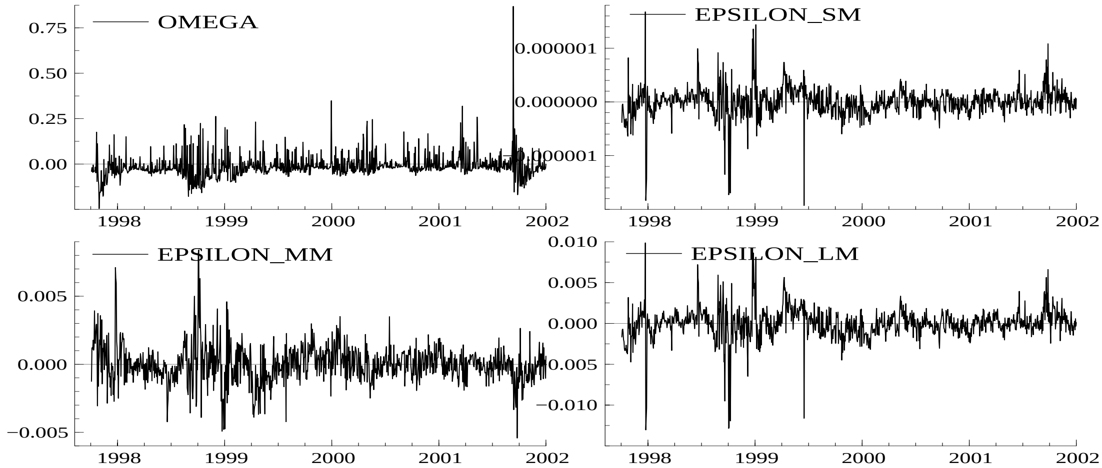


Figure 6.5: Smoothed disturbances (**1 CIR + 1 affine SV factor, case c**).

7. FTSE100-index data: Results for 3-factor affine SV

The specification analysis for the estimated 2-factor SV model (6.2) naturally motivates the consideration of 3-factor SV models. This section reports estimation results for 3-factor SV models that allow for volatility feedback. Results for the 3-factor OU SV model are discussed in section 5.3 of the previous chapter.

We first briefly discuss results for the 1 OU + 2 affine SV factors case, then for the 1 CIR + 2 affine SV factors case. Within the class of affine SV models, the latter specification appears most adequate for the FTSE100-index data. As such, we end with an extensive specification analysis to investigate if the 3-factor affine SV model with 1 CIR and 2 affine factors still lacks important features that are present in the data, but which it cannot (sufficiently well) describe.

7.1 Results 3-factor SV model with 1 OU and 2 affine factors

We first assume that 1 OU and 2 affine factors are driving the stock volatility. In light of Chernov et al. (2003), our main interest here is to investigate if the persistent factor features volatility feedback or not, now that we take account of the dynamic misspecification of the 2-factor models. We consider the following SV specification [28]

$$\sigma_t^2 = x_{1t} + x_{2t} + x_{3t}, \quad d\sigma_t^2 = dx_{1t} + dx_{2t} + dx_{3t} \qquad (7.1)$$

with

$$dx_{1t} = k_1(\theta_1 - x_{1t})dt + \sigma_{11}dW_{1t} \qquad (\mathbb{P})$$

$$dx_{2t} = -k_2 x_{2t}dt + \sigma_{21}dW_{1t} + \sqrt{\alpha_2 + \beta_{22}x_{2t}}\,dW_{2t} \qquad (\mathbb{P})$$

$$dx_{3t} = -k_3 x_{3t}dt + \sigma_{31}dW_{1t} + \sigma_{32}\sqrt{\alpha_2 + \beta_{22}x_{2t}}\,dW_{2t} + \sqrt{\alpha_3 + \beta_{33}x_{3t}}\,dW_{3t}\,. \qquad (\mathbb{P})$$

[28] Imposing $n = 3$ and $\theta_2 = \theta_3 = 0, \sigma_{12} = \sigma_{13} = \sigma_{23} = 0, \sigma_{22} = \sigma_{33} = 1, \alpha_1 = 1, \boldsymbol{\beta}_1 = \mathbf{0}$, together with $\beta_{21} = \beta_{23} = \beta_{31} = \beta_{32} = 0$ in the general multifactor SV specification (2.2)-(2.3) leads to this special case. Note that this specification is close in line with the 3-factor OU SV specification considered in the last chapter, which facilitates a comparison. In the 3-factor OU case we imposed $\theta_2 = \theta_3 = 0$ and took $\boldsymbol{\Sigma}$ lower diagonal for identification reasons. Here this seems not necessary, but given our experience with the 2-factor estimations from the last section we advocate this choice. We impose the restrictions $\sigma_{22} = \sigma_{33} = 1$ again for identification reasons.

The OU factor x_1 has constant volatility, the affine factors x_2 and x_3 feature level-dependent volatility. The factors are independent if $\sigma_{21} = \sigma_{31} = \sigma_{32} = 0$.

Table 7.1: Estimation results **3-factor affine SV models (7.1)** and **(7.2)** using all data

	Case 1	*Case 2*	*Case 3*		*Case 1*	*Case 2*	*Case 3*
θ_1	**0.0413**	**0.0413**	**0.0413**	Vol. returns	20.3%	20.3%	20.3%
k_1	**0.0017** (0.0018)	**0.0017** (0.0393)	**0.0066** (0.0068)	Vol-of-var.	0.607	0.969	0.3049
k_2	**3.65** (2.17)	**1.73** (24.2)	**2.73** (0.59)	Std.dev. x_{1t}	0.605	0.970	0.300
k_3	**11.9** (5.66)	**15.9** (67.1)	**16.5** (4.9)	Std.dev. x_{2t}	0.040	0.062	0.048
σ_{11}	**0.0354** (0.0061)	**0.0568** (0.3184)	**0.170** (0.035)	Std.dev. x_{3t}	0.028	0.027	0.027
σ_{21}	0	**-0.0738** (0.8383)	0	corr(x_{1t}, x_{2t})	0	-0.04	0
σ_{31}	0	**0.0204** (1.026)	0	corr(x_{1t}, x_{3t})	0	0.003	0
σ_{32}	0	**0.390** (5.274)	0	corr(x_{2t}, x_{3t})	0	0.05	0
α_2	**0.0117** (0.0051)	**0.0077** (0.154)	**0.0128** (0.0060)	Persist. x_1	0.99999	0.99999	0.99998
β_{22}	**0.302** (0.057)	**0.170** (0.653)	**0.290** (0.046)	Persist. x_2	0.9860	0.9934	0.9895
α_3	**0.0186** (0.0049)	**0.0214** (0.0621)	**0.0240** (0.0049)	Persist. x_3	0.9553	0.9407	0.9386
β_{33}	**0.319** (0.122)	**0.346** (0.306)	**0.283** (0.109)	Half-life x_1	405y	404y	105y
γ_1	**0.309** (0.533)	**-0.081** (3.99)	**-0.548** (1.29)	Half-life x_2	49	104	66
γ_2	**-2.58** (6.77)	**0.748** (168)	**1.25** (1.17)	Half-life x_3	15	11	11
γ_3	**-2.48** (12.5)	**-9.02** (116)	**-9.12** (13.7)	Std.dev. u_{1t}	0.0022	0.0035	0.0021
σ_{ω}	**0.0670** (0.0058)	**0.0672** (0.0063)	**0.0668** (0.0057)	Std.dev. u_{2t}	0.0067	0.0071	0.0070
$\sigma_{\varepsilon,SM}$	**0.00111** (0.00052)	**0.00011** (0.00184)	**0.00013** (0.00063)	Std.dev. u_{3t}	0.0083	0.0091	0.0093
$\sigma_{\varepsilon,MM}$	**0.00145** (0.00014)	**0.00154** (0.00076)	**0.00147** (0.00012)				
$\sigma_{\varepsilon,LM}$	**0.00195** (0.00028)	**0.00178** (0.00329)	**0.00178** (0.00024)	Loglikelih.	14,719	14,743	14,793

The table reports (restricted) parameter estimates (in boldface) with robust White (1982) QML standard errors in parentheses, resulting from estimating the conditional state space model associated with the 3-factor SV assumption using the combination of return, SM, MM and LM option data for three cases (see main text), together with the QMLEs of some other quantities of interest. Half-life is either reported in days or in years (if extension "y" is after the number). **Case 1** and **case 2** assume specification (7.1) with **1 OU + 2 affine** SV factors. **Case 3** assumes specification (7.2) with **1 CIR + 2 affine** SV factors.

The first two columns of table 7.1 report estimation results for two cases. *Case 1* assumes independent factors and zero correlation between the option errors $\{\varepsilon_{it}\}; i = SM, MM, LM$ at all times. *Case 2* assumes correlated factors, but still maintains the assumption of zero option error correlation. As motivated in the previous section, we restrict θ_1 to its moment estimate of 0.0413 prior to estimation. (*Case 3* is discussed in the next section.)

Allowing for correlated factors does not lead to a distinct optimum: The BFGS optimization routine does not converge and indicates that the quasi-loglikelihood is flat near the optimum. [29] In the "optimum", half of the gradients associated with the 18 parameters to be estimated is of order 10^{-5}, the other half is of order 10^{-4}. This explains the large standard errors in the estimates. Notice that the quasi-loglikelihood in the "optimum" does not seem much larger than in the independent factors case (1). Apparently, the signal in the data with respect to possible factor dependence is too weak to be convincingly filtered out (if it exists at all). Allowing for factor dependence seems to lead to an over-specification of the model for this dataset. Given our earlier results on insignificant option error correlation, we do not further pursue this extension here, as we expect it to lead to even more imprecision without further raising the loglikelihood much.

In some sense it is perhaps not so surprising that the signal with regard to factor dependence (if it exists) is so weakly present in the data. We have four observed time series, the return, SM, MM and LM option series. We assume these series to have been generated by our multifactor SV model, for which misspecification tests show that three factors are needed to obtain an adequate data description. These observed times series are obviously correlated. This correlation can be generated even if the three SV factors are independent, simply because each of the factors is implicitly present in each time series. The point is basically that we try to map four *observed* correlated time series to a complicated SV structure that features three *unobserved* time series (the factors), which need not necessarily be correlated to make the observed correlation possible. This may intuitively explain the weak dependence signal. Remember that the estimated 2-factor SV models yielded small but "significant" factor correlation. Given the illustrated lack of dynamics of those models however, this "significance" may well deceptively have been caused by this misspecification.

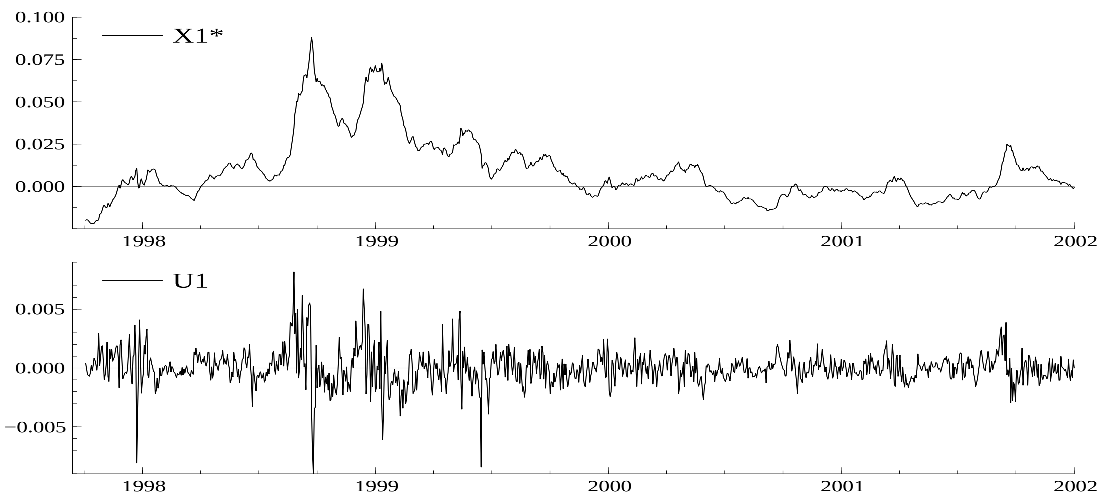


Figure 7.1: Smoothed OU factor x_{1t}^{*} and its daily unpredictable shock u_{1t} in **case 1**.

As in the 3-factor OU SV case, we focus on the independent factors case, case (1), as we feel most confident with the results obtained for that specification. Figure 7.1 shows the smoothed OU factor x_1^{*} and its daily unpredictable shock

[29] Technically speaking, there is *no improvement in line search* possible in the "optimum". Different starting values, reparametrizations and setting the convergence criteria looser do not lead to a clear-cut optimum either.

u_{1t}. The plot of u_{1t} reveals conditional heteroskedasticity. The OU assumption for the persistent factor x_1 is therefore not appropriate. A more realistic modeling of x_1 is by a CIR factor. To save space we do not further comment on the results for case (1). Instead, we directly consider the results for the 1 CIR + 2 affine SV factors specification in the next section.

7.2 Results 3-factor SV model with 1 CIR and 2 affine factors

If 1 CIR and 2 affine SV factors are driving the volatility, and if these factors are independent, the SV specification reads

$$\sigma_t^2 = x_{1t} + x_{2t} + x_{3t}, \; d\sigma_t^2 = dx_{1t} + dx_{2t} + dx_{3t} \tag{7.2}$$

with

$$dx_{1t} = k_1(\theta_1 - x_{1t})dt + \sigma_{11}\sqrt{x_{1t}}dW_{1t} \qquad (\mathbb{P})$$

$$dx_{2t} = -k_2 x_{2t}dt + \sqrt{\alpha_2 + \beta_{22}x_{2t}}dW_{2t} \qquad (\mathbb{P})$$

$$dx_{3t} = -k_3 x_{3t}dt + \sqrt{\alpha_3 + \beta_{33}x_{3t}}dW_{3t}\,. \qquad (\mathbb{P})$$

The third column of table 7.1, *Case 3*, reports estimation results for this specification assuming no option error correlation (recall our earlier discussion). Subsequently moving from the 3-factor OU SV assumption (Restrictions a) to the 1 OU + 2 affine SV assumption (case 1), to the 1 CIR + 2 affine SV assumption raises the quasi-loglikelihood from 14,500 to 14,719 to 14,793. Modeling x_1 as a CIR factor instead of an OU factor leads to a rather substantial rise it seems.

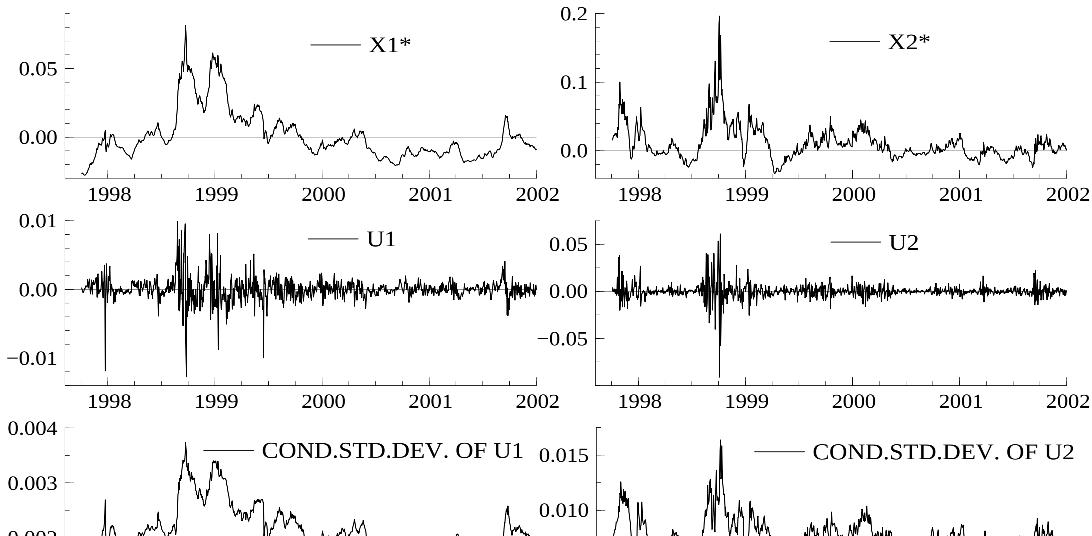


Figure 7.2: Smoothed x_{1t}^*, x_{2t}^*, daily factor shocks u_{1t} and u_{2t}, and conditional standard deviation of u_{1t} and u_{2t}. (SV model (7.2) with **1 CIR + 2 affine SV factors, case 3**).

Long-term, medium-term and short-term volatility trends

As in the 3-factor OU SV case (and case (1)), factors x_1, x_2 and x_3 can be interpreted as determining the long-term, medium-term and short-term volatility trends respectively. Shocks to the long-memory trend are virtually permanent. Shocks to the medium-term trend have a half-life of three months. The half-life of shocks to the fast mean-reverting factor x_3 is 11 days. Factor x_3 captures large volatility fluctuations in short periods of time: The shocks u_{3t} have largest variance. As in the 2-factor SV estimations, due to its high persistence, the CIR factor x_1 does not obey the Feller condition. x_1 virtually behaves as a random walk in discrete time, though features conditional heteroskedasticity.

Figure 7.2 shows the smoothed factors x_{1t}^* and x_{2t}^*, the daily unpredictable factor shocks u_{1t} and u_{2t}, and their conditional standard deviations (computed using (3.7)). Figure 7.3 shows x_{3t}^*, u_{3t}, $\sqrt{\text{var}[u_{3,t+\Delta t} \mid \mathcal{F}_t]}$, the smoothed stock variance $\sigma_t^2 = x_{1t} + x_{2t} + x_{3t}$ and stock volatility σ_t. The course of the factors looks similar as in the 3-factor OU SV case.

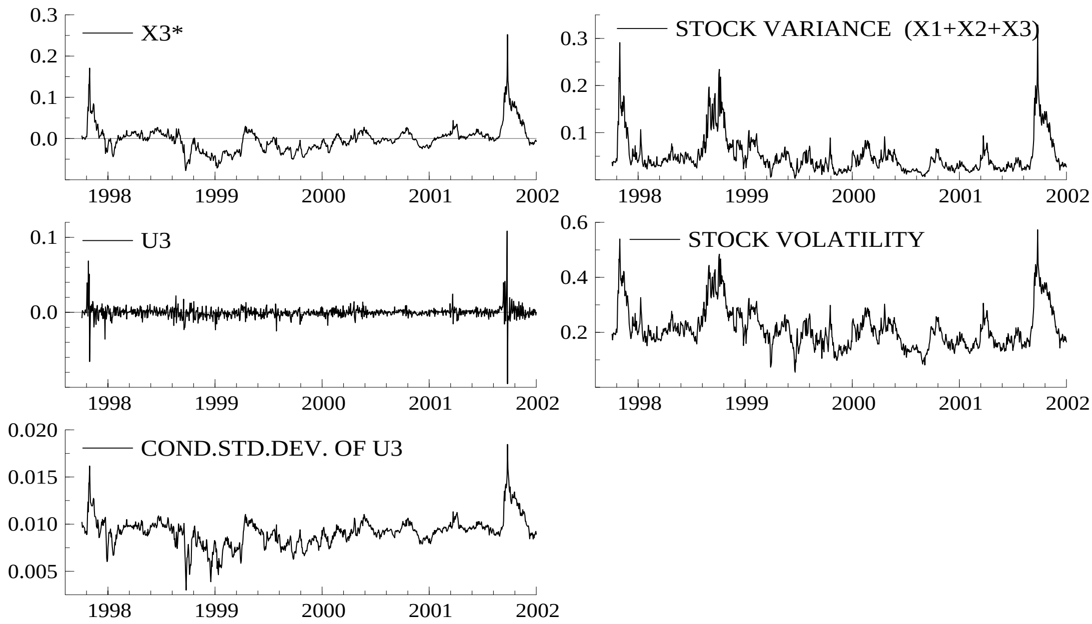


Figure 7.3: Smoothed x_{3t}^*, u_{3t} and conditional standard deviation of u_{3t}. Smoothed stock variance $\sigma_t^2 = x_{1t} + x_{2t} + x_{3t}$ and volatility σ_t .(**1 CIR + 2 affine SV factors, case (3)**).

Impact on option prices; VTS level, slope and curvature factors

A second interpretation of the factors concerns their impact on the prices of options of different maturity and the dynamics of the ATM volatility term structure. The interpretation is similar as in the 3-factor OU SV case. Figure 7.4 shows the factor loading functions $\partial\sigma_{implied,t}^2 / \partial x_{jt} = B_{1j}(\tau_t)/(\tau_t)$; $j = 1,2,3$, as functions of maturity (in years). These functions measure how option prices of different maturity respond to changes in the underlying volatility factors. The middle plot shows the instantaneous ceteris-paribus response of the BS implied variance to a one-standard-deviation shock in each of the factors, as a function of maturity, i.e. $(\mathbf{I}_n \odot \text{var}_{\mathbb{P}}[\mathbf{x}_t])^{1/2}\, \mathbf{B}_1(.)/(.)$. The lower plot shows the responses to factors x_2 and x_3 again, but without the response to factor x_1.

Due to its high persistence, long-memory factor x_1 affects option prices of all maturities in a similar way. A shock to x_1 results in parallel VTS shifts in which all implied volatilities increase with approximately the same amount (ceteris paribus). Factor x_2 impacts on all options as well, but its effect gradually tempers off the longer the option maturity. Fast mean-reverting factor x_3 has biggest impact on the prices of short-maturity options, and does not virtually impact on options with a maturity longer than half a year. As explained earlier (see section 5.3.2 of the last chapter), given its rapid mean reversion this makes sense.

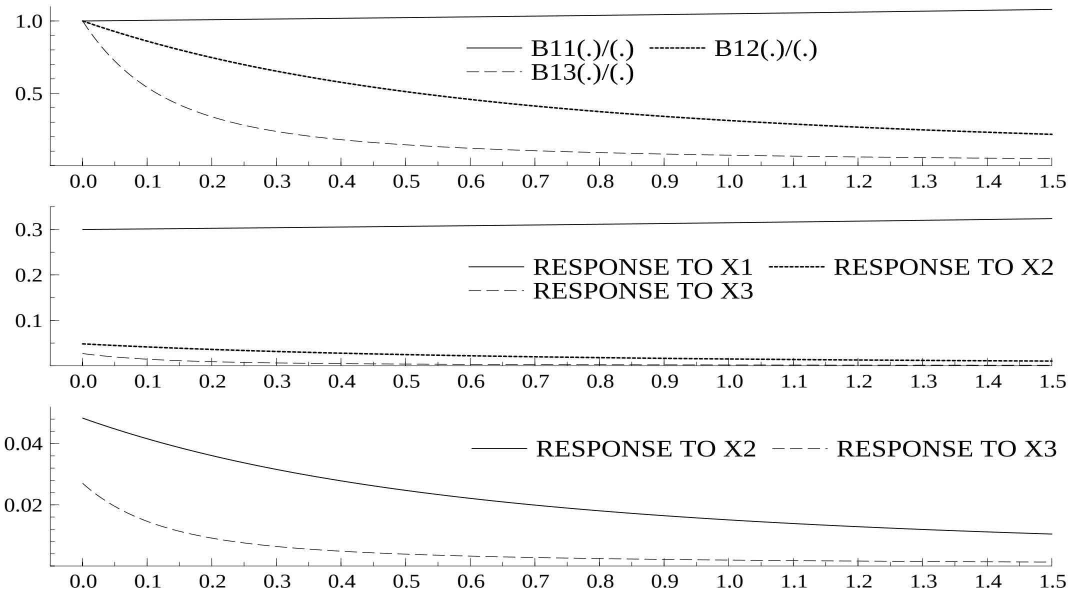


Figure 7.4. Upper plot: Reaction coefficients $B_{1j}(.)/(.); j=1,2,3$ as a function of option maturity (in years). Middle: Instant responses of the BS implied variances to 1-standard-deviation shocks in each of the factors, as a function of maturity. Lower: Same as middle plot, but without the response to x_1. (**1 CIR + 2 affine SV factors, case 3**).

Figure 7.5 confirms the earlier given interpretation to factors x_1, x_2 and x_3 as being largely associated with the level, slope and convexity dynamics of the ATM VTS respectively. The magnitude of the correlations between x_1, $-x_2$, x_3 and the level, slope and curvature is similar as in the 3-factor OU SV case.

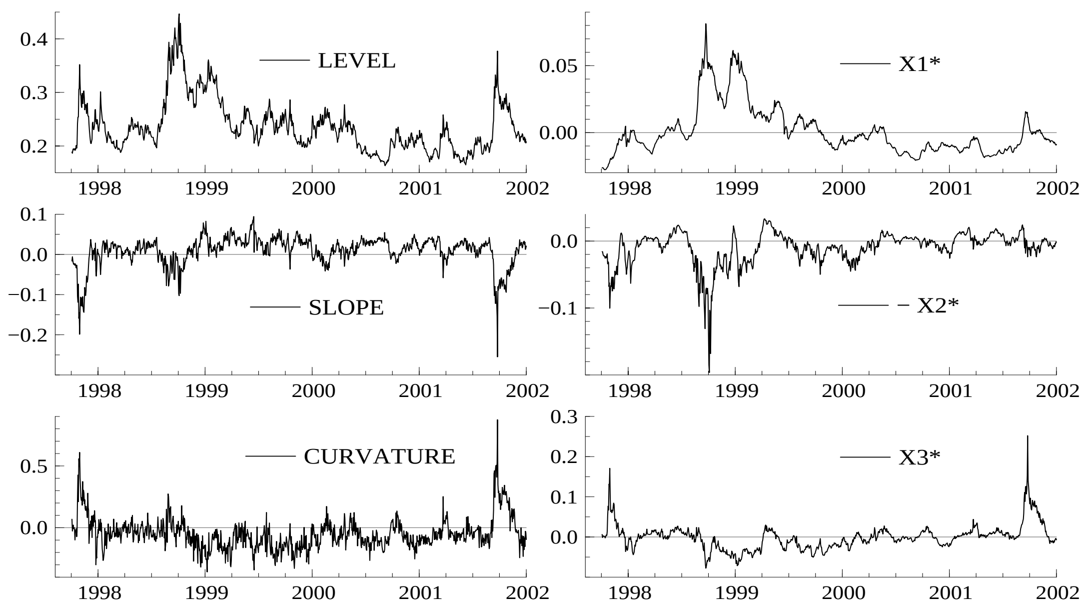


Figure 7.5. Left: Evolution of level, slope and curvature of the VTS through time. Right: Smoothed x_1^*, $-x_2^*$ and x_3^*. (**1 CIR + 2 affine SV factors, case 3**).

Market prices of risk and factor risk premia
In the general multifactor affine SV model, the expected spot return to be earned on an arbitrary, path-independent European-style derivative F is given by

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = \left[r_t + \beta_{F,S,t}\left(\sigma_t \gamma_{S,t}\right) + \boldsymbol{\beta}_{F,x,t}{}'\left(\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t}\right)\right] dt \,, \qquad (7.3)$$

with $\boldsymbol{\gamma}_{x,t} = \boldsymbol{\Lambda}_t \boldsymbol{\gamma}$ the vector of market prices of factor risk, and $\boldsymbol{\Sigma\Lambda}_t \boldsymbol{\gamma}_{x,t} = \boldsymbol{\Sigma\Lambda}_t^2 \boldsymbol{\gamma}$ the vector of volatility-factor risk premia. For the current special case, the market prices of risk associated with the CIR and affine factors are given by, respectively, $\gamma_{1t} = \gamma_1 \sqrt{x_{1t}}$, $\gamma_{2t} = \gamma_2 \sqrt{\alpha_2 + \beta_{22} x_{2t}}$ and $\gamma_{3t} = \gamma_3 \sqrt{\alpha_3 + \beta_{33} x_{3t}}$. The factor risk premia are given by $\sigma_{11} \gamma_1 x_{1t}$ for CIR factor x_1, and $\gamma_2 [\alpha_2 + \beta_{22} x_{2t}]$ respectively $\gamma_3 [\alpha_3 + \beta_{33} x_{3t}]$ for the affine factors x_2 and x_3.

As in the 3-factor OU SV case, γ_1 and γ_3 are estimated negative, γ_2 positive, and none of these parameters differs significantly from zero. Neither do they jointly: The Wald test statistic for testing the joint significance of γ_1, γ_2 and γ_3 equals 2.14, which is smaller than the 95%-critical value of 7.82 of its asymptotic chi-squared(3) distribution under the null hypothesis. Figure 7.6 shows the factor risk premia over time in percents per annum. According to the estimation results, the QMLE of the average risk premium associated with factor x_1, respectively x_2 and x_3, equals –0.38%, 1.60% and –22%. In the 3-factor OU SV case we found risk premia of –0.14%, 3.9% and –26%, which are rather comparable in magnitude. We refer to section 5.3.3 of the previous chapter for an interpretation and discussion of these risk premia.

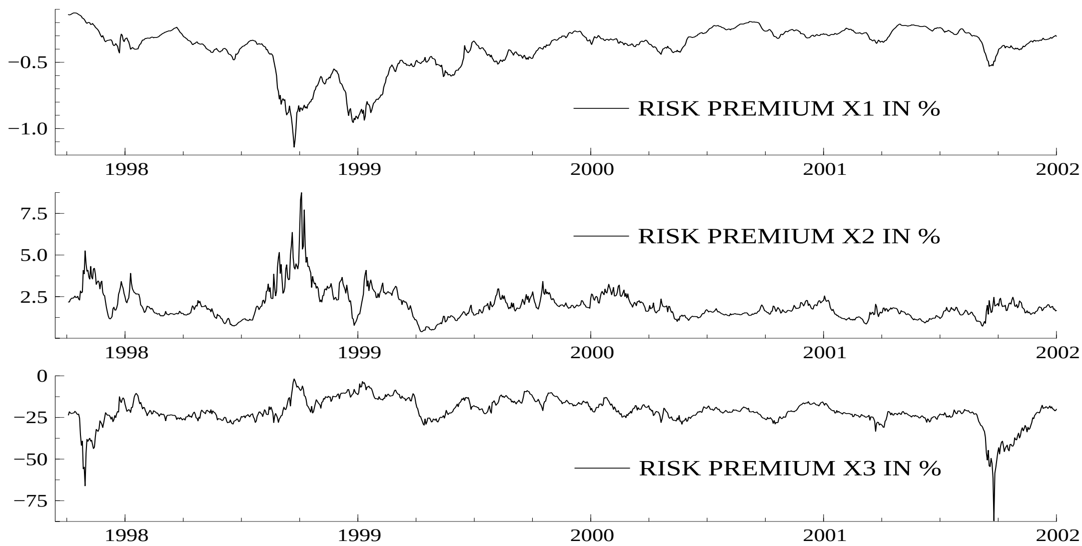


Figure 7.6. Risk premia (in % per annum) associated with the volatility factors x_1, x_2 and x_3 . (3-factor SV model (7.2) with **1 CIR + 2 affine SV factors, case 3**).

Fit of the volatility term structure
Figure 7.7 shows the in-sample fit of the VTS. In contrast to the 1-factor CIR SV results (figure 5.8) and the 2-factor SV results with 1 CIR and 1 affine factor (figure 6.3), the option pricing errors seem randomly distributed now, with no apparent systematic structure left. The improved fit is also obvious from table 7.2 as compared to tables 5.6 and 6.3. All pricing errors concentrate around zero now

and are generally small. Remarkably, the statistics are not much different from those obtained in the 3-factor OU SV case (see table 5.6 in the previous chapter). We refer to section 5.3.4 in the last chapter for an explanation of the logical pattern that has emerged in the option pricing errors, from having considered 1, 2 and 3 SV factors in succession.

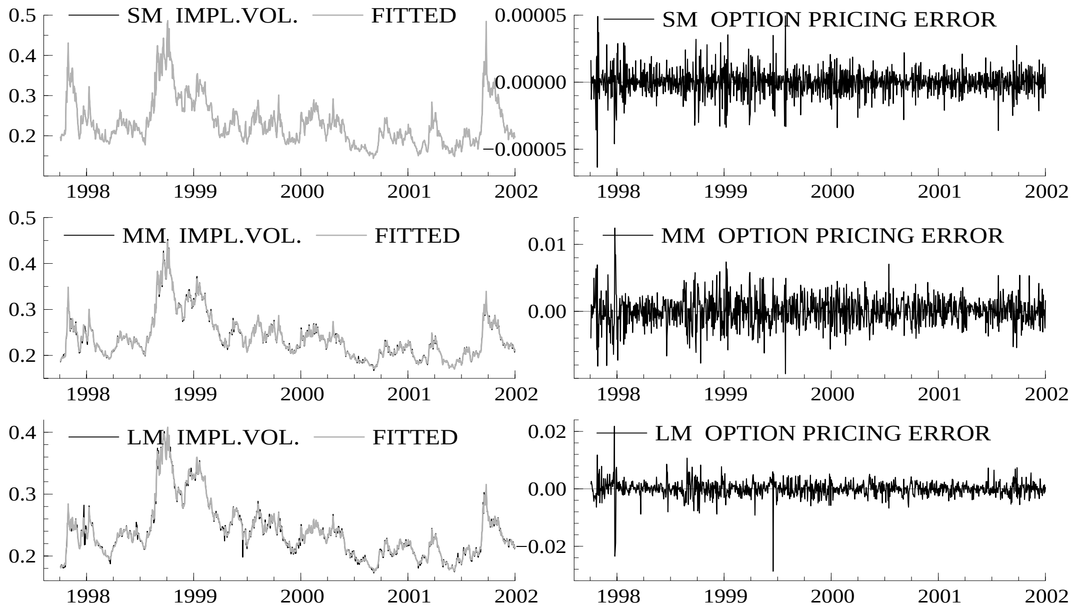


Figure 7.7: Fit of the VTS. Left: observed and fitted BS implied volatilities for each maturity. Right: difference, the pricing errors. (**1 CIR + 2 affine SV factors, case 3**).

Table 7.2: Summary stat. (abs.) pricing errors (**1 CIR + 2 affine SV factors, case 3**)

	Error SM	*Error MM*	*Error LM*	*|Error SM|*	*|Error MM|*	*|Error LM|*
Mean	0.00%	0.00%	0.00%	0.00%	0.17%	0.19%
Std.dev.	0.00%	0.22%	0.29%	0.00%	0.14%	0.22%

7.2.1 Specification analysis of the 3-factor affine SV model with 1 CIR and 2 affine volatility factors

This section considers diagnostic checking. Does the 3-factor affine SV model with one CIR and two affine volatility factors provide a satisfactory description of the joint FTSE100 stock-index and option data, or has the model still some important shortcomings? If it does, what modeling features does it lack? Given the evidence, we revisit compensation for volatility risk and outline some promising areas for future research.

Leverage effect

The leverage effect is not modeled. To get an idea of its magnitude, the correlation between the observed daily FTSE100-index returns and daily changes in the smoothed stock variance equals –0.66. We found an equal correlation in the 3-factor OU SV case. As already mentioned in the previous chapter, it seems unlikely that this misspecification invalidates most of our findings, given our focus on ATM options only. The next chapter shows by simulation that the VTS of ATM options is hardly affected by the leverage effect. (In contrast, leverage does affect the prices of in and out-of-the-money options.)

Remaining conditional heteroskedasticity

Figure 7.8 plots the smoothed disturbances of the state space model associated with the squared return and options equations. Compared to the 1-factor CIR SV case (figure 5.7) and the 2-factor SV case with 1 CIR and 1 affine factor (figure 6.5), the disturbances resemble white noise closest in the 3-factor SV case.

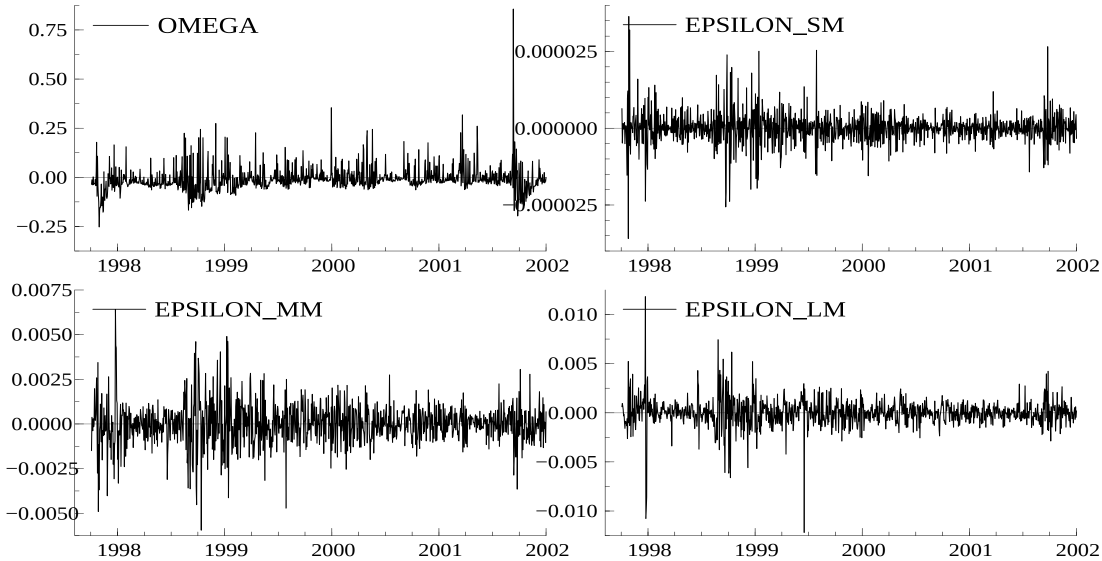


Figure 7.8: Smoothed disturbances (**1 CIR + 2 affine SV factors, case 3**).

Comparing figure 7.9 to figures 5.6 and 6.4 reveals that the standardized options innovations resemble white noise closest in the 3-factor SV case as well. Table 7.3 confirms this. Most autocorrelations are near zero now, although the Ljung-Box Q-test still rejects the null hypothesis of zero autocorrelation up to order 5 for all but the SM option series, with respective p-values of 0.000, 0.653, 0.000 and 0.000. The magnitude of the autocorrelations is similar as in the 3-factor OU SV case, which is as expected: Explicit modeling of level-dependent volatility-of-volatility (heteroskedasticity) does not influence the autocorrelation structure in any major way.

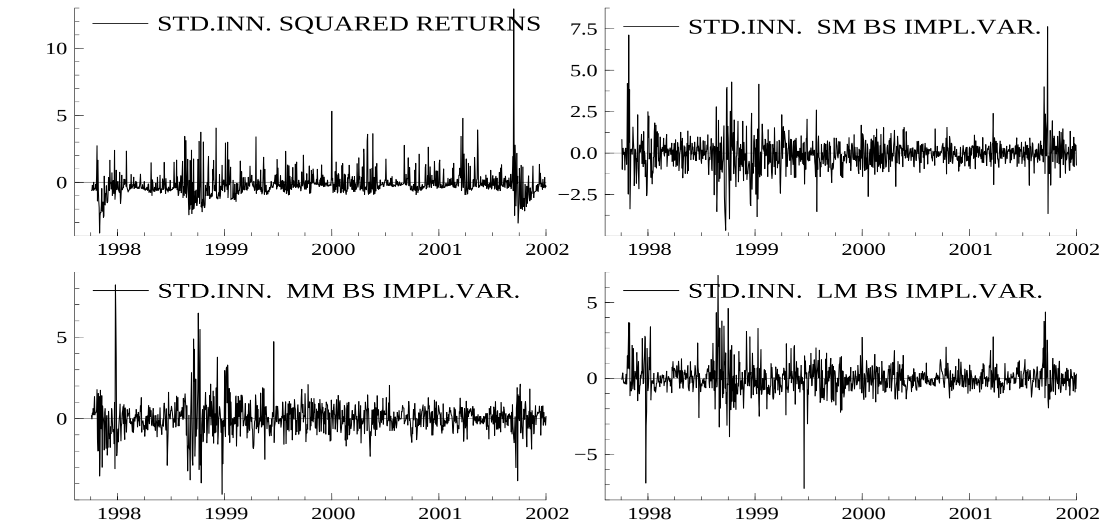


Figure 7.9: Standardized innovations (**1 CIR + 2 affine SV factors, case 3**).

Table 7.3: Summary stat. std.innovations (**1 CIR + 2 affine SV factors, case 3**)

	Std.inn. Sq. return	*Std.inn. SM series*	*Std.inn. MM series*	*Std.inn. LM series*
Mean	-0.14	-0.03	0.03	0.01
Std.deviation	0.99	0.98	1.02	0.97
AC(1)	0.09	0.04	0.16	0.10
AC(2)	0.06	-0.02	0.03	0.03
AC(3)	0.14	-0.01	-0.03	-0.08
AC(4)	0.11	-0.03	-0.01	0.05
AC(5)	0.05	0.00	0.05	0.03
Cont.correlation	1.00	0.10	0.04	0.22
matrix		1.00	-0.06	0.13
			1.00	-0.08
				1.00

The table reports the mean, standard deviation, autocorrelation coefficients (AC) up to order 5, and the contemporaneous correlation matrix of the standardized innovations.

The graphs of the smoothed disturbances and standardized options innovations nevertheless still seem to feature periods of moderate conditional heteroskedasticity, with the occasional "extreme" value occurring. This is the case notably in the fall of 1997 and 1998, and around September 11, 2001. These periods were characterized by unusual turmoil in world financial markets. If the 3-factor affine SV model with 1 CIR and 2 affine factors were correctly specified, conditional heteroskedasticity ought to be absent from the standardized options innovations.

Jumps in returns?

The plots suggest that the 3-factor affine SV specification does not capture all features of the joint FTSE100 stock-index and option data. For example, it might be the case that extending the SV specification with jumps in returns results in a more realistic data description.

Let us explore if such jumps are "really needed". [30] The upper graph of figure 7.10 shows the observed prewhitened FTSE100-index returns, i.e., $r_{t+\Delta t} - \hat{\mu}_t \Delta t$. Any real jumps in the daily returns do not seem to occur, the one exception perhaps being the larger return at September 11, 2001. To examine this in more depth, recall that if our model is correctly specified, the daily stock returns are conditionally approximately normally distributed, $r_{t+\Delta t} \mid \mathcal{F}_t \approx \mathcal{N}(\mu_t \Delta t, \sigma_t^2 \Delta t)$. The standardized returns are therefore approximately standard Gaussian,

$$\frac{r_{t+\Delta t} - \mu_t \Delta t}{\sigma_t \sqrt{\Delta t}} \mid \mathcal{F}_t \approx \mathcal{N}(0,1), \qquad (7.4)$$

but with a kurtosis somewhat larger than 3 due to SV. As such, the distribution of the standardized FTSE100-index returns provides intuitive insight if it is probable that jumps in the (daily) returns have occurred. The middle graph of figure 7.10 shows the estimated density of the standardized returns (computed using the parameter estimates and smoothed factors, which obviously has its drawbacks,

[30] Arguably, a realistic model should include jumps in returns. For example, Black Monday (October 19, 1987) may be considered a *jump event*. Moreover, our model generates continuous stock price paths, whereas real stock prices change discontinuously as well. From a practical point of view however, jumps create problems with hedging. Our modest aim here is to explore only if such jumps occur that frequently at the daily horizon, that they should evidently not be left out from the specification, if the model is used for derivative pricing and hedging in practice. Problem remains of course, how to identify such jumps: Is an unusual return truly a jump, or is it just a low-probability realization of the DGP? We will not perform a formal analysis here, but instead we look at the data and do some intuitive analysis.

but seems the best we can do), with a Gaussian density with the same mean and variance superimposed (dashed). The standardized returns have a mean of 0.04, a standard deviation of 0.97, a skewness of 0.33 and a kurtosis of 4.2. [31] The lower graph shows the standardized returns and the lines 1.96 and –1.96, which are exceeded on 4.4% of the days (which is rather close to 5%). These days seem rather randomly distributed over the sample. Extreme realizations of the standardized returns do not seem to occur, except perhaps at September 11. This suggests that the non-modeling of jumps in returns is not a serious shortcoming for the FTSE100-index data.

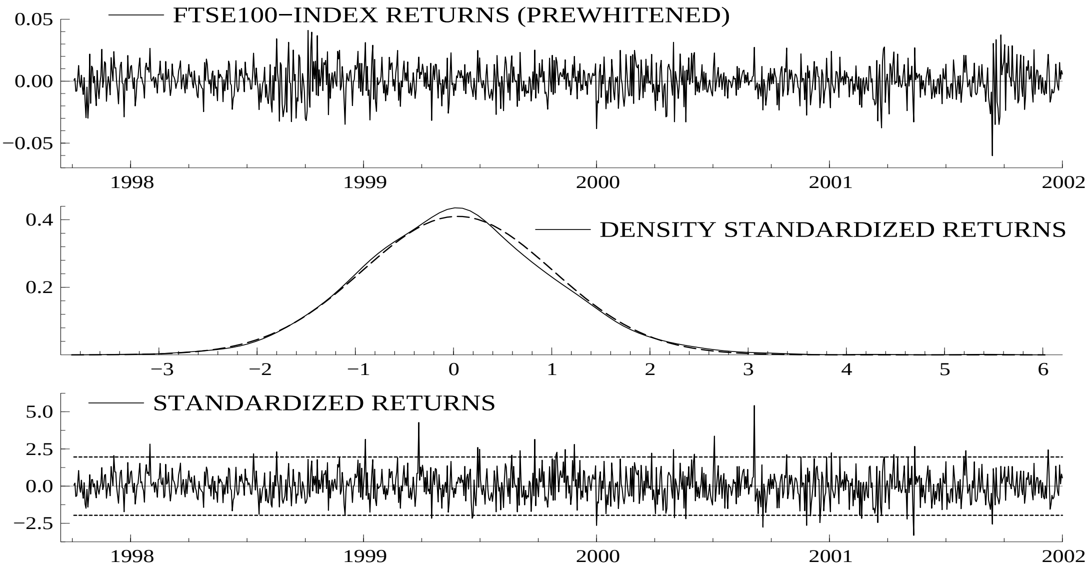


Figure 7.10: Prewhitened FTSE100-index returns, estimated density of the standardized returns (7.4) with a Gaussian density (dashed) superimposed. Lower plot: Standardized returns and the lines 1.96 and –1.96. (**1 CIR + 2 affine SV factors, case 3**).

Jump or volatility (jump) risk premium? Pan (2002) revisited

Given this jumps-in-returns analysis, if one is willing to believe in the absence of clear jumps in returns (or at least, in the presence of too infrequent jumps), this has important implications for risk compensation in practice. If there are no jumps in returns, there is no possible compensation for jumps-in-returns risk. We found the risk premium associated with the fast-mean reverting volatility factor to be largest negative, -22% on average. This suggests that investors are paying for short-term volatility risk. Pushing this further, based on the lower plot of figure 7.13 (which is explained below), this risk premium may perhaps partly be attributed to possible *jumps in volatility*. [32]

Now recall that Pan (2002), who fits the (1-factor SV) Heston (1993) model with jumps in returns (the Bates (2000) model) to weekly 1989-1996 S&P500 stock index and option data, finds that the jump risk premium dominates by far the volatility risk premium. However, Pan finds evidence indicating the possibility of jumps in volatility (which she did not model).

[31] The Jarque-Bera test statistic equals 83, which is larger than 5.99, indicating non-normality.

[32] Bates (2000) conjectures the presence of jumps in volatility. Eraker et al. (2003) estimate a 1-factor affine SV model with both jumps in returns and in volatility using S&P500 and Nasdaq100-index returns (but do not include options) for the periods 1980-1999 and 1985-1999 respectively. They find evidence of the presence of both types of jumps in the data.

So, may it be the case that these possible volatility jumps are actually responsible for the large negative jumps-in-returns risk premium that Pan finds? Pan (2002) ends her article with "*In other words, simply adding jumps in volatility will not replace the role of premia for price jumps.*" [..], but mentions a potential scope for jumps-in-volatility risk premia. However, in a footnote, Pan also mentions that "*Statistically, jumps in volatility and jumps in prices could result in similar price movements. Investor's aversion to these jump risks, however, could be quite different.*" [..], thereby pointing to her co-authored paper of Liu et al. (2001), who find that both types of jumps have distinctively different implications for investor's asset allocation.

Given our own empirical evidence, and given the evidence in Pan (2002), the last word on volatility-risk, jumps-in-returns risk and jumps-in-volatility risk compensation has evidently not been said. This forms a promising (and obviously econometrically challenging) area for future research.

3-factor affine SV still does not allow for fast enough changing volatility
The plot (7.9) of the standardized options innovations suggests that the 3-factor affine SV model with 1 CIR and 2 affine volatility factors cannot capture all conditional heteroskedasticity present in the data. We now examine the nature of this finding by performing additional volatility analysis.

Consider the change in the stock variance from day t to the next day, day $t+\Delta t$, defined as $\Delta\sigma_{t+\Delta t}^2 \equiv \sigma_{t+\Delta t}^2 - \sigma_t^2$. Given time-$t$ information, its conditional expectation equals

$$\mathbb{E}_{\mathbb{P}}\left[\Delta\sigma_{t+\Delta t}^2 \mid \mathcal{F}_t\right] = \mathbf{1}'\left(\exp\left[-\mathbf{K}_d\Delta t\right] - \mathbf{I}_n\right)\mathbf{x}_t^*, \tag{7.5}$$

which follows from (2.11) and the fact that $\sigma_t^2 = \mathbf{1}'\mathbf{x}_t = \mathbf{1}'\mathbf{x}_t^* + \mathbf{1}'\boldsymbol{\theta}$. The conditional variance of the stock variance change equals

$$\text{var}_{\mathbb{P}}\left[\Delta\sigma_{t+\Delta t}^2 \mid \mathcal{F}_t\right] = \mathbf{1}'\left[\mathbf{G}(\Delta t) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' + \sum_{i=1}^{n}\left(\left\{\left[\boldsymbol{\beta}_i \odot \mathbf{x}_t^*\right]' \otimes \mathbf{I}_n\right\}\mathcal{H}(\Delta t)\right) \odot \boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'\right]\mathbf{1}, \tag{7.6}$$

as $\text{var}_{\mathbb{P}}[\Delta\sigma_{t+\Delta t}^2 \mid \mathcal{F}_t] = \text{var}_{\mathbb{P}}[\sigma_{t+\Delta t}^2 \mid \mathcal{F}_t]$ with $\text{var}_{\mathbb{P}}[\sigma_{t+\Delta t}^2 \mid \mathcal{F}_t]$ given in (2.12). Figure 7.11 shows the smoothed stock variance σ_t^2, the day-to-day change in the stock variance $\Delta\sigma_{t+\Delta t}^2$, and the conditional standard deviation of $\Delta\sigma_{t+\Delta t}^2$ (and thus σ_t^2).

The evolution of the conditional standard deviation of σ_t^2 confirms the ability of the affine SV model to describe volatility feedback: If the volatility is large (resp. small), the volatility-of-volatility (the conditional standard deviation) is also large (small). What is not clear from this picture however, is the *extent* in which it can describe this feedback. In other words, can the volatility-of-volatility (i.e., the conditional standard deviation) under the affine SV specification change fast enough to be able to capture the occasional periods of very fast changing volatility?

To explore this in more detail, consider the conditional distribution of the daily stock variance change in the multifactor affine SV model. Although the mean and variance are known, the exact distribution of $\Delta\sigma_{t+\Delta t}^2$ is unknown. Its conditional distribution is approximately Gaussian however, because

$$\Delta\sigma_{t+\Delta t}^2 = \mathbf{1}'(\mathbf{x}_{t+\Delta t} - \mathbf{x}_t) \tag{7.7}$$

$$= \mathbf{1}'\left(\exp[-\mathbf{K}_d \Delta t] - \mathbf{I}_n\right)\mathbf{x}_t^* + \mathbf{1}' \int_t^{t+\Delta t} \exp\left[-\mathbf{K}_d (t + \Delta t - u)\right] \mathbf{\Sigma \Lambda}_u d\mathbf{W}_{x,u} \ ,$$

which follows from appendix B, (2.6). Given time-t information, for small Δt the integral in (7.7) -and hence $\Delta\sigma^2_{t+\Delta t}$ - is approximately normally distributed. The exact distribution may be regarded as a "random continuous mixture of normals", and is therefore characterized by (somewhat) fatter tails than the normal distribution. This analysis shows that the *z-score* of $\Delta\sigma^2_{t+\Delta t}$, which we define as

$$\text{z-score}_{t+\Delta t} \equiv \left(\Delta\sigma^2_{t+\Delta t} - \mathbb{E}_{\mathbb{P}}\left[\Delta\sigma^2_{t+\Delta t} \mid \mathcal{F}_t\right]\right) / \sqrt{\text{var}_{\mathbb{P}}\left[\Delta\sigma^2_{t+\Delta t} \mid \mathcal{F}_t\right]} \approx \mathcal{N}(0,1) \, , \quad (7.8)$$

is approximately standard Gaussian, given time-t information.

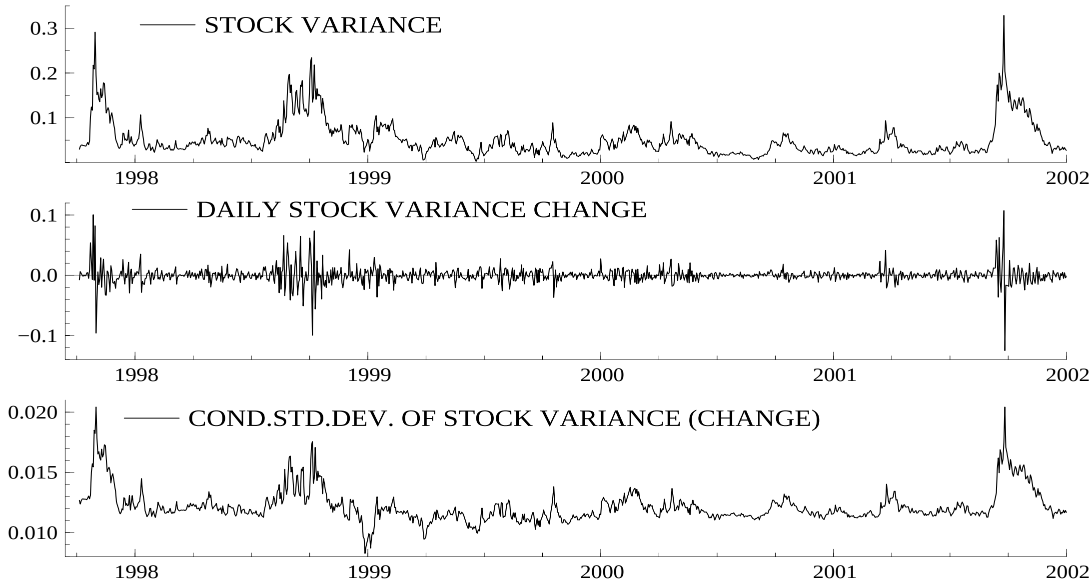


Figure 7.11: Smoothed stock variance σ_t^2, daily change in stock variance $\Delta\sigma_t^2$, and conditional standard deviation of σ_t^2 and $\Delta\sigma_t^2$ (which coincide, i.e. $\sqrt{(7.6)}$). (3-factor SV model (7.2) with **1 CIR + 2 affine SV factors, case 3**).

Given the estimation results for the 3-factor affine SV model, realized z-scores can be computed. These z-scores can be used to (intuitively) examine if the model is correctly specified for the FTSE100-index data, and in particular, if the volatility-of-volatility under the affine SV specification allows for sufficiently fast-changing volatility. If the distribution of the z-scores deviates too much from Gaussianity, then this suggests misspecification. [33]

[33] Evidently, care is in place with this procedure of detecting misspecification. First, to compute z-scores we use estimates of the parameters and unobserved factors (i.e. the smoothed factors) and not their true values. This seems the best we can do, and is common practice in other areas of econometrics; think about e.g. performing tests on the standardized returns of a GARCH model, or (indeed) our jumps-in-returns analysis above. There is a difference however: In the GARCH case we do at least observe the returns, whereas here we do not observe the realized $\Delta\sigma^2_{t+\Delta t}$, but use an estimate as well. Second, the z-score is conditionally only *approximately* normally distributed. Hence, it is not clear if small deviations from normality should either be attributed to misspecification, or to the approximate normality, and can therefore not be considered a sign of misspecification. Last, we will not perform any formal statistical tests to examine the misspecification, but base ourselves on "eyeball evidence".

Figure 7.12 shows the realized z-scores (together with the lines 1.96 and −1.96) and their estimated density, with a Gaussian density (dashed) of the same mean and variance superimposed. The density is much more peaked, and has fatter tails than the normal density. Table 7.4 reports summary statistics. The positive skewness of 1.15 indicates that positive stock variance changes occur more frequently than negative changes. The large kurtosis of 13.9 clearly indicates that the z-scores are far from Gaussian. [34] Large day-to-day stock variance changes are much more common than is actually allowed under the 3-factor affine SV specification. Notably in the fall of 1997 and 1998 and around September 11, 2001 do these large changes occur, which are exactly the periods for which the standardized options innovations in figure 7.9 still show remaining conditional heteroskedasticity.

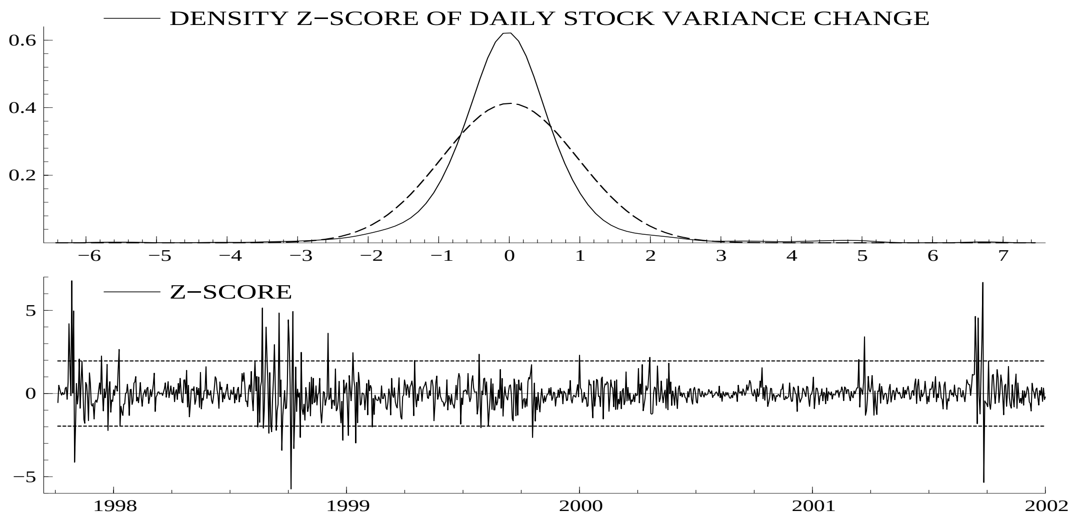


Figure 7.12: Estimated density of the z-scores (7.8) with a Gaussian density (dashed) superimposed. Lower plot: z-scores and the lines +1.96 and −1.96. (3-factor SV model (7.2) with **1 CIR + 2 affine SV factors, case 3**).

Table 7.4: Summary statistics z-scores

z-scores:	$\Delta\sigma_t^2$	Δx_{1t}	Δx_{2t}	Δx_{3t}
% abs.val.>1.96	4.5%	1.5%	2.8%	2.2%
Mean	0.005	0.015	0.026	-0.034
Std.deviation	0.97	0.65	0.79	0.76
Skewness	1.15	-0.35	0.88	0.62
Kurtosis	13.9	12.0	12.5	18.0

Statistics are reported for the z-scores associated with daily smoothed stock variance and factor changes. The first row reports the percentage of days on which the z-scores exceed −1.96 or 1.96.

The days on which the z-scores exceed the lines −1.96 and 1.96 (which occurs on 4.5% of the days, which is close to the Gaussian distribution-predicted 5%) are not randomly distributed, but clearly *cluster*. This clustering indicates that although the conditional standard deviation of σ_t^2 (the volatility-of-volatility) changes in these periods (see figure 7.11), it does not change fast enough, and therefore cannot sufficiently accommodate the large swings in volatility present in these periods.

[34] The associated Jarque Bera normality test statistic equals 5471, indicating that Gaussianity is strongly rejected. Note however that the right conditions for this test to be properly applied are not satisfied.

Is any of the factors specifically responsible for this remaining conditional heteroskedasticity? To explore this, we can compute z-scores for the daily changes in the smoothed volatility factors in a similar way as in (7.8). The distribution of these z-scores may again be contrasted to the standard normal distribution as an indication of misspecification. (However, the normal approximation becomes worse the less persistent the factor.) The summary statistics reported in table 7.4 indicate severe deviations from normality. The z-scores shown in figure 7.13 reveal that the medium-term and short-term volatility trend factors x_2 and x_3 are largely responsible for the remaining heteroskedasticity.

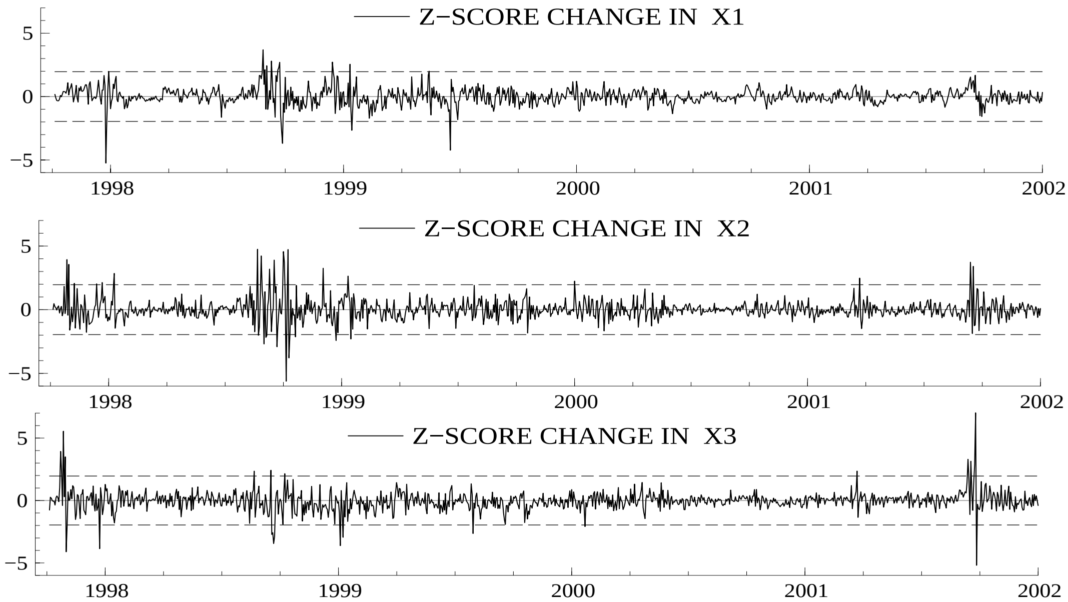


Figure 7.13: z-scores associated with daily changes in the smoothed factors and the lines +1.96 and –1.96. (3-factor SV model (7.2) with **1 CIR + 2 affine SV factors, case 3**).

Abandon the affine SV specification?

The analysis above may be regarded as evidence that the affine SV model cannot adequately describe relatively brief periods of large volatility changes in the FTSE100-index data, not even with three volatility factors. This complements the evidence in Chernov et al. (2003), who find that a 2-factor *exponential-affine* SV model without jumps yields a better fit of daily 1953-1999 DJIA stock-index data (but not option data), than do 1 or 2-factor affine SV models, or SV models with jumps in returns. Chernov et al. argue that jumps in returns are not needed because the exponential function allows more violent volatility than under the square-root (the affine) specification, which is especially appealing in times of increased market stress. (Note that this may additionally explain why price jumps are found to be needed in 1-factor affine SV models to improve on the fit.)

So, for practical (banking) purposes, should we abandon the affine SV specification and focus on e.g. the exponential-affine SV specifications instead? [35] It seems not. There are two important problems with the 2-factor exponential-

[35] Another class of multifactor volatility models based on a superposition of non-Gaussian, pure jump OU Lévy processes has recently been introduced by Barndorff-Nielsen and Shephard (2001b, 2002), see section 3.3.4 of chapter III. Brockwell and Davis (2001) and Brockwell (2001) propose an extension to these models, termed CARMA(p,q) models, which are continuous-time, pure jump Lévy process-driven ARMA SV models. Tauchen and Todorov (2004) discuss simulation methods for such models.

affine SV stock price model considered by Chernov et al. (2003). First, the specification cannot be used directly, due to problems with sufficiency conditions for the existence of a solution to the stock price SDE. Second, the affine SV specification yields almost “closed-form” option pricing formulas (see also Heston (1993) and Duffie et al. (2000)). The exponential-affine SV specification does not. This seriously complicates option pricing (and probably also estimation).

A multifactor affine SV model with jumps in volatility
A reinspection of figure 7.13 shows a number of very large z-scores associated with fast mean-reverting factor x_3 around the times of the beginning of the Asian crisis and September 11, 2001. This suggests that a 3-factor affine SV specification with jumps in volatility (which does not have the problems associated with the exponential-affine SV specification) may be a promising candidate as an even more realistic model for the FTSE100-index data. We leave this for future research.

8. Summary

Level-dependent volatility-of-volatility or volatility feedback, is a distinct characteristic of the FTSE100-index data, as evidenced in the last chapter. OU SV does not feature volatility shocks that are conditionally heteroskedastic. This chapter therefore considers affine SV processes that feature volatility feedback. The simplest example is the square root or CIR SV process, also known as Heston (1993) volatility. The 1-factor CIR SV option pricing model coincides with the Heston (1993) model, except that it does not model the leverage effect.

To be able to use the information in the data more efficiently regarding periods of observed conditional heteroskedasticity, we extend the state space estimation method (as developed in chapter III) to better exploit the characteristics of volatility feedback-featuring SV processes. We present the conditional state space model, which can be analyzed by the Extended Kalman filter, and which can be estimated by Extended Kalman filter QML.

Monte Carlo study based on the 1-factor CIR SV model
We perform an extensive Monte Carlo study towards the 1-factor CIR SV derivative pricing model, based on five sets of parameter values. We simulate time series of daily squared stock returns, 10-minute-based realized volatilities (48 intraday sampling points), and short-maturity at-the-money call options. We estimate the conditional state space model associated with 1-factor CIR SV using five different types of simulated data: Only squared return data, only RV data, only option data, both return - option data, and both RV and option data.

The main results are summarized as follows. Using squared returns only for estimation performs worst. Squared returns are noisy. The signal with regard to the DGP appears particularly weakly present in the data. QML performs poorly in this case, which is also attributed to the asymmetric distribution of the squared returns. The use of squared return data only for estimation is generally not to be advocated, not even if we restrict θ to its moment estimate $\hat{\theta}_{SqR}$. (Besides, the market price of volatility risk cannot be estimated in this case.)

Using RV data only, means a substantial reduction in bias and MSE. Nevertheless, both are still large. The volatility is much better filtered out than in the squared return case. If RV data only is used, θ should be restricted to its moment estimate $\hat{\theta}_{RV}$, computed as the mean realized variance over the sample. (In this case the market price of volatility risk cannot be estimated either however.)

Using option data only, means a dramatic improvement to using either only squared returns or RV data. Option data is much more informative on all parameters. The bias and MSE of the estimates are dramatically less, and the volatility evaluation criteria show much better volatility extraction.

The combination squared return – option data in turn, outperforms the use of option data only. The bias and MSE decrease further, and the volatility evaluation criteria further improve. The bias in the mean stock variance parameter θ, the volatility mean-reversion parameter k and the volatility-of-volatility parameter σ seems modest. Regarding the bias in the price of volatility risk parameter γ, the results are more mixed for the different parameter sets. There are efficiency gains from including squared return in addition to option data.

Obvious additional efficiency gains are obtained from using the combination RV and option data for estimation. This combination yields the most favorable volatility evaluation criteria, which indicate that the state space model performs well in extracting the underlying latent volatilities. Overall speaking, the combination of RV and option data yields the best estimation results.

The simulation results further show that the precision in the estimates of the market price of volatility risk parameter γ is poor. The combination RV – option data yields the most precise estimates. This confirms our own empirical findings and those of others in the literature.

Empirical results for affine SV models based on the FTSE100-index data
Confronting the 1-factor CIR SV model to the real-world FTSE100-index data shows the following. Using return and SM option data jointly, yields results that are largely similar as those obtained in the comparable 1-factor OU SV case. As such, we focus on the main differences. The CIR process obeys the Feller condition. The volatility persistence of 0.97 is estimated lower than the 0.993 found in the OU case. The smoothed CIR volatilities show signs of a quicker response to news than the OU volatilities.

Investors investing in FTSE100-index derivatives pay for volatility risk. This confirms our theory on volatility risk compensation, regarding their aim of consumption smoothing. The time-varying volatility risk premium averages at –21% per annum (OU: -15%). Writing SM ATM straddles yields an expected return of 195% per annum (OU: 174%), but involves substantial risk.

Diagnostic checks confirm the improved data description by 1-factor CIR SV as compared to 1-factor OU SV. Volatility clustering is better taken account of. Nonetheless, 1-factor CIR SV is still not capable of fully capturing the periods of rapidly changing volatility at the times of the Asian crisis, the near collapse of LTCM and September 11, 2001. As in the 1-factor OU SV case, the model overprices the longer-dated options out of sample. Modeling conditional heteroskedasticity does not resolve this. The reason is insufficient volatility dynamics.

Extending to a 2-factor SV specification with 1 OU and 1 affine factor, shows that the OU assumption for the persistent factor is misspecified. This contrasts to the findings of Chernov et al. (2003). A 2-factor SV specification with 1 CIR and

1 affine factor does much better. Due to its slow mean-reversion, the Feller condition associated with the CIR factor is violated. Specification tests show however, that this 2-factor volatility model still lacks sufficient dynamic structure.

Estimating a 3-factor SV model with 1 OU and 2 affine volatility factors, indicates that the shocks to the persistent, OU factor feature conditional heteroskedasticity as well. All volatility factors thus feature level-dependent volatility.

A 3-factor affine SV specification with 1 CIR and 2 affine, independent volatility factors appears to fit the joint FTSE100-index data best. The interpretations of the factors are the same as in the 3-factor OU SV case. Long-memory factor x_1 determines the long-term volatility trend, factor x_2 the medium-term, and fast mean-reverting factor x_3 the short-term volatility trend. Moreover, x_1, x_2 and x_3 are largely associated with the level, slope and convexity dynamics of the volatility term structure, respectively. Consequently, their impact on option prices of different maturity is qualitatively also the same. This is as expected. Modeling level-dependent volatility-of-volatility should and does not alter the earlier OU SV findings. The fit of the VTS is adequate. The option pricing errors do not show any obvious unmodeled systematics, concentrate near zero and are typically small. The average risk premia associated with factors x_1, x_2 and x_3 equal –0.38%, 1.60% and –22% per annum respectively (OU: -0.14%, 3.9% and –26%).

Within the affine class of SV models, the 3-factor model with 1 CIR and 2 affine volatility factors is thus found to be most appropriate for the joint FTSE100-index data. An extensive specification analysis shows that the model can still be improved upon however. First, the leverage effect is not modeled. It is clearly present in the data. (This misspecification is unlikely to invalidate most of our findings, given our focus on ATM options only. The next chapter deals with the leverage effect in a 1-factor SV model.) Second, although the affine SV model describes volatility feedback, it cannot fully capture the periods of fast-changing volatility present in the data, not even with three volatility factors. Large day-to-day volatility changes occur more often than the 3-factor affine SV specification allows for. Factors x_2 and x_3 are largely held responsible for the remaining unmodeled conditional heteroskedasticity. Extending the affine SV specification with jumps in volatility seems promising in this respect. Future research should clarify this.

Appendix

Va. System matrices conditional state space model

The conditional state space model associated with the general multifactor affine SV derivative pricing model, in which both the squared returns and the three option series are jointly used for estimation reads

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{a}_t + \boldsymbol{H}_t{}'\boldsymbol{\xi}_t + \boldsymbol{w}_t & \boldsymbol{w}_t &\sim (\boldsymbol{0}, \boldsymbol{R}) \quad t = \Delta t, \ldots, T\Delta t \\ \boldsymbol{\xi}_{t+\Delta t} &= \boldsymbol{d} + \boldsymbol{F}\boldsymbol{\xi}_t + \boldsymbol{v}_{t+\Delta t} & \boldsymbol{v}_{t+\Delta t} &\sim (\boldsymbol{0}, \boldsymbol{Q}) \\ & & \boldsymbol{v}_{t+\Delta t} \mid \mathcal{F}_t &\sim (\boldsymbol{0}, \boldsymbol{Q}_t), \end{aligned} \tag{I.1}$$

in which the error series $\{\boldsymbol{w}_t\}$ and $\{\boldsymbol{v}_{t+\Delta t}\}$ are both serially and mutually uncorrelated at all points in time, and in which the system matrices read[36] (with $m = 4,\ n = n,\ r = n + 1$):

$$(mx1)\,\boldsymbol{y}_t \equiv \begin{bmatrix} \frac{1}{\Delta t}(r_t - \hat{\mu}_{t-\Delta t}\Delta t)^2 \\ \sigma^2_{implied,1t} \\ \sigma^2_{implied,2t} \\ \sigma^2_{implied,3t} \end{bmatrix}, \qquad (mx1)\,\boldsymbol{a}_t \equiv \begin{bmatrix} 0 \\ \mu_v + \frac{1}{\tau_{1t}}[A_1(\tau_{1t}) + \mathbf{B}_1(\tau_{1t})'\boldsymbol{\theta}] \\ \mu_v + \frac{1}{\tau_{2t}}[A_1(\tau_{2t}) + \mathbf{B}_1(\tau_{2t})'\boldsymbol{\theta}] \\ \mu_v + \frac{1}{\tau_{3t}}[A_1(\tau_{3t}) + \mathbf{B}_1(\tau_{3t})'\boldsymbol{\theta}] \end{bmatrix}$$

$$(mxr)\,\boldsymbol{H}_t' \equiv \begin{bmatrix} 1 & \mathbf{0}' \\ 0 & \frac{1}{\tau_{1t}}\mathbf{B}_1(\tau_{1t})' \\ 0 & \frac{1}{\tau_{2t}}\mathbf{B}_1(\tau_{2t})' \\ 0 & \frac{1}{\tau_{3t}}\mathbf{B}_1(\tau_{3t})' \end{bmatrix}, \quad (rx1)\,\boldsymbol{\xi}_t \equiv \begin{bmatrix} \frac{1}{\Delta t}(r_t - \hat{\mu}_{t-\Delta t}\Delta t)^2 \\ \mathbf{x}_t^* \end{bmatrix}, \quad (mx1)\,\boldsymbol{w}_t \equiv \begin{bmatrix} 0 \\ \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}$$

$$(mxm)\,\boldsymbol{R} \equiv \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{\varepsilon 1} & \rho_{12}\sigma_{\varepsilon 1}\sigma_{\varepsilon 2} & \rho_{13}\sigma_{\varepsilon 1}\sigma_{\varepsilon 3} \\ 0 & \rho_{12}\sigma_{\varepsilon 1}\sigma_{\varepsilon 2} & \sigma^2_{\varepsilon 2} & \rho_{23}\sigma_{\varepsilon 2}\sigma_{\varepsilon 3} \\ 0 & \rho_{13}\sigma_{\varepsilon 1}\sigma_{\varepsilon 3} & \rho_{23}\sigma_{\varepsilon 2}\sigma_{\varepsilon 3} & \sigma^2_{\varepsilon 3} \end{bmatrix}, \qquad (rx1)\,\boldsymbol{d} \equiv \begin{bmatrix} \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} \\ \mathbf{0} \end{bmatrix},$$

$$(rxr)\,\boldsymbol{F} \equiv \begin{bmatrix} 0 & \boldsymbol{\delta}' \\ \mathbf{0} & \exp[-\mathbf{K}_d\Delta t] \end{bmatrix}, \qquad (rx1)\,\boldsymbol{v}_{t+\Delta t} \equiv \begin{bmatrix} \omega_{t+\Delta t} \\ \mathbf{u}_{t+\Delta t} \end{bmatrix},$$

$$(rxr)\,\boldsymbol{Q} \equiv \begin{bmatrix} \sigma^2_\omega & \mathbf{0}' \\ \mathbf{0} & \mathbf{G}(\Delta t) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' \end{bmatrix}, \qquad (rxr)\,\boldsymbol{Q}_t \equiv \begin{bmatrix} \sigma^2_\omega & \mathbf{0}' \\ \mathbf{0} & \mathrm{var}_{\mathbb{P}}[\mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t] \end{bmatrix}.$$

The expression for $\mathrm{var}_{\mathbb{P}}[\mathbf{u}_{t+\Delta t} \mid \mathcal{F}_t]$ is given in (3.7). We take μ_v the same for each option series, as they are all ATMF. We moreover allow for possible non-zero contemporaneous correlation between the option errors $\{\varepsilon_{SM,t}\}$, $\{\varepsilon_{MM,t}\}$ and $\{\varepsilon_{LM,t}\}$.

[36] For notational convenience we index the option series with 1, 2 and 3 instead of SM, MM, and LM.

- VI -

Leverage Effect and the Heston (1993) Model

1. Introduction

So far we have assumed a zero correlation between stock returns and volatility shocks. This assumption is generally violated in practice. When studying stock (market) data, one typically finds an asymmetric response of volatility to good and bad news. [1] Negative stock returns tend to increase volatility more than positive returns of the same magnitude. This effect, known as the *leverage effect* (Black (1976)), translates into a negative correlation between stock returns and volatility shocks.

In option prices, the leverage effect is apparent from the *volatility skew* or *smirk*: If the associated Black-Scholes implied volatilities of a certain maturity are plotted against moneyness, the typical pattern since the 1987-crash is skew-shaped. [2] SV with no leverage effect implies a U or smile-shaped BS implied volatility curve (Renault and Touzi (1996)). In contrast, SV with leverage effect induces a smirk or skew-shaped curve (Heston (1993)). [3]

We found strong indications of the presence of the leverage effect in the FTSE100-index data. The reason why we ignored leverage in the multifactor SV model in the first place was because it complicates matters dramatically, mainly with regard to our estimation method (when considering multiple factors). Nonetheless, a valid issue to rise is, if negligence of leverage has influenced our empirical results in any major way.

The main aim of this chapter is to study the impact of the false assumption of no leverage effect. Our focus on ATM FTSE100-index options only was implicitly driven by our expectation that leverage is not really important for the pricing of ATM options (besides the fact that these options are most liquid and have largest

[1] See, e.g., Black (1976), French, Schwert and Stambaugh (1987), Nelson (1991), Bollerslev, Chou and Kroner (1992), Heston (1993), Glosten, Jagannathan and Runkle (1993), Bollerslev, Engle and Nelson (1994), and Franses and Van Dijk (2000) among many others. In contrast, in foreign exchange markets the assumption of a zero correlation between exchange rate returns and exchange rate volatility shocks seems adequate; see chapter II, section 8. Indeed, our multifactor affine SV derivative pricing model and estimation method seem promising for empirical research in foreign exchange markets as well.

[2] See, e.g., Pan (2002), Eraker, Johannes and Polson (2003), and Jones (2003) for such plots.

[3] We illustrate this by simulation in sections 5.2.2 and 5.2.4. Recent evidence suggests that the Heston (1993) model cannot sufficiently adequately fit empirical volatility skews, especially not for short-maturity options. The observed skews tend to be (much) steeper; see Pan (2002), Jones (2003) and Eraker et al. (2003). Jumps in returns and jumps in volatility further steepen the volatility skew (as explained below). This forms a promising area for future research.

vega). The Monte Carlo evidence reported in this chapter suggests that this seems indeed the case. Our focus hereby is on the Heston (1993) model.

Specifically, we show that the volatility term structure of ATM options is not much affected by the leverage effect. In contrast, the prices (and hence VTS) of out-of-the-money and in-the-money options are strongly affected by leverage. Moreover, we perform a Monte Carlo study in which we simulate squared return, RV and ATM option data from the Heston model (hence with leverage effect), and next use the conditional state space model for estimation (which assumes no leverage) to investigate what the bias is. As we use the same parameter sets I-V as for the Monte Carlo study towards the 1-factor CIR SV model (i.e., the Heston model without leverage) in the previous chapter, this allows for a clear bias comparison. Although we do find additional bias in general, overall speaking, the additional bias seems rather modest. Monte Carlo evidence thus at least suggests that our multifactor SV estimation results reported in previous chapters seem not very much biased towards negligence of the leverage effect.

In section 7 of chapter II, we gave an economic rationale based on the multifactor SV model for why investors appear willing to pay for volatility risk. The argument given there builds on the existence of a leverage effect, which the multifactor SV model assumes absent. As such, the argument given there is theoretically not fully just. A second aim of this chapter is to show that the argument still stands if leverage is explicitly taken into account.

The outline of this chapter is as follows. Section 2 presents a 1-factor affine SV derivative pricing model with leverage effect. If there is no leverage, the model coincides with the 1-factor SV special case of the multifactor SV model. Moreover, the Heston (1993) model is a special case.

Section 3 analyzes the model in a similar way as we analyzed the multifactor SV model in chapter II. We consider arbitrage-free derivative pricing and hedging, and derive pricing formulas for European call and put options. We give expressions for the stock and volatility betas of a derivative, and explore the returns investors are expected to earn when trading assets in this market. We explain the differences induced by the presence of leverage as opposed to assuming it absent. We present our theory for the existence of a negative volatility risk premium, based on the existence of a leverage effect.

Section 4 derives a linear state space representation from the model to examine how leverage impacts on the estimation method (in the 1-factor SV case). As before, we linearize the model-implied call price around a suitably chosen linearization point. Simulation evidence next directs us to eventually use the same conditional state space model for estimation as derived from the multifactor no-leverage SV model.

Section 5 specializes to the Heston model and performs a Monte Carlo study. We start with examining how the model parameters affect the stock return distribution, the volatility process and what their impact is on option prices, i.e., the shape of the *implied volatility surface.* We next investigate the performance of the conditional state space model based on simulated data from the Heston model, in a similar way as in previous chapters. We comment on the additional bias induced by neglecting leverage in the estimation method. Again, using squared return data only performs worst, the combination RV – option data best.

Section 6 summarizes. The appendix states the Heston (1993) call price formula and derives expressions for the derivative betas in the model with leverage effect.

2. A 1-factor affine SV option pricing model with leverage effect, and the Heston (1993) model

Consider an arbitrage-free, frictionless financial market in which two basic assets are traded, a cash bond B and a dividend-paying stock S. [4] Uncertainty is driven by a 2-dimensional generalized $\mathbb{P}$-Brownian motion process $\{\bar{\mathbf{W}}_t; t \geq 0\}$ given by

$$\bar{\mathbf{W}}_t = \begin{pmatrix} \bar{W}_{S,t} \\ \bar{W}_{x,t} \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2} & \rho \\ 0 & 1 \end{pmatrix} \begin{pmatrix} W_{S,t} \\ W_{x,t} \end{pmatrix} \Leftrightarrow \begin{cases} d\bar{W}_{S,t} = \sqrt{1-\rho^2} dW_{S,t} + \rho dW_{x,t} \\ d\bar{W}_{x,t} = dW_{x,t} \end{cases} \tag{2.1}$$

which is derived from the standard $\mathbb{P}$-Brownian motion process $\{\mathbf{W}_t; t \geq 0\}$ with $\mathbf{W}_t = (W_{S,t}, W_{x,t})'$, and which is defined on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. Although the Brownian motions W_S and W_x are independent, the Brownian motions $\bar{W}_S$ and $\bar{W}_x$ are correlated, with correlation coefficient $\rho = \text{corr}_{\mathbb{P}}[\bar{W}_{S,t}, \bar{W}_{x,t}]$. [5] Notice that $\bar{W}_x = W_x$.

The value of the cash bond evolves as

$$dB_t = r_t B_t dt, \tag{2.2}$$

whereas, under market measure $\mathbb{P}$, the ex-dividend stock price evolves as

$$dS_t = \mu_t S_t dt + \sigma_t S_t d\bar{W}_{S,t}. \qquad (\mathbb{P}) \tag{2.3}$$

The stock pays a (deterministic) dividend yield of q_t per annum at time t. Under $\mathbb{P}$ the value of the tradable reinvestment portfolio S^r follows the SDE

$$dS_t^r = (\mu_t + q_t) S_t^r dt + \sigma_t S_t^r d\bar{W}_{S,t}. \qquad (\mathbb{P}) \tag{2.4}$$

The stock volatility evolves randomly over time, according to the 1-factor affine SV specification $\sigma_t^2 = x_t$, with

$$dx_t = k(\theta - x_t)dt + \sigma\sqrt{\alpha + \beta x_t} d\bar{W}_{x,t} = k(\theta - x_t)dt + \sigma\lambda_t d\bar{W}_{x,t}, \quad (\mathbb{P}) \tag{2.5}$$

in which $\lambda_t \equiv \sqrt{\alpha + \beta x_t}$. This specification corresponds to the 1-factor SV special case of our multifactor SV model, but differs in one important aspect: It allows for correlation between stock returns and volatility shocks. In particular,

$$\text{corr}_{\mathbb{P}}\left[\frac{dS_t}{S_t}, d\sigma_t^2 \mid \mathcal{F}_t\right] = \rho, \tag{2.6}$$

---

[4] Our model description here is deliberately brief to save space. It is largely similar as for the multifactor SV model given in section 2 of chapter II, where further details can be found.

[5] From the *covariation process* of the Brownian motions involved, it follows that $(d\bar{W}_{S,t})(d\bar{W}_{x,t}) = \rho dt$, $(dW_{S,t})(dW_{x,t}) = 0$ and $\text{cov}_{\mathbb{P}}[d\bar{W}_{S,t}, d\bar{W}_{x,t} \mid \mathcal{F}_t] = \text{cov}_{\mathbb{P}}[d\bar{W}_{S,t}, d\bar{W}_{x,t}] = \rho dt$. Decomposition (2.1) could be "reversed", i.e., $d\bar{W}_{S,t} = dW_{S,t}$, $d\bar{W}_{x,t} = \sqrt{1-\rho^2} dW_{x,t} + \rho dW_{S,t}$ Decomposition (2.1) is most suited for our purposes however.

which follows from $\text{var}_{\mathbb{P}}[dS_t / S_t \mid \mathcal{F}_t] = \sigma_t^2 dt$ and $\text{cov}_{\mathbb{P}}[dS_t / S_t, d\sigma_t^2 \mid \mathcal{F}_t] = \rho\sigma\sigma_t\lambda_t dt$, and $\text{var}_{\mathbb{P}}[d\sigma_t^2 \mid \mathcal{F}_t] = \sigma^2\lambda_t^2 dt$. The leverage effect implies a negative value for ρ.

Risk-neutral measure

The absence of arbitrage requires the value of the discounted reinvestment portfolio to follow a martingale process under an equivalent probability measure $\mathbb{Q}$, i.e., the risk-neutral measure. As the bond - stock market is incomplete due to SV there exists an infinite number of such measures $\mathbb{Q}$. In order to guarantee unique derivative prices, an assumption on the market price of volatility risk is required. As before, we model the price of volatility risk as being proportional to the volatility function of factor x:

$$\gamma_{x,t} = \gamma\sqrt{\alpha + \beta x_t} = \gamma\,\lambda_t . \tag{2.7}$$

Risk-neutral measure $\mathbb{Q}$ is next defined from the Radon-Nikodym derivative

$$L_T \equiv \left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_T} = \exp\left(-\int_0^T \boldsymbol{\gamma}_{\rho,u}{}' d\mathbf{W}_u - \tfrac{1}{2}\int_0^T \boldsymbol{\gamma}_{\rho,u}{}'\boldsymbol{\gamma}_{\rho,u} du \right), \tag{2.8}$$

in which $\boldsymbol{\gamma}_{\rho,t} \equiv (\gamma_{1t}, \gamma_{2t})'$, and

$$\gamma_{1t} \equiv \frac{1}{\sqrt{1-\rho^2}}\left(\gamma_{S,t} - \rho\,\gamma_{x,t}\right), \qquad \gamma_{2t} \equiv \gamma_{x,t}, \qquad \gamma_{S,t} = \frac{\mu_t + q_t - r_t}{\sigma_t}, \tag{2.9}$$

with $\gamma_{S,t}$ the usual market price of stock risk (or Sharpe ratio). Under market measure $\mathbb{P}$, the process $\{\mathbf{W}_t^{\mathbb{Q}}; t \geq 0\}$ with $\mathbf{W}_t^{\mathbb{Q}} = (W_{S,t}^{\mathbb{Q}}, W_{x,t}^{\mathbb{Q}})'$, defined as

$$\mathbf{W}_t^{\mathbb{Q}} = \mathbf{W}_t + \int_0^t \boldsymbol{\gamma}_{\rho,u} du \Leftrightarrow \begin{cases} dW_{S,t}^{\mathbb{Q}} = dW_{S,t} + \gamma_{1t} dt \\ dW_{x,t}^{\mathbb{Q}} = dW_{x,t} + \gamma_{2t} dt \end{cases} \tag{2.10}$$

is a Brownian motion with drift function $\{\boldsymbol{\gamma}_{\rho,t}\}$. Under probability measure $\mathbb{Q}$, $\{\mathbf{W}_t^{\mathbb{Q}}; t \geq 0\}$ is a standard Brownian motion however.

The bond - stock market is arbitrage-free

Given this measure $\mathbb{Q}$, the bond – stock market is arbitrage-free, as we will now show. Under $\mathbb{P}$, the discounted reinvestment portfolio value $Z_t^r \equiv B_t^{-1} S_t^r$ follows

$$\begin{aligned} dZ_t^r &= (\mu_t + q_t - r_t) Z_t^r dt + \sigma_t Z_t^r d\bar{W}_{S,t} \\ &= (\mu_t + q_t - r_t) Z_t^r dt + \sigma_t Z_t^r \left(\sqrt{1-\rho^2} dW_{S,t} + \rho dW_{x,t}\right), \qquad (\mathbb{P}) \end{aligned} \tag{2.11}$$

whereas, under this particular choice of $\mathbb{Q}$, it follows from (2.10) and (2.9) that

$$\begin{aligned} dZ_t^r &= \left(\mu_t + q_t - r_t\right) Z_t^r dt + \sigma_t Z_t^r \left(\sqrt{1-\rho^2}\left[dW_{S,t}^{\mathbb{Q}} - \gamma_{1t} dt\right] + \rho\left[dW_{x,t}^{\mathbb{Q}} - \gamma_{2t} dt\right]\right) \\ &= \left(\mu_t + q_t - r_t - \sigma_t\left[\sqrt{1-\rho^2}\,\gamma_{1t} + \rho\gamma_{2t}\right]\right) Z_t^r dt + \sigma_t Z_t^r \left[\sqrt{1-\rho^2} dW_{S,t}^{\mathbb{Q}} + \rho dW_{x,t}^{\mathbb{Q}}\right] \\ &= \sigma_t Z_t^r \left(\sqrt{1-\rho^2} dW_{S,t}^{\mathbb{Q}} + \rho dW_{x,t}^{\mathbb{Q}}\right) \\ &\equiv \sigma_t Z_t^r d\bar{W}_{S,t}^{\mathbb{Q}} . \qquad (\mathbb{Q}) \end{aligned} \tag{2.12}$$

As $\mathbb{E}_{\mathbb{Q}}[\exp(\frac{1}{2}\int_0^T \sigma_u^2 du)] < \infty$, the driftless process $\{Z_t^r\}$ is a $(\mathbb{Q},\{\mathcal{F}_t\})$ – martingale, as required by the absence of arbitrage. The process $\{\bar{\mathbf{W}}_t^{\mathbb{Q}}; t \geq 0\}$ with

$$\bar{\mathbf{W}}_t^{\mathbb{Q}} = \begin{pmatrix} \bar{W}_{S,t}^{\mathbb{Q}} \\ \bar{W}_{x,t}^{\mathbb{Q}} \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2} & \rho \\ 0 & 1 \end{pmatrix} \begin{pmatrix} W_{S,t}^{\mathbb{Q}} \\ W_{x,t}^{\mathbb{Q}} \end{pmatrix} \Leftrightarrow \begin{cases} d\bar{W}_{S,t}^{\mathbb{Q}} = \sqrt{1-\rho^2} dW_{S,t}^{\mathbb{Q}} + \rho dW_{x,t}^{\mathbb{Q}} \\ d\bar{W}_{x,t}^{\mathbb{Q}} = dW_{x,t}^{\mathbb{Q}} \end{cases} \tag{2.13}$$

is a generalized $\mathbb{Q}$-Brownian motion. The correlation between the $\mathbb{Q}$-Brownian motions $\bar{W}_S^{\mathbb{Q}}$ and $\bar{W}_x^{\mathbb{Q}}$ equals $\rho = \text{corr}_{\mathbb{Q}}[\bar{W}_{S,t}^{\mathbb{Q}}, \bar{W}_{x,t}^{\mathbb{Q}}]$. [6]

The model under $\mathbb{Q}$
The following expressions directly link the correlated $\mathbb{P}$ and $\mathbb{Q}$-Brownian motions with each other (and follow from substituting (2.10) in (2.13), rewriting, and using (2.1) and (2.9)):

$$d\bar{W}_{S,t}^{\mathbb{Q}} = d\bar{W}_{S,t} + \gamma_{S,t} dt, \qquad d\bar{W}_{x,t}^{\mathbb{Q}} = d\bar{W}_{x,t} + \gamma_{x,t} dt. \tag{2.14}$$

Under $\mathbb{Q}$, the stock price and reinvestment portfolio value evolve according to

$$dS_t = (r_t - q_t) S_t dt + \sigma_t S_t d\bar{W}_{S,t}^{\mathbb{Q}} \qquad (\mathbb{Q}) \tag{2.15}$$

$$dS_t^r = r_t S_t^r dt + \sigma_t S_t^r d\bar{W}_{S,t}^{\mathbb{Q}}, \qquad (\mathbb{Q}) \tag{2.16}$$

the latter SDE, which illustrates the name risk-neutral measure for $\mathbb{Q}$. (These expressions follow from (2.3) and (2.4) and use (2.14) and (2.9).) Substituting (2.14) and (2.7) in the SDE (2.5) for x, followed by rewriting, yields the SDE the unobserved volatility factor follows under $\mathbb{Q}$:

$$dx_t = \tilde{k}(\tilde{\theta} - x_t) dt + \sigma \lambda_t d\bar{W}_{x,t}^{\mathbb{Q}} \qquad (\mathbb{Q}) \tag{2.17}$$

with

$$\tilde{k} = k + \sigma \gamma \beta, \quad \tilde{\theta} = (k\theta - \sigma \gamma \alpha) / \tilde{k}. \tag{2.18}$$

The risk-neutral parameters $\tilde{k}$ and $\tilde{\theta}$ exactly coincide with those in the 1-factor SV special case of our multifactor SV model, and are thus not affected by the leverage effect.

Heston (1993) model
If we take $\alpha = 0, \beta = 1$, the model considered here essentially reduces to the Heston (1993) model, with CIR volatility $dx_t = k(\theta - x_t)dt + \sigma\sqrt{x_t} d\bar{W}_{x,t}$ under $\mathbb{P}$. The market price of volatility risk reduces to $\gamma_{x,t} = \gamma\sqrt{x_t}$, and the risk-neutral parameters to $\tilde{k} = k + \sigma\gamma$ and $\tilde{\theta} = k\theta / \tilde{k}$. Heston assumes constant interest rates and no dividends however. Moreover, Heston's risk-neutral parameters are given by $\tilde{k} = k + \vartheta$ and $\tilde{\theta} = k\theta / \tilde{k}$, i.e. parameter σ is subsumed in ϑ.

3. Analysis of the model with leverage effect

This section analyzes the 1-factor affine SV model with leverage effect in a similar way as we analyzed the multifactor affine SV model (without leverage) in chapter II. Section 3.1 considers derivative pricing. Section 3.2 derives call and put option

[6] Girsanov's theorem allows for a drift change only; the variance and hence correlation structure are the same under $\mathbb{P}$ and $\mathbb{Q}$. (If this were not the case $\mathbb{P}$ and $\mathbb{Q}$ would not be equivalent probability measures.)

pricing formulas. Section 3.3 considers hedging. Section 3.4 deals with stock and volatility betas and asset returns. Section 3.5 shows the volatility risk premium to be negative on theoretical grounds, using an argument based on leverage.

3.1 Derivative pricing

Consider a general, tradable European-style path-independent derivative F written on the stock S, that pays off an amount $F_T = f(S_T)$ at some fixed future maturity date T. Our interest is in deriving the arbitrage-free time-t price F_t of such a claim for $t < T$. This price is given by some function $F_t = F(t, S_t, x_t)$ of t, S_t and x_t only. [7] From Itô's formula, under $\mathbb{P}$ F_t follows the SDE [8]

$$\begin{aligned} dF_t &= \left[\frac{\partial F_t}{\partial t} + \tfrac{1}{2}\sigma_t^2 S_t^2 \frac{\partial^2 F_t}{\partial S_t^2} + \tfrac{1}{2}\sigma^2\lambda_t^2 \frac{\partial^2 F_t}{\partial x_t^2} + \rho\sigma\sigma_t\lambda_t S_t \frac{\partial^2 F_t}{\partial S_t \partial x_t}\right]dt + \frac{\partial F_t}{\partial S_t}dS_t + \frac{\partial F_t}{\partial x_t}dx_t \\ &= \left[\frac{\partial F_t}{\partial t} + \mu_t S_t \frac{\partial F_t}{\partial S_t} + \frac{\partial F_t}{\partial x_t}k(\theta - x_t) + \tfrac{1}{2}\sigma_t^2 S_t^2 \frac{\partial^2 F_t}{\partial S_t^2} + \tfrac{1}{2}\sigma^2\lambda_t^2 \frac{\partial^2 F_t}{\partial x_t^2} + \rho\sigma\sigma_t\lambda_t S_t \frac{\partial^2 F_t}{\partial S_t \partial x_t}\right]dt \\ &\qquad + \sigma_t S_t \frac{\partial F_t}{\partial S_t} d\bar{W}_{S,t} + \sigma\lambda_t \frac{\partial F_t}{\partial x_t} d\bar{W}_{x,t}\,. \qquad (\mathbb{P}) \qquad (3.1) \end{aligned}$$

Enlarging the marketed assets with the derivative should not lead to arbitrage opportunities. If the derivative price were to evolve according to the SDE (3.1), then arbitrage would be possible. An arbitrage-free market restricts the drift of the SDE (3.1) as follows. Using (2.14), under $\mathbb{Q}$ F_t follows

$$\begin{aligned} dF_t &= \left[\frac{\partial F_t}{\partial t} + \left(\mu_t - \sigma_t\gamma_{S,t}\right)S_t \frac{\partial F_t}{\partial S_t} + \frac{\partial F_t}{\partial x_t}\left[k(\theta - x_t) - \sigma\lambda_t\gamma_{x,t}\right] + \tfrac{1}{2}\sigma_t^2 S_t^2 \frac{\partial^2 F_t}{\partial S_t^2} + \tfrac{1}{2}\sigma^2\lambda_t^2 \frac{\partial^2 F_t}{\partial x_t^2}\right]dt \\ &\quad + \left[\rho\sigma\sigma_t\lambda_t S_t \frac{\partial^2 F_t}{\partial S_t \partial x_t}\right]dt + \sigma_t S_t \frac{\partial F_t}{\partial S_t} d\bar{W}_{S,t}^{\mathbb{Q}} + \sigma\lambda_t \frac{\partial F_t}{\partial x_t} d\bar{W}_{x,t}^{\mathbb{Q}}\,. \qquad (\mathbb{Q}) \qquad (3.2) \end{aligned}$$

In the absence of arbitrage, the relative derivative price $\{B_t^{-1}F_t\}$ must be a $(\mathbb{Q}, \{\mathcal{F}_t\})-$martingale, and hence the drift function of $d(B_t^{-1}F_t) = B_t^{-1}[dF_t - r_t F_t dt]$ must equal zero. This leads to the following no-arbitrage restriction

$$\begin{aligned} r_t F_t &= \frac{\partial F_t}{\partial t} + \left(\mu_t - \sigma_t\gamma_{S,t}\right)S_t \frac{\partial F_t}{\partial S_t} + \frac{\partial F_t}{\partial x_t}\left[k(\theta - x_t) - \sigma\lambda_t\gamma_{x,t}\right] + \tfrac{1}{2}\sigma_t^2 S_t^2 \frac{\partial^2 F_t}{\partial S_t^2} \\ &\quad + \tfrac{1}{2}\sigma^2\lambda_t^2 \frac{\partial^2 F_t}{\partial x_t^2} + \rho\sigma\sigma_t\lambda_t S_t \frac{\partial^2 F_t}{\partial S_t \partial x_t}\,. \qquad (3.3) \end{aligned}$$

This restriction implicitly defines a parabolic PDE that the pricing function $F(t, S_t, x_t)$ must satisfy in the absence of arbitrage. Taking the claim payoff $F_T = f(S_T)$ into account, the unique pricing function can numerically be solved for. Under $\mathbb{Q}$, the arbitrage-free derivative price thus follows

$$dF_t = r_t F_t dt + \sigma_t S_t \frac{\partial F_t}{\partial S_t} d\bar{W}_{S,t}^{\mathbb{Q}} + \sigma\lambda_t \frac{\partial F_t}{\partial x_t} d\bar{W}_{x,t}^{\mathbb{Q}}\,. \qquad (\mathbb{Q}) \qquad (3.4)$$

---

[7] See section 3.1 of chapter II for a discussion on this.

[8] Here we used $(dS_t)(dx_t) = \rho\sigma\sigma_t\lambda_t S_t dt$. This follows from the covariation process $\{[S, x]_t\}$ of the semimartingales $\{S_t\}$ and $\{x_t\}$, which is derived from the covariation process $\{[\bar{W}_S, \bar{W}_x]_t\}$ of the Brownian motions $\bar{W}_S$ and $\bar{W}_x$, yielding $(d\bar{W}_{S,t})(d\bar{W}_{x,t}) = \rho dt$.

Finally, transforming back to the market measure using (2.14), yields the SDE the arbitrage-free derivative price follows under $\mathbb{P}$:

$$dF_t = \left[r_t + \left(\frac{S_t}{F_t} \frac{\partial F_t}{\partial S_t} \right) \sigma_t \gamma_{S,t} + \left(\frac{1}{F_t} \frac{\partial F_t}{\partial x_t} \right) \sigma \lambda_t \gamma_{x,t} \right] F_t dt + \frac{\partial F_t}{\partial S_t} \sigma_t S_t d\bar{W}_{S,t} + \frac{\partial F_t}{\partial x_t} \sigma \lambda_t d\bar{W}_{x,t} \,. \qquad (\mathbb{P}) \qquad (3.5)$$

PDE, martingale and SDF approaches to derivative pricing

Instead of solving the PDE (3.3), the derivative price may also be obtained from the martingale approach to derivative pricing. As no-arbitrage requires the relative prices of all tradable assets to follow martingale processes under $\mathbb{Q}$ (such that $F_t / B_t = \mathbb{E}_{\mathbb{Q}}[F_T / B_T \mid \mathcal{F}_t]$), the *risk-neutral valuation formula* applies:

$$F_t = B_t \, \mathbb{E}_{\mathbb{Q}} \left[B_T^{-1} F_T \mid \mathcal{F}_t \right] = \exp\left(-\int_t^T r_u du \right) \mathbb{E}_{\mathbb{Q}} \left[F_T \mid \mathcal{F}_t \right]. \qquad (3.6)$$

Alternatively, the stochastic discount factor (SDF) approach to derivative pricing may be used instead. This yields the *SDF valuation formula*

$$F_t = \exp\left(-\int_t^T r_u du \right) \mathbb{E}_{\mathbb{P}} \left[\frac{L_T}{L_t} F_T \mid \mathcal{F}_t \right] = \mathbb{E}_{\mathbb{P}} \left[\frac{M_T}{M_t} F_T \mid \mathcal{F}_t \right]. \qquad (3.7)$$

Here, $\{M_t; t \geq 0\}$ with $M_t = B_t^{-1} L_t$ is the stochastic discount factor process, in which $\{L_t; t \geq 0\}$ is the Radon-Nikodym process, as defined from (2.8) in an obvious way. All three approaches yield an identical arbitrage-free price for derivative F. Although the risk-neutral and SDF valuation formulas look exactly the same as in our multifactor SV option pricing model without leverage effect, the difference lies in the fact that the $\mathbb{Q}$-measures are different in both settings.

3.2 Call and put option valuation

Consider pricing a European call option C having strike K and maturity $T > t$, in the model with leverage effect. Heston (1993) derives the call price using characteristic functions (and essentially *Fourier transform analysis*), a method also pursued by Bakshi et al. (1997) and Scott (1997). Appendix VIa states the complicated-looking "closed-form" Heston call price formula. This section derives the call price from a different perspective, using the *law of iterated expectations*. The resulting formula is more useful for our purposes in later sections. Although the derivation is similar as in the multifactor SV model, it is more complicated.

From the risk-neutral valuation formula, the arbitrage-free call price C_t is given by

$$C_t = \exp\left(-\int_t^T r_u du \right) \mathbb{E}_{\mathbb{Q}} \left[\max\{0, S_T - K\} \mid \mathcal{F}_t \right]. \qquad (3.8)$$

Itô's formula yields the SDE for $\{\ln S_t\}$ under $\mathbb{Q}$ from (2.15). Integrating this SDE over $[t, T]$, using (2.13), and transforming back to $\{S_t\}$, yields for S_T under $\mathbb{Q}$

$$S_T = S_t \exp\left[\left(\bar{r}_t - \bar{q}_t - \tfrac{1}{2} \bar{\sigma}_t^2 \right) \tau_t + \sqrt{1 - \rho^2} \int_t^T \sigma_u dW_{S,u}^{\mathbb{Q}} + \rho \int_t^T \sigma_u dW_{x,u}^{\mathbb{Q}} \right] \quad (\mathbb{Q}) \qquad (3.9)$$

with

$$\tau_t = T - t\,, \quad \bar{r}_t = \frac{1}{\tau_t}\int_t^T r_u du\,, \quad \bar{q}_t = \frac{1}{\tau_t}\int_t^T q_u du\,, \quad \bar{\sigma}_t^2 = \frac{1}{\tau_t}\int_t^T \sigma_u^2 du\,, \qquad (3.10)$$

being the remaining option maturity, the average risk-free rate, dividend yield and stock variance, respectively. By (2.17) and decomposition (2.13), conditioning on the path of the $\mathbb{Q}$-Brownian motion $W_x^{\mathbb{Q}}$ over the interval $[t,T]$, implies that the path of $\{x_t\}$ and thus $\{\sigma_t^2\}$ is known over $[t,T]$ under $\mathbb{Q}$, and hence

$$\int_t^T \sigma_u dW_{S,u}^{\mathbb{Q}} \mid \left(\mathcal{F}_t, \{W_{x,u}^{\mathbb{Q}}\}\right) \sim \mathcal{N}\left(0, \int_t^T \sigma_u^2 du\right) \sim \bar{\sigma}_t \sqrt{\tau_t}\,\varepsilon \text{ with } \varepsilon \overset{(\mathbb{Q})}{\sim} \mathcal{N}(0,1)\,. \quad (3.11)$$

(Here $\{W_{x,u}^{\mathbb{Q}}\} \equiv \{W_{x,u}^{\mathbb{Q}}; t \le u \le T\}$.) Moreover, the second Itô integral in (3.9) is then a given constant. The distribution $S_T \mid \left(\mathcal{F}_t, \{W_{x,u}^{\mathbb{Q}}\}\right)$ is thus lognormal under $\mathbb{Q}$.

Given this lognormality, express the conditional call payoff in terms of ε to get

$$\max\{0, S_T - K\} \mid \left(\mathcal{F}_t, \{W_{x,u}^{\mathbb{Q}}\}\right) \qquad (3.12)$$
$$= \begin{cases} S_t \exp\left[\left(\bar{r}_t - \bar{q}_t - \tfrac{1}{2}\bar{\sigma}_t^2\right)\tau_t + \sqrt{1-\rho^2}\,\bar{\sigma}_t\sqrt{\tau_t}\,\varepsilon + \rho\int_t^T \sigma_u dW_{x,u}^{\mathbb{Q}}\right] - K, & \text{if } \varepsilon \ge -d_{2t,\rho} \\ 0, \text{ if } \varepsilon < -d_{2t,\rho} & \end{cases}$$

with

$$d_{2t,\rho} \equiv \frac{\ln\dfrac{S_t}{K} + \left(\bar{r}_t - \bar{q}_t - \tfrac{1}{2}\bar{\sigma}_t^2\right)\tau_t + \rho\displaystyle\int_t^T \sigma_u dW_{x,u}^{\mathbb{Q}}}{\sqrt{1-\rho^2}\,\bar{\sigma}_t\sqrt{\tau_t}}\,. \qquad (3.13)$$

It next follows that

$$\begin{aligned} &\mathbb{E}_{\mathbb{Q}}\left[\max\{0, S_T - K\} \mid \left(\mathcal{F}_t, \{W_{x,u}^{\mathbb{Q}}\}\right)\right] \\ &= S_t \exp\left[\left(\bar{r}_t - \bar{q}_t - \tfrac{1}{2}\bar{\sigma}_t^2\right)\tau_t + \rho\int_t^T \sigma_u dW_{x,u}^{\mathbb{Q}}\right] \int_{-d_{2t,\rho}}^{\infty} \exp\left[\sqrt{1-\rho^2}\,\bar{\sigma}_t\sqrt{\tau_t}\,\varepsilon\right] \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\varepsilon^2}\, d\varepsilon - K\,\Phi(d_{2t,\rho}) \\ &= S_t \exp\left[\left(\bar{r}_t - \bar{q}_t - \tfrac{1}{2}\bar{\sigma}_t^2\right)\tau_t + \rho\int_t^T \sigma_u dW_{x,u}^{\mathbb{Q}}\right] \exp\left[\tfrac{1}{2}\left(1-\rho^2\right)\bar{\sigma}_t^2\tau_t\right]\, \Phi(d_{1t,\rho}) - K\,\Phi(d_{2t,\rho}) \\ &= S_t \exp\left[(\bar{r}_t - \bar{q}_t - \tfrac{1}{2}\rho^2\bar{\sigma}_t^2)\tau_t + \rho\int_t^T \sigma_u dW_{x,u}^{\mathbb{Q}}\right] \Phi(d_{1t,\rho}) - K\Phi(d_{2t,\rho})\,, \qquad (3.14) \end{aligned}$$

with

$$d_{1t,\rho} \equiv d_{2t,\rho} + \sqrt{1-\rho^2}\,\bar{\sigma}_t\sqrt{\tau_t} = \frac{\ln\dfrac{S_t}{K} + \left(\bar{r}_t - \bar{q}_t + (\tfrac{1}{2} - \rho^2)\bar{\sigma}_t^2\right)\tau_t + \rho\displaystyle\int_t^T \sigma_u dW_{x,u}^{\mathbb{Q}}}{\sqrt{1-\rho^2}\,\bar{\sigma}_t\sqrt{\tau_t}}\,, \quad (3.15)$$

and $\Phi(.)$ being the cumulative Gaussian distribution function. The expression for the Riemann integral after the first equality in (3.14) follows from changing the integration variable ε to the variable $\xi \equiv \varepsilon - \sqrt{1-\rho^2}\,\bar{\sigma}_t\sqrt{\tau_t}$, such that an integral over the standard normal density is obtained.

The call price in the 1-factor SV model with leverage effect then becomes:

$$\begin{aligned} C_t &= \exp(-\bar{r}_t\tau_t)\mathbb{E}_{\mathbb{Q}}\left[\max\{0, S_T - K\} \mid \mathcal{F}_t\right] \qquad (3.16) \\ &= \exp(-\bar{r}_t\tau_t)\mathbb{E}_{\mathbb{Q}}\left(\mathbb{E}_{\mathbb{Q}}\left[\max\{0, S_T - K\} \mid \mathcal{F}_t, \{W^{\mathbb{Q}}_{x,u}\}\right] \mid \mathcal{F}_t\right) \\ &= \mathbb{E}_{\mathbb{Q}}\left[S_t \exp\left(\left[-\bar{q}_t - \tfrac{1}{2}\rho^2\bar{\sigma}_t^2\right]\tau_t + \rho\int_t^T \sigma_u dW^{\mathbb{Q}}_{x,u}\right)\Phi(d_{1t,\rho}) - K\exp(-\bar{r}_t\tau_t)\Phi(d_{2t,\rho}) \mid \mathcal{F}_t\right]. \end{aligned}$$

Rewriting the call price in terms of the BS pricing function
Call price formula (3.16) can be rewritten in terms of the Black-Scholes function. Defining

$$S_{t,\rho} \equiv S_t \exp\left(-\tfrac{1}{2}\rho^2\bar{\sigma}_t^2\tau_t + \rho\int_t^T \sigma_u dW^{\mathbb{Q}}_{x,u}\right), \quad \bar{\sigma}^2_{t,\rho} \equiv (1-\rho^2)\bar{\sigma}_t^2, \qquad (3.17)$$

it follows

$$\begin{aligned} C_t &= \mathbb{E}_{\mathbb{Q}}\left[S_{t,\rho}\exp(-\bar{q}_t\tau_t)\Phi(d_{1t,\rho}) - K\exp(-\bar{r}_t\tau_t)\Phi(d_{2t,\rho}) \mid \mathcal{F}_t\right] \qquad (3.18) \\ &= \mathbb{E}_{\mathbb{Q}}\left[BS(S_{t,\rho}, K, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}^2_{t,\rho}) \mid \mathcal{F}_t\right]. \end{aligned}$$

Here, $BS(S_{t,\rho}, K, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}^2_{t,\rho}) = S_{t,\rho}\exp(-\bar{q}_t\tau_t)\Phi(d_{1t,\rho}) - K\exp(-\bar{r}_t\tau_t)\Phi(d_{2t,\rho})$ is the usual Black-Scholes pricing function. In terms of (3.17), $d_{1t,\rho}$ and $d_{2t,\rho}$ have the familiar form

$$d_{1t,\rho} = \frac{\ln\dfrac{S_{t,\rho}}{K} + (\bar{r}_t - \bar{q}_t + \tfrac{1}{2}\bar{\sigma}^2_{t,\rho})\tau_t}{\bar{\sigma}_{t,\rho}\sqrt{\tau_t}}, \quad d_{2t,\rho} = d_{1t,\rho} - \bar{\sigma}_{t,\rho}\sqrt{\tau_t}. \qquad (3.19)$$

If there is no leverage effect ($\rho = 0$), $S_{t,\rho} = S_t$, $\bar{\sigma}^2_{t,\rho} = \bar{\sigma}^2_t$, $d_{1t,\rho} = d_{1t}$, $d_{2t,\rho} = d_{2t}$, and the call price formula of section 3.2 in chapter II is obtained, i.e., $C_t = \mathbb{E}_{\mathbb{Q}}[BS(S_t, K, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_t^2) \mid \mathcal{F}_t]$.

Given time-t information, the only random quantities in call price formula (3.16) are the average variance $\bar{\sigma}_t^2$ under $\mathbb{Q}$ and an Itô integral, which both depend on the $\mathbb{Q}$-Brownian motion $W_x^{\mathbb{Q}}$ only. So rather than simulating the full correlated SDE system (2.15) and (2.17) to obtain the simulated risk-neutral call payoff distribution necessary for evaluating the risk-neutral valuation formula (3.6), there is a much quicker way to obtain the call price. As formula (3.16) shows, simulated paths of $W_x^{\mathbb{Q}}$ are needed only essentially. As such, implementing (3.16) means a big advantage in computing time over naïve implementation of the risk-neutral valuation formula (3.6).

Put option valuation
A European put option with strike K and maturity T can be valued using the put - call parity $P_t = C_t + K\exp(-\bar{r}_t\tau_t) - \exp(-\bar{q}_t\tau_t)S_t = C_t + \exp(-\bar{r}_t\tau_t)[K - F_{t,T}]$, in which $F_{t,T} = S_t\exp[(\bar{r}_t - \bar{q}_t)\tau_t]$ is the forward price of the stock for delivery at time T. These expressions are the same as in case of no leverage effect.

3.3 Hedging

This section considers the hedging of derivatives in the 1-factor affine SV model with leverage effect. Hedging in the model without leverage was discussed in section 4.3 of chapter II. How, if at all, does the analysis change once the leverage effect is taken into account?

Consider hedging a general, path-independent European-style derivative F written on the stock S that pays off an amount $F_{T_F} = f(S_{T_F})$ at its maturity T_F. From (3.5), the arbitrage-free price $F_t = F(t, S_t, x_t)$ evolves according to

$$dF_t = \mu_{F,t}F_t dt + \Delta_{F,t}\sigma_t S_t d\bar{W}_{S,t} + \mathcal{V}_{F,t}\sigma\lambda_t d\bar{W}_{x,t}, \qquad (\mathbb{P}) \qquad (3.20)$$

with $\mu_{F,t} \equiv r_t + (S_t / F_t)\Delta_{F,t}\sigma_t\gamma_{S,t} + (1/F_t)\mathcal{V}_{F,t}\sigma\lambda_t\gamma_{x,t}$, and with $\Delta_{F,t} = \partial F_t / \partial S_t$ the delta and $\mathcal{V}_{F,t} = \partial F_t / \partial x_t = \partial F_t / \partial \sigma_t^2$ the vega of derivative F.[9] Consider next another, longer-maturity derivative G that pays off $G_{T_G} = g(S_{T_G})$ at its maturity $T_G > T_F$. Derivative G serves as a hedging instrument for F which hedges price changes due to unobserved volatility fluctuations. Its price $G_t = G(t, S_t, x_t)$ obeys

$$dG_t = \mu_{G,t}G_t dt + \Delta_{G,t}\sigma_t S_t d\bar{W}_{S,t} + \mathcal{V}_{G,t}\sigma\lambda_t d\bar{W}_{x,t}, \qquad (\mathbb{P}) \qquad (3.21)$$

in which $\mu_{G,t} \equiv r_t + (S_t / G_t)\Delta_{G,t}\sigma_t\gamma_{S,t} + (1/G_t)\mathcal{V}_{G,t}\sigma\lambda_t\gamma_{x,t}$.

A self-financing hedging strategy

We aim at deriving a dynamic, self-financing replicating hedging strategy for derivative F. If we start hedging F at time 0 for maturity T_F, the hedging strategy is given by the instrument holdings $(\psi_t, \phi_t^r, \xi_t) \in \mathcal{F}_t$ for $t \in [0, T_F]$. Here, ψ_t governs the bond holding, ϕ_t^r is the number of reinvestment portfolios and ξ_t is the number of $G-$derivatives to hold in the hedge portfolio at time t. The time-t value of the hedge portfolio, V_t, equals

$$V_t = \psi_t B_t + \phi_t^r S_t^r + \xi_t G_t. \qquad (3.22)$$

As we want the strategy to be replicating, we impose $V_{T_F} = F_{T_F}$. A self-financing strategy requires

$$dV_t = \psi_t dB_t + \phi_t^r dS_t^r + \xi_t dG_t. \qquad (3.23)$$

By no arbitrage, it then must hold that $V_t = F_t \ \forall t$, and hence $dV_t = dF_t$. The instrument holdings ψ_t, ϕ_t^r and ξ_t next follow from equating the coefficients of the dt, $d\bar{W}_{x,t}$ and $d\bar{W}_{S,t}$ terms in the equality $dV_t = dF_t$, which is given by

$$\begin{aligned} dV_t &= \psi_t dB_t + \phi_t^r dS_t^r + \xi_t dG_t \\ &= \left[\psi_t r_t B_t + \phi_t^r(\mu_t + q_t)S_t^r + \xi_t\mu_{G,t}G_t\right]dt + \left[\phi_t^r\sigma_t S_t^r + \xi_t\Delta_{G,t}\sigma_t S_t\right]d\bar{W}_{S,t} + \left[\xi_t\mathcal{V}_{G,t}\sigma\lambda_t\right]d\bar{W}_{x,t} \\ &\equiv \mu_{F,t}F_t dt + \Delta_{F,t}\sigma_t S_t d\bar{W}_{S,t} + \mathcal{V}_{F,t}\sigma\lambda_t d\bar{W}_{x,t}. \end{aligned} \qquad (3.24)$$

The number of G-derivatives to hold in the hedge portfolio equals the ratio of vegas, $\xi_t = \mathcal{V}_{F,t} / \mathcal{V}_{G,t}$. The number of reinvestment portfolios should equal

$$\phi_t^r = \frac{S_t}{S_t^r}\left[\Delta_{F,t} - \xi_t\Delta_{G,t}\right] = \exp(-\bar{q}_{0t}t)\left[\Delta_{F,t} - \frac{\mathcal{V}_{F,t}}{\mathcal{V}_{G,t}}\Delta_{G,t}\right], \qquad (3.25)$$

as $S_t^r = \exp(\bar{q}_{0t}t)S_t$. As $\phi_t^r S_t^r = [\Delta_{F,t} - (\mathcal{V}_{F,t} / \mathcal{V}_{G,t})\Delta_{G,t}]S_t$, a self-financing strategy of ϕ_t^r invested in the reinvestment portfolio corresponds to a strategy of $\phi_t \equiv \Delta_{F,t} - (\mathcal{V}_{F,t} / \mathcal{V}_{G,t})\Delta_{G,t} = \exp(\bar{q}_{0t}t)\phi_t^r$ invested in the stock, because $\phi_t^r S_t^r = \phi_t S_t$. The bond holding ψ_t is most easily computed from realizing that, by no arbitrage, $V_t = \psi_t B_t + \phi_t^r S_t^r + \xi_t G_t \equiv F_t$, such that $\psi_t = B_t^{-1}[F_t - \xi_t G_t - \phi_t^r S_t^r]$.

[9] Recall that (our definition of) the vega measures price changes resulting from stock variance changes. The price sensitivity with respect to volatility changes is given by $\partial F_t / \partial \sigma_t = 2\sigma_t\mathcal{V}_{F,t}$.

The appearance of this hedging strategy is exactly the same as for the hedging strategy in the 1-factor affine SV model without leverage effect; see section 4.3.1 of chapter II. The difference lies in the fact that both models generate different derivative prices, and hence deltas and vegas.

3.4 Derivative betas, asset returns and volatility risk premium

This section defines the stock and volatility betas of a general, path-independent European-style derivative F for which $F_t = F(t, S_t, x_t)$, in the 1-factor SV model with leverage effect. We consider asset returns and the volatility risk premium.

Stock and volatility betas
Derivative prices are sensitive to stock price and volatility fluctuations. One way of measuring these price sensitivities is by the delta $\Delta_{F,t} = \partial F_t / \partial S_t$ and vega $\mathcal{V}_{F,t} = \partial F_t / \partial x_t = \partial F_t / \partial \sigma_t^2$ of derivative F. Another way of measuring these sensitivities is by the *stock beta* $\beta_{F,S}$ and *volatility beta* $\beta_{F,x}$ of derivative F, which we collect in the vector $\boldsymbol{\beta}_F$, and define as (remember that $\sigma_t^2 = x_t$):

$$\boldsymbol{\beta}_{F,t} = \begin{bmatrix} \beta_{F,S,t} \\ \beta_{F,x,t} \end{bmatrix} \equiv \left(\text{var}_{\mathbb{P}} \left[\begin{pmatrix} \frac{dS_t}{S_t} \\ dx_t \end{pmatrix} \mid \mathcal{F}_t \right] \right)^{-1} \text{cov}_{\mathbb{P}} \left[\frac{dF_t}{F_t}, \begin{pmatrix} \frac{dS_t}{S_t} \\ dx_t \end{pmatrix} \mid \mathcal{F}_t \right] = \begin{bmatrix} \frac{S_t}{F_t} \frac{\partial F_t}{\partial S_t} \\ \frac{1}{F_t} \frac{\partial F_t}{\partial x_t} \end{bmatrix}. \quad (3.26)$$

A proof is given in appendix VIb. The intuitive interpretation of these betas is best understood if we assume there to be no leverage effect for the moment, such that $\text{cov}_{\mathbb{P}}[\frac{dS_t}{S_t}, dx_t \mid \mathcal{F}_t] = 0$. In that case F's stock and volatility beta become $\beta_{F,S,t} = \text{cov}_{\mathbb{P}}[\frac{dF_t}{F_t}, \frac{dS_t}{S_t} \mid \mathcal{F}_t] / \text{var}_{\mathbb{P}}[\frac{dS_t}{S_t} \mid \mathcal{F}_t]$, $\beta_{F,x,t} = \text{cov}_{\mathbb{P}}[\frac{dF_t}{F_t}, dx_t \mid \mathcal{F}_t] / \text{var}_{\mathbb{P}}[dx_t \mid \mathcal{F}_t]$, which correspond to our beta definitions given in section 5.1 of chapter II. The betas of derivative F then measure the covariance between F's return and the (relative) change in its underlying risk factors, as a fraction of the variance of the (relative) change in the underlying risk factors, being the stock price and volatility. If a leverage effect is present, the interpretation essentially still stands, but a "correction" is needed to account for the correlation between stock returns and volatility shocks.

Asset returns
What returns are investors expected to earn when investing in the assets traded in this market with leverage? The spot bond return $dB_t / B_t = r_t dt$ quantifies the time value of money. A stock investment automatically generates dividend income. From (2.4), the spot return on the reinvestment portfolio equals

$$\mathbb{E}_{\mathbb{P}} \left[\frac{dS_t^r}{S_t^r} \mid \mathcal{F}_t \right] = (\mu_t + q_t) dt = \left[r_t + \sigma_t \gamma_{S,t} \right] dt . \quad (3.27)$$

The market price of stock price risk $\gamma_{S,t}$ determines the risk premium on the stock, $\sigma_t \gamma_{S,t}$. From (3.5), trading in a general derivative F yields an expected return of

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left[\frac{dF_t}{F_t} \mid \mathcal{F}_t \right] &= \left[r_t + \left(\frac{S_t}{F_t} \frac{\partial F_t}{\partial S_t} \right) \sigma_t \gamma_{S,t} + \left(\frac{1}{F_t} \frac{\partial F_t}{\partial x_t} \right) \sigma \lambda_t \gamma_{x,t} \right] dt \quad (3.28) \\ &= \left[r_t + \boldsymbol{\beta}_{F,t}' \begin{pmatrix} \sigma_t \gamma_{S,t} \\ \sigma \lambda_t \gamma_{x,t} \end{pmatrix} \right] dt = \left[r_t + \beta_{F,S,t} \left(\sigma_t \gamma_{S,t} \right) + \beta_{F,x,t} \left(\sigma \lambda_t \gamma_{x,t} \right) \right] dt . \end{aligned}$$

This is the *expected return – beta*, or reward – risk relationship that holds for derivative assets in the 1-factor affine SV model with leverage effect. The expected derivative return consists of three parts. The first part is due to the time value of money. The second and third parts respectively, are compensation for the risk of derivative price changes induced by unpredictable stock price and volatility fluctuations. Compensation for inherent stock price risk equals the product of the stock beta $\beta_{F,S,t}$ and the stock risk premium $\sigma_t \gamma_{S,t}$. Compensation for inherent volatility risk equals the product of the volatility beta $\beta_{F,x,t}$ and (what we define as) the *volatility risk premium*, $\sigma \lambda_t \gamma_{x,t} = \sigma \gamma (\alpha + \beta x_t)$.

Comparing (3.28) to formula (5.5) in chapter II (and taking (5.1)-(5.2) in that chapter into account) reveals that the expressions for the expected derivative return are exactly the same, whether there is a leverage effect or not.

3.5 Volatility risk premium is negative on theoretical grounds

In section 7.3 of chapter II on the multifactor affine SV model, we proposed an economic theory that predicts the volatility risk premium to be negative. The argument given there was based on the existence of a leverage effect (which the multifactor SV model assumes absent) and is therefore theoretically not fully correct. This section shows however, that if the leverage effect is explicitly modeled, the argument still stands. The reason is that the expected return–beta relationship (3.28) for derivatives is the same in both model settings.

Equation (3.28) is one expression for the arbitrage-free derivative return. Following a similar analysis as in section 5.5 of chapter II, it follows from the no-arbitrage pricing relation $M_t F_t = \mathbb{E}_{\mathbb{P}}\left[M_T F_T \mid \mathcal{F}_t\right]$ (see (3.7)), that the derivative return can also be expressed in terms of the SDF as [10]

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = r_t dt - \mathrm{cov}_{\mathbb{P}}\left[\frac{dM_t}{M_t}, \frac{dF_t}{F_t} \mid \mathcal{F}_t\right]. \tag{3.29}$$

Similar as in chapter II, section 7, we may introduce investor preferences and investor behavior. The equilibrium behavior of a risk-averse, expected-utility maximizing investor, who has to decide how much to consume, save and invest, leads again to the derivative price expression [11]

$$F_t = \mathbb{E}_{\mathbb{P}}\left[e^{-\rho(T-t)} \frac{u'(c_T)}{u'(c_t)} F_T \mid \mathcal{F}_t\right], \tag{3.30}$$

which defines the stochastic discount factor as the marginal rate of intertemporal consumption substitution.

Combining the absence of arbitrage with optimal investor behavior shows that the SDF process equals $M_t = e^{-\rho t} u'(c_t)$, from which the consumption - derivative return formula follows

---

[10] Perhaps stressed superfluously, although this result does not depend on the leverage effect, the SDF processes $\{M_t; t \geq 0\}$ differ in both model settings.

[11] Do not confuse our earlier-used notation ρ for the investor's time-preference rate with the (common) notation ρ for the correlation between stock returns and volatility shocks, as used in this chapter. Except in (3.30) and in $M_t = e^{-\rho t} u'(c_t)$, ρ stands for this correlation in this chapter.

$$\mathbb{E}_{\mathbb{P}}\left[\frac{dF_t}{F_t} \mid \mathcal{F}_t\right] = r_t dt - c_t \frac{u''(c_t)}{u'(c_t)} \operatorname{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, \frac{dc_t}{c_t} \mid \mathcal{F}_t\right]. \tag{3.31}$$

Recall that $-c_t u''(c_t)/u'(c_t) > 0$ is the investor's coefficient of relative risk aversion (CRRA). The CRRA measures the investor's aversion to having different consumption opportunities in different states of the world. Combining (3.28) with (3.31) yields

$$-c_t \frac{u''(c_t)}{u'(c_t)} \operatorname{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, \frac{dc_t}{c_t} \mid \mathcal{F}_t\right] = \left[\beta_{F,S,t}\left(\sigma_t \gamma_{S,t}\right) + \beta_{F,x,t}\left(\sigma \lambda_t \gamma_{x,t}\right)\right] dt\,. \tag{3.32}$$

To make a theoretical prediction on the sign of the volatility risk premium, recall that the investor strives at consumption smoothing over time (the permanent-income hypothesis). The investor may therefore be willing to pay more for assets that pay off when his consumption is low, and less for assets that pay off when his consumption is already high. Assets that correlate negatively with his consumption pattern are therefore more valuable to him.

The typical investor has invested most of his wealth in long stock and bond positions (in practice). As such, if markets go down these positions devaluate, limiting his consumption opportunities. Vice versa, if markets go up, these positions appreciate, expanding his consumption opportunities. The investor's consumption is thus directly tied to general market movements, which he cannot influence himself.

To be able to smooth his consumption over time (i.e., to have his consumption be less dependent on markets movements), the investor can invest in a long delta-neutral positive-vega derivative G such as an ATM straddle or variance swap. As $\Delta_{G,t} = \partial G_t / \partial S_t = 0$ and hence $\beta_{G,S,t} = 0$, it follows from (3.32)

$$-c_t \frac{u''(c_t)}{u'(c_t)} \operatorname{cov}_{\mathbb{P}}\left[\frac{dG_t}{G_t}, \frac{dc_t}{c_t} \mid \mathcal{F}_t\right] = \beta_{G,x,t}\left(\sigma \lambda_t \gamma_{x,t}\right) dt\,. \tag{3.33}$$

Suppose then that the investor holds such a derivative G in his portfolio, and suppose now that markets are dropping. His long stock and bond positions still devaluate. However, due to the leverage effect, dropping markets are generally accompanied by increasing volatility, increasing the value of derivative G. As such, this increase (partly) offsetts the drop in value in the investor's stock and bond portfolio (and hence his consumption opportunities). A delta-neutral positive-vega derivative G thus indeed helps the investor pursuing a smooth consumption pattern over time. The correlation between derivative G's return and the investor's consumption growth is therefore likely to be negative. Now, as

$$\underbrace{-c_t \frac{u''(c_t)}{u'(c_t)}}_{>0} \underbrace{\operatorname{cov}_{\mathbb{P}}\left[\frac{dG_t}{G_t}, \frac{dc_t}{c_t} \mid \mathcal{F}_t\right]}_{<0} = \underbrace{\beta_{G,x,t}}_{>0} \underbrace{\left(\sigma \lambda_t \gamma_{x,t}\right)}_{\Rightarrow <0} dt\,, \tag{3.34}$$

it follows that our theory predicts the market price of volatility risk (which equals $\gamma_{x,t} = \gamma \lambda_t = \gamma \sqrt{\alpha + \beta x_t}$) to be negative (and thus $\gamma < 0$), leading to a *negative* volatility risk premium of $\sigma \lambda_t \gamma_{x,t} = \sigma \gamma (\alpha + \beta x_t)$.

Our own empirical evidence and the evidence on negative market volatility risk compensation discussed in section 6.3 of chapter II underwrites this theory.

(Note: A similar clear-cut argument based on e.g. a long call option (which also rises in value if volatility rises) instead of a *delta-neutral* positive-vega derivative seems impossible to make: Reconsidering (3.32), dropping markets *may* induce a (temporarily) negative stock risk premium $\sigma_t \gamma_{S,t}$. Moreover, the value of a call option decreases if the underlying declines in value. Hence, basing the argument on a non-delta-neutral derivative, we cannot distinctively disentangle whether the volatility risk premium should be positive or negative theoretically.)

4. State space estimation approach

This section derives a linear state space model for the 1-factor affine SV model with leverage effect, in a similar way as we previously did for the multifactor no-leverage SV model. Section 4.1 explains how to extract information from stock returns and RV, section 4.2 from option prices. We arrive at a somewhat different options measurement equation. However, as explained in section 4.3, Monte Carlo evidence directs us to eventually use the same conditional state space model as used in the previous chapter (which is based on no leverage). To estimate correlation ρ, we propose an alternative estimator based on (2.6).

4.1 Extracting information from stock prices

Consider first affine SV specification (2.5) with decomposition (2.1). As such, the usual exact discrete-time equivalent of the SDE for x applies,

$$x_{t+\Delta t}^* = \exp(-k\Delta t) x_t^* + u_{t+\Delta t}\,, \qquad (\mathbb{P}) \qquad (4.1)$$

with $x_t^* = x_t - \theta$, and from appendix B, section 6,

$$u_{t+\Delta t} \sim \left(0, \sigma^2(\alpha + \beta\theta)\frac{1-\exp(-2k\Delta t)}{2k}\right) \qquad (\mathbb{P}) \qquad (4.2)$$

$$u_{t+\Delta t} \mid \mathcal{F}_t \sim \left(0, \sigma^2(\alpha + \beta\theta)\frac{1-\exp(-2k\Delta t)}{2k} + \beta\sigma^2 x_t^* \frac{\exp(-k\Delta t) - \exp(-2k\Delta t)}{k}\right).$$

The error series $\{u_t\}$ is unconditionally white noise, though conditionally heteroskedastic, through level-dependent volatility-of-volatility.

Extracting information from squared returns and RV

How to extract information from stock prices in the model with leverage? A derivation entirely analogous as in section 3.2 of chapter III based on an Euler discretization of the stock price SDE, shows that, despite the presence of the leverage effect, the squared returns still satisfy the equation

$$\frac{1}{\Delta t}(r_t - \mu_{t-\Delta t}\Delta t)^2 = x_{t-\Delta t} + \omega_t = \theta + x_{t-\Delta t}^* + \omega_t\,, \quad \omega_t \sim (0, \sigma_\omega^2)\,, \qquad (4.3)$$

with $\{\omega_t\}$ a white noise series. There is one difference. In the multifactor SV model without leverage, $\{\omega_t\}$ is uncorrelated with the volatility-factor innovations

$\{u_t\}$ at all points in time. Here, although the series $\{\omega_t\}$ and $\{u_t\}$ do not correlate at different points in time, they are contemporaneously correlated. [12]

If high-frequency intraday data is available, one may opt for information extraction from RV data. In the multifactor SV model we derived the measurement equation for RV from an equation involving the average stock variance (see section 3.3.3 of chapter III). The stock price dynamics did not play a role in the derivation. Given (2.5) and (2.1), exactly the same measurement equation for RV holds in the model with leverage effect,

$$\sigma_{RV,t}^2 = \theta + \frac{1-\exp(-k\Delta t)}{k\Delta t} x_{t-\Delta t}^* + \omega_{RV,t}, \quad \omega_{RV,t} \sim (0, \sigma_{\omega,RV}^2), \tag{4.4}$$

in which the white noise error series $\{\omega_{RV,t}\}$ is contemporaneously correlated with the volatility factor shocks $\{u_t\}$, but is uncorrelated at all other time points.

4.2 Extracting information from option prices

How can we extract information from call option prices in the 1-factor affine SV model with leverage effect? In terms of the Black-Scholes pricing function, the call price formula is given by (3.18) as

$$C_t = \mathbb{E}_{\mathbb{Q}}\left[BS(S_{t,\rho}, \bar{\sigma}_{t,\rho}^2) \mid \mathcal{F}_t\right] \tag{4.5}$$

with

$$S_{t,\rho} = S_t \exp\left(-\tfrac{1}{2}\rho^2\bar{\sigma}_t^2\tau_t + \rho\int_t^T \sigma_u dW_{x,u}^{\mathbb{Q}}\right) \tag{4.6}$$

$$\begin{aligned}\bar{\sigma}_{t,\rho}^2 &= (1-\rho^2)\bar{\sigma}_t^2 \qquad (4.7)\\ &= (1-\rho^2)\frac{1}{\tau_t}\ln\left[Y_t\right] = \bar{\sigma}_{t,\rho}^2(Y_t), \quad Y_t \equiv \exp\left(\tau_t\bar{\sigma}_t^2\right).\end{aligned}$$

For simplicity we have shortened $BS(S_{t,\rho}, K, \tau_t, \bar{r}_t, \bar{q}_t, \bar{\sigma}_{t,\rho}^2)$ to $BS(S_{t,\rho}, \bar{\sigma}_{t,\rho}^2)$, because given time-$t$ information, the only quantities that are random in the BS pricing function are $S_{t,\rho}$ and $\bar{\sigma}_{t,\rho}^2$.

We pursue a similar approximation method as for the multifactor affine SV model using the result $\mathbb{E}_{\mathbb{Q}}[Y_t \mid \mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}}[\exp(\tau_t\bar{\sigma}_t^2) \mid \mathcal{F}_t] = \exp[A_1(\tau_t) + B_1(\tau_t)x_t]$, which still holds in case of leverage. Due to leverage however, the BS pricing function is now a function of two random variables instead of one, $S_{t,\rho}$ and $\bar{\sigma}_{t,\rho}^2(Y_t)$. These random variables are correlated, though not perfectly. (In the multifactor model, the BS pricing function was a function of the average variance $\bar{\sigma}_t^2 = \bar{\sigma}_t^2(Y_t)$ only; remember the implied call price $C_t = \mathbb{E}_{\mathbb{Q}}\left[BS(\bar{\sigma}_t^2) \mid \mathcal{F}_t\right]$ for that model.)

[12] This is seen as follows. From an Euler discretization of the stock price SDE (2.3), it follows that $\omega_{t+\Delta t} = \sigma_t^2[(\Delta\bar{W}_{S,t})^2 / \Delta t - 1]$, in which $\Delta\bar{W}_{S,t} = \bar{W}_{S,t+\Delta t} - \bar{W}_{S,t} = \sqrt{1-\rho^2}\Delta W_{S,t} + \rho\Delta W_{x,t}$; see decomposition (2.1). Moreover, $u_{t+\Delta t} = \int_t^{t+\Delta t}\exp[-k(t+\Delta t - s)]\sigma\lambda_s dW_{x,s}$ with (recall) $\lambda_s = \sqrt{\alpha + \beta x_s}$, such that the innovation $u_{t+\Delta t}$ involves the full path of $\{W_{x,t}\}$ over the interval $[t, t+\Delta t]$. Evidently, $\omega_{t+\Delta t}$ and $u_{t+\Delta t}$ are correlated. The non-contemporaneous errors are uncorrelated, due to independent Brownian increments.

Linearizing the call price formula
We aim at linearizing around suitable choices for $S_{t,\rho}$ and Y_t such that we end up with a linear options measurement equation. We denote this linearization point by $(S_{t,\rho}^*, Y_t^*)$. From (4.7), a choice for Y_t^* implies a choice for $\bar{\sigma}_{t,\rho}^2$ and vice versa. We are looking for a $\mathcal{F}_t$ – measurable linearization point, i.e. a non-random point given $\mathcal{F}_t$-information. Linearizing the $BS(.,.)$ function around such a point $(S_{t,\rho}^*, Y_t^*)$ yields the rewritten call price formula

$$\begin{aligned} C_t &= \mathbb{E}_{\mathbb{Q}}\left[BS\left(S_{t,\rho}, \bar{\sigma}_{t,\rho}^2(Y_t)\right) \mid \mathcal{F}_t\right] \\ &= \mathbb{E}_{\mathbb{Q}}\left[BS\left(S_{t,\rho}^*, \sigma_{t,\rho}^2(Y_t^*)\right) + \left(S_{t,\rho} - S_{t,\rho}^*\right)\frac{\partial BS}{\partial S_{t,\rho}} + \left(Y_t - Y_t^*\right)\frac{\partial BS}{\partial Y_t} + HOT_t \middle| \mathcal{F}_t\right] \\ &= BS\left(S_{t,\rho}^*, \sigma_{t,\rho}^2(Y_t^*)\right) + \left(\mathbb{E}_{\mathbb{Q}}\left[S_{t,\rho} \mid \mathcal{F}_t\right] - S_{t,\rho}^*\right)\frac{\partial BS}{\partial S_{t,\rho}} + \left(\mathbb{E}_{\mathbb{Q}}\left[Y_t \mid \mathcal{F}_t\right] - Y_t^*\right)\frac{\partial BS}{\partial Y_t} + \mathbb{E}_{\mathbb{Q}}\left[HOT_t \mid \mathcal{F}_t\right] \end{aligned} \tag{4.8}$$

in which the $BS(.,.)$ partial derivatives are evaluated in $\left(S_{t,\rho}^*, \bar{\sigma}_{t,\rho}^2(Y_t^*)\right)$ (and hence are $\mathcal{F}_t$ – measurable, from which the third equality follows), and with HOT_t denoting higher-order terms.

To motivate the choice for the linearization point, consider the first term in (4.8). The model-implied Black-Scholes implied variance $\sigma_{implied,t}^2$ is defined implicitly as $C_t = BS(S_t, \sigma_{implied,t}^2)$. This suggests the following convenient linearization point:

$$\left(S_{t,\rho}^*, \sigma_{t,\rho}^2(Y_t^*)\right) = \left(S_t, \sigma_{implied,t}^2\right) \Rightarrow Y_t^* = \exp\left[\frac{\tau_t \sigma_{implied,t}^2}{1-\rho^2}\right], \tag{4.9}$$

in which the implied Y_t^* follows from (4.7). Reconsidering (4.8), the conditional expectation of Y_t under $\mathbb{Q}$ equals $\mathbb{E}_{\mathbb{Q}}[Y_t \mid \mathcal{F}_t] = \exp[A_1(\tau_t) + B_1(\tau_t)x_t]$, which is exponential-affine in the latent factor.

But what about the conditional expectation $\mathbb{E}_{\mathbb{Q}}[S_{t,\rho} \mid \mathcal{F}_t]$ that also appears in (4.8)? $S_{t,\rho}$ can be written as (see (4.6) and (3.10))

$$S_{t,\rho} = S_t \exp(U_T - U_t), \tag{4.10}$$

in which the process $\{U_t; t \geq 0\}$ is defined by

$$U_t \equiv -\tfrac{1}{2}\rho^2 \int_0^t \sigma_u^2 du + \rho \int_0^t \sigma_u dW_{x,u}^{\mathbb{Q}} \Leftrightarrow dU_t = -\tfrac{1}{2}\rho^2 \sigma_t^2 dt + \rho\sigma_t dW_{x,t}^{\mathbb{Q}}. \tag{4.11}$$

Consider next the process $\{V_t; t \geq 0\}$ with $V_t \equiv \exp(U_t)$. Itô's lemma shows that $dV_t = \rho\sigma_t V_t dW_{x,t}^{\mathbb{Q}}$, i.e. $\{V_t; t \geq 0\}$ is a driftless process. As the technical condition $\mathbb{E}_{\mathbb{Q}}[\exp(\frac{1}{2}\int_0^T \rho^2 \sigma_s^2 ds)] < \infty$ holds (see appendix B, section 10), $\{V_t; t \geq 0\}$ is a $(\mathbb{Q}, \{\mathcal{F}_t\})$ – martingale. As such, $\mathbb{E}_{\mathbb{Q}}[V_T \mid \mathcal{F}_t] = V_t$ or $\mathbb{E}_{\mathbb{Q}}[V_T / V_t \mid \mathcal{F}_t] = 1$, and hence $\mathbb{E}_{\mathbb{Q}}[\exp(U_T - U_t) \mid \mathcal{F}_t] = 1$. This implies

$$\mathbb{E}_{\mathbb{Q}}[S_{t,\rho} \mid \mathcal{F}_t] = S_t. \tag{4.12}$$

Given linearization point (4.9), the first term in (4.8) thus equals C_t, the second term equals zero, and hence (4.8) can be rewritten as

$$\ln\left(Y_t^* - \frac{\mathbb{E}_{\mathbb{Q}}[HOT_t \mid \mathcal{F}_t]}{\partial BS / \partial Y_t}\right) = \ln\left(\mathbb{E}_{\mathbb{Q}}\left[Y_t \mid \mathcal{F}_t\right]\right) = A_1(\tau_t) + B_1(\tau_t)x_t\,. \qquad (4.13)$$

This equation is linear in the latent volatility factor x.

A linear options measurement equation
Similar as before, neglecting higher-order terms, rewriting using expression (4.9) for Y_t^*, introducing a constant μ_v and noise in the form of an additive error term ε_t, yields

$$\sigma^2_{implied,t} = \mu_v + \frac{(1-\rho^2)A_1(\tau_t)}{\tau_t} + \frac{(1-\rho^2)B_1(\tau_t)}{\tau_t}x_t + \varepsilon_t,\ \ \varepsilon_t \sim (0,\sigma^2_\varepsilon)\,. \qquad (4.14)$$

This equation may serve as the input to a linear state space model with time-varying coefficients. [13] (If $\rho = 0$, the measurement equation obtained from the 1-factor affine SV model without leverage effect results. Note that moneyness does not play a role in (4.14); see section 3.4.2 of chapter III for discussion.) For volatility specification (2.5), the functions $A_1(.)$ and $B_1(.)$ satisfy the following system of Ricatti ODEs, with $\tilde{k} = k + \sigma\gamma\beta$ and $\tilde{\theta} = (k\theta - \sigma\gamma\alpha)/\tilde{k}$ (see (2.18)):

$$\frac{dA_1(\tau)}{d\tau} = \tilde{\theta}\,\tilde{k}\,B_1(\tau) + \tfrac{1}{2}\alpha\sigma^2\left[B_1(\tau)\right]^2, \qquad A_1(0) = 0\,, \qquad (4.15)$$
$$\frac{dB_1(\tau)}{d\tau} = -\tilde{k}\,B_1(\tau) + \tfrac{1}{2}\beta\sigma^2\left[B_1(\tau)\right]^2 + 1\,, \qquad B_1(0) = 0\,.$$

Correlation parameter ρ seems identifiable from (4.14): Factor $(1-\rho^2)$ cannot be subsumed in just one of the parameters, not even for Heston (CIR) SV, for which $\alpha = 0$ and $\beta = 1$. Namely, defining the functions $A_{1,\rho}(.) \equiv (1-\rho^2)A_1(.)$ and $B_{1,\rho}(.) \equiv (1-\rho^2)B_1(.)$, these functions satisfy $dA_{1,\rho}(\tau)/d\tau \equiv (1-\rho^2)dA_1(\tau)/d\tau$, $dB_{1,\rho}(\tau)/d\tau \equiv (1-\rho^2)dB_1(\tau)/d\tau$, and (taking (4.15) into account) cannot be rewritten in terms of simple adjusted parameters. Nor can $(1-\rho^2)$ be considered a simple scaling factor acting on latent factor x. [14]

4.3 Conditional state space model

Our specific choice (4.9) for the linearization point results in the linear options measurement equation (4.14). This is convenient, as this equation can be cast in a (relatively simple) linear state space representation, together with equations (4.1), (4.3) and (4.4) for the discrete-time volatility-factor evolution, the squared returns and RV. As a Monte Carlo study shows however (for the five parameter sets given in table 5.2), correlation parameter ρ is routinely estimated at zero. We attribute this to linearization point (4.9). [15] (Restricting $\rho = 0$ in succession

---

[13] A parabolic approximation of the $BS(.,.)$ pricing function does not lead to an equation that is linear in x_t. Such an approximation includes Y_t^2, $S^2_{t,\rho}$ and $Y_tS_{t,\rho}$, and requires expressions for the conditional $\mathbb{Q}$-expectations of these random variables, given $\mathcal{F}_t$. From appendix B, section 10 we have $\mathbb{E}_{\mathbb{Q}}[Y_t^2 \mid \mathcal{F}_t] = \exp[A_2(\tau_t) + B_2(\tau_t)x_t]$. "Nice" analytical expressions for $\mathbb{E}_{\mathbb{Q}}[S^2_{t,\rho} \mid \mathcal{F}_t]$ and $\mathbb{E}_{\mathbb{Q}}[Y_tS_{t,\rho} \mid \mathcal{F}_t]$ in terms of x_t seem impossible to derive however.

[14] If functions $A_1(.)$ and $B_1(.)$ were constant such that $\sigma^2_{implied,t} = \mu_v + (1-\rho^2)[A_1 + B_1x_t] + \varepsilon_t$ then ρ, A_1 and B_1 would not separately be identified. However, they do vary with maturity τ_t.

[15] A simple, clear-cut Jensen's inequality-based argument in which we take call price formula (4.8) into account seems difficult (if not impossible) to make for this finding of ρ being estimated at 0: Although the $BS(.,.)$-pricing function is concave in its variance argument, matters here are more complicated than in a no-leverage-effects model, as the $BS(.,.)$ function is now a function of *two* random variables (given $\mathcal{F}_t$), with not only $\bar{\sigma}^2_{t,\rho}$ but also $S_{t,\rho}$ depending on $\bar{\sigma}^2_t$. As such, we rather take this finding for granted, than give an intuitive but possibly incorrect reason for this finding.

leads to the same optimal loglikelihood value and estimates for θ, k, σ, and γ as when leaving ρ free.)

Restricting $\rho = 0$ and a different estimator for ρ
We adopt the following strategy to overcome this hurdle and still obtain an estimate of ρ. First, we restrict $\rho = 0$ prior to estimation. Hence, this essentially boils down to assuming the leverage effect to be absent. The state space model derived from the 1-factor affine SV model with leverage then reduces to the conditional state space model derived from the multifactor no-leverage-effects affine SV model (the 1-factor special case). The conditional state space model is given in section 3.3 of chapter V and is estimated by Extended Kalman filter QML. Second, given the smoothed volatility series extracted from the data at the optimum, we next obtain an estimate for ρ from (2.6). That is, we compute ρ as the correlation between daily stock returns and daily smoothed stock variance changes. This ρ-estimator is particularly easy to compute.

Indeed, this is essentially the strategy pursued for the multifactor affine SV model without leverage effect in previous chapters.

A 2-step estimation method
The bias in this $\rho-$estimator appears modest when option data is included for estimation (see sections 5.3.3-5.3.5). However, restricting ρ to zero whereas its true value differs from zero may result in bias in the other estimates. The following 2-step estimation method seems intuitively appealing to (partly) overcome this source of bias. In the first step, we estimate the conditional state space model by Extended Kalman filter QML with the restriction $\rho = 0$ imposed. At the optimum, we compute an estimate of ρ using (2.6). (The first step thus corresponds to the estimation procedure just proposed.) In the second step, we plug this estimate of ρ into the options measurement equation (4.14), and perform a second, restricted estimation round to determine the estimates of θ, k, σ and γ (while keeping ρ restricted).

Although appealing at first sight, this 2-step method performs much worse than simply taking the estimates obtained in the first step as the final estimates. The Monte Carlo evidence in section 5.3.6 illustrates this. As such, the device is to restrict $\rho = 0$, estimate the conditional state space model to obtain estimates for θ, k, σ and γ, and then compute an estimate for ρ based on (2.6).

5. A Monte Carlo study towards the Heston model

This section reports the results of a Monte Carlo study towards the Heston (1993) model. A first aim is to investigate each parameter's impact on the stock return and volatility distributions, and on option prices. A second aim is to examine the performance of the state space estimation method as outlined in section 4.3 (with $\rho = 0$ imposed), and to illustrate that the 2-step method performs worse.

Section 5.1 details assumptions with regard to the simulated data and discusses the set up of the Monte Carlo experiment. Section 5.2 examines each parameter's impact on stock returns and option prices. Section 5.3 investigates the performance of the estimation method based on (a combination of) squared return, RV and option data.

5.1 Simulating from the Heston model

The Heston model is obtained from taking $\alpha = 0$ and $\beta = 1$ in the model with leverage effect discussed in section 2. Under $\mathbb{P}$, the Heston volatility specification reads $\sigma_t^2 = x_t$ with $dx_t = k(\theta - x_t)dt + \sigma\sqrt{x_t}d\bar{W}_{x,t}$. Under $\mathbb{Q}$, the factor SDE reads $dx_t = \tilde{k}(\tilde{\theta} - x_t)dt + \sigma\sqrt{x_t}d\bar{W}_{x,t}^{\mathbb{Q}}$, in which the risk-neutral parameters are given by $\tilde{k} = k + \sigma\gamma$ and $\tilde{\theta} = k\theta / \tilde{k}$. The price of volatility risk is given by $\gamma_{x,t} = \gamma\sqrt{x_t}$. The volatility risk premium equals $\sigma\gamma x_t$, and averages at $\sigma\gamma\theta$. If there is no leverage effect, the Heston model reduces to the 1-factor CIR SV model.

5.1.1 Assumptions

We assume 260 trading days per annum and denote the timing of the daily data points by $t = \Delta t, 2\Delta t, .., T\Delta t$, with $\Delta t = 1/260$. We simulate a total of $T = 1058$ daily observations, corresponding to the UK data used in our empirical work. As in our previous Monte Carlo studies, we assume $S_0 = 5300$, $\mu_t = \mu = 8.25\%$, $q_t = q = 3.5\%$, and $r_t = r = 6\%$, which correspond to the information in the FTSE100-index data for the period 6 Oct 1997 – 28 Dec 2001. We simulate time series of squared returns, 10-minute realized volatilities (i.e. 48 intraday time points per trading day of eight hours) and short-maturity (close-to) ATMF call options. The maturity and moneyness of the simulated option series exactly match those of the empirical SM ATMF FTSE100-index option series, for each t. The average maturity is 1.4 months. (See section 2.1 of chapter IV for details.)

Table 5.1: Heston (1993) model estimation results in the literature

	Bakshi, Cao & Chen (1997)	*Chernov & Ghysels (2000)*	*Pan (2002)*	*Jones (2003)*	*Van der Ploeg (2004)*
θ	-	0.0154	0.0137	0.0351	0.0441
k	-	0.931	7.1	0.478	7.86
σ	0.40	0.0615	0.32	0.377	0.689
γ *(implied)*	-	-3.92	-24	-9.17	-6.91
ρ	-0.70	-0.018	-0.53	-0.679	-0.67
Volatility of returns	-	12.4%	11.7%	18.7%	21.0%
Persistence	-	0.9964	0.9731	0.9982	0.970
Std.dev. x_t	-	0.0060	0.0099	0.0722	0.0365
$\tilde{k}$	0.99	0.690	-0.50	-2.98	3.10
$\tilde{\theta}$	0.0404	0.00956	-0.20	-0.0563	0.112
Volatility risk prem. *(implied)*	-	-0.37%	-10.4%	-27%	-21%
Data	Daily panel of S&P500 call options	Daily S&P500 returns + 1 SM-ATM call series	Weekly S&P500 returns + 1 SM-ATM call series	Daily S&P100 returns + impl.vols VIX	Daily FTSE100 returns + 1 SM-ATM call series
Period	1988 - 1991	1986 - 1993	1989 - 1996	1988 - 2000	1997 - 2001
Estimation method	Daily calibration (averaged estimates)	Efficient Method of Moments	Implied-state GMM	Bayesian approach	Extended Kalman filter QML

The table reports recently estimated Heston model values and calculated implied results. Return volatility is computed as $\sqrt{\theta}$, persistence as $\exp(-k\Delta t)$, std.dev. of x_t as $\sqrt{\sigma^2\theta / 2k}$, implied γ as $\gamma = (\tilde{k} - k)/\sigma$ and mean volatility risk premium as $\sigma\gamma\theta$.

A decision needs to be made on the parameter values with which to simulate data from the Heston model. Table 5.1 summarizes Heston model values researchers have recently estimated and reports implied results, including our own results from the previous chapter. The specific data, period and estimation method are also tabulated. Sections 6.3.1 and 2.1.1 of chapter II and III respectively, contain a more elaborate discussion of the papers by Chernov and Ghysels (2000), Pan (2002) and Jones (2003), and comment on their findings. Bakshi, Cao and Chen (1997) calibrate the Heston model to a panel of option prices, for which only the model under $\mathbb{Q}$ is needed. We therefore cannot infer the implied θ, k and γ.

The estimates reported in table 5.1 may guide us in what parameter values to choose. There are some issues regarding these estimates that deserve attention however. Although Chernov and Ghysels (2000) and Pan (2002) use a largely overlapping dataset, their results differ enormously on ρ, k and the implied volatility risk premium. We feel more confident with Pan's results. Moreover, both Pan (2002) and Jones (2003) find an explosive risk-neutral volatility process (as $\tilde{k}, \tilde{\theta} < 0$). We do not advocate simulating from such an explosive (and probably ill or improperly-defined) process. This restricts possible parameter choices, mainly with regard to γ.

Table 5.2: Parameter sets

Parameter set:	*I*	*II*	*III*	*IV*	*V*
θ	0.04	0.04	0.04	0.04	0.02
k	1.50	2.61	5.25	7	7
σ	0.20	0.30	0.45	0.60	0.30
γ	-4	-5	-6	-7	-20
ρ	-0.40	-0.50	-0.55	-0.60	-0.50
Volatility of returns	20%	20%	20%	20%	14.14%
Persistence	0.9942	0.9900	0.9800	0.9734	0.9734
Std.dev. x_t	0.0231	0.0263	0.0278	0.0321	0.0113
$\tilde{k}$	0.700	1.11	2.55	2.80	1.00
$\tilde{\theta}$	0.0857	0.0941	0.0824	0.100	0.140
Av.vol. risk premium	-3.2%	-6%	-10.8%	-16.8%	-12%

The table reports five parameter sets with implied quantities which will be used for simulating data from the Heston model (i.e. the 1-factor CIR SV model with leverage).

Eventually we have chosen for five parameter sets, numbered I – V, given in table 5.2. Set IV closely corresponds to our own empirical estimates. Set V largely corresponds to Pan's findings, with the exception that a negative $\tilde{k}$ and $\tilde{\theta}$ are circumvented, while keeping the average volatility risk premium similar. Note the same choice for θ (unconditional stock variance) for parameter sets I - IV. We intuitively expect the bias of its estimate to be not much dependent on its magnitude. This allows us to focus (when moving from set I to IV) on the effects on bias of a gradual increase in mean-reversion (k), accompanied by increasing volatility-of-volatility (σ), and a stronger leverage effect (ρ), which is often observed in practice (see e.g., Jones (2003)).

Parameter sets I–V were also used in the Monte Carlo study towards the 1-factor CIR SV model (see section 4 of chapter V), except that there $\rho = 0$ (no leverage effect). This allows for a bias comparison induced by the leverage effect.

5.1.2 Set up of the Monte Carlo experiment

Below we outline, in chronological order, how to obtain one simulated dataset from the Heston model. For each of the parameter sets I - V we simulate a total of 150 datasets, each consisting of $T = 1058$ daily observations. [16]

Step 1: Simulate a correlated stock price and volatility path under $\mathbb{P}$; compute squared returns and realized volatilities

The first step consists of simulating a sample path of the correlated system of SDEs (2.3) and (2.5) for $\{S_t\}$ and $\{x_t\}$ under $\mathbb{P}$, over the interval $[0, T\Delta t]$, given $S_0 = 5300$ and some value for x_0. We draw x_0 from the stationary distribution of $\{x_t\}$. Rather than using a simple Euler scheme, we use the refined *1D-Milstein approximation scheme* for numerically solving the system. [17] It reads

$$x_{[0]} \sim \frac{\sigma^2}{4k}\,\chi^2\left(\frac{4k\theta}{\sigma^2}\right) \qquad (\mathbb{P}) \qquad (5.1)$$

$$x_{[(i+1)\delta t]} = x_{[i\delta t]} + k(\theta - x_{[i\delta t]})\delta t + \sigma\sqrt{x_{[i\delta t]}}\sqrt{\delta t}\,\epsilon_i + \tfrac{1}{4}\sigma^2\delta t\left[\epsilon_i^2 - 1\right],\ i = 0,..,TI-1,$$

and

$$S_{[0]} = S_0$$

$$S_{[(i+1)\delta t]} = S_{[i\delta t]}\left(1 + \mu\,\delta t + \sqrt{x_{[i\delta t]}}\sqrt{\delta t}\left(\sqrt{1-\rho^2}\,\eta_i + \rho\epsilon_i\right) + \tfrac{1}{2}x_{[i\delta t]}\delta t\left[\left(\sqrt{1-\rho^2}\,\eta_i + \rho\epsilon_i\right)^2 - 1\right]\right),$$

for $i = 0,..,TI-1$, in which $\epsilon_i, \eta_i \sim$ i.i.d. $\mathcal{N}(0,1)$ under $\mathbb{P}$. [18] We simulate $\{S_t\}$ and $\{x_t\}$ on $I = 48$ equidistant intraday time points. Since a trading day lasts for approximately 8 hours, this means that a value is simulated every 10 minutes. The discretization step of the SDEs then equals $\delta t = \Delta t / I = 8.0 * 10^{-5}$. The associated $\mathbb{P}$-volatility path over $[0, T\Delta t]$ is given by $\sigma_{[i\delta t]} = \sqrt{x_{[i\delta t]}}$. For all parameter sets I–V, the Feller condition associated with the CIR process for $\{x_t\}$ holds under both $\mathbb{P}$ and $\mathbb{Q}$. [19]

The annualized squared returns in deviation from their mean are computed as $(r_{[j\Delta t]} - \mu\Delta t)^2 / \Delta t$ for $j = 1,..,T$, with $r_{[j\Delta t]} = (S_{[j\Delta t]} - S_{[(j-1)\Delta t]}) / S_{[(j-1)\Delta t]}$. The simulated 10-minute realized variance over day $[t, t+\Delta t]$ is calculated as

---

[16] Obviously, the more datasets we simulate, the better for drawing conclusions on the state space estimation method. To keep the amount of computations to a reasonable level however, we choose to simulate 150 datasets. This takes 41 hours on a Pentium-III computer (see also our discussion below).

[17] An explicit solution to the SDE system does not exist, such that a numerical solution is called for. A *pathwise* approximation is usually called a *strong* numerical solution, whereas a *weak* numerical solution only aims at approximating the *moments* (i.e., the finite-dimensional distributions) of the solution. The Euler approximation scheme converges strongly with order 0.5 and weakly with order 1. The 1D-Milstein scheme converges strongly (and hence weakly) with order 1. It is an improvement to Euler's scheme.

[18] This 1D-Milstein scheme follows from $\Delta S_t \approx \mu S_t \Delta t + \sigma_t S_t \Delta\bar{W}_{S,t} + \frac{1}{2}\sigma_t^2 S_t[(\Delta\bar{W}_{S,t})^2 - \Delta t]$, and $\Delta x_t \approx k(\theta - x_t)\Delta t + \sigma\sqrt{x_t}\Delta\bar{W}_{x,t} + \frac{1}{4}\sigma^2[(\Delta\bar{W}_{x,t})^2 - \Delta t]$, with all differences being forward, e.g., $\Delta S_t \equiv S_{t+\Delta t} - S_t$. Based on decomposition (2.1), we next perform the decomposition $\Delta\bar{W}_{x,t} = \sqrt{\Delta t}\,\epsilon_t$ and $\Delta\bar{W}_{S,t} = \sqrt{\Delta t}[\sqrt{1-\rho^2}\,\eta_t + \rho\epsilon_t]$. In contrast, the so-called *2D-Milstein scheme* starts from (2.3) and (2.5) with decomposition (2.1) being imposed *before* discretizing the SDEs. Although the 2D-scheme means an improvement to the 1D-scheme, it is computationally more involved: The resulting discretized system involves a certain double stochastic integral, which may be approximated by e.g. the *subdivision method* of Kloeden (2002). See Schmitz (2004) for more details.

[19] Due to the discretization it may nevertheless occur that $\{x_t\}$ turns negative at some point. If this occurs our simulation procedure replaces such a negative value by 10^{-5}. Tracking the number of times this happens, shows that this never occurs however, for none of the simulated datasets.

$$\sigma^2_{RV,[t+\Delta t]} = \frac{1}{\Delta t}\sum_{i=1}^{I}\left(\ln\frac{S_{[t+i\delta t]}}{S_{[t+(i-1)\delta t]}}\right)^2, \ t = 0, \Delta t, .., (T-1)\Delta t\ . \qquad (\mathbb{P}) \qquad (5.2)$$

Step 2: Simulate forward prices and option strike prices

We compute the forward price for delivery on the same date as the expiration date of the option by $F_{[t,T]} = S_{[t]}\exp[(r-q)\tau_t]$, for each $t = \Delta t, .., T\Delta t$. Here, $\{\tau_t; t = \Delta t, .., T\Delta t\}$ is the option maturity series. As before, we simulate the strike prices of the ATMF option series in such a way that the moneyness of the simulated and real-world SM ATMF FTSE100-index option series exactly coincide for each t. (We use the same procedure as in section 3.2.2 of chapter IV).

Step 3: Simulate Heston call prices

Given r, q, τ_t and the simulated $\sigma_{[t]}, S_{[t]}, F_{[t,T]}, K_t$, we next simulate the theoretical Heston call price C_t for days $t = \Delta t, 2\Delta t, .., T\Delta t$. Given time-$t$ information, the only random quantities in the call price formula (3.16) are an Itô integral and the average variance $\bar{\sigma}_t^2$ under $\mathbb{Q}$, i.e.,

$$\bar{\sigma}_t^2 = \frac{1}{\tau_t}\int_t^{T_t}\sigma_u^2 du = \frac{1}{\tau_t}\int_t^{T_t} x_u du, \quad \int_t^{T_t}\sigma_u dW^{\mathbb{Q}}_{x,u} = \int_t^{T_t}\sqrt{x_u}\, dW^{\mathbb{Q}}_{x,u}\ . \qquad (\mathbb{Q}) \qquad (5.3)$$

Suppose we have simulated N independent combinations of the average variance and the Itô integral at time t. Each of these $i = 1, .., N$ combinations yields a simulated Black-Scholes call price value $BS(S_{[t,\rho]}(i), K_t, \tau_t, r, q, \bar{\sigma}^2_{[t,\rho]}(i))$ as in (3.18). We take the average of these N Black-Scholes values, $\bar{C}_{[t]}$, as an approximation of the true call price C_t at time t. This is justified by the weak law of large numbers. This approximation improves for larger N, but comes at the cost of increased computing time. As before, we take $N = 2000$ in our simulations, as this number appears to result in a sufficiently small Monte Carlo standard error of the simulated call price. Repeating this procedure for each $t = \Delta t, .., T\Delta t$, yields the simulated call price series $\{C_{[\Delta t]}, C_{[2\Delta t]}, .., C_{[T\Delta t]}\}$.

But how exactly can we obtain $N = 2000$ realizations of $\bar{\sigma}_t^2$ and the Itô integral under $\mathbb{Q}$, given the initial simulated stock volatility $\sigma_{[t]} = \sqrt{x_{[t]}}$ under $\mathbb{P}$ at time t? Each realization requires a simulated $\mathbb{Q}$-path of $\{x_t\}$ from (2.17) over the interval $(t, T_t]$, given the simulated $\mathbb{P}$-value of x_t obtained in step 1. A Milstein approximation of SDE (2.17) yields the recursive scheme

$x_{[t]} = x_{[t]}$, in which $x_{[t]}$ is the simulated time-t factor value under $\mathbb{P}$, obtained in step 1 $\qquad (\mathbb{Q}) \qquad (5.4)$

$$x_{[t+(i+1)\delta t]} = x_{[t+i\delta t]} + \tilde{k}(\tilde{\theta} - x_{[t+i\delta t]})\delta t + \sigma\sqrt{x_{[t+i\delta t]}}\sqrt{\delta t}\,\epsilon_i + \tfrac{1}{4}\sigma^2\delta t\left[\epsilon_i^2 - 1\right],$$

for $i = 0, .., n_t - 1$, with $n_t \equiv \tau_t I/\Delta t$ and $\epsilon_i \sim$ i.i.d. $\mathcal{N}(0,1)$ under $\mathbb{Q}$. To simulate the call price series, we use this Milstein scheme based on $I = 2$ intraday time points (and not $I = 48$ [20]), implying an SDE-discretization step of $\delta t = \Delta t / I = 1/(260 * 2) = 0.00192$ in this case.

[20] In step 1, we simulate the SDE system (2.3)-(2.5) using $I = 48$ intraday time points, implying a 10-minute RV measure. Taking $I = 48$ in step 4 results in a 15-second duration for simulating 1 call price, or 4.4 hours for 1 full simulated dataset, which consists of 1058 call prices (for each day one.) As we aim for simulating 150 datasets for each of the 5 parameter sets, this would imply a duration of 20 weeks(!). If we take $I = 2$ instead, this latter duration reduces to 41 hours per parameter set, or 8.5 days in total. (Notice that this duration is only the time it takes to simulate the datasets; the duration of

Given such a simulated $\mathbb{Q}$-path of $\{x_t\}$, we next approximate the average variance at time t using the *trapezoidal rule* of numerical integration (see e.g. Atkinson (1993)), by

$$\bar{\sigma}_t^2 \approx \frac{1}{\tau_t} \delta t \left[\tfrac{1}{2} x_{[t]} + x_{[t+\delta t]} + .. + x_{[t+(n_t-1)\delta t]} + \tfrac{1}{2} x_{[t+n_t \delta t]} \right]. \qquad (\mathbb{Q}) \qquad (5.5)$$

(The trapezoidal rule means an improvement to a conventional approximating Riemann sum.) We approximate the Itô integral in (5.3) at time t by the quantity $\text{Integral}[t + n_t \delta t]$, which is the final evaluation point of the recursion

$$\text{Integral}[t] = 0 \qquad (\mathbb{Q}) \qquad (5.6)$$

$$\text{Integral}[t + (i+1)\delta t] = \text{Integral}[t + i\delta t] + \sqrt{x_{[t+i\delta t]}}\sqrt{\delta t}\, \epsilon_i + \tfrac{1}{4}\delta t \left[\epsilon_i^2 - 1 \right], \; i = 0,..,n_t - 1.$$

Step 4: Obtain the Black-Scholes implied variance series
Given the call price series $\{C_{[t]}; t = \Delta t,..,T\Delta t\}$, we obtain the associated simulated Black-Scholes implied variance series $\{\sigma^2_{implied,[t]};\ t = \Delta t,..,T\Delta t\}$ by numerically solving for $\sigma^2_{implied,[t]}$ from the equation $C_{[t]} = BS(S_{[t]}, K_t, \tau_t, r, q, \sigma^2_{implied,[t]})$.

Repeating step 1 to 4 a 150 times results in 150 simulated datasets of the Heston model for a chosen set of parameter values. These datasets serve to investigate the performance of our state space estimation method in section 5.3.

Illustrating the implied volatility surface, smile/skew and term structure
In the next section, we first examine the various shapes of the (Black-Scholes) *implied volatility surface* that the Heston model can generate on a particular day. Simulation of Heston prices of option contracts that differ in maturity and strikes is therefore required.

Suppose first that, at day t, we desire to simulate the prices of a cross-section of $j = 1,2,..,nc$ call options that differ in strikes, but all have equal remaining time to maturity $\tau_t = T_t - t$. Suppose that the current stock price equals $S_t = 100$, whereas the current stock variance equals $\sigma_t^2 = x_t$ (under $\mathbb{P}$). We fix the moneyness of these options as $M_1, M_2, ..., M_{nc}$, with $M_j = K_j / S_t$, i.e., the ratio of strike and stock price. [21] Their strike prices thus equal $K_j = M_j S_t$. As all options have the same time to maturity, their Heston call prices are given by $C_{tj} = \mathbb{E}_{\mathbb{Q}}[BS(S_{t,\rho}, K_j, \tau_t, r, q, \bar{\sigma}^2_{t,\rho}) \mid \mathcal{F}_t]$. Notice that $S_{t,\rho}$ and $\bar{\sigma}^2_{t,\rho}$ are the same for each option. We therefore need to simulate N average variances $\bar{\sigma}_t^2$ and Itô integrals (5.3) only, to be able to compute the nc Heston call option prices simultaneously, as outlined in step 3. These Heston call prices can next be transformed to Black-Scholes implied volatilities, as explained in step 4. The plot of these implied volatilities against moneyness is referred to as the *volatility smile* or *volatility skew* (or *smirk*) for that maturity.

If we repeat this procedure for different maturities and next plot the BS implied volatilities on the resulting moneyness – maturity grid, we obtain a graph of the *implied volatility surface* on day t. Selecting a certain moneyness M_j, and

subsequent state space estimation is not yet added to this.) One reason to focus on a refined Milstein scheme instead of a simple Euler scheme is to "partly compensate" for taking $I = 2$ only.

[21] Defining moneyness as $M_j = K_j / S_t$ results in a downward-sloping volatility skew, which is the plot practitioners have in mind. (Defining $M_j = S_t / K_j$ results in an upward-sloping "skew".)

plotting the associated BS implied volatilities against maturity, results in the *volatility term structure* for that moneyness.

It follows from the put-call parity that European call and put options with the same strike and maturity have an identical BS implied volatility. The implied volatility surfaces of calls and puts thus coincide.

5.2 Simulation results for the Heston model: Parameter influence on stock returns and option prices

The Heston model has five parameters: θ, k, σ, γ and ρ. In what way do these parameters determine the stock return and volatility distributions? How do they influence option prices, i.e., how do they impact on the shape of the smiles and term structures?

5.2.1 Parameters effects on stock returns and volatility

From (2.3), the conditional distribution of the instantaneous stock return is Gaussian, $dS_t / S_t \mid \mathcal{F}_t \sim \mathcal{N}(\mu_t dt, \sigma_t^2 dt)$, with $\{\sigma_t^2 = x_t\}$ representing the conditional variance process of the spot stock return. (Due to SV, the distribution of the stock return over any finite horizon is not Gaussian however.)

Parameter θ determines the average level of the stock volatility: $\mathbb{E}_{\mathbb{P}}[\sigma_t^2] = \theta$. The larger θ, the larger the mean volatility level.

Parameter k determines the speed at which the stock variance $\sigma_t^2 = x_t$ reverts back to its mean: As $\mathbb{E}_{\mathbb{P}}[dx_t \mid \mathcal{F}_t] = k(\theta - x_t)dt$, it is clear that (for $k, \theta > 0$) the stock volatility is always drawn back to its mean of θ, at speed k. Equivalently, k measures the volatility persistence ($\exp[-k\Delta t]$), and determines the half-life of a volatility shock (which is given by $\ln 2 / k\Delta t$). If $k = 0$, there is no mean reversion, volatility shocks are persistent and impact on future volatility for ever. The volatility shows random walk behavior in that case. The larger k, the less persistent volatility is, the faster the mean reversion, and the faster the impact of shocks dies out.

Parameter σ determines the volatility of volatility shocks: $\text{var}_{\mathbb{P}}[dx_t \mid \mathcal{F}_t] = \sigma^2 x_t dt$. The larger σ, the larger subsequent volatility fluctuations (ceteris paribus), and the larger the unconditional volatility-of-volatility: $\text{var}_{\mathbb{P}}[\sigma_t^2] = \sigma^2\theta / 2k$. A larger θ implies a larger $\text{var}_{\mathbb{P}}[\sigma_t^2]$ as well, as the volatility-of-volatility is level-dependent. Moreover, the more persistent volatility is (the smaller k), the larger the unconditional volatility-of-volatility. Intuitively, smaller k's imply that the volatility tends to wander away from its mean more than for larger k's, resulting in more volatility variation. Parameter σ also controls the kurtosis of the stock return distribution. If stock volatility were constant, stock returns over any horizon would be normally distributed, and hence have a kurtosis of 3. The larger σ, the more the stock volatility fluctuates, the larger the deviation from normality, the larger the excess kurtosis (ceteris paribus).

Parameter ρ determines the correlation between stock returns and volatility shocks (see (2.6)), with the leverage effect indicated by $\rho < 0$. A negative ρ also induces negative skewness in the stock return distribution: Declining stock prices (negative returns) are generally accompanied by increasing volatility (the leverage effect) and thus by larger return variation, than if stock prices go up (positive returns). The return distribution will therefore be skewed to the left. [22] (More properties of the Heston (CIR) SV process are in appendix B, section 14.)

5.2.2 Implied volatility surface if there is no leverage effect

First, recall that the constant-volatility Black-Scholes model implies a horizontal, flat implied volatility surface. SV without leverage effect implies a smile across equal-maturity, different-moneyness options. (See Renault and Touzi (1996) for a mathematical proof.)

As a benchmark, figure 5.1 plots the BS implied volatility surface on a *medium-volatility day* (i.e. $\sigma_t^2 = \theta = \mathbb{E}_{\mathbb{P}}[\sigma_t^2]$) for parameter set IV, but assuming no leverage effect. Hence, $\theta = 0.04$, $k = 7$, $\sigma = 0.60$, $\gamma = -7$ and $\rho = 0$. The current stock price equals $S_t = 100$. The moneyness K / S_t ranges from 0.70 to 1.30 (in steps of 0.05), and the maturity from 1 month to 1 year (in steps of 1 month). Figure 5.2 plots the associated volatility smiles for the maturities 1, 3, 6, 9 and 12 months. The locations of currently *at-the-money* ($S_t = K$) calls and puts, *in-the-money* calls ($S_t > K$) and puts ($S_t < K$), and *out-of-the-money* calls ($S_t < K$) and puts ($S_t > K$) are highlighted.

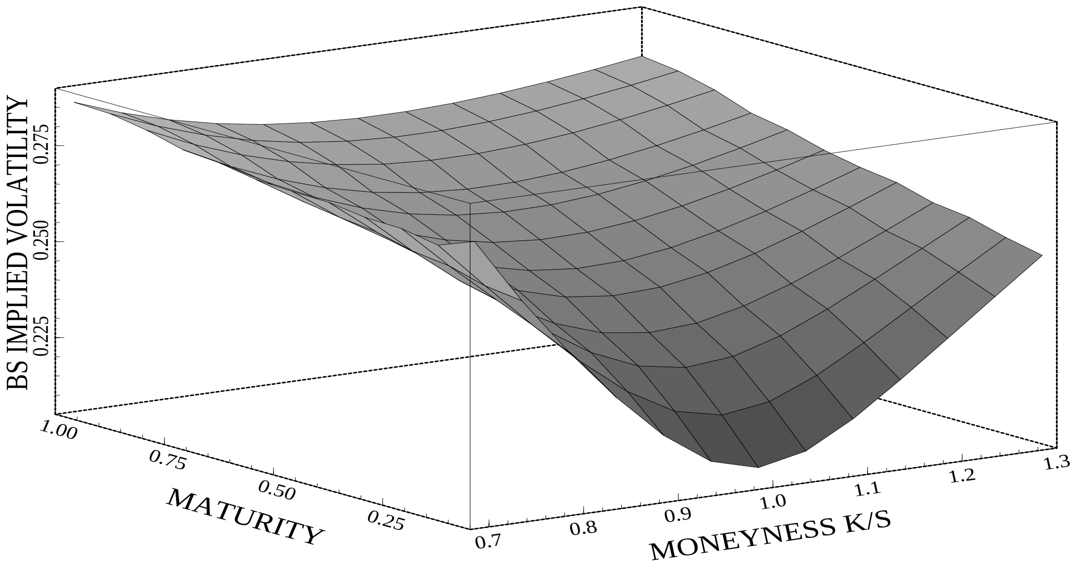


Figure 5.1: BS implied volatility surface on a medium-volatility day for parameter set IV, but with $\rho = 0$ (no leverage effect), assuming $S_t = 100$ and $\sigma_t^2 = \mathbb{E}_{\mathbb{P}}[\sigma_t^2] = \theta$.

Volatility smile flattens out the longer the option maturity
Relative to ATM options, stochastic volatility results in larger prices for ITM and OTM options: the smile effect. The volatility smile is most pronounced for short-maturity options, and *flattens out* the longer the maturity of the options. An intuitive explanation is as follows. From (3.17)-(3.18), if $\rho = 0$ the call price becomes $C_t = \mathbb{E}_{\mathbb{Q}}[BS(S_t, K, \tau_t, r, q, \bar{\sigma}_t^2) \mid \mathcal{F}_t]$. From (9.17)-(9.22) in appendix B, given $\mathcal{F}_t$, the average variance $\bar{\sigma}_t^2$ converges to $\tilde{\theta}$ under $\mathbb{Q}$ for $T \to \infty$. This is due to volatility mean reversion (i.e. stationarity); shocks tend to "average out" over time. Therefore, the longer the maturity, the less $\mathbb{E}_{\mathbb{Q}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]$ depends on the

[22] See e.g. Franses and Van Dijk (2000) for recent empirical stock return evidence that supports this model implication. The distribution of empirical exchange rate returns is typically much more symmetric, which corresponds to $\rho = 0$.

current stock volatility $\sigma_t^2 = x_t$, the closer its value is to its limiting constant value $\tilde{\theta}$, and thus the more horizontal the volatility smile will be (corresponding to the constant volatility BS model).

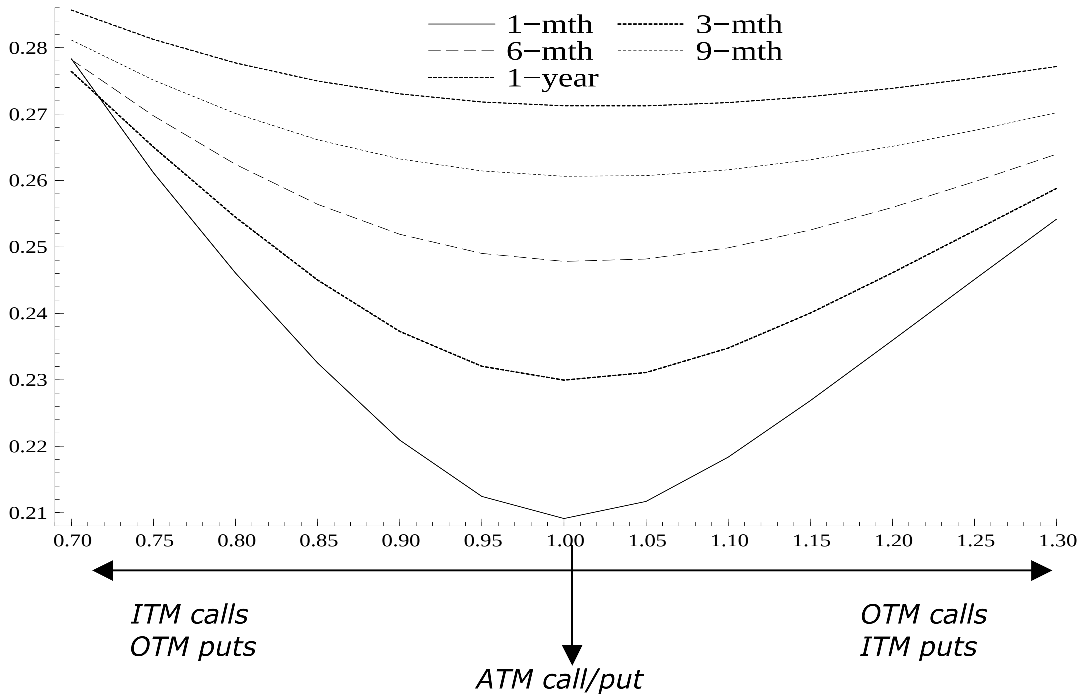


Figure 5.2: Volatility smiles associated with the implied volatility surface of figure 5.1. Locations of currently at-the-money (ATM), in-the-money (ITM) and out-of-the-money (OTM) call and put options are indicated. Horizontal axis: moneyness $M = K / S_t$. Vertical axis: BS implied volatility.

ATM VTS is upward sloping on medium-volatility days, due to $\gamma < 0$

On this medium-volatility day for which $\sigma_t^2 = \mathbb{E}_{\mathbb{P}}[\sigma_t^2] = \theta$, the VTS is upward sloping, and steepest for ATM options. Similar pictures for *low-volatility days* reveal steeper upward-sloping term structures, whereas for *high-volatility days*, the term structures are downward sloping (and the smiles remain). These findings may also be explained by volatility mean reversion. They confirm the in practice often-assigned interpretation of BS implied volatilities being market forecasts of future volatility, at least on low and high-volatility days. Namely, on a medium-volatility day, the volatility is as likely to go up as down, as in that case $\mathbb{E}_{\mathbb{P}}[\sigma_s^2 \mid \mathcal{F}_t] = \theta$ for $s > t$ (see (14.7) in appendix B). The upward-sloping ATM volatility term structure on medium-volatility days is therefore misleading.

Now, why is the ATM VTS upward sloping on a medium-volatility day and not more or less flat? We attribute this to the assumption of a *negative* market price of volatility risk, which results in a mean $\mathbb{Q}$-volatility $\tilde{\theta}$ larger than the mean $\mathbb{P}$-volatility θ. In this case, $\theta = 0.04$, $\tilde{\theta} = 0.10$. Option prices, and thus BS implied volatilities, are determined by the (mean-reverting) risk-neutral volatility process. So on a medium-($\mathbb{P}$)-volatility day for which $\sigma_t^2 = \theta < \tilde{\theta}$, it is not surprising that the VTS is upward-sloping: The $\mathbb{Q}$-volatility tends to revert to its larger mean $\tilde{\theta}$, starting from $\sigma_t^2 = \theta$, which results in larger option prices (and hence larger implied volatilities) for longer maturities. This reasoning is confirmed by setting the current volatility to $\sigma_t^2 = \tilde{\theta}$, which results in a close-to-flat ATM VTS. (There is a difference of 0.75% only between the 1-mth and 1-year implied volatility in that case). Moreover, taking $\sigma_t^2 = \theta$ and setting $\gamma = 0$ (no volatility risk premium), such that the $\mathbb{P}$ and $\mathbb{Q}$ volatility processes coincide and have mean θ, the ATM VTS is also virtually flat (a difference of 0.25% only). The ATM VTS appears steeper, the larger the deviation between the current ($\mathbb{P}$-) stock volatility σ_t^2 and the mean $\mathbb{Q}$-volatility $\tilde{\theta}$. [23] (Indeed, recall that for ATM options it holds that $\sigma_{implied,t}^2 \approx \mathbb{E}_{\mathbb{Q}}[\bar{\sigma}_t^2 \mid \mathcal{F}_t]$ in the multifactor SV model without leverage effect.)

This analysis indicates that if the ATM VTS is upward sloping on a medium-volatility day in practice, this suggests a negative volatility risk premium. (This confirms the results based on the FTSE100-index data: We found the median ATM term structure to be upward sloping and a negative volatility risk premium.)

5.2.3 Parameter effects on shape implied volatility surface

How does each of the parameters θ, k, σ, γ, ρ influence the shape of the implied volatility surface and thus option prices? Recall first the risk-neutral mean-reversion and mean-volatility parameters, $\tilde{k} = k + \sigma\gamma$ and $\tilde{\theta} = k\theta / \tilde{k}$. Starting from the benchmark *medium*-volatility day surface ($\sigma_t^2 = \mathbb{E}_{\mathbb{P}}[\sigma_t^2] = \theta$) plotted in figure 5.1, a ceteris-paribus analysis reveals the following.

The main implication of changing θ is its indirect effect on the general level of the surface. As option prices depend on the $\mathbb{Q}$-volatility process, increasing θ leads to an increase in $\tilde{\theta}$, hence option prices, and thus BS implied volatilities. [24]

If volatility risk is not priced ($\gamma = 0$), the $\mathbb{P}$ and $\mathbb{Q}$-volatility processes coincide: $\tilde{k} = k$ and $\tilde{\theta} = \theta$. The ATM VTS is then virtually flat, and averages a bit below the mean stock volatility. Volatility smiles are present (due to SV) and the VTS of ITM and OTM options is downward sloping. Negative values of γ decrease the volatility mean reversion under $\mathbb{Q}$ ($\tilde{k}$ becomes smaller) as opposed to under $\mathbb{P}$, and increase $\tilde{\theta}$. As a result, the surface shifts upward. The ATM VTS becomes upward sloping first, whereas the OTM and ITM VTS gradually change from downward to upward sloping (this occurs first for options closest ATM). If γ becomes too large negative resulting in an *explosive* $\mathbb{Q}$-volatility process (if $\tilde{k}, \tilde{\theta} < 0$), the whole surface becomes very steeply upward-sloping. (See also the empirical evidence reported in section 6.3.1 of chapter II.)

Low values of volatility-of-volatility parameter σ result in flat volatility surfaces, hence flat smiles and term structures. The larger σ, the more pronounced the volatility smiles, and the steeper-sloping VTS result. Intuitively, the smaller σ, the less σ_t^2 varies, the closer the model is to the constant-volatility BS model, which implies a flat surface. [25] (The steeper-sloping VTS are caused by the fact that an increase in σ results in an increase in $\tilde{\theta}$ (because $\tilde{k} = k + \sigma\gamma$ decreases), enlarging the difference between the current $\mathbb{P}$-volatility $\sigma_t^2 = \theta$, and $\tilde{\theta}$.)

[23] Intuitively, as $\mathbb{E}_{\mathbb{Q}}[dx_t \mid \mathcal{F}_t] = \tilde{k}(\tilde{\theta} - x_t)dt$, the larger the difference $\tilde{\theta} - x_t$, the larger the subsequent correction towards the mean volatility $\tilde{\theta}$ tends to be.

[24] A side-effect of increasing θ is the steeper and steeper becoming VTS (and it moreover flattens). We attribute this to the difference between the current $\mathbb{P}$-volatility $\sigma_t^2 = \theta$ and the mean $\mathbb{Q}$-volatility $\tilde{\theta} = k\theta / \tilde{k}$, which becomes larger and larger, as $k / \tilde{k} = 7 / 2.8 = 2.5$ remains constant.

[25] This analysis intuitively clarifies why adding *jumps to returns or volatility* results in even more pronounced smiles (recall footnote 3): Increasing σ enlarges the deviation from lognormality (the BS assumption), i.e., the kurtosis of the stock returns increases, which results in more pronounced smiles. Now, as adding jumps to returns or volatility further increase the kurtosis, this explains why such jumps lead to stronger smiles (and thus skews, if the leverage effect is taken into account.)

If mean-reversion parameter k is small enough to have explosive $\mathbb{Q}$-volatility ($\tilde{k}, \tilde{\theta} < 0$), the smile is present for all maturities, but most prominent are the very steeply-sloping VTS for all maturities. Small values of k that still imply stationary $\mathbb{Q}$-volatility (with small $\tilde{k} > 0$), result in a still steeply-sloped smiled surface. The larger k, the more the surface flattens, except for SM options, for which the smile remains prominent. Intuitively, the quicker the mean-reversion (the larger k's and thus $\tilde{k}$'s), the more transient a shock, the more the $\mathbb{Q}$-volatility will stay and move around its mean $\tilde{\theta}$ (recall $\text{var}_{\mathbb{Q}}[\sigma_t^2] = \sigma^2\tilde{\theta}/2\tilde{k}$) once it has been drawn back to it (from its $\mathbb{P}$-starting value $\sigma_t^2 = \theta$), and thus the more constant the $\mathbb{Q}$-volatility, and hence the flatter the surface. (We attribute the still prominent smile for SM options to the time it takes for the volatility to mean-revert from $\sigma_t^2 = \theta$ to its risk-neutral mean $\tilde{\theta}$.) In particular, a large k combined with a small σ, implies a very flat-shaped implied volatility surface.

5.2.4 Impact of the leverage effect

The analysis so far was based on the absence of leverage. Negatively correlating stock and volatility changes lead to a volatility *skew* across same-maturity different-moneyness options. Figures 5.3 and 5.4 plot the implied volatility surface and corresponding skews on a medium-volatility day ($\sigma_t^2 = \mathbb{E}_{\mathbb{P}}[\sigma_t^2] = \theta$), for parameter set IV, for which $\theta = 0.04$, $k = 7$, $\sigma = 0.60$, $\gamma = -7$ and $\rho = -0.60$.

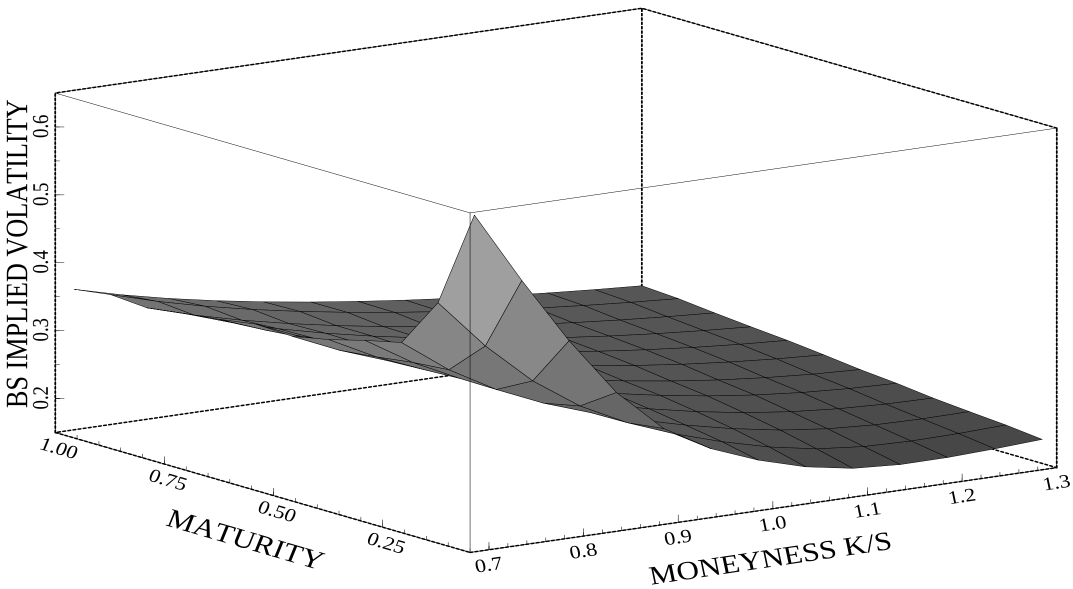


Figure 5.3: BS implied volatility surface on a medium-volatility day for parameter set IV, for which $\rho = -0.60$, assuming $S_t = 100$ and $\sigma_t^2 = \mathbb{E}_{\mathbb{P}}[\sigma_t^2] = \theta$. (Compare to figure 5.1.)

The skew is most pronounced for short-maturity options, and flattens out the longer the maturity. (Compare to figures 5.1 and 5.2 which assume no leverage effect. Although the ATM VTS is still upward sloping, for other moneyness it may now be upward or downward sloping.)

Parameter ρ truly controls the skew: As a ceteris-paribus analysis shows, the more negative ρ is, the skewer and steeper the skews, for all maturities.

On low-volatility days, both the skews and term structures are steeper. On high-volatility days, the skews are less steep, and the VTS are all downward sloping. Given our earlier results based on no leverage effect, this makes sense.

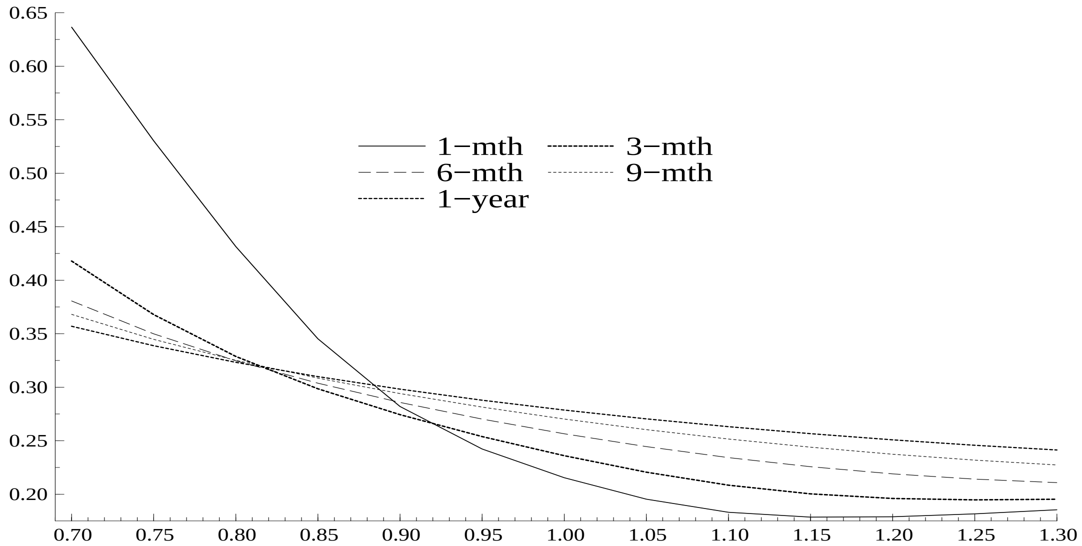


Figure 5.4: Volatility skews associated with the implied volatility surface of figure 5.3. Horizontal axis: moneyness $M = K / S_t$. Vertical axis: Black-Scholes implied volatility.

ATM VTS is not much affected by the leverage effect

Correlation parameter ρ controls the skew, but how does ρ affect the volatility term structure of ATM options? The empirical analysis of the UK market in previous chapters focused on the VTS of ATM options. The estimated multifactor SV models ignored the leverage effect, and are therefore misspecified (given the supporting evidence of its presence). Besides their largest liquidity and vega, our exclusive focus on ATM options was driven by the expectation that leverage mostly impacts on the prices of OTM and ITM options, and is of minor importance for ATM options. Let us now investigate if this is reasonable by simulation.

Figure 5.5 plots the ATM VTS on a medium-volatility day ($\sigma_t = 20\%$) for parameter set IV, for ρ = 0, -0.20, -0.40, -0.60 and -0.80. Table 5.3 reports the corresponding BS implied volatilities for maturities 1, 2, 3, 6, 9 and 12 months, for each value of ρ, and similarly for a low ($\sigma_t = 5\%$) and high ($\sigma_t = 35\%$) volatility day. Recall that set IV is based on our FTSE100-index estimation results for Heston's model, based on the combination squared return–SM option data. [26]

The ATM VTS is indeed not much affected by the leverage effect. Take a medium-volatility day for example. As opposed to $\rho = 0$, a correlation of –0.60 shifts the VTS upward about 0.7% only. A similar analysis for the other datasets confirms the in general minor impact of ρ on the at-the-money VTS. In contrast, and as a comparison of figures 5.2 and 5.4 illustrates, the leverage effect impacts heavily on the prices of ITM and OTM options.

[26] The reader may note the much steeper-sloping medium-volatility-day *simulated* ATM VTS, as compared to the median *FTSE100-index* ATM VTS (recall table 2.1 in chapter IV). Here, there is a simulated implied-volatility difference of 6.5% between 1-month and 9-month options (for $\rho = -0.60$), in the FTSE100-index data this difference is about 1.4% only. We attribute this to the severe misspecification of the (1-factor SV) Heston model for the joint FTSE100-index data. We needed a 3-factor SV specification, which can generate much more complicated ATM VTS than the Heston model can.

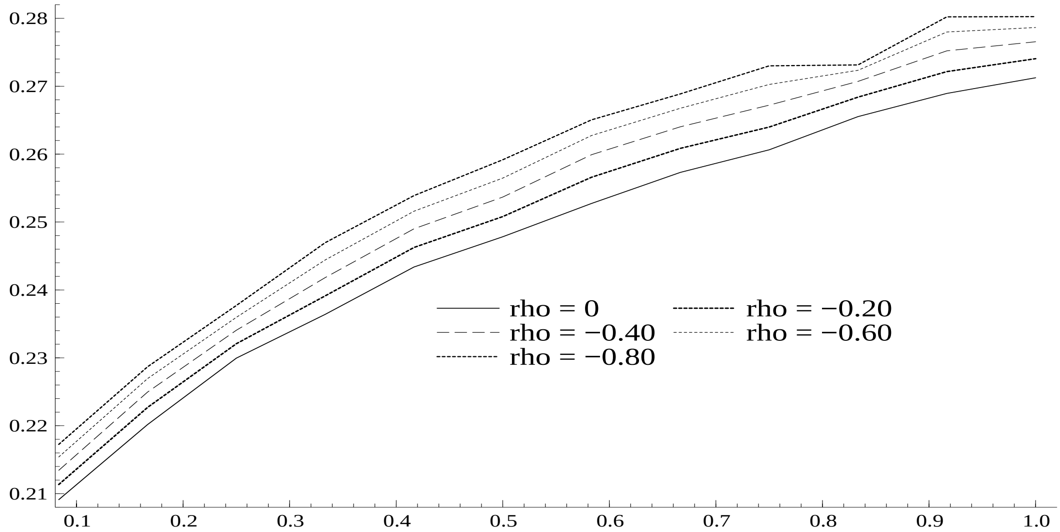


Figure 5.5: ATM VTS on a medium-volatility day ($\sigma_t = 20\%$) for parameter set IV, for ρ =0, -0.20, -0.40, -0.60, -0.80. Horizontal axis: maturity (years), vertical: BS impl.vol.

Table 5.3: Simulated ATM VTS on a medium, low and high volatility day for set IV

Medium-volatility day ($\sigma_t = 20\%$):

Maturity	$\rho = 0$	$\rho = -0.20$	$\rho = -0.40$	$\rho = -0.60$	$\rho = -0.80$
1	20.9%	21.1%	21.3%	21.5%	21.7%
2	22.0%	22.3%	22.5%	22.7%	22.9%
3	23.0%	23.2%	23.4%	23.6%	23.8%
6	24.8%	25.1%	25.4%	25.7%	25.9%
9	26.1%	26.4%	26.7%	27.0%	27.3%
12	27.1%	27.4%	27.7%	27.9%	28.0%

Low-volatility day ($\sigma_t = 5\%$):

Maturity	$\rho = 0$	$\rho = -0.20$	$\rho = -0.40$	$\rho = -0.60$	$\rho = -0.80$
1	10.7%	10.9%	11.0%	11.2%	11.3%
2	14.1%	14.3%	14.5%	14.7%	14.8%
3	16.5%	16.7%	16.9%	17.1%	17.3%
6	20.7%	20.9%	21.2%	21.4%	21.6%
9	23.1%	23.4%	23.7%	23.9%	24.2%
12	24.9%	25.1%	25.4%	25.6%	25.8%

High-volatility day ($\sigma_t = 35\%$):

Maturity	$\rho = 0$	$\rho = -0.20$	$\rho = -0.40$	$\rho = -0.60$	$\rho = -0.80$
1	34.1%	34.4%	34.7%	35.0%	35.2%
2	33.7%	34.0%	34.3%	34.6%	34.8%
3	33.4%	33.6%	33.8%	34.0%	34.2%
6	32.2%	32.6%	32.9%	33.3%	33.6%
9	31.7%	32.1%	32.5%	32.9%	33.2%
12	31.6%	31.9%	32.2%	32.4%	32.5%

The table reports the ATM VTS on a medium, low and high-volatility day for parameter set IV, for different values of ρ . Maturity is in months. Numbers are BS implied volatilities.

This Monte Carlo evidence for the Heston model thus at least suggests that our multifactor SV estimation results based on the UK data are not much biased towards having neglected leverage. The next section provides additional evidence suggesting this.

5.3 Simulation results for the Heston model: Performance of the state space estimation method

This section examines the performance of the estimation method based on simulated data from the Heston (1993) model. (See section 5.1 for assumptions, and section 4.3 for the state space method.) We thus restrict $\rho = 0$ prior to estimation such that the conditional state space model results (see section 3.3 of chapter V), which is based on no leverage effect, and estimate the parameters by Extended Kalman filter QML. We next estimate ρ by (2.6) using the smoothed volatilities extracted from the data at the optimum.

In the current Heston setting the simulated data thus takes the leverage effect into account (read: empirical FTSE100-index data), though the conditional state space model is based on the absence of leverage. In the Monte Carlo study towards the 1-factor CIR SV model (section 4 of chapter V), both the simulated data and the conditional state space model were based on the absence of leverage. Comparing the tables for this latter study with the Heston tables below thus allows for an investigation of the additional bias induced by leverage.

As before, we consider estimation based on (combinations of) squared return, RV and SM ATMF option data. How close are the estimates and smoothed volatilities to their true values? What data yields the best results in this respect? [27] As before, squared returns perform worst, the combination RV - option data best. Overall speaking, the additional bias due to the leverage effect is rather modest. This holds in particular for the combination RV – option data. We attribute this to the leverage effect not really being important for pricing ATM options, as shown in the last section. Once option data is included for estimation, the estimator for ρ performs seemingly well: it has a small positive bias only.

5.3.1 Results using squared returns for estimation

Table 5.4 reports the estimation results for the Heston model if only squared return data is used for estimation. Parameter γ cannot be estimated in this case.

The results are very similar to the corresponding squared-return results found in the Monte Carlo study towards the 1-factor CIR SV model (which assumes no leverage effect); see table 4.8 in section 4.4.1 of the last chapter. As such, we keep the discussion brief. The bias in the estimates is large and the volatility evaluation criteria RMSPE and MAPE for comparing the smoothed with the true underlying volatilities indicate poor volatility extraction. Also in this case it sometimes occurs that θ is estimated near zero, which is generally accompanied by very large estimates of k and σ, and this becomes worse when subsequently going from set I to IV. The estimator of ρ performs poorly in this case.

Table 5.5 reports estimation results based on squared returns only, but with θ restricted to its method-of-moments estimate $\hat{\theta}_{SqR}$ prior to estimation (computed as the average of the annualized squared returns in deviation from their mean). Again, the results are very similar as found in the Monte Carlo study towards the 1-factor CIR SV model; see table 4.9 in section 4.4.1 of the previous chapter.

[27] To keep the discussion brief, we do not separately consider the quality of the approximations carried out to arrive at a linear state space model, neither do we separately investigate the assumptions with regard to the state space errors. For practical purposes, the method's performance matters most.

To repeat: Using squared return data only for QML estimation is generally not to be advocated, not even if θ is restricted to its moment estimate.

Table 5.4: Estimation results Heston model based on **squared return data**

	I	*II*	*III*	*IV*	*V*
θ	**0.039**	**0.035**	**0.031**	**0.023**	**0.018**
	(0.012)	(0.014)	(0.016)	(0.017)	(0.006)
Bias	-0.001	-0.005	-0.009	-0.017	-0.002
MSE	1.4e-4	2.3e-4	3.4e-4	5.8e-4	3.4e-5
k	**7.0**	**18.4**	**29.0**	**61.0**	**29.3**
	(18.6)	(46.6)	(49.9)	(85.0)	(50.1)
Bias	5.5	15.8	23.8	54.0	22.3
MSE	375	2423	3060	10143	3011
σ	**0.41**	**1.27**	**2.31**	**5.85**	**1.03**
	(1.12)	(3.23)	(4.91)	(13.7)	(2.12)
Bias	0.21	0.97	1.86	5.25	0.73
MSE	1.30	11.4	27.5	215	5.03
ρ	**-0.05**	**-0.06**	**-0.06**	**-0.05**	**-0.06**
	(0.04)	(0.05)	(0.06)	(0.07)	(0.05)
Bias	0.36	0.44	0.49	0.55	0.45
MSE	0.13	0.20	0.25	0.31	0.20
RMSPE	**14.6**	**22.5**	**29.8**	**42.9**	**23.5**
%	(5.3)	(10.4)	(9.4)	(11.4)	(7.6)
MAPE	**10.7**	**15.9**	**20.4**	**28.6**	**17.1**
%	(4.0)	(8.8)	(8.5)	(11.0)	(6.9)

For each parameter set I-V, the estimation results are summarized over the 150 simulated datasets jointly. The conditional state space model was estimated using **squared return data** only. The average estimates of θ, k, σ, ρ are reported (std.dev. in parentheses), with associated sample bias and MSE. Based on (2.6), ρ is estimated as the correlation between returns and smoothed stock variance changes. The average RMSPE and MAPE, for comparing the smoothed and true volatilities, are also reported (std.dev. in (.)).

Table 5.5: Results Heston model based on **squared returns**, under restriction $\theta = \hat{\theta}_{SqR}$

	I	*II*	*III*	*IV*	*V*
$\theta = \hat{\theta}_{SqR}$	**0.040**	**0.041**	**0.040**	**0.040**	**0.020**
	(0.012)	(0.011)	(0.009)	(0.009)	(0.003)
Bias	0.000	0.001	0.000	0.000	0.000
MSE	1.5e-5	1.3e-4	7.8e-5	8.3e-5	1.0e-5
k	**5.8**	**11.8**	**19.2**	**24.5**	**23.5**
	(13.3)	(37.2)	(38.5)	(46.4)	(38.7)
Bias	4.3	9.2	14.0	17.5	16.5
MSE	195	1471	1680	2462	1765
σ	**0.32**	**0.64**	**1.08**	**1.65**	**0.67**
	(0.33)	(1.02)	(1.37)	(1.59)	(0.83)
Bias	0.12	0.34	0.63	1.05	0.37
MSE	0.12	1.15	2.26	3.63	0.83
ρ	**-0.05**	**-0.06**	**-0.06**	**-0.06**	**-0.06**
	(0.04)	(0.04)	(0.05)	(0.06)	(0.05)
Bias	0.35	0.44	0.49	0.54	0.44
MSE	0.13	0.19	0.24	0.30	0.20
RMSPE	**14.4**	**21.0**	**28.5**	**41.3**	**22.5**
%	(3.8)	(7.2)	(7.7)	(11.5)	(5.3)
MAPE	**10.5**	**14.4**	**18.6**	**23.9**	**16.1**
%	(2.1)	(4.9)	(5.4)	(5.3)	(4.2)

See table 5.4 for further explanatory legend.

5.3.2 Results using realized volatilities for estimation

Table 5.6 reports Heston model results based on RV data only. Parameter γ cannot be estimated in this case (and $c = \text{corr}_{\mathbb{P}}[\varpi_{RV,t+\Delta t}, u_{t+\Delta t}]$ is not identified).

The results are very similar to the corresponding RV results found in the Monte Carlo study towards the 1-factor CIR SV model (though the bias in k is somewhat larger for sets I – II; see table 4.10 in the last chapter). This is not surprising. The RV measurement equation is based on the equation the average variance satisfies over day $[t, t+\Delta t]$: $\frac{1}{\Delta t}\int_t^{t+\Delta t} \sigma_u^2 du = \mathbb{E}_{\mathbb{P}}[\frac{1}{\Delta t}\int_t^{t+\Delta t} \sigma_u^2 du \mid \mathcal{F}_t] + \varpi_{RV,t+\Delta t}$. This equation does not depend on the leverage effect (see (2.5) and (2.1)). We next essentially replaced the average variance with the realized variance to obtain the RV equation. The Heston and 1-factor CIR SV results are not completely the same, because the realized variance measures the average variance with error, and is computed as the sum of squared intraday stock returns (in which the leverage effect is present).

Using RV data means a substantial reduction in bias and MSE in the k, σ and ρ-estimates as opposed to using squared returns. Nonetheless, the bias and MSEs are still large. The bias in θ is even larger. The volatility is much better filtered out than if squared returns are used. Here as well it sometimes happens that θ is estimated near zero, though less frequently than in the squared return case.

Table 5.6: Estimation results Heston model based on **RV data**

	I	*II*	*III*	*IV*	*V*
θ	**0.036**	**0.029**	**0.022**	**0.017**	**0.016**
	(0.014)	(0.015)	(0.011)	(0.009)	(0.004)
Bias	-0.005	-0.011	-0.018	-0.023	-0.004
MSE	2.1e-4	3.5e-4	4.3e-4	6.1e-4	2.8e-5
k	**6.65**	**9.39**	**16.9**	**15.5**	**16.5**
	(9.90)	(6.10)	(5.7)	(5.1)	(6.3)
Bias	5.2	6.8	11.7	8.5	9.5
MSE	125	83	168	98	129
σ	**0.377**	**0.632**	**1.01**	**1.15**	**0.537**
	(0.297)	(0.304)	(0.21)	(0.11)	(0.167)
Bias	0.18	0.33	0.56	0.55	0.24
MSE	0.12	0.20	0.36	0.32	0.08
ρ	**-0.10**	**-0.12**	**-0.12**	**-0.14**	**-0.14**
	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
Bias	0.30	0.38	0.43	0.46	0.36
MSE	0.09	0.15	0.19	0.21	0.13
RMSPE	**5.9**	**8.2**	**11.4**	**14.1**	**8.8**
%	(2.4)	(2.5)	(1.8)	(1.3)	(2.0)
MAPE	**4.6**	**6.3**	**8.9**	**10.6**	**6.9**
%	(2.0)	(2.0)	(1.4)	(0.7)	(1.6)

For each parameter set I-V, the estimation results are summarized over the 150 simulated datasets jointly. The conditional state space model was estimated using **RV data** only. The average estimates of θ, k, σ, ρ are reported (std.dev. in parentheses), with associated sample bias and MSE. Based on (2.6), ρ is estimated as the correlation between returns and smoothed stock variance changes. The average RMSPE and MAPE, for comparing the smoothed and true volatilities, are also reported (std.dev. in (.)).

Table 5.7 reports estimation results based on RV data, but with θ restricted to its RV-based estimate $\hat{\theta}_{RV}$ for each dataset, computed as the mean realized variance. Again, the results are very similar as found in the Monte Carlo study for the 1-factor CIR SV model (see table 4.11 in section 4.4.2 of the last chapter).

Table 5.7: Estimation results Heston model based on **RV data**, under restriction $\theta = \hat{\theta}_{RV}$

	I	*II*	*III*	*IV*	*V*
$\theta = \hat{\theta}_{RV}$	**0.041**	**0.041**	**0.040**	**0.040**	**0.020**
	(0.012)	(0.011)	(0.008)	(0.009)	(0.003)
Bias	0.001	0.001	0.000	0.000	0.000
MSE	1.4e-4	1.2e-4	7.1e-5	7.3e-5	9.1e-6
k	**5.33**	**5.93**	**10.2**	**8.15**	**12.8**
	(8.53)	(3.72)	(5.0)	(5.42)	(4.4)
Bias	3.8	3.3	5.0	1.2	5.8
MSE	87	25	50	31	53
σ	**0.373**	**0.603**	**0.961**	**1.11**	**0.505**
	(0.286)	(0.281)	(0.193)	(0.10)	(0.147)
Bias	0.17	0.30	0.51	0.51	0.21
MSE	0.11	0.17	0.30	0.27	0.06
ρ	**-0.10**	**-0.12**	**-0.13**	**-0.15**	**-0.14**
	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
Bias	0.30	0.38	0.43	0.46	0.36
MSE	0.09	0.15	0.18	0.21	0.13
RMSPE	**5.9**	**8.0**	**11.2**	**14.1**	**8.5**
%	(2.5)	(2.5)	(1.8)	(1.3)	(1.9)
MAPE	**4.6**	**6.2**	**8.8**	**10.6**	**6.7**
%	(2.0)	(2.0)	(1.4)	(0.7)	(1.5)

See table 5.6 for further explanatory legend.

Imposing $\theta = \hat{\theta}_{RV}$ leads to a large bias and MSE reduction in the estimates of θ and k, only modestly affects the estimates of σ, but has virtually no effect on the $\rho-$estimates and the volatility evaluation criteria. An explanation for this latter finding is given in the last chapter; it applies here as well.

5.3.3 Results using option data for estimation

Table 5.8 reports Heston model estimation results if only SM at-the-money option data is used for estimation. Once option data is included for estimation, all five parameters θ, k, σ, γ , ρ of the Heston model can be estimated.

Using option data is clearly preferable to using only squared return or RV data. Option data is much more informative: The bias and MSE are dramatically less, and the RMSPE and MAPE reveal much better volatility extraction.

The estimation bias in θ, σ and ρ is small for all sets I-V. Their seemingly small standard deviations indicate that the estimates are generally close to their true values. The absolute biases do not vary much across the different parameter sets. In contrast, the relative bias decreases when moving from persistent, calmly fluctuating volatility (set I), to less persistent, stronger fluctuating volatility and increased leverage effect (set IV, which is close to our empirical estimates).

The estimator (2.6) of the correlation ρ between stock returns and unobserved volatility shocks seems a particularly useful estimator if option data is used. It has a small positive bias, but is very easy to compute.

The estimates of the mean-reversion and market price of volatility risk determining parameters k and γ are biased upward. The relative bias of k decreases when moving from set I to IV, whereas for γ it remains more constant. As encountered before, the relatively large standard deviations of the γ-estimates indicate that this price-of-risk parameter is hard to pin down precisely.

There is no additional bias induced by the leverage effect in the θ-estimates, a small extra bias in the σ-estimates, though quite some additional bias in the k and γ estimates. (Compare table 5.8 with table 4.12 in the last chapter for the 1-factor CIR SV model).

Table 5.8: Estimation results Heston model based on (short-maturity ATMF) **option data**

	I	*II*	*III*	*IV*	*V*
θ	**0.040**	**0.040**	**0.040**	**0.039**	**0.020**
	(0.011)	(0.011)	(0.008)	(0.008)	(0.003)
Bias	0.000	0.000	0.000	-0.001	0.000
MSE	1.1e-4	1.1e-4	6.6e-5	6.6e-5	7.8e-6
k	**2.11**	**3.22**	**5.73**	**7.31**	**7.67**
	(0.61)	(0.87)	(1.13)	(1.36)	(1.18)
Bias	0.61	0.61	0.48	0.31	0.67
MSE	0.73	1.13	1.50	1.95	1.85
σ	**0.225**	**0.335**	**0.490**	**0.643**	**0.326**
	(0.011)	(0.015)	(0.020)	(0.024)	(0.016)
Bias	0.025	0.035	0.040	0.043	0.026
MSE	7.3e-4	0.001	0.002	0.002	9.5e-4
γ	**-3.4**	**-4.1**	**-4.8**	**-5.7**	**-18.0**
	(2.4)	(2.6)	(2.6)	(2.4)	(3.6)
Bias	0.6	0.9	1.2	1.3	2.0
MSE	5.9	7.5	8.2	7.4	17
ρ	**-0.35**	**-0.45**	**-0.51**	**-0.57**	**-0.46**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Bias	0.05	0.05	0.04	0.03	0.04
MSE	0.0035	0.0040	0.0026	0.0020	0.0022
RMSPE	**2.0**	**3.5**	**5.2**	**8.4**	**3.8**
%	(0.7)	(1.3)	(1.9)	(2.7)	(1.3)
MAPE	**1.5**	**2.3**	**3.1**	**4.3**	**2.6**
%	(0.4)	(0.8)	(1.1)	(1.4)	(0.8)

For each parameter set I-V, the results are summarized over the 150 simulated datasets jointly. The conditional state space model was estimated using **option data** only, with $\mu_V = 0$ imposed. The average estimates of θ, k, σ, γ, ρ are reported (std.dev. in (.)), with associated sample bias and MSE. Based on (2.6), ρ is estimated as the correlation between returns and smoothed stock variance changes. The average RMSPE and MAPE, for comparing the smoothed and true volatilities, are also reported (std.dev. in (.)).

5.3.4 Results using squared return and option data for estimation

Table 5.9 reports Heston model estimation results based on squared return and short-maturity ATM option data.

This combination of data yields better results than using option data only. The bias in notably the k and γ-estimates has decreased further. The extra inclusion of the squared return data does not seem to affect the θ and ρ-estimates, whereas the bias and MSE reduction in the $\sigma-$estimates modestly improves.

Table 5.9: Estimation results Heston model based on **squared return - option data**

	I	*II*	*III*	*IV*	*V*
θ	**0.040**	**0.040**	**0.040**	**0.039**	**0.020**
	(0.011)	(0.011)	(0.008)	(0.009)	(0.003)
Bias	0.000	0.000	0.000	-0.001	0.000
MSE	1.2e-4	1.1e-4	6.5e-5	7.6e-5	7.7e-6
k	**2.06**	**3.15**	**5.64**	**7.32**	**7.48**
	(0.59)	(0.84)	(1.10)	(1.41)	(1.14)
Bias	0.56	0.54	0.39	0.32	0.48
MSE	0.66	1.00	1.35	2.09	1.53
σ	**0.224**	**0.333**	**0.486**	**0.639**	**0.322**
	(0.011)	(0.016)	(0.020)	(0.029)	(0.017)
Bias	0.024	0.033	0.036	0.039	0.022
MSE	6.8e-4	0.001	0.002	0.002	7.9e-4
γ	**-3.6**	**-4.3**	**-5.0**	**-6.0**	**-18.3**
	(2.4)	(2.6)	(2.5)	(2.4)	(3.6)
Bias	0.4	0.7	1.0	1.0	1.7
MSE	6.0	7.1	7.4	7.0	15.6
ρ	**-0.35**	**-0.45**	**-0.51**	**-0.57**	**-0.46**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Bias	0.05	0.05	0.04	0.03	0.04
MSE	0.0035	0.0040	0.0025	0.0020	0.0022
RMSPE	**2.0**	**3.3**	**4.8**	**8.1**	**3.5**
%	(0.7)	(1.3)	(1.8)	(2.6)	(1.2)
MAPE	**1.5**	**2.2**	**2.9**	**4.1**	**2.4**
%	(0.4)	(0.7)	(0.9)	(1.2)	(0.8)

For parameter set I-V, the results are summarized over the 150 simulated datasets jointly. The conditional state space model was estimated using **squared return-option data** with $\mu_V = 0$ imposed. The average estimates of θ, k, σ, γ, ρ are reported (std.dev. in (.)), with associated sample bias and MSE. Based on (2.6), ρ is estimated as the correlation between returns and smoothed stock variance changes. The average RMSPE and MAPE, for comparing the smoothed and true volatilities, are also reported (std.dev. in (.)).

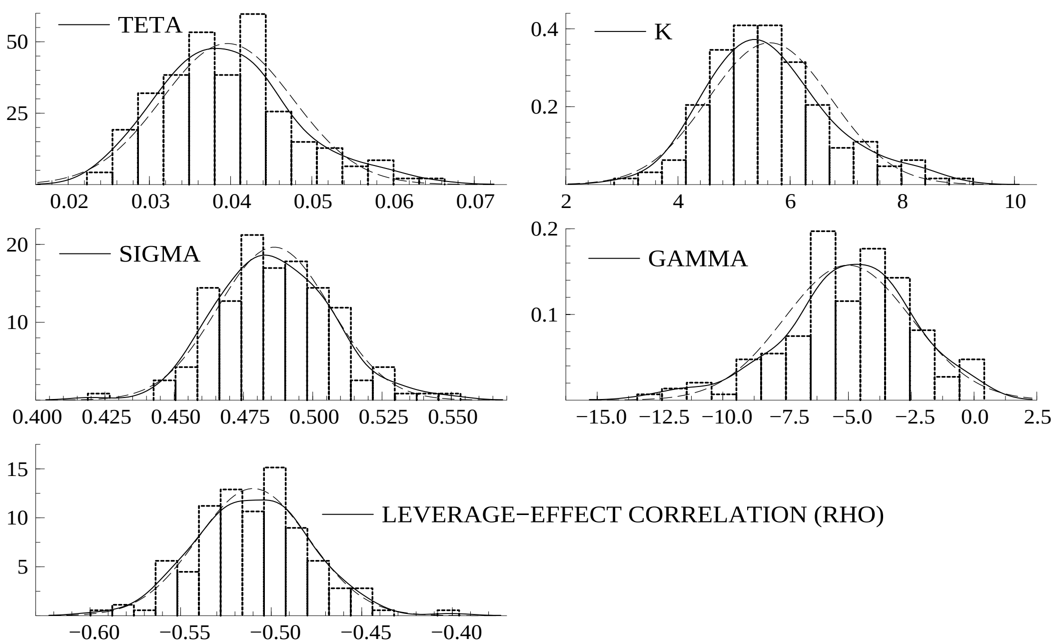


Figure 5.6: Histogram, estimated density, and associated normal density (with same mean and variance) of the 150 estimates (based on **squared return-option data**) of θ, k, σ, γ, ρ for set III. True values: $\theta = 0.04$, $k = 5.25$, $\sigma = 0.45$, $\gamma = -6$ and $\rho = -0.55$.

There is no additional bias induced by the leverage effect in the θ-estimates, a seemingly small extra bias in the σ-estimates, though some additional bias in the k and γ-estimates. (Compare table 5.9 to table 4.14 in the previous chapter for the 1-factor CIR SV model).

For illustration, figure 5.6 plots the distributions of the estimates for set III ($\theta = 0.04$, $k = 5.25$, $\sigma = 0.45$, $\gamma = -6$, $\rho = -0.55$). Using the 150 estimates of each parameter, the histogram, estimated continuous density, and a superimposed normal density are plotted. Compare figure 5.6 to figure 4.6 in the last chapter.

5.3.5 Results using RV and option data for estimation

Table 5.10 reports Heston model estimation results if both RV and short-maturity at-the-money option data is used for estimation.

Table 5.10: Estimation results Heston model based on **RV - option data**

	I	*II*	*III*	*IV*	*V*
θ	**0.039**	**0.039**	**0.040**	**0.038**	**0.020**
	(0.010)	(0.011)	(0.008)	(0.008)	(0.003)
Bias	-0.001	-0.001	0.000	-0.002	0.000
MSE	1.1e-4	1.2e-4	7.2e-5	6.5e-5	7.9e-6
k	**1.70**	**2.71**	**5.13**	**6.85**	**6.80**
	(0.46)	(0.70)	(1.06)	(1.33)	(1.03)
Bias	0.20	0.10	-0.12	-0.15	-0.20
MSE	0.25	0.49	1.14	1.80	1.10
σ	**0.218**	**0.323**	**0.472**	**0.621**	**0.311**
	(0.011)	(0.013)	(0.016)	(0.021)	(0.012)
Bias	0.018	0.023	0.022	0.021	0.011
MSE	4.4e-4	7.1e-4	7.1e-4	8.6e-4	2.6e-4
γ	**-3.7**	**-4.3**	**-4.8**	**-5.8**	**-17.9**
	(2.1)	(2.3)	(2.2)	(2.1)	(3.2)
Bias	0.3	0.7	1.2	1.2	2.1
MSE	4.3	5.6	6.4	6.0	14.4
ρ	**-0.35**	**-0.45**	**-0.51**	**-0.57**	**-0.46**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Bias	0.05	0.05	0.04	0.03	0.04
MSE	0.0035	0.0039	0.0025	0.0020	0.0021
RMSPE	**1.3**	**2.0**	**2.9**	**5.7**	**2.2**
%	(0.4)	(0.5)	(0.9)	(2.8)	(0.4)
MAPE	**1.0**	**1.4**	**1.8**	**2.6**	**1.6**
%	(0.2)	(0.2)	(0.3)	(0.5)	(0.5)

For each parameter set I-V, the results are summarized over the 150 simulated datasets jointly. The conditional state space model was estimated using **RV - option data** with $\mu_V = 0$ imposed. The average estimates of θ, k, σ, γ, ρ are reported (std.dev. in (.)), with associated sample bias and MSE. Based on (2.6), ρ is estimated as the correlation between returns and smoothed stock variance changes. The average RMSPE and MAPE, for comparing the smoothed and true volatilities, are also reported (std.dev. in (.)).

θ is accurately estimated, with only a small bias and MSE. There is a large bias reduction in the estimates of mean-reversion parameter k as opposed to using squared return – option data. The estimation bias in k seems small now. For the other data types we found over-estimation of k on average, here the bias is negative for sets III - V. Parameter σ is also more accurately estimated with only a small bias. For sets I and II, γ is more precisely estimated than in the return–

option data case, though for sets III - V it is the other way round. The positive bias and MSE in the correlation estimates ρ seems small and is similar as in the squared return - option data case. The combination RV – option data yields most favorable volatility evaluation criteria. Their magnitudes indicate that the state space model performs well in extracting the underlying latent volatilities.

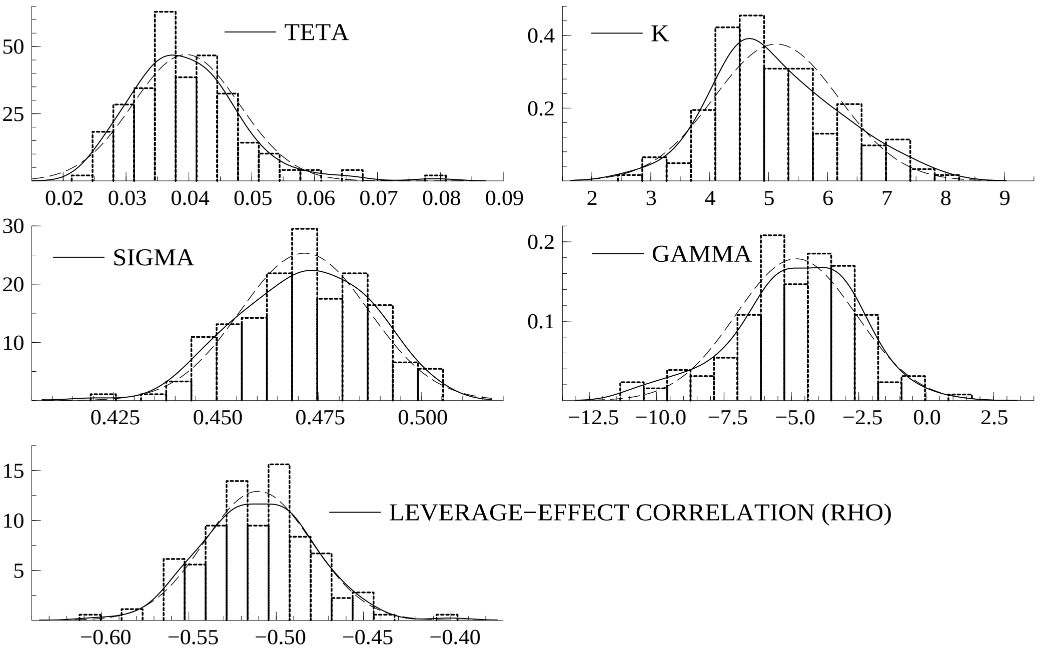


Figure 5.7: Histogram, estimated density, and associated normal density (with same mean and variance) of the 150 estimates (based on **RV - option data**) of $\theta, k, \sigma, \gamma, \rho$ for parameter set III. True values: $\theta = 0.04$, $k = 5.25$, $\sigma = 0.45$, $\gamma = -6$ and $\rho = -0.55$.

Figure 5.7 plots the histogram, estimated continuous density and associated normal density of the 150 estimates for set III, for which $\theta = 0.04$, $k = 5.25$, $\sigma = 0.45$, $\gamma = -6$, $\rho = -0.55$. Compare figure 5.7 to figure 4.8 in the last chapter.

We conclude that the combination RV – option data yields the best estimation results (also) for the Heston model. Moreover, the additional estimation bias induced by the leverage effect is particularly modest in this case. (Compare table 5.10 to table 4.16 in the previous chapter.) We attribute this to the only minor impact of the leverage effect on the prices of at-the-money options.

5.3.6 Results 2-step estimation method based on RV – option data

Recall the reason for restricting $\rho = 0$ prior to state space estimation of the Heston model: If ρ is left free, it is routinely estimated at zero. We attribute this to linearization point (4.9). We chose this specific linearization point because it leads to a linear options measurement equation.

As the Monte Carlo evidence has shown, the bias in the $\rho-$estimator based on (2.6) appears modest if option data is included for estimation. Restricting ρ to zero whereas its true value differs from zero may nonetheless result in bias in the other estimates. An intuitively appealing 2-step estimation method to deal with this possible bias is the following. In the first step, the conditional state space model is estimated with $\rho = 0$ imposed. At the optimum, an estimate of ρ is computed based on (2.6). (This procedure is followed in sections 5.3.1–5.3.5.) In the second step, this estimate of ρ is plugged into the options measurement equation (4.14), after which a second, restricted estimation round is performed, to determine the estimates of θ, k, σ and γ (while keeping ρ restricted).

Table 5.11: **2-step** estimation results Heston model based on **RV - option data**

	I	*II*	*III*	*IV*	*V*
θ	**0.039**	**0.038**	**0.037**	**0.035**	**0.019**
	(0.011)	(0.010)	(0.008)	(0.008)	(0.003)
Bias	-0.001	-0.002	-0.003	-0.005	-0.001
MSE	1.2e-4	1.1e-4	6.9e-5	8.6e-5	8.3e-6
k	**1.74**	**3.01**	**6.02**	**8.57**	**7.95**
	(0.53)	(0.81)	(1.34)	(1.87)	(1.31)
Bias	0.24	0.40	0.77	1.57	0.95
MSE	0.34	0.82	2.38	5.98	2.62
σ	**0.228**	**0.346**	**0.503**	**0.664**	**0.332**
	(0.012)	(0.016)	(0.020)	(0.027)	(0.014)
Bias	0.028	0.046	0.053	0.064	0.032
MSE	9.3e-4	0.002	0.003	0.005	0.001
γ	**-11.9**	**-14.8**	**-15.9**	**-17.3**	**-30.9**
	(3.2)	(3.4)	(3.2)	(3.4)	(4.1)
Bias	-7.9	-9.8	-9.9	-10.3	-10.9
MSE	72	106	109	118	137
ρ	**-0.34**	**-0.42**	**-0.48**	**-0.52**	**-0.44**
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
Bias	0.06	0.08	0.07	0.08	0.06
MSE	0.0045	0.0071	0.0061	0.0065	0.0040
RMSPE	**2.3**	**3.4**	**5.3**	**9.6**	**4.1**
%	(0.9)	(0.6)	(1.2)	(3.5)	(0.6)
MAPE	**1.8**	**2.7**	**3.9**	**5.9**	**3.3**
%	(0.5)	(0.4)	(0.6)	(1.0)	(0.4)

For each parameter set I-V, the estimation results are summarized over the 150 simulated datasets jointly. The state space model resulting from (4.1)-(4.4), (4.14)-(4.15) was estimated by the **2-step method** using **RV - option data**, with $\mu_V = 0$ imposed. The average *second-step* estimates of θ, k, σ, γ, ρ are reported (std.dev. in (.)), with associated sample bias and MSE. Based on (2.6), ρ is estimated as the correlation between returns and smoothed stock variance changes. The average RMSPE and MAPE, for comparing the smoothed and true volatilities, are also reported (std.dev. in (.)). *First-step* estimates are in table 5.10.

This 2-step method performs much worse than simply taking the estimates obtained in the first step as the final estimates. Table 5.11 illustrates this. It reports the 2-step estimation results in case both RV and option data is used for estimation. (NB: The ρ-estimates in table 5.11 are computed using (2.6) at the optimum of the *second* estimation round. This is driven by our curiosity to what extent the ρ-estimates obtained in the first and second round differ.) Comparing table 5.11 to table 5.10 shows the larger bias and MSE, and the worse volatility evaluation criteria. Especially notable is the dramatic (negative) increase in bias and MSE of the $\gamma-$estimates. This is perhaps not so surprising, given that our earlier results have shown that γ is hardest to pin down precisely. As such, its estimate seems most sensitive to picking up distortions.

We conclude that this 2-step estimation method should not be adopted. Instead, take the estimates obtained in the first step as the final estimates.

6. Summary

We found supporting evidence of the presence of the leverage effect in the FTSE100-index data in previous chapters. We ignored leverage in the multifactor affine SV model because it complicates matters greatly, mainly with regard to our estimation method (and especially if multiple SV factors are considered). Nonetheless, one might wonder if this misspecification has affected the empirical results in any major way. This chapter provides Monte Carlo evidence suggesting that the impact of having neglected leverage seems modest, given our exclusive focus on ATM options. We focus hereby on a 1-factor affine SV derivative pricing model with leverage effect, of which the Heston (1993) model is a special case.

The chapter starts with analyzing the model in a similar way as the multifactor no-leverage-effects affine SV model was analyzed in chapter III, and explains the differences induced by leverage. We show how to value and hedge derivatives, and consider expected asset returns. Both the hedging strategy and the expected return - beta relationship for derivatives have the same appearance as in case of no leverage (but the implied derivative prices differ for both models).

Our economic rationale for negative volatility risk compensation is explicitly based on the existence of a leverage effect. To summarize: Risk-averse investors who aim at consumption smoothing are willing to pay a premium for volatility risk. As dropping markets typically lead to less consumption opportunities, but simultaneously generally increase volatility (the leverage effect), holding delta-neutral positive-vega derivatives provides consumption insurance. Namely, such derivatives are sure to increase in value then, (partially) offsetting the drop in the investor's consumption opportunities, leading to a smoother consumption pattern.

To examine the effects of leverage on our estimation method (in the 1-factor SV case), we next derive a linear state space representation from the model. However, Monte Carlo evidence next directs us to still use the conditional state space model from the previous chapter (which is based on no leverage) to estimate the model with leverage effect. We estimate the correlation between unobserved volatility changes and stock returns by using the smoothed volatilities obtained at the optimum.

We next focus on the Heston model, for which we perform a Monte Carlo study to investigate the effects of leverage. Parameters θ, k, σ, γ, and ρ of the Heston model have the following interpretation regarding their impact on the volatility process, the stock return distribution, and option prices. Parameter θ determines the average stock volatility level, k the volatility mean reversion or persistence, σ the excess return kurtosis and volatility-of-volatility, and ρ the leverage-effects correlation and induces negative skewness in the stock returns. The Heston model without leverage effect ($\rho = 0$) yields a U-shaped implied volatility surface with the smile becoming less pronounced the longer the option maturity. The Heston model with leverage effect ($\rho < 0$) results in a skew-shaped surface. ρ controls the skew: the stronger the leverage effect, the skewer the skews.

Based on simulation, the effects of the other parameters on the shape of the implied volatility surface may roughly be summarized as follows. Parameter γ largely determines the average $\mathbb{Q}$-volatility, and as such, the general level of the surface. Larger negative values result in higher surface levels. Parameter σ largely determines the extent in which the surface smiles or smirks. Larger values imply stronger smiles and skews. Parameter k also determines the extent of smileyness or smirkiness, and impacts on the slope of the surface. Low values result in steep-sloping, bent surfaces, whereas large values typically imply a more flat surface (except for short maturities).

Another important VTS slope-determining factor is the current ($\mathbb{P}$) stock volatility $\sigma_t^2 = x_t$. The more it deviates from the risk-neutral mean volatility $\tilde{\theta}$, the steeper the term structures tend to be. If the market price of volatility risk is negative, the at-the-money VTS is upward sloping on a medium volatility day. If volatility risk is not priced, the ATM VTS is more or less flat.

The ATM VTS is not much affected by the leverage effect, as simulation shows. In contrast, the VTS of OTM and ITM options is strongly affected by leverage.

We also perform a Monte Carlo study in which we simulate squared return, RV and SM ATM option data from the Heston model (hence with leverage), and next use the conditional state space model (which assumes no leverage) for estimation. We use the same parameter sets I-V as in the Monte Carlo study for the 1-factor CIR SV model (the Heston model without leverage) in the previous chapter. As such, this yields insight in the additional bias induced by having neglected the leverage effect. As found in the other Monte Carlo studies, also for estimating the Heston model using squared return data only performs worst, and the combination RV - option data best. Overall speaking, the additional bias due to the leverage effect seems rather modest. This holds in particular for the combination RV – option data. We attribute this to the leverage effect not being really important for pricing ATM options. Our proposed estimator for ρ performs well if option data is included for estimation; it has a small positive bias only.

The Monte Carlo evidence reported in this chapter thus strengthens and underwrites our motivation for having focused exclusively on ATM options in the empirical analysis of the FTSE100-index market. It at least suggests that the empirical results seem not very much distorted by having neglected the leverage effect.

Appendix

VIa. Heston call price formula

Below we state the Heston (1993) call price formula for a European call option written on a stock S having strike K and maturity T, assuming a constant risk-free interest rate r (such that $dB_t = rB_t dt$), and no dividend payments (such that $q_t = 0 \ \forall t$).

The Heston model is stated under the risk-neutral measure as

$$\begin{aligned} dS_t &= rS_t dt + \sigma_t S_t d\bar{W}_{S,t}^{\mathbb{Q}} \qquad (\mathbb{Q}) \qquad \text{(a.1)} \\ \sigma_t^2 &= x_t \\ dx_t &= \tilde{k}(\tilde{\theta} - x_t)dt + \sigma\sqrt{x_t}\, d\bar{W}_{x,t}^{\mathbb{Q}}, \end{aligned}$$

in which the $\mathbb{Q}$-Brownian motions $\bar{W}_S^{\mathbb{Q}}$ and $\bar{W}_x^{\mathbb{Q}}$ are correlated with correlation coefficient $\rho = \mathrm{corr}_{\mathbb{Q}}[\bar{W}_{S,t}^{\mathbb{Q}}, \bar{W}_{x,t}^{\mathbb{Q}}]$.

Heston (1993) shows that the time-t price of the call option is given by

$$C_t = S_t\,\Pi_1(t,S_t,x_t) - K\exp(-r\tau_t)\,\Pi_2(t,S_t,x_t)\,, \tag{a.2}$$

in which $\tau_t = T - t$ is the current remaining time to maturity, and $\Pi_j(t,S_t,x_t);\; j = 1,2$ are functions of t, S_t and x_t. These functions are given by

$$\Pi_j(t,S_t,x_t) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\,\frac{\exp[-i\phi\ln K]\; f_j(t,S_t,x_t;\phi)}{i\phi}\,d\phi\,, \tag{a.3}$$

with $i \equiv \sqrt{-1}$, and Re(.) stands for the real part of a complex number, and

$$\begin{aligned} f_1(t,S_t,x_t;\phi) \equiv \exp\Bigg\{ & i\phi\, r\tau_t - \frac{\tilde{k}\tilde{\theta}}{\sigma^2}\left[2\ln\left(1 - \frac{\left[\xi - \tilde{k} + (i\phi+1)\rho\sigma\right]\left(1-\exp(-\xi\tau_t)\right)}{2\xi}\right)\right] \\ & - \frac{\tilde{k}\tilde{\theta}}{\sigma^2}\left[\xi - \tilde{k} + (i\phi+1)\rho\sigma\right]\tau_t + i\phi\ln S_t \\ & + \frac{i\phi(i\phi+1)(1-\exp(-\xi\tau_t)x_t}{2\xi - \left[\xi - \tilde{k} + (i\phi+1)\rho\sigma\right]\left(1-\exp(-\xi\tau_t)\right)}\Bigg\}, \end{aligned}$$

in which $\xi \equiv \sqrt{\left[\tilde{k} - (i\phi+1)\rho\sigma\right]^2 - i\phi(i\phi+1)\sigma^2}$, and

$$\begin{aligned} f_2(t,S_t,x_t;\phi) \equiv \exp\Bigg\{ & i\phi\, r\tau_t - \frac{\tilde{k}\tilde{\theta}}{\sigma^2}\left[2\ln\left(1 - \frac{\left[\psi - \tilde{k} + i\phi\rho\sigma\right]\left(1-\exp(-\psi\tau_t)\right)}{2\psi}\right)\right] \\ & - \frac{\tilde{k}\tilde{\theta}}{\sigma^2}\left[\psi - \tilde{k} + i\phi\rho\sigma\right]\tau_t + i\phi\ln S_t \\ & + \frac{i\phi(i\phi-1)(1-\exp(-\psi\tau_t)x_t}{2\psi - \left[\psi - \tilde{k} + i\phi\rho\sigma\right]\left(1-\exp(-\psi\tau_t)\right)}\Bigg\}, \end{aligned}$$

in which $\psi \equiv \sqrt{\left[\tilde{k} - i\phi\rho\sigma\right]^2 - i\phi(i\phi-1)\sigma^2}$.

VIb. Derivative betas

This appendix derives expressions for the stock beta $\beta_{F,S}$ and volatility beta $\beta_{F,x}$ of a general, path-independent European-style derivative F for which $F_t = F(t,S_t,x_t)$, in the 1-factor affine SV derivative pricing model with leverage effect.

Betas $\beta_{F,S}$ and $\beta_{F,x}$ are collected in vector $\boldsymbol{\beta}_F$, which we define as:

$$\begin{aligned} \boldsymbol{\beta}_{F,t} = \begin{bmatrix} \beta_{F,S,t} \\ \beta_{F,x,t} \end{bmatrix} & \equiv \left(\mathrm{var}_{\mathbb{P}}\left[\begin{pmatrix} \frac{dS_t}{S_t} \\ dx_t \end{pmatrix} \mid \mathcal{F}_t\right]\right)^{-1} \mathrm{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, \begin{pmatrix} \frac{dS_t}{S_t} \\ dx_t \end{pmatrix} \mid \mathcal{F}_t\right] \\ & = \begin{pmatrix} \mathrm{var}_{\mathbb{P}}\left[\frac{dS_t}{S_t} \mid \mathcal{F}_t\right] & \mathrm{cov}_{\mathbb{P}}\left[\frac{dS_t}{S_t}, dx_t \mid \mathcal{F}_t\right] \\ \mathrm{cov}_{\mathbb{P}}\left[\frac{dS_t}{S_t}, dx_t \mid \mathcal{F}_t\right] & \mathrm{var}_{\mathbb{P}}\left[dx_t \mid \mathcal{F}_t\right] \end{pmatrix}^{-1} \begin{bmatrix} \mathrm{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, \frac{dS_t}{S_t} \mid \mathcal{F}_t\right] \\ \mathrm{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, dx_t \mid \mathcal{F}_t\right] \end{bmatrix}. \end{aligned} \tag{b.1}$$

We next compute each variance and covariance appearing in (b.1) separately.

From (2.3), it follows $\text{var}_{\mathbb{P}}[dS_t / S_t \mid \mathcal{F}_t] = \text{var}_{\mathbb{P}}[\mu_t + \sigma_t d\bar{W}_{S,t} \mid \mathcal{F}_t] = \sigma_t^2 dt$. From (2.5), $\text{var}_{\mathbb{P}}[dx_t \mid \mathcal{F}_t] = \sigma^2 \lambda_t^2 dt$. Using $\text{cov}_{\mathbb{P}}[d\bar{W}_{x,t}, d\bar{W}_{S,t} \mid \mathcal{F}_t] = \rho dt$, we obtain from (2.3) and (2.5) $\text{cov}_{\mathbb{P}}[\frac{dS_t}{S_t}, dx_t \mid \mathcal{F}_t] = \text{cov}_{\mathbb{P}}[\sigma_t d\bar{W}_{S,t}, \sigma\lambda_t d\bar{W}_{x,t} \mid \mathcal{F}_t] = \rho\sigma\sigma_t\lambda_t dt$. As such,

$$\left(\text{var}_{\mathbb{P}}\left[\begin{pmatrix} \frac{dS_t}{S_t} \\ dx_t \end{pmatrix} \mid \mathcal{F}_t\right]\right)^{-1} = \left(\begin{bmatrix} \sigma_t^2 & \rho\sigma\sigma_t\lambda_t \\ \rho\sigma\sigma_t\lambda_t & \sigma^2\lambda_t^2 \end{bmatrix} dt\right)^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} \frac{1}{\sigma_t^2} & \frac{-\rho}{\sigma\sigma_t\lambda_t} \\ \frac{-\rho}{\sigma\sigma_t\lambda_t} & \frac{1}{\sigma^2\lambda_t^2} \end{bmatrix}\frac{1}{dt}. \quad \text{(b.2)`}$$

From (2.3) and (3.5), the covariance between the derivative and stock returns equals:

$$\begin{aligned} \text{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, \frac{dS_t}{S_t} \mid \mathcal{F}_t\right] &= \text{cov}_{\mathbb{P}}\left[\left(\frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t}\right)\sigma_t d\bar{W}_{S,t} + \left(\frac{1}{F_t}\frac{\partial F_t}{\partial x_t}\right)\sigma\lambda_t d\bar{W}_{x,t}, \sigma_t d\bar{W}_{S,t} \mid \mathcal{F}_t\right] \\ &= \left(\frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t}\right)\sigma_t^2 \,\text{var}_{\mathbb{P}}\left[d\bar{W}_{S,t}\right] + \left(\frac{1}{F_t}\frac{\partial F_t}{\partial x_t}\right)\sigma\sigma_t\lambda_t \,\text{cov}_{\mathbb{P}}\left[d\bar{W}_{x,t}, d\bar{W}_{S,t}\right] \\ &= \left[\left(\frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t}\right)\sigma_t^2 + \left(\frac{1}{F_t}\frac{\partial F_t}{\partial x_t}\right)\rho\sigma\sigma_t\lambda_t\right] dt. \qquad \text{(b.3)} \end{aligned}$$

From (2.5) and (3.5), the covariance between the derivative return and stock variance change (recall $\sigma_t^2 = x_t$) equals:

$$\begin{aligned} \text{cov}_{\mathbb{P}}\left[\frac{dF_t}{F_t}, dx_t \mid \mathcal{F}_t\right] &= \text{cov}_{\mathbb{P}}\left[\left(\frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t}\right)\sigma_t d\bar{W}_{S,t} + \left(\frac{1}{F_t}\frac{\partial F_t}{\partial x_t}\right)\sigma\lambda_t d\bar{W}_{x,t},\ \sigma\lambda_t d\bar{W}_{x,t} \mid \mathcal{F}_t\right] \\ &= \left(\frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t}\right)\sigma\sigma_t\lambda_t \,\text{cov}_{\mathbb{P}}\left[d\bar{W}_{S,t}, d\bar{W}_{x,t}\right] + \left(\frac{1}{F_t}\frac{\partial F_t}{\partial x_t}\right)\sigma^2\lambda_t^2 \,\text{var}_{\mathbb{P}}\left[d\bar{W}_{x,t}\right] \\ &= \left[\left(\frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t}\right)\rho\sigma\sigma_t\lambda_t + \left(\frac{1}{F_t}\frac{\partial F_t}{\partial x_t}\right)\sigma^2\lambda_t^2\right] dt. \qquad \text{(b.4)} \end{aligned}$$

Straightforward matrix multiplication and some further manipulations then finally yield for the stock and volatility beta of derivative F:

$$\boldsymbol{\beta}_{F,t} = \begin{bmatrix} \beta_{F,S,t} \\ \beta_{F,x,t} \end{bmatrix} = \begin{bmatrix} \frac{S_t}{F_t}\frac{\partial F_t}{\partial S_t} \\ \frac{1}{F_t}\frac{\partial F_t}{\partial x_t} \end{bmatrix}. \qquad \text{(b.5)}$$

- VII -

Summary and Directions for Future Research

This thesis considers derivative pricing under stochastic volatility. Driven by recent empirical evidence, we assume the volatility of the underlying stock to be driven by a multifactor SV specification. Specifically, we model the stock volatility as an affine function of an arbitrary number of unobserved affine factors, which follow mean-reverting Markov diffusions. We label the model *the multifactor affine stochastic volatility derivative pricing model*. Investors are exposed to two sources of risk in this model, stock price fluctuations and volatility changes. We explain how the model can be used for pricing and hedging stock and foreign-exchange derivatives.

We develop an (Extended) Kalman filter-based QML estimation method for the multifactor SV model. Our method is transparent, circumvents simulation of option prices during estimation, generates a volatility forecast and is above all fast. We show how to include time series of squared returns, realized volatilities and option prices for estimation, and combinations hereof.

Monte Carlo studies for a number of subcases of our general SV model consistently show that if ATM option data is included for estimation, our method performs seemingly well. Option data appears very informative. More specific, using only squared return data for QML estimation is generally not to be advocated. The estimation bias and MSEs are large, and the underlying latent volatility is not very well filtered out. Using RV data instead yields considerably better results, though the bias and MSEs are still large. Option data means a dramatic improvement to using only RV data for estimation. The bias and MSE of the estimates are dramatically less, and the volatility is much better extracted. Combining squared return and option data for estimation outperforms the use of option data only in turn. The bias seems modest in general. There are efficiency gains from including squared returns in addition to option data. The combination RV - option data contains the most precise information: The estimation bias is typically small, it yields the most efficient estimates and the extracted and true volatilities are generally close. As such, combining RV and option data for estimation is preferable. The simulation results further show that the market price of volatility risk is hard to pin down precisely, irrespective of the data used for estimation. This confirms findings by other researchers in the literature.

We study the nature of volatility risk. We consider investment strategies that are especially prone to volatility risk, such as straddles and variance swaps. We investigate how investors are compensated for the risk of derivative price changes due to uncertain volatility fluctuations. We derive an expected return – beta relationship for derivatives and define the volatility risk premium. Our empirical work shows that investors appear willing to pay for market volatility risk. This is in line with other recent evidence in the literature, and seems counterintuitive at first sight.

Our economic rationale for the existence of a negative volatility risk premium is summarized as follows. According to the permanent-income hypothesis, risk-averse investors aim at consumption smoothing. As such, they are willing to pay a premium for volatility risk: As dropping markets typically lead to less consumption opportunities, but simultaneously typically increase volatility (the leverage effect), holding delta-neutral positive-vega derivatives provides consumption insurance. Such derivatives are certain to increase in value then, (partially) offsetting the drop in the investor's consumption opportunities, which leads to a smoother consumption pattern. As the return on such derivatives is likely to negatively correlate with the investor's change in consumption (due to the leverage effect), this translates into a negative volatility risk premium.

Our empirical work confronts the multifactor affine SV derivative pricing model to joint FTSE100 stock-index and option data for Oct 1997 – Dec 2001. These years were characterized by three periods of unusual turmoil in financial markets: the Asian crisis (fall 1997), the near collapse of LTCM and the Soviet Union (and continuing Asian) crisis (fall 1998), and September 11, 2001 with its aftermath. The data consists of time series of squared index returns and three ATM index-option series of different maturities, a short, medium and long-maturity series.

We initially assume the FTSE100-index data to have been generated by the 1-factor OU and CIR SV special cases of our model. Fitting the model to only the squared return and SM option data reveals the following. The observed BS implied volatilities are typically larger than the underlying hidden volatilities, due to a negative market price of volatility risk. The common market interpretation of BS implied volatilities being forecasts of future volatility should therefore be taken with care if volatility risk is priced. Investing in FTSE100-index derivatives yields a negative volatility risk premium of approximately –17.5% per annum. Writing SM ATM straddles yields a substantial expected return of about 185%, though involves considerable risk. Straddles are true bets on volatility. The volatility factor features level-dependent volatility. The 1-factor CIR SV specification (Heston volatility) does not allow for sufficiently fast-changing volatility however, during the three periods of increased market stress mentioned above. The misspecification of the 1-factor CIR SV model estimated with squared return – SM option data seems modest at first sight. However, the estimated model severely overprices the longer-maturity options out of sample. Estimating the model using the squared return, SM, MM and LM option data jointly unveils the cause of this finding: 1-factor SV option pricing models cannot sufficiently adequately describe the rich volatility dynamics present in the joint data.

A subsequent search for the multifactor SV option pricing model within the affine class that best fits the data reveals the following. A considerable improvement in fit is obtained when 2-factor SV specifications are considered. Both factors feature volatility feedback. Dynamic volatility misspecification still remains however.

A 3-factor affine SV specification with one CIR and two affine, independent volatility factors appears to fit the joint FTSE100-index data best. Each of the three factors thus features level-dependent volatility. The volatility dynamics present in the data is adequately accounted for and the fit of the VTS is good. The option pricing errors concentrate near zero and are typically small. The 3-factor SV volatilities respond quicker to news than the 1-factor SV volatilities, especially in times of sudden market stress.

The three volatility factors x_1, x_2 and x_3 differ greatly in their characteristics. Factor x_1 is extremely persistent and shows near random walk behavior. x_1 fluctuates around the long-term mean stock volatility and determines the long-term volatility trend. Factor x_2 is much faster mean reverting. Shocks to x_2 have a half-life of approximately three months. x_2 determines the medium-term volatility trend. Factor x_3 is very rapidly mean reverting. Shocks to x_3 lose half their impact in about ten days and have largest variance. x_3 governs large volatility changes in relatively short periods of time. Factor x_3 is associated with short-term volatility trends.

Volatility factors x_1, x_2 and x_3 impact on the prices of options of different maturity in different ways. A shock to long-memory factor x_1 impacts on all option prices similarly, irrespective of maturity. Medium-term volatility trend factor x_2 affects all option prices as well, but gradually by smaller amounts the longer the option's lifetime. Shocks to fast mean-reverting factor x_3 mainly affect the prices of short-maturity options. The impact of x_3 quickly diminishes as the option maturity increases. As x_3 is so rapidly mean reverting, its shocks tend to average out over a sufficiently long time span, resulting in an only marginal impact on the prices of long-dated options.

Factors x_1, x_2 and x_3 impact differently on the shape and evolution of the volatility term structure over time. x_1 is mainly responsible for changes in the general level of the VTS. x_2 is largely associated with changes in the slope. x_3 is closely related to dynamic changes in the curvature of the VTS.

The market prices of risk associated with the SV factors are estimated imprecisely. We lack an intuition why the price of risk of medium-term volatility trend factor x_2 is estimated positive; both others are negative. The risk premium associated with x_2 equals 3.9% per annum. The risk premium on fast mean-reverting factor x_3 is much more negative (–22% per annum) than the premium on long-memory factor x_1 (–0.4%). These premia imply a much more negative expected return to be earned on short-maturity ATM straddles than on long-maturity ATM straddles, which agrees with short-maturity straddles being riskier. The finding of negative expected (long) straddle returns agrees with investors being negatively compensated for volatility risk.

Directions for future research

Within the affine class of SV models, we have found the 3-factor affine SV derivative pricing model with one CIR and two affine volatility factors to be most appropriate for the joint FTSE100 stock-index and option data. As diagnostic checking has revealed however, there is still room left for improvement. (See section 7.2.1. of chapter V for a more complete discussion.)

First, we did not model the leverage effect. Given our focus on ATM options only, having neglected leverage is unlikely to invalidate most of our findings: The ATM

VTS is not much affected by leverage as we have shown in chapter VI. However, if the model is to be used for pricing of OTM and ITM options, the leverage effect cannot be neglected. As such, the multifactor affine SV derivative pricing model with associated estimation method may be extended to allow for leverage.

Second, the 3-factor affine SV model cannot fully capture the three periods of rapidly changing volatility present in the FTSE100-index data, not even with three volatility factors. Extending the affine SV specification with jumps in volatility seems promising in this respect. At this stage however, it is unclear to us how to incorporate the modeling of jumps in volatility (as well as in returns) in a state space estimation strategy.

Other directions for future research include the following.

As recent evidence has shown, empirical (short-maturity) volatility skews are typically steeper than can be described by diffusive SV or SV with jumps in returns. Jumps in volatility further steepen the skews. As such, much more remains to be learned about how exactly investors are compensated for volatility-factor risk, jumps-in-volatility risk and jumps-in-returns risk. (We refer to section 7.2.1. of chapter V for a discussion.) Without doubt it is econometrically challenging to estimate such a fully specified multifactor SV option pricing model that includes jumps in returns and volatility, based on multiple time series of stock and option prices. From an economic perspective more important however, a problem likely to be encountered is imprecision in the estimates of the prices of risk. This makes it difficult to disentangle for which sources of risk investors are exactly compensated for and for which they are (possibly) not. Relatedly, the imprecision complicates a precise numerical interpretation of the results.

Given the Monte Carlo evidence, an interesting extension is to include RV data in the analysis of the FTSE100-index (or any other) market, and to compare the estimation results to those obtained from including squared return data instead.

As a number of papers has recently shown, direct modeling of time series of realized volatilities using ARMA-type models yields much better volatility forecasts than GARCH and SV stock price models (see section 3.3.2. of chapter III). Given our simulation evidence, it would be interesting to compare state space, RV – option data-based volatility forecasts to forecasts obtained from direct modeling of RV, and to investigate if the latter ones still yield superior forecasts.

The hedging strategies outlined in this thesis assume continuous rebalancing of the hedging portfolio and no transaction costs. In that case the strategies replicate the derivative payoff perfectly, as desired. These assumptions are violated in practice. It may be of interest to examine the hedging error incurred from only periodically rebalancing of the hedging portfolio by simulation.

Our multifactor affine SV derivative pricing model and associated state space estimation method seem promising in a foreign-exchange markets context. Section 8 of chapter II discusses this extension. No further mathematical analysis is required for casting the model in a FX option pricing and hedging framework. A main advantage is that the assumption of a zero correlation between FX returns and FX volatility shocks (i.e. no "leverage effect") seems adequate. The Monte Carlo evidence reported in chapter V for the 1-factor CIR SV model (which assumes no leverage) is suggestive for how well our state space estimation method will perform in a FX setting.

Combining time series of squared FX returns or realized FX volatilities with FX options data for estimation yields insight in how investors are compensated for exchange-rate volatility risk (if at all). Our rationale for negative stock volatility-risk compensation is explicitly based on the existence of the leverage effect. If a zero correlation between FX returns and FX volatility shocks is indeed reasonable, how are domestic and foreign-country investors then compensated for FX volatility risk? If FX volatility risk is systematic, both types of investors ought to be compensated for bearing this risk at least to some extent.

As this incomplete list already indicates, there are numerous interesting directions for future research based on our multifactor affine SV derivative pricing model and associated state space estimation strategy.

- Appendix A -

Matrix definitions

This appendix serves as a convenient reference. It collects matrix definitions associated with the statistical properties of the factors $(nx1)\,\mathbf{x}$ which are derived in appendix B. It also states a lemma and derives some intermediate results which are often invoked in appendix B.

Definition: (exponent of a diagonal matrix)
The exponent of a square diagonal matrix $(nxn)\,\mathbf{A}_d = \text{diag}\left[a_1,\ldots,a_n\right]$ is defined as $\exp\left[\mathbf{A}_d\right] \equiv \text{diag}\left[\exp(a_1),\ldots,\exp(a_n)\right]$.

Lemma:
Let $(nxn)\mathbf{D}_d = \text{diag}\left[d_1,..,d_n\right]$ and $(nxn)\mathbf{F}_d = \text{diag}\left[f_1,..,f_n\right]$ be square diagonal matrices. Collect their diagonal elements in the vectors $\mathbf{d} = (d_1,..,d_n)' = \mathbf{D}_d\mathbf{1}$ and $\mathbf{f} = (f_1,..,f_n)' = \mathbf{F}_d\mathbf{1}$ respectively, in which $(nx1)\,\mathbf{1}$ is a vector of ones. Let $(nxn)\mathbf{A}$ be a general square matrix. Straightforward matrix multiplication shows that $\mathbf{D}_d\mathbf{A}\mathbf{F}_d = \mathbf{d}\mathbf{f}'\odot\mathbf{A} = \mathbf{D}_d\mathbf{1}\mathbf{1}'\mathbf{F}_d\odot\mathbf{A}$, in which $\odot$ denotes the Hadamard product. [1]

Matrix definitions:

$(nxn)\mathbf{K}_d = \text{diag}\left[k_1,..,k_n\right]$

$(nxn)\,\mathbf{J}$ with $\left[\mathbf{J}\right]_{ij} = \dfrac{1}{k_i + k_j}$

$(nxn)\mathbf{M}_d = \text{diag}\left[\alpha_1 + \boldsymbol{\beta}_1'\boldsymbol{\theta},..,\alpha_n + \boldsymbol{\beta}_n'\boldsymbol{\theta}\right]$

$(nx1)\,\boldsymbol{\alpha} = (\alpha_1,..,\alpha_n)'$

$(nxn)\mathcal{B} = \left[\boldsymbol{\beta}_1,..,\boldsymbol{\beta}_n\right]$

$(nxn)\boldsymbol{\Gamma} = \text{diag}\left[\gamma_1,..,\gamma_n\right]$

$(nxn)\,\boldsymbol{\Lambda}_t = \text{diag}\left[\sqrt{\alpha_1 + \boldsymbol{\beta}_1'\mathbf{x}_t},\ldots,\sqrt{\alpha_n + \boldsymbol{\beta}_n'\mathbf{x}_t}\right]$

$(nxn)\;\tilde{\mathbf{K}} = \mathbf{K}_d + \boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathcal{B}'$

$(nx1)\;\tilde{\boldsymbol{\theta}} = \tilde{\mathbf{K}}^{-1}\left(\mathbf{K}_d\boldsymbol{\theta} - \boldsymbol{\Sigma}\boldsymbol{\Gamma}\boldsymbol{\alpha}\right)$

$(nxn)\boldsymbol{\Sigma} = \left[\boldsymbol{\sigma}_1,..,\boldsymbol{\sigma}_n\right]$ (column partition).

Scalar, vector and matrix functions
Below we define a list of scalar, vector and matrix functions. These functions appear in the matrix integral manipulations in appendix B.

Definitions: $c_i(.,.)$, $(nx1)$ $\mathbf{c}(.,.)$, (nxn) $\mathbf{C}_d(.,.)$

$$c_i(p,q) \equiv \exp\left[-k_i(q-p)\right],\;\; i = 1,..,n.$$

[1] For matrices $(mxn)\mathbf{A}$ and $(mxn)\mathbf{B}$, the Hadamard product of $\mathbf{A}$ and $\mathbf{B}$ is defined as the (mxn) matrix $\mathbf{A}\odot\mathbf{B}$ which *ij*-th element equals the product of the *ij*-th elements of $\mathbf{A}$ and $\mathbf{B}$. That is, element-by-element multiplication: $[\mathbf{A}\odot\mathbf{B}]_{ij} = [\mathbf{A}]_{ij}\,[\mathbf{B}]_{ij}$.

$$\mathbf{C}_d(p,q) \equiv \exp\left[-\mathbf{K}_d(q-p)\right] = \text{diag}\left[c_1(p,q),..,c_n(p,q)\right]$$

$$\mathbf{c}(p,q) \equiv \left[c_1(p,q),..,c_n(p,q)\right]' = \mathbf{C}_d(p,q)\mathbf{1}.$$

Definition: (nxn) $\mathbf{G}(.,.)$

$$\mathbf{G}(p,q) \equiv \int_p^q \exp\left[-\mathbf{K}_d(q-u)\right]\mathbf{1}\mathbf{1}'\exp\left[-\mathbf{K}_d(q-u)\right]du = \int_p^q \mathbf{c}(u,q)\mathbf{c}(u,q)'du.$$

The ij-th element of $\mathbf{G}(.,.)$ equals

$$\begin{aligned}
\left[\mathbf{G}(p,q)\right]_{ij} &= \int_p^q \exp\left[-(k_i+k_j)(q-u)\right]du \\
&= \frac{1-\exp\left[-(k_i+k_j)(q-p)\right]}{k_i+k_j} \\
&= \left[\mathbf{J}\right]_{ij} \odot \left[\mathbf{1}\mathbf{1}' - \exp\left[-\mathbf{K}_d(q-p)\right]\mathbf{1}\mathbf{1}'\exp\left[-\mathbf{K}_d(q-p)\right]\right]_{ij} \\
&= \left[\mathbf{J}\right]_{ij} \odot \left[\mathbf{1}\mathbf{1}' - \mathbf{c}(p,q)\mathbf{c}(p,q)'\right]_{ij},
\end{aligned}$$

which reveals that $(nxn)\,\mathbf{G}(p,q)$ can be written as

$$\begin{aligned}
\mathbf{G}(p,q) &= \mathbf{J} \odot \left[\mathbf{1}\mathbf{1}' - \exp\left[-\mathbf{K}_d(q-p)\right]\mathbf{1}\mathbf{1}'\exp\left[-\mathbf{K}_d(q-p)\right]\right] \\
&= \mathbf{J} \odot \left[\mathbf{1}\mathbf{1}' - \mathbf{c}(p,q)\mathbf{c}(p,q)'\right].
\end{aligned}$$

The following holds

$$\mathbf{G}(p,q) = \mathbf{G}(0,q-p),$$

such that $(nxn)\,\mathbf{G}(p,q)$ is a function of $q-p$ and not of p or q separately. This observation yields another result which turns out to be convenient in the derivations in appendix B:

$$\begin{aligned}
\mathbf{G}(p,q) &= \mathbf{G}(0,q-p) \\
&= \int_0^{q-p} \exp\left[-\mathbf{K}_d(q-p-u)\right]\mathbf{1}\mathbf{1}'\exp\left[-\mathbf{K}_d(q-p-u)\right]du \\
&= -\int_{q-p}^{0} \exp[-\mathbf{K}_d v]\mathbf{1}\mathbf{1}'\exp[-\mathbf{K}_d v]dv \\
&= \int_0^{q-p} \exp[-\mathbf{K}_d v]\mathbf{1}\mathbf{1}'\exp[-\mathbf{K}_d v]dv,
\end{aligned}$$

in which the third equality follows from the change of variable towards $v \equiv q-p-u$, such that $dv = -du$, and u running from 0 to $q-p$ implies that v is running from $q-p$ to 0.

As $\mathbf{G}(p,q)$ is a function of $q-p$ only, and not of p and q separately, we sometimes write the following short cut:

$$\mathbf{G}(q) \equiv \mathbf{G}(0,q) = \int_0^q \exp\left[-\mathbf{K}_d v\right]\mathbf{1}\mathbf{1}'\exp\left[-\mathbf{K}_d v\right]dv = \mathbf{J} \odot \left[\mathbf{1}\mathbf{1}' - \exp\left[-\mathbf{K}_d q\right]\mathbf{1}\mathbf{1}'\exp\left[-\mathbf{K}_d q\right]\right]$$

Definitions: $d_i(.,.)$, $(nx1)\mathbf{d}$, (nxn) $\mathbf{D}(.,.)$

$$d_i(p,q) \equiv \int_p^q \exp\left[-k_i(u-p)\right]du = \frac{1-\exp\left[-k_i(q-p)\right]}{k_i}, \quad i = 1,..,n$$

$$\mathbf{D}(p,q) \equiv \text{diag}\left[d_1(p,q),..,d_n(p,q)\right] = \int_p^q \exp\left[-\mathbf{K}_d(u-p)\right]du = \mathbf{K}_d^{-1}\left(\mathbf{I}_n - \exp\left[-\mathbf{K}_d(q-p)\right]\right)$$

$$\mathbf{d}(p,q) \equiv \left[d_1(p,q),..,d_n(p,q)\right]' = \mathbf{D}(p,q)\mathbf{1}.$$

The following holds:

$$d_i(p,q) = d_i(0,q-p), \quad i = 1,..,n$$

$$\mathbf{d}(p,q) = \mathbf{d}(0,q-p)$$

$$\mathbf{D}(p,q) = \mathbf{D}(0,q-p).$$

As $d_i(p,q)$, $\mathbf{d}(p,q)$, $\mathbf{D}(p,q)$ are functions of $q-p$ only, and not of p and q separately, we sometimes write the following short cuts:

$$d_i(q) \equiv d_i(0,q) = \int_0^q \exp\left[-k_i u\right]du = \frac{1-\exp(-k_i q)}{k_i}; \quad i = 1,..,n$$

$$\mathbf{d}(q) \equiv \mathbf{d}(0,q) = \left[d_1(q),..,d_n(q)\right]' = \mathbf{D}(0,q)\mathbf{1}$$

$$\mathbf{D}(q) \equiv \mathbf{D}(0,q) = \text{diag}\left[d_1(q),..,d_n(q)\right] = \int_0^q \exp\left[-\mathbf{K}_d u\right]du = \mathbf{K}_d^{-1}\left(\mathbf{I}_n - \exp\left[-\mathbf{K}_d q\right]\right).$$

Definition: (nxn) $\mathbf{N}(.,.)$

$$\mathbf{N}(p,q) \equiv \int_p^q \mathbf{d}(u,q)\mathbf{d}(u,q)'du.$$

$\mathbf{N}(p,q)$ can be rewritten as:

$$\mathbf{N}(p,q) \equiv \int_p^q \mathbf{d}(u,q)\mathbf{d}(u,q)'du$$

$$= \int_p^q \mathbf{d}(0,q-u)\mathbf{d}(0,q-u)'du$$

$$= -\int_{q-p}^{0} \mathbf{d}(0,v)\mathbf{d}(0,v)'dv = \int_{0}^{q-p} \mathbf{d}(0,v)\mathbf{d}(0,v)'dv = \int_{0}^{q-p} \mathbf{d}(v)\mathbf{d}(v)'dv ,$$

in which the third equality follows from a change of variable towards the variable v defined as $v \equiv q - u$, such that $dv = -du$, and as u runs from p to q, then v runs from $q - p$ to 0. The latter expression reveals

$$\mathbf{N}(p,q) = \mathbf{N}(0,q-p) ,$$

such that $\mathbf{N}(.,.)$ is a function of $q - p$ only, and not of p and q separately. We therefore write the following short cut

$$\mathbf{N}(q) \equiv \mathbf{N}(0,q) = \int_{0}^{q} \mathbf{d}(u)\mathbf{d}(u)'du ,$$

which can be simplified towards

$$\begin{aligned}
\mathbf{N}(q) &= \int_{0}^{q} \mathbf{d}(u)\mathbf{d}(u)'du \\
&= \int_{0}^{q} \mathbf{D}(u)\mathbf{1}\mathbf{1}'\mathbf{D}(u)du \\
&= \int_{0}^{q} \mathbf{K}_d^{-1}\left(\mathbf{I}_n - \exp\left[-\mathbf{K}_d u\right]\right)\mathbf{1}\mathbf{1}'\left(\mathbf{I}_n - \exp\left[-\mathbf{K}_d u\right]\right)\mathbf{K}_d^{-1}du \\
&= \int_{0}^{q} \mathbf{K}_d^{-1}\left[\mathbf{1}\mathbf{1}' - \mathbf{1}\mathbf{1}'\exp\left[-\mathbf{K}_d u\right] - \exp\left[-\mathbf{K}_d u\right]\mathbf{1}\mathbf{1}' + \exp\left[-\mathbf{K}_d u\right]\mathbf{1}\mathbf{1}'\exp\left[-\mathbf{K}_d u\right]\right]\mathbf{K}_d^{-1}du \\
&= \mathbf{K}_d^{-1}\left[\int_{0}^{q} \mathbf{1}\mathbf{1}'du - \mathbf{1}\mathbf{1}'\int_{0}^{q} \exp\left[-\mathbf{K}_d u\right]du - \int_{0}^{q} \exp\left[-\mathbf{K}_d u\right]du\,\mathbf{1}\mathbf{1}' + \int_{0}^{q} \exp\left[-\mathbf{K}_d u\right]\mathbf{1}\mathbf{1}'\exp\left[-\mathbf{K}_d u\right]du\right]\mathbf{K}_d^{-1} \\
&= \mathbf{K}_d^{-1}\left[q\mathbf{1}\mathbf{1}' - \mathbf{1}\mathbf{1}'\mathbf{D}(q) - \mathbf{D}(q)\mathbf{1}\mathbf{1}' + \mathbf{G}(q)\right]\mathbf{K}_d^{-1} .
\end{aligned}$$

The ij – th element of this matrix $(i = 1,..,n; j = 1,..,n)$ equals

$$\left[\mathbf{N}(q)\right]_{ij} = \frac{1}{k_i k_j}\left[q - d_j(q) - d_i(q) + \frac{1 - \exp\left[-(k_i + k_j)q\right]}{k_i + k_j}\right].$$

I

Definitions: (nxn) $\mathcal{H}_j(.,.)$, (n^2xn) $\mathcal{H}(.,.)$

$$\mathcal{H}_j(t,s) \equiv \int_{t}^{s} c_j(t,u)\mathbf{c}(u,s)\mathbf{c}(u,s)'du , \quad j = 1,..,n$$

$$\mathcal{H}(t,s) \equiv \left[\mathcal{H}_1(t,s)',...,\mathcal{H}_n(t,s)'\right]' .$$

The pq – th element of $\mathcal{H}_j(t,s)$ equals

$$\left[\mathcal{H}_j(t,s)\right]_{pq} = \int_t^s c_j(t,u)c_p(u,s)c_q(u,s)du = \int_t^s \exp[-k_j(u-t)-(k_p+k_q)(s-u)]du$$
$$= \frac{\exp[-k_j(s-t)]-\exp[-(k_p+k_q)(s-t)]}{k_p+k_q-k_j}.$$

As $\mathcal{H}_j(t,s)$ and $\mathcal{H}(t,s)$ are functions of $s-t$ only, and not of t or s separately, we typically write the following short cuts:

$$\mathcal{H}_j(s-t) \equiv \mathcal{H}_j(t,s) = \mathcal{H}_j(0,s-t)$$
$$\mathcal{H}(s-t) \equiv \mathcal{H}(t,s) = \mathcal{H}(0,s-t).$$

Definitions: (nxn) $\mathcal{I}_j(.,.)$, $(n^2xn)\,\mathcal{I}(.,.)$

$$\mathcal{I}_j(t,T) \equiv \int_t^T c_j(t,u)\mathbf{d}(u,T)\mathbf{d}(u,T)'du, \; j=1,..,n$$
$$\mathcal{I}(t,T) \equiv \left[\mathcal{I}_1(t,T)',...,\mathcal{I}_n(t,T)'\right]'.$$

The pq – th element of matrix $\mathcal{I}_j(t,T)$ equals

$$\left[\mathcal{I}_j(t,T)\right]_{pq} = \int_t^T c_j(t,u)d_p(u,T)d_q(u,T)du$$
$$= \int_t^T \exp[-k_j(u-t)]\left(\frac{1-\exp[-k_p(T-u)]}{k_p}\right)\left(\frac{1-\exp[-k_q(T-u)]}{k_q}\right)du$$
$$= \frac{1}{k_pk_q}\left[\int_t^T \exp[-k_j(u-t)]du - \int_t^T \exp[(k_q-k_j)u+k_jt-k_qT]du\right.$$
$$\left.-\int_t^T \exp[(k_p-k_j)u+k_jt-k_pT]du + \int_t^T \exp[-k_j(u-t)-(k_p+k_q)(T-u)]du\right]$$
$$= \frac{1}{k_pk_q}\left[d_j(\tau) - \int_t^T \exp[(k_q-k_j)u+k_jt-k_qT]du\right.$$
$$\left.-\int_t^T \exp[(k_p-k_j)u+k_jt-k_pT]du + \left[\mathcal{H}_j(\tau)\right]_{pq}\right],$$

with $\tau = T-t$, and in which the integrals in this latter expression equal

$$\int_t^T \exp[(k_q - k_j)u + k_j t - k_q T]du = \begin{cases} \dfrac{\exp[-k_j\tau] - \exp[-k_q\tau]}{k_q - k_j}; & k_j \neq k_q \\ \tau \exp[-k_j\tau]; & k_j = k_q \end{cases}$$

$$\int_t^T \exp[(k_p - k_j)u + k_j t - k_p T]du = \begin{cases} \dfrac{\exp[-k_j\tau] - \exp[-k_p\tau]}{k_p - k_j}; & k_j \neq k_p \\ \tau \exp[-k_j\tau]; & k_j = k_p. \end{cases}$$

As $\mathcal{I}_j(t,T)$ and $\mathcal{I}(t,T)$ are functions of $\tau = T - t$ only, and not of t or T separately, we typically write the following short cuts:

$$\mathcal{I}_j(T-t) \equiv \mathcal{I}_j(t,T) = \mathcal{I}_j(0, T-t)$$
$$\mathcal{I}(T-t) \equiv \mathcal{I}(t,T) = \mathcal{I}(0, T-t).$$

Limits of matrix functions

The limits below are invoked in section 3.3.5 of chapter III and in appendix IIIc. It holds that

$$\lim_{\delta t \to 0} \mathbf{D}(\delta t) = \lim_{\delta t \to 0} \mathbf{K}_d^{-1}\left(\mathbf{I}_n - \exp\left[-\mathbf{K}_d \delta t\right]\right) = \mathbf{0}$$
$$\lim_{\delta t \to 0} \mathbf{D}(\delta t)/\delta t = \lim_{\delta t \to 0} \left(\mathbf{K}_d \delta t\right)^{-1}\left(\mathbf{I}_n - \exp\left[-\mathbf{K}_d \delta t\right]\right) = \mathbf{I}_n,$$

where the latter limit follows from an application of l'Hôspital's rule. To derive the limit $\lim_{\delta t \to 0} \mathbf{N}(\delta t)/\delta t$, note first that

$$\frac{\mathbf{N}(\delta t)}{\delta t} = \mathbf{K}_d^{-1}\left(\mathbf{1}\mathbf{1}' - \mathbf{1}\mathbf{1}'\frac{\mathbf{D}(\delta t)}{\delta t} - \frac{\mathbf{D}(\delta t)}{\delta t}\mathbf{1}\mathbf{1}' + \frac{\mathbf{G}(\delta t)}{\delta t}\right)\mathbf{K}_d^{-1}.$$

By l'Hôspital's rule, the limit of the ij-th element of the matrix function $\mathbf{G}(\delta t)/\delta t$ for δt approaching zero equals

$$\lim_{\delta t \to 0}\left[\frac{\mathbf{G}(\delta t)}{\delta t}\right]_{ij} = \lim_{\delta t \to 0} \frac{1 - \exp\left[-(k_i + k_j)\delta t\right]}{(k_i + k_j)\delta t} = \lim_{\delta t \to 0} \frac{(k_i + k_j)\exp\left[-(k_i + k_j)\delta t\right]}{k_i + k_j} = 1.$$

As such, $\lim_{\delta t \to 0} \mathbf{G}(\delta t)/\delta t = \mathbf{1}\mathbf{1}'$ and hence $\lim_{\delta t \to 0} \mathbf{N}(\delta t)/\delta t = \mathbf{0}$, the zero-matrix.

- Appendix B -

Statistical Properties of the Multifactor Affine SV Process

This appendix derives statistical properties of the volatility-driving factors $\mathbf{x}$ and the stock volatility process. The multifactor affine SV specification and its properties are stated under some arbitrary probability measure $\mathbb{M}$. Taking the specific evolutions of the factors under the market and risk-neutral measures $\mathbb{P}$ and $\mathbb{Q}$ into account, the results below may subsequently be translated to these specific measures. Sections 1 to 10 exclusively deal with the properties of the factors and the stock variance process, without involving the specific evolution of the stock price as a geometric Brownian motion-type process (with SV). Sections 11 and 12 do involve the stock price evolution; these sections look at stock returns.

1. The SV specification under measure $\mathbb{M}$

Consider the multifactor affine SV specification, in which the instantaneous stock variance σ_t^2 is driven by n possibly correlated, unobservable factors $\mathbf{x} = (x_1,..,x_n)'$ in an affine way,

$$\sigma_t^2 = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t, \tag{1.1}$$

in which δ_0 and $(nx1)\,\boldsymbol{\delta} = (\delta_1,..,\delta_n)'$ are positively valued. Under $\mathbb{M}$ the latent factors evolve according to stationary mean-reverting Markov diffusions,

$$d\mathbf{x}_t = \mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t)dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t}, \qquad (\mathbb{M}) \tag{1.2}$$

in which $(nx1)\,\boldsymbol{\theta} > \mathbf{0}$ represents the mean of the factors (as shown below), $(nxn)\,\mathbf{K}_d, \boldsymbol{\Sigma}$ are matrices of constants with $\mathbf{K}_d = \text{diag}[k_1,..,k_n]$ being diagonal and positive definite, and $\boldsymbol{\Lambda}_t$ is a diagonal matrix given by

$$\boldsymbol{\Lambda}_t = \text{diag}\left[\sqrt{\alpha_1 + \boldsymbol{\beta}_1'\mathbf{x}_t},..,\sqrt{\alpha_n + \boldsymbol{\beta}_n'\mathbf{x}_t}\right], \tag{1.3}$$

in which $\boldsymbol{\alpha} = (\alpha_1,..,\alpha_n)'$ and $\boldsymbol{\beta}_i = (\beta_{i1},..,\beta_{in})'$, $i = 1,..,n$ are $(nx1)$ vectors of positive real-valued constants. Uncertainty is resolved by an n-dimensional standard Brownian motion process $\{\mathbf{W}_{x,t}; t \geq 0\}$, given by

$$\mathbf{W}_{x,t} = (W_{1t},..,W_{nt})', \qquad (\mathbb{M}) \tag{1.4}$$

which is defined on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{M})$, satisfying the "usual conditions"; see e.g. Protter (1990). $\{\mathcal{F}_t; t \geq 0\}$ denotes the natural Brownian filtration.

As $\mathbb{E}_{\mathbb{M}}[d\mathbf{x}_t \mid \mathcal{F}_t] = \mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t)dt$, with $\boldsymbol{\theta} > \mathbf{0}$, $\mathbf{K}_d > \mathbf{0}$, it is clear that whenever the factors are below (resp. above) their mean, there is an upward (downward) drift tendency. As such, the factors tend to revert back to their mean at all times, which is characteristic for a (covariance) stationary process. The matrix $\mathbf{K}_d$ governs the speed of adjustment of the factors towards their mean $\boldsymbol{\theta}$. We assume the dynamics of the factors to be well defined, which requires $\alpha_i + \boldsymbol{\beta}_i{}'\mathbf{x}_t \geq 0$ for all i and t. [1] The Heston (1993) (i.e., CIR) volatility specification is a special case of this model; see section 14. If the factors follow a multifactor Ornstein-Uhlenbeck (OU) process (in case $\boldsymbol{\beta}_i = \mathbf{0}\ \forall i$), the stock variance is not guaranteed to stay positive at all times. This is theoretically inconsistent. (We refer to the main text for further discussion.) The volatility specification allows for level-dependent volatility-of-volatility, also called volatility feedback.

As becomes apparent in section 6, in discrete-time, this SV specification boils down to the stock variance being modeled as a superposition of n (possibly correlated) autoregressive processes of order 1 (i.e., AR(1)-processes). Each AR(1) process features its own mean, and has an error term which is unconditionally white noise, but which conditional variance depends on the current level of the AR(1) process(es).

2. Expressing $\mathbf{x}_s$ in terms of $\mathbf{x}_t$

From the SDE of the factors $(nx1)\,\mathbf{x}$ under $\mathbb{M}$,

$$d\mathbf{x}_t = \mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t)dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t}, \qquad (\mathbb{M}) \qquad (2.1)$$

the basic equation from which to derive the factor moments is one in which $\mathbf{x}_s$ is expressed in terms of $\mathbf{x}_t$, for $s > t$. To achieve this, consider the transformation

$$\mathbf{y}_t \equiv \exp[\mathbf{K}_d t]\mathbf{x}_t, \qquad (2.2)$$

with $\exp[\mathbf{K}_d t] = \mathrm{diag}[\exp(k_1 t),..,\exp(k_n t)]$; see appendix A. By Itô's lemma,

$$d\mathbf{y}_t = \mathbf{K}_d \exp[\mathbf{K}_d t]\mathbf{x}_t dt + \exp[\mathbf{K}_d t]\,d\mathbf{x}_t = \exp(\mathbf{K}_d t)[\mathbf{K}_d\boldsymbol{\theta}dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t}], \qquad (2.3)$$

in which the second equality uses the diagonality of both $\mathbf{K}_d$ and $\exp[\mathbf{K}_d t]$, such that these matrices are commutative. Integrating the increments over the interval $[t,s]$ yields

$$\mathbf{y}_s = \mathbf{y}_t + \int_t^s \exp[\mathbf{K}_d u]\mathbf{K}_d\boldsymbol{\theta}du + \int_t^s \exp[\mathbf{K}_d u]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u}. \qquad (2.4)$$

Transforming back to $\mathbf{x}_s$ by premultiplying with $\exp[-\mathbf{K}_d s]$ yields

$$\mathbf{x}_s = \exp[-\mathbf{K}_d(s-t)]\mathbf{x}_t + \int_t^s \exp[-\mathbf{K}_d(s-u)]\mathbf{K}_d\boldsymbol{\theta}du + \int_t^s \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u}$$

$$= \exp[-\mathbf{K}_d(s-t)]\mathbf{x}_t + (\mathbf{I}_n - \exp[-\mathbf{K}_d(s-t)])\boldsymbol{\theta} + \int_t^s \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u}. \qquad (2.5)$$

such that

$$\mathbf{x}_s = \boldsymbol{\theta} + \exp[-\mathbf{K}_d(s-t)](\mathbf{x}_t - \boldsymbol{\theta}) + \int_t^s \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u}. \qquad (\mathbb{M}) \qquad (2.6)$$

[1] Parameter restrictions that ensure a unique strong solution to the SDE (1.2) are in Duffie and Kan (1996) and Dai and Singleton (2000).

3. (Conditional) mean of factors, and the mean-reverting property

Consider the integral in equation (2.6). Like ordinary Riemann and Riemann Stieltjes integrals, Itô integrals are *linear on adjacent intervals*. Therefore

$$\begin{aligned}
&\mathbb{E}_{\mathbb{M}}\left[\int_t^s \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} \mid \mathcal{F}_t\right] \qquad (3.1)\\
&\quad = \mathbb{E}_{\mathbb{M}}\left[\int_0^s \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} - \int_0^t \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} \mid \mathcal{F}_t\right]\\
&\quad = \mathbb{E}_{\mathbb{M}}\left[\int_0^s \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} \mid \mathcal{F}_t\right] - \int_0^t \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u}\\
&\quad = \int_0^t \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} - \int_0^t \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} = \mathbf{0},
\end{aligned}$$

in which the third equality follows from the *martingale property* of Itô stochastic integrals. The conditional expectation of $\mathbf{x}_s$ given $\mathcal{F}_t$ then follows from (2.6) as

$$\mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_s \mid \mathcal{F}_t\right] = \boldsymbol{\theta} + \exp[-\mathbf{K}_d(s-t)]\left(\mathbf{x}_t - \boldsymbol{\theta}\right). \tag{3.2}$$

As the factor process is a *Markov process*, conditioning on the information filtration $\mathcal{F}_t$ is equivalent to conditioning on the value the process assumes at time t. Therefore $\mathbb{E}_{\mathbb{M}}[\mathbf{x}_s \mid \mathbf{x}_t] = \mathbb{E}_{\mathbb{M}}[\mathbf{x}_s \mid \mathcal{F}_t]$. Given (3.2), (2.6) yields the following interpretation

$$\mathbf{x}_s = \mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_s \mid \mathcal{F}_t\right] + \mathbf{u}_{t,s}, \tag{3.3}$$

in which we label the deviation of the factors from their conditional mean, i.e.,

$$\mathbf{u}_{t,s} \equiv \int_t^s \exp\left[-\mathbf{K}_d(s-u)\right]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} = \mathbf{x}_s - \mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_s \mid \mathcal{F}_t\right], \tag{3.4}$$

the *disturbance*- or *error term*. Evidently (but as also shown in (3.1)), $\mathbf{u}_{t,s}$ has (conditional) expectation zero. The unconditional mean of the factors equals

$$\mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_s\right] = \mathbb{E}_{\mathbb{M}}\left(\mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_s \mid \mathcal{F}_t\right]\right) = \boldsymbol{\theta} + \exp[-\mathbf{K}_d(s-t)]\left(\mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_t\right] - \boldsymbol{\theta}\right). \tag{3.5}$$

Given stationarity, $\mathbb{E}_{\mathbb{M}}[\mathbf{x}_s] = \mathbb{E}_{\mathbb{M}}[\mathbf{x}_t]\ \forall t, s$. This in turn implies

$$\left(\mathbf{I}_n - \exp[-\mathbf{K}_d(s-t)]\right)\left(\mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_s\right] - \boldsymbol{\theta}\right) = \mathbf{0}. \tag{3.6}$$

Apparently, the vector $\left(\mathbb{E}_{\mathbb{M}}[\mathbf{x}_s] - \boldsymbol{\theta}\right)$ lies in the null space of the diagonal matrix $\mathbf{I}_n - \exp[-\mathbf{K}_d(s-t)]$. As $k_i > 0\ \forall i = 1,..,n$, the matrix $\mathbf{I}_n - \exp[-\mathbf{K}_d(s-t)]$ has full rank, such that (3.6) implies that the unconditional mean of the factors equals

$$\mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_s\right] = \boldsymbol{\theta}. \tag{3.7}$$

Below it will often turn out convenient to state results in terms of the factors in deviation from their mean, which we define as

$$\mathbf{x}_s^* \equiv \mathbf{x}_s - \boldsymbol{\theta} \ \forall s \,. \tag{3.8}$$

The conditional factor expectation (3.2) represents the sum of two parts, one being the unconditional factor expectation $\boldsymbol{\theta}$, the other the current deviation of the factors from their mean value, $\mathbf{x}_t - \boldsymbol{\theta}$, scaled by the matrix $\exp[-\mathbf{K}_d(s-t)]$. It holds that

$$\lim_{s\to\infty} \mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_s \mid \mathcal{F}_t\right] = \boldsymbol{\theta}\,, \tag{3.9}$$

as expected, given stationarity. Hence, given current information, the best forecast of the factors in the indefinite future equals their unconditional mean.

Mean-reversion

A characteristic property of stationary processes is *mean-reversion*: Irrespective of the current deviation of the process from its mean, the process will always return to its mean. To illustrate this property of the factors in discrete time, from (3.2) it holds that

$$\mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_s - \mathbf{x}_t \mid \mathcal{F}_t\right] = \left(\exp[-\mathbf{K}_d(s-t)] - \mathbf{I}_n\right)\left(\mathbf{x}_t - \boldsymbol{\theta}\right). \tag{3.10}$$

The diagonal matrix $\exp[-\mathbf{K}_d(s-t)] - \mathbf{I}_n$ ($s > t$) is negative definite. Equation (3.10) reveals that if the factors are currently above (resp. below) their mean, their tendency is to revert back to their mean, as the expected increment over $[t,s]$ is then negative (resp. positive). The factors thus indeed show mean-reversion.

4. Conditional variance of the factors

Appendix A states the definitions of the scalar functions $c_i(.,.); i = 1,..,n$, the vector function $(nx1)\,\mathbf{c}(.,.)$ and the matrix function $(nxn)\,\mathbf{C}_d(.,.)$. From (2.6), the conditional variance of $\mathbf{x}_s$ given $\mathcal{F}_t$ (for $s > t$) then becomes

$$\begin{aligned}
\operatorname{var}_{\mathbb{M}}\left[\mathbf{x}_s \mid \mathcal{F}_t\right] &= \operatorname{var}_{\mathbb{M}}\left[\int_t^s \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} \middle| \mathcal{F}_t\right] \\
&= \operatorname{var}_{\mathbb{M}}\left[\int_t^s \mathbf{C}_d(u,s)\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} \middle| \mathcal{F}_t\right] \\
&= \int_t^s \mathbb{E}_{\mathbb{M}}\left[\mathbf{C}_d(u,s)\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} d\mathbf{W}_{x,u}{}'\boldsymbol{\Lambda}_u\boldsymbol{\Sigma}'\mathbf{C}_d(u,s)\middle|\mathcal{F}_t\right] \\
&= \int_t^s \mathbb{E}_{\mathbb{M}}\left[\mathbf{C}_d(u,s)\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u^2\boldsymbol{\Sigma}'\mathbf{C}_d(u,s)\middle|\mathcal{F}_t\right] du \\
&= \int_t^s \mathbf{C}_d(u,s)\boldsymbol{\Sigma}\,\mathbb{E}_{\mathbb{M}}\left[\boldsymbol{\Lambda}_u^2\middle|\mathcal{F}_t\right]\boldsymbol{\Sigma}'\mathbf{C}_d(u,s)\,du\,,
\end{aligned} \tag{4.1}$$

in which the fourth equality uses $d\mathbf{W}_{x,u} d\mathbf{W}_{x,u}{}' = \mathbf{I}_n du$, which holds since the individual Brownian motions are independent. Using the lemma in appendix A , we subsequently obtain

$$\text{var}_{\mathbb{M}}\left[\mathbf{x}_s \mid \mathcal{F}_t\right] = \int_t^s \left(\mathbf{c}(u,s)\mathbf{c}(u,s)' \odot \mathbf{\Sigma}\,\mathbb{E}_{\mathbb{M}}\left[\mathbf{\Lambda}_u^2 \middle| \mathcal{F}_t\right]\mathbf{\Sigma}'\right) du\,. \tag{4.2}$$

The aim is to arrive at an analytical expression not involving integrals. As a first step, partition the matrix $\mathbf{\Sigma}$ into its columns: $\mathbf{\Sigma} = [\boldsymbol{\sigma}_1,..,\boldsymbol{\sigma}_n]$. $(nx1)$ $\boldsymbol{\sigma}_i$ thus represents the i-th column of $\mathbf{\Sigma}$. As $\mathbf{\Lambda}_t^2 = \text{diag}[\alpha_1 + \boldsymbol{\beta}_1'\mathbf{x}_t,..,\alpha_n + \boldsymbol{\beta}_n'\mathbf{x}_t]$,

$$\mathbf{\Sigma}\,\mathbb{E}_{\mathbb{M}}\left[\mathbf{\Lambda}_u^2 \middle| \mathcal{F}_t\right]\mathbf{\Sigma}' = \sum_{i=1}^n \left(\alpha_i + \boldsymbol{\beta}_i'\mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_u \mid \mathcal{F}_t\right]\right)\boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'\,, \tag{4.3}$$

and hence

$$\begin{aligned}
\text{var}_{\mathbb{M}}\left[\mathbf{x}_s \mid \mathcal{F}_t\right] &= \int_t^s \left(\mathbf{c}(u,s)\mathbf{c}(u,s)' \odot \left[\sum_{i=1}^n \left(\alpha_i + \boldsymbol{\beta}_i'\mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_u \mid \mathcal{F}_t\right]\right)\boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'\right]\right) du \\
&= \int_t^s \left(\sum_{i=1}^n \left(\alpha_i + \boldsymbol{\beta}_i'\mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_u \mid \mathcal{F}_t\right]\right)\mathbf{c}(u,s)\mathbf{c}(u,s)' \odot \boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'\right) du \\
&= \sum_{i=1}^n \left(\int_t^s \left(\alpha_i + \boldsymbol{\beta}_i'\mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_u \mid \mathcal{F}_t\right]\right)\mathbf{c}(u,s)\mathbf{c}(u,s)'\,du\right) \odot \boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'\,.
\end{aligned} \tag{4.4}$$

The integral in (4.4) becomes

$$\begin{aligned}
&\int_t^s \left(\alpha_i + \boldsymbol{\beta}_i'\mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_u \mid \mathcal{F}_t\right]\right)\mathbf{c}(u,s)\mathbf{c}(u,s)'\,du \\
&\qquad = \int_t^s \left(\alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\theta} + \boldsymbol{\beta}_i'\mathbf{C}_d(t,u)\,\mathbf{x}_t^*\right)\mathbf{c}(u,s)\mathbf{c}(u,s)'\,du \\
&\qquad = (\alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\theta})\int_t^s \mathbf{c}(u,s)\mathbf{c}(u,s)'\,du + \int_t^s \boldsymbol{\beta}_i'\mathbf{C}_d(t,u)\;\mathbf{x}_t^*\,\mathbf{c}(u,s)\mathbf{c}(u,s)'\,du \\
&\qquad = (\alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\theta})\,\mathbf{G}(t,s) + \int_t^s \boldsymbol{\beta}_i'\mathbf{C}_d(t,u)\;\mathbf{x}_t^*\;\mathbf{c}(u,s)\mathbf{c}(u,s)'\,du\,,
\end{aligned} \tag{4.5}$$

in which the last equality invokes the definition of the matrix function $(nxn)\,\mathbf{G}(.,.)$ as stated in appendix A. As $\mathbf{x}_s^* = \mathbf{x}_s - \boldsymbol{\theta}$ has elements $x_{js}^* = x_{js} - \theta_j$; $j = 1,..,n$, the integral in (4.5) becomes

$$\begin{aligned}
\int_t^s \boldsymbol{\beta}_i'\mathbf{C}_d(t,u)\,\mathbf{x}_t^*\,\mathbf{c}(u,s)\mathbf{c}(u,s)'\,du &= \int_t^s \left[\sum_{j=1}^n \beta_{ij}c_j(t,u)\;x_{jt}^*\right]\mathbf{c}(u,s)\mathbf{c}(u,s)'\,du \\
&= \sum_{j=1}^n \left[\beta_{ij}\;x_{jt}^*\int_t^s c_j(t,u)\mathbf{c}(u,s)\mathbf{c}(u,s)'\,du\right] \\
&= \sum_{j=1}^n \beta_{ij}\;x_{jt}^*\;\mathcal{H}_j(t,s)\,,
\end{aligned} \tag{4.6}$$

in which the matrix functions (nxn) $\mathcal{H}_j(t,s)$, $j = 1,..,n$ are defined as

$$\mathcal{H}_j(t,s) \equiv \int_t^s c_j(t,u)\mathbf{c}(u,s)\mathbf{c}(u,s)'du. \tag{4.7}$$

The pq – th element of (nxn) $\mathcal{H}_j(t,s)$, $j = 1,..,n$ equals

$$\left[\mathcal{H}_j(t,s)\right]_{pq} = \int_t^s c_j(t,u)c_p(u,s)c_q(u,s)du = \int_t^s \exp[-k_j(u-t)-(k_p+k_q)(s-u)]du$$
$$= \frac{\exp[-k_j(s-t)] - \exp[-(k_p+k_q)(s-t)]}{k_p+k_q-k_j}. \tag{4.8}$$

These matrices $\mathcal{H}_j(t,s)$ are functions of the time difference $s-t$ only, and not of t or s separately. Therefore, if convenient, we will sometimes write $\mathcal{H}_j(s-t) \equiv \mathcal{H}_j(t,s)$ $= \mathcal{H}_j(0,s-t)$. Expression (4.6) can further be simplified towards

$$\sum_{j=1}^n \beta_{ij}\, x_{jt}^*\, \mathcal{H}_j(t,s) = \left[\beta_{i1}\, x_{1t}^*\, \mathbf{I}_n, \ldots, \beta_{in}\, x_{nt}^*\, \mathbf{I}_n\right] \begin{bmatrix} \mathcal{H}_1(t,s) \\ \vdots \\ \mathcal{H}_n(t,s) \end{bmatrix}$$
$$= \left(\left[\boldsymbol{\beta}_i \odot \mathbf{x}_t^*\right]' \otimes \mathbf{I}_n\right) \mathcal{H}(t,s), \tag{4.9}$$

in which $(n^2xn)\,\mathcal{H}(t,s) \equiv \left[\mathcal{H}_1(t,s)',\ldots,\mathcal{H}_n(t,s)'\right]'$.

Collecting the intermediate results together, we find

$$\text{var}_{\mathbb{M}}\left[\mathbf{x}_s \mid \mathcal{F}_t\right] = \sum_{i=1}^n \left[\left(\left[\alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\theta}\right] \mathbf{G}(t,s) + \left\{\left[\boldsymbol{\beta}_i \odot \mathbf{x}_t^*\right]' \otimes \mathbf{I}_n\right\}\mathcal{H}(t,s)\right) \odot \boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'\right]$$
$$= \sum_{i=1}^n \left[(\alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\theta})\,\mathbf{G}(t,s) \odot \boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'\right] + \sum_{i=1}^n \left(\left\{\left[\boldsymbol{\beta}_i \odot \mathbf{x}_t^*\right]' \otimes \mathbf{I}_n\right\}\mathcal{H}(t,s)\right) \odot \boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'$$
$$= \mathbf{G}(t,s) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' + \sum_{i=1}^n \left(\left\{\left[\boldsymbol{\beta}_i \odot \mathbf{x}_t^*\right]' \otimes \mathbf{I}_n\right\}\mathcal{H}(t,s)\right) \odot \boldsymbol{\sigma}_i\boldsymbol{\sigma}_i', \tag{4.10}$$

in which (recall) $\mathbf{M}_d \equiv \text{diag}\left[\alpha_1 + \boldsymbol{\beta}_1'\boldsymbol{\theta},\ldots,\alpha_n + \boldsymbol{\beta}_n'\boldsymbol{\theta}\right]$. As $\{\mathbf{x}_t\}$ is a Markov process, it also holds that $\text{var}_{\mathbb{M}}[\mathbf{x}_s \mid \mathbf{x}_t] = \text{var}_{\mathbb{M}}[\mathbf{x}_s \mid \mathcal{F}_t]$.

Note that $\text{var}_{\mathbb{M}}[\mathbf{x}_s \mid \mathcal{F}_s] = \mathbf{0}$ (as it should). Moreover, as $\lim_{s\to\infty} \mathbf{G}(t,s) = \lim_{s\to\infty} \mathbf{G}(0,s-t)$ $= \lim_{s\to\infty} \mathbf{J} \odot \left[\mathbf{11}' - \exp[-\mathbf{K}_d(s-t)]\mathbf{11}'\exp[-\mathbf{K}_d(s-t)\right] = \mathbf{J} \odot \mathbf{11}' = \mathbf{J}$, and as it can be seen that $\lim_{s\to\infty} \mathcal{H}(t,s) = \mathbf{0}$, we also have that

$$\lim_{s\to\infty} \text{var}_{\mathbb{M}}\left[\mathbf{x}_s \mid \mathcal{F}_t\right] = \mathbf{J} \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'. \tag{4.11}$$

As shown in the next section, $\mathbf{J} \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' = \text{var}_{\mathbb{M}}\left[\mathbf{x}_t\right]$. Result (4.11) is precisely what we expect for a stationary process. Namely, its current value has no impact on its conditional mean (recall (3.9)) or variance if prediction of the process in the infinite future is the concern. In that case, the current value of the process does not yield any extra information.

5. Variance of the factors & contemporaneous correlation matrix

This section computes the unconditional variance of the factors in two ways. First, due to stationarity, $\text{var}_{\mathbb{M}}[\mathbf{x}_s] = \text{var}_{\mathbb{M}}[\mathbf{x}_t]$. It then follows that

$$\begin{aligned}\text{var}_{\mathbb{M}}\left[\mathbf{x}_s\right] &= \mathbb{E}_{\mathbb{M}}\left[\text{var}_{\mathbb{M}}\left(\mathbf{x}_s \mid \mathcal{F}_t\right)\right] + \text{var}_{\mathbb{M}}\left[\mathbb{E}_{\mathbb{M}}\left(\mathbf{x}_s \mid \mathcal{F}_t\right)\right] \\ &= \mathbf{G}(t,s) \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}' + \text{var}_{\mathbb{M}}\left[\boldsymbol{\theta} + \exp[-\mathbf{K}_d(s-t)](\mathbf{x}_t - \boldsymbol{\theta})\right] \\ &= \mathbf{G}(t,s) \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}' + \mathbf{C}_d(t,s)\ \text{var}_{\mathbb{M}}\left[\mathbf{x}_s\right]\ \mathbf{C}_d(t,s).\end{aligned} \tag{5.1}$$

To solve for the variance, vectorize this expression: [2]

$$\begin{aligned}\text{vec}\left[\text{var}_{\mathbb{M}}\left(\mathbf{x}_s\right)\right] &= \text{vec}\left[\mathbf{G}(t,s) \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right] + \text{vec}\left[\mathbf{C}_d(t,s)\,\text{var}_{\mathbb{M}}\left(\mathbf{x}_s\right)\mathbf{C}_d(t,s)\right] \\ &= \text{vec}\left[\mathbf{G}(t,s) \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right] + \left[\mathbf{C}_d(t,s) \otimes \mathbf{C}_d(t,s)\right]\text{vec}\left[\text{var}_{\mathbb{M}}\left(\mathbf{x}_s\right)\right],\end{aligned} \tag{5.2}$$

in which we exploit the linearity of the vec operator. It next follows that

$$\text{vec}\left[\text{var}_{\mathbb{M}}\left(\mathbf{x}_s\right)\right] = \left[\mathbf{I}_{n^2} - \left(\mathbf{C}_d(t,s) \otimes \mathbf{C}_d(t,s)\right)\right]^{-1}\ \text{vec}\left[\mathbf{G}(t,s) \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right]. \tag{5.3}$$

Equation (5.3) defines the factor variance matrix *implicitly*. It is not clear how to back out this matrix *explicitly*. We nonetheless desire an explicit expression for $\text{var}_{\mathbb{M}}[\mathbf{x}_s]$, for e.g. computing the correlation matrix of the factors.

Let us therefore compute $\text{var}_{\mathbb{M}}[\mathbf{x}_s]$ in a different way. From appendix A, $\mathbf{G}(t,s)$ can be written as

$$\mathbf{G}(t,s) = \mathbf{J} \odot \left[\mathbf{1}\mathbf{1}' - \mathbf{c}(t,s)\mathbf{c}(t,s)'\right]. \tag{5.4}$$

Using (5.4) and invoking the lemma stated in appendix A, we obtain from (5.1)

$$\begin{aligned}\text{var}_{\mathbb{M}}\left[\mathbf{x}_s\right] &= \mathbf{G}(t,s) \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}' + \mathbf{C}_d(t,s)\ \text{var}_{\mathbb{M}}\left[\mathbf{x}_s\right]\ \mathbf{C}_d(t,s) \\ &= \mathbf{J} \odot \left[\mathbf{1}\mathbf{1}' - \mathbf{c}(t,s)\mathbf{c}(t,s)'\right] \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}' + \mathbf{c}(t,s)\mathbf{c}(t,s)' \odot \text{var}_{\mathbb{M}}\left[\mathbf{x}_s\right],\end{aligned} \tag{5.5}$$

such that

$$\left[\mathbf{1}\mathbf{1}' - \mathbf{c}(t,s)\mathbf{c}(t,s)'\right] \odot \text{var}_{\mathbb{M}}\left[\mathbf{x}_s\right] = \mathbf{J} \odot \left[\mathbf{1}\mathbf{1}' - \mathbf{c}(t,s)\mathbf{c}(t,s)'\right] \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'. \tag{5.6}$$

This implies

$$\text{var}_{\mathbb{M}}\left[\mathbf{x}_s\right] = \mathbf{J} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}' \quad \forall s, \tag{5.7}$$

which is an explicit expression for the variance matrix of the factors.

The *contemporaneous correlation matrix* of the factors is next given by

$$\text{corr}_{\mathbb{M}}\left[\mathbf{x}_s\right] = \left[\mathbf{I}_n \odot \mathbf{J} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right]^{-1/2}\left[\mathbf{J} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right]\left[\mathbf{I}_n \odot \mathbf{J} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right]^{-1/2}. \tag{5.8}$$

[2] The second equality in expression (5.2) uses the fact that for $(mxn)\mathbf{A}$, $(nxp)\mathbf{B}$, $(pxq)\mathbf{C}$, it holds that $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec}(\mathbf{B})$. See e.g., Magnus and Neudecker (1988).

6. Properties of $\mathbf{u}_{t,s}$

The error term $\mathbf{u}_{t,s}$ defined in (3.4) plays a pivotal role in the state space estimation method of our multifactor SV model. Its properties are therefore important. First, recall equations (2.6) and (3.4), repeated here for convenience,

$$\mathbf{x}_s = \boldsymbol{\theta} + \exp[-\mathbf{K}_d(s-t)](\mathbf{x}_t - \boldsymbol{\theta}) + \int_t^s \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u}, \qquad (\mathbb{M}) \qquad (6.1)$$

$$\mathbf{u}_{t,s} = \int_t^s \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u}. \qquad (\mathbb{M}) \qquad (6.2)$$

As $\mathbf{x}_s^* = \mathbf{x}_s - \boldsymbol{\theta}\ \forall s$, (6.1) can be rewritten as

$$\mathbf{x}_s^* = \exp[-\mathbf{K}_d(s-t)]\mathbf{x}_t^* + \mathbf{u}_{t,s}. \qquad (\mathbb{M}) \qquad (6.3)$$

From (3.1) it is clear that

$$\mathbb{E}_{\mathbb{M}}\left[\mathbf{u}_{t,s} \mid \mathcal{F}_t\right] = \mathbf{0}, \qquad \mathbb{E}_{\mathbb{M}}\left[\mathbf{u}_{t,s}\right] = \mathbf{0}. \qquad (6.4)$$

From equation (6.3),

$$\begin{aligned}\operatorname{var}_{\mathbb{M}}\left[\mathbf{u}_{t,s} \mid \mathcal{F}_t\right] &= \operatorname{var}_{\mathbb{M}}\left[\mathbf{x}_s \mid \mathcal{F}_t\right] \qquad (6.5)\\ &= \mathbf{G}(t,s) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' + \sum_{i=1}^n \left(\left\{\left[\boldsymbol{\beta}_i \odot \mathbf{x}_t^*\right]' \otimes \mathbf{I}_n\right\}\mathcal{H}(t,s)\right) \odot \boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'.\end{aligned}$$

As such,

$$\operatorname{var}_{\mathbb{M}}\left[\mathbf{u}_{t,s}\right] = \mathbb{E}_{\mathbb{M}}\left[\operatorname{var}_{\mathbb{M}}\left(\mathbf{u}_{t,s} \mid \mathcal{F}_t\right)\right] + \operatorname{var}_{\mathbb{M}}\left[\mathbb{E}_{\mathbb{M}}\left(\mathbf{u}_{t,s} \mid \mathcal{F}_t\right)\right] = \mathbf{G}(t,s) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'. \quad (6.6)$$

What about the correlation between disturbances that do not overlap in time? Consider the covariance between $\mathbf{u}_{t,s}$ and $\mathbf{u}_{u,v}$ for $t < s < u < v$. By first invoking the *law of iterated expectations* and then the *tower property* of conditional expectations, yields

$$\begin{aligned}\operatorname{cov}_{\mathbb{M}}\left[\mathbf{u}_{t,s}, \mathbf{u}_{u,v}\right] &= \mathbb{E}_{\mathbb{M}}\left[\mathbf{u}_{t,s}\mathbf{u}_{u,v}'\right]; \qquad t < s < u < v \qquad (6.7)\\ &= \mathbb{E}_{\mathbb{M}}\left(\mathbb{E}_{\mathbb{M}}\left[\mathbf{u}_{t,s}\mathbf{u}_{u,v}' \mid \mathcal{F}_s\right]\right)\\ &= \mathbb{E}_{\mathbb{M}}\left(\mathbf{u}_{t,s}\, \mathbb{E}_{\mathbb{M}}\left[\mathbf{u}_{u,v}' \mid \mathcal{F}_s\right]\right)\\ &= \mathbb{E}_{\mathbb{M}}\left(\mathbf{u}_{t,s}\mathbb{E}_{\mathbb{M}}\left[\mathbb{E}_{\mathbb{M}}\left(\mathbf{u}_{u,v}' \mid \mathcal{F}_u\right) \mid \mathcal{F}_s\right]\right) = \mathbf{0},\end{aligned}$$

as $\mathbb{E}_{\mathbb{M}}\left(\mathbf{u}_{u,v}' \mid \mathcal{F}_u\right) = \mathbf{0}$ by (6.4). Hence, non-overlapping disturbances do not correlate.

7. Covariance and correlation function of the factors

To examine the dependence structure in the factors over time, we consider their correlation function. The covariance between $\mathbf{x}_s$ and $\mathbf{x}_t$ for $s > t$ equals

$$\begin{aligned}\mathrm{cov}_{\mathbb{M}}\left[\mathbf{x}_s, \mathbf{x}_t\right] &= \mathrm{cov}_{\mathbb{M}}\left[\mathbf{x}_s^*, \mathbf{x}_t^*\right] \\ &= \mathrm{cov}_{\mathbb{M}}\left[\exp[-\mathbf{K}_d(s-t)]\,\mathbf{x}_t^* + \mathbf{u}_{t,s},\ \mathbf{x}_t^*\right] \\ &= \exp[-\mathbf{K}_d(s-t)]\,\mathrm{var}_{\mathbb{M}}\left[\mathbf{x}_t\right] + \mathrm{cov}_{\mathbb{M}}\left[\mathbf{u}_{t,s},\ \mathbf{x}_t^*\right] \\ &= \exp[-\mathbf{K}_d(s-t)]\left(\mathbf{J} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right) + \mathbb{E}_{\mathbb{M}}\left[\mathbf{u}_{t,s}\mathbf{x}_t^{*\prime}\right] \\ &= \exp[-\mathbf{K}_d(s-t)]\left(\mathbf{J} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right) + \mathbb{E}_{\mathbb{M}}\left[\mathbb{E}_{\mathbb{M}}\left(\mathbf{u}_{t,s} \mid \mathcal{F}_t\right)\mathbf{x}_t^{*\prime}\right] \\ &= \exp[-\mathbf{K}_d(s-t)]\left(\mathbf{J} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right),\end{aligned} \tag{7.1}$$

in which the second equality uses (6.3), and the fourth and sixth equality exploit (6.4). The correlation between $\mathbf{x}_s$ and $\mathbf{x}_t$ for $s > t$ then follows as

$$\begin{aligned}\mathrm{corr}_{\mathbb{M}}\left[\mathbf{x}_s, \mathbf{x}_t\right] &= \left(\mathbf{I}_n \odot \mathrm{var}_{\mathbb{M}}\left[\mathbf{x}_s\right]\right)^{-1/2}\ \mathrm{cov}_{\mathbb{M}}\left[\mathbf{x}_s, \mathbf{x}_t\right]\ \left(\mathbf{I}_n \odot \mathrm{var}_{\mathbb{M}}\left[\mathbf{x}_t\right]\right)^{-1/2} \\ &= \left(\mathbf{I}_n \odot \mathbf{J} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right)^{-1/2}\exp[-\mathbf{K}_d(s-t)]\left(\mathbf{J} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right)\left(\mathbf{I}_n \odot \mathbf{J} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right)^{-1/2} \\ &= \exp[-\mathbf{K}_d(s-t)]\left(\mathbf{I}_n \odot \mathbf{J} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right)^{-1/2}\left(\mathbf{J} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right)\left(\mathbf{I}_n \odot \mathbf{J} \odot \mathbf{\Sigma M}_d\mathbf{\Sigma}'\right)^{-1/2} \\ &= \exp[-\mathbf{K}_d(s-t)]\,\mathrm{corr}_{\mathbb{M}}\left[\mathbf{x}_s\right],\end{aligned} \tag{7.2}$$

in which the matrix-interchange after the second equality is allowed due to the diagonality of the matrices involved, and in which the latter equality uses (5.8).

The covariance and correlation between $\mathbf{x}_s$ and $\mathbf{x}_t$ depend on the time distance $s-t$ only and not on their individual time stamps s or t. The dependence structure in $\{\mathbf{x}_t; t \geq 0\}$ is thus invariant under shifts of time. Moreover, for fixed t

$$\lim_{s\to\infty} \mathrm{cov}_{\mathbb{M}}\left[\mathbf{x}_s, \mathbf{x}_t\right] = \mathbf{0}, \qquad \lim_{s\to\infty} \mathrm{corr}_{\mathbb{M}}\left[\mathbf{x}_s, \mathbf{x}_t\right] = \mathbf{0}, \tag{7.3}$$

as $\mathbf{K}_d$ is positive definite. The dependence thus fades away over time: A current shock to $\{\mathbf{x}_t; t \geq 0\}$ has a diminishing effect on the future evolution of the process as time goes by. Shocks are thus not persistent. Indeed, these properties are due to stationarity of $\{\mathbf{x}_t; t \geq 0\}$.

8. Properties of $\{\sigma_t^2\}$ and volatility mean reversion

From (3.2),

$$\mathbb{E}_{\mathbb{M}}\left[\sigma_s^2 \mid \mathcal{F}_t\right] = \delta_0 + \boldsymbol{\delta}'\ \mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_s \mid \mathcal{F}_t\right] = \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} + \boldsymbol{\delta}'\exp[-\mathbf{K}_d(s-t)]\,\mathbf{x}_t^*. \tag{8.1}$$

The mean stock variance equals

$$\mathbb{E}_{\mathbb{M}}\left[\sigma_s^2\right] = \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta}. \tag{8.2}$$

The conditional mean stock variance thus equals the sum of the unconditional mean stock variance and a term representing the scaled factors in deviation from their mean. This makes sense intuitively. For the conditional variance of the instantaneous stock variance, by (4.10) we find

$$\begin{aligned}\mathrm{var}_{\mathbb{M}}\left[\sigma_s^2 \mid \mathcal{F}_t\right] &= \boldsymbol{\delta}'\mathrm{var}_{\mathbb{M}}\left[\mathbf{x}_s \mid \mathcal{F}_t\right]\boldsymbol{\delta} \\ &= \boldsymbol{\delta}'\left[\mathbf{G}(t,s)\odot\boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' + \sum_{i=1}^{n}\left(\left\{\left[\boldsymbol{\beta}_i\odot\mathbf{x}_t^*\right]'\otimes\mathbf{I}_n\right\}\mathcal{H}(t,s)\right)\odot\boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'\right]\boldsymbol{\delta}.\end{aligned} \tag{8.3}$$

From (5.7), the unconditional variance of the stock variance equals

$$\mathrm{var}_{\mathbb{M}}\left[\sigma_s^2\right] = \boldsymbol{\delta}'\mathrm{var}_{\mathbb{M}}\left[\mathbf{x}_s\right]\boldsymbol{\delta} = \boldsymbol{\delta}'\left[\mathbf{J}\odot\boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right]\boldsymbol{\delta}. \tag{8.4}$$

What about the conditional distribution of the stock variance? Notice that

$$\sigma_s^2 \mid \mathcal{F}_t = \delta_0 + \boldsymbol{\delta}'\left(\mathbf{x}_s \mid \mathcal{F}_t\right), \tag{8.5}$$

with

$$\mathbf{x}_s \mid \mathcal{F}_t = \boldsymbol{\theta} + \exp[-\mathbf{K}_d(s-t)]\ \mathbf{x}_t^* + \left(\int_t^s \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\boldsymbol{\Lambda}_u d\mathbf{W}_{x,u} \mid \mathcal{F}_t\right). \ (\mathbb{M}) \tag{8.6}$$

The conditional distribution of the integral in this expression is essentially a random "continuous mixture of normals" (except in the OU case, for which it is Gaussian). As such, the distribution $\sigma_s^2 \mid \mathcal{F}_t$ is a random continuous mixture of normals as well, and will therefore be characterized by *fat tails* (except in the OU case).

Regarding the dependence structure in the volatility process, for $s > t$

$$\mathrm{cov}_{\mathbb{M}}\left[\sigma_s^2,\sigma_t^2\right] = \boldsymbol{\delta}'\mathrm{cov}_{\mathbb{M}}\left[\mathbf{x}_s,\mathbf{x}_t\right]\boldsymbol{\delta} = \boldsymbol{\delta}'\exp[-\mathbf{K}_d(s-t)]\left(\mathbf{J}\odot\boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right)\boldsymbol{\delta}, \tag{8.7}$$

and

$$\begin{aligned}\mathrm{corr}_{\mathbb{M}}\left[\sigma_s^2,\sigma_t^2\right] &= \mathrm{cov}_{\mathbb{M}}\left[\sigma_s^2,\sigma_t^2\right]/\mathrm{var}_{\mathbb{M}}\left[\sigma_s^2\right] \\ &= \boldsymbol{\delta}'\exp[-\mathbf{K}_d(s-t)]\left(\mathbf{J}\odot\boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right)\boldsymbol{\delta}/\boldsymbol{\delta}'\left(\mathbf{J}\odot\boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right)\boldsymbol{\delta}.\end{aligned} \tag{8.8}$$

For t fixed,

$$\lim_{s\to\infty}\mathrm{cov}_{\mathbb{M}}\left[\sigma_s^2,\sigma_t^2\right] = 0, \qquad \lim_{s\to\infty}\mathrm{corr}_{\mathbb{M}}\left[\sigma_s^2,\sigma_t^2\right] = 0. \tag{8.9}$$

Evidently, as the stock variance process is an affine function of $\{\mathbf{x}_t; t \geq 0\}$ which is stationary, the stock variance process itself is stationary. Therefore, the dependence in the stock variance process dies out over time as well. To illustrate volatility mean reversion, it holds

$$\begin{aligned}\mathbb{E}_{\mathbb{M}}\left[\sigma_s^2-\sigma_t^2 \mid \mathcal{F}_t\right] &= \mathbb{E}_{\mathbb{M}}\left[\delta_0+\boldsymbol{\delta}'\mathbf{x}_s-(\delta_0+\boldsymbol{\delta}'\mathbf{x}_t) \mid \mathcal{F}_t\right] \\ &= \boldsymbol{\delta}'\mathbb{E}_{\mathbb{M}}\left[\mathbf{x}_s-\mathbf{x}_t \mid \mathcal{F}_t\right] \\ &= \boldsymbol{\delta}'\left(\exp[-\mathbf{K}_d(s-t)]-\mathbf{I}_n\right)\left(\mathbf{x}_t-\boldsymbol{\theta}\right),\end{aligned} \tag{8.10}$$

which uses (3.10). As the diagonal matrix $\exp[-\mathbf{K}_d(s-t)]-\mathbf{I}_n$ is negative definite, (8.10) shows that if $\mathbf{x}_t > \mathbb{E}_{\mathbb{M}}[\mathbf{x}_t] = \boldsymbol{\theta}$ such that $\sigma_t^2 > \mathbb{E}_{\mathbb{M}}[\sigma_t^2] = \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta}$, then $\mathbb{E}_{\mathbb{M}}[\sigma_s^2-\sigma_t^2 \mid \mathcal{F}_t] < 0$ and vice versa. Hence, $\{\sigma_t^2; t \geq 0\}$ has the tendency to revert back to its mean at all times.

9. (Conditional) mean and variance of integrated variance $\int_t^T \sigma_s^2 ds$

The *integrated stock variance* over the interval $[t,T]$ equals

$$\int_t^T \sigma_s^2 ds = \delta_0 \tau + \boldsymbol{\delta}' \int_t^T \mathbf{x}_s ds \,, \tag{9.1}$$

in which we define $\tau \equiv T - t$. Recall (2.6), which expresses $\mathbf{x}_s$ in terms of $\mathbf{x}_t$ for $s > t$:

$$\mathbf{x}_s = \boldsymbol{\theta} + \exp[-\mathbf{K}_d(s-t)](\mathbf{x}_t - \boldsymbol{\theta}) + \mathbf{u}_{t,s} \,, \qquad (\mathbb{M}) \tag{9.2}$$

with $\mathbf{u}_{t,s}$ defined in (3.4) as

$$\mathbf{u}_{t,s} = \int_t^s \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma\Lambda}_u d\mathbf{W}_{x,u} \,, \qquad (\mathbb{M}) \tag{9.3}$$

and which we labeled the disturbance or error term since $\mathbf{u}_{t,s} = \mathbf{x}_s - \mathbb{E}_{\mathbb{M}}[\mathbf{x}_s \mid \mathbf{x}_t]$; see (3.3). Integrating the process $\{\mathbf{x}_s; s \geq 0\}$ over $[t,T]$ yields

$$\int_t^T \mathbf{x}_s ds = \boldsymbol{\theta}\tau + \int_t^T \exp[-\mathbf{K}_d(s-t)]ds \; \mathbf{x}_t^* + \int_t^T \mathbf{u}_{t,s} ds = \boldsymbol{\theta}\tau + \mathbf{D}(\tau)\mathbf{x}_t^* + \int_t^T \mathbf{u}_{t,s} ds \,, \quad (\mathbb{M}) \tag{9.4}$$

with $\mathbf{D}(\tau) = \mathbf{K}_d^{-1}\left(\mathbf{I}_n - \exp[-\mathbf{K}_d\tau]\right)$, as defined in appendix A. Next,

$$\int_t^T \sigma_s^2 ds = (\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\tau + \boldsymbol{\delta}'\mathbf{D}(\tau)\mathbf{x}_t^* + \boldsymbol{\delta}' \int_t^T \mathbf{u}_{t,s} ds \,. \qquad (\mathbb{M}) \tag{9.5}$$

Equation (9.5) has a fairly natural interpretation. The integrated stock variance over $[t,T]$, expressed in terms of time-t information, i.e. $\mathbf{x}_t$, consists of the sum of three parts. The first part is the unconditional stock variance weighted by the length of the interval. The second part measures the current deviation of the variance-driving factors from their mean, $\mathbf{x}_t^*$, scaled by the matrix $\mathbf{D}(\tau)$. The third part is interpreted as the *integrated disturbance term* over $[t,T]$, which is the random component in the integrated stock variance, given time-t information.

The integrated disturbance term $\int_t^T \mathbf{u}_{t,s} ds$ and its moments

To derive the moments of the integrated stock variance, the moments of the integrated disturbance term are needed. Invoking the *stochastic Fubini theorem*, which enables to interchange the order of classical Riemann-Stieltjes integrals and Itô stochastic integrals, yields

$$\int_t^T \mathbf{u}_{t,s} ds = \int_t^T \left(\int_t^s \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma\Lambda}_u d\mathbf{W}_{x,u} \right) ds \qquad (\mathbb{M}) \tag{9.6}$$

$$= \int_t^T \left(\int_u^T \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma\Lambda}_u ds \right) d\mathbf{W}_{x,u}$$

$$= \int_t^T \left(\int_u^T \exp[-\mathbf{K}_d (s-u)]\, ds \right) \mathbf{\Sigma \Lambda}_u d\mathbf{W}_{x,u} = \int_t^T \mathbf{D}(u,T) \mathbf{\Sigma \Lambda}_u d\mathbf{W}_{x,u} \, .$$

As Itô integrals are martingales,

$$\mathbb{E}_{\mathbb{M}} \left[\int_t^T \mathbf{u}_{t,s} ds \mid \mathcal{F}_t \right] = \mathbf{0} , \tag{9.7}$$

and thus

$$\mathbb{E}_{\mathbb{M}} \left[\int_t^T \mathbf{u}_{t,s} ds \right] = \mathbb{E}_{\mathbb{M}} \left(\mathbb{E}_{\mathbb{M}} \left[\int_t^T \mathbf{u}_{t,s} ds \mid \mathcal{F}_t \right] \right) = \mathbf{0} \, . \tag{9.8}$$

Intuitively, this result is immediately clear, as the integrated disturbance term is essentially nothing but a "continuous sum" of disturbances which each have mean zero.

To derive its conditional variance, we follow an analysis largely in line with the work performed to obtain $\text{var}_{\mathbb{M}}[\mathbf{x}_s \mid \mathcal{F}_t]$. As such, we skip many of the intermediate steps below. Due to the *Itô isometry* and the lemma stated in appendix A, it holds that

$$\begin{aligned} \text{var}_{\mathbb{M}} \left[\int_t^T \mathbf{u}_{t,s} ds \mid \mathcal{F}_t \right] &= \text{var}_{\mathbb{M}} \left[\int_t^T \mathbf{D}(u,T) \mathbf{\Sigma \Lambda}_u d\mathbf{W}_{x,u} \mid \mathcal{F}_t \right] \\ &= \int_t^T \mathbf{D}(u,T) \mathbf{\Sigma} \, \mathbb{E}_{\mathbb{M}} \left[\mathbf{\Lambda}_u^2 \mid \mathcal{F}_t \right] \mathbf{\Sigma}' \mathbf{D}(u,T) du \\ &= \int_t^T \left(\mathbf{d}(u,T) \mathbf{d}(u,T)' \odot \mathbf{\Sigma} \, \mathbb{E}_{\mathbb{M}} \left[\mathbf{\Lambda}_u^2 \mid \mathcal{F}_t \right] \mathbf{\Sigma}' \right) du \\ &= \int_t^T \left(\mathbf{d}(u,T) \mathbf{d}(u,T)' \odot \left[\sum_{i=1}^n \left(\alpha_i + \boldsymbol{\beta}_i' \mathbb{E}_{\mathbb{M}} \left[\mathbf{x}_u \mid \mathcal{F}_t \right] \right) \boldsymbol{\sigma}_i \boldsymbol{\sigma}_i' \right] \right) du \\ &= \sum_{i=1}^n \left(\int_t^T \left(\alpha_i + \boldsymbol{\beta}_i' \mathbb{E}_{\mathbb{M}} \left[\mathbf{x}_u \mid \mathcal{F}_t \right] \right) \mathbf{d}(u,T) \mathbf{d}(u,T)' du \right) \odot \boldsymbol{\sigma}_i \boldsymbol{\sigma}_i' , \end{aligned} \tag{9.9}$$

in which the fourth equality uses the partition of $\mathbf{\Sigma}$ into its columns, $\mathbf{\Sigma} = [\boldsymbol{\sigma}_1, .., \boldsymbol{\sigma}_n]$, with $(nx1)$ $\boldsymbol{\sigma}_i$ the i-th column of $\mathbf{\Sigma}$. Using the expression for $\mathbb{E}_{\mathbb{M}}[\mathbf{x}_s \mid \mathcal{F}_t]$ in (3.2), the integral in the latter expression can be written as

$$\begin{aligned} &\int_t^T \left(\alpha_i + \boldsymbol{\beta}_i' \mathbb{E}_{\mathbb{M}} \left[\mathbf{x}_u \mid \mathcal{F}_t \right] \right) \mathbf{d}(u,T) \mathbf{d}(u,T)' du \\ &\quad = (\alpha_i + \boldsymbol{\beta}_i' \boldsymbol{\theta}) \int_t^T \mathbf{d}(u,T) \mathbf{d}(u,T)' du + \int_t^T \boldsymbol{\beta}_i' \mathbf{C}_d(t,u) \mathbf{x}_t^* \, \mathbf{d}(u,T) \mathbf{d}(u,T)' du \\ &\quad = (\alpha_i + \boldsymbol{\beta}_i' \boldsymbol{\theta}) \mathbf{N}(\tau) + \int_t^T \boldsymbol{\beta}_i' \mathbf{C}_d(t,u) \mathbf{x}_t^* \, \mathbf{d}(u,T) \mathbf{d}(u,T)' du , \end{aligned} \tag{9.10}$$

in which the final equality invokes the definition of the matrix function $(nxn)\mathbf{N}(.,.)$ from appendix A. The integral in (9.10) is simplified as follows:

$$\begin{aligned}
\int_t^T \boldsymbol{\beta}_i\,'\mathbf{C}_d(t,u)\,\mathbf{x}_t^*\,\mathbf{d}(u,T)\mathbf{d}(u,T)'\,du &= \int_t^T \left[\sum_{j=1}^n \beta_{ij}c_j(t,u)\,x_{jt}^*\right]\mathbf{d}(u,T)\mathbf{d}(u,T)'\,du \\
&= \sum_{j=1}^n \left[\beta_{ij}\,x_{jt}^* \int_t^T c_j(t,u)\mathbf{d}(u,T)\mathbf{d}(u,T)'\,du\right] \\
&= \sum_{j=1}^n \beta_{ij}\,x_{jt}^*\,\mathcal{I}_j(t,T) \\
&= \left[\beta_{i1}\,x_{1t}^*\,\mathbf{I}_n,\ldots,\beta_{in}\,x_{nt}^*\,\mathbf{I}_n\right]\begin{bmatrix}\mathcal{I}_1(t,T)\\ \vdots \\ \mathcal{I}_n(t,T)\end{bmatrix} \\
&= \left(\left[\boldsymbol{\beta}_i \odot \mathbf{x}_t^*\right]' \otimes \mathbf{I}_n\right)\mathcal{I}(t,T), \qquad (9.11)
\end{aligned}$$

with $(n^2xn)\,\mathcal{I}(t,T) \equiv \left[\mathcal{I}_1(t,T)',\ldots,\mathcal{I}_n(t,T)'\right]'$, and in which the matrices $(nxn)\ \mathcal{I}_j(t,T)$, $j=1,..,n$ are defined as

$$\mathcal{I}_j(t,T) \equiv \int_t^T c_j(t,u)\mathbf{d}(u,T)\mathbf{d}(u,T)'\,du\,. \qquad (9.12)$$

The pq – th entry of the matrix $\mathcal{I}_j(t,T)$ is given by

$$\begin{aligned}
\left[\mathcal{I}_j(t,T)\right]_{pq} &= \int_t^T c_j(t,u)d_p(u,T)d_q(u,T)du \qquad (9.13)\\
&= \int_t^T \exp[-k_j(u-t)]\left(\frac{1-\exp[-k_p(T-u)]}{k_p}\right)\left(\frac{1-\exp[-k_q(T-u)]}{k_q}\right)du \\
&= \frac{1}{k_pk_q}\left[\int_t^T \exp[-k_j(u-t)]du - \int_t^T \exp[(k_q-k_j)u+k_jt-k_qT]du\right. \\
&\quad \left. -\int_t^T \exp[(k_p-k_j)u+k_jt-k_pT]du + \int_t^T \exp[-k_j(u-t)-(k_p+k_q)(T-u)]du\right] \\
&= \frac{1}{k_pk_q}\left[d_j(\tau) - \int_t^T \exp[(k_q-k_j)u+k_jt-k_qT]du\right. \\
&\quad \left. -\int_t^T \exp[(k_p-k_j)u+k_jt-k_pT]du + \left[\mathcal{H}_j(\tau)\right]_{pq}\right].
\end{aligned}$$

The integrals in this latter expression become

$$\int_t^T \exp[(k_q - k_j)u + k_j t - k_q T]du = \begin{cases} \dfrac{\exp[-k_j\tau] - \exp[-k_q\tau]}{k_q - k_j}; & k_j \neq k_q \\ \tau\ \exp[-k_j\tau]; & k_j = k_q \end{cases} \tag{9.14}$$

$$\int_t^T \exp[(k_p - k_j)u + k_j t - k_p T]du = \begin{cases} \dfrac{\exp[-k_j\tau] - \exp[-k_p\tau]}{k_p - k_j}; & k_j \neq k_p \\ \tau\ \exp[-k_j\tau]; & k_j = k_p. \end{cases}$$

Note that the matrix function $\mathcal{I}_j(t,T);\ j = 1,..,n$ is a function of $T - t = \tau$ only, and not of t or T separately. Therefore we will sometimes write $\mathcal{I}_j(\tau) \equiv \mathcal{I}_j(t,T) = \mathcal{I}_j(0, T-t)$.

Collecting the intermediate results together, yields

$$\begin{aligned} \text{var}_{\mathbb{M}}\left[\int_t^T \mathbf{u}_{t,s}ds \mid \mathcal{F}_t\right] &= \sum_{i=1}^n \Big(\left[\alpha_i + \boldsymbol{\beta}_i{}'\boldsymbol{\theta}\right] \mathbf{N}(\tau) + \Big\{ \left[\boldsymbol{\beta}_i \odot \mathbf{x}_t^*\right]' \otimes \mathbf{I}_n \Big\} \mathcal{I}(\tau) \Big) \odot \boldsymbol{\sigma}_i \boldsymbol{\sigma}_i{}' \\ &= \sum_{i=1}^n \left[\alpha_i + \boldsymbol{\beta}_i{}'\boldsymbol{\theta}\right] \mathbf{N}(\tau) \odot \boldsymbol{\sigma}_i \boldsymbol{\sigma}_i{}' + \sum_{i=1}^n \Big(\Big\{ \left[\boldsymbol{\beta}_i \odot \mathbf{x}_t^*\right]' \otimes \mathbf{I}_n \Big\} \mathcal{I}(\tau) \Big) \odot \boldsymbol{\sigma}_i \boldsymbol{\sigma}_i{}' \\ &= \mathbf{N}(\tau) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' + \sum_{i=1}^n \Big(\Big\{ \left[\boldsymbol{\beta}_i \odot \mathbf{x}_t^*\right]' \otimes \mathbf{I}_n \Big\} \mathcal{I}(\tau) \Big) \odot \boldsymbol{\sigma}_i \boldsymbol{\sigma}_i{}', \end{aligned} \tag{9.15}$$

and thus

$$\begin{aligned} \text{var}_{\mathbb{M}}\left[\int_t^T \mathbf{u}_{t,s}ds\right] &= \mathbb{E}_{\mathbb{M}}\left[\text{var}_{\mathbb{M}}\left(\int_t^T \mathbf{u}_{t,s}ds \mid \mathcal{F}_t\right)\right] + \text{var}_{\mathbb{M}}\left[\mathbb{E}_{\mathbb{M}}\left(\int_t^T \mathbf{u}_{t,s}ds \mid \mathcal{F}_t\right)\right] \\ &= \mathbf{N}(\tau) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'. \end{aligned} \tag{9.16}$$

Returning to the integrated stock variance

Given these moments, from (9.5) it next follows that

$$\mathbb{E}_{\mathbb{M}}\left[\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t\right] = (\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\tau + \boldsymbol{\delta}'\mathbf{D}(\tau)\mathbf{x}_t^*, \tag{9.17}$$

and

$$\mathbb{E}_{\mathbb{M}}\left[\int_t^T \sigma_s^2 ds\right] = \mathbb{E}_{\mathbb{M}}\left[\mathbb{E}_{\mathbb{M}}\left(\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t\right)\right] = (\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\tau. \tag{9.18}$$

(Note again the intuitive logic behind these results.) The conditional variance of the integrated stock variance becomes

$$\begin{aligned} \text{var}_{\mathbb{M}}\left[\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t\right] &= \boldsymbol{\delta}'\text{var}_{\mathbb{M}}\left[\int_t^T \mathbf{u}_{t,s}ds \mid \mathcal{F}_t\right]\boldsymbol{\delta} \\ &= \boldsymbol{\delta}'\left[\mathbf{N}(\tau) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' + \sum_{i=1}^n \Big(\Big\{ \left[\boldsymbol{\beta}_i \odot \mathbf{x}_t^*\right]' \otimes \mathbf{I}_n \Big\} \mathcal{I}(\tau) \Big) \odot \boldsymbol{\sigma}_i \boldsymbol{\sigma}_i{}'\right]\boldsymbol{\delta}, \end{aligned} \tag{9.19}$$

and thus

$$\begin{aligned}\mathrm{var}_{\mathbb{M}}\left[\int_t^T \sigma_s^2 ds\right] &= \mathbb{E}_{\mathbb{M}}\left[\mathrm{var}_{\mathbb{M}}\left(\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t\right)\right] + \mathrm{var}_{\mathbb{M}}\left[\mathbb{E}_{\mathbb{M}}\left(\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t\right)\right] \\ &= \boldsymbol{\delta}'\left[\mathbf{N}(\tau) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' + \mathbf{D}(\tau)\left(\mathbf{J} \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right)\mathbf{D}(\tau)\right]\boldsymbol{\delta} \\ &= \boldsymbol{\delta}'\left[\mathbf{N}(\tau) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}' + \mathbf{D}(\tau)\mathbf{1}\mathbf{1}'\mathbf{D}(\tau) \odot \left(\mathbf{J} \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right)\right]\boldsymbol{\delta} \\ &= \boldsymbol{\delta}'\left[\left(\mathbf{N}(\tau) + \mathbf{D}(\tau)\mathbf{1}\mathbf{1}'\mathbf{D}(\tau) \odot \mathbf{J}\right) \odot \boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right]\boldsymbol{\delta}.\end{aligned} \tag{9.20}$$

Define the *average stock variance* over the interval $[t,T]$ as

$$\bar{\sigma}^2 \equiv \frac{1}{\tau}\int_t^T \sigma_s^2 ds. \tag{9.21}$$

From the results just derived, it follows that the average stock variance converges in mean-square to the unconditional mean stock variance, for $T \to \infty$ and t fixed, which is again intuitively a logical result due to stationarity:

$$\bar{\sigma}^2 \overset{2}{\to} \delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} = \mathbb{E}[\sigma_t^2]. \qquad (\mathbb{M}) \tag{9.22}$$

10. cMGF of the integrated stock variance

This section derives the *conditional moment generating function* (cMGF) of the integrated stock variance by generalizing the results of Duffie and Kan (1996). From its cMGF, any arbitrary moment of $\int_t^T \sigma_s^2 ds$ can be computed (if it exists).

Section 9 derived closed-form expressions for the (conditional) mean and variance of $\int_t^T \sigma_s^2 ds$, for the general multifactor affine SV model. This section shows that the moments of $\int_t^T \sigma_s^2 ds$ can also be derived from solving and differentiating a system of ordinary differential equations (ODEs). However, only in the multifactor OU (the Gaussian) case this yields closed-form expressions. As such, the work in section 9 is not superfluous.

We consider the same SV specification as before (see section 1), but with one important exception. Previously we assumed the speed-of-adjustment matrix to be diagonal (and positive definite); i.e., $\mathbf{K}_d = \mathrm{diag}[k_1,..,k_n]$ with $k_1,..,k_n > 0$. Now we allow the speed-of-adjustment matrix to be an arbitrary (positive definite) matrix $(nxn)\mathbf{K}$. (Positive definiteness seems required for the factor process to be stationary.) Specifically, in this section the factors obey

$$d\mathbf{x}_t = \mathbf{K}(\boldsymbol{\theta} - \mathbf{x}_t)dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t}. \qquad (\mathbb{M}) \tag{10.1}$$

Given this specification, from the results of Duffie and Kan (1996), it follows that [3]

[3] Duffie and Kan (1996) model the *short interest rate* by $r_t = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t$, with the latent factors $\mathbf{x}$ following the same SDE as in (10.1) (but then under the risk-neutral measure $\mathbb{Q}$). Given this set-up, they show that the time-t price $P(t,T)$ of a zero coupon bond maturing at time T is given by

$$P(t,T) = \mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_t^T r_u du\right) \mid \mathcal{F}_t\right] = \exp[A(\tau) + \mathbf{B}(\tau)'\mathbf{x}_t],$$

$$\mathbb{E}_{\mathbb{M}}\left[\exp\left(\int_t^T \sigma_s^2 ds\right) | \mathcal{F}_t\right] = \exp\left[A_1(\tau) + \mathbf{B}_1(\tau)'\mathbf{x}_t\right], \tag{10.2}$$

in which $A_1(.)$ is some deterministic function of $\tau \equiv T - t$ with $T \geq t$, and $\mathbf{B}_1(.)$ is an $(nx1)$ deterministic vector function of τ, which satisfy the following system of Ricatti ODEs:

$$\begin{aligned} \frac{dA_1(\tau)}{d\tau} &= \boldsymbol{\theta}'\mathbf{K}'\mathbf{B}_1(\tau) + \tfrac{1}{2}\sum_{i=1}^{n}\left[\boldsymbol{\Sigma}'\mathbf{B}_1(\tau)\right]_i^2 \alpha_i + \delta_0 \\ \frac{d\mathbf{B}_1(\tau)}{d\tau} &= -\mathbf{K}'\mathbf{B}_1(\tau) + \tfrac{1}{2}\sum_{i=1}^{n}\left[\boldsymbol{\Sigma}'\mathbf{B}_1(\tau)\right]_i^2 \boldsymbol{\beta}_i + \boldsymbol{\delta}, \end{aligned} \tag{10.3}$$

with boundary conditions $A_1(0) = 0$ and $\mathbf{B}_1(0) = \mathbf{0}$. In the OU (Gaussian) case, in which the volatility function of the factor SDE is deterministic, closed-form expressions for $A_1(\tau)$ and $\mathbf{B}_1(\tau)$ exist; see section 13. In other cases, the system can be solved numerically.

We now further generalize these results. In particular, our interest is in the cMGF of the integrated stock variance, denoted by $\phi_{\tau\bar{\sigma}^2|\mathcal{F}_t}(\mathrm{m})$:

$$\phi_{\tau\bar{\sigma}^2|\mathcal{F}_t}(\mathrm{m}) \equiv \mathbb{E}_{\mathbb{M}}\left[\exp\left(\mathrm{m}\int_t^T \sigma_s^2 ds\right) | \mathcal{F}_t\right], \tag{10.4}$$

in which m is an arbitrary multiple of the integrated stock variance. Given (10.2), consider the following reparametrization: $\ddot{\sigma}_t^2 \equiv \mathrm{m}\sigma_t^2 = (\mathrm{m}\delta_0) + (\mathrm{m}\boldsymbol{\delta})'\mathbf{x}_t \equiv \ddot{\delta}_0 + \ddot{\boldsymbol{\delta}}'\mathbf{x}_t$ with $\ddot{\delta}_0 \equiv \mathrm{m}\delta_0$ and $\ddot{\boldsymbol{\delta}} \equiv \mathrm{m}\boldsymbol{\delta}$. It is then immediately clear that it holds that

$$\phi_{\tau\bar{\sigma}^2|\mathcal{F}_t}(\mathrm{m}) = \mathbb{E}_{\mathbb{M}}\left[\exp\left(\mathrm{m}\int_t^T \sigma_s^2 ds\right) | \mathcal{F}_t\right] = \exp\left[A_{\mathrm{m}}(\tau) + \mathbf{B}_{\mathrm{m}}(\tau)'\mathbf{x}_t\right], \tag{10.5}$$

in which $A_{\mathrm{m}}(.)$ and $(nx1)$ $\mathbf{B}_{\mathrm{m}}(.)$ are deterministic functions of $\tau \equiv T - t$, which depend on the choice of m, and which satisfy the following system of Ricatti ODEs

$$\begin{aligned} \frac{dA_{\mathrm{m}}(\tau)}{d\tau} &= \boldsymbol{\theta}'\mathbf{K}'\mathbf{B}_{\mathrm{m}}(\tau) + \tfrac{1}{2}\sum_{i=1}^{n}\left[\boldsymbol{\Sigma}'\mathbf{B}_{\mathrm{m}}(\tau)\right]_i^2 \alpha_i + \mathrm{m}\delta_0 \\ \frac{d\mathbf{B}_{\mathrm{m}}(\tau)}{d\tau} &= -\mathbf{K}'\mathbf{B}_{\mathrm{m}}(\tau) + \tfrac{1}{2}\sum_{i=1}^{n}\left[\boldsymbol{\Sigma}'\mathbf{B}_{\mathrm{m}}(\tau)\right]_i^2 \boldsymbol{\beta}_i + \mathrm{m}\,\boldsymbol{\delta}, \end{aligned} \tag{10.6}$$

with boundary conditions $A_{\mathrm{m}}(0) = 0$ and $\mathbf{B}_{\mathrm{m}}(0) = \mathbf{0}$. (The reason is that the parameters $\mathbf{K}, \boldsymbol{\theta}, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_i$, $i = 1,..,n$ are not affected by the reparametrization.)

From the cMGF (10.5), any arbitrary moment (both conditional and unconditional) of $\int_t^T \sigma_s^2 ds$ can be computed (if it exists). Specifically, the i-th conditional moment follows from evaluating

---

in which $\tau \equiv T - t$, and $A(.)$ and $\mathbf{B}(.)$ satisfy a similar -but not exactly equal- system of ODEs as in (10.3). The ODE system (10.3) is obtained from a reparametrization. Indeed, the inspiration for our multifactor SV model arose when reading their and related papers in the interest rate literature.

$$\mathbb{E}_{\mathbb{M}}\left[\left(\int_t^T \sigma_s^2 ds\right)^i \mid \mathcal{F}_t\right] = \left.\frac{d^i \phi_{\tau\bar{\sigma}^2|\mathcal{F}_t}(\mathrm{m})}{d\mathrm{m}^i}\right|_{\mathrm{m}=0} = \phi^{(i)}_{\tau\bar{\sigma}^2|\mathcal{F}_t}(0). \qquad (10.7)$$

By the law of iterated expectations, the i-th unconditional moment can next be obtained from

$$\mathbb{E}_{\mathbb{M}}\left[\left(\int_t^T \sigma_s^2 ds\right)^i\right] = \mathbb{E}_{\mathbb{M}}\left[\left.\frac{d^i \phi_{\tau\bar{\sigma}^2|\mathcal{F}_t}(\mathrm{m})}{d\mathrm{m}^i}\right|_{\mathrm{m}=0}\right] = \mathbb{E}_{\mathbb{M}}\left[\phi^{(i)}_{\tau\bar{\sigma}^2|\mathcal{F}_t}(0)\right]. \qquad (10.8)$$

Computing the derivatives of the cMGF with respect to m involves differentiating the system of ODEs (10.6) with respect to m. This does not seem to be a trivial exercise, and seems in general only possible numerically, except in the Gaussian case (see section (13)).

11. Moments of the stock returns

Consider the multifactor affine SV model for a stock price, as covered in the main text (but now stated under measure $\mathbb{M}$ to maintain consistency within this appendix). Assuming a constant stock price drift of $\mu_t = \mu \ \forall t$, it reads:

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma_t S_t dW_{S,t} \qquad (\mathbb{M}) \qquad (11.1)\\ \sigma_t^2 &= \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t \\ d\mathbf{x}_t &= \mathbf{K}_d(\boldsymbol{\theta} - \mathbf{x}_t)dt + \boldsymbol{\Sigma}\boldsymbol{\Lambda}_t d\mathbf{W}_{x,t}. \end{aligned}$$

Uncertainty is resolved by the $(n+1)$-dimensional standard Brownian motion $\{\mathbf{W}_t; t \geq 0\}$, given by $\mathbf{W}_t = (W_{S,t}, \mathbf{W}_{x,t}')' = (W_{S,t}, W_{1t}, .., W_{nt})'$, which is defined on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{M})$, satisfying the usual conditions. (Recall that the leverage effect is not modeled; this facilitates the calculations considerably.)

This section derives the cMGF of the *logreturns* on stock S. Moreover, we indicate how the exact moments of the *relative* stock returns can be obtained. This appears much more involved than for the logreturns.

Deriving the cMGF of the logreturns
From Itô's lemma, the log stock price follows the SDE

$$d\ln S_t = (\mu - \tfrac{1}{2}\sigma_t^2)dt + \sigma_t dW_{S,t}. \qquad (\mathbb{M}) \qquad (11.2)$$

The logreturn $R_{t+\delta t}$ over the interval $[t, t+\delta t]$ equals

$$R_{t+\delta t} \equiv \ln\frac{S_{t+\delta t}}{S_t} = \mu\,\delta t - \tfrac{1}{2}\int_t^{t+\delta t} \sigma_u^2 du + \int_t^{t+\delta t} \sigma_u dW_{S,u}. \qquad (\mathbb{M}) \qquad (11.3)$$

The cMGF of the logreturn, denoted by $\phi_{R_{t+\delta t}|\mathcal{F}_t}(\mathfrak{m})$, follows as

$$\phi_{R_{t+\delta t}|\mathcal{F}_t}(\mathfrak{m}) \equiv \mathbb{E}_{\mathbb{M}}\left[\exp(\mathfrak{m}R_{t+\delta t}) \mid \mathcal{F}_t\right] \qquad (11.4)$$

$$= \mathbb{E}_{\mathbb{M}}\left[\exp\left(\mathrm{m}\mu\,\delta t - \tfrac{1}{2}\mathrm{m}\int_t^{t+\delta t}\sigma_u^2 du + \mathrm{m}\int_t^{t+\delta t}\sigma_u dW_{S,u}\right) \mid \mathcal{F}_t\right]$$

$$= \exp(\mathrm{m}\mu\delta t)\,\mathbb{E}_{\mathbb{M}}\left\{\mathbb{E}_{\mathbb{M}}\left[\exp\left(-\tfrac{1}{2}\mathrm{m}\int_t^{t+\delta t}\sigma_u^2 du + \mathrm{m}\int_t^{t+\delta t}\sigma_u dW_{S,u}\right) \mid \{\sigma_u\}\right] \mid \mathcal{F}_t\right\}$$

$$= \exp(\mathrm{m}\mu\delta t)\,\mathbb{E}_{\mathbb{M}}\left\{\exp\left(-\tfrac{1}{2}\mathrm{m}\int_t^{t+\delta t}\sigma_u^2 du\right)\mathbb{E}_{\mathbb{M}}\left[\exp\left(\mathrm{m}\int_t^{t+\delta t}\sigma_u dW_{S,u}\right) \mid \{\sigma_u\}\right] \mid \mathcal{F}_t\right\},$$

in which the third equality follows from the law of iterated expectations. (The condition $\mid \{\sigma_u\}$ stands for conditioning on the $\sigma -$field generated by the volatility over the interval $[t, t+\delta t]$.) Now,

$$\int_t^{t+\delta t}\sigma_u dW_{S,u} \mid \{\sigma_u\} \sim \mathcal{N}\left(0, \int_t^{t+\delta t}\sigma_u^2 du\right). \qquad (\mathbb{M}) \qquad (11.5)$$

By invoking the moment generating function of a Gaussian variate, it thus holds that

$$\mathbb{E}_{\mathbb{M}}\left[\exp\left(\mathrm{m}\int_t^{t+\delta t}\sigma_u dW_{S,u}\right) \mid \{\sigma_u\}\right] = \exp\left(\tfrac{1}{2}\mathrm{m}^2\int_t^{t+\delta t}\sigma_u^2 du\right). \qquad (11.6)$$

The cMGF of the logreturn $R_{t+\delta t}$ over $[t, t+\delta t]$ then becomes

$$\phi_{R_{t+\delta t}|\mathcal{F}_t}(\mathrm{m}) = \exp(\mathrm{m}\mu\delta t)\,\mathbb{E}_{\mathbb{M}}\left[\exp\left(\tfrac{1}{2}\mathrm{m}(\mathrm{m}-1)\int_t^{t+\delta t}\sigma_u^2 du\right) \mid \mathcal{F}_t\right] \qquad (11.7)$$

$$= \exp\left[\mathrm{m}\mu\delta t + A_{\frac{1}{2}\mathrm{m}(\mathrm{m}-1)}(\delta t) + \mathbf{B}_{\frac{1}{2}\mathrm{m}(\mathrm{m}-1)}(\delta t)'\mathbf{x}_t\right],$$

in which $A_{\mathrm{m}(\mathrm{m}-1)/2}(.)$ and $(nx1)\mathbf{B}_{\mathrm{m}(\mathrm{m}-1)/2}(.)$ satisfy the ODE system (10.6) with m replaced by $\mathrm{m}(\mathrm{m}-1)/2$ and τ by δt. (Note that this result hinges on the cMGF of the integrated stock variance, which we derived in section 10.)

Given (11.7), the i-th conditional moment of the log stock return follows as

$$\mathbb{E}_{\mathbb{M}}\left[R_{t+\delta t}^i \mid \mathcal{F}_t\right] = \left.\frac{d^i\phi_{R_{t+\delta t}|\mathcal{F}_t}(\mathrm{m})}{d\mathrm{m}^i}\right|_{\mathrm{m}=0} = \phi_{R_{t+\delta t}|\mathcal{F}_t}^{(i)}(0). \qquad (11.8)$$

By the law of iterated expectations, the i-th unconditional moment can subsequently be computed from

$$\mathbb{E}_{\mathbb{M}}\left[R_{t+\delta t}^i\right] = \mathbb{E}_{\mathbb{M}}\left[\left.\frac{d^i\phi_{R_{t+\delta t}|\mathcal{F}_t}(\mathrm{m})}{d\mathrm{m}^i}\right|_{\mathrm{m}=0}\right] = \mathbb{E}_{\mathbb{M}}\left[\phi_{R_{t+\delta t}|\mathcal{F}_t}^{(i)}(0)\right]. \qquad (11.9)$$

The derivatives of the cMGF with respect to m can be determined in a recursive-type fashion. In particular:

$$\phi'(\mathfrak{m}) = \phi(\mathfrak{m})\left[\mu\,\delta t + \frac{dA}{d\mathfrak{m}} + \mathbf{x}_t{}'\frac{d\mathbf{B}}{d\mathfrak{m}}\right] \tag{11.10}$$
$$\phi''(\mathfrak{m}) = \phi'(\mathfrak{m})\left[\mu\,\delta t + \frac{dA}{d\mathfrak{m}} + \mathbf{x}_t{}'\frac{d\mathbf{B}}{d\mathfrak{m}}\right] + \phi(\mathfrak{m})\left[\frac{d^2A}{d\mathfrak{m}^2} + \mathbf{x}_t{}'\frac{d^2\mathbf{B}}{d\mathfrak{m}^2}\right]$$
$$\phi'''(\mathfrak{m}) = ...,$$

etcetera, in which the following abbreviations are used for ease of notation: $\phi(\mathfrak{m}) \equiv \phi_{R_{t+\delta t}|\mathcal{F}_t}(\mathfrak{m})$, $dA/d\mathfrak{m} \equiv dA_{\mathfrak{m}(\mathfrak{m}-1)/2}(\delta t)/d\mathfrak{m}$, $d\mathbf{B}/d\mathfrak{m} \equiv d\mathbf{B}_{\mathfrak{m}(\mathfrak{m}-1)/2}(\delta t)/d\mathfrak{m}$, and similarly for the higher-order derivatives.

As closed-form expressions exist for $A_{\mathfrak{m}(\mathfrak{m}-1)/2}(.)$ and $(nx1)\mathbf{B}_{\mathfrak{m}(\mathfrak{m}-1)/2}(.)$ in the Gaussian, OU case (see section 13), the moments of the logreturns have closed-from expressions in the OU case. In the non-OU case, the moments of the logreturns can in principle numerically be computed, though this does not seem to be a trivial exercise.

The moments of the relative stock returns
The *relative* stock return over the interval $[t, t+\delta t]$ is given by $r_{t+\delta t} \equiv (S_{t+\delta t} - S_t)/S_t$. Integrating the log stock price increments given in (11.2) over $[t, t+\delta t]$, and transforming back yields

$$S_{t+\delta t} = S_t \exp\left[\mu\,\delta t - \tfrac{1}{2}\int_t^{t+\delta t}\sigma_u^2 du + \int_t^{t+\delta t}\sigma_u dW_{S,u}\right]. \tag{11.11}$$

The i – th power of the relative return can thus be written as

$$r_{t+\delta t}^i = \left(\exp\left[\mu\,\delta t - \tfrac{1}{2}\int_t^{t+\delta t}\sigma_u^2 du + \int_t^{t+\delta t}\sigma_u dW_{S,u}\right] - 1\right)^i, \tag{11.12}$$

which may be expanded. Having done so, the moments can in principle be computed by following a similar analysis as for the logreturns. Nonetheless, it is clear that this is much more involved than for the logreturns. We therefore do not further pursue this here.

From a first-order Taylor series expansion, $R_{t+\delta t} = \ln(S_{t+\delta t}/S_t) \approx (S_{t+\delta t} - S_t)/S_t = r_{t+\delta t}$. Logreturns and relative returns will therefore be close in value for small δt. As such, we expect their moments to be close in value as well.

12. Correlation between non-overlapping logreturns

This section investigates the correlation between non-overlapping logreturns, assuming the same SV stock price model as in section 11.

First, a remark: If the stock volatility would vary in a deterministic way instead, then it readily follows from the independent increments property of Brownian motion that non-overlapping logreturns (and relative returns) are uncorrelated. (Remember that the *Efficient Markets Hypothesis* implies that non-overlapping stock returns should be uncorrelated.)

Given that there is SV, does this still hold? Based on an Euler discretization of the (log-) stock price SDE, and pretending this discretization to be exact, it can be shown that non-overlapping stock returns are uncorrelated indeed. As the Euler discretization is not exact however, it may well be the case that the true stock returns *are* correlated. If they are found to be correlated, we nevertheless expect this correlation to be small, given the Euler discretization result.

Let us examine the exact correlation. Consider the covariance between the non-overlapping logreturns $R_{t+\delta t}$ and $R_{t+(1-p)\delta t}$ for $p = 1, 2, \ldots$ under measure $\mathbb{M}$:

$$\begin{aligned}
\operatorname{cov}\left[R_{t+\delta t}, R_{t+(1-p)\delta t}\right] &= \operatorname{cov}\left[-\tfrac{1}{2}\int_t^{t+\delta t}\sigma_u^2 du + \int_t^{t+\delta t}\sigma_u dW_{S,u}\,, -\tfrac{1}{2}\int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u^2 du + \int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u dW_{S,u}\right] \\
&= -\tfrac{1}{2}\operatorname{cov}\left[\int_t^{t+\delta t}\sigma_u^2 du\,, \int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u dW_{S,u}\right] - \tfrac{1}{2}\operatorname{cov}\left[\int_t^{t+\delta t}\sigma_u dW_{S,u}\,, \int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u^2 du\right] \\
&\quad + \operatorname{cov}\left[\int_t^{t+\delta t}\sigma_u dW_{S,u}\,, \int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u dW_{S,u}\right] + \tfrac{1}{4}\operatorname{cov}\left[\int_t^{t+\delta t}\sigma_u^2 du\,, \int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u^2 du\right]. \quad (12.1)
\end{aligned}$$

(All expectations and covariances are taken under measure $\mathbb{M}$ here and below. We omit this for notational simplicity.) We tackle each covariance term in (12.1) individually. As Itô integrals have expectation zero, we find for the first covariance

$$\begin{aligned}
\operatorname{cov}\left[\int_t^{t+\delta t}\sigma_u^2 du\,, \int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u dW_{S,u}\right] &= \mathbb{E}\left[\int_t^{t+\delta t}\sigma_u^2 du \int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u dW_{S,u}\right] \qquad (12.2) \\
&= \mathbb{E}\left[\mathbb{E}\left(\int_t^{t+\delta t}\sigma_u^2 du \int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u dW_{S,u} \mid \{\sigma_u\}\right)\right] \\
&= \mathbb{E}\left[\int_t^{t+\delta t}\sigma_u^2 du \; \mathbb{E}\left(\int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u dW_{S,u} \mid \{\sigma_u\}\right)\right] = \mathbb{E}\left[\int_t^{t+\delta t}\sigma_u^2 du * 0\right] = 0\,,
\end{aligned}$$

in which $\mid \{\sigma_u\}$ is short-hand notation for the condition $\mid \{\sigma_u ; t - p\delta t \le u \le t + \delta t\}$. A similar argument shows that the second covariance in (12.1) equals zero as well. The third covariance equals

$$\begin{aligned}
\operatorname{cov}\left[\int_t^{t+\delta t}\sigma_u dW_{S,u}\,, \int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u dW_{S,u}\right] &= \mathbb{E}\left[\int_t^{t+\delta t}\sigma_u dW_{S,u} \int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u dW_{S,u}\right] \qquad (12.3) \\
&= \mathbb{E}\left[\mathbb{E}\left(\int_t^{t+\delta t}\sigma_u dW_{S,u} \int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u dW_{S,u} \mid \mathcal{F}_t\right)\right] \\
&= \mathbb{E}\left[\int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u dW_{S,u} \; \mathbb{E}\left(\int_t^{t+\delta t}\sigma_u dW_{S,u} \mid \mathcal{F}_t\right)\right] = \mathbb{E}\left[\int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u dW_{S,u} * 0\right] = 0\,,
\end{aligned}$$

in which the last equality exploits the martingale property of Itô stochastic integrals. As $\sigma_t^2 = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t$, the last covariance term in (12.1) equals

$$\operatorname{cov}\left[\int_t^{t+\delta t}\sigma_u^2 du, \int_{t-p\delta t}^{t+(1-p)\delta t}\sigma_u^2 du\right] = \boldsymbol{\delta}'\operatorname{cov}\left[\int_t^{t+\delta t}\mathbf{x}_u du, \int_{t-p\delta t}^{t+(1-p)\delta t}\mathbf{x}_u du\right]\boldsymbol{\delta}\,. \qquad (12.4)$$

As $\mathbb{E}_{\mathbb{M}}[\int_t^{t+\delta t} \mathbf{x}_u du] = \boldsymbol{\theta}\delta t$, the covariance in (12.4) becomes

$$\operatorname{cov}\left[\int_t^{t+\delta t} \mathbf{x}_u du, \int_{t-p\delta t}^{t+(1-p)\delta t} \mathbf{x}_u du\right] = -\boldsymbol{\theta}\boldsymbol{\theta}'\delta t^2 + \mathbb{E}\left[\int_t^{t+\delta t} \mathbf{x}_u du \int_{t-p\delta t}^{t+(1-p)\delta t} \mathbf{x}_u' du\right] \tag{12.5}$$

$$= -\boldsymbol{\theta}\boldsymbol{\theta}'\delta t^2 + \mathbb{E}\left[\mathbb{E}\left(\int_t^{t+\delta t} \mathbf{x}_u du \mid \mathcal{F}_{t+(1-p)\delta t}\right) \int_{t-p\delta t}^{t+(1-p)\delta t} \mathbf{x}_u' du\right],$$

in which the second equality invokes the law of iterated expectations. Using result (3.2), the inner expectation in this latter expression equals

$$\mathbb{E}\left(\int_t^{t+\delta t} \mathbf{x}_u du \mid \mathcal{F}_{t+(1-p)\delta t}\right) = \int_t^{t+\delta t} \mathbb{E}\left[\mathbf{x}_u \mid \mathcal{F}_{t+(1-p)\delta t}\right] du \tag{12.6}$$

$$= \int_t^{t+\delta t} \left(\boldsymbol{\theta} + \exp[-\mathbf{K}_d(u - t + (p-1)\delta t)]\, \mathbf{x}^*_{t+(1-p)\delta t}\right) du$$

$$= \boldsymbol{\theta}\,\delta t + \left(\int_t^{t+\delta t} \exp[-\mathbf{K}_d(u-t)] du\right) \exp[\mathbf{K}_d(1-p)\delta t]\, \mathbf{x}^*_{t+(1-p)\delta t}$$

$$= \boldsymbol{\theta}\,\delta t + \mathbf{D}(\delta t)\, \exp[\mathbf{K}_d(1-p)\delta t]\, \mathbf{x}^*_{t+(1-p)\delta t},$$

with (recall) $\mathbf{x}_t^* = \mathbf{x}_t - \boldsymbol{\theta}$. Hence,

$$\mathbb{E}\left[\mathbb{E}\left(\int_t^{t+\delta t} \mathbf{x}_u du \mid \mathcal{F}_{t+(1-p)\delta t}\right) \int_{t-p\delta t}^{t+(1-p)\delta t} \mathbf{x}_u' du\right] \tag{12.7}$$

$$= \mathbb{E}\left[\left(\boldsymbol{\theta}\,\delta t + \mathbf{D}(\delta t)\, \exp[\mathbf{K}_d(1-p)\delta t]\, \mathbf{x}^*_{t+(1-p)\delta t}\right) \int_{t-p\delta t}^{t+(1-p)\delta t} \mathbf{x}_u' du\right]$$

$$= \boldsymbol{\theta}\boldsymbol{\theta}'\delta t^2 + \mathbf{D}(\delta t)\, \exp[\mathbf{K}_d(1-p)\delta t]\; \mathbb{E}\left[\mathbf{x}^*_{t+(1-p)\delta t} \int_{t-p\delta t}^{t+(1-p)\delta t} \mathbf{x}_u' du\right].$$

Collecting the intermediate results together, it holds that

$$\operatorname{cov}\left[\int_t^{t+\delta t} \mathbf{x}_u du, \int_{t-p\delta t}^{t+(1-p)\delta t} \mathbf{x}_u du\right] = \mathbf{D}(\delta t)\, \exp[\mathbf{K}_d(1-p)\delta t]\; \mathbb{E}\left[\mathbf{x}^*_{t+(1-p)\delta t} \int_{t-p\delta t}^{t+(1-p)\delta t} \mathbf{x}_u' du\right], \tag{12.8}$$

such that the last covariance term in (12.1) becomes

$$\operatorname{cov}\left[\int_t^{t+\delta t} \sigma_u^2 du, \int_{t-p\delta t}^{t+(1-p)\delta t} \sigma_u^2 du\right] = \boldsymbol{\delta}'\mathbf{D}(\delta t)\, \exp[\mathbf{K}_d(1-p)\delta t]\; \mathbb{E}\left[\mathbf{x}^*_{t+(1-p)\delta t} \int_{t-p\delta t}^{t+(1-p)\delta t} \mathbf{x}_u' du\right]\boldsymbol{\delta}. \tag{12.9}$$

Finally, for the covariance between the non-overlapping logreturns $R_{t+\delta t}$ and $R_{t+(1-p)\delta t}$ for $p = 1, 2, \ldots$ under measure $\mathbb{M}$, we find that

$$\text{cov}\left[R_{t+\delta t}, R_{t+(1-p)\delta t}\right] = \tfrac{1}{4}\boldsymbol{\delta}'\mathbf{D}(\delta t)\ \exp[\mathbf{K}_d(1-p)\delta t]\ \mathbb{E}\left[\mathbf{x}^*_{t+(1-p)\delta t} \int_{t-p\delta t}^{t+(1-p)\delta t} \mathbf{x}_u\,'du\right]\boldsymbol{\delta}. \quad (12.10)$$

It is not obvious how to further simplify the expectation in (12.10) in an exact way. To examine its approximate magnitude, notice that for small δt

$$\begin{aligned}\mathbb{E}\left[\mathbf{x}^*_{t+(1-p)\delta t} \int_{t-p\delta t}^{t+(1-p)\delta t} \mathbf{x}_u\,'du\right] &\approx \mathbb{E}\left[\mathbf{x}^*_{t+(1-p)\delta t}\ \mathbf{x}_{t+(1-p)\delta t}\,'\delta t\right] \quad (12.11)\\ &= \delta t\ \text{cov}\left[\mathbf{x}^*_{t+(1-p)\delta t}, \mathbf{x}_{t+(1-p)\delta t}\right]\\ &= \delta t\ \text{cov}\left[\mathbf{x}_{t+(1-p)\delta t}, \mathbf{x}_{t+(1-p)\delta t}\right] = \left(\mathbf{J}\odot\boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}'\right)\delta t,\end{aligned}$$

in which the $\approx$ follows from approximating the Riemann integral in (12.11) by the product of the integrand at the end of the interval, times the length of the interval, δt.

What do we learn from these computations? First, non-overlapping logreturns *do* correlate, due to SV. Second, the autocorrelation in the returns dies out exponentially, due to the appearance of the matrix $\exp[\mathbf{K}_d(1-p)\delta t]$ in (12.10), in which $\mathbf{K}_d$ is positive definite. It holds that $\lim_{p\to\infty}\ \text{cov}[R_{t+\delta t}, R_{t+(1-p)\delta t}] = 0$. Third, as for small δt, it holds that $\mathbf{D}(\delta t)\approx\mathbf{0}$, $\exp[\mathbf{K}_d(1-p)\delta t]\approx\mathbf{I}_n$ and $(\mathbf{J}\odot\boldsymbol{\Sigma}\mathbf{M}_d\boldsymbol{\Sigma}')\delta t\approx\mathbf{0}$, this autocorrelation is close to zero for small δt. So for δt small, non-overlapping stock returns are virtually uncorrelated.

13. Special case: *n*-factor OU SV

This section states the most important results obtained in the previous sections for the OU SV special case of the multifactor affine SV model. The OU SV case is obtained by imposing

$$\boldsymbol{\alpha} = \mathbf{1},\quad \boldsymbol{\beta}_i = \mathbf{0},\ i = 1,..,n, \quad (13.1)$$

in the general multifactor model (1.1)-(1.4). The factors thus follow the SDE

$$d\mathbf{x}_t = \mathbf{K}_d(\boldsymbol{\theta}-\mathbf{x}_t)dt + \boldsymbol{\Sigma}d\mathbf{W}_{x,t} \qquad (\mathbb{M}) \qquad (13.2)$$

under $\mathbb{M}$, as in the OU case $\boldsymbol{\Lambda}_t = \mathbf{I}_n\ \forall t$. Matrix $\mathbf{M}_d$ reduces to the *n*-dimensional unity matrix as well: $\mathbf{M}_d = \mathbf{I}_n$. Imposing these restrictions leads to the following results.

First, the n-factor OU process is a Gaussian process. The factors are normally distributed. To illustrate this, consider the disturbance term $\mathbf{u}_{t,s}$ (defined in (3.4)). In the OU case it equals

$$\mathbf{u}_{t,s} = \int_t^s \exp[-\mathbf{K}_d(s-u)]\boldsymbol{\Sigma}\,d\mathbf{W}_{x,u}. \qquad (\mathbb{M}) \qquad (13.3)$$

As the integrand of the Itô integral in (13.3) is non-stochastic, $\mathbf{u}_{t,s}$ is Gaussian unconditionally. However, as the information filtration $\mathcal{F}_t$ does not provide any information about $\mathbf{u}_{t,s}$, $\mathbf{u}_{t,s}$ is conditionally Gaussian as well, with the same mean and variance. Given this observation, and invoking (6.4), (6.5) and (6.6), it holds that

$$\mathbf{u}_{t,s} \sim \mathbf{u}_{t,s}\mid\mathcal{F}_t \sim \mathcal{N}\left[\mathbf{0},\ \mathbf{G}(t,s)\odot\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right]. \qquad (\mathbb{M}) \qquad (13.4)$$

Hence, in the OU SV case, the conditional and unconditional distribution of $\mathbf{u}_{t,s}$ are the same. Given (13.4) and from (3.2) and (3.3), the conditional distribution of the factors is given by

$$\mathbf{x}_s \mid \mathcal{F}_t \sim \mathcal{N}\left[\boldsymbol{\theta} + \exp[-\mathbf{K}_d(s-t)]\left(\mathbf{x}_t - \boldsymbol{\theta}\right);\ \mathbf{G}(t,s) \odot \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right]. \qquad (\mathbb{M}) \qquad (13.5)$$

And what about the unconditional, or *invariant* (also called *stationary*) *distribution* of the factors? Suppose the initial value $\mathbf{x}_0$ of the process is drawn from the unconditional factor distribution. Alternatively, suppose that the factor process $\{\mathbf{x}_t; t \geq 0\}$ has been evolving for a long time yet (or, to be precise, started evolving in the infinite far past). As $\{\mathbf{x}_t; t \geq 0\}$ is stationary, the initial value $\mathbf{x}_0$ no longer influences the current distribution of the factors. Then, from (3.9), (4.11) and (13.5), it is clear that

$$\mathbf{x}_s \sim \mathcal{N}\left[\boldsymbol{\theta};\ \mathbf{J} \odot \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right]. \qquad (\mathbb{M}) \qquad (13.6)$$

(The mean and variance of $\mathbf{x}_s$ coincide with what (3.2), (3.7), (4.10) and (5.7) yield in the OU case, as it should of course.) From (5.8), the correlation matrix of the factors becomes

$$\text{corr}_{\mathbb{M}}\left[\mathbf{x}_s\right] = \left[\mathbf{I}_n \odot \mathbf{J} \odot \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right]^{-1/2}\left[\mathbf{J} \odot \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right]\left[\mathbf{I}_n \odot \mathbf{J} \odot \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right]^{-1/2}. \qquad (13.7)$$

As the instantaneous stock variance $\sigma_t^2 = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t$ is an affine function of the factors, its conditional distribution is Gaussian as well,

$$\sigma_s^2 \mid \mathcal{F}_t \sim \mathcal{N}\left[\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta} + \boldsymbol{\delta}'\exp[-\mathbf{K}_d(s-t)]\left(\mathbf{x}_t - \boldsymbol{\theta}\right);\ \boldsymbol{\delta}'\left(\mathbf{G}(t,s) \odot \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right)\boldsymbol{\delta}\right]. \ (\mathbb{M}) \qquad (13.8)$$

This also holds for the invariant distribution of the stock variance:

$$\sigma_s^2 \sim \mathcal{N}\left[\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta};\ \boldsymbol{\delta}'\left(\mathbf{J} \odot \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right)\boldsymbol{\delta}\right]. \qquad (\mathbb{M}) \qquad (13.9)$$

(An obvious drawback of the n-factor OU SV model is that the stock variance can become negative, which is theoretically inconsistent. We refer to the main text for a discussion; see section 2 of chapter II and section 1 of chapter III.) In the OU SV case, the integrated disturbance term (9.6) becomes

$$\int_t^T \mathbf{u}_{t,s} ds = \int_t^T \mathbf{D}(u,T)\boldsymbol{\Sigma} d\mathbf{W}_{x,u}\ . \qquad (\mathbb{M}) \qquad (13.10)$$

Its conditional and unconditional distributions coincide, and is Gaussian:

$$\int_t^T \mathbf{u}_{t,s} ds \sim \int_t^T \mathbf{u}_{t,s} ds \mid \mathcal{F}_t \sim \mathcal{N}\left[\mathbf{0};\ \mathbf{N}(\tau) \odot \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right]. \qquad (\mathbb{M}) \qquad (13.11)$$

As the integrated stock variance (9.5) is a linear combination of the integrated disturbance term, it is both conditionally and unconditionally normally distributed as well:

$$\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t \sim \mathcal{N}\left[(\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\tau + \boldsymbol{\delta}'\mathbf{D}(\tau)\left(\mathbf{x}_t - \boldsymbol{\theta}\right);\ \boldsymbol{\delta}'\left(\mathbf{N}(\tau) \odot \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right)\boldsymbol{\delta}\right], \ (\mathbb{M}) \qquad (13.12)$$

$$\int_t^T \sigma_s^2 ds \sim \mathcal{N}\left[(\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\tau\,;\;\; \boldsymbol{\delta}'\left[\left(\mathbf{N}(\tau) + \mathbf{D}(\tau)\mathbf{1}\mathbf{1}'\mathbf{D}(\tau) \odot \mathbf{J}\right) \odot \boldsymbol{\Sigma\Sigma}'\right]\boldsymbol{\delta}\right]. \quad (\mathbb{M}) \qquad (13.13)$$

When discussing the cMGF of $\int_t^T \sigma_s^2 ds$, we mentioned that the system of Ricatti ODEs (10.6) (which is based on (10.1) with arbitrary $\mathbf{K}$) yields closed-form expressions for $A_{\mathrm{m}}(.)$ and $\mathbf{B}_{\mathrm{m}}(.)$ in the Gaussian, OU case. Having derived the conditional distribution of $\int_t^T \sigma_s^2 ds$ in (13.12) based on (13.2) with diagonal $\mathbf{K}_d$, obtaining the functions $A_{\mathrm{m}}(.)$ and $\mathbf{B}_{\mathrm{m}}(.)$ in (10.6) for the n-factor OU SV case with diagonal $\mathbf{K} = \mathbf{K}_d$ is easy. Namely, from the MGF of a Gaussian variate,

$$\begin{aligned}\phi_{\tau\bar{\sigma}^2|\mathcal{F}_t}(\mathrm{m}) &= \mathbb{E}_{\mathbb{M}}\left[\exp\left(\mathrm{m}\int_t^T \sigma_s^2 ds\right) \mid \mathcal{F}_t\right] \qquad (13.14)\\ &= \exp\left[\mathrm{m}(\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\tau + \mathrm{m}\,\boldsymbol{\delta}'\mathbf{D}(\tau)(\mathbf{x}_t - \boldsymbol{\theta}) + \tfrac{1}{2}\mathrm{m}^2\,\boldsymbol{\delta}'\left(\mathbf{N}(\tau) \odot \boldsymbol{\Sigma\Sigma}'\right)\boldsymbol{\delta}\right]\end{aligned}$$

which is indeed of the exponential-affine form, with

$$A_{\mathrm{m}}(\tau) = \mathrm{m}\,(\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\tau - \mathrm{m}\,\boldsymbol{\delta}'\mathbf{D}(\tau)\boldsymbol{\theta} + \tfrac{1}{2}\mathrm{m}^2\,\boldsymbol{\delta}'\left(\mathbf{N}(\tau) \odot \boldsymbol{\Sigma\Sigma}'\right)\boldsymbol{\delta}, \qquad (13.15)$$

$$\mathbf{B}_{\mathrm{m}}(\tau) = \mathrm{m}\;\mathbf{D}(\tau)\boldsymbol{\delta}. \qquad (13.16)$$

cMGF of the logreturns

Consider next the logreturn $R_{t+\delta t} = \ln(S_{t+\delta t} / S_t)$ over the interval $[t, t+\delta t]$ in the SV stock price model stated in section 11. For the n-factor OU SV process, the cMGF of the logreturn exists in closed form. From (11.7),

$$\phi_{R_{t+\delta t}|\mathcal{F}_t}(\mathrm{m}) = \exp\left[\mathrm{m}\mu\delta t + A_{\frac{1}{2}\mathrm{m}(\mathrm{m}-1)}(\delta t) + \mathbf{B}_{\frac{1}{2}\mathrm{m}(\mathrm{m}-1)}(\delta t)'\mathbf{x}_t\right], \qquad (13.17)$$

in which the functions $A_{\mathrm{m}(\mathrm{m}-1)/2}(.)$ and $(nx1)\,\mathbf{B}_{\mathrm{m}(\mathrm{m}-1)/2}(.)$ are given by

$$\begin{aligned}A_{\frac{1}{2}\mathrm{m}(\mathrm{m}-1)}(\delta t) &= \tfrac{1}{2}\mathrm{m}(\mathrm{m}-1)\left[(\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\delta t - \boldsymbol{\delta}'\mathbf{D}(\delta t)\boldsymbol{\theta}\right] + \tfrac{1}{8}\mathrm{m}^2(\mathrm{m}-1)^2\boldsymbol{\delta}'\left[\mathbf{N}(\delta t) \odot \boldsymbol{\Sigma\Sigma}'\right]\boldsymbol{\delta}\\ \mathbf{B}_{\frac{1}{2}\mathrm{m}(\mathrm{m}-1)}(\delta t) &= \tfrac{1}{2}\mathrm{m}(\mathrm{m}-1)\mathbf{D}(\delta t)\boldsymbol{\delta}. \qquad (13.18)\end{aligned}$$

Using the same abbreviations as in (11.10), straightforward algebra shows

$$\left.\frac{dA}{d\mathrm{m}}\right|_{\mathrm{m}=0} = -\tfrac{1}{2}(\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\delta t + \tfrac{1}{2}\boldsymbol{\delta}'\mathbf{D}(\delta t)\boldsymbol{\theta}, \qquad (13.19)$$

$$\left.\frac{d^2A}{d\mathrm{m}^2}\right|_{\mathrm{m}=0} = (\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\delta t - \boldsymbol{\delta}'\mathbf{D}(\delta t)\boldsymbol{\theta} + \tfrac{1}{4}\boldsymbol{\delta}'\left[\mathbf{N}(\delta t) \odot \boldsymbol{\Sigma\Sigma}'\right]\boldsymbol{\delta},$$

and

$$\left.\frac{d\mathbf{B}}{d\mathrm{m}}\right|_{\mathrm{m}=0} = -\tfrac{1}{2}\mathbf{D}(\delta t)\boldsymbol{\delta}\,; \quad \left.\frac{d^2\mathbf{B}}{d\mathrm{m}^2}\right|_{\mathrm{m}=0} = \mathbf{D}(\delta t)\boldsymbol{\delta}. \qquad (13.20)$$

(Un)conditional mean and variance of the logreturns

Notice that $\phi(0) = 1$ in the OU case. The conditional mean logreturn is given by $\phi'(0)$, and equals

$$\mathbb{E}_{\mathbb{M}}\left[R_{t+\delta t} \mid \mathcal{F}_t\right] = \left[\mu - \tfrac{1}{2}(\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\right]\delta t - \tfrac{1}{2}\boldsymbol{\delta}'\mathbf{D}(\delta t)\,\mathbf{x}_t^*. \qquad (13.21)$$

The unconditional mean logreturn is next determined from $\mathbb{E}_{\mathbb{M}}[\phi'(0)]$. It equals

$$\mathbb{E}_{\mathbb{M}}\left[R_{t+\delta t}\right] = \left[\mu - \tfrac{1}{2}(\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\right]\delta t \,. \tag{13.22}$$

Some algebra shows that the conditional second moment of the logreturn $R_{t+\delta t}$ (which equals $\phi''(0)$) is given by

$$\begin{aligned}\mathbb{E}_{\mathbb{M}}\left[R_{t+\delta t}^2 \mid \mathcal{F}_t\right] &= (\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\delta t + \left[\mu - \tfrac{1}{2}(\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\right]^2 \delta t^2 + \tfrac{1}{4}\boldsymbol{\delta}'\mathbf{D}(\delta t)\,\mathbf{x}_t^*\mathbf{x}_t^{*}{}'\mathbf{D}(\delta t)\boldsymbol{\delta} \\ &+ \left(1 - \left[\mu - \tfrac{1}{2}(\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\right]\delta t\right)\boldsymbol{\delta}'\,\mathbf{D}(\delta t)\,\mathbf{x}_t^* + \tfrac{1}{4}\boldsymbol{\delta}'\left[\mathbf{N}(\delta t)\odot\boldsymbol{\Sigma\Sigma}'\right]\boldsymbol{\delta}\,.\end{aligned} \tag{13.23}$$

As $\text{var}_{\mathbb{M}}[\mathbf{x}_t] = \mathbb{E}_{\mathbb{M}}[\mathbf{x}_t^*\mathbf{x}_t^{*}{}'] = \mathbf{J}\odot\boldsymbol{\Sigma\Sigma}'$, the unconditional second moment equals

$$\begin{aligned}\mathbb{E}_{\mathbb{M}}\left[R_{t+\delta t}^2\right] &= (\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\delta t + \left[\mu - \tfrac{1}{2}(\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\right]^2 \delta t^2 + \tfrac{1}{4}\boldsymbol{\delta}'\mathbf{D}(\delta t)\left[\mathbf{J}\odot\boldsymbol{\Sigma\Sigma}'\right]\mathbf{D}(\delta t)\boldsymbol{\delta} \\ &+ \tfrac{1}{4}\boldsymbol{\delta}'\left[\mathbf{N}(\delta t)\odot\boldsymbol{\Sigma\Sigma}'\right]\boldsymbol{\delta}\,.\end{aligned} \tag{13.24}$$

It then follows that, in the OU SV case,

$$\text{var}_{\mathbb{M}}\left[R_{t+\delta t} \mid \mathcal{F}_t\right] = (\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\delta t + \boldsymbol{\delta}'\,\mathbf{D}(\delta t)\,\mathbf{x}_t^* + \tfrac{1}{4}\boldsymbol{\delta}'\left[\mathbf{N}(\delta t)\odot\boldsymbol{\Sigma\Sigma}'\right]\boldsymbol{\delta} \tag{13.25}$$

$$\text{var}_{\mathbb{M}}\left[R_{t+\delta t}\right] = (\delta_0 + \boldsymbol{\delta}'\boldsymbol{\theta})\delta t + \tfrac{1}{4}\boldsymbol{\delta}'\left[\mathbf{D}(\delta t)\left(\mathbf{J}\odot\boldsymbol{\Sigma\Sigma}'\right)\mathbf{D}(\delta t) + \mathbf{N}(\delta t)\odot\boldsymbol{\Sigma\Sigma}'\right]\boldsymbol{\delta} \tag{13.26}$$

13.1 1-factor OU SV

Let us further specialize to 1-factor OU SV. In that case there is only one factor driving the volatility which evolves according to an Ornstein-Uhlenbeck process, such that

$$dx_t = k(\theta - x_t)dt + \sigma dW_{x,t}\,. \qquad (\mathbb{M}) \tag{13.27}$$

(1-factor OU SV is obtained by imposing the restrictions $n = 1$, $\alpha = 1$, $\beta = 0$ in the general multifactor affine SV model.) This section further restricts $\delta_0 = 0$, $\delta_1 = 1$ such that $\sigma_t^2 = x_t$, and thus

$$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \sigma dW_{x,t}\,. \qquad (\mathbb{M}) \tag{13.28}$$

Specializing the results for n-factor OU SV to 1-factor OU SV, yields (recall that $s > t$):

$$u_{t,s} \sim u_{t,s} \mid \mathcal{F}_t \sim \mathcal{N}\left[0;\ \sigma^2\left(\frac{1-\exp[-2k(s-t)]}{2k}\right)\right] \qquad (\mathbb{M}) \tag{13.29}$$

$$x_s \mid \mathcal{F}_t \sim \mathcal{N}\left[\theta + \exp[-k(s-t)]\left(x_t - \theta\right);\ \sigma^2\left(\frac{1-\exp[-2k(s-t)]}{2k}\right)\right] \qquad (\mathbb{M}) \tag{13.30}$$

$$x_s \sim \mathcal{N}\left[\theta;\ \frac{\sigma^2}{2k}\right] \qquad (\mathbb{M}) \tag{13.31}$$

$$\sigma_s^2 \mid \mathcal{F}_t \sim \mathcal{N}\left[\theta + \exp[-k(s-t)]\left(\sigma_t^2 - \theta\right);\ \sigma^2\left(\frac{1-\exp[-2k(s-t)]}{2k}\right)\right] \qquad (\mathbb{M}) \tag{13.32}$$

$$\sigma_s^2 \sim \mathcal{N}\left[\theta;\, \frac{\sigma^2}{2k}\right]. \qquad (\mathbb{M}) \qquad (13.33)$$

The conditional distribution of the integrated stock variance $\int_t^T \sigma_s^2 ds$ is Gaussian, with

$$\mathbb{E}_{\mathbb{M}}\left[\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t\right] = \theta\tau + \left(\frac{1-\exp[-k\tau]}{k}\right)(\sigma_t^2 - \theta), \qquad (13.34)$$

$$\mathrm{var}_{\mathbb{M}}\left[\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t\right] = \frac{\sigma^2}{k^2}\left[\tau - 2\left(\frac{1-\exp[-k\tau]}{k}\right) + \left(\frac{1-\exp[-2k\tau]}{2k}\right)\right]. \qquad (13.35)$$

The unconditional distribution of the integrated stock variance is also Gaussian, with

$$\mathbb{E}_{\mathbb{M}}\left[\int_t^T \sigma_s^2 ds\right] = \theta\tau\,, \qquad (13.36)$$

$$\begin{aligned}\mathrm{var}_{\mathbb{M}}\left[\int_t^T \sigma_s^2 ds\right] &= \mathbb{E}_{\mathbb{M}}\left[\mathrm{var}_{\mathbb{M}}\left(\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t\right)\right] + \mathrm{var}_{\mathbb{M}}\left[\mathbb{E}_{\mathbb{M}}\left(\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t\right)\right] \qquad (13.37)\\ &= \frac{\sigma^2}{k^2}\left[\tau - 2\left(\frac{1-\exp[-k\tau]}{k}\right) + \frac{1-\exp[-2k\tau]}{2k}\right] + \frac{\sigma^2}{2k}\left(\frac{1-\exp[-k\tau]}{k}\right)^2.\end{aligned}$$

The functions $A_{\mathrm{m}}(.)$ and $B_{\mathrm{m}}(.)$ in (13.15)-(13.16) become in the 1-factor OU SV case

$$\begin{aligned}A_{\mathrm{m}}(\tau) &= \mathrm{m}\theta\left[\tau - \left(\frac{1-\exp[-k\tau]}{k}\right)\right] + \tfrac{1}{2}\left(\frac{\mathrm{m}\sigma}{k}\right)^2\left[\tau - 2\left(\frac{1-\exp[-k\tau]}{k}\right) + \left(\frac{1-\exp[-2k\tau]}{2k}\right)\right]\\ B_{\mathrm{m}}(\tau) &= \mathrm{m}\left(\frac{1-\exp[-k\tau]}{k}\right). \qquad (13.38)\end{aligned}$$

14. Special case: 1-factor CIR SV (Heston (1993) volatility process)

This section states the main results for the 1-factor Cox-Ingersoll-Ross (CIR) SV specification, also known as Heston (1993) volatility. In this case there is one volatility-driving factor that follows a CIR process: [4]

$$dx_t = k(\theta - x_t)dt + \sigma\sqrt{x_t}\,dW_{x,t}\,. \qquad (\mathbb{M}) \qquad (14.1)$$

The CIR SV case is obtained by restricting $n = 1$, $\alpha = 0$, $\beta = 1$ in the general multifactor affine SV model. In this section we furthermore restrict $\delta_0 = 0$, $\delta_1 = 1$ such that $\sigma_t^2 = x_t$. The instantaneous stock variance thus follows

$$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \sigma\sqrt{\sigma_t^2}\,dW_{x,t}\,. \qquad (\mathbb{M}) \qquad (14.2)$$

[4] This process is also known as the *square root* or *Feller* process, due to W. Feller. In the SV literature it is also referred to as the Heston (1993) stock variance process, or in short, Heston volatility.

This specification precludes negative σ_t^2's if k, θ, $\sigma_0^2 > 0$. As Brownian sample paths are continuous, so are the paths $\{\sigma_t^2\}$ follows. σ_t^2 therefore first has to become zero for a path to possibly become negative. But if σ_t^2 becomes zero, then the diffusion component of the SDE drops out, and hence there is temporarily no stochastic variation in σ_t^2. As $\theta, k > 0$ are positive, the process will then naturally be pulled away from zero towards θ (due to mean reversion). As such, σ_t^2 will never become negative. If the *Feller condition* $2k\theta \geq \sigma^2$ holds (Feller (1951)), the upward drift in the CIR stock variance process is sufficiently large for σ_t^2 to never exactly reach zero.

From (3.2) and (3.7) it follows that (recall that $s > t$):

$$\mathbb{E}_{\mathbb{M}}\left[x_s \mid \mathcal{F}_t\right] = \theta + \exp[-k(s-t)]\, x_t^*, \qquad \mathbb{E}_{\mathbb{M}}\left[x_s\right] = \theta. \tag{14.3}$$

From (4.10) and (5.7),

$$\begin{aligned} \text{var}_{\mathbb{M}}\left[x_s \mid \mathcal{F}_t\right] &= \sigma^2\theta\left(\frac{1-\exp[-2k(s-t)}{2k}\right) + \sigma^2\, x_t^*\left(\frac{\exp[-k(s-t)] - \exp[-2k(s-t)]}{k}\right) \\ &= \frac{\sigma^2\theta}{2k}\left(1-\exp[-k(s-t)]\right)^2 + x_t\frac{\sigma^2}{k}\left(\exp[-k(s-t)] - \exp[-2k(s-t)]\right), \end{aligned} \tag{14.4}$$

$$\text{var}_{\mathbb{M}}\left[x_s\right] = \frac{\sigma^2\theta}{2k}. \tag{14.5}$$

As $\sigma_t^2 = x_t$, the unconditional and conditional moments of the stock variance follow directly from the expressions above. In contrast to the 1-factor OU SV case (see (13.8)), the conditional variance of the CIR stock variance depends on the current level of the volatility. In particular,

$$\frac{\partial\, \text{var}_{\mathbb{M}}[\sigma_s^2 \mid \mathcal{F}_t]}{\partial \sigma_t^2} = \frac{\sigma^2}{k}\left(\exp[-k(s-t)] - \exp[-2k(s-t)]\right) > 0, \tag{14.6}$$

as $s - t > 0$ and $k > 0$. Therefore, if the currently stock volatility increases, the conditional variance of the stock variance increases. The CIR process is thus able to describe level-dependent volatility-of-volatility.

Regarding the properties of the disturbance term $u_{t,s}$, from (6.4), (6.5) and (6.6):

$$\mathbb{E}_{\mathbb{M}}\left[u_{t,s} \mid \mathcal{F}_t\right] = 0, \qquad \mathbb{E}_{\mathbb{M}}\left[u_{t,s}\right] = 0, \tag{14.7}$$

$$\text{var}_{\mathbb{M}}\left[u_{t,s} \mid \mathcal{F}_t\right] = \sigma^2\theta\left(\frac{1-\exp[-2k(s-t)]}{2k}\right) + \sigma^2\, x_t^*\left(\frac{\exp[-k(s-t)] - \exp[-2k(s-t)]}{k}\right) \tag{14.8}$$

$$\text{var}_{\mathbb{M}}\left[u_{t,s}\right] = \sigma^2\theta\left(\frac{1-\exp[-2k(s-t)}{2k}\right). \tag{14.9}$$

From (9.17) and (9.18),

$$\mathbb{E}_{\mathbb{M}}\left[\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t\right] = \theta\tau + \left(\frac{1-\exp[-k\tau]}{k}\right)(\sigma_t^2 - \theta) \tag{14.10}$$

$$\mathbb{E}_{\mathbb{M}}\left[\int_t^T \sigma_s^2 ds\right] = \theta\tau . \qquad (14.11)$$

From (9.19) and (9.20),

$$\begin{aligned}\text{var}_{\mathbb{M}}\left[\int_t^T \sigma_s^2 ds \mid \mathcal{F}_t\right] &= \frac{\sigma^2\theta}{k^2}\left[\tau - 2\left(\frac{1-\exp[-k\tau]}{k}\right) + \frac{1-\exp[-2k\tau]}{2k}\right] \qquad (14.12)\\ &+ \frac{\sigma^2}{k^2}(\sigma_t^2 - \theta)\left[\frac{1-\exp[-k\tau]}{k} - 2\tau\exp[-k\tau] + \frac{\exp[-k\tau]-\exp[-2k\tau]}{k}\right]\end{aligned}$$

$$\text{var}_{\mathbb{M}}\left[\int_t^T \sigma_s^2 ds\right] = \sigma^2\theta\left\{\frac{1}{k^2}\left[\tau - 2\left(\frac{1-\exp[-k\tau]}{k}\right) + \frac{1-\exp[-2k\tau]}{2k}\right] + \frac{1}{2k}\left(\frac{1-\exp[-k\tau]}{k}\right)^2\right\}$$

(14.13)

In the CIR SV case, the functions $A_1(.)$ and $B_1(.)$ can numerically be solved for, after first having tailored the ODE system (10.3) to this particular case.

The analytical results in Hull (2003)

The observant reader may perhaps think (due to the similarities) that the analytical results in Hull (2003) p.542 can be transformed to our specific SV interests. Hull discusses the CIR *interest rate* model, and gives an analytical expression for the implied zero-coupon bond prices resulting from that model. Translating his results to our SV setting in which $x_t = \sigma_t^2$ follows a CIR process, yields

$$\mathbb{E}_{\mathbb{M}}\left[\exp\left(-\int_t^T \sigma_s^2 ds\right) \mid \mathcal{F}_t\right] = \exp\left[\mathfrak{F}_1(\tau,k,\theta,\sigma) + \mathfrak{F}_2(\tau,k,\sigma)\, x_t\right], \qquad (14.14)$$

in which $\mathfrak{F}_1(\tau,k,\theta,\sigma)$ and $\mathfrak{F}_2(\tau,k,\sigma)$ are some closed-form expressions of their respective arguments. However, our interest is *not* in the conditional expectation of the exponent of *minus* the integrated stock variance, but instead is in the conditional expectation of the exponent of *plus* the integrated variance. This latter quantity *cannot* be obtained by a simple transformation or reparametrization however. The argument is as follows. Notice that

$$\mathbb{E}_{\mathbb{M}}\left[\exp\left(\int_t^T \sigma_s^2 ds\right) \mid \mathcal{F}_t\right] = \mathbb{E}_{\mathbb{M}}\left[\exp\left(-\int_t^T \tilde{\sigma}_s^2 ds\right) \mid \mathcal{F}_t\right], \qquad (14.15)$$

in which $\tilde{\sigma}_s^2 \equiv -\sigma_s^2 \ \forall s$. Now, if it can be shown that $\{\tilde{\sigma}_t^2; t \geq 0\}$ is a CIR process (with possibly different parameters), then we can resort to the closed-form result. However, as $\tilde{\sigma}_t^2 = -x_t < 0 \ \forall t$, as $\{x_t; t \geq 0\}$ is a CIR process itself, $\{\tilde{\sigma}_t^2; t \geq 0\}$ *cannot* be a CIR process, as it is always negative. As such, to obtain the functions $A_1(.)$ and $B_1(.)$ in the Heston SV case, requires numerically solving the system of Ricatti ODEs.

References

Aït-Sahalia, Y., P.A. Mykland and L. Zhang, 2005a, *How often to sample a continuous-time process in the presence of market microstructure noise*, The Review of Financial Studies 18, 351-416.

Aït-Sahalia, Y., P.A. Mykland and L. Zhang, 2005b, *Ultra high frequency volatility estimation with dependent microstructure noise*, working paper, Princeton University.

Alizadeh, S., M.W. Brandt, and F.X. Diebold, 2002, *Range-based estimation of stochastic volatility models*, Journal of Finance, vol. LVII, no.3, 1047-1091.

Andersen, T.G., L. Benzoni and J. Lund, 2002, *An empirical investigation of continuous-time equity return models*, Journal of Finance, vol. LVII, no. 3, 1239-1284.

Andersen, T.G. and T. Bollerslev, 1998, *Answering the skeptics: Yes, standard volatility models do provide accurate forecasts*, International Economic Review, vol. 39, no. 4, 885-905.

Andersen, T.G., T. Bollerslev and F.X Diebold, 2003, *Some like it smooth, and some like it rough: Untangling continuous and jump components in measuring, modeling and forecasting asset return volatility*, unpublished paper, Duke University.

Andersen, T.G., T. Bollerslev, F.X. Diebold and H. Ebens, 2001a, *The distribution of realized stock return volatility*, Journal of Financial Economics 61, 43-76.

Andersen, T.G., T. Bollerslev, F.X. Diebold and H. Ebens, 2001b, *The distribution of realized stock return volatility: Appendix*, Working paper, North Western University, USA.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys, 2001a, *The distribution of realized exchange rate volatility*, Journal of the American Statistical Association 96, 42-55.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys, 2001b, *Exchange rate returns standardized by realized volatility are (nearly) Gaussian*, Multinational Finance Journal, forthcoming.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys, 2003, *Modelling and forecasting realized volatility*, Econometrica 71, 579, 626.

Andersen, T.G. and J. Lund, 1997, *Estimating continuous-time stochastic volatility models of the short-term interest rate*, Journal of Econometrics 77, 343-377.

Anderson, B.D.O and J.B. Moore, 1979, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, New Jersey.

Atkinson, K., 1993, *Elementary numerical analysis*, second edition, Wiley.

Ball, C.A. and A. Roma, 1994, *Stochastic volatility option pricing*, Journal of Financial and Quantitative Analysis 29, 589-607.

Bakshi, G., C. Cao and Z. Chen, 1997, *Empirical performance of alternative option pricing models*, Journal of Finance, vol. 52, issue 5, 2003 – 2049.

Bakshi, G. and N. Kapadia, 2003, *Delta-hedged gains and the negative market volatility risk premium*, The Review of Financial Studies, vol.16, no.2, 527-566.

Bandi, F.M. and J.R. Russel, 2004, *Microstructure noise, realized volatility, and optimal sampling*, working paper, Graduate School of Business, University of Chicago.

Barndorff-Nielsen, O., P.R. Hansen, A. Lunde and N. Shephard, 2004, *Regular and modified kernel-based estimators of integrated variance: the case with independent noise*, working paper, University of Oxford.

Barndorff, O.E., E. Nicolato and N. Shephard, 2002, *Some recent developments in stochastic volatility modelling*, Quantitative Finance, 2, 11-23.

Barndorff, O.E. and N. Shephard, 2001a, *Realised power variation and stochastic volatility models*, working paper series no. 91, Centre for Analytical Finance, University of Aarhus, Aarhus, Denmark.

Barndorff-Nielsen, O.E. and N. Shephard, 2001b, *Non-Gaussian Ornstein-Uhlenbeck–based models and some of their uses in financial economics*, Journal of the Royal Statistical Society, Series B, 63, 167-241.

Barndorff-Nielsen, O.E. and N. Shephard, 2002, *Econometric analysis of realized volatility and its use in estimating stochastic volatility models*, Journal of the Royal Statistical Society, Series B, 64, 253-280.

Barndorff, O.E. and N. Shephard, 2003, *Realised power variation and stochastic volatility*, Bernoulli 9, 243-265.

Barndorff, O.E. and N. Shephard, 2004a, *Power and bipower variation with stochastic volatility and jumps*, Journal of Financial Econometrics 2, 1-48.

Barndorff, O.E. and N. Shephard, 2004b, *Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics*, Econometrica 72, 885-925.

Barndorff, O.E. and N. Shephard, 2005, *Econometrics of testing for jumps in financial economics using bipower variation*, Journal of Financial Econometrics, forthcoming.

Bates, D. S., 1996, *Testing option pricing models*, in Maddala, G.S., C.R. Rao, eds., Handbook of Statistics, Vol. 14, Elsevier, Amsterdam.

Bates, D.S., 1996b, *Jumps and stochastic volatility: exchange rate processes implicit in Deutsche Mark options*, Review of Financial Studies, vol.9, no.1.

Bates, D.S., 2000, *Post-'87 crash fears in S&P500 futures options*, Journal of Econometrics 94 (1-2), 181 – 238.

Bates, D.S., 2003, *Empirical option pricing: a retrospection*, Journal of Econometrics 116, 387 – 404.

Bauwens, L., S. Laurent and J.V.K Rombouts, 2003, *Multivariate GARCH models: A survey*, CORE discussion paper 2003/31, Katholieke Universiteit Leuven.

Baxter, M. and A. Rennie, 1996, *Financial calculus: an introduction to derivative pricing*, Cambridge University Press, Cambridge UK.

Bera, A.K., and M.L. Higgins, 1993, *ARCH models: properties, estimation and testing*, Journal of Economic Surveys, Vol.7, No.4, 215-269.

Bertsimas, D., L. Kogan and A.W. Lo, 2000, *When is time continuous?*, Journal of Financial Economics 55, 173-204.

Black, F., 1976, *Studies of stock price volatility changes*, Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association, 177-181.

Black, F. and M. Scholes, 1973, *The pricing of options and corporate liabilities*, Journal of Political Economy 81, 637-54.

Bodie, Z., A. Kane and A.J. Marcus, 1996, *Investments*, Richard D. Irwin, A Times Mirror Higher Education Group, Inc. company, USA.

Bollerslev, T., 1986, *Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics 31, 307-327.

Bollerslev, T., 1990, *Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model*, The Review of Economics and Statistics 72, 498 – 505.

Bollerslev, T., R.Y. Chou and K.F. Kroner, 1992, *ARCH modeling in finance: a review of the theory and empirical evidence*, Journal of Econometrics 52, 5-59.
Bollerslev, T., R.F. Engle and D. Nelson, 1994, *ARCH models*, Handbook of Econometrics, Volume IV, Elsevier Science.
Bollerslev, T, and H. Zhou, 2002, *Estimating stochastic volatility diffusion using conditional moments of integrated volatility*, Journal of Econometrics 109, 33 - 65.
Breeden, D., 1979, *An intertemporal asset pricing model with stochastic consumption and investment opportunities*, Journal of Financial Economics 7, 265-296.
Brockwell, P., 2001, *Continuous-time ARMA processes*, in Shanbhag, D. and C. Rao (eds.), Handbook of Statistics, vol.19, North-Holland.
Brockwell, P. and R. Davis, 2001, *Comment on Barndorff-Nielsen, O.E. and N. Shephard*, Journal of the Royal Statistical Society, Series B, Methodological 64.
Buraschi, A. and J. Jackwerth, 2001, *The price of a smile: hedging and spanning in option markets*, The Review of Financial Studies, vol.14, no. 2, 495-527.
Campbell, J.Y., A.W. Lo and A.C. MacKinlay, 1997, *The econometrics of financial markets*, Princeton University Press, Princeton NJ.
Carnero, M. A., D. Pena and E. Ruiz, 2001, *Is stochastic volatility more flexible than GARCH?*, Proceedings of the International Conference on Modelling and Forecasting Financial Volatility, September 2001, Perth, Western Australia.
Carr, P. and L. Wu, 2004, *Variance risk premia*, Working paper, Bloomberg L.P. and Courant Institute, Zicklin School of Business, New York.
Carr, P. and L. Wu, 2004b, *Stochastic Skew in Currency Options*, Working paper, Bloomberg L.P. and Courant Institute, Zicklin School of Business, New York.
Chernov, M., A.R. Gallant, E. Ghysels and G. Tauchen, 2003, *Alternative models for stock price dynamics*, Journal of Econometrics, vol. 116, issues 1-2, 225-257.
Chernov, M. and E. Ghysels, 2000, *A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation*, Journal of Financial Economics 56, 407 – 458.
Chesney, M. and L. Scott, 1989, *Pricing European currency options: A comparison of the modified Black-Scholes model and a random variance model*, Journal of Financial and Quantitative Analysis 24, 267-84.
Christensen, B. and N. Prabhala, 1998, *The relation between implied and realized volatility*, Journal of Financial Economics 50, 125-150.
Cochrane, J.H., 2001, *Asset pricing*, Princeton University Press, Princeton, New Yersey, USA.
Cont, R. and J. Fonseca (2002), *Dynamics of implied volatility surfaces*, Quantitative Finance, vol. 2, 45-60.
Coval, J.D. and T. Shumway, 2001, *Expected option returns*, Journal of Finance, vol. LVI, no.3, 983-1009.
Cox, J.C., J.E. Ingersoll and S.A. Ross, 1985, *An intertemporal general equilibrium model of asset prices*, Econometrica 53, 363-384.
Dai, Q. and K.J. Singleton, 2000, *Specification analysis of affine term structure models*, The Journal of Finance, Vol. LV, no. 5, 1943 – 1978.
Danielsson, J., 1994, *Stochastic volatility in asset prices, estimation with simulated maximum likelihood*, Journal of Econometrics 64, 375-400.
Danielsson, J., 1998, *Multivariate stochastic volatility models: estimation and a comparison with VGARCH models*, Journal of Empirical Finance 5, 155-173.
Delbaen, F. and W. Schachermayer, 1994, *The fundamental theorem of asset pricing*, Mathematische Annalen 300, 463-520.
Doornik, J.A., G. Draisma and M. Ooms, 1998, *Introduction to Ox, an object-orientated matrix programming language*, Timberlake Consultants Ltd, UK.

Driessen, J. and P. Maenhout, 2003, *A portfolio perspective of option pricing anomalies*, Working paper, University of Amsterdam and INSEAD (France).
Duffee, G. R., 2002, *Term premia and interest rate forecast in affine models*, Journal of Finance, vol. LVII, no. 1, p. 405 – 443.
Duffee, G. R. and R. H. Stanton, 2001, *Estimation of dynamic term structure models*, Unpublished paper, Haas School of Business, U.C. Berkeley, California, USA.
Duffie, D., 2001, *Dynamic asset pricing theory*, Princeton University Press, Princeton, New Jersey, USA.
Duffie, D. and R. Kan, 1996, *A yield-factor model of interest rates*, Mathematical Finance 6, 379 – 406.
Duffie, D., J. Pan and K. Singleton, 2000, *Transform analysis and asset pricing for affine jump diffusions*, Econometrica 68, 1343-1376.
Duffie, D. and K.J. Singleton, 1993, *Simulated moments estimation of Markov models of asset prices*, Econometrica 61, 929-52.
Durbin, J. and S.J. Koopman, 1997, *Monte Carlo maximum likelihood estimation of non-Gaussian state space models,* Biometrika 84, 669-84.
Durbin, J. and S.J. Koopman, 2000, *Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion),* Journal of The Royal Statistical Society B 62, 3-56.
Durbin, J. and S.J. Koopman, 2001, *Time series analysis by state space methods*, Oxford University Press, Oxford, UK.
Durham, G.B., and R. Gallant, 2002, *Numerical techniques for simulated maximum likelihood estimation of stochastic differential equations*, Journal of Business and Economic Statistics 20, 297-316.
Ebens, H., 1999, *Realized stock index volatility*, working paper no. 420, Department of Economics, Johns Hopkins University, USA.
Engle, R.F., 1982, *Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation*, Econometrica 50, 987-1008.
Engle, R. F., 1995, *ARCH: selected readings*, Oxford University Press, Oxford, United Kingdom.
Engle, R.F. and T. Bollerslev, 1986, *Modelling the persistence of conditional variances*, Econometric Reviews 5, 1-50, 81-87.
Engle, R.F. and G.G.J. Lee, 1999, *A permanent and transitory component model of stock return volatility*, in R.F. Engle and H. White (eds.), Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger, 475-497, Oxford University Press, Oxford UK.
Eraker, B., M. Johannes and N. Polson, 2003, *The impact of jumps in volatility and returns*, Journal of Finance, vol. LVIII, no.3, 1269-1299.
Etheridge, A.E., 2002, *A course in financial calculus*, Cambridge university press, Cambridge, UK.
Figlewski, S., 1989, *Options arbitrage in imperfect markets*, Journal of Finance vol. XLIV, no.5, 1289-1311.
Fornari, F. and A. Mele, 1997, *Sign- and volatility switching ARCH models: Theory and applications to international stock markets*, Journal of Applied Econometrics 12, 49-65.
Franses, P.H., and D. van Dijk, 2000, *Non-linear time series models in empirical finance*, Cambridge University Press, Cambridge.
French, K., G. Schwert and R. Stambaugh, 1987, *Expected stock returns and volatility*, Journal of Financial Economics, 19, 3 – 30.
Friedman, M., 1957, *A theory of the consumption function*, Princeton, NJ: Princeton University Press.

Gallant, A.R., D. Hsieh and G. Tauchen, 1997, *Estimation of stochastic volatility models with diagnostics*, Journal of Econometrics 81, 159-92.

Gallant, A.R., C. Hsu and G. Tauchen, 1999, *Using daily range data to calibrate volatility diffusions and extract the forward integrated variance*, The Review of Economics and Statistics 81, vol.4, 617-631.

Gallant, A.R. and G. Tauchen, 1996, *Which moments to match?*, Econometric Theory 12, 657 – 681.

Garcia, R., E. Ghysels and E. Renault, 2003, *The econometrics of option pricing*, working paper, University of Montreal, University of North Carolina, CIRANO and CIREQ.

Garman, M.B. and S.W. Kohlhagen, 1983, *Foreign currency option values*, Journal of International Money and Finance 2, 231 – 237.

Ghysels, E., Harvey, A.C. and E. Renault, 1996, *Stochastic volatility*, in: Statistical Methods in Finance (North-Holland, Amsterdam), 119-191.

Glosten, L.R., Jagannathan, R. and D.E. Runkle, 1993, *On the relation between the expected value and the volatility of the nominal excess returns on stocks*, Journal of Finance 48, 1779-1801.

Gonzalez-Rivera, G., 1998, *Smooth transition GARCH models,* Studies in Nonlinear Dynamics and Econometrics 3, 61-78.

Gouriéroux, C., A. Montfort and E. Renault, 1993, *Indirect inference*, Journal of Applied Econometrics 8, 85–118.

Greene, W.H., 1997, *Econometric Analysis*, third edition, Prentice-Hall International Editions.

Hamilton, J. D., 1994, *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Hansen, P.R, and A. Lunde, 2004a, *Realized variance and market microstructure noise*, working paper, Stanford University.

Hansen, P.R, and A. Lunde, 2004b, *An unbiased measure of realized variance*, working paper, Stanford University.

Harrison, M.J. and D. Kreps, 1979, *Martingales and arbitrage in multiperiod securities markets*, Journal of Economic Theory 20, 381-408.

Harrison, M.J. and S.R. Pliska, 1981, *Martingales and stochastic integrals in the theory of continuous trading*, Stochastic Processes and Their Applications 11, 215-260.

Harvey, A.C., 1989, *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press.

Harvey, A.C., E. Ruiz and N. Shephard, 1994, *Multivariate stochastic variance models*, The Review of Economic Studies 61, 247 – 264.

Harvey, A.C. and N. Shephard, 1996, *Estimation of an asymmetric stochastic volatility model for asset returns*, Journal of Economics and Business Statistics 14, 429-34.

Heston, S.L., 1993, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, The Review of Financial Studies 6, 327-43.

Ho, T. and S. Lee, 1986, *Term structure movements and pricing interest rate contingent claims*, Journal of Finance 41, 1011-1029.

Huang, X. and G. Tauchen, 2004, *The relative contribution of jumps to total price variance*, working paper, Duke University.

Hull, J.C., 2003, *Options, futures and other derivatives*, Prentice Hall International.

Hull, J.C. and A. White, 1987, *The pricing of options on assets with stochastic volatilities*, Journal of Finance 42, 2, 281-300.

Jackwerth, J. and M. Rubinstein, 1996, *Recovering probability distributions from option prices*, Journal of Finance 51, 1611-1631.

Jacquier, E., N.G. Polson and P.E. Rossi, 1994, *Bayesian analysis of stochastic volatility models*, Journal of Business and Economic Statistics 12, 371-389.

Johnson, H. and D. Shanno, 1987, *Option pricing when the variance is changing*, Journal of Financial and Quantitative Analysis 22, 143-51.

Jones, C.S., 2003, *The dynamics of stochastic volatility: evidence from underlying and options markets*, Journal of Econometrics 116, 181-224.

de Jong, F., 2000, *Time series and cross-section information in affine term-structure models*, Journal of Business and Economic Statistics, Vol. 18, No. 3, 300 – 314.

Kalman, R.E., 1960, *A new approach to linear filtering and prediction problems*, J. Basic Engineering, Transactions ASMA, Series D, 82, 35-45.

Kim, S., Shephard, N. and S. Chib, 1998, *Stochastic volatility: likelihood inference and comparison with ARCH models*, The Review of Economic Studies 65, 361-93.

Kloeden, P.E., 2002, *The systematic derivation of higher-order numerical schemes for stochastic differential equations*, Milan Journal of Mathematics.

Kloeden, P.E., and E. Platen, 1999, *Numerical solution of stochastic differential equations*, Springer.

Koopman, S.J., 1993, *Disturbance smoother for state space models*, Biometrika 80, p. 117 – 126.

Koopman, S.J., B. Jungbacker and E. Hol, 2005, *Forecasting daily variability of the S&P100 stock index using historical, realised and implied volatility measurements*, Journal of Empirical Finance, Volume 12, Issue 3, p. 445-475.

Koopman, S.J., N. Shephard and J.A. Doornik, 1999, *Statistical algorithms for models in state space using SsfPack 2.2*, Econometrics Journal, volume 2, p. 113 – 166.

Langetieg, T., 1980, *A multivariate model of the term structure of interest rates*, Journal of Finance, 35, 71-97.

Liesenfeld, R. and R.C. Jung, 2000, *Stochastic volatility models: conditional normality versus heavy-tailed distributions*, Journal of Applied Econometrics 15, 137-60.

Lintner, J., 1965, *The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets*, The Review of Economics and Statistics 47, 13-37.

Liu, J., Longstaff, F. and J. Pan, 2001, *Dynamic asset allocation with event risk*, Working paper, UCLA, Anderson School and MIT Sloan School.

Lund, J., 1997, *Econometric analysis of continuous-time arbitrage-free models of the term-structure of interest rates*, Working paper, Aarhus School of Business, Aahrus, Denmark.

Mahieu, R.J. and P.C. Schotman, 1998, *An empirical application of stochastic volatility models*, Journal of Applied Econometrics 13, 333-60.

Martens, M., D. van Dijk, and M. de Pooter, 2003, *Modeling and forecasting S&P500 volatility: long memory, structural breaks and nonlinearity*, working paper, Erasmus University, Rotterdam, The Netherlands.

Meddahi, N., 2001, *An eigenfunction approach for volatility modeling*, working paper, Université de Montréal, Cahiers de Recherche 2001-29.

Meddahi, N., 2002, *Moments of continuous time stochastic volatility models*, Working paper, CIRANO.

Melino, A. and M. Turnbull, 1990, *Pricing foreign currency options with stochastic volatility*, Journal of Econometrics 45, 239-65.

Merton, R., 1976, *Option pricing when underlying stock returns are discontinuous*, Journal of Financial Economics 3, 125 – 144.

Mikosch, T., 1998, *Elementary Stochastic Calculus with Finance in View*, World Scientific Publishing.

Modigliani, F. and R. Brumberg, 1954, *Utility analysis and the consumption function: An interpretation of cross-section data*, in Kenneth K. Kurihari, ed., Post-Keynesian Economics, 388-436, New Brunswick, NJ: Rutgers University Press.

Mossin, J., 1966, *Equilibrium in a capital asset market*, Econometrica 35, 768-793.

Nelson, D.B., 1990, *ARCH models as diffusion approximations*, Journal of Econometrics 45, 7-39.

Nelson, D.B., 1991, *Conditional heteroskedasticity in asset returns: a new approach*, Econometrica 59, p.347-370.

Nicolato, E. and E. Vernardos (2002), *Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type*, working paper, Centre for Analytical Finance, University of Aarhus, Denmark.

Pan, J., 2002, *The jump-risk premia implicit in options: evidence from an integrated time-series study*, Journal of Financial Economics 63, 3-50.

Poon, S.H., 2005, *Forecasting volatility in financial markets: a practical guide*, Wiley, to appear in 2005.

Poon, S.H. and C. W. J. Granger, 2003, *Forecasting volatility in financial markets: a review*, Journal of Economic Literature, American Economic Association, vol. 41 (2), 478-539.

Protter, P., 1990, *Stochastic integration and differential equations*, New York, Springer-Verlag.

Renault, E., 1997, *Econometric models of option pricing errors*, in Advances in Economics and Econometrics: Theory and Applications, vol.3, eds. Kreps, D. and Wallis, K. Cambridge University Press.

Renault, E. and N. Touzi, 1996, *Option hedging and implied volatilities in a stochastic volatility model*, Mathematical Finance 6, 279 - 302.

Romer, D., 2001, *Advanced macroeconomics*, 2nd edition, McGraw-Hill Higher Education, New York.

Ruiz, E., 1994, *Quasi maximum likelihood estimation of stochastic volatility models*, Journal of Econometrics 63, 289 - 306.

Rabemananjara, R. and J. M. Zakoian, 1993, *Threshold ARCH models and asymmetries in volatility*, Journal of Applied Econometrics, vol. 8, 31-49.

Ross, S., 1976, *The arbitrage theory of capital asset pricing*, Journal of Economic Theory 13, 341-360.

Sandmann, G. and S.J. Koopman, 1998, *Estimation of stochastic volatility models via Monte Carlo maximum likelihood*, Journal of Econometrics 87, 271 - 301.

Sentana, E., 1995, *Quadratic ARCH models*, The Review of Economic Studies 62, 639 - 61.

Schmitz, K., 2004, *Strong Taylor schemes for stochastic volatility*, http://www.maths.ox.ac.uk/~schmitz/project2.htm.

Scott, L.O., 1987, *Option pricing when the variance changes randomly: theory, estimation and an application*, Journal of Financial and Quantitative Analysis 22, 419 - 38.

Sharpe, W., 1964, *Capital asset prices: a theory of market equilibrium under conditions of risk*, Journal of Finance, 19, 425 - 442.

Shephard, N., 1996, *Statistical aspects of ARCH and stochastic volatility*, in Time series Models in Econometrics, Finance and Other Fields, London, Chapman & Hall, 1-67.

Shephard, N., 2005, *Stochastic Volatility: Selected Readings*, edited volume, Oxford University Press.

Shreve, S.E., 1997, *Stochastic calculus and finance*, draft (book in preparation), Carnegie Mellon University.

Stein, E.M. and J.C. Stein, 1991, *Stock price distributions with stochastic volatility: an analytic approach*, The Review of Financial Studies 4, 727-52.

Tauchen, G., 2004, *Remarks on recent developments in stochastic volatility: Statistical modelling and general equilibrium analysis*, Paper prepared for the joint 59th European Meeting of the Econometric Society and 19th annual congress of the European Economics Association, Universidad Carlos III de Madrid.

Tauchen, G., and V. Todorov, 2004, *Simulation methods for Lévy-driven CARMA stochastic volatility models*, Proceedings of The First Symposium on Econometric Theory and Applications (SETA), Academia Sinica, Taiwan.

Taylor, S.J., 1994, *Modeling stochastic volatility: a review and comparitive study*, Mathematical Finance 4, 183-204.

Taylor, S.J., 1986, *Modelling Financial Time Series*, Chichester: John Wiley publishing, John Wiley and sons.

Van der Ploeg, Antoine P.C., 1999, *New developments in Value-at-Risk techniques for stock portfolios* (in Dutch), M.Sc. thesis, University of Amsterdam, The Netherlands.

Van der Ploeg, Antoine P.C., 2002, *Volatility and derivative pricing in financial markets*, master thesis M.Phil. in Economics, Tinbergen Institute and University of Amsterdam, The Netherlands.

Van der Ploeg, Antoine P.C., 2004, *Multifactor volatility models: evidence from stock and option markets*, working paper, Tinbergen Institute and University of Amsterdam, The Netherlands.

Van der Ploeg, Antoine P.C., H.P. Boswijk and F.C.J.M. de Jong, 2003, *A state space approach to the estimation of multifactor affine stochastic volatility option pricing models*, working paper, Tinbergen Institute and University of Amsterdam, The Netherlands.

Van der Sluis, P.J., 1999, *Estimation and inference with the efficient method of moments with applications to stochastic volatility models and option pricing*, Ph.D. Thesis, Tinbergen Institute, Amsterdam, The Netherlands.

Watson, M.W., 1989, *Recursive solution methods for dynamic linear rational expectations models*, Journal of Econometrics 41, 65 – 89.

Vasicek, O.A., 1977, *An equilibrium characterization of the term structure*, Journal of Financial Economics 5, 177 - 88.

White, H., 1982, *Maximum likelihood estimation of misspecified models*, Econometrica 50, 1 - 25.

Wiggins, J.B., 1987, *Option values under stochastic volatility*, Journal of Financial Economics 19, 351 - 72.

Zhang, L., 2004, *Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach*, manuscript, Carnegie Mellon University.

Zhang, L., P.A. Mykland and Y. Aït-Sahalia, 2002, *A tale of two time-scales: determining integrated volatility with noisy high-frequency data*, working paper, Princeton University, forthcoming in the Journal of the American Statistical Association.

Stochastische Volatiliteit en de Waardering van Financiële Derivaten:

Samenvatting

(Summary in Dutch)

Dit proefschrift beschouwt het waarderen van derivaten onder stochastische volatiliteit (SV). Gedreven door recent empirisch bewijs, nemen we aan dat de volatiliteit van het onderliggende aandeel aangedreven wordt door een *multifactor* SV specificatie. Meer specifiek, we modelleren de volatiliteit als een affiene functie van een willekeurig aantal niet-waarneembare affiene factoren, welke *mean-reverting* Markov diffusies volgen. We noemen het model *het multifactor affiene stochastische volatiliteit derivatenwaarderingsmodel*. Beleggers worden blootgesteld aan twee bronnen van risico in dit model, aandeelkoersbewegingen en veranderingen in volatiliteit. Het model kan worden gebruikt voor het prijzen en *hedgen* van derivaten geschreven op aandelen en vreemde valuta's.

Het proefschrift ontwikkelt een op de (Extended) Kalman filter gebaseerde quasi-maximum likelihood (QML) schattingsmethode voor het derivatenwaarderingsmodel. De methode is transparant, omzeilt het simuleren van optieprijzen tijdens het schatten, genereert een volatiliteitvoorspelling en is bovenal snel. We leggen uit hoe (combinaties van) tijdreeksgegevens van gekwadrateerde aandeelrendementen, *realized volatilities* (RV) en optieprijzen kunnen worden gebruikt voor het destilleren van informatie met betrekking tot de modelparameters en onderliggende volatiliteit.

Verscheidene Monte Carlo studies voor een aantal speciale gevallen van het algemene multifactor affiene SV model laten zien dat indien *at-the-money* (ATM) optiedata wordt meegenomen tijdens het schatten, de methode schijnbaar goed werkt. Optiedata blijkt erg informatief te zijn. Meer specifiek, het gebruik van alleen gekwadrateerde rendementen voor QML schatting is in het algemeen niet aan te raden. De schattingsonzuiverheid en de *mean-squared errors* (MSEs) zijn groot, en de onderliggende latente volatiliteit wordt niet goed uit de data gefilterd. Het gebruik van realized volatilities leidt tot aanzienlijk betere resultaten, hoewel de onzuiverheid en MSEs nog altijd groot zijn. Optiedata leidt tot een dramatische verbetering ten opzichte van het gebruik van alleen RV data. De schattingsonzuiverheid en MSEs zijn substantieel minder, en de onderliggende volatiliteit wordt veel beter blootgelegd. Het combineren van gekwadrateerde rendementen en optiedata leidt vervolgens tot betere schattingsresultaten dan alleen het gebruik van optiedata. In het algemeen lijkt de onzuiverheid bescheiden en is de efficiëntie van de schattingen groter. De combinatie RV – optiedata bevat de meest precieze informatie: De schattings-onzuiverheid is in het algemeen klein, het leidt tot de meest efficiënte schattingen en de geschatte en werkelijke volatiliteiten liggen in het algemeen dicht bij elkaar. Indien voorhanden, is deze combinatie van data dus te prefereren voor het schatten van het derivatenwaarderingsmodel. De simulatieresultaten laten verder zien dat de marktprijs van volatiliteitrisico moeilijk precies te schatten is, ongeacht de data die gebruikt wordt voor het schatten. Dit bevestigt de bevindingen van andere onderzoekers in de literatuur.

Het proefschrift bestudeert tevens de aard van volatiliteitrisico. We beschouwen beleggingsstrategieën die in het bijzonder onderhevig zijn aan volatiliteitrisico, zoals *straddles* en *variance swaps*. Onderzocht wordt hoe beleggers worden gecompenseerd voor het risico van derivatenprijsveranderingen als gevolg van onzekere volatiliteitfluctuaties. We leiden een *verwacht rendement – beta* relatie voor derivaten af, en definiëren de risicopremie op volatiliteit. In overeenstemming met recente bevindingen in de literatuur, laat een empirische analyse zien dat beleggers bereid zijn te betalen voor marktvolatiliteitrisico. In eerste instantie lijkt dit tegenintuïtief.

Onze economische verklaring voor het bestaan van een negatieve volatiliteit-risicopremie kan als volgt samengevat worden. Volgens de *permanent-inkomen hypothese* streven beleggers naar een zo stabiel mogelijke consumptiestroom. Met andere woorden, hun doel is *consumption smoothing*. Hierdoor zijn zij bereid een premie te betalen voor volatiliteitrisico: Aangezien een dalende markt in het algemeen leidt tot minder consumptiemogelijkheden, maar tegelijkertijd ook tot een toename in de volatiliteit (het *leverage effect*), verschaft het aanhouden van delta-neutrale positieve-vega derivaten een verzekering tegen consumptieveranderingen. Immers, daar de waarde van dergelijke derivaten in dat geval toeneemt, doet dit de daling in de consumptiemogelijkheden van de belegger (gedeeltelijk) teniet, wat leidt tot een gelijkmatiger consumptiepatroon. Daar het waarschijnlijk is dat het rendement op dergelijke derivaten negatief correleert met veranderingen in de consumptie van de belegger (als gevolg van het leverage effect), vertaalt dit zich in een negatieve risicopremie op volatiliteit.

De empirische analyse in dit proefschrift confronteert het multifactor affiene SV derivatenwaarderingsmodel met FTSE100 aandelenindex- en optiedata, betreffende de periode oktober 1997 – december 2001. Deze jaren kenmerkten zich door drie ongebruikelijk roerige perioden in de financiële markten: de Azië crisis (herfst 1997), de bijna-teneergang van hedgefonds Long Term Capital Management en de Rusland en continuerende Azië crises (herfst 1998), en 11 september 2001 met zijn nasleep. De data bestaat uit tijdreeksen van gekwadrateerde FTSE100-indexrendementen en drie ATM indexoptiereeksen met verschillende looptijden, een korte-termijn (KT), een middellange-termijn (MT) en een lange-termijn (LT) optiereeks.

Aanvankelijk nemen we aan dat de FTSE100-index data gegenereerd is door de 1-factor OU en 1-factor CIR SV speciale gevallen van ons multifactor model. Wanneer het model gefit wordt aan alleen de gekwadrateerde indexrendementen en de korte-termijn optiereeks, vinden we het volgende. De geobserveerde Black-Scholes impliciete volatiliteiten zijn in het algemeen groter dan de onderliggende verborgen volatiliteiten, dankzij de negatieve marktprijs van volatiliteitrisico. De in de praktijk gangbare interpretatie van een Black-Scholes impliciete volatiliteit zijnde een voorspelling van toekomstige volatiliteit, dient daarom met voorzichtigheid te worden benaderd indien de markt volatiliteitrisico prijst. Het beleggen in FTSE100-index derivaten levert een negatieve volatiliteitrisicopremie op van ongeveer −17.5% op jaarbasis. Het schrijven (*shorten*) van korte-termijn ATM straddles genereert een substantieel verwacht rendement van ongeveer 185%. Deze strategie is niettemin zeer risicovol; straddles zijn echte *bets on volatility*. De volatiliteit van de SV factor hangt duidelijk af van het huidige niveau van de factor zelf. Anders gezegd, de SV factor kenmerkt zich door *level-dependent volatility*. De 1-factor CIR SV specificatie (Heston (1993) volatiliteit) schiet echter tekort in het voldoende adequaat kunnen beschrijven van de zeer snel veranderende volatiliteit die de drie eerdergenoemde perioden van heftige koersschommelingen karakteriseert. Niettemin lijkt de misspecificatie van het 1-factor CIR SV model (geschat met de gekwadrateerde rendementen en de KT optiereeks) in eerste instantie bescheiden. Nadere analyse laat echter zien dat het geschatte model te hoge prijzen genereert voor de langere-termijn opties buiten de steekproef. Wanneer het model geschat wordt met gelijktijdig zowel de gekwadrateerde rendementen als de KT, MT en de LT optiereeksen, wordt de oorzaak van deze overwaardering duidelijk: 1-factor SV optiewaarderingsmodellen zijn niet in staat de rijke volatiliteitdynamiek die impliciet aanwezig is in de gezamenlijke data, adequaat te beschrijven.

Dit leidt ons tot het ondernemen van een econometrische speurtocht op zoek naar het multifactor SV optiewaarderingsmodel binnen de affiene klasse, dat het beste de gezamenlijke data beschrijft. Een aanzienlijke verbetering in de fit wordt verkregen door 2-factor SV specificaties te beschouwen. Beide factoren kenmerken zich door level-dependent volatility. Niettemin is er nog steeds sprake van misspecificatie met betrekking tot onvoldoende volatiliteitdynamiek.

Een 3-factor affiene SV specificatie met één CIR en twee affiene, onafhankelijke volatiliteitfactoren blijkt de gezamenlijke FTSE100 aandelenindex en optiedata het best te beschrijven. Elk van de drie factoren kenmerkt zich dus door level-dependent volatility. Het model beschrijft de in de data aanwezige volatiliteitdynamiek adequaat, en de fit van de termijnstructuur van de volatiliteit is goed. De optieprijsfouten concentreren zich rondom nul en zijn in het algemeen klein. De geschatte 3-factor SV volatiliteiten reageren sneller op nieuws dan de 1-factor SV volatiliteiten, met name in tijden van plotselinge marktstress.

De karakteristieken van de drie volatiliteitfactoren x_1, x_2 en x_3 verschillen in belangrijke mate. Factor x_1 is extreem persistent en vertoont vrijwel random-walk gedrag. x_1 fluctueert rondom de lange-termijn gemiddelde FTSE100-indexvolatiliteit, en bepaalt de lange-termijn trend in de volatiliteit. Factor x_2 is veel sneller *mean-reverting*. Schokken in x_2 hebben een *half-life* van ongeveer 3 maanden. x_2 bepaalt de middellange-termijn volatiliteittrend. Factor x_3 keert zeer snel naar zijn gemiddelde terug. Schokken in x_3 verliezen de helft van hun uitwerking in ongeveer tien dagen en hebben de grootste variantie. Factor x_3 bepaalt grote volatiliteit-veranderingen in relatief korte tijdsperioden, en wordt geassocieerd met de korte-termijn trend in de volatiliteit.

De volatiliteitfactoren x_1, x_2 en x_3 beïnvloeden de prijzen van opties van verschillende looptijd op verschillende wijze. Een schok in de lange-termijn SV factor

x_1 heeft soortgelijke invloed op alle optieprijzen, ongeacht looptijd. Ook de middellange-termijn SV factor x_2 beïnvloedt alle optieprijzen, maar in afnemende mate naarmate de looptijd van de optie langer is. Schokken in de zeer snelle mean-reverting factor x_3 hebben voornamelijk invloed op de prijzen van korte-termijn opties. De impact van x_3 neemt snel af zodra de looptijd van de optie toeneemt. Aangezien x_3 zo snel naar zijn gemiddelde terugkeert, hebben de schokken in x_3 de neiging uit te middelen over een voldoende lange tijdshorizon, met als gevolg een slechts marginale uitwerking op de prijzen van opties met lange looptijden.

De SV factoren x_1, x_2 en x_3 hebben een verschillende invloed op de vorm en de evolutie van de volatiliteittermijnstructuur door de tijd. x_1 is met name verantwoordelijk voor veranderingen in het algemene niveau van de termijnstructuur. Veranderingen in de helling worden voornamelijk geassocieerd met x_2. Factor x_3 is sterk gerelateerd aan de dynamiek in de convexiteit van de termijnstructuur.

Helaas zijn de schattingen van de marktprijzen van risico behorende bij elk van de SV factoren onnauwkeurig. We missen een intuïtie voor het vinden van een positieve prijs van risico behorende bij de middellange-termijn SV factor x_2; de prijzen van risico geassocieerd met x_1 en x_3 worden beide negatief geschat. De risicopremie behorende bij x_2 bedraagt 3.9% op jaarbasis. De risicopremie op de snelle mean-reverting factor x_3 is veel meer negatief (-22%) dan de premie op de lange-termijn factor x_1 (-0.4%). Deze premies impliceren een veel groter negatief verwacht rendement op korte-termijn dan op lange-termijn ATM straddles. Dit is in overeenstemming met het grotere risico dat gepaard gaat met beleggen in korte-termijn straddles. Het feit dat we negatief verwachte rendementen op long straddles vinden komt overeen met beleggers die negatief gecompenseerd worden voor volatiliteitrisico.

Binnen de affiene klasse van SV modellen, is het 3-factor affiene SV optiewaarderingsmodel met één CIR en twee affiene volatiliteitfactoren dus het meest geschikt voor de gezamenlijke FTSE100 aandelenindex- en optiedata. Specificatietoetsen laten echter zien dat er nog steeds ruimte voor modelverbetering mogelijk is. (Zie sectie 7.2.1 van hoofdstuk V voor een meer complete discussie.) Ten eerste hebben we het leverage effect niet gemodelleerd. Gegeven onze focus op louter ATM opties, is het onwaarschijnlijk dat het hebben genegeerd van leverage het merendeel van onze bevindingen ongeldig verklaart. Zoals hoofdstuk VI laat zien, wordt de ATM volatiliteittermijnstructuur maar in beperkte mate beïnvloed door leverage. Niettemin, indien het doel het prijzen en hedgen van *in-* en *out-of-the-money* aandelenopties is, kan leverage niet zomaar worden genegeerd. Het multifactor affiene SV derivatenwaarderingsmodel met bijbehorende state space schattingsmethode kan eventueel worden uitgebreid ten einde het leverage effect expliciet mee te nemen. Ten tweede blijkt het multifactor affiene SV model niet in staat de in de data aanwezige drie perioden van zeer snel veranderende volatiliteit volledig te beschrijven, zelfs niet met drie factoren. Een uitbreiding van het model dat sprongen (*jumps*) in de volatiliteit toelaat lijkt veelbelovend in dit opzicht. Samen met de overige suggesties voor verder onderzoek geopperd in hoofdstuk VII, vormt dit interessant werk voor toekomstig onderzoek.

Tinbergen Institute Research Series

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

314. T. VERHAGEN, *Towards understanding online purchase behavior.*
315. R. HOEKSTRA, *Structural change of the physical economy. Decomposition analysis of physical and hybrid-units input-output tables.*
316. R.K. AIDIS, *By law and by custom: Factors affecting small and medium sized enterprises during the transition in Lithuania.*
317. S. MANZAN, *Essays in nonlinear economic dynamics.*
318. K. OLTMER, *Agricultural policy, land use and environmental effects: Studies in quantitative research synthesis.*
319. H.J. HORBEEK, *The elastic workfloor. About the implementation of internal flexibility arrangements.*
320. P. HOUWELING, *Empirical studies on credit markets.*
321. E. MENDYS, *Essays on network economics.*
322. J. NOAILLY, *Coevolutionary modeling for sustainable economic development.*
323. Y.V. VELD-MERKOULOVA, *Essays on futures markets and corporate spin-offs.*
324. F.J. WILLE, *Auditing using Bayesian decision analysis.*
325. D. HUISMAN, *Integrated and dynamic vehicle and crew scheduling.*
326. S. WANG, *Global climate change policies: An analysis of CDM policies with an adapted GTAP model.*
327. M.W. VAN GELDEREN, *Opportunity entry performance. Studies of entrepreneurship and small business.*
328. W. VAN WINDEN, *Essays on urban ICT policies.*
329. G.J. KULA, *Optimal retirement decision.*
330. R.J. IMESON, *Economic analysis and modeling of fisheries management in complex marine ecosystems.*
331. M. DEKKER, *Risk, resettlement and relations: Social security in rural Zimbabwe.*

332. A. MULATU, *Relative stringency of environmental regulation and international competitiveness.*
333. C.M. VAN VEELEN, *Survival of the fair: Modelling the evolution of altruism, fairness and morality.*
334. R. PHISALAPHONG, *The impact of economic integration programs on inward foreign direct investment.*
335. A.H. NÖTEBERG, *The medium matters: The impact of electronic communication media and evidence strength on belief revision during auditor-client inquiry.*
336. M. MASTROGIACOMO, *Retirement, expectations and realizations. Essays on the Netherlands and Italy.*
337. E. KENJOH, *Balancing work and family life in Japan and four European countries: Econometric analyses on mothers' employment and timing of maternity.*
338. A.H. BRUMMANS, *Adoption and diffusion of EDI in multilateral networks of organizations.*
339. K. STAAL, *Voting, public goods and violence.*
340. R.H.J. MOSCH, *The economic effects of trust. Theory and empirical evidence.*
341. F. ESCHENBACH, *The impact of banks and asset markets on economic growth and fiscal stability.*
342. D. LI, *On extreme value approximation to tails of distribution functions.*
343. S. VAN DER HOOG, *Micro-economic disequilibrium dynamics.*
344. B. BRYS, *Tax-arbitrage in the Netherlands evaluation of the capital income tax reform of January 1, 2001.*
345. V. PRUZHANSKY, *Topics in game theory.*
346. P.D.M.L. CARDOSO, *The future of old-age pensions: Its implosion and explosion.*
347. C.J.H. BOSSINK, *To go or not to go…? International relocation willingness of dual-career couples.*
348. R.D. VAN OEST, *Essays on quantitative marketing models and Monte Carlo integration methods.*
349. H.A. ROJAS-ROMAGOSA, *Essays on trade and equity.*
350. A.J. VAN STEL, *Entrepreneurship and economic growth: Some empirical studies.*
351. R. ANGLINGKUSUMO, *Preparatory studies for inflation targeting in post crisis Indonesia.*
352. A. GALEOTTI, *On social and economic networks.*
353. Y.C. CHEUNG, *Essays on European bond markets.*

354.A. ULE, *Exclusion and cooperation in networks.*
355.I.S. SCHINDELE, *Three essays on venture capital contracting*.
356.C.M. VAN DER HEIDE, *An economic analysis of nature policy.*
357.Y. HU, *Essays on labour economics: Empirical studies on wage differentials across categories of working hours, employment contracts, gender and cohorts.*
358.S. LONGHI, *Open regional labour markets and socio-economic developments: Studies on adjustment and spatial interaction.*
359.K.J. BENIERS, *The quality of political decision making: Information and motivation.*
360.R.J.A. LAEVEN, *Essays on risk measures and stochastic dependence: With applications to insurance and finance.*
361.N. VAN HOREN, *Economic effects of financial integration for developing countries.*
362.J.J.A. KAMPHORST, *Networks and learning*.
363.E. PORRAS MUSALEM, *Inventory theory in practice: Joint replenishments and spare parts control.*
364.M. ABREU, *Spatial determinants of economic growth and technology diffusion.*
365.S.M. BAJDECHI-RAITA, *The risk of investment in human capital*.