Essays on Variance Risk

Peter H. Gruber

Dissertation submitted for the degree of
Ph.D. in Economics

Institute of Finance
Faculty of Economics
Università della Svizzera Italiana

June 10, 2015

Thesis Committee:
Prof. David Bates, University of Iowa
Prof. Patrick Gagliardini, University of Lugano
Prof. Markus Leippold, University of Zurich
Prof. Fabio Trojani, University of Lugano

Abstract

My PhD thesis consists of three papers which study the nature, structure, dynamics and price of variance risks. As tool I make use of multivariate affine jump-diffusion models with matrix-valued state spaces.

The first chapter proposes a new three-factor model for index option pricing. A core feature of the model are unspanned skewness and term structure effects, i.e., it is possible that the structure of the volatility surface changes without a change in the volatility level. The model reduces pricing errors compared to benchmark two-factor models by up to 22%. Using a decomposition of the latent state, I show that this superior performance is directly linked to a third volatility factor which is unrelated to the volatility level.

The second chapter studies the price of the smile, which is defined as the premia for individual option risk factors. These risk factors are directly linked to the variance risk premium (VRP). I find that option risk premia are spanned by mid-run and long-run volatility factors, while the large high-frequency factor does not enter the price of the smile. I find the VRP to be unambiguously negative and decompose it into three components: diffusive risk, jump risk and jump intensity risk. The distinct term structure patterns of these components explain why the term structure of the VRP is downward sloping in normal times and upward sloping during market distress. In predictive regressions, I find an economically relevant predictive power over returns to volatility positions and S&P 500 index returns.

The last chapter introduces several numerical methods necessary for estimating matrix-valued affine option pricing models, including the Matrix Rotation Count algorithm and a fast evaluation scheme for the Likelihood function.

Acknowledgments

The completion of this thesis marks the end of an exiting journey that turned a nuclear physicist into a financial economist. This itinerary would not have been possible without the support of many great people whom I have I met on the way. First and foremost I want to thank my advisor and friend Fabio Trojani, who offered me a PhD position over coffee one evening in St. Gallen. Many more coffees followed, over which he shared his ideas, his wisdom and his precise way of thinking with me. Having a polymath like Fabio as supervisor raised the bar for my thesis substantially, but I would not want to have it any other way.

I am also grateful to the members of my dissertation committee: David Bates, Patrick Gagliardini and Markus Leippold for their willingness to discuss my work and for their kind support.

I want to express my gratitude to my current and former colleagues, friends and coauthors: Andrea Vedolin, Elisa Ossola, Ilaria Piatti, Andras Sali, Claudio Ortelli, Claudio Tebladi, Davide Lavecchia, Evert Wipplinger, Paul Schneider, Piotr Orlowski and all the others. I am also grateful to the Swiss National Science Foundation for the financial support that I have received.

This thesis is dedicated to Nicole, who has always supported me, especially in the inevitable low moments. Nicole, this has not been our first adventure together and it will certainly not be our last one!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

OVER the last 40 years, variance has emerged from a phenomenon largely confined to the context of stock prices to an entirely new, separate asset class. My PhD thesis consists of three papers which study the nature, structure, dynamics and price of variance risks. My tools of choice are multivariate affine jump-diffusion models with matrix-valued state spaces.

A growing body of literature has tried to solve a long list of issues in option pricing. Returns are usually negatively correlated with changes in volatility (Hull and White (1988) and Heston (1993)), they are non-gaussian and subject to jumps (Merton 1976). Volatility is nowadays understood as a multi-factor phenomenon (Bates 2000) and a fraction of the variation of the volatility surface is orthogonal to changes in the volatility level (Christoffersen, Heston and Jacobs 2009). There is clear empirical evidence that standard observable properties of the implied volatility surface such as implied volatility level, the option-implied skew or the volatility term structure, have time-varying correlations. So far, however, the literature has focused on lower-dimensional models with an independent factor structure that links the volatility term structure and the option-implied skew in a restrictive way to the volatility level.

My first paper, *Three make a Dynamic Smile – Unspanned Skewness and Interacting Volatility Components in Option Valuation*, joint work with Fabio Trojani, proposes an affine three-factor model for index options with novel structural properties. It is based on the Matrix Affine Jump Diffusion (MAJD) process introduced in Leippold and Trojani (2008). The model allows for dynamic interactions between risk factors, which gives rise to mutually-exciting risks. It also allows for unspanned skewness and volatility term structure effects. The model has closed form solutions and allows for pricing via transform methods. It naturally nests many important affine models in the literature, such as Bates (2000) or Heston (1993), both of which have independent volatility components and a stochastic skewness that is linearly related to the volatility.

A preliminary empirical investigation of S&P 500 option data, identifies unspanned skewness and term structure effects, i.e., a significant variability of implied volatility skewness and term structure conditional on the volatility level. We then estimate our model under the risk-neutral measure on five years of option data using a nonlinear least squares (NLLS) approach with error group specific heteroskedasticity, as pioneered in Bates (2000). Unlike Bates, the problem of joint estimation of parameters and states is solved using a computationally intensive nested algorithm, which optimizes over the parameter vector given the optimal state. We then evaluate our model on 15 years of option data, pricing some 600'000 option contracts using a single set of parameters.

In order to better understand the interplay of state components and parameters in creating unspanned volatility effects, we introduce a decomposition of the state matrix into a volatility level factor and two bounded volatility structure factors. Based on this decomposition,it is possible to draw the set of admissible combinations of skewness and volatility term structure for given volatility levels for different models.

We also study the pure diffusive version of our model (see da Fonseca, Grasselli and Tebaldi (2008)). While such a stylized model is less realistic, we can use the method of Durrleman (2010) to calculate closed-form approximations of important properties of the volatility surface and link them to specific parameters and state components.

The main findings are the following. Our model reduces pricing errors relative to the benchmark two-factor Bates (2000) model by between 20% and 30%, depending on the metric. These improvements follow from the failure of two-factor models to generate sufficient volatility-unrelated variability in the implied volatility skew and term structure. There are several indicators for the important contribution of the matrix specification of the state space to these results. First, improvements increase out-of sample, i.e., the choice of parameters has less impact on the pricing errors. Second, there is little variability in the overall fit when varying in- and out-of sample periods, as long as the in-sample period contains sufficiently varied economic conditions.

When trying to explain the magnitude of the relative improvement, no apparent link to volatility can be found, but a positive link to the size of the benchmark model's daily RMS error and to a proxy of unspanned skewness implied by our model. This suggests that our model improves precisely on the specification of volatility-unrelated smile dynamics. In a time series analysis, the largest gains in pricing performance are found in periods of financial distress.

Understanding the properties of the market price of volatility risk is an important issue in financial economics. The literature has reached the consensus that unexpected shocks in aggregate market uncertainty are priced in the form of a time-varying negative variance risk premium (VRP), see Carr and Wu (2009a). However, less is known about which volatility factors generate the premium.

My second paper, *The Price of the Smile and Variance Risk Premia*, joint work with Fabio Trojani and Claudio Tebaldi, aims at exploiting the information embedded in index options to study (i) the dynamics and price of individual option risk factors, (ii) the size, structure and dynamics of the VRP and (iii) the link between option risk factors, their price, the VRP, and equity returns. We perform a joint estimation of the physical and risk-neutral dynamics of a matrix affine jump diffusion (MAJD) model in order to study option-implied risks, i.e., risks generated by unexpected variations of the state variables that drive the volatility surface. From the risk-neutral and physical dynamics of the individual risk factors it is possible to calculate the risk premia paid by investors for an exposure to such risks. We call these premia the price of the smile, as they span the risk premium of a shock in the price of volatility. We then study the size and dynamics of the variance risk premium (VRP), and its term structure. We decompose the VRP into three constituents: compensation for diffusive volatility, for pure jump risk (i.e. exposure to jumps in returns) and for jump intensity risk (i.e. exposure to the risk that the rate of jumps in returns may change). We are then able to link each constituent of the VRP to the option-implied risk factors and their prices, as identified by our model.

The starting point of our model is the risk-neutral dynamics formulated in "Three make a Dynamic Smile" to which an affine change of measure is added. As our interest is focused on the information embedded in options, we deliberately define a minimal stochastic discount factor and leave the equity risk premium unspecified. This allows makes it possible to make predictions on the equity premium, based exclusively on second-moment information.

The estimation of our model is based on a simple two-step procedure, which exploits the information in a panel of option prices and excess returns to variance positions, but contains no direct information on index returns. In a first step, we use an extended Kalman filter to estimate the physical and risk neutral dynamics of the latent variance process and the risk-neutral properties of the return jump component. In the second step, we estimate the parameters of the pure jump VRP via a simple arbitrage-free regression.

Our results are as follows. First, despite the additional specification of a $\mathbb{P}$-dynamics, the updated estimation procedure does not degrade pricing performance or fit relative to "Three make a Dynamic Smile". Each of the three state components still owns its useful interpretation in terms of observable properties of the volatility surface. The largest and least persistent factor can be interpreted as a volatility level, the most persistent one as a term structure factor, while the remaining factor captures changes in skewness unrelated to the volatility level.

Second, the market price of the smile is completely spanned by the two most persistent volatility factors in a cascading structure, where the market price of a factor is always more persistent than the factor itself. Interestingly, the high-frequency level factor, which is the largest volatility factor, does not enter the price of the smile.

Third, our risk factors imply a highly time-varying and unambiguously negative VRP, ranging between annualized $-16\%^2$ and zero for a monthly investment horizon. We confirm that the VRP is largest in times of market turmoil, when the price of insurance is large. Our decomposition of the VRP also shows that the variance risk premium is largely driven by jump risk, in line with the results of Bollerslev and Todorov (2011). More precisely, the VRP is almost exclusively attributed to pure jump variance risk over short horizons, while diffusive volatility risk and jump intensity together risk explain about half the VRP for a twelve month horizon.

Fourth, the term structure of VRP is mostly downward sloping, reflecting a higher relative price of long-run market insurance. In contrast to, e.g., Ait-Sahalia, Karaman and Mancini (2012) we find that the term structure of the VRP can change sign in a systematic way and can become strongly upward sloping, especially in times of market distress. This changing sign of the VRP term structure is explained by the interplay of its components: the term structure of pure jump risk premia is almost always increasing, while the term structures of diffusive and jump intensity risk premia are usually decreasing.

Finally, we investigate the predictive power of our option-implied risk factors for realized returns on variance positions and market returns. We find an economically relevant predictive power for S&P 500 index returns, consistent with the results in Bollerslev, Tauchen and Zhou (2009a), and a significant predictive power over volatility returns, with a dominating contribution from the two most persistent risk factors. Our affine specification of variance risk premia is preferred by the out-of-sample predictability results, in which the largest degree of predictability is obtained for the model-implied variance risk premium predictions. This evidence suggests that low-frequency volatility factors, which span the market price of the smile and the term structure of the VRP, also contain useful information about the price of market risk.

While *Three make a Smile* and *The Price of the Smile and Variance Risk Premia* are both based on matrix affine jump diffusion models, they differ in scope and methodology. *Three Make a Smile*, the first of the two papers, lays the foundations and focuses on the number and role of volatility factors, on unspanned skewness and on cross-sectional pricing implications, by estimating the model under the risk-neutral measure via nonlinear least squares. The main goal is to understand the *structure* of index-option volatility.

On the other hand, *The Price of the Smile* concentrates on the *dynamics* of volatility factors, on their price and on the variance risk premium, estimating the physical and risk-neutral dynamics using an extended Kalman filter. The main objective is to understand the link between option risk factors, the variance risk premium and equity returns.

The matrix nature of MAJD models creates a unique set of computational difficulties for analysis and estimation. Using off-the-shelf methods, the estimation of an MAJD option pricing is numerically unstable and would take years. My third paper, *Eliciting* *a Smile – Numerical Methods for Option Pricing with Matrix Affine Jump Diffusions*, focuses on the computational aspects of the estimation of this class of models. It proposes a collection of new and improved numerical methods that render their estimation feasible with good precision in a reasonable amount of time. They fall into three categories:

Numerics. The characteristic function of the matrix affine jump diffusion model is available in closed form as long as no jumps in volatility occur (see Leippold and Trojani (2008)). Its evaluation involves a matrix logarithm, which is not uniquely defined, as the complex parts of its eigenvalues are periodic. I show that standard numerical platforms choose the wrong plane by default, which leads to a significant bias in the option price when using transform methods. The error is larger for longer maturities and when the high-frequency volatility component is dominant. A similar phenomenon has been documented for Heston-type models by Lord and Kahl (2010). To solve the problem of selecting the correct branch on the complex matrix logarithm, I propose the Matrix Rotation Count (MRC) algorithm. In an unrelated numerical problem, I identify a numerical instability of the above-mentioned matrix logarithm for extreme arguments of the Laplace transform. This problem is solved by the correct choice of integration limits for the Fourier integral.

Optimization. When estimating the model in a maximum likelihood context, little is ex ante known about the plausible range for the parameter matrices, especially their out-of-diagonal components. This makes it difficult to find useful starting values for classical optimization algorithms. I also show with a counter-example that the maximization problem is non-convex. Both challenges are overcome by a two step hybrid optimization scheme composed of Differential Evolution (Storn and Price 1997) and the Nelder-Mead simplex algorithm. To enforce admissibility constraints of the parameter matrices, such as negative definiteness, I introduce two convenient reparametrizations of the parameter and state matrices.

Execution time. With modern computer hardware, speed is normally not an issue in estimating option pricing models. The MAJD model is an exception, as its characteristic function involves matrix exponentials and logarithms, which are computationally more costly than their scalar counterparts by a factor of 30 to 40. Beyond the MRC algorithm, which enables the use of transform methods, I identify three directions for improvement, each of which reduces the execution time by at least an order of magnitude.

First, I replace the Fast Fourier Transform of Carr and Madan (1999) by the Cosine algorithm of Fang and Oosterlee (2008). The latter reduces the number of required evaluations of the characteristic function from 4096 to 200 by avoiding oscillating terms. Second, I introduce a fast evaluation scheme for the likelihood function. This scheme uses an efficient decomposition of the characteristic function into a fast part depending exclusively on the state and a computationally costly part depending on parameters and maturity. In order to save time, the latter is pre-calculated, which requires a choice of integration limits that is independent of the volatility level, unlike the original algorithm of Fang and Oosterlee (2008). I propose an approximation for these limits and show that the overall approximation error is below one tenth of a volatility basis point for all admissible state combinations and volatilities between 10% and 63%. Third, I devise a parallel version of the Differential Evolution algorithm of Storn and Price (1997) and run it on a 64-core cluster, reducing the execution time by a further factor of 50.

Chapter 2

Three make a Dynamic Smile

Unspanned skewness and interacting volatility components in option valuation

We study a new class of three-factor affine option pricing models, featuring interdependent volatility risks and a stochastic skewness component conditionally unrelated to the volatility. These properties allow us to improve on the modelling of the implied volatility smile of standard affine models along two main dimensions. First, they enhance the specification of short and long term skew dynamics largely unrelated to volatility shocks. Second, they produce a broader range of term structures of implied volatility skews, which are potentially more consistent with the data.

We specify our models using three distinct stochastic components for the joint dynamics of return volatility and skewness: Two components capture short and long run volatility risks, while the third component captures stochastic skewness effects not related to volatility shocks. In contrast to standard affine models, we introduce interdependent risks, specified by a multivariate dynamics, in which the persistence and local variance of the volatility components depends on the degree of return skewness, and vice versa. Methodologically, we borrow from Leippold and Trojani (2008) and specify our model using a matrix affine jump diffusion (AJD). In this way, we achieve two objectives at the same time. First, we preserve a good degree of model tractability, with efficient pricing formulae for plain vanilla options, computed by means of standard transform methods. Second, we can nest as special cases a number of two- and three-factor affine models in the literature, such as Bates (2000) jump diffusion model or two-factor Heston (1993)-type models. These benchmark models have (i) independent volatility components and (ii) a skewness dynamics that is a function only on the volatility dynamics, which excludes volatility-unrelated stochastic skewness effects. Therefore, they are natural models against which we can benchmark the incremental pricing accuracy of our option valuation framework.

We estimate our three-factor model together with the benchmark models, using S&P 500 index options data from January 1996 to September 2009, and obtain a number of novel findings for the option pricing literature. First, interdependent volatility and volatility-unrelated skewness dynamics are well consistent with S&P 500 index option smiles: A test of the null hypothesis that these features are not present in S&P 500 index option data is rejected with a high degree of statistical significance. Second, our models improve on benchmark affine two-factor option valuation models, by reducing pricing errors on average by 20% out-of-sample. The reliability in pricing performance is also improved, with on average lower standard deviations of pricing errors by 27% out-of-sample. Therefore, these fit improvements are unlikely a consequence of overfitting effects related to the higher-dimensional (three-factor) state space of our model. Third, standard affine three-factor models with independent volatility components tend to have poor pricing performance out-of-sample, indicating a likely misspecification of their state dynamics. For instance, we find that the improvement in out-of-sample pricing performance of our three-factor models relative to a three-factor Heston-type model is on average about 21%. Fourth, while the improvements in pricing performance relative to benchmark models are quite consistent over time, the largest improvements tend to arise during periods of financial crises or market distress, like, e.g., during the Russia debt crisis in 1998 and the more recent subprime crisis: While the average improvement in out-of-sample model fit is about 20%, daily pricing improvements in such crisis periods are often larger that 30% and can be, in some cases, even above 50%. This finding suggests volatility-unrelated skewness as a potentially useful reduced-form risk factor, which is better able to reproduce some of the crisis-related dynamics of index option smiles.

To understand the motivation of modelling (i) a stochastic skewness component unrelated to volatility shocks and (ii) interdependent volatility dynamics, we start by analyzing the implied volatility surface of S&P 500 index options. Figure 2.1 depicts the relation between one-month maturity skew and at-the-money volatility term structure[1] in the data (grey points). In order to better isolate smile components that are largely unrelated to variations of the level of volatility, each plot is stratified with respect to the volatility level, ranging from 0.16 to 0.28.

First, we observe that in the data, the variability of the skew and term structure proxies is quite substantial for all volatility levels. This feature suggests that a fraction of these variations is not exclusively explained by the level of volatility, indicating the potential presence of a volatility dynamics driven by additional sources of time-varying

---

[1] See Sec. A.2.1 for details on the definition and the computation of the implied volatility skewness and term structure proxies used in Figure 2.1.

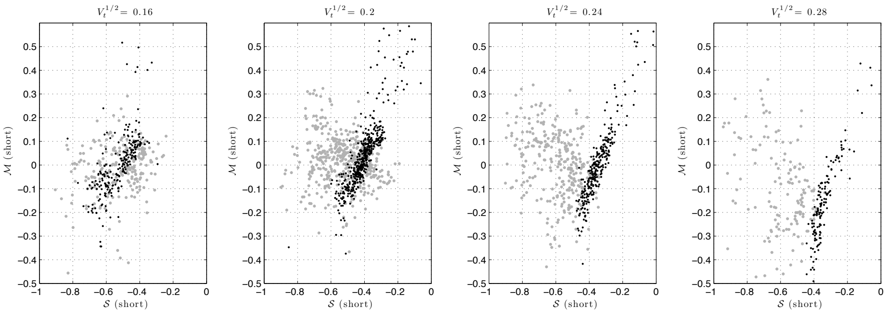
risk. Second, stochastic volatility models without unspanned skewness factors tend to fail in generating (i) the large variability of the skew of index options and (ii) the joint relation between slope of the smile and term structure across different volatility states. As a first illustration of this argument, we estimate two factor models of the Heston- and Bates-type and plot the model-implied values for volatility skew and term structure in Figure 2.1 as black points. While these models generate a degree of term structure variation similar to the one in the data, we find that they tend to imply (i) a rather limited degree of variation along the skewness dimension and (ii) a comovement of skewness and term structure that is not fully supported by the data. This empirical evidence suggests the presence of a skew dynamics that is only weakly linked to either the level or the term structure of the implied volatility smile. Therefore models in which return skewness is completely spanned by shocks to the volatility might be overly restrictive for an adequate specification of these dynamic aspects of the smile. Third, the tight link between skew and term structure of two-factor option valuation models with independent volatility components might also imply an overly simplified term structure of volatility skews. This feature is illustrated in Figure 2.2, top panels, where we plot the twelve month skew of the smile against the one month skew in the data (gray points). Each scatter plot is again stratified with respect to different levels of the at-the-money implied volatility.

In the data, the degree of variability of the skew at maturities of one and twelve months is similar, especially for the low volatility state (left Panels in Figure 2.2). The model-implied twelve month and one month skews of the two-factor models (black points in Panels A and B) feature, as expected, a lower variability than the data. At the same time, it appears that these models have an even larger difficulty to generate (i) a sufficient variability of twelve month skews and (ii) a degree of comovement between short and long term skews similar to the data.

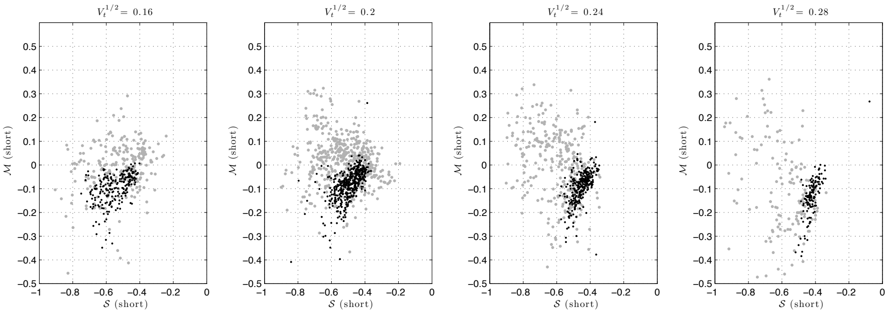
Panel A and B of Figures 2.1 and 2.2 highlight the difficulty of two-factor models to fit relevant features of the volatility surface. To which extent do our models improve the fit along these particular dimensions? Panels C and D indicate that our models improve on the two factor models mainly in two directions. First, Figure 2.1 shows that they generate a fitted co-movement of short term skews and term structures more consistent with the data. Second, Figure 2.2 shows that the models also imply an additional degree of variability in fitted

Our work borrows from a large literature documenting the time variation of the equity volatility and its negative co-movement with returns. The state of the art in the option pricing literature specifies the underlying return dynamics as driven by several components that follow independent volatility processes, each negatively related to re-

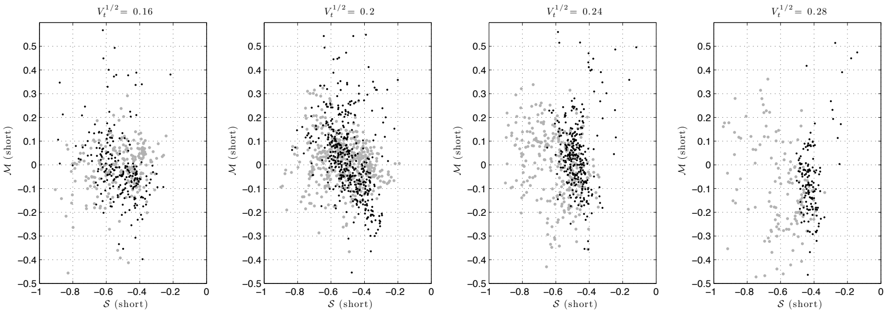
Panel A: Fit of model $SV_{2,0}$



Panel B: Fit of model $SVJ_{2,0}$



Panel C: Fit of model $SV_{3,1}$



Panel D: Fit of model $SVJ_{3,1}$

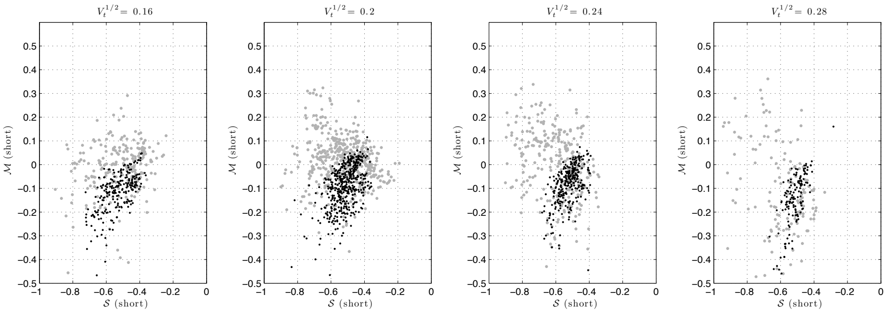
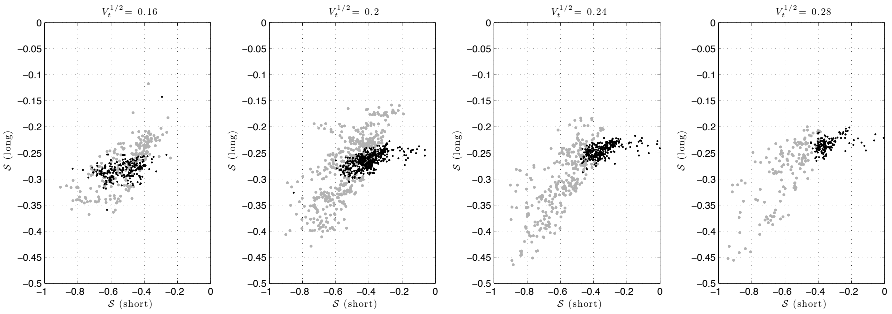
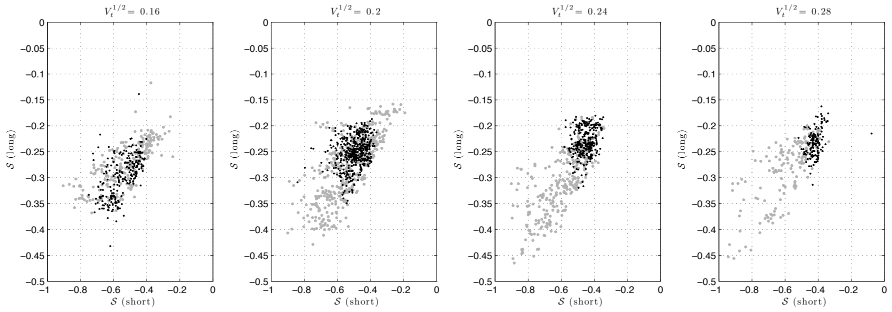


Figure 2.1: Short term skew $\mathcal{S}(short)$ versus term structure $\mathcal{M}(short)$ per trading day. Grey dots: data. Black dots: Fitted values of a two factor Heston-type model ($SV_{2,0}$), Bates (2000) model ($SVJ_{2,0}$), our pure diffusion model ($SV_{3,1}$) and our model with jumps in returns ($SVJ_{3,1}$). In each plot, we select observations corresponding to a short term at the money implied volatility of ±5% around the specified level, i.e., 19%-21% for the second plot of each panel.

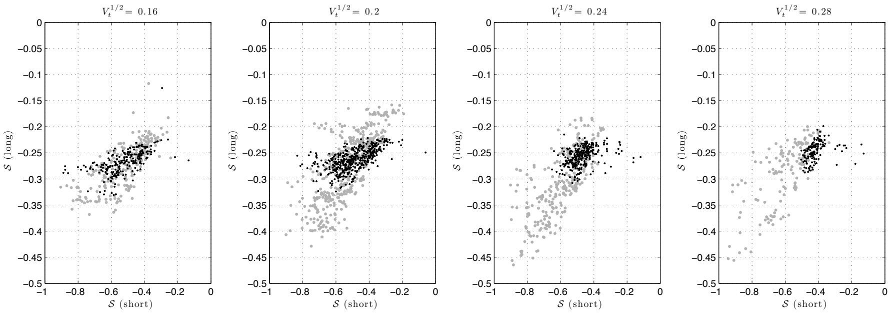
Panel A: Fit of model $SV_{2,0}$



Panel B: Fit of model $SVJ_{2,0}$



Panel C: Fit of model $SV_{3,1}$



Panel D: Fit of model $SVJ_{3,1}$

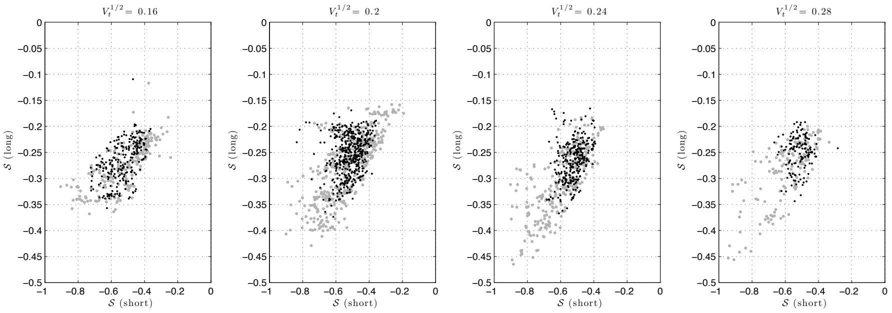


Figure 2.2: Short term skew $\mathcal{S}(short)$ versus long term skew $\mathcal{S}(long)$ per trading day. Grey dots: Black dots: Fitted values of a two factor Heston-type model ($SV_{2,0}$), Bates (2000) model ($SVJ_{2,0}$) and our pure diffusion model ($SV_{3,1}$) and our model with jumps in returns ($SVJ_{3,1}$). In each plot, we select observations corresponding to a short term at the money implied volatility of ±5% around the specified level, i.e., 19%-21% for the second plot of each panel.

turn shocks.[2] A number of recent studies shows that these models tend to perform better than single factor stochastic volatility models in pricing equity index options. In contrast to single-factor models, multiple component models can generate a degree of stochastic skewness that can help to capture part of the time variation of the smile along the moneyness dimension. Moreover, when the distinct volatility components feature different persistence properties, these models also tend to better capture the behaviour of the implied volatility surface along the maturity dimension. Bates (2000) specifies two jump-diffusion components driven by independent volatility processes and studies empirically the relative performance of pure diffusion models and models augmented by Poisson-normal jumps. Using S&P 500 futures option data from 1988 to 1993, he documents the negative skew of the smile after the 1987 crash and concludes that models with jumps better reconcile return and option data. Christoffersen et al. (2009) focus on the ability of a pure diffusion version of the Bates (2000) model to explain the option implied volatility dynamics. They document that these models improve the pricing performance relative to single factor volatility settings, both in-sample and out-of-sample, because they imply a higher degree of flexibility in modelling the conditional skewness and kurtosis of returns in dependence of the overall level of the volatility. Using time-changed Levy processes, Huang and Wu (2004) study two-component jump diffusion models with different types of jump specifications. They document that models with high frequency jumps and volatility variations deriving from both the instantaneous variance of the diffusion component and the arrival rate of the jump component better capture the behaviour of S&P 500 index options. They also find that the diffusion induced volatility exhibits a larger instantaneous variation, but the jump induced volatility features a much higher persistence. Finally, Carr and Wu (2009b) propose a three-component model based on three different sources of variation in volatility: Time varying financial leverage, time-varying business risk and self-exciting market behaviour. The first component follows a CEV-type dynamics in order to model a dependence of the volatility on the level of financial leverage. The second component specifies volatility feedback effects modelled by a Heston (1993)-type volatility model. The third component models self-exciting market behaviour using a high frequency pure-jump Levy-process. The model

---

[2] The literature in this domain is too large to be reviewed exhaustively here. Early papers introducing single-factor volatility models with a correlation between returns and volatility include (Heston 1993), (Hull and White 1988) and (Melino and Turnbull 1990). Leverage effects and jump driven skewness for modelling volatility have been studied in (Bakshi, Cao and Chen 1997), (Bates 1996), Backus, Foresi, Li and Wu (1997), (Nandi 1998), (Chernov and Ghysels 2000), (Pan 2002), (Jones 2003), (Eraker, Johannes and Polson 2003), (Carr and Wu 2004), (Eraker 2004), Li and Pearson (2008), (Broadie, Chernov and Johannes 2007) and (Carr and Wu 2007), among others. Specification based on multi-factor volatility were addressed in (Bates 2000), (Duffie, Pan and Singleton 2000), (Huang and Wu 2004), (Christoffersen et al. 2009), (Carr and Wu 2009b), da Fonseca et al. (2008) and Leippold and Trojani (2008), among others.

is estimated using about a decade of over-the-counter equity index options data and is shown to perform well in pricing equity index options.

All these specifications model the equity index return as a sum of independent components driven by independent volatility processes. The specification in Carr and Wu (2009b) models three distinct sources of volatility variation, while all other models are based on a two factor volatility dynamics. Their empirical results support three factor models as convenient settings to describe the overall shape of the implied volatility surface of index options. We borrow from this insight and specify a three factor state dynamics for the volatility, but we use a completely different approach with distinct implications, starting from the family of MAJD introduced in Leippold and Trojani (2008) with a matrix-valued latent state variable. This symmetric 2×2 matrix contains two volatility factors and a third component linked to stochastic skewness variations that are not spanned by shocks to the two volatility factors. In this sense, our model comprises a component for volatility-unrelated (unspanned) stochastic skewness. This model feature allows us in the first place to obtain a wider range of model-implied degrees of risk neutral skewness, thus improving along the moneyness dimension the fit of the implied volatility smile. Furthermore, our dynamics for the volatility components admits feedback effects. The volatility risk factors in our model interact dynamically and imply a more flexible specification for the dynamics of the skewness term structure. Finally, we introduce a new representation of the state space of MAJD, which allows for a convenient interpretation of the structural and pricing implications of our framework.

The article proceeds as follows. Section 2 introduces our modelling approach and discusses key properties of our model specifications. Starting from a MAJD, it derives a class of three-component option valuation models, with interacting two-factor volatility dynamics and a volatility-unrelated stochastic skewness component. It also shows that a variety of multi-factor affine option pricing models in the literature are special cases of our setting. Section 3 introduces our model estimation procedure, presents estimation results, as well as the in-sample and out-of-sample model fit analysis. Using the estimated parameters, Section 4 analyzes in more detail the main structure and theoretical features of our setting, relative to a number of benchmark affine models. Section 5 concludes.

2.1 Model

In this section, we propose an affine three-factor model for index options based on the Matrix Affine Jump Diffusion (MAJD) process introduced in Leippold and Trojani (2008). The model has several novel structural properties: it allows for dynamic interactions between risk factors and for mutually-exciting risks, as well as for unspanned skewness and dynamic volatility term structure effects. It has closed form solutions and allows for pricing via transform methods. It naturally nests many important affine models in the literature, such as Bates (2000) or Heston (1993), both of which have independent volatility components and a stochastic skewness that is linearly related to the volatility.

2.1.1 Return Dynamics and Volatility Process

Our model features a matrix-valued volatility dynamics and Poisson-Normal return jumps with a stochastic intensity. The volatility components are potentially correlated with returns. It follows that our model has two different channels for generating a stochastic return skewness: The standard leverage effect (i.e. a correlation between innovations in returns and volatility) and a time-varying probability of (mostly negative) return jumps. Contrary to independent-component models such as Bates (2000) or Christoffersen et al. (2009), the volatility factors in our model interact with each other and shocks to skewness are not completely spanned by shocks to diffusive volatility. As in the aforementioned models, the potentially different persistence of the volatility components can generate interesting term structure of volatility patterns.

Let S_t denote the value of an equity index at time t, r and q be the (constant) interest rate and dividend yield, and X_t a 2×2 matrix valued volatility factor. Under the risk neutral probability measure, the return dynamics is summarized in the following assumption.

Assumption 1. Under the risk neutral probability measure, the dynamics of S_t is given by:

$$\frac{dS_t}{S_t} = (r - q - \lambda_t \bar{k})dt + tr(\sqrt{X_t} dZ_t) + k dN_t \tag{2.1}$$

where the matrix-valued volatility dynamics X_t is specified in assumption 2, Z is a matrix Brownian motion defined in (2.6), and return jumps follow a Poisson-Normal process kdN_t featuring a time-varying jump intensity λ_t and an iid jump size k specified in (2.4) and (2.3)

The latent state X_t is a symmetric, positive definite 2×2 matrix

$$X_t = \begin{pmatrix} X_{11t} & X_{12t} \\ X_{12t} & X_{22t} \end{pmatrix} .$$

The dynamics of X_t is detailed in the next assumption.

Assumption 2. The symmetric 2×2 matrix process X_t follows the affine dynamics

$$dX_t = [\Omega\Omega' + MX_t + X_t M']dt + \sqrt{X_t} dB_t Q + Q' dB_t' \sqrt{X_t} \tag{2.2}$$

where Ω, M, Q are 2×2 parameter matrices and B_t is a 2×2 standard Brownian motion. $\sqrt{X_t}$ denotes the unique symmetric square root of X_t.

X_t is the Wishart process introduced by (Bru 1991). This process is a matrix-valued extension of the univariate square-root process that is widely used in the term structure and stochastic volatility literatures; see, e.g., Cox, Ingersoll and Ross (1985) and Heston (1993). Process (2.2) has many convenient properties that make it ideal to model multivariate financial risks. First, if $\Omega\Omega' \gg Q'Q$, then X_t is positive semi-definite. Under this condition, the diagonal elements X_{11t}, X_{22t} are well-defined non-negative volatility processes with potentially different persistence features. If $\Omega\Omega' \gg 3Q'Q$, then X_t is positive definite and no volatility component can reach the zero bound.[3] Second, when matrices Ω, Q and M are diagonal, then (X_{11t}, X_{22t}) defines two autonomous Markov processes with components distributed as independent Heston (1993)-type volatility models. Under these constraints, Bates' (2000) two factor state dynamics is nested by the dynamics of the diagonal elements in (2.2), see Appendix. A.1 for more details. More general, when matrices Ω, M or Q are not diagonal, the joint dynamics of X_{11t}, X_{12t} and X_{22t} features conditional interactions, both in the drift and diffusion parts of equation (2.2). Third, the process is affine and has a closed-form Laplace transform that implies convenient and tractable expressions for the prices of European options.

The specification of the jumps in returns is as follows: Jumps are independent of the volatility process with an iid jump size k distributed as:

$$\ln(1+k) \sim N(\ln(1+\bar{k}) - \frac{\delta^2}{2}\,,\, \delta^2) \tag{2.3}$$

The stochastic intensity λ_t is an affine function of the state X_t:

$$\lambda_t = \lambda_0 + tr(\Lambda X_t)\ , \tag{2.4}$$

where $\lambda_0 \geq 0$ and Λ is a 2×2 positive definite matrix and $tr(\cdot)$ is the trace operator.

Our process nests Bates (2000) insofar as the diffusive part of the return variance is the sum of two diagonal state components $X_{11t} + X_{22t}$. The total variance obtains as:

$$\begin{aligned} Var_t\left(\frac{dS_t}{S_t}\right) &= tr(X_t) + \lambda_t E(k^2) \\ &= tr\Big[X_t\left(Id_2 + \Lambda E(k^2)\right)\Big] + \lambda_0 E(k^2). \end{aligned} \tag{2.5}$$

In order to generate the feedback effect between returns and volatility, it is convenient to correlate shocks between returns and state dynamics (2.2). To preserve an affine structure, we specify as in da Fonseca et al. (2008) a 2×2 standard Brownian motion

---

[3] We will later on assume $\Omega\Omega' = \beta Q'Q$ with the scalar $\beta > 1$. This guarantees the existence of a closed-form solution without reducing the flexibility of the model.

as follows:

$$Z_t = B_t R + W_t \sqrt{I_2 - RR'} \ , \tag{2.6}$$

where W is another 2×2 standard Brownian motion, independent of B, and R is a 2×2 matrix such that $I_2 - RR'$ is positive semi-definite.

2.1.2 Unspanned Skewness

Our model follows the Bates (2000) option valuation setting by specifying the diagonal state components X_{11t}, X_{22t} as short and long run volatility components, allowing for rich volatility term structure effects. The sum $V_t^{diff} = tr(X_t)$ is the diffusive variance of returns. We extend the Bates (2000) setting by adding the state variable X_{12t}. Shocks to the diffusive volatility are not fully spanned by shocks in X_{12t}, and vice versa. By linking X_{12t} to a return component that models skewness, it is possible to introduce a stochastic skewness component that is unspanned by volatility shocks. In order to highlight the role of X_{12t}, we reconsider the two channels for generating a stochastic return skewness:

(i) For the leverage effect, equation (2.6) implies, after simple calculations:

$$\begin{aligned} Cov_t\left(\frac{dS_t}{S_t}, dtr(X_t)\right) &= tr(R'QX_t) \\ &= (R'Q)_{11}X_{11t} + (R'Q)_{22}X_{22t} + (R'Q)_{12}X_{12t} \end{aligned} \tag{2.7}$$

i.e. the feedback effect between return and diffusive volatility depends on X_{12t}, which itself is not part of the diffusive volatility.

(ii) For the jumps in returns, we first note that the jump size distribution (2.3) is usually negatively skewed. Furthermore, if we write the jump intensity (2.4) explicitly as

$$\lambda_t = \lambda_0 + tr(\Lambda X_t) = \lambda_0 + \Lambda_{11}X_{11t} + \Lambda_{22}X_{22t} + \Lambda_{12}X_{12t} \ ,$$

we note once again the role of X_{12t} in this skewness channel.

2.1.3 Model Classification and Nested Models

Model (2.1-2.2) belongs to the class of matrix affine jump diffusions (MAJD) introduced in (Leippold and Trojani 2008). The affine return and state dynamics implies closed form Laplace transforms and an efficient computation of plain vanilla option prices by transform methods; see also (Carr and Madan 1999) and (Duffie et al. 2000), among others. The model naturally nests a number of affine jump diffusions in the literature. Therefore, it provides a consistent framework for studying and comparing the performance of

r	q	$SV_{r,q}$	**Pure diffusive models**	$SVJ_{r,q}$	**Jump-diffusion models**
1	0	$SV_{1,0}$	Heston (1993)	$SVJ_{1,0}$	Bates (1996)
2	0	$SV_{2,0}$	Christoffersen et al. (2009)	$SVJ_{2,0}$	Bates (2000)
3	0	$SV_{3,0}$	This paper	$SVJ_{3,0}$	(not considered)
3	1	$SV_{3,1}$	da Fonseca et al. (2008)	$SVJ_{3,1}$	Leippold and Trojani (2008)

Table 2.1: Models considered in our study. r is the total number of state variables and q the number of unspanned stochastic skewness components.

these models in capturing the behavior of the index option implied volatility surface. Unspanned skewness features arise when either Λ or R are not diagonal. Similarly, dynamic interactions between volatility and skewness components can emerge when either M or Q are not diagonal. When matrices M, Q, Ω, R and Λ are all diagonal, model (2.1) collapses to a Bates (2000) model. If in addition the jump component is removed ($\lambda_0 = 0$ and $\Lambda = 0$), then we obtain the two factor Heston (1993)-type volatility model studied by Christoffersen et al. (2009). By construction, these diagonal models feature independent volatility components with

no dynamic interaction, together with a return skewness spanned exclusively by X_{11t} and X_{22t}. The pure diffusion model in da Fonseca et al. (2008) arises for $\lambda_0 = 0$ and $\Lambda = 0$ when either M, Q, Ω or R are not diagonal. This setting provides dynamic interactions and volatility feedbacks not spanned by volatility shocks. In addition, the full jump diffusion setting features an unspanned skewness component related to X_{12t} via the time varying jump probability.[4]

We can classify all models nested within our framework, in dependence of their total number r of state variables and their number q of stochastic skewness components unspanned by the volatility components. Pure diffusion models are denoted by SV and jump diffusion models by SVJ. In general, any model can be then classified as $SV_{r,q}$ or $SVJ_{r,q}$. Table 2.1 provides an overview of the models nested by our Assumptions 1 and 2.

Option Valuation

The model yields closed-form transform expressions for returns, which are useful in order to efficiently compute the prices of plain vanilla options by transform methods, as proposed by (Carr and Madan 1999) and (Duffie et al. 2000), among others. Closed form expressions for the risk neutral Laplace transform of returns are available when

---

[4] In Appendix A.1 we provide the parameter constraints under which different nested models obtain.

$\Omega\Omega' = \beta Q'Q$ for some $\beta > 1$.[5] In this case, Assumption 1 implies an exponentially affine conditional Laplace transform for $Y_T := \log(S_T)$, given by (see (Leippold and Trojani 2008)):

$$\Psi(\tau;\gamma) := E_t\left[\exp\left(\gamma Y_T\right)\right] = \exp\Big(\gamma Y_t + tr\big[A(\tau)X_t\big] + B(\tau)\Big) \quad , \tag{2.8}$$

where $\tau = T - t$, $A(\tau) = C_{22}(\tau)^{-1}C_{21}(\tau)$ and the 2×2 matrices $C_{ij}(\tau)$ are the $ij-$th blocks of the matrix exponential:

$$\left(\begin{array}{cc} C_{11}(\tau) & C_{12}(\tau) \\ C_{21}(\tau) & C_{22}(\tau) \end{array}\right) = \exp\left[\tau\left(\begin{array}{cc} M + \gamma Q'R & -2Q'Q \\ C_0(\gamma) & -(M' + \gamma R'Q) \end{array}\right)\right] . \tag{2.9}$$

The explicit expressions for the 2×2 matrix C_0 is:

$$C_0(\gamma) = \frac{\gamma(\gamma-1)}{2} I_2 + \Lambda\left[(1+\bar{k})^{\gamma}\exp\left(\gamma(\gamma-1)\frac{\delta^2}{2}\right) - 1 - \gamma\bar{k}\right] \tag{2.10}$$

and real-valued function $B(\tau)$ is given by:

$$\begin{aligned} B(\tau) \quad = \quad & \left\{r - q + \lambda_0\left[(1+\bar{k})^{\gamma}\exp\left(\gamma(\gamma-1)\frac{\delta^2}{2}\right) - 1 - \gamma\bar{k}\right]\right\}\tau \\ & -\frac{\beta}{2} tr[\log C_{22}(\tau) + \tau(M' + \gamma R'Q)] \end{aligned} \tag{2.11}$$

where $\log(\cdot)$ is the matrix logarithm. In contrast to diagonal Bates (2000)-type models, computation of the return transform in the full model cannot be reduced to calculations that involve only scalar exponential and logarithmic functions, because coefficients $C_{ij}(\tau)$ and $B(\tau)$ depend on a matrix exponential and logarithm, respectively. This feature makes the computation of Laplace transform (2.8) typically two orders of magnitude more costly than in diagonal models. We obtain an efficient computation of the pricing transform for the full model using the Cosine-Fourier Transform method of Fang and Oosterlee (2008). For a detailed description of the numerical issues in evaluating our model, see Gruber (2015).

2.2 Empirical Analysis

We estimate the models listed in Table 2.1, using about fourteen years of S&P 500 index option data, and study the added value of models with unspanned skewness components

---

[5] Precisely, in order to nest the diagonal Bates (2000) and two-factor Heston-type models when both Q and M are diagonal, we allow β to be a diagonal matrix K .

or dynamic volatility interactions in explaining the cross sectional behaviour of S&P 500 option implied volatility surfaces.

2.2.1 Data Source and Characteristics

We collect from OptionMetrics daily data of end-of-day prices of S&P 500 index options, traded at the Chicago Board Options Exchange, for the sample period January 1996 to September 2009 and maturities up to one year.[6] We then apply a number of standard filtering procedures outlined in Bakshi et al. (1997). First, we eliminate options with midquote premia below 0.375 dollars and options with zero bid price or with bid price larger than the ask price. Second, we eliminate options with stale quotes (i.e., prices identical to the prices of the previous trading day), options with prices that violate arbitrage bounds, options with duplicate entries and options where the bid-ask spread is smaller than the minimum tick size (i.e., five cents for options having prices below 3 dollars and ten cents for all other options). Third, we drop options with a time to maturity less than 10 days, in order to avoid pricing effects largely driven by short term liquidity features. Note that we do not apply additional filters that cut options with extreme moneyness, in order to obtain a data set as rich and challenging as possible, with respect to the empirical features of the term structure of implied volatility skews. On average, we obtain about 185 option prices per trading day, having an average time to maturity of 133.5 days and an average moneyness $S/K = 1.07$. The interest rate r is computed by linearly interpolating the US treasuries yield curve supplied by OptionMetrics. The dividend yield q is computed by minimizing, separately for each maturity τ and each day, the put-call parity error of near-the-money options ($0.9 \leq K/S \leq 1.1$): $q = \arg\min_{q} \left(C - P - Se^{-\tau q} + Ke^{-\tau r}\right)^2$, where K is the option strike price, C and P the prices of call and put options, S the underlying spot price and τ the time to maturity of the option.

We estimate all models by Non Linear Least Squares (NLLS) and obtain estimates of each model's risk neutral dynamics and latent state variables. Parameter estimation is based on monthly observations of S&P 500 index option prices in the sample period from January 2000 to December 2004 (in-sample data set). Each month, we select the Wednesday of the week before the expiry date. Thus, the shortest time to maturity in our estimation sample is fixed at ten days.[7] Focusing on the monthly sample for estimation purposes has several reasons. First, it reduces the computational costs implied by NLLS estimation of the models in Table 2.1, which can have up to three latent state components

---

[6] We obtain end-of-day midquotes as simple averages of end-of-day bid and ask call prices and force the put-call parity to hold when calculating the implied dividend yields.

[7] There is no observation for September 2001, as US exchanges were closed from September 11 to September 16, 2001.

Panel A: Summary statistics

Sample	**“monthly”**	**“full”**
Time frame	2000-2004	1996-09/2009
Sampling interval	monthly	daily
Trading days T	59	3460
Total number of observations	21’993	638’365
Average time to maturity	130 days	133.5 days
Average moneyness (S/K)	1.06	1.07
Average option price	\$107	\$115

Panel B: Number of contracts stratified by moneyness and maturity

	$\tau \leq 20$	$20 < \tau \leq 80$	$80 < \tau \leq 180$	$\tau > 180$	all
$S/K < 0.80$	82	2’113	5’791	16’933	24’919
$0.80 < S/K < 0.90$	956	16’426	20’016	29’928	67’326
$0.90 < S/K < 1.00$	14’562	78’895	37’364	37’065	167’886
$1.00 < S/K < 1.10$	19’199	77’512	33’340	32’301	162’352
$1.10 < S/K < 1.20$	5270	41’087	21’124	23’406	90’887
$1.20 < S/K < 1.30$	1139	18’084	14’568	16’525	50’316
$1.30 < S/K < 1.40$	390	8’258	9’565	10’943	29’156
$S/K > 1.40$	251	8’590	14’230	22’452	45’523
all	41’849	250’965	155’998	189’553	638’365

Table 2.2: Panel A: Summary statistics of the data. The “monthly” column refers to the data set used for parameter estimation and in-sample performance analysis. The “full” column refers to the data set used for out-of sample evaluation. Panel B: Number of contracts stratified by moneyness S/K and maturity in days.

and sixteen parameters. Second, it leaves a large fraction of our data available for out-of-sample model evaluation. This aspect is important in our context, in order to allow for a fair comparison of models in Table 2.1, which can feature different dimensions of both state dynamics and parameter space. The evaluation of out-of-sample performance also controls for possible overfitting that tends to favour higher dimensional models, when purely assessing in-sample pricing performance. Out-of-sample evaluation is performed by (i) fixing the parameter estimates at the values estimated using the monthly data set and (ii) estimating by NLLS only the missing latent state for each daily observation from January 1996 to September 2009. Note that for our out-of-sample analysis we keep the model parameter values fixed at one single set of parameters, implying that the out-of-sample pricing performance is completely driven by the ability of the specified state space to account for all variations in out-of-sample option prices, both cross-sectionally and in the time series. Summary statistics of our data sets are reported in Table 2.2.

Panel A presents basic aggregate statistics, while Panel B describes the structure of our data set along the maturity and moneyness dimensions.

Overall, our sample consists of 638'365 contracts with average time to maturity of about 133.5 days and moneyness ranges between $S/K = 0.7$ and $S/K = 1.5$. The most frequently traded options have maturities between 20 and 80 days and a moneyness $0.9 < S/K < 1.1$.

2.2.2 Estimation Method

The main challenge when estimating the stochastic volatility models in Table 2.1 it to estimate the model structural parameters together with the time series of latent states X_t ($t = 1 \ldots T$). Several approaches are available in the literature and have been applied to a variety of option pricing models with independent volatility components. A popular approach in single-factor models it to treat the spot volatility as an additional parameter that has to be re-estimated with a recursive procedure; see (Bakshi et al. 1997), among others. Other approaches filter the volatility states using time series information on underlying returns, thus ensuring consistency of physical and risk-neutral probabilities. This is achieved for a number of single-factor stochastic volatility models in (Jones 2003) and (Eraker 2004), using Monte Carlo Markov Chain methods, in (Chernov and Ghysels 2000), who apply a version of the Efficient Method of Moments, in (Pan 2002), who introduces filtered state Generalized Method of Moments estimation, and in (Christoffersen, Jacobs and Mimouni 2010) and (Johannes, Polson and Stroud 2010), who make use of particle filtering techniques, in order to better account for model non-linearities. Multi-factor volatility models with independent components have been studied in (Carr and Wu 2007), who apply the standard Kalman filter to estimate a two-factor model with high-frequency jumps, and in (Carr and Wu 2009b), who estimate with the unscented Kalman filter a three-factor volatility model with self-exciting volatility.

We choose a modification of the NLLS approach taken in Bates (2000), (Huang and Wu 2004) and Christoffersen et al. (2009), among others, to infer model parameter values and state realizations. This allows us to test whether the inferred risk neutral distributions of models in Tab. 2.1 are consistent with the observed cross-sectional and time-series behaviour of S&P 500 index option prices. We maximize a Gaussian pseudo likelihood function for observed pricing errors, while allowing for conditional error heteroskedasticity driven by group specific and idiosyncratic shocks.

Let $\theta = \{M, R, Q, \beta, \lambda_0, \Lambda, \bar{k}, \delta\}$ be the parameter vector of interest[8] and $e_{i,t}(\theta) =$

---

[8] More precisely, we vectorize matrices M, R, Q and Λ. See section A.3.2 in the appendix for details on parameter encoding and identification.

$\big(\widehat{C}_i(\theta, X_t^*(\theta)) - C_{it}\big)/F_t$ be the relative pricing error, where C_{it} and $\hat{C}_{it}$ are observed and model-implied option prices of option i at time t, respectively, and F_t is the S&P 500 index future of corresponding maturity at time t. The conditional implied state $X_t^*(\theta)$ is defined by:

$$X_t^*(\theta) = \arg\min_{X_t} \sum_{i=1}^{N_t} \left[\left(\widehat{C}_i(\theta, X_t) - C_{it} \right) / F_t \right]^2 \; . \tag{2.12}$$

For any given $t = 1, \ldots, T$, vector $\mathbf{e}_t(\theta) = (e_{1,t}(\theta), \ldots, e_{N_t,t}(\theta))'$ denotes the vector of pricing errors at time t. Our point estimate for parameter θ is given by the following pseudo Maximum Likelihood estimator:[9]

$$\widehat{\theta} = \arg\max_{\theta} \mathcal{L}_T(\theta) := \arg\max_{\theta} \; -\frac{1}{2} \sum_{t=1}^{T} \Big(\ln|\Omega_t| + \mathbf{e}_t'(\theta)\, \Omega_t^{-1}\, \mathbf{e}_t(\theta) \Big) \; . \tag{2.13}$$

where the $N_t \times N_t$ matrix Ω_t is the conditional covariance matrix of these errors. We obtain the elements of Ω_t from calculating group-specific error covariances in the three by three grid with maturity groups $\tau \leq 2; 2 < \tau \leq 6$ and $\tau > 6$ months and moneyness groups $K/S \leq 0.9; 0.9 < K/S \leq 1.1$ and $K/S > 1.1$, similar to Bates (2000). We do not perform an additional vega-weighting of the pricing errors as options with extreme moneyness values usually have larger pricing errors and receive a small weight from Ω_t^{-1}.

We solve the problem of jointly estimating the model implied states $\{X_t^*(\theta)\}_{t=1,\ldots,T}$ and parameter θ using a full nested optimization. In the first step of the optimization, we compute for any candidate parameter vector θ optimal state $\{X_t^*(\theta)\}_{t=1,\ldots,T}$. In the second step, we maximize the pseudo likelihood criterion $\mathcal{L}_T(\theta)$ over θ. We have also investigated a less computationally demanding two-step optimization approach, used in (Huang and Wu 2004) and Christoffersen et al. (2009), among others, which iterates between parameter estimation for a given state and state estimation for a fixed parameter. We find that this method does not produce good convergence properties and stable estimation results.

2.2.3 Parameter Estimates and In-Sample Results

In this section, we first present parameter estimates and in-sample pricing results for the two- and three-factor models in the context of Assumption 1. In a second step, we study the distinct role of model parameters and latent states in generating implied-volatility skew and term structures effects largely unrelated to the level of the volatility.

---

[9] Under the given conditions, pseudo Maximum Likelihood and NLLS estimators coincide.

Panel A: Estimated diagonal diffusion parameters

	M_{11}	M_{22}	Q_{11}	Q_{22}	R_{11}	R_{22}	β	β_2
$SV_{3,1}$	-0.3426	-4.4856	0.0136	0.4116	-0.4878	-0.6279	1.2039	
	(0.0120)	(0.0149)	(0.0114)	(0.0041)	(0.5340)	(0.01781)	(0.0339)	
$SV_{2,0}$	-0.0155	-5.9240	0.1583	0.5671	-0.7849	-0.6835	0.0934	0.8997
	(0.0154)	(0.0364)	(0.0040)	(0.0019)	(0.0116)	(0.0124)	(0.0312)	(0.0101)
$SVJ_{3,1}$	-0.0793	-1.9428	0.0001	0.1690	-0.7912	-0.9854	1.0368	
	(0.0166)	(0.0432)	(0.0030)	(0.0011)	(1.5804)	(0.0829)	(0.1848)	
$SVJ_{2,0}$	-0.3055	-1.5245	0.0757	0.3468	-0.8517	-0.7367	3.2122	0.0007
	(0.0159)	(0.312)	(0.0028)	(0.0221)	(0.0356)	(0.0191)	(0.0844)	(0.0008)

Panel B: Out-of-diagonal diffusion parameters

	M_{21}	Q_{12}	R_{21}
$SV_{3,1}$	8.8713	-0.0137	-0.5900
	(0.0506)	(0.3229)	(0.3747)
$SVJ_{3,1}$	2.1498	0.0366	-0.1039
	(0.0491)	(0.0272)	(0.4037)

Panel C: Estimated parameters of the jump component

	λ_0	Λ_{11}	Λ_{12}	Λ_{22}	$\bar{k}$	δ
$SVJ_{3,1}$	0.0023	51.77	-55.09	55.04	-0.0699	0.0993
	(0.0587)	(1.18)	(3.62)	(1.35)	(0.0028)	(0.0010)
$SVJ_{2,0}$	0.2013	84.56	–	0.94	-0.0436	0.0954
	(0.0455)	(8.18)		(2.36)	(0.0024)	(0.0038)

Table 2.3: Point estimates and corresponding standard errors for parameters of different models. Panel A: point estimates and standard errors, in parentheses, for the diagonal components of the diffusion parameter matrices. Panel B: point estimates and standard errors, in parentheses, for the out-of-diagonal components of the diffusion parameter matrices. Panel C: point estimates and standard errors, in parentheses, for the parameters in the jump component.

Estimated Risk Neutral Dynamics

Estimated risk neutral parameters for two- and three-factor models in Table 2.1 are presented in Table 2.3.[10] NLLS estimation procedure (2.13) is applied to the in-sample data, consisting of monthly observations of S&P 500 index options from 2000 to 2004.

Panel A of Table 2.3 presents the point estimates for the diffusive parameters of diagonal models $SV_{2,0}$ and $SVJ_{2,0}$ and the diagonal elements of matrices M, Q and R of models $SV_{3,1}$ and $SVJ_{3,1}$. Estimates for the out-of-diagonal elements are presented in Panel B, while panel C summarizes estimation results for the parameters in the jump components of models $SVJ_{2,0}$ and $SVJ_{3,1}$.

Both diagonal models $SV_{2,0}$ and $SJV_{2,0}$ feature two volatility factors with very dif-

---

[10] For brevity, we omit results for single-factor models. They are available on request. For a simple comparison of results across the nested models, we present parameter estimates using the notation of Assumption 1. The Appendix provides the link to Bates (2000) notation in the context of diagonal model $SVJ_{2,0}$.

ferent mean reversions and volatilities of volatility, and slightly different leverage effect parameters: In both models, the less persistent volatility component v_{2t} ($M_{22} < M_{11}$) also features a higher volatility of volatility ($Q_{22} > Q_{11}$) and a stronger leverage effect, since $|R_{11}Q_{11}| > |R_{22}Q_{22}|$, similar to findings in previous studies; see Bates (2000) and Christoffersen et al. (2009), among others. Despite the different data sets, the estimated jump component for model $SVJ_{2,0}$ in our study is consistent with the results in Bates (2000). We obtain an average jump size $\overline{k} = -0.043$, a jump size volatility $\delta = 0.0954$ and the following sensitivities of λ_t to the volatility components: $\Lambda_{11} = 84.56$ and $\Lambda_{22} = 0.94$.[11] The non significant point estimate for Λ_{22} indicates that, in the context of diagonal model $SVJ_{2,0}$, linear specifications of time varying intensities based on a single volatility component are not rejected by the data.

The point estimates of the out-of-diagonal elements of matrices M, Q and Λ for the models $SV_{3,1}$ and $SVJ_{3,1}$ in Panels B and C, respectively, provide a number of interesting results. First, we can reject the null hypothesis of a diagonal mean reversion matrix M, since point estimate M_{21} is statistically highly significant in both the $SV_{3,1}$ and $SVJ_{3,1}$ models. This finding is a first indication that volatility components with dynamic interactions are well supported by the data. Second, all point estimates for the components of matrix Λ in model $SVJ_{3,1}$ are similar in absolute value and highly statistically significant: While volatility components X_{11t}, X_{22t} load positively on λ_t, the unspanned skewness component X_{12t} loads negatively. Finally, we test the hypothesis that the leverage effect in models $SV_{3,1}$ and $SVJ_{3,1}$ is not driven by the unspanned component X_{12t}, i.e. that $\frac{1}{2dt}Cov_t(dS_t/S_t, dtr(X_t)) = (R'Q)_{11}X_{11t} + (R'Q)_{22}X_{22t}$. According to the leverage effect expression (2.7), this is equivalent to a test of null hypothesis $\mathcal{H}_0$ against alternative hypothesis $\mathcal{H}_A$:

$$\mathcal{H}_0 \; : \; (R'Q)_{21} = 0 \quad ; \quad \mathcal{H}_A \; : \; (R'Q)_{21} \neq 0 \; , \tag{2.14}$$

where $(R'Q)_{21} = R_{12}Q_{11} + R_{22}Q_{12}$. Using a standard Wald test, we reject $\mathcal{H}_0$ in favor of $\mathcal{H}_A$ with a $p-$value below 0.005. In summary, the estimation results in Table 2.3 support a $SVJ_{3,1}$ model specification with non diagonal matrices M, R and Λ, i.e., a MAJD option valuation model with dynamically interacting volatility factors and unspanned skewness features.

In-Sample Pricing Results

The in-sample pricing results for the different models are presented in Table 2.4, Panel A. Consistently with our NNLS estimation criterion (2.13), we rank pricing performance across models according to the root mean square dollar pricing error ($RMSE$).The $RMSE$

---

[11] For comparison, Bates (2000) estimates are $\overline{k} = -0.057$, $\delta = 0.102$, $\Lambda_{11} = 81.56$ and $\Lambda_{22} = 0.28$.

Panel A: Pricing performance for the “monthly” sample (2000-2004)

	$SV_{2,0}$	$SV_{3,0}$	$SV_{3,1}$	$SVJ_{2,0}$	$SVJ_{3,1}$
State space dimension	2	3	3	2	3
RMSE	1.180	1.127	1.048	1.115	0.913
σ_{RMSE}	(0.370)	(0.348)	(0.285)	(0.446)	(0.324)
Within bid-ask spread	0.603	0.617	0.640	0.635	0.633

Panel B: Pricing performance for the “full” sample (1996-09/2008)

	$SV_{2,0}$	$SV_{3,0}$	$SV_{3,1}$	$SVJ_{2,0}$	$SVJ_{3,1}$
State space dimension	2	3	3	2	3
RMSE	1.937	1.844	1.570	1.862	1.457
σ_{RMSE}	(1.101)	(1.027)	(0.808)	(1.129)	(0.809)
RMSIVE	2.060	1.974	1.824	1.935	1.772
σ_{RMSIVE}	(0.754)	(0.660)	(0.537)	(0.700)	(0.519)
Within bid-ask spread	0.437	0.461	0.540	0.452	0.527

Table 2.4: Performance comparison. *RMSE* is the sample average of the daily root-mean-squared dollar pricing errors. σ_{RMSE} is the sample standard deviation of daily *RMSE*. *RMSIVE* is the sample average of the daily root-mean-squared implied volatility error. σ_{RMSIVE} is the standard deviation of daily *RMSIVE*. The row “Within bid-ask spread” reports the fraction of fitted prices within the bid-ask spread.

for day $t = 1, \ldots, T$, denoted by ϵ_t, is defined as:

$$\epsilon_t = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}(\hat{C}_{t,i} - C_{t,i})^2} \quad ; \quad t = 1 \ldots T \quad , \tag{2.15}$$

where $\hat{C}_{t,i}$ and $C_{t,i}$ are the model-implied and the observed option prices, respectively, of option $i = 1, \ldots, N_t$ on day t. The sample *RMSE* is simply the time series average of daily root mean square errors:

$$RMSE = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t = \frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}(\hat{C}_{t,i} - C_{t,i})^2} \quad . \tag{2.16}$$

As a measure of overall model reliability, we also compute the standard deviation of daily root mean square errors: $\sigma_{RMSE} := \sqrt{\sum_{t=1}^{T}(\epsilon_t - RMSE)^2}$. The implied volatility root mean square error *IVRMSE* and σ_{IVRMSE} are defined in the same way.

Overall, Panel A of Table 2.4 confirms the superiority of models with jumps in returns. Adding jumps in returns reduced the *RMSE* by between 6% and 14%. Moreover, our results show that models with dynamic interactions and unspanned skewness substantially reduce the *RMSE* of diagonal models. Model $SV_{3,1}$ lowers the *RMSE* of model $SV_{2,0}$ by approximately 12%, while model $SVJ_{3,1}$ lowers the *RMSE* of model $SVJ_{2,0}$ by approximately 19%.

It is important to note that the *RMSE* reduction of models $SV_{3,1}$ and $SVJ_{3,1}$ is not simply due to their higher dimensional state space relative to the $SV_{2,0}$ and $SVJ_{2,0}$ models (with three latent components instead of two), but rather to the particular features of their state dynamics. For the pure diffusive case, we study the $SV_{2,0}$-model of Christoffersen et al. (2009) as a reference model and compare it to our $SV_{3,1}$-model and a three-factor Heston $SV_{3,0}$ model. This allows us to disentangle the effects of adding a third factor from the choice of state space (i.e., independent or interacting factors). While the $SV_{3,0}$ model is only 4.5% better than $SV_{2,0}$, our model $SV_{3,1}$, which has the same number of factors, but a matrix-valued state space, improves the fit by 12% over $SV_{2,0}$.

Models with dynamic interactions and unspanned skewness components also reduce the in-sample variability of the pricing errors. The *RMSE* standard deviation of model $SV_{3,1}$ is 23% (18%) lower than the one in model $SV_{2,0}$ ($SV_{3,0}$), while the RMSE standard deviation of model $SVJ_{3,1}$ is 28% lower than the one in model $SVJ_{2,0}$. This finding shows that the matrix state space of $SV_{3,1}$ and $SVJ_{3,1}$ models produces more reliable results also in terms of a less volatile in-sample pricing performance.

2.2.4 Out of Sample Results and Estimated State Dynamics

We fix the vector of estimated model parameters $\hat{\theta}$ (see Table 2.3) and compute for every day the latent state $X_t^*(\hat{\theta})$ in equation (2.12). In this way, we assess the out-of-sample pricing performance of the models in Table 2.1, by relying exclusively on the ability of each model's state space to reproduce the cross-sectional and time-series patterns of the implied volatility smile of index options. This out-of-sample implementation is similar to (Huang and Wu 2004) and Christoffersen et al. (2009).

Out-of-Sample Pricing Performance

Panel B of Table 2.4 summarizes aggregate out-of-sample pricing results. Overall, $SV_{3,1}$ and $SVJ_{3,1}$ models clearly outperform diagonal models: The *RMSE* of model $SV_{3,1}$ ($SVJ_{3,1}$) is about 19% (22%) lower than the $RMSE$ of model $SV_{2,0}$ ($SVJ_{2,0}$). Similarly, the $RMSE$ of model $SVJ_{3,1}$ is about 21% lower than the $RMSE$ of model $SV_{3,0}$. These out-of-sample $RMSE$ reductions are on average larger than the in-sample $RMSE$ reductions in Panel A of Table 2.4. The $SV_{3,1}$ and $SVJ_{3,1}$ models also imply a higher pricing reliability, relative to the $SV_{2,0}$ and $SVJ_{2,0}$ benchmarks, with out-of-sample RMSE standard deviations that are 27% and 28% lower, respectively.

The higher performance of $SV_{3,1}$ and $SVJ_{3,1}$ models relative to benchmark models is

Label	**Date**	**Event description**
(1)	1997-07-02	Devaluation of Thai Bhat (begin of Asian Crisis)
(2)	1998-08-17	Begin of Russian Crisis
(3)	2000-03-10	NASDAQ maximum
(4)	2008-05-30	Bear Sterns bailout
(5)	2008-09-15	Lehman bankruptcy
(6)	2011-08-05	US downgrade and EU debt crisis

Table 2.5: Description of crisis events indicated in the time-series plots with (1) to (6).

quite consistent over time. Figure 2.3, Panel A (Panel B) plots the time series of daily RMSE for model $SV_{2,0}$ ($SVJ_{2,0}$), together with the percentage reduction in daily RMSE of $SV_{3,1}$ and $SV_{3,0}$ ($SVJ_{3,1}$) models.

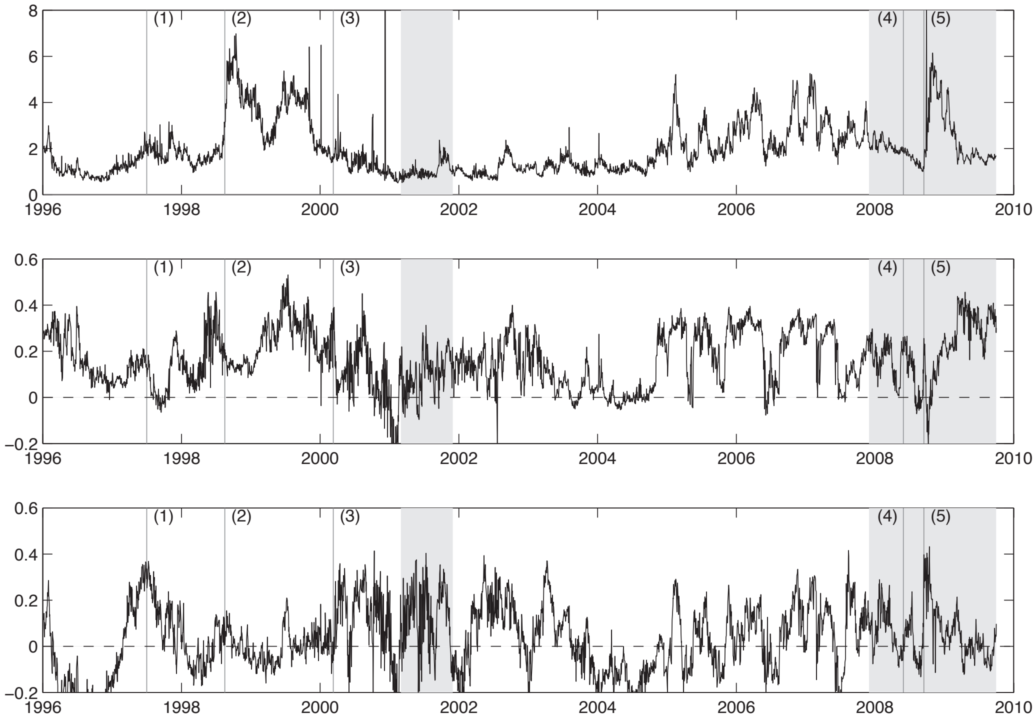
The middle plot of Panel A shows that model $SV_{3,1}$ almost always outperforms model $SV_{2,0}$, with a few rare exceptions at the beginning of 2001 and the end of 2008. Relative performance improvements can be large: While their average is about 20%, they often exceed 30%. Large improvements of pricing performance can arise during some periods of financial crises or market distress, including the Russian debt crisis, the collapse of LTCM, the bursting of the dot-com bubble and the recent Subprime Crisis. These extreme market events are denoted with (1) to (5) in the time series plots and listed in Table 2.5.

In contrast to the results for the $SV_{3,1}$ model, pricing improvements implied by the $SV_{3,0}$ model in the bottom plot of Panel A are often negative, rarely above 30% and substantially more volatile, which is a potential indication of model overfitting. The bottom plot of Panel B shows that model $SVJ_{3,1}$ virtually always outperforms model $SVJ_{2,0}$: There is no distinct period in which the $SVJ_{3,1}$ model systematically behaves worse than the $SVJ_{2,0}$ model. Outperformance is often large and the reduction in daily $RMSE$ can in some cases exceed 50%.

The better pricing performance of $SV_{3,1}$ and $SVJ_{3,1}$ models relative to benchmark models is consistent across moneyness regions and times to maturity. Panels A1 and B1 of Table 2.6 present out-of-sample $RMSE$ of $SV_{2,0}$ and $SVJ_{2,0}$ models, respectively, stratified by moneyness and maturity. Panels A2 and B2 summarize the improvements of the models $SV_{3,1}$ and $SVJ_{3,1}$ relative to the benchmark models.

Model $SV_{3,1}$ outperforms model $SV_{2,0}$ for all options with maturity above 20 days. For options with time to maturity of up to 20 days, we find a large outperformance with respect to out-of-the-money put options and a moderate underperformance for out-of-the-money call options, indicating that for short-maturity options the model faces a pricing tradeoff between calls and puts. Model $SVJ_{3,1}$ outperforms model $SVJ_{2,0}$ for all listed option classes. In particular, the stratification of pricing error improvements across moneyness in Panel B2 shows that model $SVJ_{3,1}$ captures much better the skew patterns

Panel A: Pure diffusion models



Panel B: Jump diffusion models

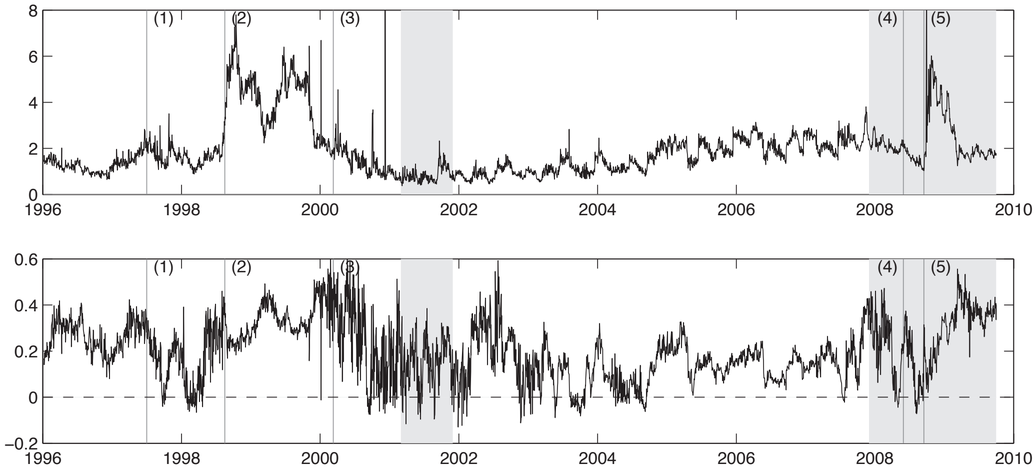


Figure 2.3: Time series of daily $RMSE$ and daily $RMSE$ improvements. Panel A compares pure diffusive models. Top graph: $RMSE$ of the benchmark $SV_{2,0}$ model. Middle graph: relative improvement of the $SV_{3,1}$ model over $SV_{2,0}$. Bottom graph: relative $RMSE$ improvement of $SV_{3,0}$ over $SV_{2,0}$. Panel B compares jump diffusion models. Top graph: $RMSE$ of the benchmark $SVJ_{2,0}$ model. Bottom graph: relative improvement of the $SVJ_{3,1}$ model over $SVJ_{2,0}$. Grey areas depict NBER recessions; crisis events indicated as (1) to (5) are listed in Table 2.5.

Panel A1: ***RMSE*** **for $SV_{2,0}$ model**

	$\tau < 20$	$20 < \tau < 80$	$80 < \tau < 180$	$\tau > 180$	all
$S/K < 0.80$	0.549	2.517	2.521	2.401	2.436
$0.80 < S/K < 0.90$	1.514	2.367	2.269	2.462	2.371
$0.90 < S/K < 1.00$	1.766	1.898	1.901	2.654	2.079
$1.00 < S/K < 1.10$	1.193	1.420	1.959	2.859	1.883
$1.10 < S/K < 1.20$	1.452	1.984	2.721	3.190	2.499
$1.20 < S/K < 1.30$	2.117	2.080	2.667	3.507	2.785
$1.30 < S/K < 1.40$	2.128	1.952	2.550	3.701	2.901
$S/K > 1.40$	1.550	1.750	2.428	5.019	3.854
all	1.495	1.834	2.277	3.216	2.410

Panel A2: ***RMSE*** **improvement of model $SV_{3,1}$ over model $SV_{2,0}$**

	$\tau < 20$	$20 < \tau < 80$	$80 < \tau < 180$	$\tau > 180$	all
$S/K < 0.80$	0.229	0.184	0.080	0.040	0.062
$0.80 < S/K < 0.90$	0.482	0.281	0.204	0.026	0.131
$0.90 < S/K < 1.00$	0.360	0.456	0.333	0.305	0.369
$1.00 < S/K < 1.10$	-0.002	0.171	0.395	0.342	0.283
$1.10 < S/K < 1.20$	-0.130	0.121	0.340	0.232	0.217
$1.20 < S/K < 1.30$	-0.071	0.084	0.286	0.188	0.186
$1.30 < S/K < 1.40$	-0.033	0.040	0.186	0.124	0.127
$S/K > 1.40$	-0.010	0.003	0.029	0.026	0.025
all	0.140	0.235	0.259	0.146	0.189

Panel B1: ***RMSE*** **of model $SVJ_{2,0}$**

	$\tau < 20$	$20 < \tau < 80$	$80 < \tau < 180$	$\tau > 180$	all
$S/K < 0.80$	0.439	2.402	3.020	2.437	2.578
$0.80 < S/K < 0.90$	0.997	2.376	2.695	2.225	2.398
$0.90 < S/K < 1.00$	1.492	1.909	2.154	1.643	1.880
$1.00 < S/K < 1.10$	1.346	1.348	1.483	2.843	1.772
$1.10 < S/K < 1.20$	1.625	1.645	2.303	3.856	2.537
$1.20 < S/K < 1.30$	2.224	1.850	2.381	4.144	2.937
$1.30 < S/K < 1.40$	2.180	1.855	2.353	4.073	3.013
$S/K > 1.40$	1.557	1.733	2.319	4.956	3.791
all	1.468	1.741	2.220	3.227	2.373

Panel B2: ***RMSE*** **improvement of model $SVJ_{3,1}$ over model $SVJ_{2,0}$**

	$\tau < 20$	$20 < \tau < 80$	$80 < \tau < 180$	$\tau > 180$	all
$S/K < 0.80$	0.021	0.322	0.291	0.384	0.348
$0.80 < S/K < 0.90$	0.169	0.335	0.328	0.372	0.346
$0.90 < S/K < 1.00$	0.016	0.271	0.369	0.367	0.297
$1.00 < S/K < 1.10$	0.056	0.190	0.246	0.414	0.293
$1.10 < S/K < 1.20$	0.120	0.112	0.276	0.368	0.289
$1.20 < S/K < 1.30$	0.101	0.114	0.231	0.296	0.252
$1.30 < S/K < 1.40$	0.060	0.098	0.153	0.191	0.172
$S/K > 1.4$	0.026	0.055	0.020	0.014	0.016
all	0.055	0.210	0.259	0.233	0.229

Table 2.6: $RMSE$ and $RMSE$ improvements over benchmark models stratified by maturity and moneyness. For diffusion and jump diffusion models, we present out-of-sample RMSE of benchmark models and percentage out-of-sample RMSE improvements of $SV_{3,1}$ and $SVJ_{3,1}$ models, stratified by moneyness and maturity in days. All performance computations are based on the "full" sample (1996-01/2009-09).

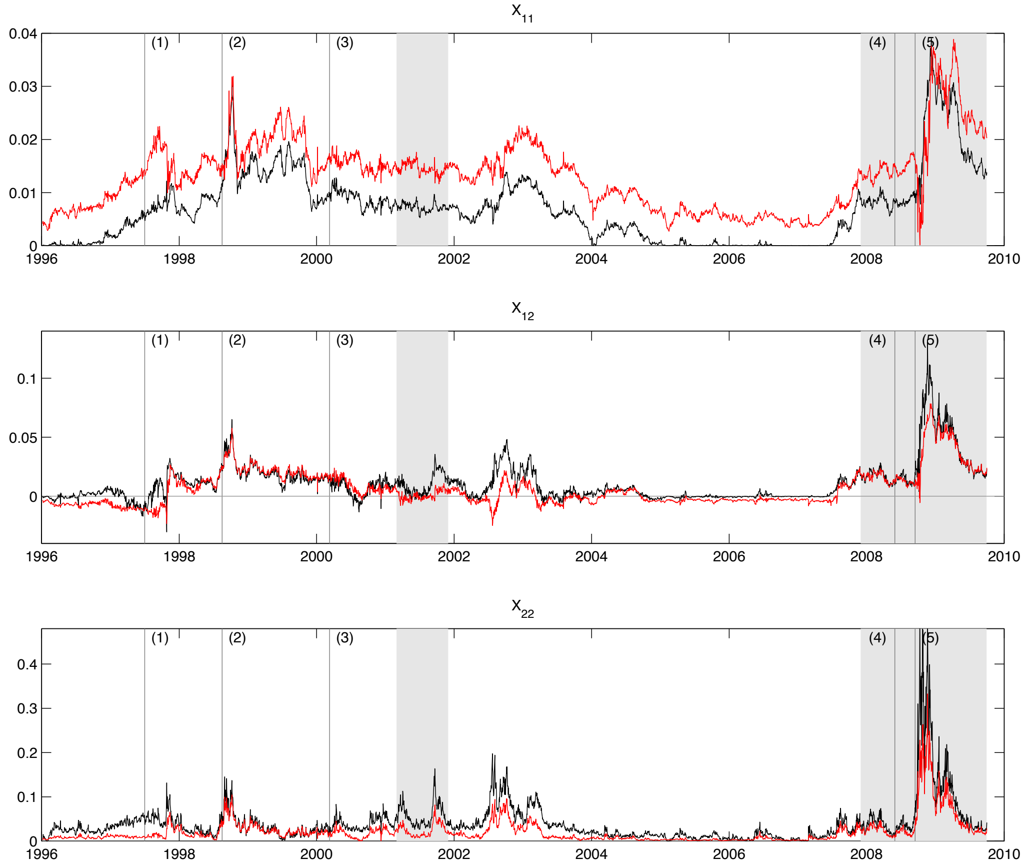


Figure 2.4: Estimated time series of implied states $\left(X_{11t}, X_{12t}, X_{22t}\right)$ in models $SV_{3,1}$ and $SVJ_{3,1}$. Black (red) lines correspond to states estimated for $SV_{3,1}$ pure diffusion ($SVJ_{3,1}$ jump diffusion) model. Grey areas in each plot depict NBER recessions; crisis events, indicated as (1) to (5), are listed in Table 2.5.

across maturities, with monotonically decreasing pricing improvements from option of moneyness $S/K < 0.80$ (average improvements of 35%) to options of moneyness $1.30 < S/K < 1.4$ (average improvements of 17%). In summary, these findings indicate that the better pricing performance of $SV_{3,1}$ and $SVJ_{3,1}$ models is unlikely due to overfitting, but rather the consequence of a state space specification that is better able to reproduce the structural dynamics of S&P 500 index option smiles.

Features of Latent State Dynamics

A key feature of models $SV_{3,1}$ and $SVJ_{3,1}$, relative to benchmarks $SV_{2,0}$ and $SVJ_{2,0}$, is the form of their (matrix) state dynamics, which allows us to model dynamic volatility interactions and unspanned skewness features. In Figure 2.4, we take a closer look at

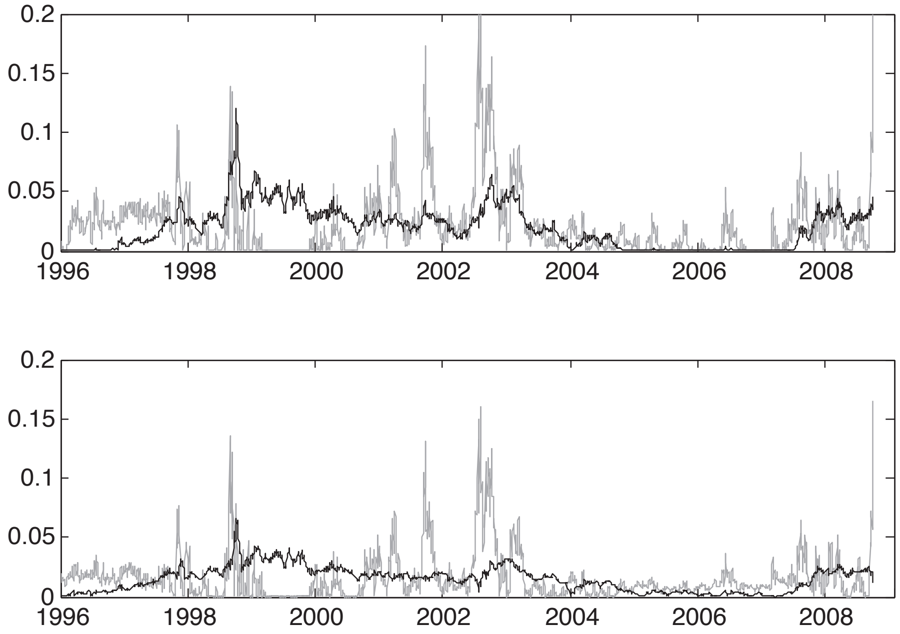


Figure 2.5: Implied state for the 2-factor reference models. Top panel: $SV_{2,0}$. Bottom panel: $SVJ_{2,0}$. The black line depicts the first (slowly mean reverting) and the grey one the second (fast mean reverting) factor.

the latent states X_{11t}, X_{22t} and X_{12t}, estimated for models $SV_{3,1}$ and $SVJ_{3,1}$ (plotted in black and red, respectively).

Remarkably, the time series for the $SV_{3,1}$ (black) and $SVJ_{3,1}$ (red) models are quite similar and show little noise, although the state has been obtained by computing (2.12) separately for each of the 3208 trading days and without penalizing large innovations. As a comparison, Figure 2.5 depicts the state for the two factor reference models. Especially for the model $SV_{2,0}$, the long-term volatility factor vanishes during the conundrum period.

Overall, we estimate a volatility component X_{22} that is on average larger and less persistent than component X_{11}. For instance, in the $SV_{3,1}$ model, 85% of the average instantaneous variance is generated by X_{22t}. At the same time, X_{22t} has an unconditional half life of about 16 days, which is about one tenth the half life of the unconditionally more persistent component X_{11t}. In the $SVJ_{3,1}$ model, component X_{11t} (X_{22t}) is on average responsible for about 39% (24%) of the total variance $V_t = X_{11t}+X_{22t}+\lambda_t E(k^2)$. As a consequence, jump-driven volatility tends to generate on average a quite substantial fraction (37%) of total variance. While the average total variance in the $SV_{3,1}$ model is dominated by a large not very persistent factor, in $SVJ_{3,1}$ model total risk is more evenly distributed across different potential sources of uncertainty. These distinct variance

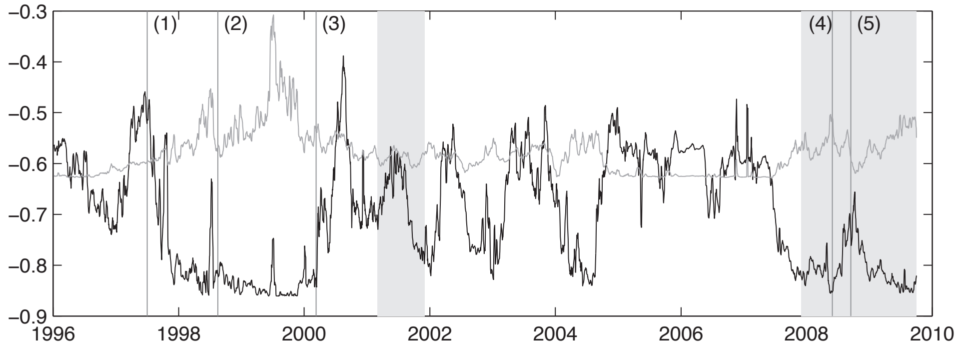


Figure 2.6: Unspanned stochastic skewness effects in the $SV_{3,1}$ model. We plot the implied time series of volatility feedback effects $corr_t(dS_t/S_t, d(v_{1t} + v_{2t}))$ for the unconstrained model (black line) and under the additional constraint that $X_{12t} = 0$ (grey line). Grey areas in each plot depict NBER recessions in our sample period; important crisis events, indicated as (1) to (5) in each plot, are listed in Table 2.5.

components are associated with a broader variety of persistence features: While λ_t has a short unconditional half life of about 18 days, which is comparable to the half life of factor X_{11t} in $SV_{3,1}$ model, X_{11t} and X_{22t} have average half lives of about 24 and 78 days, respectively.

Overall, these findings highlight a more pronounced multi-frequency volatility structure in model $SVJ_{3,1}$, which is potentially useful in order to generate more flexible term structures of implied volatility skews from very short to longer maturities.

The estimated state X_{12t} of the $SV_{3,1}$ and $SVJ_{3,1}$ models in the middle panel of Figure 2.4 highlights additional useful features of our model. First, X_{12t} can be both positive or negative: It tends to be positive during phases of high volatility, but it can turn sightly negative in other periods. In our models, X_{12t} can play qualitatively different roles, because it potentially jointly determines (i) the local persistence of volatility factors X_{11t}, X_{22t}, (ii) the feedback effects of returns and volatility and (iii) the time varying features of jump intensity λ_t in the $SVJ_{3,1}$ model. We find that, at the models' estimated parameters, the largest quantitative implications of the out-of-diagonal component X_{12t} arise for the dynamics of risk neutral skewness. Within the $SV_{3,1}$ model, state X_{12t} produces a substantial additional degree of variability of volatility feedback effects, which is useful to better capture time varying skewness patterns within these models.

To understand this point more concretely, Figure 2.6 plots the volatility feedback coefficient $corr_t(dS_t/S_t, d(X_{11t} + X_{22t}))$ implied by model $SV_{3,1}$, both for the case where the latent state is fictitiously restricted to be diagonal ($X_{12t} = 0$, grey line) and in the unconstrained case ($X_{12t} \neq 0$, black line). The large additional variability of volatility feedback effects generated by component X_{12t} quantifies the impact of unspanned skewness factors in generating (volatility-unrelated) stochastic skewness in model $SV_{3,1}$. For the $SVJ_{3,1}$ model, the impact of X_{12} is split into a volatility feedback channel and a jump intensity channel, resulting in a similar effect.

2.3 Model Analysis

We develop a more formal analysis of the $SV_{3,1}$ and $SVJ_{3,1}$ models, in order to better isolate the distinct effects of the components of the matrix state X_t on the shape of the implied-volatility smile. To achieve this goal, we introduce a state reparametrization in terms of (i) a volatility level factor $V_t := tr(X_t)$, (ii) a bounded volatility composition factor ξ_t and (iii) a further bounded factor α_t. Using this decomposition, we achieve several purposes. First, we can study the volatility-level unrelated tradeoff between option implied volatility skew and term structure ($\mathcal{S}$ and $\mathcal{M}$) in our models. Second, we can isolate the contribution of each factor V_t, ξ_t, α_t to the different pieces of the implied volatility smile. Third, we show that with the new parametrization, the $SV_{3,1}$ and $SVJ_{3,1}$ models can be reinterpreted as two-factor volatility models with stochastic coefficients, which depend exclusively on the third state variable α_t. Therefore, α_t is a useful factor in order to isolate (i) new dynamic volatility interactions and (ii) additional unspanned skewness effects produced by our setting, relative to, e.g, the $SV_{2,0}$ and $SVJ_{2,0}$ models.

2.3.1 Volatility Structure and a Useful State Reparameterization

We reparameterize state X_t in a more convenient coordinate system, using its standard spectral decomposition:

$$X_t = \mathcal{O}_t \mathcal{V}_t \mathcal{O}_t' \ , \tag{2.17}$$

where $\mathcal{V}_t = \begin{pmatrix} \mathcal{V}_{1t} & 0 \\ 0 & \mathcal{V}_{2t} \end{pmatrix}$ is a diagonal matrix of positive eigenvalues $\mathcal{V}_{1t} \geq \mathcal{V}_{2t}$ and $\mathcal{O}_t = [\mathcal{O}_{1t}, \mathcal{O}_{2t}]$ is an orthonormal rotation matrix with eigenvectors $\mathcal{O}_{1t}, \mathcal{O}_{2t}$. The matrix of eigenvectors can be expressed in polar coordinates by means of a single parameter $\alpha_t \in (-\pi/2, \pi/2]$: $\mathcal{O}_t = \begin{pmatrix} \cos(\alpha_t) & -\sin(\alpha_t) \\ \sin(\alpha_t) & \cos(\alpha_t) \end{pmatrix}$.

In this way we introduce the state variables:

$$\begin{aligned} V_t &= \mathcal{V}_{1t} + \mathcal{V}_{2t} \ , \\ \xi_t &= \frac{\mathcal{V}_{1t} - \mathcal{V}_{2t}}{\mathcal{V}_{1t} + \mathcal{V}_{2t}} \ , \end{aligned}$$

and α_t. V_t is diffusive variance, while $\xi_t \in [0, 1]$ can be interpreted as the composition of the diffusive variance. State component α_t finally captures volatility effects distinct from volatility level or composition effects. The reparametrization of X_t in terms of (V_t, ξ_t, α_t) reads as follows.

Lemma 1 (V-ξ-α decomposition)**.** The symmetric 2×2 state matrix X_t is decomposed into a volatility part V_t, a structural part ξ_t, and a third state variable $\mathcal{U}(\alpha_t)$, as follows:

$$X_t = \frac{V_t}{2}\Big[Id_2 + \xi_t \cdot \mathcal{U}(\alpha_t)\Big] \, , \tag{2.18}$$

where Id_2 is the 2×2 identity matrix and

$$\mathcal{U}(\alpha_t) = \begin{pmatrix} \cos(2\alpha_t) & \sin(2\alpha_t) \\ \sin(2\alpha_t) & -\cos(2\alpha_t) \end{pmatrix} \, . \tag{2.19}$$

Identity (2.18) features a number of convenient properties. First, is homogenous in V_t. Second, it further decomposes the volatility unrelated part of X_t into volatility structure ξ_t and the component that depends only on α_t, which captures dynamic properties not linked to the level and the structure of the volatility. Fourth, identity (2.18) is convenient to study expressions of the form $tr(HX_t)$, for some given 2×2 matrix H, because of the rotation invariance properties of the trace, as discussed in (A.10). Such expressions drive key quantities in the $SV_{3,1}$ and $SVJ_{3,1}$ models, such as the stochastic covariance of returns and volatility or the time-varying jump intensity.

Remark 3. Since ξ_t and α_t are bounded, identity (2.18) produces the *feasible set* of admissible values of X_t for a given variance level V_t, by varying ξ_t and α_t on a compact grid.

Remark 4. Reparametrization (2.18) allows us to isolate within models $SV_{3,1}$ and $SVJ_{3,1}$ the incremental pricing effects of (i) dynamic volatility interactions (i.e., non diagonal matrices M, Q, R or Λ) and (ii) a higher dimensional state space (i.e., non diagonal matrix X_t), in comparison, e.g., to benchmark lower-dimensional dynamics with independent factors. For instance, when $\alpha_t = 0$:

$$X_t \;\; = \;\; \begin{pmatrix} \mathcal{V}_{1t} & 0 \\ 0 & \mathcal{V}_{2t} \end{pmatrix} \, , \tag{2.20}$$

we obtain the state of a two-factor model with dynamic interactions. If, furthermore, matrices M, R, Q, Λ are diagonal, we obtain the state space of models $SV_{2,0}$ or $SVJ_{2,0}$. Similarly, if we also restrict ξ_t to 1, then

$$X_t \;\; = \;\; \begin{pmatrix} \mathcal{V}_{1t} & 0 \\ 0 & 0 \end{pmatrix} \, , \tag{2.21}$$

we obtain models $SV_{1,0}$ and $SVJ_{1,0}$ by letting only the upper diagonal elements of matrices M, R, Q and Λ to be different from zero, see Appendix A.1.

2.3.2 The Feasible Set of Volatility Skew and Volatility Term Structure

In the data, there is a substantial variability of implied volatility skew and term structure, which is largely unrelated to the overall level of the volatility; see again Figures 2.1 and 2.2 for a simple illustration. Intuitively, models with dynamic volatility interactions and unspanned skewness components help reproducing such a variability, because the skew of the smile is less tied to volatility shocks. We study whether this flexibility is actually useful for better modelling the volatility surface. For brevity, we focus on pure diffusive models $SV_{2,0}$, $SV_{3,0}$ and $SV_{3,1}$. Models with jumps imply similar results.

We proceed as follows. First, we fix three benchmark variance levels, corresponding to low ($\sqrt{V_1} = 0.1$), average ($\sqrt{V_2} = 0.2$) and high volatility ($\sqrt{V_3} = 0.3$). Second we use identity (2.18) to produce all admissible state matrices X_t given the variance levels V_i. To do so, we vary ξ from 0 to 1 and α from $-\pi/2$ to $\pi/2$. Third, we compute the model implied one-month skew ($\mathcal{S}$) and at-the-money term structure ($\mathcal{M}$),[12] drawing the boundaries of the set of attainable combinations of $(\mathcal{M}, \mathcal{S})$ as black lines. We call this set the *feasible set*. Fourth, we add grey dots for observations in our data that lie within ±5% around the respective volatility level. In this way, we obtain a model-implied version of Figure 2.1, which can be used to quantify in the $(\mathcal{M}, \mathcal{S})$-coordinate system the degree of volatility-unrelated smile variability that is produced by a model. Panel A of Figure 2.7 illustrates the feasible set for the $SV_{3,1}$ model.

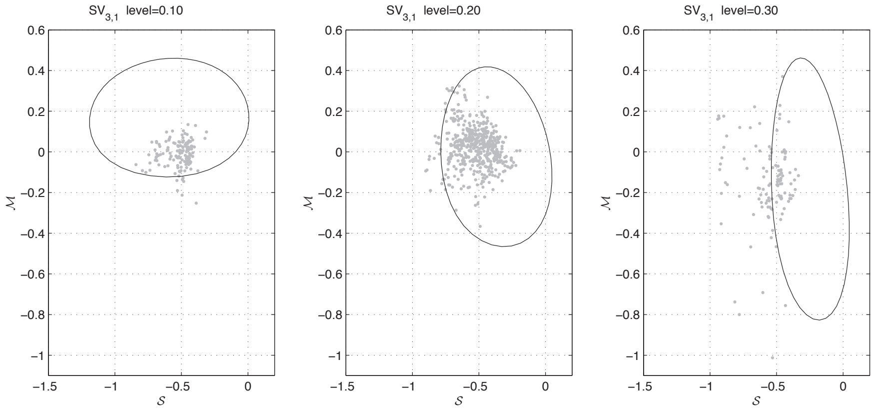
The feasible set of the $SV_{3,1}$ model is a surface bounded by an ellipse. Low volatility levels are associated, on average, with a more (less) limited range of possible model-implied $\mathcal{M}$ ($\mathcal{S}$) shapes. As we move towards higher volatility, from the left to the right panel in Figure 2.7, we obtain a flatter skew and a more negatively-sloped term structure. It is useful to compare the feasible sets of model $SV_{3,1}$ to the data. Overall, we find that the degree of $(\mathcal{M}, \mathcal{S})$-data variability within each volatility regime is quite substantial. For instance, when $\sqrt{V_t} = 0.20$, we observe term structures of both $\mathcal{M} = 0.2$ and $\mathcal{M} = -0.2$ for an implied volatility skew of about $\mathcal{S} = -0.7$. Such $(\mathcal{M}, \mathcal{S})-$structures are difficult to explain accurately within $SV_{2,0}-$type models using a single set of estimated parameters, because of the deterministic, approximate linear relation between $\mathcal{M}$ and $\mathcal{S}$ in these models.

How do the feasible sets of the $SV_{3,1}$ and a three-factor model with independent volatility dynamics compare? To investigate this question, we plot in panel B of Figure 2.7 the feasible set of a three-factor Heston-type model $SV_{3,0}$. Despite the identical dimension of their state spaces, the feasible set of models $SV_{3,0}$ and $SV_{3,1}$ are very different. The feasible set of model $SV_{3,0}$ is bounded by triangular-like shape, in which for any given skewness $\mathcal{S}$ the range of term structures $\mathcal{M}$ is restricted. For instance,

---

[12] Appendix A.2.1 provides details on the computation of $\mathcal{M}$ and $\mathcal{S}$.

Panel A: Feasible set of model $SV_{3,1}$



Panel B: Feasible set of model $SV_{3,0}$

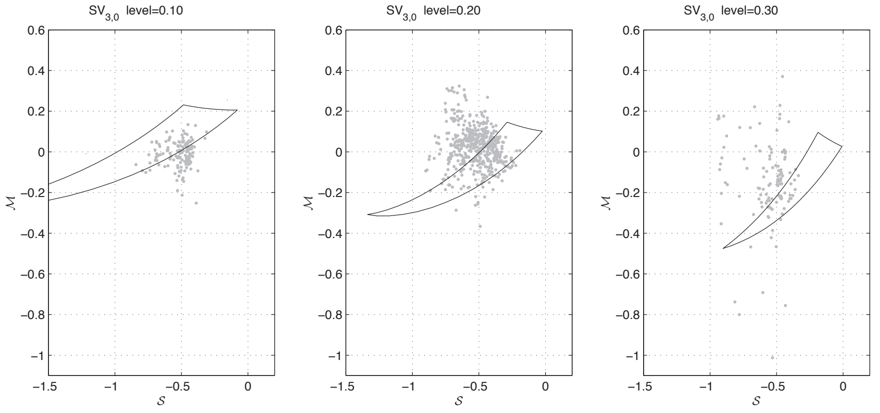


Figure 2.7: Feasible set of ($\mathcal{M}$,$\mathcal{S}$)-combinations for the $SV_{3,1}$ and $SV_{3,0}$ models. We plot the feasible set of skewness and term structure combinations for three volatility levels. The grey dots in each panel represent ($\mathcal{M}$,$\mathcal{S}$) combinations observed in the data for a a range of ±5% around the volatility level, e.g. 19%-21% for the middle panel.

when $\sqrt{V}_t = 0.2$ and $\mathcal{S} = -0.5$, the feasible range of term structures is between about $\mathcal{M} = 0$ and $\mathcal{M} = 0.2$. In contrast, the feasible range of term structures for model $SV_{3,1}$ in Panel A is between $\mathcal{M} = 0.4$ and -0.4, which is more similar to the range of term structures between about $\mathcal{M} = 0.3$ and $\mathcal{M} = -0.4$ in the data.

Comparative Statics

Which is the relation between model parameters and volatility-unrelated skewness and term structure effects? To understand this link, we study the comparative statics of model-implied proxies $\mathcal{M}$ and $\mathcal{S}$, with respect to model parameters M, Q, R, Λ and β.

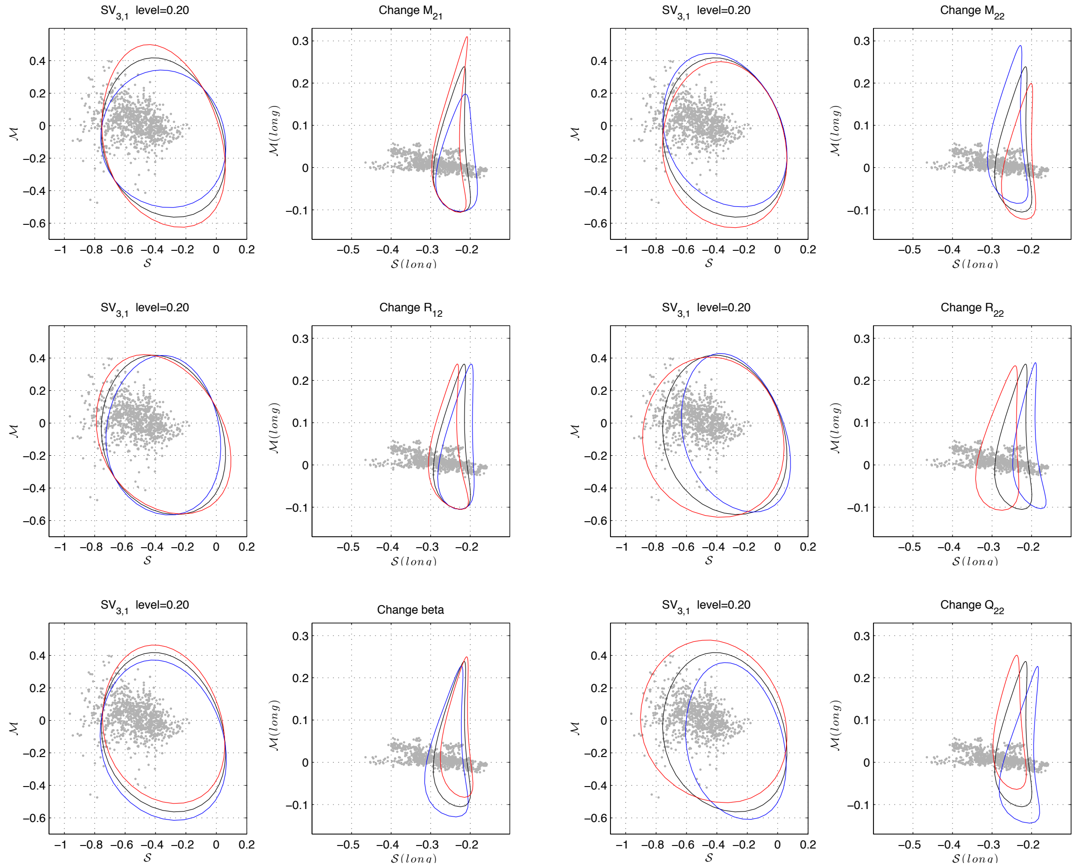


Figure 2.8: Comparative statics of the feasible set with respect to selected model parameters in $SV_{3,1}$ model. For a volatility level $\sqrt{V_t} = 20\%$, we plot in black the ellipse of admissible points for the combinations of short term (long-term) skewness $\mathcal{S}$ and term structure $\mathcal{M}$, in the left (right) panel. We then vary in each plot one parameter by -20% ($+20\%$) and plot in blue (red) the resulting feasible set. The light grey dots in each plot represent $(\mathcal{M}, \mathcal{S})$-combinations observed in the data.

For brevity, we focus on the most significant parameters M_{21}, M_{22}, Q_{22}, R_{12}, R_{22} and β, estimated in Table 2.3 for model $SV_{3,1}$. Similar insights are obtained for model $SVJ_{3,1}$.

For $\sqrt{V_t} = 0.17$, we present in Figure 2.8 the feasible sets, both for the short and the long term segments of the smile. Comparative statics for other volatility levels are similar. In each Panel, we vary the model parameter by ±20% and plot the resulting feasible set in red and blue, respectively. We find that the comparative statics of the different parameters are broadly consistent with the main model intuition. The mean reversion parameters M_{21}, M_{22} mainly influence the range of feasible term structures ($\mathcal{M}$). This is intuitive, as these parameters are directly linked to the mean reversion of the volatility factors. Parameter Q_{22}, which is linked both to the leverage effect and the volatility of volatility affects the admissible range of both $\mathcal{M}$ and $\mathcal{S}$. Finally, volatility-feedback parameters R_{12}, R_{22}, virtually only influence the implied volatility skew ($\mathcal{S}$), while long term volatility parameter β essentially affects only the admissible range of implied volatility term structures ($\mathcal{M}$).

Volatility Structure and the Shape of the Feasible set

How does the elliptical shape of the feasible set in Figures 2.7 and 2.8 relate to the volatility structure captured by (ξ_t, α_t)? We use once more decomposition (2.18) to study the exact shape and composition of the feasible set of $(\mathcal{M}, \mathcal{S})$ combinations for an illustrative volatility level of $\sqrt{V_t} = 0.1$. In Figure 2.9, we plot different combinations of skewness and term structure for the following parameter values: $\xi = (0, 0.2, \ldots 1)$ and $\alpha = (-\pi/4, 0, \pi/4, \pi/2)$.

Given the fixed volatility level, we find that each admissible $(\mathcal{M}, \mathcal{S})$ combination lies on an elliptical curve, in which ξ_t parameterizes the distance from the center of the ellipse and α_t parameterizes the location of each point on the ellipse, i.e., each point inside the ellipse is fully described by the coordinate pair (ξ, α). In particular, given a volatility composition $\xi_t = 1$, say, we see that α_t can generate a wide degree of variations in $\mathcal{M}$ and $\mathcal{S}$, which is by construction unrelated to the volatility level and the volatility composition. For instance, as α_t moves from $-\pi/4$ to $\pi/4$ in Figure 2.9, we obtain points that are moved counter-clockwise from regions of strong negative skew $\mathcal{S}$ and strong positive term structure $\mathcal{M}$ to regions of slightly negative skew $\mathcal{S}$ and strong negative term structure $\mathcal{M}$. Recalling that $\alpha_t = 0$ identifies the state space of a diagonal two-factor model, we can also directly see the more restricted set of admissible $(\mathcal{M}, \mathcal{S})$ combinations, which are all on an approximately piecewise linear function in $(\mathcal{M}, \mathcal{S})$ space, produced by these models as volatility composition ξ_t changes.

2.3.3 Implied Volatility Skew and Term Structure Approximations

For the pure diffusion model $SV_{3,1}$, the shape of the feasible set in Figure 2.9 can be studied analytically using the following proposition.

Proposition 1 (Feasible set of the model SV_{31})**.** In the limit $\tau \to 0$, the feasible set of combinations of $\mathcal{S}_t$ and $\mathcal{M}_t$ in the model $SV_{3,1}$ has the exact form of an ellipse.

Proof. We use the short term asymptotics of (Durrleman and Karoui 2008):

$$\mathcal{S}_t \quad := \quad \lim_{\tau \to 0} \frac{\partial IV_t}{\partial K} = \frac{1}{2V_t^{3/2}} Cov_t(dS_t/S_t, dV_t) \tag{2.22}$$

$$\mathcal{M}_t \quad := \quad \lim_{\tau \to 0} \frac{\partial IV_t}{\partial \tau} = \frac{1}{2V_t^{1/2}} \left(\frac{1}{dt} E_t(dV_t) - V_t^{3/2} \mathcal{C}_t - 3V_t \mathcal{S}_t^2 \right) \tag{2.23}$$

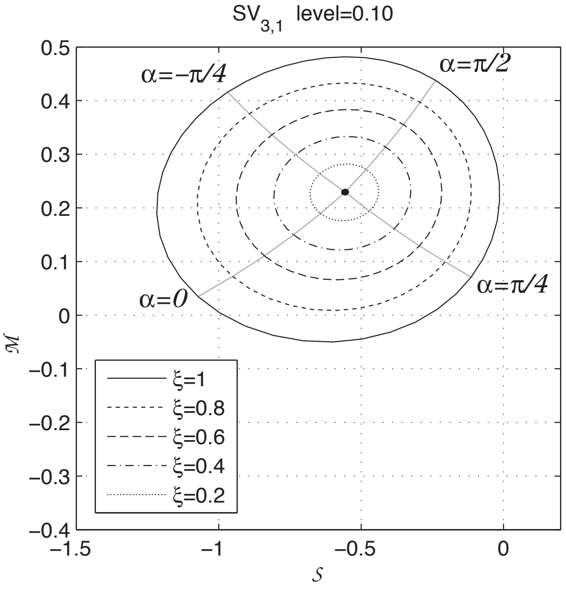


Figure 2.9: Admissible set of skewness and term structure combinations in model $SV_{3,1}$ for a volatility level $\sqrt{V_t} = 10\%$. We plot the admissible model-implied combinations of short term skewness $\mathcal{S}$ and term structure $\mathcal{M}$, implied for different volatility compositions $\xi := \frac{\mathcal{V}_1 - \mathcal{V}_2}{\mathcal{V}_1 + \mathcal{V}_2} = 0.2, 0.4, 0.6, 0.8, 1$, and different unspanned skewness parameters $\alpha = -\pi/4, 0, \pi/4, \pi/2$ in model $SV_{3,1}$. For $\xi = 0$, the set collapses to a single point in the center of the ellipse. Due to the periodicity of $\cos(2\alpha)$, model-implied combinations for $\pi/2$ are equal to those for $-\pi/2$. Shape and location of the ellipse depend on the volatility level.

where K is the strike price, IV_t the Black-Scholes implied volatility and $\mathcal{C}_t$ is a convexity adjustment which has little quantitative impact in out setting. Expression (2.22) has already been obtained as $R'QX_t$ in (2.7). From dynamics (2.2) we obtain:

$$\frac{1}{dt} E_t(dV_t) = \beta Q'Q + 2tr(MX_t), \tag{2.24}$$

which is the leading term of (2.23) the limit $\tau \to 0$. We can now use decomposition (2.18) to rewrite the pair $(\mathcal{M}_t, \mathcal{S}_t)$ in polar coordinates:

$$\begin{aligned} \mathcal{M}_t &= V_t^{1/2} \Big(\mathcal{M}_0 + \xi_t \big[f_3 \cos(2\alpha_t) + f_4 \sin(2\alpha_t) \big] \Big) && (2.25) \\ \mathcal{S}_t &= \frac{1}{V_t^{1/2}} \Big(\mathcal{S}_0 + \xi_t \big[f_1 \cos(2\alpha_t) + f_2 \sin(2\alpha_t) \big] \Big) && (2.26) \end{aligned}$$

$$\begin{array}{ll} \mathcal{S}_0 = \frac{1}{2}\left((RQ)_{11} + (RQ)_{22}\right) & \mathcal{M}_0 = \frac{1}{2}\left(M_{11} + M_{22} + Tr[\beta Q'Q] V_t^{-1}\right) \\ f_1 = \frac{1}{2}\left((RQ)_{11} - (RQ)_{22}\right) & f_3 = \frac{1}{2}\left(M_{11} - M_{22}\right) \\ f_2 = \frac{1}{2}\left((RQ)_{12} + (RQ)_{12}\right) & f_4 = \frac{1}{2}\left(M_{12} + M_{21}\right) \end{array}$$

Observing that (2.25) and (2.26) describe a general ellipse with center $(\mathcal{S}_0, \mathcal{M}_0)$ concludes the proof. $\square$

The functional forms (2.25) and (2.26) explain our first empirical results in Figure 2.7 and 2.9. The center of the ellipse $(\mathcal{S}_0, \mathcal{M}_0)$ is completely determined by the level V_t of the volatility. Consistently with Figure 2.9, volatility composition ξ_t parametrizes the distance of each point from the center.

We can now use the nesting relationships (2.20) and (2.21), to see how the feasible set shrinks for lower dimensional models. The restriction for the two factor Heston model $SV_{2,0}$ is $\alpha_t = 0$. In this case, (2.25) and (2.26) parametrize a line in $(\mathcal{S}, \mathcal{M})$-space. For the one-factor Heston model $SV_{1,0}$, the feasible set shrinks to the point $(\mathcal{S}_0, \mathcal{M}_0)$.

Overall, these findings confirm the key role of volatility interactions and unspanned skewness features for an improved modelling of volatility-unrelated skewness and term structure tradeoffs.

Volatility-Unrelated Term Structure of Implied Volatility Skews

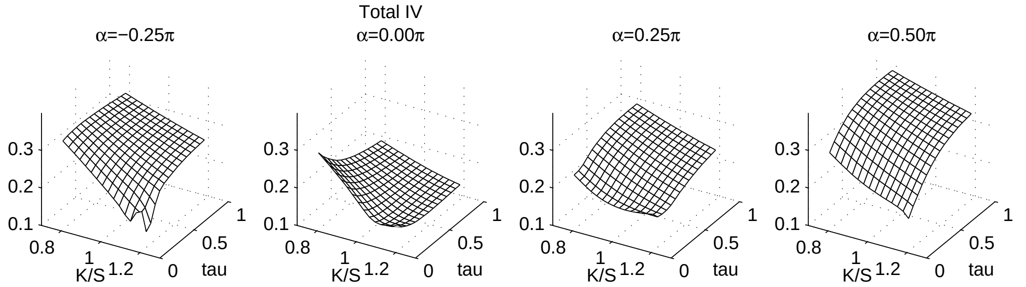
In models $SV_{3,1}$ and $SVJ_{3,1}$, dynamic volatility interactions and unspanned skewness produce a dynamic term structure of implied volatility skews. Using our (V_t, ξ_t, α_t) parametrization, we isolate more clearly these effects. Figure 2.10 illustrates the resulting model-implied mechanics for $\sqrt{V_t} = 0.17$ and $\xi_t = 1$.[13]

In Panel A variations of α_t produce a broad variety of term structures of implied volatility skews for model $SV_{3,1}$. For instance, while for $\alpha_t = \pi/2$ the model generates a steep increasing term structure, combined with a steep skew at short and long maturities, for $\alpha_t = 0$ it yields flat or decreasing term structures, combined with a pronounced (flat) skew at short (long) maturities. These effects are richer in the $SVJ_{3,1}$ model, because of the short horizon effects produced by the jump component. The decomposition of the implied volatility skew term structure into (i) A diffusive part (middle row of Panel B) and (ii) a residual generated by the jump term (bottom row of Panel B) produces additional insights. As expected, the diffusive part of the model dominates the implied volatility skew term structures for maturities above roughly three weeks, while the jump component has dominating effects for maturities roughly up to three weeks. Interestingly, the model-implied jump-driven segment of the smile can feature both strong short-term skews ($\alpha_t = 0.5$) or a short term smile ($\alpha_t = 0$), independently of the level and the structure of the volatility. This evidence confirms that the $SV_{3,1}$ and $SVJ_{3,1}$ models can produce a rich variety term structures of implied volatility skews, which are largely independent of the level and the composition of the volatility.

---

[13] The choice $\xi_t = 1$ is for illustration and produces the most pronounced implied volatility smile effects.

Panel A: Pure diffusion model $SV_{3,1}$



Panel B: Jump diffusion model $SVJ_{3,1}$

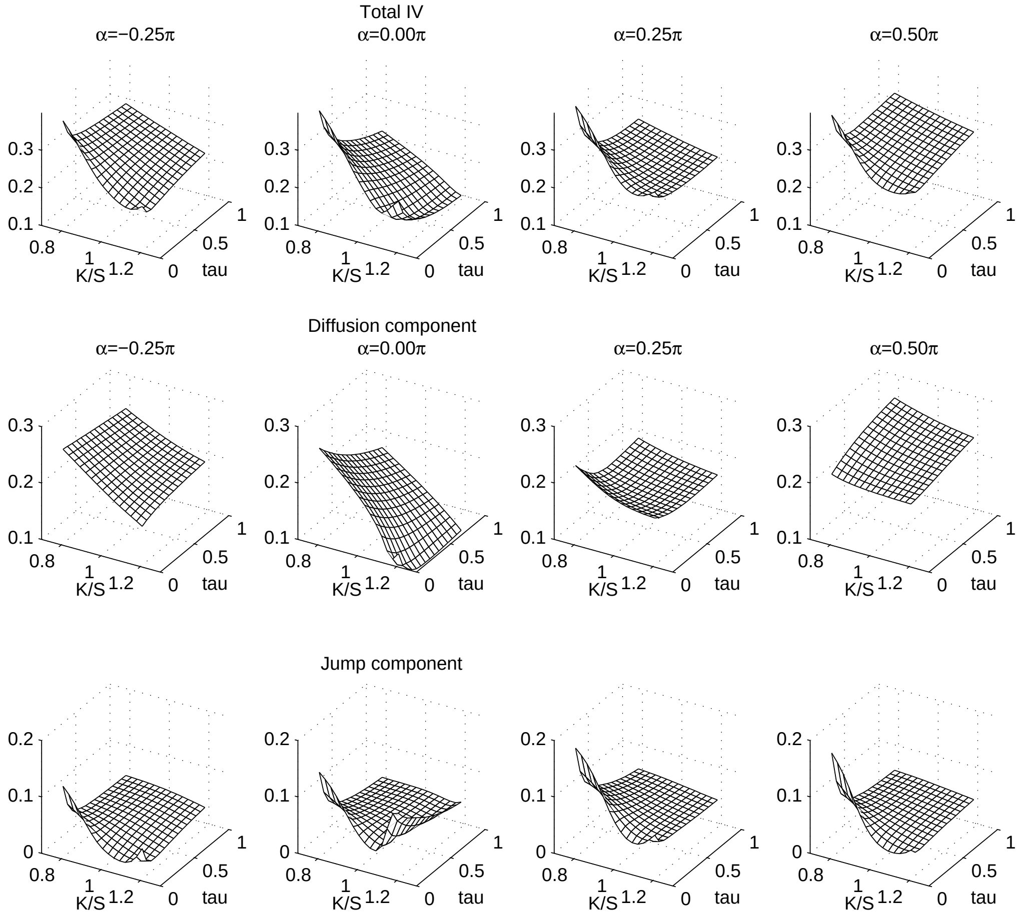
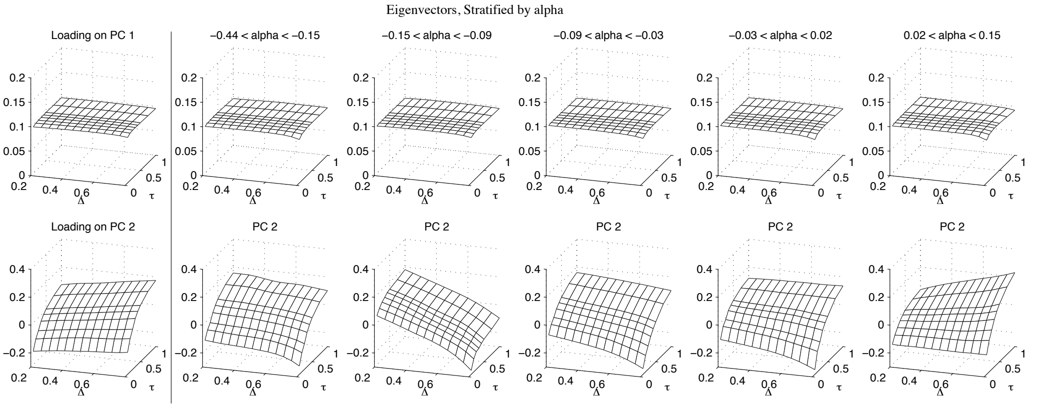


Figure 2.10: Model-implied volatility surfaces for a volatility level $\sqrt{V} = 0.20$ and a volatility composition $\xi = 1$, in dependence of unspanned skewness parameter α. Panel A present plots for the pure diffusion $SV_{3,1}$ model. Panel B presents plots for the $SVJ_{3,1}$ jump diffusion model. In panel B, the top graphs plot the total volatility surface, the middle one illustrates the contribution of diffusive volatility to the surface and the bottom one illustrates the contribution of the jump component.

Panel A: PCA stratified by $SV_{3,1}$ model-implied α



Panel B: PCA stratified by $SVJ_{3,1}$ model-implied α

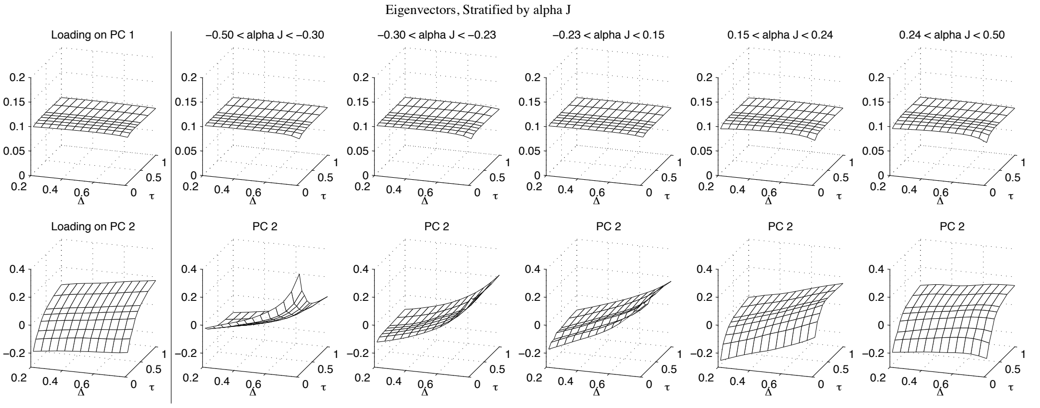


Figure 2.11: Factor loadings of the first two principal components of the implied volatility surface as a function of model-implied component α_t. The left column presents results for an unconditional PCA. The following five columns present results of 5 conditional PCAs, in which we stratify the data into quintiles of model-implied state α_t. Panel A (B) applies a stratification with respect to the model-implied α_t of $SV_{3,1}$ ($SVJ_{3,1}$).

Is there also a way to isolate volatility-unrelated effects in the term structure of implied volatility skews using a more model-free approach? To address this question, we stratify our sample in quintiles of estimated latent state component α_t for $SV_{3,1}$ and $SVJ_{3,1}$ models, respectively. We then perform within each quintile a Principal Component Analysis of the S&P 500 index option implied volatility surfaces, using a standardized grid of maturities (1, 2, 3, 4, 6, 9, 12 months) and moneynesses (Black-Scholes deltas of 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8). We find on average two principal components in each quintile, which are plotted in Figure 2.11, Panel A and B for α estimates in the $SV_{3,1}$ and $SVJ_{3,1}$ models, respectively.

Within each quintile, the first estimated principal component reflects a very stable volatility level, which is remarkably unrelated to moneyness and term structure. The second component is instead typically related to both the moneyness and maturity dimensions, indicating that it acts as a factor moving the entire term structure of implied volatility skews. Interestingly, this second component has quite different properties across quintiles, indicating that the term structure of implied volatility skews reacts differently, i.e., dynamically, in dependence of the proxy α_t for unspanned skewness. These findings are related to our alternative interpretation of the $SVJ_{3,1}$ model as stochastic coefficient models in Sec. 2.3.4.

Stochastic Feedback Effects and Volatility-Unrelated Risk Neutral Skewness

A main driver of risk neutral skewness in our models is a stochastic volatility feedback between returns and volatility. Naturally, the volatility-unrelated variations in risk neutral skewness are linked to the degree of volatility-unrelated variation of volatility feedback effects. We illustrate these aspects in Figure 2.12 for model $SV_{3,1}$, in which we have:

$$\begin{aligned} corr_t(dS_t/S_t, dV_t) &= \frac{tr(R'QX_t)}{\sqrt{tr(X_t)tr(Q'QX_t)}} \\ &= \frac{tr(R'Q(I_2+\xi_t\mathcal{U}(\alpha_t)))}{\sqrt{tr(I_2+\xi_t\mathcal{U}(\alpha_t))tr(Q'Q(I_2+\xi_t\mathcal{U}(\alpha_t)))}} \ , \end{aligned} \tag{2.27}$$

where the last equality follows from Lemma 1. Precisely, we plot correlation (2.27), for different volatility compositions ξ_t, as a function of unspanned skewness parameter α_t.

The case $\alpha_t = 0$ corresponds to the range of possible model-implied volatility compositions and volatility feedback effects in model $SV_{2,0}$. For this case, the volatility feedback effect is monotonic in the volatility composition. For instance, as ξ_t goes from 1 to 0 (i.e., $\mathcal{V}_{1t} - \mathcal{V}_{2t} \to 0$; $\mathcal{V}_{1t} \geq \mathcal{V}_{2t}$) the correlation between volatility and returns goes from an upper bound of about -0.15 to a lower bound of about -0.45. This feature generates a tight skewness–term structure tradeoff: The model tries to fit a more negative skew in the data with a stronger volatility feedback effect, which also forces the implied state to a more equal volatility composition. In doing so, the model puts higher weights on the more strongly mean reverting factor $\mathcal{V}_{2t}$, implying ceteris paribus a more negative implied volatility term structure.

The introduction of an unspanned skewness dimension in model $SV_{3,1}$ weakens this tight link and enlarges the range of admissible volatility composition and feedback effects. For instance, while in model $SV_{2,0}$ ($\alpha_t = 0$) a volatility feedback below -0.4 is accessible only with a volatility composition $\xi_t \in [0, 0.25]$, in model $SV_{3,1}$ it is achievable by any volatility composition when α_t is above approximately 0.1.

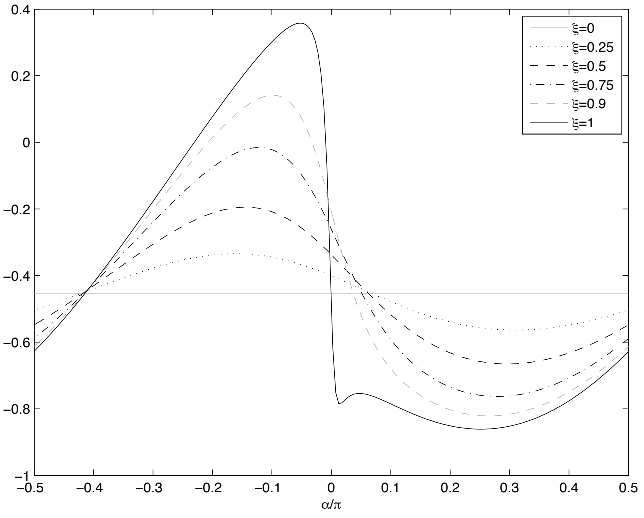


Figure 2.12: Model-implied correlations between returns and volatility in model $SV_{3,1}$, for different volatility compositions $\xi = 0, 0.25, 0.5, 0.75, 0.9, 1$, as a function of unspanned skewness parameter $\alpha \in (-\pi/2, \pi/2]$.

2.3.4 Stochastic Coefficients Model

The risk neutral dynamics of the rotated volatility factors $\mathcal{V}_{1t}$, $\mathcal{V}_{2t}$ in (2.17) provide additional insight into the role of these state variables. Using Itô's Lemma we obtain, after lengthy calculations:[14]

$$d\mathcal{V}_{1t} = \left(\beta(\tilde{Q}_t'\tilde{Q}_t)^{11} + 2(\tilde{M}_t)^{11}\mathcal{V}_{1t} + \frac{\mathcal{V}_{1t}(\tilde{Q}_t'\tilde{Q}_t)^{22} + \mathcal{V}_{2t}(\tilde{Q}_t'\tilde{Q}_t)^{11}}{\mathcal{V}_{1t} - \mathcal{V}_{2t}}\right) dt + 2\sqrt{\mathcal{V}_{1t}(\tilde{Q}_t'\tilde{Q}_t)^{11}} d\nu_{1t} \tag{2.28}$$

$$d\mathcal{V}_{2t} = \left(\beta(\tilde{Q}_t'\tilde{Q}_t)^{22} + 2(\tilde{M}_t)^{22}\mathcal{V}_{2t} - \frac{\mathcal{V}_{1t}(\tilde{Q}_t'\tilde{Q}_t)^{22} + \mathcal{V}_{2t}(\tilde{Q}_t'\tilde{Q}_t)^{11}}{\mathcal{V}_{1t} - \mathcal{V}_{2t}}\right) dt + 2\sqrt{\mathcal{V}_{2t}(\tilde{Q}_t'\tilde{Q}_t)^{22}} d\nu_{2t} \tag{2.29}$$

with $(\nu_1, \nu_2)'$ a standard bivariate Brownian motion and 2×2 random matrices $\widetilde{M}_t = \mathcal{O}_t' M \mathcal{O}_t$ and $\widetilde{Q}_t = \mathcal{O}_t' Q \mathcal{O}_t$. This shows that the $\mathcal{V}_{1t}$ and $\mathcal{V}_{2t}$ are conditionally independent stochastic volatility processes, in which the (stochastic) volatility of volatility and drift parameters depend only on the random matrix $\mathcal{O}_t$. Figure 2.13 plots the time series of $\widetilde{M}_{11}, \widetilde{M}_{22}, \widetilde{Q}_{11}$ and $\widetilde{Q}_{22}$ for model SV_{31} .

Conditional on $\mathcal{O}_t$, processes $(\mathcal{V}_{1t}, \mathcal{V}_{2t})$ behave as independent Bessel processes, in which the linear drift is perturbed by the nonstandard term:

$$\pm \frac{\mathcal{V}_{1t}(\tilde{Q}_t'\tilde{Q}_t)^{22} + \mathcal{V}_{2t}(\tilde{Q}_t'\tilde{Q}_t)^{11}}{\mathcal{V}_{1t} - \mathcal{V}_{2t}} \ . \tag{2.30}$$

---

[14] See also (Benabid, Bensusan and El Karoui 2009). The proof is available on request.

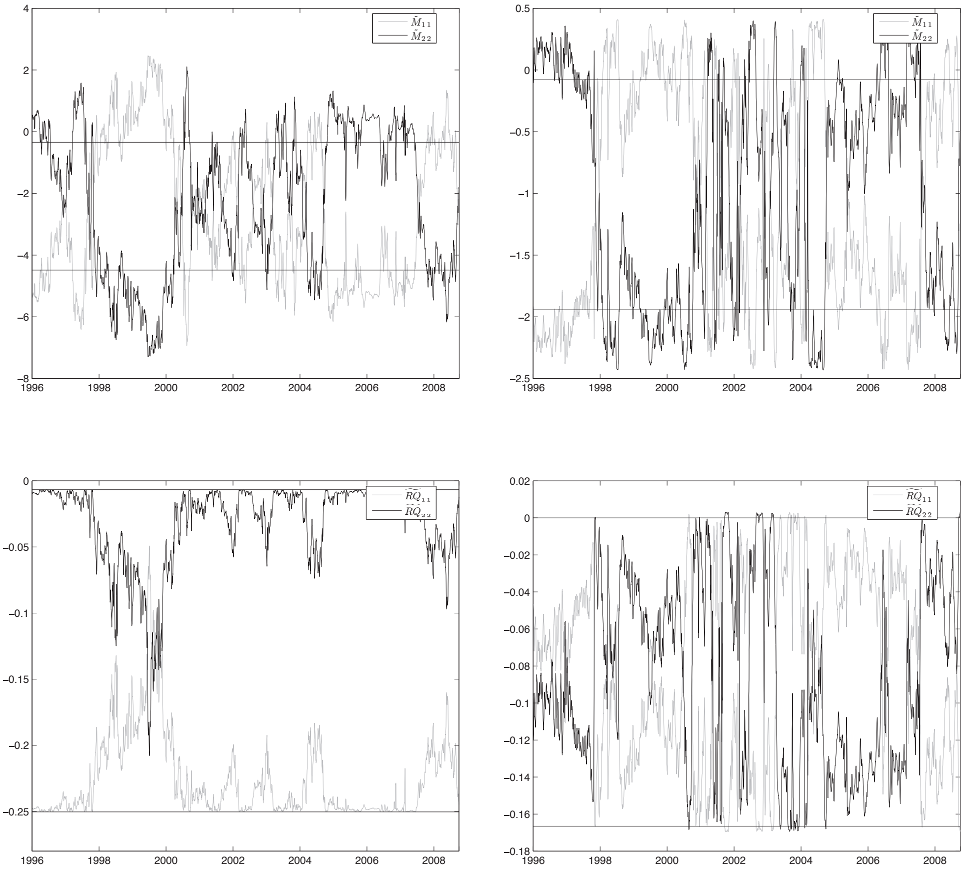


Figure 2.13: Dynamic parameter representation for model $SV_{3,1}$ (left panels) and $SVJ_{3,1}$ (right panels). Top: mean reversion parameters $\widetilde{M}_{11}$ for the small factor (grey line) and $\widetilde{M}_{22}$ for the large factor (black line). The horizontal lines depict the case for the nested two-factor model (i.e. $\alpha = 0$). Bottom: leverage parameters $\widetilde{RQ}_{11}$ for the small factor (grey line) and $\widetilde{RQ}_{22}$ for the large factor (black line). Horizontal lines depict the nested two factor case.

This term ensures that the ranking of eigenvalues $\mathcal{V}_{1t}, \mathcal{V}_{2t}$ is preserved, but it is typically small. We can therefore interpret the risk neutral dynamics of the volatility components $\mathcal{V}_{1t}$, $\mathcal{V}_{2t}$ in our model as a two-factor random coefficient stochastic volatility model, in which the random coefficients are driven by the unspanned skewness component $\mathcal{O}_t = \mathcal{O}(\alpha_t)$. Variable α_t can impact in two ways on the model-implied volatility surface. First, via the stochastic mean reversion and volatility of volatility coefficients in dynamics (2.28)-(2.29), it produces a variety of effects on the term structure of the volatility. Second, it impacts on the time varying jump intensity and the volatility feedbacks, as:

$$\lambda_t = \lambda_0 + tr(\Lambda X_t) = \lambda_0 + (\widetilde{\Lambda}_t)^{11}\mathcal{V}_{1t} + (\widetilde{\Lambda}_t)^{22}\mathcal{V}_{2t} \ , \tag{2.31}$$

and

$$\frac{1}{2}Cov_t(dS_t/S_t, d(\mathcal{V}_{1t} + \mathcal{V}_{2t})) = tr(R'QX_t) = (\widetilde{R'Q_t})^{11}\mathcal{V}_{1t} + (\widetilde{R'Q_t})^{22}\mathcal{V}_{2t} \ , \qquad (2.32)$$

where $\widetilde{\Lambda}_t = \mathcal{O}_t'\Lambda\mathcal{O}_t$ and $\widetilde{R'Q_t} = \mathcal{O}_t'R'Q\mathcal{O}_t$. Note that conditional on $\mathcal{O}_t$, equations (2.31) and (2.32) define a a time varying intensity and volatility feedback effect consisten with a two-factor Bates (2000)-type model. Relative to this model, model $SVJ_{3,1}$ produces an additional degree of skewness variability by making the coefficients $(\widetilde{\Lambda}_t)^{ii}$ and $(\widetilde{R'Q_t})^{ii}$ in formulae (2.31) a (2.32) stochastic. Since this additional variability is completely driven by matrix $\mathcal{O}_t = \mathcal{O}(\alpha_t)$, these unspanned skewness effects are largely captured by state variable α_t.

2.4 Conclusions

Using a new option valuation framework, featuring interdependent volatility risks and a stochastic skewness component unrelated to the volatility factors, we analyze the pricing of S&P 500 index options. We estimate two specifications of our model based on options data from January 1996 to September 2009 and find that they provide superior pricing performance over a number of benchmark two- and three-factor affine volatility models in the literature, with reductions in average root mean square pricing error of about 20% out-of-sample. We find that the improved fit of our model is largely due to an improved modelling of the term structure of implied-volatility skews. We explain the role of the matrix state space in generating implied volatility skews and term structures. We show that the feasible set of the diffusive version of our model is approximately an ellipse at short maturities and that it is considerably larger than the admissible set of a 3-factor Heston model.

In addition to highlighting the usefulness of multi-factor risk specifications for modelling the dynamics of implied volatility smiles, our results emphasize the key role of dynamic volatility interactions and volatility-unrelated skewness for option valuation purposes. More generally, they raise the question of how the dynamics of a multi-factor option pricing model ought to be specified. Our findings show that three-factor state dynamics based on the class of MAJD in (Leippold and Trojani 2008) can provide a powerful framework for the specification of interacting volatility components and volatility-unrelated skewness effects within a tractable specification.

Appendix A

Appendix for “Three make a Dynamic Smile”

A.1 Nested Models

Several well-studied affine option pricing models with independent factors are nested in our framework, if we allow β to be a diagonal matrix instead of a scalar. In this case, the independent volatility factors can be written as diagonal elements of X_t. Below, we show the equivalence of the processes and how the parameters can be converted from the notation in the original papers into our notation.

A.1.1 Diffusive Models

The return dynamics of the $SV_{2,0}$ two-factor Heston model in Christoffersen et al. (2009) is

$$\frac{dS}{S} = (r-q)dt + \sqrt{V_1}dz_1 + \sqrt{V_2}dz_2 \tag{A.1}$$

where r is the risk-free rate, q the dividend yield, z_i are independent Brownian motions and V_i are independent stochastic volatility factors with the following dynamics:

$$dV_i = (a_i - b_iV_i)dt + \sigma_i\sqrt{V_i}dw_i \qquad i = 1,2 \tag{A.2}$$

where the correlation between dz_i and dw_j is $\delta_{ij}\rho_i$.

Using the notation $X = \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}$and $dZ = \begin{pmatrix} dz_1 & dZ_{12} \\ dZ_{21} & dz_2 \end{pmatrix}$, dynamics (A.1) can be written as

$$\frac{dS}{S} = (r-q)dt + tr[\sqrt{X_t}dZ],$$

which is exactly the diffusive part of our return equation (2.1). To show the equality of the volatility factors, we establish that the diagonal elements of X_t in (2.2) are independent CIR processes when the parameter matrices M, R, Q are diagonal. We start by explicitly writing the diagonal elements of X_t in this case:

$$dX_{ii} = \left(\beta Q_{ii}^2 + 2M_{ii}X_{ii}\right) dt + \sum_k (\sqrt{X})_{ki}\, dB_{ki} \tag{A.3}$$

To eliminate the seeming interdependence of the diagonal elements, we introduce n new independent Brownian motions dW_i:

$$dW_i = \frac{1}{\sqrt{X_{ii}}} \sum_k (\sqrt{X})_{ki}\, dB_{ki}$$

This allows us to express (A.3) as n independent CIR processes:

$$dX_{ii} = \left(\beta Q_{ii}^2 + 2M_{ii}X_{ii}\right) dt + 2Q_{ii}\sqrt{X_{ii}} dW_i \tag{A.4}$$

To convert our notation into the notation of (A.2), simply set

$$a_i = \beta_{ii} Q_{ii}^2, \quad b_i = -2M_{ii}, \quad \sigma_i = 2Q_{ii}, \text{ and } \rho_i = R_{ii}.$$

Remark 5. Our state matrix X_t will generally not remain diagonal, even if all parameter matrices and the initial state X_0 are diagonal. This does not void the nesting argument, because $X_{12,t}$ does not enter the pricing equation in that case. There is no economic interpretation for the process $X_{12,t}$ in such a setting, it is a mere artifact of writing a two-dimensional CIR process in matrix form.

A.1.2 Jump Parameters

The jump intensity in Bates (2000) is given as $\lambda_t = \lambda_0 + \lambda_1 V_{1t} + \lambda_2 V_{2t}$, which is already identical to our jump intensity $\lambda_t = \lambda_0 + tr(\Lambda X_t)$, if we write $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$. Our choice of the jump size distribution is the same as in Bates (2000).

A.2 Short-maturity Smile Asymptotics

We choose a convenient, low-dimensional framework to study the unspanned volatility effects in our model. Let $IV_t(T, K)$ be the Black-Scholes option implied volatility at time t for maturity T and strike price K, and consider the following approximation of the implied volatility smile:[1]

$$\overline{IV}_t(T,K) = V_t^{1/2} + \mathcal{S}_t \frac{K-S_t}{S_t} + \mathcal{M}_t(T-t) + \frac{1}{2}\mathcal{C}_t \left(\frac{K-S_t}{S_t}\right)^2 \tag{A.5}$$

with

$$\mathcal{S}_t = S_t \lim_{T\to t} \frac{\partial IV(T,S_t)}{\partial K} \ , \quad \mathcal{M}_t = \lim_{T\to t} \frac{\partial IV(T,S_t)}{\partial T} \ , \mathcal{C}_t = S_t^2 \lim_{T\to t} \frac{\partial^2 IV(T,S_t)}{\partial^2 K} \ . \tag{A.6}$$

Where $V^{1/2}$ is the short-term, at the money volatility level, $\mathcal{S}_t$ is the skew for short maturities, i.e., the short maturity limit of the derivative of the at-the-money implied volatility with respect to moneyness K/S_t. $\mathcal{M}_t$ is the smile term structure for short maturities, i.e., the short maturity limit of the derivative with respect to maturity T. Finally, $\mathcal{C}_t$ is the smile convexity for short maturities.

A.2.1 Construction of Level, Skewness and Term Structure Factors from Data and Model

For our empirical studies, we construct proxies for the skewness and term structure factors using two different methods. Whenever we calculate these quantities from the data or from model fits, we perform regression (A.5), separately for each day of data (model-implied prices). To obtain the short-term skew $\mathcal{S}$ and the short term structure $\mathcal{M}$, we consider only options with $\tau < 73$ days and $0.85 \leq K/S_t \leq 1.15$. To obtain the long-term skew $\mathcal{S}_t^{\text{long}}$ and the long term structure $\mathcal{M}_t^{\text{long}}$ in Figures 2.1 and 2.8, we consider options with $\tau \geq 122$ days and $0.7 \leq K/S_t \leq 1.3$.

Whenever we calculate feasible set as in Figures 2.7, 2.8 and 2.9, we calculate the derivatives (A.6) numerically. More precisely, we approximate the skew as $\mathcal{S}_t = \frac{\partial IV(T,S_t)}{\partial K}$ at $\tau = 0.25$ (6) months for the short (long) term. We approximate the at the money term structure as $\mathcal{M}_t = \frac{IV(\tau_1)-IV(\tau_0)}{\tau_1-\tau_0}$, where $\tau_1 = 0.25$ (12) months for the short (long) term structure and $\tau_0 = 0$. In the $SV_{3,1}$-model, we calculate $IV(\tau_0) = tr(X)^{1/2}$ and use this quantity directly.

---

[1] See, among others, Dumas, Fleming and Whaley (1998), Durrleman (2010) and Durrleman and Karoui (2008).

A.3 Proofs and Additional Expressions

A.3.1 Proof of Lemma 1

Since X_t is a symmetric positive definite matrix, we can always write

$$X_t = \mathcal{O}_t \mathcal{V}_t \mathcal{O}_t' \ , \tag{A.7}$$

where $\mathcal{V}_t$ is a 2×2 diagonal matrix of positive eigenvalues $\mathcal{V}_{1t}, \mathcal{V}_{2t}$ and $\mathcal{O}_t = [\mathcal{O}_{1t}, \mathcal{O}_{2t}]$ is a 2×2 orthogonal matrix of eigenvectors $\mathcal{O}_{1t}, \mathcal{O}_{2t}$ having unit norm. A convenient parametrization of $\mathcal{O}_t$ by means of a single parameter $\alpha_t \in [-\pi/2, \pi/2]$ is obtained using standard polar coordinates:

$$\mathcal{O}_t = \begin{pmatrix} \cos(\alpha_t) & -\sin(\alpha_t) \\ \sin(\alpha_t) & \cos(\alpha_t) \end{pmatrix} \quad . \tag{A.8}$$

The sum of $\mathcal{V}_{1t}$ and $\mathcal{V}_{2t}$ naturally parametrizes the spot volatility of returns. Thus we can define the volatility level factor V_t and a dimensionless factor ξ_t that measures the composition of the volatility:

$$V_t := tr(X_t) = tr(\mathcal{V}_t). \qquad \xi_t := \frac{\mathcal{V}_{1t} - \mathcal{V}_{2t}}{\mathcal{V}_{1t} + \mathcal{V}_{2t}} = \frac{\mathcal{V}_{1t} - \mathcal{V}_{2t}}{V_t} \tag{A.9}$$

Using this notation we obtain:

$$\begin{aligned} X_t &= \mathcal{O}_t \mathcal{V}_t \mathcal{O}_t' = \mathcal{O}_t V_t \frac{\mathcal{V}_t}{V_t} \mathcal{O}'_t \\ &= V_t \left[\mathcal{O}_t \begin{pmatrix} \frac{1+\xi_t}{2} & 0 \\ 0 & \frac{1-\xi_t}{2} \end{pmatrix} \mathcal{O}'_t \right] \\ &= \frac{V_t}{2} \left[\mathcal{O}_t Id_2 \mathcal{O}'_t + \xi_t \cdot \mathcal{O}_t \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \mathcal{O}'_t \right] \\ &= \frac{V_t}{2} \left[Id_2 + \xi_t \cdot \mathcal{U}_t \right] \end{aligned}$$

where Id_2 is the 2×2 identity matrix and

$$\mathcal{U}(\alpha_t) = \begin{pmatrix} \cos(2\alpha_t) & \sin(2\alpha_t) \\ \sin(2\alpha_t) & -\cos(2\alpha_t) \end{pmatrix} \ .$$

Note that $\mathcal{U}(\alpha_t)$ is a reflection matrix with trace zero and determinant minus one such that all components are bounded in the interval $[-1, 1]$. Therefore, it can be conveniently used to specify different correlation processes, such as, for instance, those needed to specify stochastic volatility feedback effects.

Remark 6. In order to make decomposition (2.18) unique, one has to choose an ordering of the eigenvalues, and thus the sign of ξ_t. We choose $\mathcal{V}_{1t} > \mathcal{V}_{2t}$ and therefore $0 \leq \xi_t \leq 1$.

Remark 7. Lemma 1 can be used to decompose expressions of the form $Tr[HX_t]$ when H is a 2×2 parameter matrix, as follows:

$$\begin{aligned} Tr[HX_t] &= \frac{V_t}{2}\Big[Tr(H) + \xi_t \cdot Tr\big(H\,\mathcal{U}(\alpha_t)\big)\Big] \\ &= \frac{V_t}{2}\Big[Tr(H) + \xi_t \cdot \Big(\cos(2\alpha_t)(H_{11} - H_{22}) + \sin(2\alpha_t)(H_{12} + H_{21})\Big)\Big] \end{aligned} \tag{A.10}$$

A.3.2 Parameter Identification

We first discuss the identification of the diffusive parameters. Every stochastic process is uniquely characterized by its infinitesimal generator. The infinitesimal generator of the joint process for stock returns $Y_t := dS_t/S_t$ and the factor X_t is (see Leippold and Trojani (2008)):

$$\begin{aligned} \mathcal{L}_{Y,X} &= \left(r - q - \frac{1}{2}Tr[X]\right)\frac{\partial}{\partial Y} + \frac{1}{2}TrX\frac{\partial^2}{\partial Y^2} + 2Tr[XR'QD]\frac{\partial}{\partial Y} + \\ &\quad + Tr\left[\left(\beta Q'Q + MX + XM'\right)D + 2XDQ'QD\right] \end{aligned} \tag{A.11}$$

where $(D)_{ij} = \frac{\partial}{\partial X_{ij}}$ is the matrix differential operator.

The parameter set of the diffusive process is $\theta = \{\beta, M, R'Q, Q'Q\}$. Parameter identification requires that the infinitesimal generator be unique for each set of parameters given any state X_t. Maximal identification aims at achieving this goal through the minimal set of parameter restrictions. Equation (A.11) contains an ambiguity that has to be resolved. Let

$$Z_t = DX_tD^{-1},$$

then

$$\mathcal{L}(X_t, \theta) = \mathcal{L}(Z_t, \theta_Z)$$

with $\theta_Z = \{\beta, DMD^{-1}, DR'QD^{-1}, DQ'QD^{-1}\}$.

We now want to identify parameter restrictions on θ that only admit D to be the identity matrix. Without loss of generality, we can assume $|det(D)| = 1$. Next we observe that the expression $Q'Q$ is symmetric by construction, thus $DQ'QD^{-1}$ needs to be symmetric, as well. Symmetry of $DQ'QD^{-1}$ is ensured if D is orthogonal ($D' = D^{-1}$), thus D must be a rotation matrix.

In a next step, we choose M to be lower triangular. This requires D to be lower triangular, in order to ensure DMD^{-1} to be lower triangular, as well. If D is orthogonal, lower triangular and has a determinant of one, it must be a diagonal matrix $\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$ with elements $\alpha, \beta = \pm 1$. We now have $DMD^{-1} = \begin{pmatrix} M_{11} & 0 \\ \beta/\alpha \; M_{21} & M_{22} \end{pmatrix}$. By choosing the sign of $M_{2,1}$ to be positive, we exclude the case $\alpha \neq \beta$, which concludes the identification of the diffusion parameters.

Remark 8. Our choices for M implicitly identify some features of the hidden state. The choice for M to be lower triangular selects the order of the mean reversion speeds of the eigenvalues of X_t. In our setting, this implies that the dominant factor with the fast mean reversion is X_{22t}. Our choice for $M_{21} > 0$ identifies the sign of X_{12t}.

We now need to relate the composite parameters $Q'Q$ and $R'Q$ to the parameter matrices Q and R. We choose Q to be the unique Choleski decomposition of $Q'Q$, i.e. Q upper triangular and positive definite. By simple matrix algebra, we obtain $R = (Q')^{-1}(R'Q)'$. We further add the restriction that R be upper triangular. This is not an identification assumption and serves to reduce the number of parameters.

The identification of the jump parameter Λ follows a similar argument. The jump intensity (see Assumption 1) is $\lambda_0 + tr[\Lambda X_t]$, with $tr[\Lambda X_t] = \Lambda_{11} X_{11,t} + (\Lambda_{12} + \Lambda_{21}) X_{12,t} + \Lambda_{22} X_{22,t}$. To identify the out-of diagonal elements of Λ, we choose Λ to be upper triangular, i.e. $\Lambda_{21} = 0$.

Chapter 3

The Price of the Smile and Variance Risk Premia

Understanding the properties of the market price of volatility risk is an important issue in financial economics. The recent literature has reached a consensus on the fact that unexpected shocks in aggregate future market uncertainty are priced in modern financial markets, by estimating a time-varying negative volatility risk premium for long variance swap positions or long volatility option portfolios. However, less is known about (i) which characteristics of volatility risk generate the premium, (ii) how the premium depends on the investment horizon (i.e., the term structure of volatility risk premia), and (iii) the relation between volatility risk premia, option risk premia and market risk premia. To address these questions, we specify a flexible three-factor volatility model with tractable option transforms and we estimate the joint dynamics of option-implied risks, option-implied risk premia and the term structure of variance risk premia.

Option-implied risks are the risks generated by unexpected variations of the option-implied volatility smile. Therefore, they can be summarized by the vector of state variables that control the conditional uncertainty about future volatility. Option-implied risk premia are the risk premia implicitly paid by investors for an option exposure to such risks. By construction, they span the risk premium of a shock in the price of volatility, across option moneyness and maturity, i.e., they capture the market price of the option-implied volatility smile. While the market price of the smile captures the risk premium demanded for an exposure to shocks in the option-implied price of volatility, the variance risk premium is the excess return required to trade realized variance risk. Investors can trade realized variance either by trading variance swaps in over-the-counter markets or by dynamically delta-hedging static option portfolios. Therefore, using information on the dynamics of the option-implied volatility smile and the returns of synthetic option portfolios, it is possible to identify the dynamics of option-implied risks, together with the term structures of option-implied risk premia and variance risk premia.

We identify option-implied risks and the market price of the smile using a novel parsimonious three-factor stochastic volatility model, in the class of matrix affine jump diffusion (AJD) proposed by Leippold and Trojani (2008). In contrast to benchmark models in the literature, our specification is characterized by (i) the presence of interdependent, potentially mutually-exciting, volatility risks, (ii) a skewness component disconnected from volatility and (iii) a compensation for variance risk that can vary independently of the level of the volatility. The first property enables a more direct identification and interpretation of hidden, potentially correlated and highly time-varying, volatility risks. The second property helps to generate a variability of the option-implied skewness more consistent with the data, because the link between option-implied volatility and skewness in our specification is more loose than in most benchmark volatility models. The third property allows the excess returns of volatility strategies to depend on option-implied skew dimensions that are partially disconnected from the volatility, making the term structure of variance risk premia partly disconnected from the level of the volatility.[1]

We estimate the dynamics of the option-implied volatility smile and the term structure of variance risk premia using a simple two-step procedure, which exploits the joint information of a panel of S&P500 option prices and a panel of excess returns of option volatility portfolios, in the sample period from January 1996 to January 2013. In the first step, we exploit the information from the panel of option prices to estimate the physical and risk neutral dynamics of the hidden volatility risks driving the S&P500 implied volatility surface, together with the risk-neutral properties of the jump component in returns. This allows us to identify the dynamics of option-implied risks and the market price of the smile. In the second step, we estimate the parameters of the jump volatility risk premium, from a simple arbitrage-free regression of the payoffs of option volatility portfolios on the option-implied risks estimated in the first step. This identification strategy allows us to isolate the joint dynamics of option-implied risks, option-implied risk premia and the term structure of variance risk premia, without relying on direct information about market excess returns.

We first find that our model produces an excellent pricing performance and fit, relative to other benchmark two- and three-factor models in the literature. For instance, the pricing improvements generated by our model relative to a benchmark two-factor Bates (2000)-type model, are between 20% and 30%, using different metrics of in- and out-of-sample pricing performance.[2] We document that such improvements in pricing

---

[1] Our model specification nests a number of important affine stochastic volatility models in the literature, such as Bates (2000) two-factor jump diffusion or Heston (1993)-type two-factor volatility models. These models have independent volatility components and a stochastic skewness that is linearly related to the volatility components.

[2] The improvement in in-sample (out-of-sample) fit, measured by the increase of the (pseudo) likelihood

performance follow from the failure of two-factor models to generate enough volatility-unrelated variability in option-implied skews. More importantly, our model yields a sharp identification of option-implied risk and risk-premium dynamics, in terms of a small set of three interdependent volatility components, which parsimoniously span the option-implied volatility surface, the market price of the smile and the term structure of variance risk premia. We find that these components capture risk and risk premium dynamics at three distinct frequencies, inducing different contributions to the term structure of option-implied risks and risk premia. The two components with the lowest and highest degree of persistence influence both the diffusive and the jump volatility of returns, while the third component only affects the jump volatility and also has a moderate degree of persistence.[3]

Second, we find that each of the volatility risks identified by our approach is linked to a natural interpretation, in terms of the relation with observable option-implied properties. The least persistent volatility risk has a weekly correlation of 0.84 with the 30-days at-the-money implied volatility. The second least persistent volatility risk has a weekly correlation of -0.89 with the 30 days option-implied skew. Therefore, it is naturally related to short-term option-implied skewness. More importantly, while this component is dynamically correlated with the 30 days option-implied volatility, an important fraction is orthogonal to it. We find that this fraction has a correlation of -0.86 with a 30-days option-implied skew residual, defined as the residual of a regression of the 30-days option-implied skew on the 30-days option-implied volatility. In this sense, the second least persistent component of the volatility correlates with short term option-implied skew dynamics unrelated to high-frequency implied-volatility shocks. The last, most persistent, volatility risk is also dynamically correlated with the 30 days option-implied volatility. However, the fraction of this risk that is orthogonal to the volatility has a weekly correlation of 0.83 with the residual of a regression of the option-implied skew term structure on the 30-days option-implied volatility. In this sense, it captures dynamics of the option-implied skew term structure that are unrelated to high-frequency option-implied volatility shocks. According to this evidence, we interpret the least persistent volatility risk as a high-frequency implied volatility factor. Similarly, we interpret the two more persistent volatility risks as mid-run and long-run option-implied skew factors partially disconnected from high-frequency implied volatility shocks.

Third, we find that the market price of the smile is completely spanned by the mid-run and the long-run volatility components alone. To illustrate, the monthly market price of the long-run volatility risk is proportional to the level of this risk, while the

---

function, is in the order of 4.5% (9.7%).

[3] To illustrate, the half-lifes of the first two components in our sample are 0.11 years and 1.25 years, respectively. The half-life of the third component is 0.28 years.

market price of the short-run volatility risk has a correlation of about 0.99 with the level of the mid-run risk.[4] These findings disclose our two option-implied skew components as key risk premium factors that parsimoniously capture the dynamics and the term structure of option risk premia. We find that the market price of the mid-run and the long-run volatility components typically has a downward sloping term structure, while the term structure of the market price of the short-run volatility risk is flat for horizons from 3 months on.

Fourth, we study the relation between volatility risks, option risk premia and the excess returns of popular volatility strategies. The time-variation of all volatility risks implies highly time-varying and unambiguously negative conditional variance risk premia, ranging between zero and -16 percent squared (-11 percent squared), on an annualized basis, for a monthly (an annual) investment horizon. Variance risk premia are largest (in absolute value) when the price of option-implied market insurance is large, typically during phases of financial distress and market turmoil, e.g., during the Asian and Russian crises in the late nineties, shortly before the collapse of the internet bubble in 2000, shortly after the Lehmann bankruptcy in September 2008 and the US downgrade in August 2011, and during the EU debt crisis. Consistent with intuition, we find that the largest fraction of variance risk premia is explained by a premium for jump variance risk. To illustrate, at short horizons of one month variance risk premia are virtually completely explained by a time-varying premium for pure jump variance risk, while at horizons of twelve months the premium for diffusive variance explains about one fourth of the total variance risk premium. This structure of the variance risk premium is naturally related to the option-implies risks and risk premia estimated in our model. At short horizons, the pure jump variance risk premium is completely explained by the level of the probability of a jump, which depends on all option-implied risks. In contrast, at longer horizons variance risk premia are almost completely explained by option-implied risk premia, which are spanned by mid-run and long-run volatility risks.

Fifth, the structure of volatility risks and risk premia in our model has sharp implications for the dynamics of the term structure of variance risk premia. We find that while the term structure of variance risk premia is most of the time downward sloping, reflecting a higher relative price of long-run option-implied market insurance, it can be strongly upward sloping in periods of market distress. The most prominent cases, in which we observe an inversion of the term structure of variance risk premia, arise immediately after both the Lehmann default in September 2008 and the US downgrade in August 2011, when the spread between annualized 12 month and 1 month variance risk premia has been as large as $+5.8\%$ squared and $+2\%$ squared, respectively. The

---

[4] Similarly, the monthly market price of the mid-run risk has a high correlation of 0.87 with its level and of 0.90 with the level of the long-run risk.

inversion of the term structure of variance risk premia is explained by the interplay of the term structures of option-implied risks and risk premia in our estimated model. While the first term structure is decreasing and implies the increasing term structure of pure jump variance risk premia, the second term structure is decreasing and implies a decreasing term structure of diffusive and jump intensity variance risk premia. During periods of market turmoil, the term structure of pure jump variance risk premia can become strongly upward sloping for short periods of time, when high-frequency volatility risks escalates, causing the inversion of the term structure of variance risk premia.

Finally, we investigate the predictive ability of option-implied volatility risks, both for market returns and for the realized returns of option volatility portfolios, based on forecasting horizons between one month and twelve months. Predictive regressions results suggest that option-implied risks have an economically relevant predictive power, both for S&P500 index and for volatility returns, with a dominating contribution to the predictive power deriving from mid-run and long-run volatility risks.[5] The affine specification of variance risk premia in our model is preferred by the out-of-sample predictability results, in which the largest degree of predictability is obtained for the model-implied variance risk premium predictions. In contrast, affine specifications of market risk premia are dominated by a simple nonlinear specification that allows the coefficients of the predictive relation to depend on the composition of the volatility. Consistent with a long-run risk explanation for market risk premia, we find that a simple trading strategy, which balances the option-implied premium for mid-run and long-run volatility risks with the frequency-composition of the volatility, produces economically large out-of-sample excess returns.

Review of the Literature. Our work draws from a large and important literature that has studied the economic sources of volatility variations, the dynamics of the option-implied price of volatility, the origins of a negative variance premium and the relation with market risk premia. We contribute to this literature along several dimensions.

First, we adopt a novel multivariate specification of stochastic volatility, which allows us to parsimoniously identify multi-frequency volatility risks, option risk premia and variance risk premia using three mutually exciting volatility components. Following Heston (1993)'s seminal single-factor volatility model, it has been early recognized that volatility is a multi-frequency phenomenon, potentially dependent on several risks with distinct persistence and variability properties. Bates (2000) was among the first to

---

[5] To illustrate, we find that the in-sample R^2 for future returns, using the two option-implied skew components alone as predictive variables, ranges between 5% and 25%, for prediction horizons between 1 and 12 months, with a peak at the quarterly forecasting horizon. The in-sample R^2 for the excess returns of option volatility portfolios are larger, ranging between 25% and 60%, with a peak at the 9 months forecasting horizon.

estimate with a panel of option prices a tractable two-factor model for index returns, in which the diffusive and the jump volatility are affine functions of two independent sources of risk having different persistence and variability. Some of the subsequent papers, such as Huang and Wu (2004) and Christoffersen et al. (2009), have quantified in more detail the improvements in the fit of the option-implied volatility smile provided by two-factor models with independent volatility components. The more recent literature has explored reacher multi-frequency specifications, in which volatility is driven by three sources of risk. Carr and Wu (2009b) estimate a three-factor model with independent components and a self-exciting jump volatility. Andersen, Fusari and Todorov (2015) specify a flexible three-factor model with self-exciting diffusive and jump volatilities that can jump simultaneously. Both approaches clearly improve on the fit of option-implied volatility dynamics provided by two-factor models. Our three-factor specification of stochastic volatility is different and complements these approaches, based on three mutually exciting volatility risks and risk premium factors that follow an affine jump diffusion on the state space of symmetric positive definite matrices.[6] This modeling approach allows us to obtain a direct and natural identification of potentially correlated option-implied state variables. At the same time, it produces a parsimonious specification of a multi-frequency dynamics for interdependent option-implied risks and risk premia, in which option-implied skew variations are note entirely spanned by volatility shocks.[7]

While the goal of our study is naturally related to Carr and Wu (2009b) and Andersen et al. (2015), it is also very different in several important dimensions. Carr and Wu (2009b) use information from a panel of option prices to identify in a semi-structural model the economic channels underlying the variations of equity volatility, in order to map them onto a leverage, a volatility feedback and a self-exciting component. Andersen et al. (2015) specify a very flexible option-pricing model to identify with a penalized nonlinear least-squares approach the option-implied risks revealed by a panel of S&P500 index options. They remain agnostic about the specification of option and volatility risk premia and study the role of option-implied risks as factor risk premia for the excess returns of index and volatility strategies. We start from a different modelling approach for the dynamics of option-implied risks and premia, in order to understand the implications for the dynamics of the term structures of option and variance risk premia. Given our need for a parsimonious multi-frequency specification of volatility risks and risk premia, we adopt a three-factor matrix jump diffusion with jumps in

---

[6] See, among others, Gourieroux (2006), da Fonseca et al. (2008) and Buraschi, Porchia and Trojani (2010) for examples and applications of affine matrix-valued diffusions, as well as Leippold and Trojani (2008) for a broad class of affine matrix jump diffusion processes.

[7] To illustrate, while the half-life of the most persistent option-implied risk estimated in our model is about one year and a quarter, the half-life of the least persistent option-implied risk is about three weeks.

index returns. While it would be in principle possible to introduce a jump component also in the volatility state variables, such an extended model would require in general a set of not innocuous additional assumptions, in order to preserve parameter parsimony and to identify the jump and the diffusion components in volatility risks and volatility risk premia. For instance, to identify the risk-neutral and the physical dynamics in their model, Carr and Wu (2009b) assume that the time-varying probability of a co-jump in returns and volatility follows a pure-jump single-factor dynamics. Such an assumption restricts the jump variance risk premia of different horizons to be perfectly correlated, which we feel excessively constrains the term structure of variance risk premia for our analysis.

Second, our paper borrows from a large literature that has studied the trading of realized nonlinear risks, the market price of volatility and the term structure of variance risk premia. In a first strand of this literature, Dupire (1993) and Neuberger (1994) where among the first to propose synthetic option portfolio strategies for trading proxies of realized variance, followed by Carr and Madan (1998), Demeterfi, Derman, Kamal and Zou (1999) and Britten-Jones and Neuberger (2000), among others. From the price of such portfolios, the price of volatility risk can be measured in a model-free way, giving rise to a variety of synthetic variance swap contracts; Carr and Lee (2009) provide an excellent review of this literature.[8] More recent papers have focused on the properties of different variance swaps in presence of jumps, as well as on the definition of swap contracts for trading higher-order risks linked, e.g., to skewness or kurtosis. Martin (2012), (Neuberger 2012) and (Bondarenko 2014) consider definitions of variance swap payoff that are robust to jumps. (Kozhan, Neuberger and Schneider 2010) analyze synthetic skew option portfolios and investigate the relation between skewness and variance risk premia, while (Schneider and Trojani 2014b) make use of Hellinger skew swaps to trade and price fear. (Schneider and Trojani 2014a) introduce a general class of divergence swaps for trading nonlinear risks and study in a model-free way the relation between the premia for divergence risks of different orders. Our paper draws from the insights produced in this literature, by identifying the term structure of variance risk premia with a two-step procedure that extracts risk premium information from a panel of returns of synthetic variance swaps. This approach allows us to exploit in a more comprehensive way the information generated by S&P500 index options for the term structures of option and variance risk premia, without relying on a complete specification of market risk

---

[8] Variance risk was made publicly tradable in 1993 with the introduction of VIX futures (now called VXO) by the CBOE. The definition of the VIX contract was revised in 2003, because hedging the old VXO contract turned out to be difficult, as pointed out also by Carr and Wu (2006). Jiang and Tian (2007) study the discretization errors implied by the VIX definition, when only a finite grid of option strikes is available, and propose an interpolation/extrapolation scheme for computing variance swap rates synthetically.

premia.

A second strand of this literature has established the existence of a negative risk premium for market volatility and has studied its properties. Bakshi and Kapadia (2003) provide first evidence on a negative variance risk premium using delta-hedge call option positions. Similar evidence is obtained by Wu (2011), using the payoffs of (over-the-counter) variance swaps, and by Carr and Wu (2009a), using synthetic variance swaps on several underlyings, who show that variance risk premia are significant and negative, both in bull and bear markets, and partly unspanned by the level of variance swap rates. Bondarenko (2014) also finds a negative and economically significant variance risk premium using synthetic option portfolios. Todorov (2010) and Bollerslev and Todorov (2011), among others, analyze the composition of variance risk premia, concluding that they are dominated by a premium for jump variance risk, which tends to increase after a negative jump has occurred. Our findings and model implications are consistent with the evidence provided by this literature, in particular the negativity of the premium for market volatility at different horizons and the dominating contribution of jump variance risk premia to the total variance risk premium. Additionally, our model implies an explicit decomposition of variance risk premia into the contribution of three option-implied risks with distinct persistence features. We find that short term variance risk premia are partly disconnected from option risk premia in periods of financial market distress, when they are significantly dependent on high-frequency option-implied volatility shocks and mid-frequency option-implied skewness shocks. In contrast, longer term variance risk premia are spanned by option risk premia and dominated by mid-term and long-term option-implied skewness shocks.

A third strand of this literature has studied the dynamics of the term structure of variance risk premia. Ait-Sahalia et al. (2012) and Filipovic, Gourier and Mancini (2015) estimate an affine and a quadratic two-factor volatility model, based on (over-the-counter) variance swap rates of maturities between two months and two years, focusing on the implications for the term structure of variance risk premia and for optimal portfolio choice with variance swaps, respectively. The first paper finds a negative and downward sloping term structure of variance risk premia. The term structure of jump variance risk premia is usually downward sloping, but it can be upward sloping during turbulent times, when the estimated contribution of the jump component becomes large at the short end of the term structure. The second paper shows that the optimal portfolio including variance swaps contains an economically relevant long-short position in long term versus short term variance swaps, which allows investors to earn the premium implied by a decreasing term structure of variance risk premia and to simultaneously hedge a short term increase in volatility. While we identify variance risk premia from S&P500 index option prices in centrally organized exchanges, our results can be useful for interpreting also the findings on (over-the-counter) variance swap risk premia.

Using our three-factor volatility model, we parsimoniously summarize the information embedded in S&P500 index option of maturities between one year and about two weeks, thus spanning also the short end of the option-implied volatility smile. In this way, we identify three option-implied state variables with very distinct implications for the term structures of option and variance risk premia. While we find a term structure of variance risk premia that is usually decreasing, we document that it can be strongly upward sloping in economically relevant periods of financial distress. The inversion of the curve is driven by the high-frequency component in the term structure of pure jump variance risk premia, which is strongly upward sloping when high-frequency option-implied volatility and skewness risks are large. In contrast, the long end of the term structure of variance risk premia is more directly related to the long end of the term structure of option-implied risk premia, which is dominated by the long-term components of the option-implied skewness.[9] The multi-frequency structure of variance risk premia in our model is consistent with a market price of volatility risk that is related at the short-end to high frequency shocks generated by situations of financial distress. Adrian and Rosenberg (2008) decompose market volatility into two weakly persistent components with a half-life of less than a quarter, which are priced in the cross-section of stock returns. They explicitly interpret their highest frequency volatility component as a proxy of skewness risk reflecting the tightness of financial constraints. Adrian and Shin (2010) show that expansions and contractions of repo and commercial paper funding predict variations in option-implied volatility, while Adrian, Moench and Shin (2013) document empirically the link between financial intermediaries balance sheets and asset prices. Muir (2013) emphasizes the high-frequency character of financial crises and explains in a theoretical model with financial intermediation why the term structure of the price of volatility and the term structure of variance risk premia are inverted in phases of financial turmoil. The dynamics of variance risk premia estimated by our model, in particular the high-frequency character of the inverted term structure of variance risk premia in periods of financial distress, is consistent with the economic intuition motivating this literature.

Finally, our paper is related to the literature studying in a predictive regression context the risk premium factors for market returns and their relation to variance risk premia. Bollerslev, Tauchen and Zhou (2009b) where the first to document the predictive power of variance risk premia, proxied by the difference of implied and realized volatilities, for future S&P500 index returns, finding predictive regression R^2s of about 7% for prediction horizons of about one quarter. Similarly, in an international context,

---

[9] Such a rich term structure dynamics has likely interesting implications also for optimal portfolio choice with variance swaps, as according to our findings shorting long term against short term variance swaps can be very expensive in periods of turbulent markets.

Bollerslev, Marrone, Xu and Zhou (2014) introduce a proxy for a global variance risk premium and obtain predictive regression R^2s of up to 12% for S&P500 index returns at horizons of 4 months. Bekaert and Hoerova (2014) decompose the VIX into an expected volatility and a risk premium component. They find that the latter predicts stock returns, that the expected volatility forecasts economic activity and that both are early indicators of financial instability. Chung, Tsai, Wang and Weng (2011) find no predictive power of the VIX volatility index alone for future S&P500 index returns, with an estimated predictive regression coefficient that changes sign in different sub-samples, while Johnson (2012) shows that the VIX term structure produces adjusted R^2 of about 5.2% at a 3 month horizon. Andersen, Fusari and Todorov (forthcoming) specify a flexible stochastic volatility model and estimate from a panel of S&P500 options three option-implied state variables, in order to study their relation to index and volatility risk premia. They find that a single option-implied skewness factor unspanned by volatility shocks has a large predictive power for variance and S&P500 index returns.

Our predictive regression results are consistent with the findings in this literature and provide a number of additional insights. First, we find that mid- and low-frequency option-implied premia for volatility jointly produce predictive power for future market returns. Second, such predictive power is virtually completely exhausted by mid- and low-frequency volatility shocks that are orthogonal to high-frequency volatility shocks. Third, mid-frequency and low-frequency option risk premia are naturally related to the risk premium factor for unspanned option-implied skewness in Andersen et al. (forthcoming), as they correlate quite exhaustively with the latter in our sample.[10] Fourth, our mid-frequency and low-frequency option risk premium factors are also naturally linked to observable option-implied skew components unspanned by volatility, as they largely correlate with two volatility-residuals of option-implied skew and option-implied skew term structure. Finally, while option-implied unspanned skewness is likely related to market risk premia, we find that its relation to market equity premia is possibly not affine and dependent on the frequency-composition of the volatility. This evidence supports identification procedures for the market price of the smile and the term structure of variance risk premia that do not depend on a complete affine specification of market excess returns.

The rest of the paper is organized as follows. Section 3.1 introduces our three-factor stochastic volatility model for the dynamics of S&P500 option-implied risks and risk premia. It then provides the closed-form expressions for the term structure of

---

[10] To illustrate, full sample regressions of the predictive factor for unspanned skewness in Andersen et al. (forthcoming) on our two option-implied risk premium factors produce significant predictive regression results, with predictive R^2's of about 95%. We are grateful to Nicola Fusari for having provided us with the time series of the option implied state variables estimated in Andersen et al. (forthcoming).

variance risk premia used in our two-step identification strategy. Section 3.2 presents our empirical findings. We first report results on in- and out-of-sample pricing accuracy and fit, and compare them with those of benchmark two- and three-factor models in the literature. In a second step, we analyze the properties of the option-implied risks estimated by our approach, we highlight their link to observable option-implied risk factors, and we disclose the relation between option-implied risks, the market price of the smile and the term structure of variance risk premia. Finally, we study the predictive ability of option-implied risks for future market returns and for the returns of popular volatility strategies. Section 3.3 concludes and highlights avenues of future research.

3.1 Model

Our model is characterized by the presence of (i) mutually exciting risks, (ii) a skewness component disconnected from volatility and (iii) a compensation for variance risk that can vary independently of the volatility. To link our approach to well-known benchmarks in the literature, we embed Bates (2000) two-factor jump diffusion in a more general state dynamics, within the class of matrix affine jump diffusions (AJD) proposed in Leippold and Trojani (2008). We first specify the index return dynamics under the risk-neutral measure. In a second step, we specify an affine market price of risk for the volatility components and for the jump risk in index returns. This characterizes the dynamics of option-implied risks, option-implied risk premia and the term structure of variance risk premia. We leave the specific form of the market price of diffusive return shocks unspecified, as it is unnecessary to identify the smile and variance risk premium dynamics in our two-step identification approach.

3.1.1 A Two-Component Benchmark Volatility Model

In Bates (2000) model, returns are driven by two independent volatility factors and follow a Poisson-Normal jump process with a stochastic intensity. Two different important channels generate a stochastic skewness of returns: The standard feed-back effect between returns and volatility and a time-varying probability of return jumps. We denote by S_t the value of an equity index at time t, by r and q the (constant) interest rate and dividend yield, and by v_{1t}, v_{2t} the two volatility components. Under the risk-neutral probability measure $\mathbb{Q}$, the return dynamics is:

$$\frac{dS_t}{S_{t-}} = (r - q - \lambda_t \bar{k})dt + \sqrt{v_{1t}}dz_{1t} + \sqrt{v_{2t}}dz_{2t} + kdN_t \ , \tag{3.1}$$

where z_1, z_2 are independent standard Brownian motions and the volatility components have the dynamics:

$$dv_{it} = (\alpha_i - \beta_i v_{it})\, dt + \sigma_i \sqrt{v_{it}} dw_{it} \quad ; \quad i = 1, 2 \quad , \tag{3.2}$$

where w_1 and w_2 are independent standard Brownian motions, having correlation ρ_1 and ρ_2 with z_1 and z_2, respectively. Return jumps kdN_t feature an affine jump intensity

$$\lambda_t := P_t(dN_t = 1)/dt = \lambda_0 + \lambda_1 v_{1t} + \lambda_2 v_{2t} \ , \tag{3.3}$$

and a jump size k with expected value $\bar{k} = E^{\mathbb{Q}}(k)$.[11] The well-known volatility feedback effect is captured by the (stochastic) correlation between returns and diffusive volatility $v_{1t} + v_{2t}$:

$$Corr_t \left(\frac{dS_t}{S_{t-}}, d(v_{1t} + v_{2t}) \right) = \frac{\rho_1 v_{1t} + \rho_2 v_{2t}}{\sqrt{(v_{1t} + v_{2t} + \lambda_t E^{\mathbb{Q}}(k^2))(\sigma_1^2 v_{1t} + \sigma_2^2 v_{2t})}} \quad , \tag{3.4}$$

where the first expression in the denominator, $V_t := v_{1t} + v_{2t} + \lambda_t E^{\mathbb{Q}}(k^2)$, is the total spot return variance. In addition to the volatility feedback effect, the time varying jump intensity (3.3) generates a second direct channel for a stochastic jump-driven return skewness. Moreover, distinct mean reversion speeds or volatilities of volatility associated with volatility components v_{1t} and v_{2t} generate a stochastic term structure of volatility, as the composition of the volatility varies over time.

Two features of Bates (2000) model are interesting to understand the motivation of our modeling approach. First, volatility components v_{1t}, v_{2t} are mutually independent. Therefore, they are difficult to identify and interpret directly, using observable, potentially correlated, option-implied components. Second, the jump intensity (3.3) and the volatility-feedback effect (3.4) are both functions of the volatility components v_{1t} and v_{2t} alone, meaning that shocks to risk neutral skewness are always correlated with a shock to volatility. In other words, a time-variation of the risk neutral skewness is always associated with a variation of the volatility. This feature produces a tight relation between volatility and option-implied risk-neutral skewness, which can be difficult to reconcile with the data. Figure 3.1 provides a simple illustration of this important aspect.

In Figure 3.1, top panel, we scatter plot two option-implied measures of short-term risk-neutral skewness and volatility term structure, respectively. We compute such option-implied measures both model-free (grey data points), based on the panel

---

[11] Different assumptions can be made on the risk-neutral distribution of log return jumps $\ln(1 + k)$. Bates (2000), e.g., assumes $\ln(1 + k) \sim N\left(\ln\left(1 + \bar{k}\right) - \frac{\delta^2}{2}, \delta^2\right)$. Alternative specifications include, e.g., a double exponential or similar distributions.

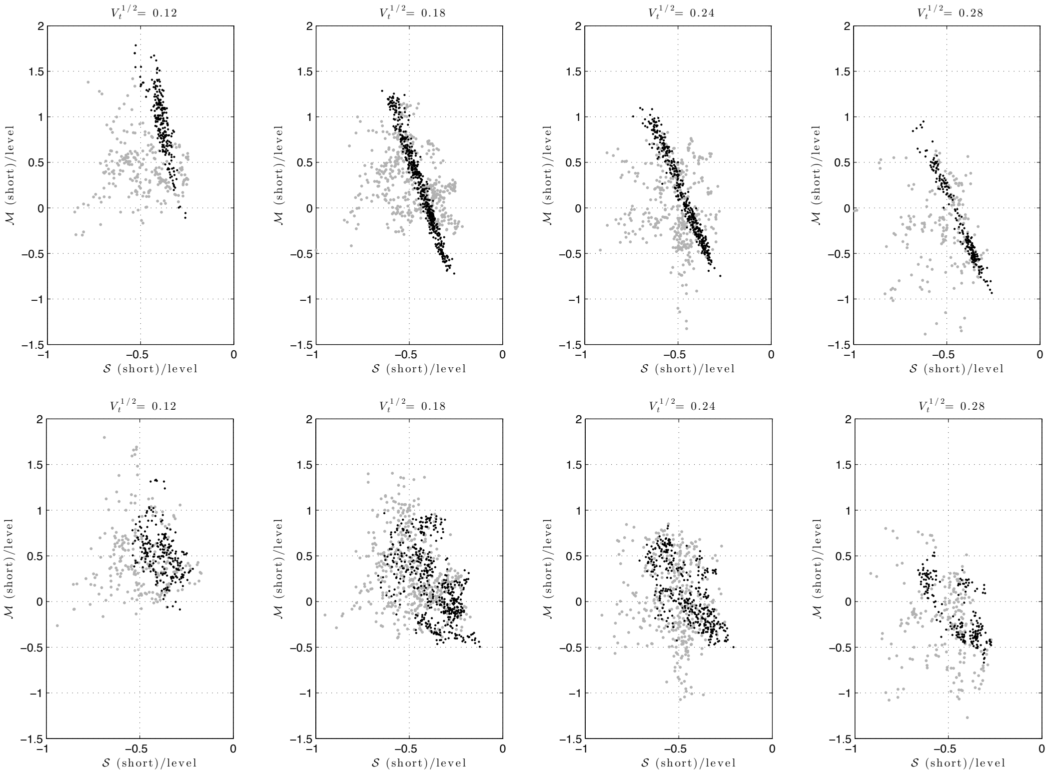


Figure 3.1: Option-implied short term (1 month) skew $\mathcal{S}_{short}$ versus term structure $\mathcal{M}$, both normalized by the volatility level. Grey dots: Data. Black dots – top panel: Fitted values of a two factor Bates model ($SVJ_{2,0}$). Black dots – bottom panel: Fitted values of our model (SVJ_{31}). We stratify by the short term at the money implied volatility at ±5% around the selected level, i.e., 17.1%–18.9% for the second panel. The exact calculation method for the option-implied skewness $\mathcal{S}$ and term structure $\mathcal{M}$ is explained in Appendix B.3.

of S&P500 index options in the time span from January 1996 to January 2013, and using the fitted parameters and volatility states of a two-factor Bates (2000) model (black data points). To isolate the effect of the volatility level on the option-implied skewness and option-implied volatility term structure, we scale each option-implied proxy by the level of the 30-days at-the-money implied volatility and we stratify the sample in four volatility bins, associated with an at-the-money volatility of 12%, 16%, 24% and 28%, respectively.

The evidence in Figure 3.1 highlights at least two interesting features. First, in each scatter plot of the top panel, we observe a large variability of model-free option-implied short-term skewness and option-implied volatility term structure. This is a direct indication of a conditional dynamics in the S&P500 option-implied volatility smile, which is partially disconnected from the level of the volatility. Second, the joint variability of option-implied skewness and option-implied volatility term structure produced by a two-factor Bates (2000) model yields a very tight, almost deterministic, relation between these two proxies of the smile. This is a simple indication of the fact that two-factor specifications of option-implied risks might excessively restrict the model-implied dynamics of the smile, relative to the evidence provided by S&P500 option data. This intuition is supported by the scatter plots in the bottom panel of Figure 3.1, which collect the evidence produced by a three-factor model with interdependent volatility risks and a skewness component disconnected from the volatility. Compared to the top panel, each scatter plot produces a more consistent evidence between model-implied and model-free findings, which better reproduces the loose link between option-implied skewness and option-implied volatility term structure in the data.[12]

3.1.2 Modelling Interdependent Risks and Skewness Components Disconnected from Volatility

Following the intuition provided by the previous section, we study a class of stochastic volatility models, in which risk-neutral skewness can feature a component disconnected from volatility, which interacts with all volatility components. We model this component using a third state variable v_{12t} and specify its interactions with v_{1t}, v_{2t} based on the dynamics of a 2×2 symmetric and positive definite matrix diffusion X_t, where:

$$X_t := \begin{pmatrix} X_{11t} & X_{12t} \\ X_{12t} & X_{22t} \end{pmatrix} = \begin{pmatrix} v_{1t} & v_{12t} \\ v_{12t} & v_{2t} \end{pmatrix} . \tag{3.5}$$

Overall, this means that we consider a family of three-factor stochastic volatility models with latent state $(X_{11t}, X_{22t}, X_{12t})$.

State Dynamics

Positive definiteness of X_t ensures positivity of X_{11t}, X_{22t}. Therefore, it is a natural choice to consider symmetric positive definite matrix processes.

Assumption 9. Symmetric positive definite matrix process X_t follows the affine dynamics

$$dX_t = [\beta Q'Q + MX_t + X_t M']dt + \sqrt{X_t}dB_t Q + Q'dB_t'\sqrt{X_t} \ , \tag{3.6}$$

where $\beta > 1$, M, Q are 2×2 parameter matrices and B is a 2×2 standard Brownian motion under the risk-neutral martingale measure $\mathbb{Q}$. $\sqrt{X_t}$ denotes the unique symmetric

---

[12] (Gruber, Tebaldi and Trojani 2010) discuss the role of dynamically interacting components of short-run volatility, long-run volatility and unspanned skewness for parsimoniously capturing the dynamics of the option-implied volatility surface.

square root of X_t.

Remark 10. X is the affine Wishart diffusion process first introduced by (Bru 1991).[13] When matrices M or Q are not diagonal, states $(X_{11t}, X_{22t}, X_{12t})$ are dynamically interconnected, because their drift and volatility functions depend on all state variables in equation (3.6). When matrices Q and M are diagonal, (X_{11t}, X_{22t}) is an autonomous Markov process with components distributed as independent Heston (1993)-type volatility models. Therefore, under these constraints the state dynamics in Bates (2000) option valuation model arises as a particular case of our setting.

Risk-Neutral Return Dynamics and Nested Models

Given the matrix state dynamics (3.6), we specify the risk-neutral return dynamics by the following matrix AJD process.

Assumption 11. Under the risk neutral probability measure $\mathbb{Q}$, the dynamics of S_t is given by:

$$\frac{dS_t}{S_{t-}} = (r - q - \lambda_t \bar{k})dt + tr(\sqrt{X_t} dZ_t) + k dN_t \ , \tag{3.7}$$

where X_t follows the dynamics (3.6),

$$Z_t = B_t R + W_t \sqrt{I_2 - RR'} \ , \tag{3.8}$$

with $tr(\cdot)$ denoting the trace operator, W another 2×2 standard Brownian motion, independent of B, and R a 2×2 matrix such that $I_2 - RR'$ is positive semi-definite. Return jumps follow a compound Poisson process kdN_t with jump intensity $\lambda_t = \lambda_0 + tr(\Lambda X_t)$, for $\lambda_0 \geq 0$, a 2×2 matrix Λ and an iid jump size k such that $\bar{k} = E^{\mathbb{Q}}[k]$. The distribution of log return jumps $J := ln(1 + k)$ is a double exponential with parameter $\lambda^+, \lambda^- > 0$ and density f given by:[14]

$$f(J) = \frac{\lambda^+ \lambda^-}{\lambda^+ + \lambda^-} \left[e^{-\lambda^- J^- - \lambda^+ J^+} \right] \ , \tag{3.9}$$

where $J^+ := \max(J, 0)$ and $J^- := \max(-J, 0)$ are the positive and negative parts of log return jumps.

Dynamics (3.7) incorporates a stochastic risk-neutral skewness disconnected from the

---

[13] Positive semi-definiteness (definiteness) of X_t follows if $\beta > 1$ $(\beta > 3)$, ensuring that the volatility components cannot cross (reach) the zero boundary.

[14] We adopt a double exponential distribution for risk-neutral log return jumps because of its parsimony and flexibility. We have estimated our model also using a normal distribution, as in Bates (2000), and we have obtained very similar results to those reported in the paper.

diffusive volatility. To see this, note that:

$$Var_t\left(\frac{dS_t}{S_{t-}}\right) = tr(X_t) + \lambda_t E(k^2) = X_{11t} + X_{22t} + \lambda_t E(k^2) \ , \tag{3.10}$$

i.e., the diffusive variance $tr(X_t) = X_{11t} + X_{22t}$ does not depend on X_{12t}. However, X_{12t} impacts the jump-driven volatility and skewness, because λ_t is a function of X_{12t}, when Λ is not diagonal. At the same time, X_{12t} also influences the volatility feed-back effect, because:

$$Cov_t\left(\frac{dS_t}{S_{t-}}, d(X_{11t} + X_{22t})\right) = 2tr(R'QX_t) \ . \tag{3.11}$$

This shows that whenever matrices RQ' and Λ are not diagonal, component X_{12t} directly influences the jump-driven volatility, the jump-driven skewness and the diffusive skewness. At the same time, X_{12t} is disconnected from the diffusive volatility.

Remark 12. When matrices R and Q are diagonal, equation (3.11) collapses to Bates (2000) specification of volatility feedbacks. Similarly, when matrix Λ is diagonal, λ_t coincides with Bates (2000) specification of a stochastic intensity. Thus, when M, Q, R and Λ are all diagonal, Assumption 11 yields Bates (2000) two-factor volatility model.[15] If, in addition, $\lambda_0 = 0$ and $\Lambda = 0$, a two factor Heston (1993)-type volatility model is obtained, which has been recently studied empirically in Christoffersen et al. (2009). All these diagonal models feature independent volatility components and a return skewness fully connected to volatility.

Assumption 11 provides a convenient framework for studying the pricing accuracy and the dynamic fit of models with interdependent components and a risk-neutral skewness disconnected from volatility. Table 2.1 provides an overview of benchmark single-factor, two-factor and three-factor models related to Assumption 11. We denote by SV_{rq} diffusion and by SVJ_{rq} jump diffusion models, according to the numbers r and q of state variables and skewness components disconnected from volatility, respectively. For comparison, we also report the total number of parameters necessary for a complete specification of the risk-neutral and the physical dynamics in our two-step estimation approach.

Note that single- and two-factor Bates (2000)-type models are all nested by Assumption 11. The single jump diffusion model not nested by our framework in Table 2.1 is the three-factor Bates (2000)-type model. Table 2.1 also provides useful information about the parsimony of our framework for specifying a model with an additional state vari-

---

[15] In order to nest two-factor Bates (2000)- and Heston (1993)-type models in our option pricing approach, we specify β to be a diagonal matrix B when both Q and M are diagonal matrices.

able driving the option-implied smile, relative to benchmark two-factor Heston (1993)- and Bates (2000)-type models. Indeed, while our three-factor matrix jump-diffusion has three additional parameters relative to two-factor Bates (2000)-type models, a three-factor Bates (2000)-type model has as many as seven additional parameters. Such a larger number of additional parameters can generate additional estimation and identification challenges, when estimating the joint dynamics of option-implied risks and option-implied risk premia from a panel of option price observations.[16]

Option Valuation

Assumption 11 and Assumption 16 yield closed-form risk-neutral transforms in our matrix AJD setting, which are useful to compute the prices of plain vanilla options with transform methods; see also, e.g., (Carr and Madan 1999) and (Duffie et al. 2000), among others. Following (Leippold and Trojani 2008), the exponentially affine conditional Laplace transform for $Y_T := \log(S_T)$ is given by:

$$\Psi(\tau;\gamma) := E_t\left[\exp\left(\gamma Y_T\right)\right] = \exp\Big(\gamma Y_t + tr\big[A(\tau)X_t\big] + B(\tau)\Big) \,, \tag{3.12}$$

where $\tau = T - t$, $A(\tau) = C_{22}(\tau)^{-1}C_{21}(\tau)$ and the 2×2 matrices $C_{ij}(\tau)$ are given in closed form in Section B.1.1 of the Appendix.

Remark 13. In contrast, e.g., to Bates (2000)-type models, the computation of the risk neutral transform cannot be reduced to calculations that involve only scalar exponential and logarithmic functions, because $C_{ij}(\tau)$ and $B(\tau)$ are functions of a matrix exponential and a matrix logarithm, respectively. This feature makes the computation of Laplace transform (3.12) typically at least one order of magnitude more computationally intensive. Furthermore, (3.12) contains a matrix logarithm, which is a multivalued function. We use the methods described in Gruber (2015) to control discontinuities of the complex matrix logarithm and to mitigate the computational cost of evaluating the Laplace transform.

Affine Market Price of Risk and Physical Dynamics

Following Leippold and Trojani (2008), we can price the different risks in our economy with a stochastic discount factor, linked to a market price of (Brownian) shocks B in Assumption 11, which preserves an affine dynamics under physical probability measure $\mathbb{P}$.

---

[16] When focusing on the risk-neutral distribution $\mathbb{Q}$, the number of risk neutral parameters in a two-factor Bates (2000)-type model is 13, while in a three-factor Bates (2000)-type model it is 18. In our matrix AJD model, the total number of risk-neutral parameters is 16.

Assumption 14. The change of measure from the physical probability $\mathbb{P}$ to the risk neutral probability $\mathbb{Q}$ is such that:

$$dB^* = dB - \left(\sqrt{X_t}\Gamma + \frac{1}{2\sqrt{X_t}}(\beta^* - \beta)Q'\right) dt \ , \tag{3.13}$$

where either $\beta^* > 3$ or $\beta^* = \beta$, Γ is a 2×2 parameter matrix and B^* is 2×2 standard Brownian motion under the physical probability measure.[17]

Remark 15. Under Assumption 14, the dynamics of process X_t with respect to the physical probability $\mathbb{P}$ is given by:

$$dX_t = [\beta^* Q'Q + M^* X_t + X_t(M^*)']dt + \sqrt{X_t} dB_t^* Q + Q' dB_t^{*\prime} \sqrt{X_t} \ , \tag{3.14}$$

where

$$M^* = M + \Gamma Q \ . \tag{3.15}$$

When $\beta \neq \beta^*$, the condition $\beta^* > 3$ implies that process X is positive definite under probability $\mathbb{P}$. This feature ensures a well-defined change of probability measure, from the physical probability $\mathbb{P}$ to the risk neutral probability $\mathbb{Q}$, associated with the market prices of risk for B^*-Brownian shocks in equation (3.13). In all other cases, we require $\beta^* = \beta$.[18] For the case where $\beta = \beta^*$, we obtain a completely affine market price of risk for our matrix AJD. Whenever $\beta \neq \beta^*$, we obtain an extended affine specification of the market price of risk that allows the market price of the volatility factors X_{11t} and X_{22_t} to be inversely related to the (diffusive) volatility. This feature can produce a large conditional market price of volatility when volatility is low. (Cheridito, Filipovic and Kimmel 2007) propose a class of yield curve models with an extended affine market price of risk, in order to improve the specification of bond excess returns.

A useful feature of the market price of risk specification in Assumption 14 is that the market price of a shock in the matrix AJD dynamics (3.14) can depend on the other option-implied components $X_{11t}, X_{22t}, X_{12t}$. This feature implies a market price of a dynamic smile with extensive interconnections, whenever matrix ΓQ is not diagonal. This is easily seen, e.g., from the difference of the $\mathbb{P}$ and $\mathbb{Q}$ expectation of a shock in state X_t, which captures the instantaneous risk premium of each option-implied component in our model:

$$\frac{1}{dt}(E^{\mathbb{P}} - E^{\mathbb{Q}})[dX_t] = (\beta^* - \beta)Q'Q + \Gamma Q X_t + X_t Q' \Gamma' \ . \tag{3.16}$$

---

[17] $1/\sqrt{X_t}$ denotes the unique inverse square root of positive definite matrix X_t.

[18] Empirically, our model estimations suggest that $\beta^* < 3$, which effectively reduces the number of parameters that need to be estimated.

This risk premium is state dependent whenever $M^* - M = \Gamma Q \neq 0$. If matrix $M^* - M$ is diagonal with diagonal components D_1, D_2, risk premia between option-implied components are disconnected:

$$\frac{1}{dt}(E^{\mathbb{P}} - E^{\mathbb{Q}}) \left[d \begin{pmatrix} X_{11t} & X_{12t} \\ X_{12t} & X_{22t} \end{pmatrix} \right] = (\beta^* - \beta)Q'Q + \begin{pmatrix} 2D_1 X_{11t} & (D_1 + D_2)X_{12t} \\ (D_1 + D_2)X_{12t} & 2D_2 X_{22t} \end{pmatrix},$$

because the risk premium of each option-implied component is proportional to the level of the component itself. This is the situation emerging, e.g., in Bates (2000)-type models. Whenever matrix $M^* - M =$ is not diagonal, the risk premium of the diagonal option-implied component X_{11t} (X_{22t}) is an affine function of both X_{11t} (X_{22t}) and X_{12t}, while the risk premium of the out-of-diagonal component X_{12t} is an affine function of all states X_{11t}, X_{22t}, X_{12t}. In this case, the compensation for diffusive volatility risk in the model can vary in a way partly disconnected from the diffusive volatility.

To identify the physical and the risk neutral parameters of the state dynamics for X in Assumptions 11 and 14, together with the risk-neutral jump intensity and the risk-neutral distribution of return jumps, we rely on a panel of observations of S&P500 option prices, in the sample period January 1996 to January 2013. This is the first step in our two-step identification procedure of the market price of the smile and the term structure of variance risk premia.

Stochastic Discount Factor

In our model, three types of shocks can be priced: (i) diffusive shocks to index returns, (ii) shocks to the state variable X_t in the smile dynamics and (iii) jump-type shocks in index returns. In Assumption 11, these sources of risk correspond to the (risk-neutral) Brownian shocks dW_t, dB_t and the (risk-neutral) compound poisson shock $(e^J - 1)dN_t$, respectively.

Given the incompleteness of our framework, a multiplicity of stochastic discount factors exists. Existence of a well-defined stochastic discount factor to price all shocks in our model is ensured by a proper density for an equivalent change of measure, from the physical to the risk neutral probability. Given suitable matrix processes $\{\Gamma_{1t}\}$, $\{\Gamma_{2t}\}$ for the market prices of Brownian shocks dW_t^*, dB_t^*, and our double-exponential specification for the distribution of log return jumps, such a density can take the form:

$$\begin{aligned} \left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_T} &= \exp\left\{ tr\left(-\int_0^T \Gamma_{1t} dW_t^* + \frac{1}{2}\int_0^T \Gamma_{1t}'\Gamma_{1t} dt - \int_0^T \Gamma_{2t} dB_t^* + \frac{1}{2}\int_0^T \Gamma_{2t}'\Gamma_{2t} dt \right) \right\} dt \\ &\quad \times \prod_{i=1}^{N_T^*} \exp\left\{ -(\lambda^- - \lambda^{*-})J_i^{*-} - (\lambda^+ - \lambda^{*+})J_i^{*+} + \ln\left(\frac{1/\lambda^{*-} + 1/\lambda^{*+}}{1/\lambda^- + 1/\lambda^+} \right) \right\}, \qquad (3.17) \end{aligned}$$

where the second line of the equality defines the change of measure for return jumps.

This choice of the change of measure implies a double exponential density (3.9) for return jumps, having parameters λ^{*+}, λ^{*-} and λ^{+}, λ^{-} with respect to the physical and the risk neutral distribution, respectively. Assumption 14 states:

$$\Gamma_{2t} = \sqrt{X_t}\Gamma + \frac{1}{2\sqrt{X_t}}(\beta^* - \beta)Q' \ . \tag{3.18}$$

Together with Assumption 9, this assumption ensures a well-defined change of probability measure for B^*-shocks in our model. It is straightforward to introduce a well-defined market price of risk also for $W-$shocks. For instance, under Assumptions 9 and 14, a choice:

$$\Gamma_{1t} = \sqrt{X_t}\Delta + \frac{\mu_0 - (r-q)}{\sqrt{X_t}} \ , \tag{3.19}$$

where Δ is a 2×2 parameter matrix and $\mu_0 > r - q$ a scalar parameter, implies a well-defined change of measure with affine dynamics for index returns, both under the physical and the risk-neutral probability measures.[19] However, this last assumption is not necessary for the validity of our identification of the market price of the smile and the term structure of variance risk premia, which is robust with respect to the particular form of the market price of risk Γ_{1t} of W^*-shocks.

Term Structure of Variance Risk Premia

We characterize the risk premia of VIX-type swap contracts. Since VIX-type swap payoffs can be synthesized using a dynamically delta hedged static option portfolio, these are natural contracts for studying the relation between the market price of variance risk and option-implied risk premia in our context. The flexible leg $RV_{t+\tau}(\tau)$ of a VIX contract is proportional to the delta-hedged payoff of a log contract:

$$\begin{aligned} RV_{t+\tau}(\tau) \ &:= \ \frac{2}{\tau}\left[-\ln(S_{t+\tau}/S_t) + \int_t^{t+\tau} dS_s/S_{s-}\right] && (3.20) \\ &= \ \frac{1}{\tau}\int_t^{t+\tau} \frac{1}{S_s^2} d[S,S]_s^c + \frac{2}{\tau}\sum_{t\le s\le t+\tau} \mathcal{E}(S_s/S_{s-}) \ , && (3.21) \end{aligned}$$

with $[S,S]_s^c$ the index continuous quadratic variation at time s and the Itakura-Saito realized divergence $\mathcal{E}(S_s/S_{s-}) := -\ln(S_s/S_{s-}) + S_s/S_{s-} - 1$ of a jump in index returns at time s.[20] By definition, the (VIX) variance risk premium is the difference of the $\mathbb{P}$ and

---

[19] Appendix B.1.3 provides a proof that under Assumptions 9 and 14 density process $\{\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_T}\}_{T\ge 0}$ defined by equations (3.17)–(3.19) is a proper martingale.

[20] See, e.g., (Schneider and Trojani 2014a).

$\mathbb{Q}$ expectations of the flexible leg in the VIX contract: $VRP_t(\tau) := (E^{\mathbb{P}} - E^{\mathbb{Q}})[RV_{t+\tau}(\tau)]$. Applying Assumptions 9 and 14 to equation (3.21), we obtain:

$$VRP_t(\tau) = tr\left((E_t^{\mathbb{P}} - E_t^{\mathbb{Q}})\left[\frac{1}{\tau}\int_t^{t+\tau} X_s ds\right]\right) + \frac{2}{\tau}\sum_{t\leq s\leq t+\tau}(E_t^{\mathbb{P}} - E_t^{\mathbb{Q}})\left[\mathcal{E}\left(\frac{S_s}{S_{s-}}\right)\right] .$$

The first term on the right hand side of this equality, which is affine in X_t, is the variance risk premium contribution deriving from continuous index shocks. The second term is the contribution deriving from jumps in returns. To preserve a tractable affine form for the second term, we can specify an affine physical jump intensity for return jumps: $\lambda_t^* = \lambda_0^* + tr(\Lambda^* X_t)$ say, for a scalar $\lambda_0^* \geq 0$ and a 2×2 parameter matrix Λ^*. In the most general case, such an intensity specification introduces five additional parameters.[21] A more parsimonious specification, with a single additional parameter, can assume a physical intensity proportional to the risk-neutral intensity. In our empirical analysis, we have investigated these different specifications, finding no incremental out-of-sample explanatory power for variance risk premia in the more general specification. Therefore, we rely on the most parsimonious specification.

Assumption 16. The intensities of return jumps are identical under probabilities $\mathbb{P}$ and $\mathbb{Q}$. We denote the ratio of the expected realized jump entropy under probabilities $\mathbb{P}$ and $\mathbb{Q}$ by $\beta_\Lambda^* = \mathbb{E}^{\mathbb{P}}[\mathcal{E}(1+k^*)]/\mathbb{E}^{\mathbb{Q}}[\mathcal{E}(1+k)] > 0$.

Remark 17. (i) Since quantities $\mathbb{E}^{\mathbb{P}}[\mathcal{E}(1+k^*)]$ and $\mathbb{E}^{\mathbb{Q}}[\mathcal{E}(1+k)]$ cannot be identified from the term structure of variance risk premia alone, Assumption 16 is equivalent to the assumption of proportional intensities under the physical and the risk-neutral probability measures. (ii) Assumption 16 yields a jump contribution to variance risk premia given by:

$$\frac{2}{\tau}\sum_{t\leq s\leq t+\tau}(E_t^{\mathbb{P}} - E_t^{\mathbb{Q}})\left[\mathcal{E}\left(\frac{S_s}{S_{s-}}\right)\right] = 2E^{\mathbb{Q}}[\mathcal{E}(1+k)]tr\left(\Lambda(\beta_\Lambda^* E_t^{\mathbb{P}} - E_t^{\mathbb{Q}})\left[\frac{1}{\tau}\int_t^{t+\tau} X_s ds\right]\right) .$$

Therefore, in our model the variance risk premium is a linear function of the expected average integrated state X, under the physical and the risk neutral probabilities.

Given the closed-form expressions for the expected average integrated state X in our setting, both under the risk-neutral and the physical probability, we obtain the following closed-form term structure of variance risk premia.

---

[21] Four parameters for the physical intensity process and one for the physical second moment of return jumps.

Proposition 2. Given Assumptions 11, 14 and 16, we can decompose for any $\tau > 0$ the variance risk premium in the contribution of diffusive and jump variance:

$$VRP_t(\tau) = VRP_t^c(\tau) + VRP_t^d(\tau) \ . \tag{3.22}$$

Each contribution is given explicitly by

$$\begin{aligned} VRP_t^c(\tau) &= tr[X_\infty^{\mathbb{P}} - X_\infty^{\mathbb{Q}} + A_\tau^{\mathbb{P}}(X_t - X_\infty^{\mathbb{P}}) - A_\tau^{\mathbb{Q}}(X_t - X_\infty^{\mathbb{Q}})], \\ VRP_t^d(\tau) &= 2E^{\mathbb{Q}}[\mathcal{E}(1+k)]tr[\Lambda(\beta_\Lambda^* X_\infty^{\mathbb{P}} - X_\infty^{\mathbb{Q}} + \beta_\Lambda^* A_\tau^{\mathbb{P}}(X_t - X_\infty^{\mathbb{P}}) - A_\tau^{\mathbb{Q}}(X_t - X_\infty^{\mathbb{Q}}))], \end{aligned}$$

where, for any 2×2 matrix H:

$$A_\tau^{\mathbb{Q}}(H) := \frac{1}{\tau}\int_0^\tau e^{Mu} H e^{M'u} du \ ; \ A_\tau^{\mathbb{P}}(H) := \frac{1}{\tau}\int_0^\tau e^{M^* u} H e^{M^{*\prime} u} du \ ,$$

and 2×2 matrices $X_\infty^{\mathbb{Q}}$, $X_\infty^{\mathbb{P}}$ are such that:

$$\beta Q'Q = X_\infty^{\mathbb{Q}} M + M' X_\infty^{\mathbb{Q}} \ ; \ \beta^* Q'Q = X_\infty^{\mathbb{P}} M^* + M^{*\prime} X_\infty^{\mathbb{P}} \ . \tag{3.23}$$

Remark 18. (i) The variance risk premium in Proposition 2 is affine in state X_t because functions $A_\tau^{\mathbb{Q}}(\cdot)$ and $A_\tau^{\mathbb{P}}(\cdot)$ are linear. (ii) With the exception of parameter β_Λ^*, all variables and parameters in the expression for $VRP_t(\tau)$ are identifiable from a first-step estimation, using exclusively information from a panel of S&P500 options. (iii) It is easy to show that whenever matrices Q, M and M^* are diagonal, the continuous variance risk premium $VRP_t^c(\tau)$ is a function only of the diffusive variance states X_{11t}, X_{22t}.[22] Whenever the intensity matrix Λ is diagonal, then also the jump variance risk premium $VRP_t^d(\tau)$ only depends on the diffusive variance states, inducing a perfect correlation between shocks to variance risk premia and shocks to the diffusive variance. This is the situation emerging in, e.g., Bates (2000)-type models. More generally, continuous and jump variance risk premia in our model can depend on all option-implied components of the smile, so that shocks in variance risk premia can be partially disconnected from shocks in the diffusive variance. This feature follows from the direct exposure of future VIX contract payoffs to future option-implied risks.

Jump variance risk premia directly depend on both the intensity and the distribution of return jumps. Since expected future jump intensities depend on all option-implied components X_{11}, X_{22}, X_{12} in our model, whenever matrix Λ is not diagonal, $VRP_t^d(\tau)$ is the sum of two economically distinct risk premium components. The first component,

---

[22] This follows from the fact that, in this case, the expected integrated diffusive variances $E_t^{\mathbb{P}}[\int_t^{t+\tau} tr(X_s)ds]$ and $E_t^{\mathbb{Q}}[\int_t^{t+\tau} tr(X_s)ds]$ only depend on X_{11t}, X_{22t}.

$VRP_t^{dc}(\tau)$, captures priced jump intensity risk and is spanned by the option-implied risk premia. The second component, $VRP_t^{dj}(\tau)$, captures priced pure jump variance risk consistently with stochastic discount factor (3.17), and is spanned by the expected jump intensity.[23] The closed-form expression for the pure jump variance risk premium is:

$$VRP_t^{dj}(\tau) = 2(\beta_\Lambda^* - 1)E^{\mathbb{Q}}[\mathcal{E}(1+k)]tr[\Lambda(X_\infty^{\mathbb{P}} + A_\tau^{\mathbb{P}}(X_t - X_\infty^{\mathbb{P}}))] \ . \tag{3.24}$$

We make use of equation (3.24) to identify the dynamics and the term structure of pure jump variance risk premia in our empirical analysis.

Model-Free VIX Payoffs and Variance Risk Premia

Denoting by F_t the S&P500 index futures price for residual maturity $\tau - t$, we can compute in a model-free way the payoff of a VIX swap contract, as the delta-hedged excess payoff of a static option portfolio:

$$\begin{aligned} RV_{t+\tau}^e(\tau) := RV_{t+\tau}(\tau) - E_t^{\mathbb{Q}}[RV_{t,t+\tau}] \quad = \quad & \frac{2}{\tau}\left[\int_0^\infty \frac{O_{t+\tau}(K)}{K^2}dK + \int_t^{t+\tau}\left(\frac{1}{F_{s-}} - \frac{1}{F_t}\right)dF_s\right] \\ & -\frac{2}{\tau}\int_0^\infty \frac{E_t^{\mathbb{Q}}[O_{t+\tau}(K)]}{K^2}dK \ , \end{aligned} \tag{3.25}$$

where for any $K < F_t$ ($K \geq F_t$) quantity $O_{t+\tau}(K) := (K - F_{t+\tau})^+$ ($O_{t+\tau}(K) := (F_{t+\tau} - K)^+$) is the terminal payoff of an out-of-the-money European put (call) option on index futures, with residual maturity τ and strike price K. In our empirical analysis, we compute $RV_{t+\tau}^e(\tau)$ in a model-free way, using the panel of S&P500 options and a time-series of high-frequency S&P500 index futures prices. This feature motivates our two-step estimation approach, in which we estimate parameter β_Λ^* from a simple linear regression of VIX option portfolio payoffs on the model-implied variance risk premia of Proposition 2. This approach is summarized by the next proposition.

Proposition 3. For any $\tau > 0$, define the following variables:

$$\begin{aligned} Y_{t+\tau}(\tau) \quad := \quad & RV_{t+\tau}^e(\tau) - tr((I + 2E^{\mathbb{Q}}[\mathcal{E}(1+k)]\Lambda)(X_\infty^{\mathbb{Q}} + A_\tau^{\mathbb{Q}}(X_t - X_\infty^{\mathbb{Q}}))) \\ & + tr(X_\infty^{\mathbb{P}} + A_\tau^{\mathbb{P}}(X_t - X_\infty^{\mathbb{P}})) \ , \end{aligned} \tag{3.26}$$

and

$$U_t(\tau) := tr(E^{\mathbb{Q}}[\mathcal{E}(1+k)]\Lambda(X_\infty^{\mathbb{Q}} + A_\tau^{\mathbb{Q}}(X_t - X_\infty^{\mathbb{Q}}))) \ . \tag{3.27}$$

Given a set of maturities $0 < \tau_1 < \ldots < \tau_n$, the following is a proper linear regression

---

[23] By definition, $VRP_t^{dj}(\tau)$ captures pure jump variance risk premia and is non zero whenever jump risk is priced, also when option-implied risks are not priced.

model,

$$\begin{pmatrix} Y_{t+\tau_1}(\tau_1) \\ \vdots \\ Y_{t+\tau_n}(\tau_n) \end{pmatrix} = \begin{pmatrix} U_t(\tau_1) \\ \vdots \\ U_t(\tau_n) \end{pmatrix} \beta^*_\Lambda + \begin{pmatrix} \eta_{t+\tau_1}(\tau_1) \\ \vdots \\ \eta_{t+\tau_n}(\tau_n) \end{pmatrix} , \tag{3.28}$$

with an error term $\eta_{t+\tau}(\tau) := (\eta_{t+\tau_1}(\tau_1), \ldots, \eta_{t+\tau_n}(\tau_n))'$ such that $E_t^{\mathbb{P}}[\eta_{t+\tau}(\tau)] = 0$.

In equation (3.28) of Proposition 3, all quantities are computable from our first-step estimation, using estimated model parameters $\hat{M}, \hat{Q}, \hat{M}^*, \hat{\Lambda}, \hat{\lambda}^+, \hat{\lambda}^-, \hat{\beta}, \hat{\beta}^*$ and filtered states $\{\hat{X}_t\}$. This is why we can estimate parameter β^*_Λ in the linear regression (3.28) separately from all other parameters and the hidden states for process $\{X_t\}$.

3.2 Empirical Analysis

3.2.1 Data and Estimation

We collect from OptionMetrics daily data of end-of-day prices of S&P500 index options, traded at the Chicago Board Options Exchange, for the sample period from January 1996 to January 2013 and maturities up to one year.[24] The sample consists of 4298 trading days, which we reduce to 883 weekly observations (each Wednesday). In order to allow for an out-of sample evaluation of our model, we further split these 883 observations into an in-sample period (from January 1996 to December 2002) with 359 observations and an out-of sample period (from January 2003 to January 2013) with 524 observations.

We apply a number of standard filtering procedures outlined, e.g., in Bakshi et al. (1997). First, we eliminate options with midquote premia below 0.375 dollars and options with zero bid price or with bid price larger than the ask price. Second, we eliminate options with stale quotes (i.e., prices unchanged from the previous trading day), observations that violate arbitrage bounds, duplicate entries and options where the bid-ask spread is smaller than the minimum tick size (i.e., five cents for options having prices below 3 dollars and ten cents for all other options). Third, we drop options with a time to maturity less than 10 days, in order to avoid pricing effects largely driven by short term liquidity features.

For our first-step estimation of the model parameters, we make use of all options with a Black-Scholes delta between 0.1 and 0.9. On average, this gives about 139 option prices per trading day, having an average time to maturity of 130 days and an average

---

[24] We obtain end-of-day midquotes as simple averages of end-of-day bid and ask option prices and force the put-call parity to hold when calculating the implied dividend yields.

Panel A: Summary statistics of the data

	In-sample	Out-of sample	Total
Time frame	1996-2002	2003-01/2013	1996-01/2013
Sampling frequency		weekly	
Trading days T	359	524	883
Number of observations	37'499	85'237	122'736
Average time to maturity (days)	141.5	124.9	130.0
Average moneyness (S/K)	0.99	0.98	0.99

Panel B: Number of observations by duration and delta

	$\tau < 30$	$30 < \tau < 75$	$75 < \tau < 180$	$180 < \tau$	all
$\lvert\Delta\rvert < 0.2$	1'761	4'647	3'679	3'858	13'945
$0.2 < \lvert\Delta\rvert < 0.4$	2'576	7'460	6'369	6'769	2'3174
$0.4 < \lvert\Delta\rvert < 0.6$	2'575	8'258	7'303	7'586	25'722
$0.6 < \lvert\Delta\rvert < 0.8$	3'479	10'808	9'399	10'446	34'132
$0.8 < \lvert\Delta\rvert$	2'981	8'651	6'947	7'184	25'763
all	13'372	39'824	33'697	35'843	122'736

Table 3.1: Main characteristics of our S&P500 option panel. We use out-of the money calls and puts.

moneyness $S/K = 0.99$. Table 3.1 presents a summary of the main characteristics of our option data set.[25]

For the calculation of the model-free VIX variance payoffs in equation (3.25), we make use of options for all available strikes. The delta hedging component in the VIX variance payoff is computed using tick-by-tick data for the S&P500 future traded at the CBOE, obtained from tickdata.com and sampled at 60 second intervals.

In the first step of our estimation procedure, we use the panel of S&P500 in-sample observations of option prices to estimate the structural model parameters, together with the time series of option-implied states X_{11t}, X_{22t} and X_{12t}. The time series of option-implied states uncovers their distinct roles as drivers of option-implied risks and risk premia over time. The parameter estimates shed light on the dynamic interactions between state variables driving option-implied risks and risk premia.

We estimate the model parameters Q, M, M^{*}, R, λ_0, Λ, λ^{+}, λ^{-}, β, β^{*}, by max-

---

[25] The interest rate r is computed by linearly interpolating the US treasuries yield curve supplied by OptionMetrics. The dividend yield q is computed by minimizing each day the put-call parity error of nearly at the money options (i.e., such that $0.9 \leq K/S \leq 1.1$): $q = \arg\min_{q} \left(C - P - Se^{-\tau q} + Ke^{-\tau r}\right)^2$, where K is the option strike price, C and P the prices of call and put options, S the underlying spot price and τ the time to maturity of the option.

imizing the likelihood defined on the option forecasting errors in a Kalman filter. For identification, we require matrices M, M^*, Λ, R and Q to be triangular, giving a total of 20 parameters to estimate. We borrow from Bates (2000) and conveniently discretize the matrix transition dynamics for state process X, accounting for the variability of conditional first and second moments. For the observation equation, we assume Gaussian errors and account for a potential autocorrelation of option pricing errors. Details on the estimation procedure are provided in the appendix B.2.

Despite the additional flexibility in allowing three interconnected channels of implied-volatility surface variation, our model adopts a reasonably parsimonious parametrization. For comparison, the physical and risk-neutral dynamics of a two-factor Bates (2000)-type model implies only three parameters less than our model, while a three-factor Bates (2000)-type model requires the identification of 25 structural parameters. This feature allows us to improve on the pricing performance of such benchmark two- and three- factor models, while preserving a good identification of the model parameters. In addition, we show that our specification of option-implied dynamics, with two dependent volatility components and a third jump volatility component disconnected from the diffusive volatility, provides a sharp identification of correlated option-implied risks and their distinct roles for the dynamics of option-implied risk premia.

In the second step of our procedure we estimate conditional variance risk-premia, by estimating parameter β^*_Λ in Assumption 16, using a simple linear regression of realized VIX option portfolio payoffs on model-implied variance risk premia, constructed from the filtered option-implied risks and the estimated parameters in the first step. Precisely, we first compute VIX option portfolio payoffs for maturities $\tau_1, \tau_2, \ldots, \tau_n = 1, 2, 3, 4, 5, 6, 9, 12$ months and construct a time series of in-sample weekly observations for variables $Z_t := (Y_{t+\tau_i}(\tau_i), U_t(\tau_i))_{i=1,\ldots,n}$ in linear model (3.28), where $t = 1, \ldots, N$ and the in-sample sample size is $N = 359$. We then estimate the single unknown parameter β^*_Λ in equation (3.28) with a pooled linear regression. To account for the overlap of VIX portfolio payoffs collected at weekly frequencies, we compute standard errors and critical values with a Newey-West correction for autocorrelation and heteroskedasticity.

3.2.2 Option Pricing Performance and Model Fit

In this section, we quantify the option pricing performance and the statistical fit of our model (model SVJ_{31}), in relation to the benchmark two- and three-factor models in Table 2.1. These models are linked to different degrees of parametrization and to state spaces of different dimensions. Moreover, while two-factor models are nested into our 2×2 matrix AJD setting, three-factor Bates (2000)-type models (model SVJ_{30}) are not. Therefore, such a comparison needs to consider also the different model dimensions, in order to avoid favoring highly parametrized models that may exhibit overfitting.

To control for overfitting, we split our sample of 883 weekly observations of S&P500

	SV_{20}	SV_{30}	SV_{31}	SVJ_{20}	SVJ_{30}	SVJ_{31}
RMSIVE						
in-sample	1.323	1.237	0.941	0.858	0.718	0.678
out-of sample	1.672	1.552	1.203	1.093	0.826	0.769
MAIVE						
in-sample	1.023	0.957	0.731	0.680	0.565	0.549
out-of sample	1.325	1.226	0.948	0.854	0.640	0.610
Average log-likelihood						
in-sample	7.288	7.359	8.001	8.100	8.315	8.491
out-of sample	6.667	6.878	7.298	7.265	7.955	8.005

Table 3.2: Indicators of pricing performance and statistical fit. We report indicators of in- and out-of-sample pricing performance and fit for model SVJ_{31} and for the benchmark models in Table 2.1. The in-sample period for estimation is January 1996 to December 2002. The out-of-sample period is from January 2003 to January 2013. For each model, we report the daily root-mean-squared implied volatility error (*RMSIVE*) and the daily mean absolute implied volatility error ($MAIVE$). These quantities are computed using the filtered states implied by the in-sample weekly parameter estimates for each day of our in- and out-of-sample periods. As a measure of statistical model fit and predictive ability, we also report the in- and the out-of-sample average value of the weekly likelihood function, evaluated at the in-sample parameter estimates.

options into an in-sample period (from January 1996 to December 2002) with 359 observations and an out-of sample period (from January 2003 to January 2013) with 524 observations. Besides focusing on in-sample pricing performance and fit, we require that higher dimensional models achieve a relatively stable pricing performance and statistical fit out-of-sample. The out-of-sample period includes phases of very low volatility and benign markets, such as the conundrum, as well as periods of very high volatility and market turmoil, such as the recent financial crisis and the EU sovereign debt crisis. Therefore, it represents a reasonably challenging benchmark for stochastic volatility models.

We estimate all models using only the in-sample weekly data. We then compute proxies of in-sample pricing accuracy, such as the weekly absolute average implied volatility error, by computing option implied volatility pricing errors for each week of our in- and out-of-sample periods, using the filtered states implied by the in-sample parameter estimates. Finally, we compare the statistical fit of different models using the in- and the out-of-sample value of the average likelihood function, evaluated at the in-sample parameter estimates. Table 3.2 summarizes the pricing performance and the statistical fit of the models in Table 2.1.

The results in Table 3.2 indicate that our matrix AJD model produces the best pricing performance and statistical fit, both in- and out-of-sample, with respect to all models in Table 2.1. The improvement in pricing performance relative to two-factor models, such as a Bates (2000)-type model (column SJV_{20} in Table 3.2), is very substantial. It amounts to about 18.5% (19.4%) in-sample and 28.8% (29.2%) out-of-sample, using the $RMSIVE$ ($MAIVE$) metric. The improvement of the in-sample (out-of-sample) value of the likelihood function is 4.5% (9.7%) and is statistically significant at conventional significance levels.

Model SVJ_{31} also improves with respect to benchmark three-factor Bates (2000)-type models (column SVJ_{30}), despite having four parameter less in its specification of the physical and risk-neutral dynamics of the smile. The improvement in pricing performance is about 3.1% in-sample and 6.7% out-of-sample, with respect to the $RMSIVE$ metric. Similarly, while there is no in-sample improvement with respect to the $MAIVE$ metric, the out-of-sample improvement in pricing performance is 4.3%. Finally, the more parsimonious model SVJ_{31} also attains a higher average likelihood in- and out-of-sample. The likelihood improvement is 2.1% in-sample and 1.6% out-of-sample. These measures of improvement in statistical fit are conservative, as model SVJ_{31} is less parametrized than model SVJ_{30}.

In addition to having the better in- and out-of sample performance, model SVJ_{31} also implies the smallest deteriorations in out-of sample performance. The out-of-sample $RMSIVE$ ($MAIVE$) is only 5.5% (6.1%) higher than the in-sample $RMSIVE$ ($MAIVE$). In contrast, in the SVJ_{20} (SVJ_{30}) models, the out-of-sample $RMSIVE$ and $MAIVE$ are 20.9% (10.2%) and 20.8% (10.7%) higher than the in-sample $RMSIVE$ and $MAIVE$, respectively. Similarly, while the average out-of-sample likelihood in model SVJ_{31} is only 6.1% lower than the in-sample likelihood, the out-of-sample likelihood of the SVJ_{20} (SVJ_{30}) model is 11.1% (6.7%) lower. In summary, the improvements implied by model SVJ_{31} are not a consequence of overfitting, as the model's pricing performance and fit are quite similar in- and out-of-sample.

The pricing performance of model SVJ_{31} is quite consistent over time, when we stratify our sample period even further. The $MAIVE$ is lower than one (two) volatility percentage points in 95.4% (99.2%) of the days and it is above 2% in only seven days, six days during the financial crisis and one day immediately after the US downgrade. In terms of relative volatility errors, the MAIVE is less that 5% (10%) of the volatility level in 78.3% (98.8%) of the days. It is more than 10% of the volatility in only 11 days, two of which are at the beginning of the sample and the remaining 9 in the conundrum period (2003-2007). In the crisis period of 2008-2009, the MAIVE never exceeds 5.5% of the volatility level.

To understand in more detail along which dimensions of the implied volatility surface model SVJ_{31} provides the largest pricing improvements, it is useful to disaggregate the pricing performance across the moneyness and the maturity dimensions. Table B.1 of the appendix compares $MAIVE$ pricing errors of models SVJ_{31} and SVJ_{20}, across different moneyness and maturity bins (in days). It shows that model SVJ_{31} especially improves on the modelling of out-of-the money options of maturities of 30 days or higher. To illustrate, the in-sample (out-of-sample) $MAIVE$ of model SVJ_{31} for maturities $\tau < 30$, $30 \leq \tau < 75$, $75 \leq \tau < 180$ and $\tau \geq 180$ is 8.1% (6.3%), 29.8% (31%), 13% (18.2%), 22.9% (33.6%) and 29.2% (21.8%) lower, respectively, than for model SVJ_{20}. Similarly, the in-sample (out-of-sample) $MAIVE$ of model SVJ_{31} for option deltas $\Delta < 0.2$, $0.2 \leq \Delta < 0.4$, $0.4 \leq \Delta < 0.6$, $0.6 \leq \Delta < 0.8$ and $\Delta \geq 0.8$ is 13.3% (29.8%), 17.8% (28.5%), 9.9% (15.6%), 19.2% (25.3%) and 28.1% (33.8%) lower, respectively.[26] In summary, this evidence shows that model SVJ_{31} clearly improves on the specification of the term structure of the option-implied volatility smile of benchmark two-factor models.

3.2.3 Mutually Exciting Option-Implied Risks

The time series of option-implied components X_{11}, X_{22} and X_{12} estimated for our model is presented in Figure 3.2, highlighting quite different persistence and variability properties. For instance, the half-lifes (volatilities) of estimated states X_{11}, X_{12} and X_{22} are 1.275, 0.277 and 0.108 years (0.0091, 0.0104 and 0.0259), respectively.

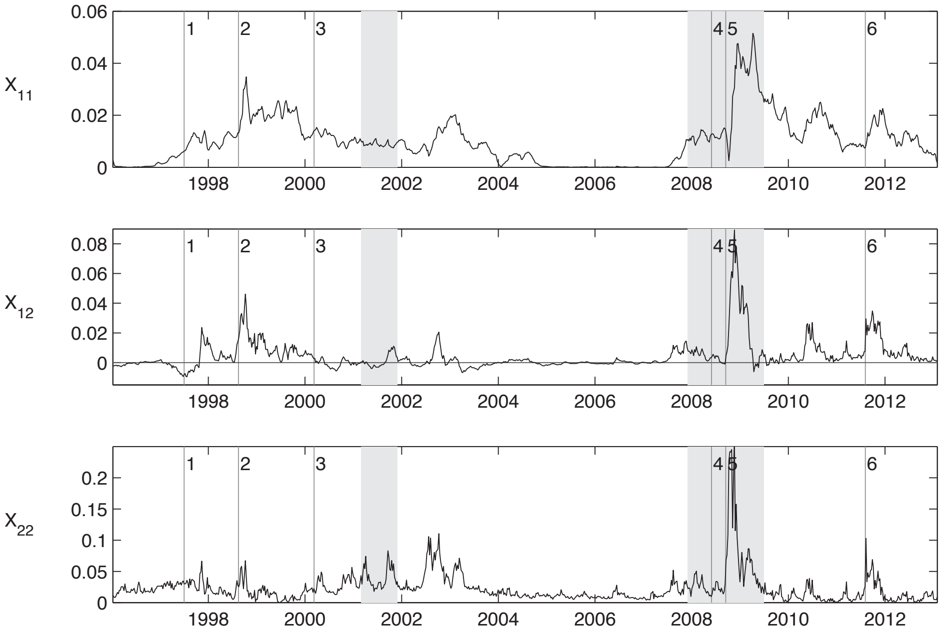
By construction, in Panel A of Figure 3.2, components X_{11} and X_{22} of the diffusive volatility are nonnegative. Component X_{22} is on average larger than X_{11}, besides being more volatile and less persistent. In this sense, the diffusive volatility is decomposed into two non-Markovian components, with significantly different persistence and volatility of volatility. Component X_{12} is most of the time positive, but it can also turn slightly negative in a number of cases. The largest absolute values of X_{22} and X_{12} are observed in connection with periods of significant market turmoil or financial distress, as for instance during the recent financial crisis and EU sovereign debt crisis. The more persistent volatility component X_{11} can also spike in periods of financial distress, but often with a lag or a lead with respect to X_{12} and X_{22}.

Such state dynamics can imply large variations in the relative importance of option-implied risks X_{11}, X_{22} and X_{12} for the structure of the conditional distribution of both jump and diffusive variance. Panel B of Figure 3.2 illustrates this aspect by scaling components X_{11t} and X_{12t} by the total diffusive variance $tr(X_t) = X_{11t} + X_{22t}$. It appears that while visually the time series of $tr(X_t)$ and X_{22t} have strong similarities, the fraction of diffusive variance generated by the most persistent component X_{22t} can

---

[26] We also find that the fitted prices of model SVJ_{31} are systematically more often inside the bid-ask price band than the fitted prices of model SVJ_{20}. On an aggregate level, model SVJ_{31} (SVJ_{20}) implies prices inside the bid-ask band in 51.7% (44.9%) and 64.7% (55.6%) of the cases, in- and out-of-sample, respectively. The disaggregated statistics across moneyness and maturity are collected in Tab. B.1 in the appendix.

Panel A: Option-implied components for model SVJ_{31}



Panel B: Scaled components X_{11}, X_{12} and diffusive variance $tr(X_t)$

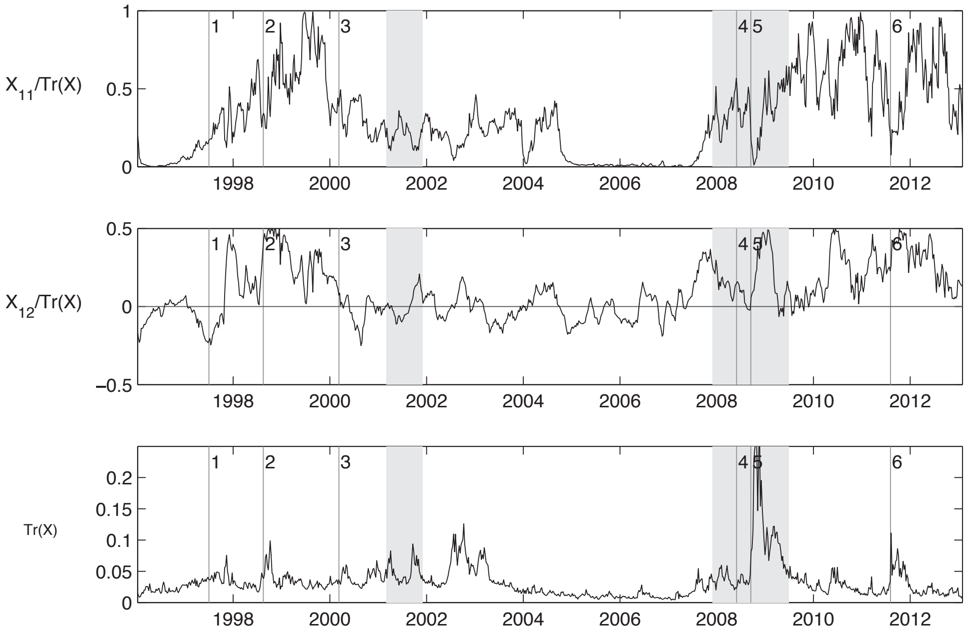


Figure 3.2: Panel A: Time series of filtered option-implied components $\left(X_{11t}, X_{12t}, X_{22t}\right)$ for model SVJ_{31}. Panel B: Time series of option-implied components X_{11}, X_{12} for model SVJ_{31}, scaled by the diffusive variance $tr(X_t)$. The range of admissible values $[-tr(X_t)/2, tr(X_t)/2]$ for X_{12t} follows from the positive definiteness of matrix X_t. Grey areas highlight NBER recessions; vertical lines indicate important crisis events in our sample period, listed in more detail in Table 2.5.

vary essentially from zero to one. The top plot of Panel B in Figure 3.2 shows that the fraction of total diffusive variance created by the most persistent component is basically zero during the whole conundrum period, it is about one shortly before the collapse of the NASDAQ bubble and it rapidly increases from about 0.2 to 0.9 shortly after the EU sovereign debt crisis. The middle plot of Panel B shows that the relative importance of component X_{12} can also vary a lot over time. X_{12t} can be as large as 50% of $X_{11t} + X_{22t}$ during phases of market turmoil, e.g, shortly after the devaluation of the Thai Baht, the begin of the Russian crisis, the Lehman default and the US downgrade.[27] Note that while X_{12} is disconnected from the diffusive variance, it has a significant contribution to the jump volatility, via the jump intensity. Therefore, this option-implied component is directly related to the decomposition of the total variance into diffusive and jump variance. Given its predominant role during periods of market distress, it is also a key driver of the jump volatility dynamics in such periods.

The time series of option-implied states uncovers the dynamic properties of such state variables as drivers of option-implied risks. In contrast, the estimated model parameters capture more directly the dynamic interactions between option-implied components and their relation with the market price of the smile. Table 3.3 presents the parameter values estimated for our model (SVJ_{31}) and a number of benchmark two- and three-factor models in Table 2.1.

All parameters are very significant, with the exception of the constant λ_0 in the intensity process, which is not significantly different from 0. Since we cannot reject the null hypothesis $\beta^* = \beta$, the data also support a completely affine specification of the market price of risk in our matrix AJD setting. In all parameter matrices M, M^*, Λ and R, the out-of-diagonal element is strongly significant, indicating that option prices are better described by our three-factor matrix AJD models than by a two-factor diagonal model with independent components. The estimated jump parameters $\lambda^- << \lambda^+$ directly reflect a negative risk-neutral skewness of the distribution of log returns jumps.

The large negative coefficient M^*_{22} indicates that option-implied component X_{22} has the strongest degree of autonomous mean reversion. This feature induces an option-implied state X_{22} with the lowest persistence in our sample. Given the positivity of parameter M^*_{12}, the mean reversion of X_{22} is dampened in phases where state X_{12} is positive, while it is reinforced when X_{12} is negative. Recalling that X_{12} is positive most of the time, with stronger excursions during phases of market distress when the option-implied smile is typically steeper, this feature induces a mutually-exciting behaviour of state variables X_{22} and X_{12} during phases of market turmoil. Given the point estimates $Q_{22} >> Q_{11}$ in the volatility of volatility matrix Q, component X_{22} also has the largest

---

[27] The ratio $|X_{12t}|/tr(X_t)$ is less than 0.5 by construction, because of the positive definiteness of X_t.

Panel A: Diffusion parameters

	SV_{20}	SV_{30}	SV_{31}	SVJ_{20}	SVJ_{30}	SVJ_{31}
M_{11}	-0.3121	-0.0844	-1.0716	-0.3242	-0.1231	-0.0079
	(0.0063)	(0.0020)	(0.0185)	(0.0067)	(0.0023)	(0.0002)
M_{22}	-5.0719	-5.4283	-4.9213	-4.4564	-4.2041	-2.6808
	(0.1040)	(0.1254)	(0.0489)	(0.0895)	(0.0582)	(0.0261)
M_{33}		-1.4410			-0.5517	
		(0.0307)			(0.0104)	
M_{12}			14.3050			1.0265
			(0.2173)			(0.0120)
Q_{11}	0.2370	0.1957	0.0556	0.0903	0.0742	0.0698
	(0.0024)	(0.0026)	(0.0006)	(0.0015)	(0.0010)	(0.0009)
Q_{22}	0.4209	0.4498	0.5256	0.4204	0.2853	0.2924
	(0.0057)	(0.0062)	(0.0033)	(0.0054)	(0.0026)	(0.0024)
Q_{33}		0.0718			0.0738	
		(0.0019)			(0.0016)	
Q_{12}			-0.1440			-0.0770
			(0.0021)			(0.0012)
R_{11}	-1.0000	-1.0000	-0.0431	-1.0000	-0.9997	-0.2970
	(0.0131)	(0.0134)	(0.0008)	(0.0227)	(0.0189)	(0.0036)
R_{22}	-0.5348	-1.0000	-0.6405	-0.3823	-0.7111	-0.4057
	(0.0087)	(0.0192)	(0.0055)	(0.0069)	(0.0117)	(0.0048)
R_{33}		0.9633			-0.1178	
		(0.0255)			(0.0026)	
R_{12}			-0.7672			-0.8708
			(0.0110)			(0.0121)
β_{11}	1.0000	1.0031	1.0000	1.0006	1.0064	1.0012
	(0.0160)	(0.0169)	(0.0118)	(0.0191)	(0.0180)	(0.0116)
β_{22}	1.0000	1.0007		1.0000	1.0042	
	(0.0187)	(0.0219)		(0.0197)	(0.0153)	
β_{33}		1.0162			1.0146	
		(0.0235)			(0.0187)	
M^*_{11}	-1.4051	-1.2204	-0.6378	-0.7395	-0.8289	-0.5467
	(0.0266)	(0.0298)	(0.0091)	(0.0172)	(0.0134)	(0.0083)
M^*_{22}	-1.8593	-2.2558	-2.7528	-1.9462	-1.2661	-2.6808
	(0.0401)	(0.0584)	(0.0435)	(0.0477)	(0.0221)	(0.0334)
M^*_{33}		-0.4869			-0.5539	
		(0.0116)			(0.0093)	
M^*_{12}			1.9200			0.3982
			(0.0284)			(0.0051)
β^*_{11}	1.0000	1.0017	1.0000	1.0006	1.0064	1.0012
	(0.0203)	(0.0216)	(0.0162)	(0.0200)	(0.0190)	(0.0124)
β^*_{22}	1.0000	1.0046		1.0000	1.0042	
	(0.0201)	(0.0199)		(0.0251)	(0.0232)	
β^*_{33}		1.0693			1.0146	
		(0.0316)			(0.0208)	

Panel B: Jump parameters

	SVJ_{20}	SVJ_{30}	SVJ_{31}
λ_0	0.0000	0.0003	0.0000
	(0.0003)	(0.0002)	(0.0002)
Λ_{11}	43.8971	57.3248	25.6671
	(0.9240)	(0.9276)	(0.3193)
Λ_{22}	1.0566	11.9429	15.9795
	(0.0265)	(0.1899)	(0.1933)
Λ_{33}		0.0454	
		(0.0008)	
Λ_{12}			40.4278
			(0.6332)
$\bar{k}$	-0.1500	-0.1500	
	(0.0030)	(0.0019)	
δ	0.1500	0.1500	
	(0.0027)	(0.0020)	
λ^-			7.1518
			(0.0372)
λ^+			58.3547
			(0.7690)
β^*_Λ			0.3230
			(0.0553)

Table 3.3: In-sample parameter point estimates with standard errors in brackets.

local volatility. Note that besides driving a high-frequency component in the diffusive volatility, state X_{22} is also related to high-frequency movements in the jump volatility, because parameter Λ_{22} in the intensity process is positive and significant. Thus, we can interpret X_{22} as a high-frequency mean reverting component of the volatility, which features a mutually exciting behaviour with X_{12} during phases of market distress. Interestingly, we also find that we cannot reject the null hypothesis $M^*_{22} - M_{22} = 0$, implying that the instantaneous risk premium of the least persistent component X_{22} is zero.

The negative coefficient M^*_{11} indicates that option-implied volatility component X_{11} is also mean-reverting, even though clearly more persistent than X_{22}, with a low local volatility since $Q_{22} >> Q_{11}$. As for X_{22}, the mean-reversion of X_{11} is dampened during phases of market distress, when X_{12} is positive, so that overall the volatility follows a mutually-exciting dynamics in such periods. State X_{11} is also positively related to low-frequency shocks in the jump volatility, because parameter Λ_{11} is positive and significant. Thus, we can also interpret X_{11} as a low-frequency mean reverting component of the volatility, featuring mutually-exciting behaviour with X_{12} in periods of market distress. Since $M^*_{11} < M_{11}$ and $M^*_{12} < M_{12}$, the instantaneous risk premium for X_{11} depends on both X_{11} and X_{12}. It is negative and larger in absolute value when X_{12} is large.[28]

The negative coefficients M^*_{11} and M^*_{22} indicate that the third option-implied component X_{12} has an autonomous mean-reversion between the mean reversion speed of the high- and low-frequency components of the volatility. The total local mean reversion of X_{12} depends also on X_{11} and X_{22} and is asymmetric. It is increase (dampened) in states where X_{12} is negative (positive), making X_{12} more persistent and mutually-exciting in phases of market distress. Note that, by construction, state X_{12} loads on the jump volatility, via the jump intensity, but not on the diffusive volatility. Moreover, the large positive estimated loading Λ_{12} of X_{12} in the intensity process indicates that this component is the most important state variable in the dynamics of the jump variance during periods of market distress. Therefore, it has the natural interpretation of a pure jump variance risk factor, with mutually-exciting dynamics during phases of turbulences in financial markets. Since X_{12} is directly related to pure jump volatility in periods of financial distress, we expect it to have a negative risk premium. Indeed, since $M^*_{22} = M_{22}$, the local risk premium of the jump volatility component X_{12} depends on X_{12} itself and the lowest-frequency volatility component X_{11}. Given that both $M^*_{11} - M_{11} < 0$ and $M^*_{12} - M_{12} < 0$, the local risk premium of X_{12} is always negative in our sample period.[29]

---

[28] Even though in states where X_{12} is negative its contribution to the risk premium for X_{11} is positive, this effect is compensated by the effect of the negative dependence of the risk premium on X_{11}.

[29] At the model parameters, the local risk premium of X_{12} is always positive when X_{12} is positive. Given that $M^*_{11} - M_{11} < 0$, the contribution of X_{12} to its premium can be positive when state X_{12t} is negative. However, in our sample we find that this effect is always compensated by the negative risk premium contribution proportional to state X_{22t}.

The largest risk premia for pure jump volatility typically arise during crisis periods, whenever simultaneously the high-frequency option-implied volatility component X_{22} and the pure jump option-implied volatility component X_{12} are large.

3.2.4 The Market Price of the Smile

At the estimated model parameters, the instantaneous risk premium for option-implied components X_{11}, X_{22}, X_{12} takes a simplified form:[30]

$$\begin{aligned} \frac{1}{dt}(E_t^{\mathbb{P}} - E_t^{\mathbb{Q}})[dX_{11t}] &= -1.0776 X_{11t} \ , & (3.29) \\ \frac{1}{dt}(E_t^{\mathbb{P}} - E_t^{\mathbb{Q}})[dX_{12t}] &= -0.6283 X_{12t} - 0.5388 X_{22t} \ , & (3.30) \\ \frac{1}{dt}(E_t^{\mathbb{P}} - E_t^{\mathbb{Q}})[dX_{22t}] &= -1.2566 X_{12t} \ . & (3.31) \end{aligned}$$

In other words, while the two more persistent option-implied components have a negative instantaneous market price of risk, high-frequency component X_{22} is not priced instantaneously. This means that the instantaneous premia for shocks in option-implied risks X_{11} and X_{12} are actually risk premia for X_{11}- and X_{12}-risks that are not conditionally spanned by X_{22}-shocks. Given the estimated loadings of components X_{11}, X_{12} and X_{22} in equations (3.29)-(3.30), the highest instantaneous absolute risk premia and the most persistent risk premium dynamics emerge for X_{11}-shocks. Finally, notice that since X_{11}, X_{12} and X_{22} satisfy an interdependent joint dynamics, the dynamics of state variable X_{22} is different under the physical and the risk neutral probabilities, even though component X_{22} has a zero instantaneous risk premium. This means that payoffs with exposure to future X_{22}-risks, such as VIX option portfolio payoffs, with exhibit in general a non-zero risk premium for such an exposure.

We can study the term structure of the market price of the smile, by measuring the risk premium associated with the single components of the average integrated option-implied risk $\frac{1}{\tau}\int_t^{t+\tau} X_s ds$. Identifying the term structure of the market price of the smile is a necessary step, in order to understand the decomposition of the term structure of variance risk premia into a risk premium component for exposure to option-implied risks and a risk premium component for pure jump variance risk. Given the distinct persistence features of option-implied risks X_{11}, X_{12} and X_{22}, we expect the risk premia depending on the most persistent components to exhibit a more pronounced term struc-

---

[30] Recalling that we cannot reject the hypotheses $\beta = \beta^*$ and $M_{22}^* = M_{22}$, we obtain (see Table 3.3):

$$\frac{1}{dt}(E_t^{\mathbb{P}} - E_t^{\mathbb{Q}})[dX_t] = \begin{pmatrix} -0.5388 & 0 \\ -0.6283 & 0 \end{pmatrix} X_t + X_t \begin{pmatrix} -0.5388 & -0.6283 \\ 0 & 0 \end{pmatrix} .$$

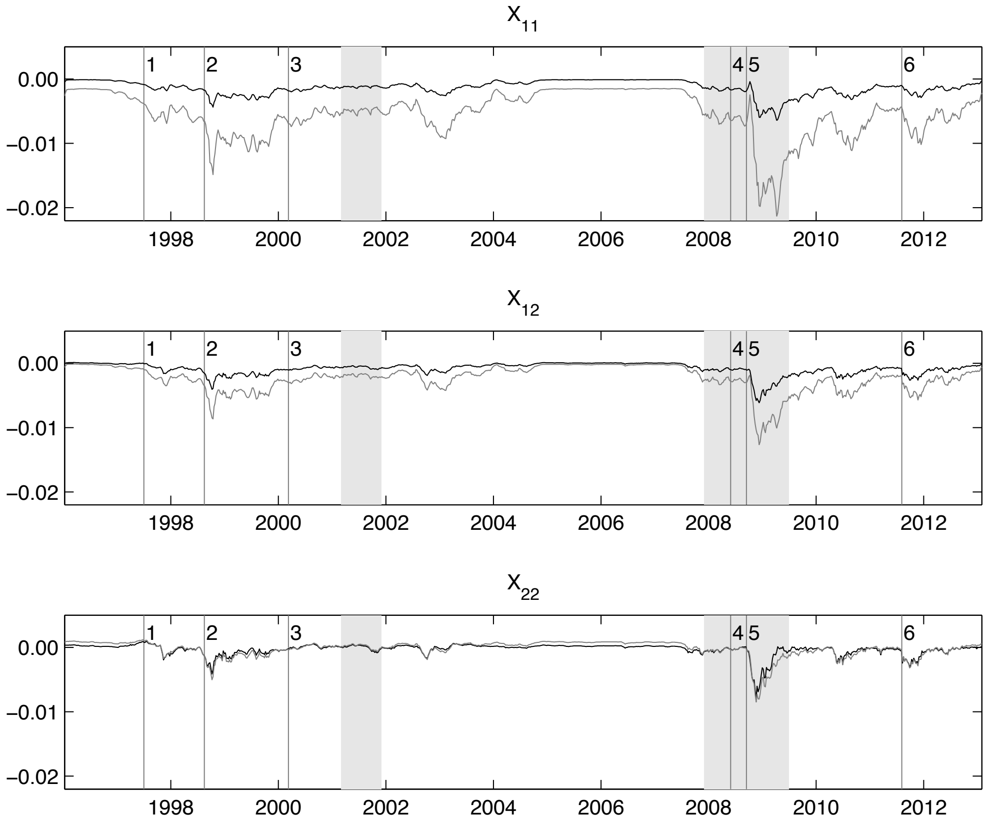


Figure 3.3: Market price of a dynamic smile. We plot the difference $(E_t^{\mathbb{P}} - E_t^{\mathbb{Q}}) \left[\frac{1}{\tau} \int_t^{t+\tau} X_s ds \right]$ of the physical and risk-neutral expectations of the integrated option-implied state X component-wise, with states X_{11}, X_{12} and X_{22} from the top to the bottom panel, for maturities $\tau = 3$ and $\tau = 12$ months (black and grey lines, respectively). Grey areas highlight NBER recessions; vertical lines indicate important crisis events, listed in Table 2.5.

ture dynamics. We follow the standard practice in the variance risk premium literature and measure the model-implied risk premium for integrated states X_{11}, X_{22} and X_{12}, using the expected excess payoff of a swap contract with flexible leg $\frac{1}{\tau} \int_t^{t+\tau} X_s ds$ and time to maturity τ. To illustrate the dynamics and the term structure of the market price of the smile, we fix two horizons $\tau_1 = 3$ and $\tau_2 = 12$ months and compute the difference of the $\mathbb{P}-$ and $\mathbb{Q}-$ expected average integrated state:

$$(E_t^{\mathbb{P}} - E_t^{\mathbb{Q}}) \left[\frac{1}{\tau_i} \int_t^{t+\tau_i} X_s ds \right] \ ; \ i = 1, 2 \ .$$

The different elements of these matrices identify the size and the term structure of the option-implied risk premia for latent states X_{11}, X_{12} and X_{22}. Figure 3.3 presents the time series of these risk premia.

Option-implied risks X_{11} and X_{12} have an unambiguously negative market price of risk. The market price of X_{22} is also most of the time negative, but it can turn slightly positive occasionally. The market prices of all components are larger in absolute value in phases of market distress, reflecting the higher price of option-implied insurance. More importantly, each component of the market price of the smile has very distinct term structure features. While the term structure of the market price of X_{22} is flat from an horizon of 3 months on, both X_{11} and X_{12} imply a decreasing term structure. The slope of these term structures is pro-cyclical, in the sense that it becomes more negative in periods of financial distress. The low-frequency option-implied risk X_{11} implies the largest absolute risk premia at long horizons of 12 months, which are about double the risk premia implied by X_{12} and more than three times those of X_{22}.

The intuition for the time-series features of the market prices of the smile in Figure 3.3 are well understood, in terms of corresponding risk premia linked to stochastic risks with distinct persistence features. To this end, Table B.3 of the Appendix decomposes each component of the market price of the smile into the single contributions of option-implied risks X_{11}, X_{12} and X_{22}. First, we find that all market prices of the smile are linear functions exclusively of option-implied risks X_{11} and X_{12}. In this sense, these two components can be interpreted as risk premium factors that span the dynamics and the term structure of option risk premia. Second, while the market price of X_{11} only depends on X_{11} itself, the market price of X_{12} and X_{22} loads on both X_{11} and X_{12}. However, while the market price of X_{12} is dominated by low frequency component X_{11} at longer horizons, the market price of X_{22} mostly depends on the less persistent component X_{12}. As a consequence, low frequency option-implied component X_{11} has the most persistent risk premia, followed by components X_{12} and X_{22}, respectively. Third, the dominant role of persistent option-implied risk X_{11} implies a negative market price and a corresponding decreasing term structure of integrated risk premia for X_{11} and X_{12}. A similar tendency arises for the highest-frequency option-implied risk X_{22}. However, in this case integrated risk premia can be occasionally marginally positive and the term structure slightly increasing, in states where the jump volatility is not persistent (X_{11} is small) and X_{12} takes negative values. Finally, the flat relation in Table B.3, between the market price of the smile and option-implied risk X_{12} at horizons above 3 months, implies in most cases a flat term structure of integrated risk premia for X_{22} between 3 months and 12 months horizons.[31]

---

[31] These features arise are a consequence of the estimated dynamic interactions of option-implied risks in our model, yielding the observed empirical patterns for the market price of the smile. Consistent with the above intuitions, we find that the risk premium for component X_{11} has a perfect weekly correlation with the level of X_{11}, both over short (1 month) and longer (6 month) horizons. Similarly, the risk premium for X_{12} has a correlation of 0.9 (0.95) with the level of X_{11}, over short (long)

3.2.5 Interpretation of Option-Implied Risks

To gain additional economic interpretation for the role of option-implied components X_{11}, X_{12}, X_{22}, it is useful to link these state variables to observable characteristics of the option-implied volatility surface. This approach allows us to interpret each option-implied component, e.g., in terms of the price of an option replicating portfolio. Given that state X spans both option-implied risks and risk premia, such an approach can help to link the price of option strategies offering protection against a market downturn to the different dimensions of the market price of the smile.

We find that high-frequency option-implied component X_{22} has a weekly correlation of 0.84 with the 30-days at-the-money implied volatility. Components X_{12} and X_{11} are obviously correlated with X_{22}. However, given their different persistence and variability properties, the weekly correlation with X_{22} is relatively low, amounting to 0.59 and 0.26, respectively. As a consequence, a relevant part of the conditional variation of X_{11} and X_{12} in our sample is orthogonal to X_{22}. We find that $X_{12}-$variations orthogonal to X_{22} have a weekly correlation of -0.85 with a 30-days option-implied skew residual, given by the residual of a regression of the 30-days option-implied skew on the 30-days option-implied volatility. In this sense, X_{12} is linked to short term option-implied skew dynamics disconnected from the 30-days option-implied volatility. We also find that $X_{11}-$variations orthogonal to X_{22} have a weekly correlation of 0.83 with the residual of a regression of the option-implied skew term structure on the 30-days option-implied volatility. In this sense, X_{11} is related to option-implied skew term structure dynamics disconnected from the 30-days option-implied volatility.

The above interpretation of option-implied risks X_{11} and X_{12} depends on the residuals of a regression of model-implied variables X_{11} and X_{12} on model-implied state X_{22}. A more direct interpretation, in terms of simple transformations of model-implied variables, is obtained by scaling X_{11} and X_{12} with a model-implied proxy of volatility. In this way, we can isolate the volatility-independent effects of X_{11} and X_{22} on the option-implied volatility smile. A convenient model-implied proxy for the volatility level is the diffusive variance $Y_{1t} := tr(X_t)$, which implies bounded scaled quantities $Y_{2t} := X_{11t}/tr(X_t)$ and $Y_{3t} := X_{12t}/tr(X_t)$. In this reparametrization, Y_1 measures the level of the diffusive variance, while Y_2 captures the composition of the diffusive variance, in terms of two components with distinct persistence and variability. Note that Y_2 is bounded between zero and one and roughly captures, on a scale from zero to one, the degree of persistence of the diffusive variance not related to the variance level. Y_3 roughly captures the fraction of jump-driven variance not due to diffusive return shocks, in percentage of the level of the diffusive variance. Therefore, it captures variations in jump variance not directly

---

horizons, while the risk premium for X_{22} has a correlation of 0.999 (0.98) with the level of X_{12}.

related to the diffusive variance. Note that Y_3 is bounded between $-1/2$ and $1/2$ and can change sign.

As expected, we find that variable Y_1 is a good proxy of the option-implied volatility level in the data, as it has a weekly correlation of 98% with the at-the-money 30-days option-implied volatility. In order to interpret the scaled variables Y_2 and Y_3, we need to consider observable option-implied quantities that do not depend on the level of the 1-month implied volatility in the data. We follow the same approach as above and we separately regress on the one-month implied volatility the one-month option-implied skew and a proxy of option-implied skew term structure, corresponding to the difference between the 12 months and the 3 months option-implied skews. We compute the time series of residuals from these two regressions and interpret them as skew and skew term structure components unrelated to the volatility level in the data. Figure 3.4 presents the time series of model-implied variables Y_1, Y_2, Y_3 and compares these to those of the 1-month implied volatility and the two skew and skew term structure residuals in the data.

We find that variable $Y_{3t} = X_{12t}/tr(X_t)$ has a correlation of 0.85 with the skew residual of our regression, suggesting a possible interpretation of Y_3 in terms of an observable option-implied skew component unrelated to the volatility level. We also find that variable $Y_{2t} = X_{11t}/tr(X_t)$ has a correlation of 0.71 with the skew term structure residual of our regression, suggesting a possible interpretation of Y_2 in terms of an observable option-implied skew term structure component unrelated to the volatility level.[32]

The identification of variables Y_1, Y_2, Y_3, as a volatility level factor and as two skew and skew term structure residuals, respectively, can help to better understand the economic content of the two risk premium components X_{11} and X_{12} that span the market price of the smile in our model. According to the results of this and the previous sections, the market price of the smile is naturally related to observed option-implied skew components, which parsimoniously measure a time-varying and horizon-dependent price of market insurance and are disconnected from the volatility level.

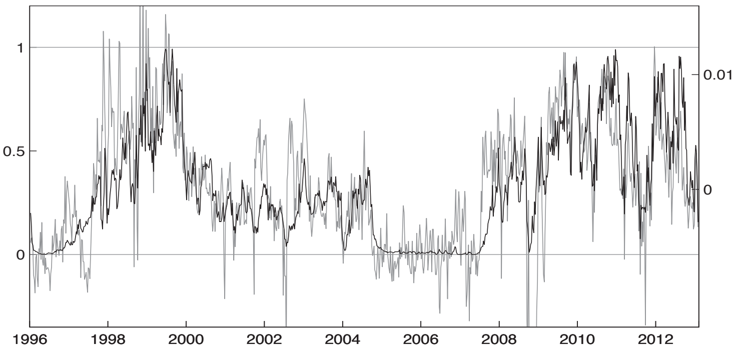
3.2.6 Term Structure of Variance Risk Premia

To identify the term structure of variance risk premia, we estimate parameter β_{Λ}^{*} in the linear model specified by Proposition 3, using a pooled linear regression. The parameter estimate $\beta_{\Lambda}^{*} = 0.427$ reported in Table 3.3 implies an expected realized Itamura Saito

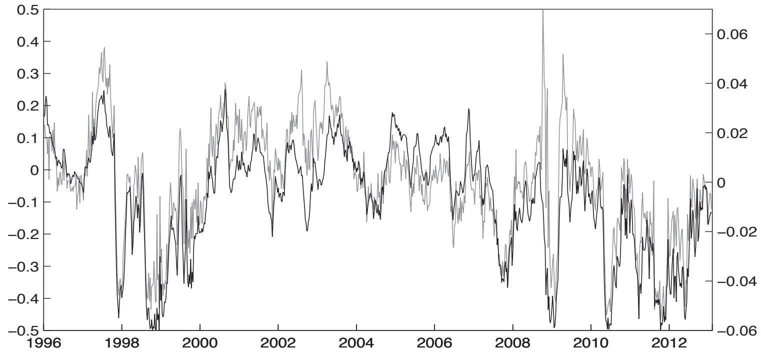
---

[32] The time series in Figure 3.4 and the corresponding correlations listed above were obtained by comparing the normalized components Y_1, Y_2, Y_3 with the observable properties of the model-implied volatility surface. We obtain qualitatively similar, albeit more noisy, results using option-implied volatilities directly computed from option data.

Panel A: Scaled state $X_{11}/tr[X_t]$ as option-implied skew term structure residual



Panel B: Scaled state $-X_{12}/tr[X_t]$ as option-implied skew residual



Panel C: Diffusive variance $tr[X_t]$ squared at-the-money implied volatility

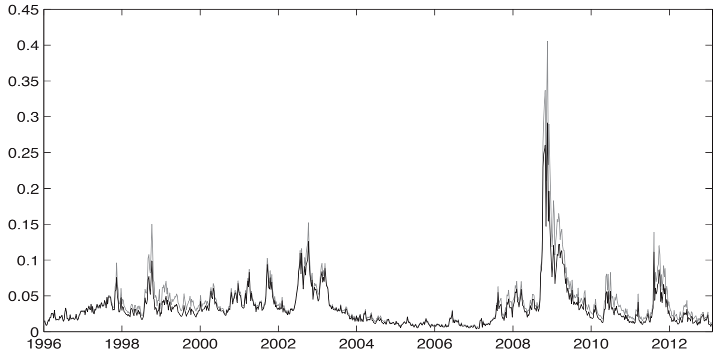


Figure 3.4: Scaled option-implied components of model SVJ_{31} as observable components of the option-implied volatility surface of S&P500 index options. Panel A: $X_{11}/tr[X_t]$ (black line) and skew term structure residual of the S&P500 option-implied volatility surface (grey line). Panel B: $-X_{12}/tr[X_t]$ (black lines) and one-month skew residual of the S&P500 option-implied volatility surface (grey lines). Panel C: $tr[X_t] = X_{11} + X_{22}$ (black line) and one-month at-the-money implied volatility^2 of S&P500 index options (grey line). The exact calculation method for the option-implied skewness and term structure is explained in Appendix B.3.

divergence of the return jump distribution under the physical measure, $E^{\mathbb{P}}[\mathcal{E}(1+k^*)]$, which is slightly less than half the expected realized divergence under the risk-neutral probability, $E^{\mathbb{Q}}[\mathcal{E}(1+k)]$. Using this value for β^*_{Λ} and the model parameter estimates from the first step of our estimation approach, we estimate the model-implied dynamics for the term structure of VIX variance risk premia.

Variance Risk Premia

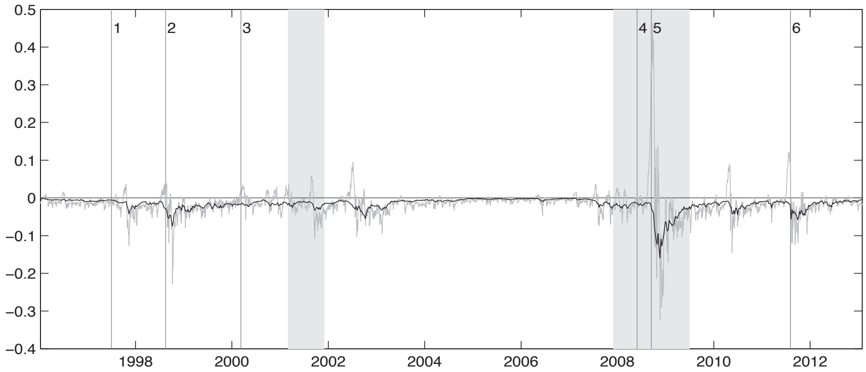
In Figure 3.5, we plot the model-implied variance risk premia for horizons $\tau = 12$ months and $\tau = 1$ month, together with their difference, as a proxy for the slope of the variance risk premium term structure.

We find that conditional variance risk premia are highly time-varying and unambiguously negative. They range between -0.1% and -16% (-0.4% and -11%) squared for horizon $\tau = 1$ month (horizon $\tau = 12$ months). We find that model-implied variance risk premia provide a plausible description for the first conditional moment of the realized payoffs of VIX option portfolios, at horizons $\tau = 1$ month and $\tau = 12$ months, respectively. The variability of such payoffs around the conditional first moment is state dependent and can be extremely high during periods of marker turmoil. Consistent with intuition, variance risk premia are largest in absolute value when the price of option-implied market insurance is large, typically during phases of financial distress and market turmoil, e.g., during the Asian and Russian crises in the late nineties, shortly before the collapse of the internet bubble in 2000, shortly after the Lehmann bankruptcy in September 2008 and the US downgrade in August 2011, and during the EU government debt crisis.

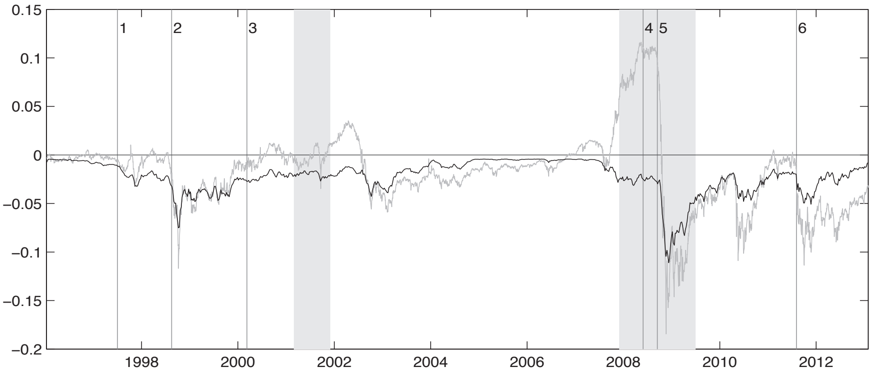
The slope of the term structure of variance risk premia is most of the time negative, reflecting a higher relative price of option-implied market insurance for longer investment horizons. However, it can be strongly upward sloping for short periods of time. Overall, the slope of the term structure is positive for about 12% of the observations in our sample. The most prominent cases in which we observe an inversion of the term structure of variance risk premia arise immediately after both the Lehmann default in September 2008 and the US downgrade in August 2011, when the spread between annualized 12 month and 1 month VIX variance risk premia has been as large as $+5.8\%$ squared and $+2\%$ squared, respectively.

Table B.4 of the Appendix shows that the term structure of variance risk premia depends on all option-implied risks X_{11}, X_{12} and X_{22}, with an absolute contribution of component X_{11} (components X_{12} and X_{22}) that monotonically increases (decreases) with the horizon. Low frequency risk X_{11} dominates variance risk premia al horizons above one quarter, while higher-frequency risks X_{22} and X_{12} dominate variance risk premia for horizon less than one quarter. These features imply a different role of low- and high-frequency components for the slope of the term structure of variance risk premia.

Panel A: 1 month VIX variance risk premium



Panel B: 12 months VIX variance risk premium



Panel C: Term structure of VIX variance risk premia

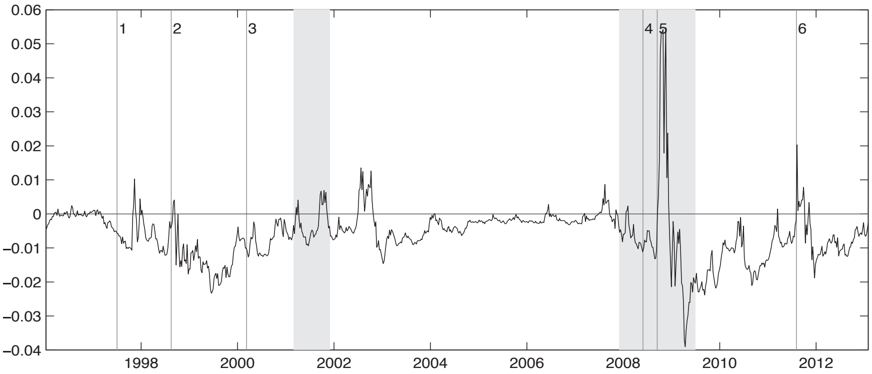


Figure 3.5: VIX variance risk premium and term structure of VIX variance risk premia. In panel A (B), we plot the annualized model-implied 1 month (12 months) VIX variance risk premium (black lines) and the payoffs of synthetic VIX option portfolios (grey lines). In panel C, we plot the slope of the model-implied term structure of VIX variance risk premia, computed as the difference of 12-months and 1-month VIX variance risk premia. Grey areas highlight NBER recessions; vertical lines indicate important crisis events in our sample period, listed in more detail in Table 2.5.

Indeed, while an increase in X_{11} lowers the slope of the term structure, and increase in X_{11} or X_{12} increases the slope.[33]

Variance Risk Premia for Diffusive and Jump Variance

To gain a deeper understanding for the multi-frequency dynamics of the term structure of VIX variance risk premia, we estimate the contributions of diffusive and jump variance risk to the VIX variance risk premium. We make use of the closed-form expressions in Proposition 2 and plot in Figure 3.6 the model-implied dynamics for diffusive and jump variance risk premia, $VRP_t^c(\tau)$ and $VRP_t^d(\tau)$, respectively, at horizons of 3 months and 12 months. As a proxy for the slope of their term structure, we also plot the difference of $VRP_t^c(\tau)$ and $VRP_t^d(\tau)$ at horizons of 12 months and 3 months.

We find that variance risk premia at all horizons are dominated by jump variance risk premia, as $VRP_t^d(\tau)$ is always at least 85% (65%) of the total variance risk premium at horizon $\tau = 3$ ($\tau = 12$) months. Moreover, the dynamic properties of the term structures of diffusive and jump variance risk premia are very different.

Diffusive variance risk premia for horizon $\tau = 3$ months are never above 1% squared in absolute value, even in the recent financial crisis, while they are consistently larger in absolute value at horizons $\tau = 12$ months, implying a maximum of about 3% squared during the 2008 financial crisis. Their conditional term structure is always decreasing. This finding is consistent with the properties of the term structure of option risk premia estimated for option-implied risks X_{11} and X_{22} in Section 3.2.4.[34] Table B.4 of the Appendix shows that the term structure of diffusive variance risk premia is completely spanned by option-implied components X_{11} and X_{12} alone, with a contribution of low-frequency component X_{11} that monotonically increases with the horizon and largely dominates the contribution of component X_{12}. These features are well reflected in the dynamics of the slope of the term structure of diffusive VIX variance risk premia, which basically only reflects variations in option-implied component X_{11}.[35] This explains both the decreasing term structure of diffusive VIX variance risk premia and its low frequency dynamics.

The term structure of jump variance risk premia is slightly decreasing most of the time, but it can be strongly upward sloping in a number of economically relevant cases, i.e., for about 28% of the observations in our sample. While the absolute differences in

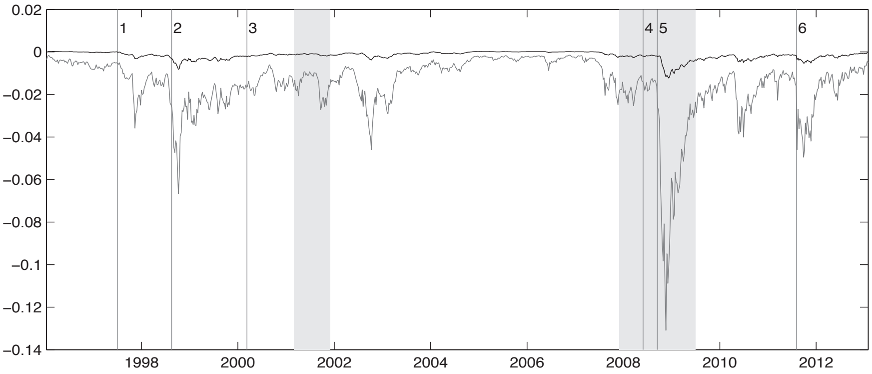
---

[33] In Table B.4 of the Appendix, the loading of component X_{11} on the slope of the term structure of $VRP_t(\tau)$ is -0.574, while the loading of X_{22} and X_{12} are 0.121 and 0.325, respectively.

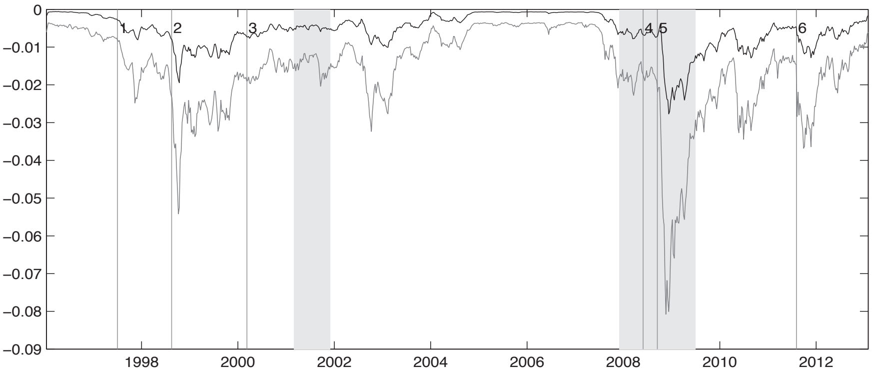
[34] $VRP_t^c(\tau)$ is the sum of the market prices of option-implied components X_{11} and X_{22}, which together span the diffusive variance $tr(X_t) = X_{11t} + X_{22t}$. As shown in Section 3.2.4, the term structure of the market price of the smile for option-implied components X_{11} and X_{22} is decreasing.

[35] In Table B.4 of the Appendix, the loading of component X_{11} on the slope of the term structure of $VRP_t^c(\tau)$ is -0.314, while the loading of X_{12} is only 0.004.

Panel A: 3 months diffusive and jump VIX variance risk premium



Panel B: 12 months diffusive and jump VIX variance risk premium



Panel C: Term structure of diffusive and jump VIX variance risk premia

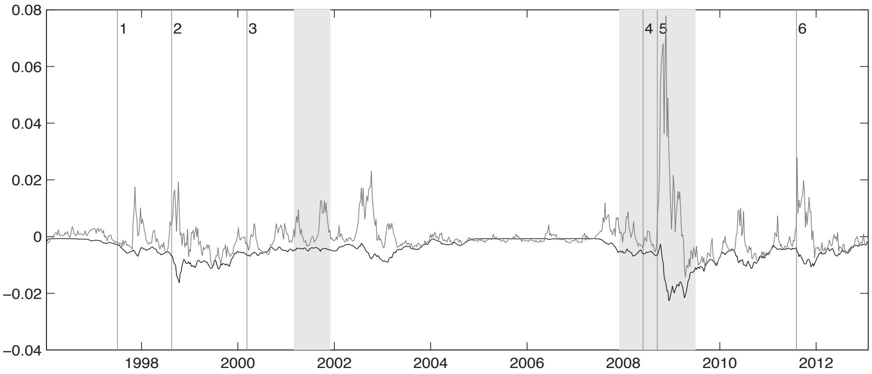


Figure 3.6: Diffusive and jump VIX variance risk premia $VRP_t^c(\tau)$ (black line) and $VRP_t^d(\tau)$ (grey line). Panel A: 3 months horizon; Panel B: 12 months horizon. In panel C, we plot the slope of the model-implied term structure of diffusive and jump VIX variance risk premia, computed as the difference of 12 months and 3 months risk premia. Grey areas highlight NBER recessions; vertical lines indicate crisis events, listed in Table 2.5.

the spread between 12 months and 3 months jump variance risk premia are small when the term structure is decreasing, they can be very substantial when the term structure is increasing. The largest absolute jump variance risk premia arise for horizon of $\tau = 1$ month, for which they can be as large as about 13% squared during the 2008 financial crisis. In contrast, jump variance risk premia for maturity $\tau = 12$ months are never larger than 8% squared in absolute value. The most prominent cases, in which the term structure of jump variance risk premia is increasing, arise in periods of market distress, e.g., immediately before the Lehmann default, with a spread of about +7% squared between 12 months and 3 months risk premia, or in correspondence of the US default, with a spread of almost +3% squared between 12 months and 3 months risk premia. In such periods, when the positive slope of the term structure of jump variance risk premia escalates, the term structure of variance risk premia is also upward sloping.

Table B.4 of the Appendix shows that the term structure of jump variance risk depends on all option-implied components. While the contribution of low-frequency component X_{11} increases with the horizon, the contributions of X_{12} and X_{22} decreases. Overall, these features produce a slope of the term structure of jump variance risk premia that decreases with X_{11} and increases with both X_{11} and X_{22}.[36] Therefore, in contrast to the term structure of diffusive variance risk premia, the term structure of jump variance risk premia reflects both high and low frequency option-implied risks.

Pure Jump Variance Risk Premia

A proper intuition for the contribution of low- and high-frequency risks to the dynamics of the term structure of jump variance risk premia is gained, by recalling that jump variance risk premia are the sum of a jump intensity risk premium and a pure jump variance risk premium:

$$VRP_t^d(\tau) = VRP_t^{dc}(\tau) + VRP_t^{dj}(\tau) \ . \tag{3.32}$$

$VRP_t^{dc}(\tau)$ is a risk premium linked to unexpected future variations in the probability of a jump in index returns. Therefore, it is spanned by the components of the market price of the smile. $VRP_t^{dj}(\tau)$ is proportional to a (constant) risk premium $E^{\mathbb{P}}[\mathcal{E}(1 + k^*)] - E^{\mathbb{Q}}[\mathcal{E}(1 + k)]$ for instantaneous jump variance risk, scaled by the (time-varying) expected average jump intensity under the physical probability:

$$VRP_t^{dj}(\tau) = 2(E^{\mathbb{P}}[\mathcal{E}(1 + k^*)] - E^{\mathbb{Q}}[\mathcal{E}(1 + k)])tr\left(\Lambda E^{\mathbb{P}}\left[\frac{1}{\tau}\int_t^{t+\tau} X_s ds\right]\right) \ . \tag{3.33}$$

---

[36] In Table B.4 of the Appendix, the loading of component X_{11} on the slope of the term structure of $VRP_t^d(\tau)$ is -0.356, while the loading of X_{22} and X_{12} are 0.536 and 0.206, respectively.

Therefore, $VRP_t^{dj}(\tau)$ is a premium for pure jump variance risk over horizon τ, which is spanned by the expected average probability of future jumps in index returns.

We can illustrate more concretely the above intuition, by plotting in Figure 3.7 both components of the model-implied jump variance risk premium in equation (3.32), together with the slope of their term structure, using the closed-form expression in equation (3.24).

Figure 3.7 shows that the term structure of $VRP_t^{dj}(\tau)$ is increasing for about 72% of the observations in our sample, while in all other cases it is virtually flat. In contrast, the term structure of $VRP_t^{dc}(\tau)$ is always decreasing. Similarly to diffusive variance risk premia, the variance risk premium component $VRP_t^{dc}(\tau)$ is spanned by the components of the market price of the smile. Therefore, it implies a decreasing term structure, with dynamics spanned by option-implied risk premium components X_{11} and X_{12} alone. In contrast, the structure of $VRP_t^{dj}(\tau)$ premia is spanned by all option-implied components X_{11}, X_{12} and X_{22}, because at the estimated parameters jump intensities follow a multi-frequency process that is a function of all option implied risks:

$$\lambda_t = 25.668 \cdot X_{11t} + 40.427 \cdot X_{12t} + 15.979 \cdot X_{22t} \ . \tag{3.34}$$

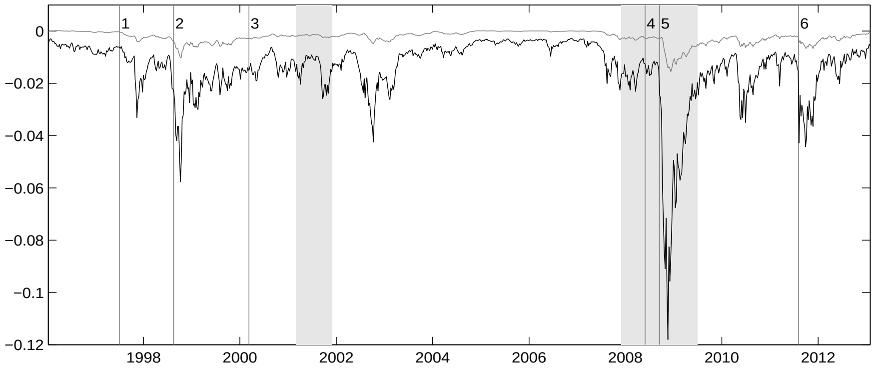
While the largest contribution to λ_t comes from option-implied skew component X_{12}, all option implied-risks are mean reverting under the physical probability, so that their contribution to expected average jump intensities decreases with the horizon. Such decay is fastest (slowest) for highest (lowest) frequency component X_{22} (X_{11}). These features induce a typically decreasing term structure of average expected jump intensities and an increasing term structure of pure jump variance risk premia, because the risk premium for instantaneous VIX jump variance risk is negative: $E^{\mathbb{P}}[\mathcal{E}(1+k^*)] - E^{\mathbb{Q}}[\mathcal{E}(1+k)] < 0$. The slope of the term structure of pure jump variance risk premia escalates in periods of market distress, because in those periods jump intensities contain a large high-frequency component, generated by large option-implied risks X_{12} and X_{22}. Table B.4 of the Appendix confirms this intuition, showing that the slope of the term structure of pure jump variance risk premia mostly depends on option-implied component X_{12}, while option implied risks X_{11} and X_{22} contribute significantly, but to a smaller extent.[37]

In summary, we find that the multifrequency dynamics of the term structure of variance risk premia is explained by the interplay of two economically distinct components. A downward sloping component, with dynamics spanned only by medium- and low-frequency option-implied risk premium factors, and an upward sloping component, with multifrequency dynamics spanned by all option-implied risks. While the

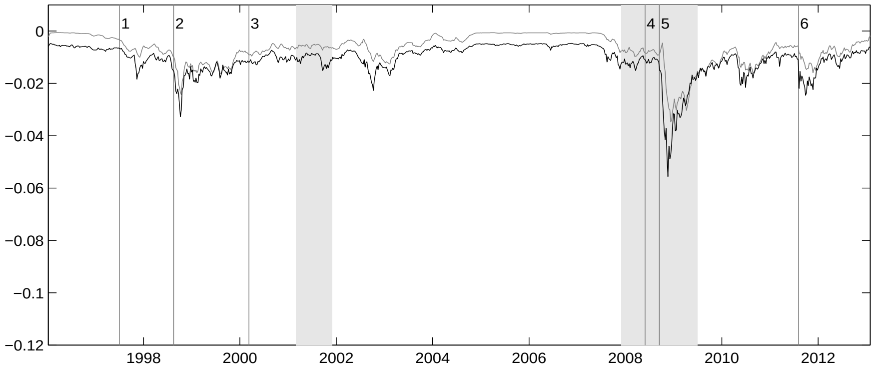
---

[37] In Table B.4, the loadings of X_{11}, X_{12} and X_{22} for the slope of the term structure of pure jump variance risk premia are 0.147, 0.480 and 0.206, respectively.

Panel A: 3 months intensity and pure jump VIX variance risk premium



Panel B: 12 months intensity and pure jump VIX variance risk premium



Panel C: Term structure of intensity and pure jump VIX VRP

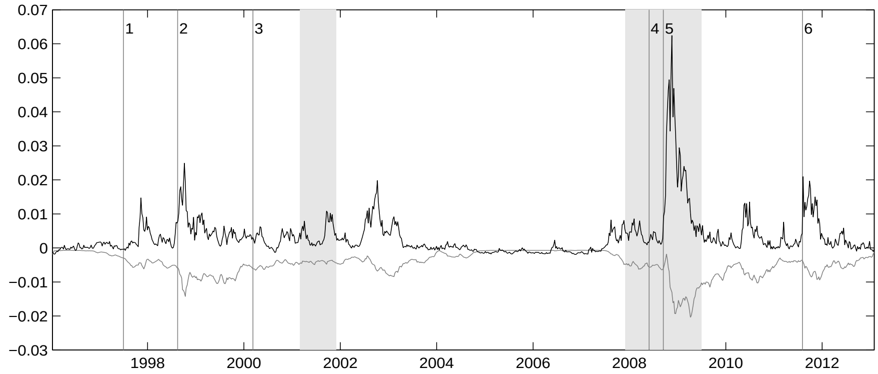


Figure 3.7: Intensity and pure jump VIX variance risk premia $VRP_t^{dc}(\tau)$ (grey line) and $VRP_t^{dj}(\tau)$ (black line). Panel A: 3 months horizon, Panel B: 12 months horizon. In panel C, we plot the slope of the model-implied term structure of intensity and pure jump VIX variance risk premia, computed as the difference of 12 months and 3 months risk premia. Grey areas highlight NBER recessions; vertical lines indicate crisis events, listed in Table 2.5.

first component mostly captures the persistent dynamics of the horizon-dependent price of option-implied market insurance, the second component more directly reflects the multi-frequencies character of expectations about future jump risk.

3.2.7 Predictive Ability of Option-Implied Risks and Risk Premia

Intuitively, the information encoded from the option-implied volatility surface can help to forecast future volatility risks, future excess returns on (VIX) option volatility portfolios or future index returns. Since option-implied risks X_{11}, X_{22}, X_{12} can be computed virtually in real time, they provide a feasible instrument for out-of-sample prediction. We study the predictive power of option-implied risks X_{11}, X_{22}, X_{12} for future VIX realized variance, future VIX option portfolio payoffs and future S&P500 index returns. We compare the predictive ability of model-implied variables with the one of nonparametric option-implied proxies of the smile, capturing the level of the implied volatility, the term structure of at-the-money implied volatilities and the one-month option-implied skew, respectively. This comparison has three purposes. First, it allows us to assess the added-value of our model for parsimoniously identifying relevant option-implied state variables, having predictive power for future risks or risk premia. Second, it allows us to test the specification of model-implied risks and risk premia. Third, its allows us to study the relation between option-implied risks and conditional market risk premia.

Predictability of Realized Variance

Figure 3.8 collects predictive regression results for future (VIX) realized variance, over forecasting horizons from $\tau = 1$ month to $\tau = 12$ months.

We find that realized variance is highly predictable over short horizons, with predictive regression R^2s up to 55% for one-month ahead forecasts, using model-implied risks X_{11}, X_{22} and X_{12}. At longer horizons, e.g., between nine and twelve months, the model-implied predictive R^2s drop to about 15%. Overall, this evidence produces a decreasing term structure of predictive R^2s. We find that the predictive power of model-implied components is virtually identical to the predictive power of nonparametric option-implied proxies, indicating that the model-implied proxies provide an accurate summary of option-implied risks. Moreover, the degree of predictability of the realized variance is similar in- and out-of-sample, implying a quite stable relation between future realized variance and option-implied risks over time. At short horizons of 1 month, about 80% of the large predictive R^2 is generated by the highest-frequency option-implied risk X_{22}, while the remaining 20% is almost completely due to the pure-jump volatility component X_{12}. At forecasting horizons of 12 months, two third of the lower predictive R^2 is still generated by X_{22}, while the remaining part is again due to X_{12}.

The term structure of predictive R^2s in Figure 3.8 is understood by recalling the

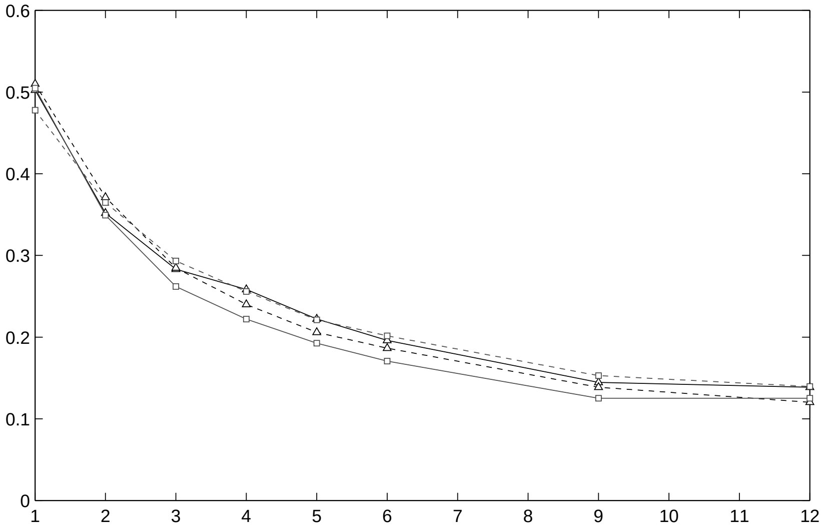


Figure 3.8: Predictive regression for VIX realized variance. We regress the future VIX realized variance for horizons between 1 month and 12 months (x-axis) on (i) the three option-implied components $X_{11,t}, X_{22,t}, X_{12,t}$ (triangles) and (ii) three standard nonparametric measures of option-implied level, skew and volatility term structure for the volatility surface (squares). Full lines correspond to in-sample R^2s for the in-sample period 1996-2002; dashed lines correspond to out-of-sample R^2s for the period 2003-2013/01. For both R^2 computations, model-implied and predictive regression parameters are fixed to the 1996-2002 point estimates. The calculation method for the option-implied skewness and term structure is explained in Appendix B.3.

structure of the average realized variance $RV_{t+\tau}$ in equation (3.21). The model-implied predictable part of $RV_{t+\tau}$ is given in closed-form by:

$$E_t^{\mathbb{P}}[RV_{t+\tau}] = tr\left((id_{2\times 2} + 2E^{\mathbb{P}}[\mathcal{E}(1+k^*)]\Lambda)E_t^{\mathbb{P}}\left[\frac{1}{\tau}\int_t^{t+\tau} X_s ds\right]\right) , \tag{3.35}$$

where $id_{2\times 2}$ is the 2×2 identity matrix, and it is a linear function of expected average integrated risks X_{11}, X_{12} and X_{22}. Note that at the model parameter estimates all option-implied risks X_{11}, X_{22} and X_{12} contribute to the expected realized variance. However, the contribution of X_{12} and X_{22} dominates, especially in the expected jump realized variance. Therefore, these risks generate the largest fraction of the predictive R^2. Moreover, while all option-implies risks are mean reverting, X_{11} and X_{12} are the least persistent risks, with an half-life of about one month and one quarter, respectively. Therefore, their predictive power for future realized variance decays quite rapidly with the forecasting horizon, from an R^2 of about 55% at horizons of one month to an R^2 of about 15% at horizons of one year. These features are directly related to the observed decreasing term structure of predictive R^2s.

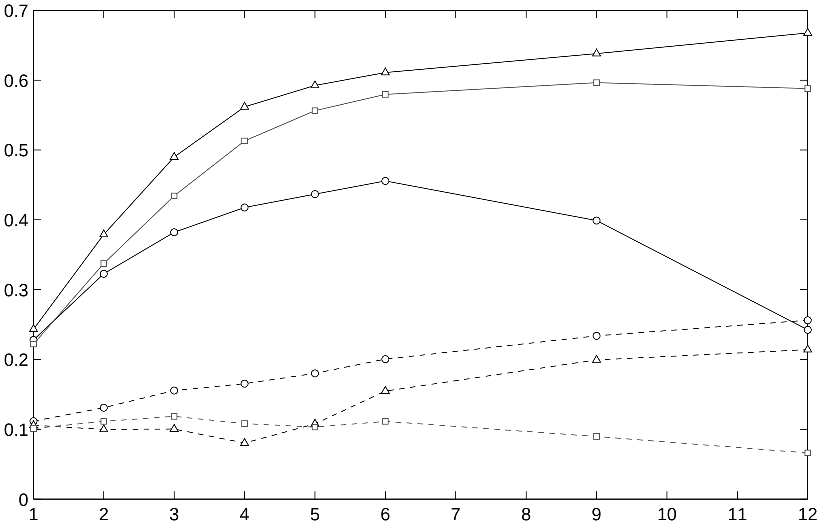


Figure 3.9: Predictive regression for VIX option portfolio payoffs. We regress future VIX option portfolio payoffs for horizons between 1 month and 12 months (reported on the x-axis) on (i) the model-implied VIX variance risk premium (circles), (ii) the three option-implied components $X_{11,t}, X_{22,t}, X_{12,t}$ (triangles) and (iii) three standard nonparametric measures of the option-implied level, skew and volatility term structure for the option-implied volatility surface (squares). Full lines correspond to in-sample R^2s for the in-sample period 1996-2002; dashed lines correspond to out-of-sample R^2s for the period 2003-2013/01. For each R^2 computation, model-implied and predictive regression parameters are fixed at the 1996-2002 point estimates.

Predictability of VIX Option Portfolio Payoffs

We next study the predictability of future VIX option portfolio payoffs, defined as the difference of realized variance $RV_{t+\tau}$ and the variance swap rate $E_t^{\mathbb{Q}}[RV_{t+\tau}]$. By definition, the expected payoff of a VIX option portfolio is the variance risk premium for horizon τ. Therefore, in this predictive regression we try to identify option-implied risks that act as variance risk premium factors. Besides studying unconstrained predictive regressions with model-implied and nonparametric proxies of the smile as predictive variables, we also quantify the direct predictive power of model-implied variance risk premia. This is an obvious test for our specification of variance risk premia. Figure 3.9 summarizes the resulting predictability evidence.

We find economically large in-sample R^2s for all predictive regressions. The largest in-sample predictive power emerges for the unconstrained regressions based on model-implied predictive variables. Such in-sample R^2s range between 25% and 60%, with a peak at the 12 months forecasting horizons. The in-sample R^2s of unconstrained predictive regressions with nonparametric proxies of the smile are only slightly lower than for the model-implied predictive variables and they peak at a forecasting horizon of about 9 months. The in-sample R^2s associated with model-implied variance risk premia range between about 25% and 45% and peak at a 6 months forecasting horizon. Interestingly, the unconstrained predictive regression produces and in-sample predictive power that is almost exclusively generated by the two more persistent model-implied risks X_{11} and X_{12}, i.e., the two risk premium factors that span option risk premia.[38] The incremental in-sample predictive power of the unconstrained predictive regression, relative to the model-implied variance risk premium regression, is virtually completely explained by a larger estimated loading of the lowest-frequency option-implied risk premium factor X_{11}. Different estimated factor loadings are possible, because the unconstrained predictive regression coefficients are estimated individually across different horizons from $\tau = 1$ month to $\tau = 12$ months, without restricting them to satisfy cross-sectional arbitrage-free restrictions, as they are in the model-implied variance risk premium. While in-sample this additional degree of freedom can produce an improved forecasting power, it is important to verify whether such additional flexibility does not come at the cost of an overfitted predictive relation. The results of our out-of-sample predictive regressions confirm the economic relevance of model-implied proxies of the smile as variance risk premium factors. They also provide additional support for the specification of variance risk premia in our model. We find that the largest out-of-sample predictive R^2s are obtained using model-implied variance risk premia. They range between 12% and 24% and imply a monotonically increasing term structure. The out-of-sample R^2s of unconstrained predictive regressions using model-implied risks are smaller, especially for forecasting horizons below one quarter, while the predictive regressions with nonparametric proxies of the smile imply a low out-of-sample predictive power, which decreases with the forecasting horizon.

In summary, these findings suggest that the large in-sample predictive power of unconstrained predictive regressions for VIX variance payoffs is partly the consequence of overfitting. At the same time, they indicate that the arbitrage-free constraints implied for model-implied variance risk premia are supported by the in-sample and the out-of-sample evidence. Model-implied variance risk premia explicitly reflect the composition of variance risk premia as the sum of a diffusive and a jump variance risk premium, where the last premium consists of an intensity and a pure jump variance risk premium:

$$VRP_t(\tau) = VRP_t^c(\tau) + VRP_t^{dc}(\tau) + VRP_t^{dj}(\tau) \ . \tag{3.36}$$

---

[38] The high-frequency component X_{22} has a small marginal contribution to the predictive power, because of its small unconstrained predictive regression parameter estimate. Indeed, unconstrained predictive regressions with predictive variables X_{11} and X_{12} alone produce virtually the same term structure of in-sample predictive R^2s in a full regression with the additional option-implied component X_{22}.

We find that while $VRP_t^c(\tau)$ and $VRP_t^{dc}(\tau)$ are spanned by two mid- and low-frequency option-implied risk premium factors X_{11} and X_{12}, $VRP_t^{dj}(\tau)$ is proportional to expected future jump probabilities and depends on all option-implied risks X_{11}, X_{22} and X_{12}. The low persistence of components X_{22} and X_{12} implies an increasing term structure for $VRP_t^{dj}(\tau)$ during periods of market distress, which can imply an increasing term structure of variance risk premia. Even though the contribution of high-frequency option-implied risks to VIX option portfolio payoffs might appear negligible, based on in-sample predictive regression results, we find that it is important to identify the relation between high-frequency risks and jump variance risk premia, in order to capture the multi-frequency dynamics of the term structure of variance risk premia.

Predictability of S&P500 Index Excess Returns

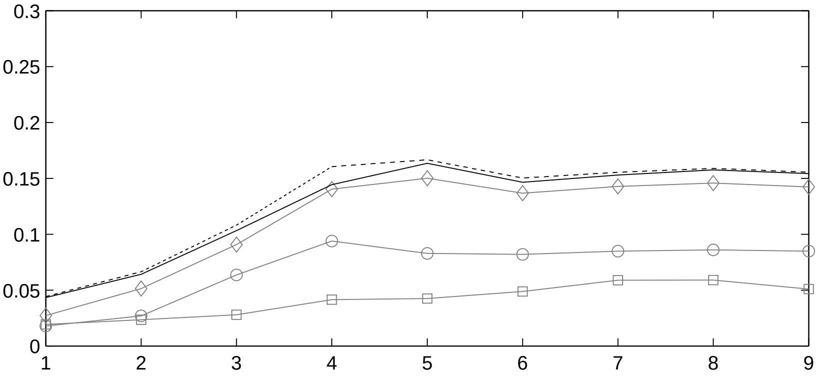
We finally address the predictability of S&P500 index excess returns $r_{t+\tau}^e$, for prediction horizons τ from one month to one year. We consider unconstrained predictive systems that include estimated option-implied risks X_{11}, X_{12} and X_{22} as predictive variables. In order to account parsimoniously also for a potential nonlinearity in the predictive relation, we estimate a simple threshold-linear regression, with an endogenous threshold $T \in [0,1]$ for the composition $X_{11t}/(X_{11t}+X_{22t})$ of the diffusive volatility:

$$r_{t+\tau}^e = [\alpha_l + tr(\beta_l X_t)]\mathbb{I}_{\{[X_{11t}/tr(X_t)]<T]\}} + [\alpha_h + tr(\beta_h X_t)]\mathbb{I}_{\{[X_{11t}/tr(X_t)]\geq T]\}} + \epsilon_{t+\tau} \quad (3.37)$$

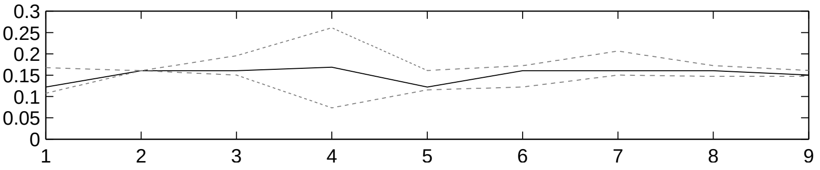
In this equation, $\mathbb{I}_{\{.\}}$ is an indicator function, while α_u and symmetric 2×2 matrix β_u contain regime-dependent regression intercepts and slope parameters, associated with model-implied conditional state $u = h$ $(u = l)$ of more (less) persistent volatility: $[X_{11t}/tr(X_t)] \geq T$ $([X_{11t}/tr(X_t)] < T)$. In this way, we achieve two goals at the same time. First, we can estimate the predictive relation between market excess returns and option-implied risks with a more general specification, which directly allows us to test in a simple way the correct specification of an affine relation between market risk premia on option-implied risks. Second, we can incorporate an economically plausible long-run risk channel for market risk premia, in which the market compensation for low-frequency volatility risks can depend on the frequency composition of the volatility. The estimation results are summarized in Figure 3.10.

We find a significant estimated threshold in the predictive relations, which is quite stable across horizons and samples and equal to about $T = 0.15$ $(T = 0.18)$ for the shorter (the longer) sample period. This finding is direct evidence of a nonlinear relation between market risk premia and option-implied risks, which is dependent on the frequency composition of the volatility. All parameters in the estimated threshold predictive regression are significant, and the degree of predictability is quite large. Over the full sample period, the predictive R^2's range between about 5% and 16%, with a peak at

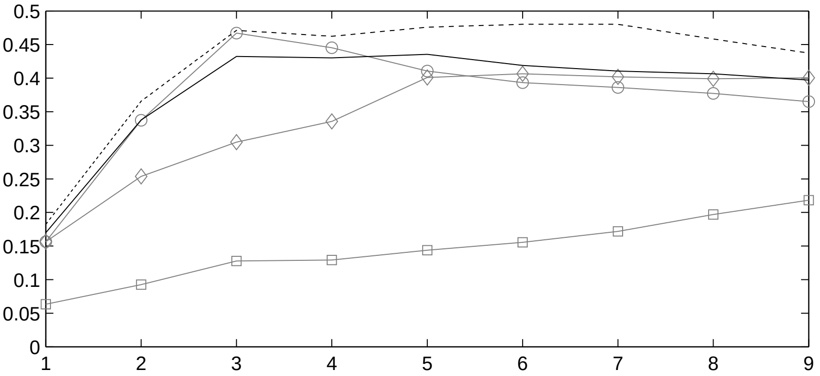
Panel A1: R^2 **1996-2013/01** (full sample)



Panel A2: threshold 1996-2013/01



Panel B1: R^2 **1996-2002** (in-Sample period)



Panel B2: threshold 1996-2002

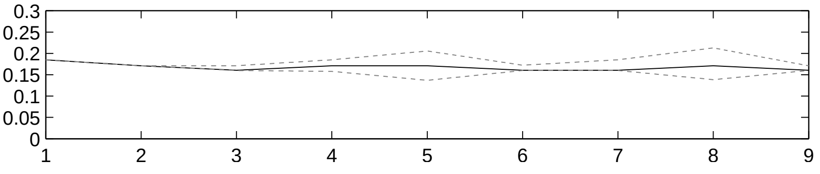


Figure 3.10: Predictive regression for future index excess returns. We perform a threshold regression of future realized S&P500 index excess returns for horizons between 1 month and 9 months (reported on the x-axis) on (i) the individual factors X_{11} (diamonds), X_{12} (circles), X_{22} (squares); (ii) our preferred model using the risk factors $\tilde{X}_{11}$ and $\tilde{X}_{12}$ (black line) and (iii) jointly all three state components (dashed line). In all regressions, we use $Y_{2t} := X_{11t}/Tr[X_t]$ as threshold variable. We report the R^2 of the predictive regressions in panels A1, B1 and estimates and 95% confidence intervals for the threshold in panels A2, B2.

	1mo	2mo	3mo	4mo	5mo	6mo	7mo	8mo	9mo
Out of sample	0.53	0.54	0.52	0.52	0.58	0.57	0.55	0.57	0.57
Conundrum	0.52	0.50	0.51	0.46	0.47	0.49	0.48	0.51	0.53
Crisis	0.63	0.60	0.50	0.63	0.85	0.85	0.86	0.87	0.88
Post crisis	0.49	0.57	0.56	0.55	0.58	0.54	0.47	0.47	0.42

Table 3.4: Out of sample sign correlations between realized excess returns of the S&P 500 index over horizons from 1 to 9 months and predicted returns from our threshold regression. We perform a predictive threshold regression for the in-sample period 1996-2002 and evaluate the signs of the predicted returns for the out of sample period 2003-Jan 2013 and three sub-periods: Conundrum (2003-2007), Financial Crisis (2008-2009) and Post-Crisis (2010-Jan 2013).

forecasting horizons of five months. Such predictive power is almost entirely generated by the mid-frequency and the low-frequency option-implied risks X_{11} and X_{12}.

Besides the detected degree of predictability, these findings highlight the difficulties in identifying a time-invariant affine relation between index risk premia and option-implied risks. From a different angle, they also support identification procedures of option-implied risks that do not rely on an affine specification of market risk premia. The economic implications of the estimated predictability for the 2003-2013 out-of-sample period are illustrated in Table 3.4, based on the sign correlations between realized returns and predicted equity premia implied by the 1996-2002 in-sample predictive regression estimates.

These correlations are all above 50% over the full sample, with a peak of 58% at the five months horizons. They are essentially indistinguishable from 50% in the conundrum period, when volatility was almost completely transient. In contrast, they can be as large as about 85% in the financial crisis period, when volatility was often very persistent, indicating an economically relevant predictive power of option-implied risk premium factors in persistent volatility states. We find that a simple trading strategy that goes long (short) the index when the predicted equity premium is positive (negative), outperforms a static market investment out-of-sample, with, e.g., a Sharpe ratio (a sample skewness) for horizons of one quarter that is about double (much less negative) than the one implied by a static index investment. Figure B.6 and Table B.5 of the Appendix provide a graphical description and a complete summary of these results, for different horizons and across different out-of-sample periods.

3.3 Conclusions and Outlook

We estimate the dynamics and the market prices of the hidden option-implied risks driving the implied volatility surface of S&P500 options, in order to study their link with the term structures of option-implied and variance risk premia. We identify option-implied risks, their market prices and the term structure of variance risk premia, following a two-step procedure that does not rely on direct information about S&P500 index returns, in which we estimate a parsimonious three-factor matrix jump diffusion characterized by (i) interdependent, mutually-exciting, risks, (ii) a skewness component disconnected from volatility and (iii) a compensation for variance risk that varies independently of the volatility.

Besides providing an excellent pricing performance and fit, deriving from an improved ability in generating a volatility-unrelated variability of option-implied skews, our model yields a sharp identification of option-implied risks and risk-premia, based on three interdependent volatility components that imply a useful multi-frequency representation of the term structure of variance risk premia. The least persistent option-implied risk largely correlates with the thirty days at-the-money option-implied volatility. The other two more persistent option-implied risks correlate with a 30-days option-implied skew residual and with an option-implied skew term structure residual, respectively, in a way that is partly disconnected from at-the-money implied-volatility shocks. We find that option-implied risk premia are spanned by the two persistent option-implied skew components alone, disclosing their interpretation as option-implied risk premium factors.

We then address the relation between option-implied risks, option-implied risk premia and the term structure of variance risk premia. We find that the dominating role of low-frequency option-implied risks for the market price of the smile implies a downward sloping term structure of option-implied risk premia. Time-variation of all option-implied risks also implies highly time-varying and unambiguously negative variance risk premia, as well as a multi-frequency term structure of variance risk premia, which can be strongly upward sloping for short periods of time, in a number of economically relevant cases. The inversion of the term structure of variance risk premia is understood by the interplay of the term structures of option-implied risks and risk premia. While the first term structure is linked to an increasing term structure of pure jump variance risk premia, the second one implies a decreasing term structure of diffusive and jump intensity variance risk premia. In phases of market distress, the term structure of pure jump variance risk premia can be pronouncedly upward sloping, when high-frequency option-implied risks escalate, and have a dominating effect on the slope of the term structure of variance risk premia.

Finally, we find a significant predictive power of option-implied risks for future S&P500 volatility and index returns, with a dominating contribution to the predictive power deriving from the mid-frequency and low-frequency option risk premium factors. While an affine specification of variance risk premia is supported by our predictability findings, affine specifications of market risk premia are dominated by nonlinear specification that allow the predictive relation to depend on the composition of the volatility. Consistent with a long-run risk explanation for market risk premia, a simple trading strategy, balancing the risk premium predictions of mid-and low-frequency option-implied risks with the frequency-composition of the volatility, produces economically large out-of-sample excess returns.

In summary, our findings indicate that a volatility specification with three interdependent risks, capturing the different dimensions of the option-implied price of market volatility, can produce a parsimonious joint representation of the market price of the smile and the term structure of variance risk premia. Such a specification can also help to improve the identification of high-frequency option-implied risks in the multi-frequency dynamics of variance risk premia, complementing the results of standard predictive regressions. Lastly, while option-implied risks exhibit and important correlation with information about future market returns, it is difficult to identify their precise relation with conditional market risk premia, using an affine joint dynamics for returns and volatility. This feature motivates identification procedures for option and variance risk premia that do not depend on complete specification of market risk premia.

Appendix B

Appendix for "The Price of the Smile and Variance Risk Premia"

B.1 Additional Results in the Matrix AJD Model

B.1.1 Pricing Transform in the Matrix AJD Model

Under Assumption 11 and Assumption 16, the closed-form exponentially affine risk-neutral transform for $Y_T := \log(S_T)$ is given by:

$$\Psi(\tau;\gamma) := E_t\left[\exp\left(\gamma Y_T\right)\right] = \exp\Big(\gamma Y_t + tr\big[A(\tau)X_t\big] + B(\tau)\Big)\ , \tag{B.1}$$

where $\tau = T - t$, $A(\tau) = C_{22}(\tau)^{-1}C_{21}(\tau)$ and the 2×2 matrices $C_{ij}(\tau)$ are the $ij-$th blocks of the matrix exponential:

$$\left(\begin{array}{cc} C_{11}(\tau) & C_{12}(\tau) \\ C_{21}(\tau) & C_{22}(\tau) \end{array}\right) = \exp\left[\tau\left(\begin{array}{cc} M+\gamma Q'R & -2Q'Q \\ C_0(\gamma) & -(M'+\gamma R'Q) \end{array}\right)\right]\ . \tag{B.2}$$

The explicit expressions for the 2×2 matrix C_0 is:

$$C_0(\gamma) = \frac{\gamma(\gamma-1)}{2}I_2 + \Lambda\left[\Theta^Y(\gamma) - 1 - \gamma\Theta^Y(1)\right]\ , \tag{B.3}$$

and real-valued function $B(\tau)$ is given by:

$$\begin{aligned} = \ & \tau\Big\{(\gamma-1)r + \lambda_0\big[\Theta^Y(\gamma) - 1 - \gamma\Theta^Y(1)\big]\Big\} \\ & -\frac{\beta}{2}tr[\ln(C_{22}(\tau)) + \tau(M'+\gamma R'Q)] \end{aligned} \tag{B.4}$$

where $\ln(\cdot)$ is the matrix logarithm and $\Theta^Y(\gamma)$ is the univariate Laplace transform of the return jump size distribution. In the case of the double exponential distribution,

$$\Theta^Y_{DX}(\gamma) = \frac{\lambda^+\lambda^-}{\lambda^+\lambda^- + \gamma(\lambda^+ - \lambda^-) - \gamma^2} \ ,$$

in the case of the lognormal distribution

$$\Theta^Y_{LN}(\gamma) = (1+\overline{k})^\gamma \exp\left(\gamma(\gamma-1)\frac{\delta^2}{2}\right) \ ,$$

see, e.g., Leippold and Trojani (2008).

B.1.2 VIX Variance Risk Premium in the Matrix AJD Model

The affine expression for the variance risk premium in Proposition 2 is obtained by recalling the relations:

$$\begin{aligned} VRP_t(\tau) &= tr\left((E^{\mathbb{P}}_t - E^{\mathbb{Q}}_t)\left[\frac{1}{\tau}\int_t^{t+\tau} X_s ds\right]\right) + (E^{\mathbb{P}}_t - E^{\mathbb{Q}}_t)\left[\frac{1}{\tau}\int_t^{t+\tau} (dS_s/S_{s-})^2\right] \\ &= tr\left((E^{\mathbb{P}}_t - E^{\mathbb{Q}}_t)\left[\frac{1}{\tau}\int_t^{t+\tau} X_s ds\right]\right) \\ &\quad + E^{\mathbb{Q}}[\mathcal{E}(1+k)] tr\left(\Lambda(\beta^*_\Lambda E^{\mathbb{P}}_t - E^{\mathbb{Q}}_t)\left[\frac{1}{\tau}\int_t^{t+\tau} X_s ds\right]\right) . \end{aligned}$$

This shows that $VRP_t(\tau)$ is the sum of two-affine functions of state X_t. To compute these functions in closed-form, we need to compute the $\mathbb{P}$ and $\mathbb{Q}$ expectation of the average integrated state X in our model. These expectations are available in closed-form:

$$E^{\mathbb{Q}}_t\left[\frac{1}{\tau}\int_t^{t+\tau} X_s ds\right] = X^{\mathbb{Q}}_\infty + \frac{1}{\tau}\int_0^\tau e^{Mu}(X_t - X^{\mathbb{Q}}_\infty)e^{M'u} du \ , \tag{B.5}$$

where the long-run mean $X^{\mathbb{Q}}_\infty$ is the unique solution of the Lyapunov equation $MX^{\mathbb{Q}}_\infty + X^{\mathbb{Q}}_\infty M' = \beta Q'Q$. Similarly,

$$E^{\mathbb{P}}_t\left[\frac{1}{\tau}\int_t^{t+\tau} X_s ds\right] = X^{\mathbb{P}}_\infty + \frac{1}{\tau}\int_0^\tau e^{M^*u}(X_t - X^{\mathbb{P}}_\infty)e^{M^{*\prime}u} du \ , \tag{B.6}$$

where $X^{\mathbb{P}}_\infty$ is such that $M^*X^{\mathbb{P}}_\infty + X^{\mathbb{P}}_\infty M^{*\prime} = \beta^* Q'Q$. This implies, for any 2×2 matrix D:

$$\begin{aligned} tr\left(DE^{\mathbb{Q}}_t\left[\frac{1}{\tau}\int_t^{t+\tau} X_s ds\right]\right) &= tr\left(D(X^{\mathbb{Q}}_\infty + A^{\mathbb{Q}}_\tau(X_t - X^{\mathbb{Q}}_\infty))\right) \ , \\ tr\left(DE^{\mathbb{P}}_t\left[\frac{1}{\tau}\int_t^{t+\tau} X_s ds\right]\right) &= tr\left(D(X^{\mathbb{P}}_\infty + A^{\mathbb{P}}_\tau(X_t - X^{\mathbb{P}}_\infty))\right) \ , \end{aligned}$$

where, for any 2×2 matrix H:

$$A_\tau^{\mathbb{Q}}(H) := \frac{1}{\tau}\int_0^\tau e^{Mu} H e^{M'u} du \ ; \ A_\tau^{\mathbb{P}}(H) := \frac{1}{\tau}\int_0^\tau e^{M^*u} H e^{M^{*\prime}u} du \ .$$

Since these two functions are linear in H, the VIX variance risk premium is affine in X_t. This conclude the proof. $\square$

B.1.3 Stochastic Discount Factor in the Matrix AJD Model

Existence of a well-defined stochastic discount factor to price all shocks in our model is ensured by a proper density for an equivalent change of measure, from the physical to the risk neutral probability. To this end, we specify matrix processes $\{\Gamma_{1t}\}$, $\{\Gamma_{2t}\}$ for the market prices of Brownian shocks dW_t^*, dB_t^*, and an appropriate distribution for return jumps. Following Assumption 11, we specify a double exponential distribution for log return jumps, with parameters $\lambda^{+*}, \lambda^{-*}$ and λ^+, λ^-, respectively, under the physical and the risk neutral probabilities.

We show that, under Assumption 14, a proper density process consistent with these properties is defined for any $T \geq 0$ by:

$$\begin{aligned} \left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_T} = \ & \exp\left\{ tr\left(-\int_0^T \Gamma_{1t} dW_t^* + \frac{1}{2}\int_0^T \Gamma_{1t}'\Gamma_{1t} dt - \int_0^T \Gamma_{2t} dB_t^* + \frac{1}{2}\int_0^T \Gamma_{2t}'\Gamma_{2t} dt \right) \right\} dt \\ & \times \prod_{i=1}^{N_T^*} \exp\left\{ -(\lambda^- - \lambda^{*-}) J_i^{*-} - (\lambda^+ - \lambda^{*+}) J_i^{*+} + \ln\left(\frac{1/\lambda^{*-} + 1/\lambda^{*+}}{1/\lambda^- + 1/\lambda^+} \right) \right\} , \end{aligned} \tag{B.7}$$

where

$$\Gamma_{1t} = \sqrt{X_t}\Gamma + \frac{1}{2\sqrt{X_t}}(\beta^* - \beta)Q' \ , \tag{B.8}$$

and

$$\Gamma_{2t} = \sqrt{X_t}\Delta + \frac{\mu_0 - (r-q)}{\sqrt{X_t}} \ , \tag{B.9}$$

with $\mu_0 - (r-q) \geq 0$ and Δ a 2×2 parameter matrix. The first (second) line of equality (B.7) defines a possible change of measure for diffusive (jump) shocks in our model.

Under Assumption 14, the stochastic exponential in the first line of (B.7) is a well-defined positive local martingale, and hence a supermartingale. Therefore, to show that this term is a martingale, it is enough to show that it has a constant expectation:

$$1 = E_0^{\mathbb{P}}\left[\exp\left\{tr\left(-\int_0^T \Gamma_{1t}dW_t^* + \frac{1}{2}\int_0^T \Gamma_{1t}'\Gamma_{1t}dt - \int_0^T \Gamma_{2t}dB_t^* + \frac{1}{2}\int_0^T \Gamma_{2t}'\Gamma_{2t}dt\right)\right\} dt\right] .$$

In our matrix AJD setting, this property does not follow from a standard Novikov-type condition. However, it follows from a localization argument; see, e.g., (Meyerhofer 2014).

We now show that the second line of (B.7) also defines a martingale process. Using the independence between IID log jump sizes J^* and counting process N^* under the physical probability, it is enough to show that:

$$1 = E_0^{\mathbb{P}}\left[\exp\left\{-(\lambda^- - \lambda^{*-})J^{*-} - (\lambda^+ - \lambda^{*+})J^{*+}\right\} \frac{1/\lambda^{*-} + 1/\lambda^{*+}}{1/\lambda^- + 1/\lambda^+}\right] . \quad \text{(B.10)}$$

Explicit calculations of the expectation on the right hand side yield:

$$\frac{\lambda^{*-}\lambda^{*+}}{\lambda^{*-} + \lambda^{*+}} \cdot \frac{1/\lambda^{*-} + 1/\lambda^{*+}}{1/\lambda^- + 1/\lambda^+} \int_{-\infty}^{\infty} \exp\left(-\lambda^- J^{*-} - \lambda^+ J^{*+}\right) dJ^* = 1 .$$

With respect to the risk-neutral probability $\mathbb{Q}$, log return jumps follows a double exponential distribution with parameters λ^- , λ^+. Indeed, for any $u \in \mathbb{R}$ it follows:

$$E^{\mathbb{Q}}\left[\exp(uJ)\right] = \frac{\lambda^-\lambda^+}{\lambda^- + \lambda^+} \int_{-\infty}^{\infty} e^{uJ} e^{-\lambda^- J^- - \lambda^+ J^+} dJ ,$$

which is the Laplace transform of a double exponential distribution with parameter λ^-, λ^+. This concludes the proof.

□

B.2 Estimation procedure

B.2.1 First step: Kalman filter

The first estimation step is performed using a Kalman filter of the linearized process, using exclusively options in the observation equation. Thus we can estimate all risk-neutral parameters via the observation equation and the physical parameters of the state dynamics via the transition equation. We denote the set of all parameters estimated in the first step by $\theta := (M, Q, R, \beta, \lambda_0, \Lambda;\ M^*, \beta^*)$.

The physical dynamics of our state variable is given in (3.14):

$$dX_t = \left[\beta^* Q'Q + M^* X_t' + X_t M^{*\prime}\right] dt + \sqrt{X_t} dB_t^* Q + Q' dB_t^{*\prime} \sqrt{X_t}$$

We discretize this process on a weekly grid with $\Delta_k = 7$ calendar days. When there is no data for a given Wednesday, we skip the respective week and set $\Delta_k = 14$.

We initialize the filtered state $\widehat{X}_t$ to be the steady state $X_\infty^{\mathbb{P}}$, which can be computed by solving the Lyapunov equation $M^* X_\infty^{\mathbb{P}} + X_\infty^{\mathbb{P}} (M^*)' = Q'Q$. We initialize the variance matrix of $\widehat{X}_t$ as $\widehat{\Sigma}_0 = 0$. At each step, we compute exact expectations of mean and variance of $X_{t+\Delta}$ given X_t from the Laplace transform (B.1)

$$\begin{aligned}
\overline{X}_{t+\Delta} &= \beta\,\overline{\mu} + \Phi \widehat{X}_t \Phi' && \text{(B.11)}\\
\overline{V}_{t+\Delta} &= (I_4 + K_4)\left(\Phi \widehat{X}_t \Phi' \otimes \overline{\mu} + \beta\, \overline{\mu} \otimes \overline{\mu} + \overline{\mu} \otimes \Phi \widehat{X}_t \Phi'\right) && \text{(B.12)}\\
\text{with} & \\
\overline{\mu} &= -\frac{1}{2} C_{12} C_{11}' \\
\Phi &= e^{\Delta M^*} \\
C &= \exp\left[\Delta \begin{pmatrix} M^* & -2Q'Q \\ 0 & -(M^*)' \end{pmatrix}\right] = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}
\end{aligned}$$

where $C_{11}, C_{12}, C_{21}, C_{22}$ are all square 2×2 matrices and K_4 is the 4×4 commutation matrix. These calculations are used in the transition equation:

$$\widetilde{X}_{t+\Delta} = \overline{X}_{t+\Delta} \tag{B.13}$$

The predicted state $\widetilde{X}_{t+\Delta}$ is then used to compute the observation equations:

$$\widehat{O}_{t+\Delta,i} = O_{t+\Delta,i}(\widetilde{X}_{t+\Delta};\theta) + \varepsilon_{t+\Delta,i}, \qquad i = 1, \ldots, N_{t+\Delta} \tag{B.14}$$

where $\widehat{O}_{t+\Delta,i}$ denotes the Black-Scholes implied volatility of the $i-$th option on day $t + \Delta$, $N_{t+\Delta}$ the total number of options observed on that day, $O_{t+\Delta,i}(\widetilde{X}_{t+\Delta};\theta)$ the model-implied option prices and $\varepsilon_{t+\Delta,i}$ is an iid noise with zero mean and variance σ_r. We also allow for autocorrelation in the noise:

$$corr(\overline{\varepsilon}_{t+\Delta}, \overline{\varepsilon}_t) = \rho_r$$

where $\overline{\varepsilon}_t$ is the mean error over all options on day t.

We finally *update* the state using a linearization of the dynamics. We first linearize the transition equation and the observation equations by computing the Jacobian matrices:

$$\begin{aligned}
F &= \frac{\partial \overline{X}_{t+\Delta}}{\partial \widehat{X}_t} = \Phi \otimes \Phi \\
G_t &= \frac{\partial O_{t+\Delta}}{\partial \widetilde{X}_t}
\end{aligned}$$

where we applied the identity $\frac{\partial}{\partial X} BXC = C' \otimes B$ to obtain F, while G is calculated via numerical differentiation. The variance matrix of the state is:

$$\widetilde{\Sigma}_{t+\Delta} = F\widehat{\Sigma}_t F' + \overline{V}_{t+\Delta} \tag{B.15}$$

Finally we update the state and the variance matrix to be used in the next step:

$$\begin{aligned} S_t &= G_t \widetilde{\Sigma}_{t+\Delta} G_t' + \sigma_r^2 I_2 \\ H_t &= \widetilde{\Sigma}_{t+\Delta} G_t' S_t^{-1} \\ \widehat{X}_{t+\Delta} &= \widetilde{X}_{t+\Delta} + H_t \left(\widehat{O}_{t+\Delta,i} - O_{t+\Delta,i}(\widetilde{X}_{t+\Delta}, \theta) \right) \\ \widehat{\Sigma}_{t+\Delta} &= (I_2 - H_t G_t) \widetilde{\Sigma}_{t+\Delta} \end{aligned}$$

For every parameter set θ, we compute the time-series of the predicted state $\{\widetilde{X}_t\}$ and the log-likelihood function

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} \left[\log det(S) + \left(\widehat{O}_{t+\Delta,i} - O_{t+\Delta,i}(\widetilde{X}_{t+\Delta}, \theta) \right)' S_t^{-1} \left(\widehat{O}_{t+\Delta,i} - O_{t+\Delta,i}(\widetilde{X}_{t+\Delta}, \theta) \right) \right] \tag{B.16}$$

The estimated parameter $\widehat{\theta}$ is the maximizer of $\mathcal{L}(\theta)$. The maximization itself is performed using differential evolution of Storn and Price (1997).

B.2.2 Model identification

Our model allows for several parameter combinations that are observationally equivalent. Parameter identification requires that the option pricing model be unique under invariant transformations. We borrow from Dai and Singleton (2000) and study invariant transformations that change the state and parameter matrices without changing the joint distribution of option prices and thus the spot variance $V_t := Tr[X_t] + E(k^2)\left(Tr[\Lambda X_t] + \lambda_0\right)$.

In the first step of our estimation process, we jointly estimate the state, all risk-neutral parameters, and the physical parameters of the state dynamics $\theta = (X_t;\ M,\ Q,\ R, \beta, \lambda_0, \Lambda;\ M^*,\ \beta^*)$. To identify these parameters, we first focus on the risk-neutral, diffusive part.

The diffusive spot volatility is $Tr[X_t]$, therefore the only class of transformations that needs to be considered are trace invariant transformations. These are first the similarity transformation $\mathcal{T}_S = \mathcal{D} X_t \mathcal{D}^{-1}$ and second the permutation $\mathcal{T}_P$ that reorders the rows (or columns) of X_t.

Applying $\mathcal{T}_S$ to (3.6) results in a transformed model with state and parameter matrices

$$\mathcal{T}_S \theta = (\mathcal{D} X_t \mathcal{D}^{-1};\ \mathcal{D} M \mathcal{D}^{-1}, \mathcal{D} Q \mathcal{D}^{-1}, \mathcal{D} R \mathcal{D}^{-1};\ \beta\ ).$$

In order to identify our model, we apply parameter restrictions that only admit $\mathcal{D} = I_2$.

Without loss of generality, we can assume $|det(\mathcal{D})| = 1$.[1] Next we observe that the state matrix X_t is symmetric by construction, thus $\mathcal{D}X_t\mathcal{D}^{-1}$ also needs to be symmetric. This requires $\mathcal{D}$ to be orthogonal ($\mathcal{D}' = \mathcal{D}^{-1}$), thus $\mathcal{D}$ must be a rotation or mirror matrix.

We choose the following restrictions: M is lower triangular and the sign of M_{21} is positive. Choosing M to be lower triangular requires $\mathcal{D}$ to be lower triangular, in order to ensure $\mathcal{D}M\mathcal{D}^{-1}$ lower triangular. If $\mathcal{D}$ is both orthogonal and lower triangular, it must be a diagonal matrix $\left(\begin{smallmatrix} d_1 & 0 \\ 0 & d_2 \end{smallmatrix}\right)$ with elements $d_i = \pm 1$. We now have $\mathcal{D}M\mathcal{D}^{-1} = \left(\begin{smallmatrix} M_{11} & 0 \\ d_2/d_1\, M_{21} & M_{22} \end{smallmatrix}\right)$. By choosing the sign of M_{21} we exclude the case $d_1 \neq d_2$.

Our choices for M implicitly identify the state and select the order of the mean reversion speeds of the eigenvalues and thus of the components of X_t. Thus we also achieve identification with respect to $\mathcal{T}_P$. A direct consequence of our identification choices is the result that $X_{22,t}$ is the leading volatility factor and the identification of the sign of $X_{12,t}$.

We now discuss the identification of Q and R. To do so, we inspect the infinitesimal generator of the joint process for stock returns $Y_t := dS_t/S_t$ and state X_t (see Leippold and Trojani (2008)):

$$\begin{aligned} \mathcal{L}_{Y,X} \;&=\; \left(r - q - \frac{1}{2}Tr[X]\right)\frac{\partial}{\partial Y} + \frac{1}{2}Tr[X]\frac{\partial^2}{\partial Y^2} + 2Tr[XR'QD]\frac{\partial}{\partial Y} + \\ &\quad + Tr\left[\left(\beta Q'Q + MX + XM'\right)D + 2XDQ'QD\right] \end{aligned} \tag{B.17}$$

where $(D)_{ij} = \frac{\partial}{\partial X_{ij}}$ is the matrix differential operator.

The matrices Q and R only appear in the expressions $Q'Q$ and $R'Q$, i.e. only seven of their eight elements are identified. We choose Q to be the unique Choleski decomposition of $Q'Q$, i.e. Q upper triangular and positive definite. In order to reduce the number of parameters, we add the ad-hoc restriction for R to be also upper triangular.

Next, we focus on the spot jump variance $E^{\mathbb{Q}}(k^2)\left(Tr[\Lambda X_t] + \lambda_0\right)$ with $Tr[\Lambda X_t] = \Lambda_{11}X_{11,t} + (\Lambda_{12} + \Lambda_{21})X_{12,t} + \Lambda_{22}X_{22,t}$. Only the sum of the out-of diagonal elements of Λ are identified and we choose Λ upper triangular.

The physical parameter M^* enters our estimation via the transition equation of the Kalman filter (B.13) in the two expressions $\bar{\mu}$ and $\Psi\hat{X}_t\Psi'$. By construction, both expressions are symmetric and therefore only three elements are identified. We choose M^* to be lower triangular, to allow for an easy comparison to M. With this step, we indirectly identify the the price of risk Γ in (3.15).

The remaining parameter β^*_Λ in (3.28), which is estimated via OLS in the second estimation step, is fully identified.

---

[1] We can always construct a $\widetilde{\mathcal{D}} = \frac{1}{\sqrt{|det(\mathcal{D})|}}\mathcal{D}$ with $|det(\widetilde{\mathcal{D}})| = 1$ such that $\mathcal{D}X_t\mathcal{D}^{-1} = \widetilde{\mathcal{D}}X_t\widetilde{\mathcal{D}}^{-1}$.

B.2.3 Admissible parameter set

In order to ensure the existence and non-explosivity of our latent process (3.14), we have to apply the following additional restrictions to the feasible parameter set. First, $M^{*\prime}M^*$ must be negative definite to ensure the non-explosivity. Second, R must satisfy $RR' < I_2$ in order to ensure the existence of $Z_t = B_t R + W_t\sqrt{I_2 - RR'}$ in (3.8). Third, to ensure the existence of processes (3.6, 3.14) and of the change of measure (3.13), we require either $\beta > 1$, $\beta^* > 1$ and $\beta = \beta^*$ or $\beta > 3$ and $\beta^* > 3$. Finally, we require $\Lambda'\Lambda$ to be positive semi-definite and $\lambda_0 \geq 0$ in order to satisfy the positivity condition of the jump intensity $\lambda_t = \lambda_0 + Tr[\Lambda X_t]$. See Gruber (2015) for the details of the implementation of the constraints.

B.3 Measures of the volatility surface

B.3.1 Definition of level $\mathcal{L}_t$, skew $\mathcal{S}_t$ and term structure $\mathcal{M}_t$

To analyze our results in terms of observable properties of the implied volatility surface, such as in Figure 3.4, we define the following proxies[2]

$$\begin{array}{cccl}
\text{level} & \mathcal{L}_t & := & IV(\tau = \frac{1}{12}, \Delta = 0.5) \\
\text{short term skew} & \mathcal{S}_t & := & \frac{1}{0.6-0.4}\left[IV(\tau = \frac{1}{12}, \Delta = 0.6) - IV(\tau = \frac{1}{12}, \Delta = 0.4)\right] \\
\text{long term skew} & \mathcal{S}_t^{long} & := & \frac{1}{0.6-0.4}\left[IV(\tau = \frac{3}{12}, \Delta = 0.6) - IV(\tau = \frac{3}{12}, \Delta = 0.4)\right] \\
\text{term structure} & \mathcal{M}_t & := & \frac{1}{\frac{3}{12}-\frac{1}{12}}\left[IV(\tau = \frac{3}{12}, \Delta = 0.5) - IV(\tau = \frac{1}{12}, \Delta = 0.5)\right] \\
\text{skew term structure} & \mathcal{M}_t^{skew} & := & \frac{1}{\frac{3}{12}-\frac{1}{12}}\left[\mathcal{S}_t^{long} - \mathcal{S}_t\right]
\end{array}$$

where IV and Δ stand for the Black-Scholes implied volatility and delta. The time to maturity τ is measured in years. In the data, we obtain the required implied volatilities through two-dimensional interpolation of the volatility surface. In the model, we calculate these quantities exactly.

---

[2] We have evaluated the regression $IV(\tau, K)_t = \mathcal{L}_t + \mathcal{S}_t \cdot K + \mathcal{M}_t \cdot \tau$ as an alternative specification. We have found similar, but more noisy results. We have also performed robustness checks with respect to our definition. The alternative term structure measure $\mathcal{M}_t^6 := \frac{1}{\frac{6}{12}-\frac{1}{12}}\left[IV(\tau = \frac{6}{12}, \Delta = 0.5) - IV(\tau = \frac{1}{12}, \Delta = 0.5)\right]$ is, for example, 92% correlated with our term structure measure.

B.3.2 Skewness and term structure residuals

The volatility level $\mathcal{L}_t$ is highly correlated to the skewness $\mathcal{S}_t$ and the term structure $\mathcal{M}_t$ with correlation coefficients of -0.80 and -0.63 respectively. In order to unmask the unspanned skewness and term structure effects, we introduce the skewness residual $\widetilde{\mathcal{S}_t}$ (term structure residual $\widetilde{\mathcal{M}_t}$) as the skewness (term structure) components that cannot be explained by the volatility level. We obtain these quantities from the simple regressions:

$$\begin{aligned} \mathcal{S}_t &= \alpha_S + \beta_S \, \mathcal{L}_t + \widetilde{\mathcal{S}_t} \\ \mathcal{M}_t &= \alpha_M + \beta_M \, \mathcal{L}_t + \widetilde{\mathcal{M}_t} \end{aligned}$$

B.4 Additional Figures

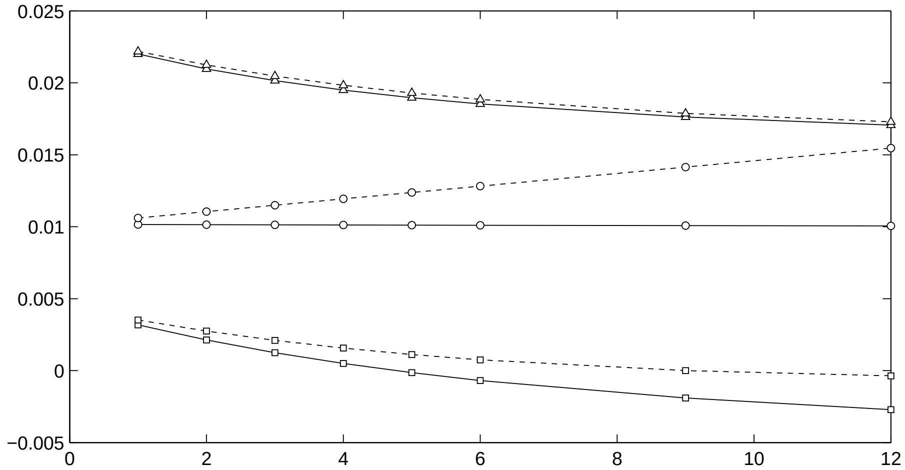


Figure B.1: The market price of the smile. We plot the sample average of the model-implied expectations, $\frac{1}{\tau} E_t^{\mathbb{P}}[\int_t^{t+\tau} X_s ds]$ and $\frac{1}{\tau} E_t^{\mathbb{Q}}[\int_t^{t+\tau} X_s ds]$, component-wise, for horizons τ from 1 to 12 months. Full lines report expectations under $\mathbb{P}$, dashed lines expectations under $\mathbb{Q}$. Circles: X_{11}; squares: X_{12}; triangles: X_{22}.

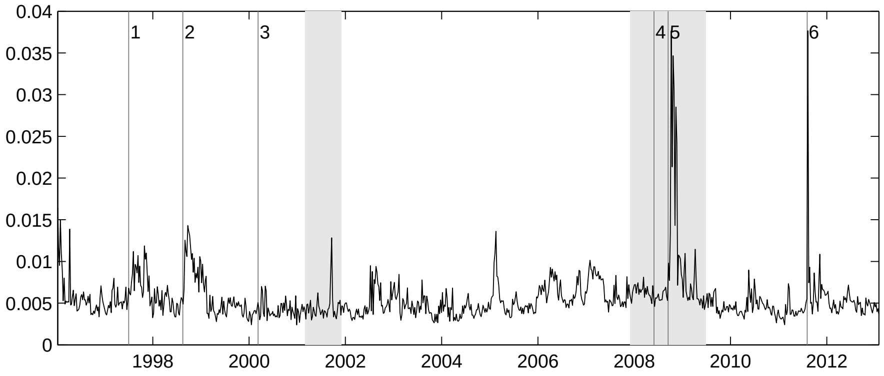


Figure B.2: Time series of mean absolute implied volatility errors ($MAIVE$) for model SVJ_{31}. For every day t in our sample, we plot the $MAIVE$ on that day, defined by $MAIVE_t := \frac{1}{N_t}\sum_{i=1}^{N_t}|IV_i - \widehat{IV_i}|$, where N_t is the number of available options on that day.

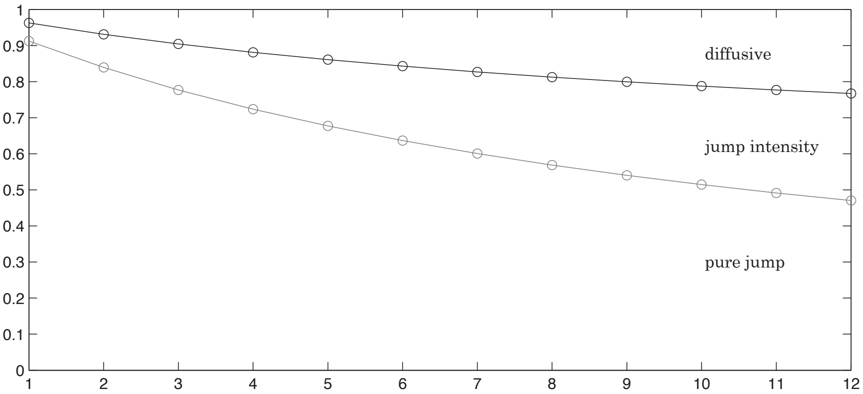


Figure B.3: Unconditional decomposition of the Variance Risk Premium for horizons of 1 to 12 months. This plot shows from bottom to top the fractions of VRP due to pure jump risk, jump intenstiy risk and diffusive volatility risk.

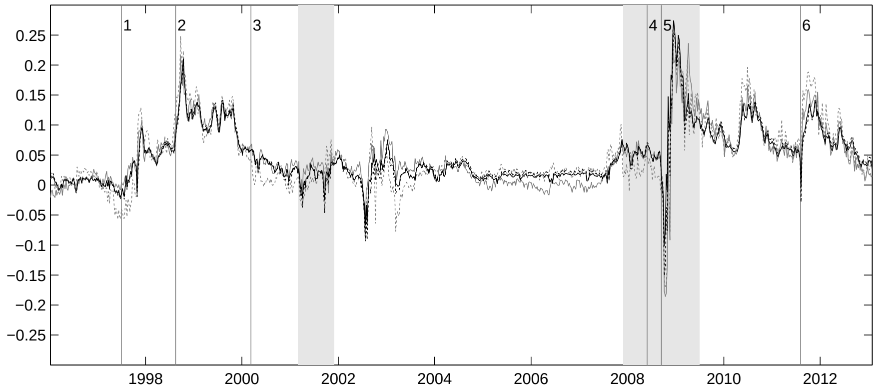
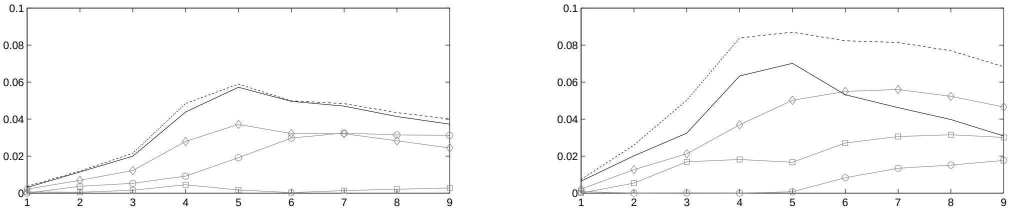


Figure B.4: Conditional equity risk premium at five month horizon. Grey line: ERP implied by the predictive regressions of Andersen, Fusari, Todorov (2013) using their $\tilde{U}$ factor. Black line: ERP implied by our model's $\tilde{X}_{11}$ and $\tilde{X}_{12}$ factors. Dashed lines depict predictive regressions using the respective full models.

Panel A: 1996-2013/01 (full sample)



Panel B: 1996-2002 (in-Sample period)

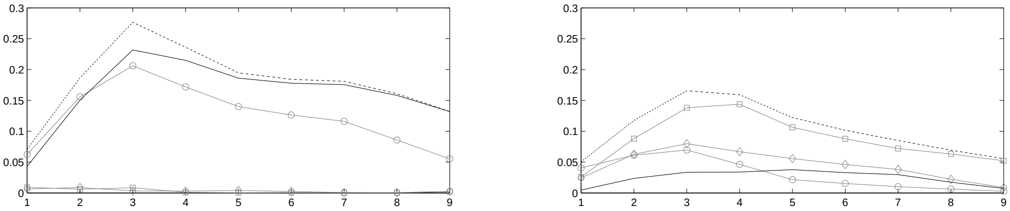


Figure B.5: R^2 of the predictive regressions of future excess returns on option-implied risk factors as function of the horizon in months. Left panels: Wishart model; grey: univariate regressions with individual factors X_{11} (diamonds), X_{12} (circles), X_{22} (squares); black: preferred model using the risk factors $\tilde{X}_{11}$ and $\tilde{X}_{12}$; black dashed: full regression. Right panels: model of Andersen, Fusari, Todorov (2013); grey: R^2 for univariate regressions with individual factors U (diamonds), V_1 (circles), V_2 (squares); black: preferred model using the risk factor $\tilde{U}$; black dashed line: full regression.

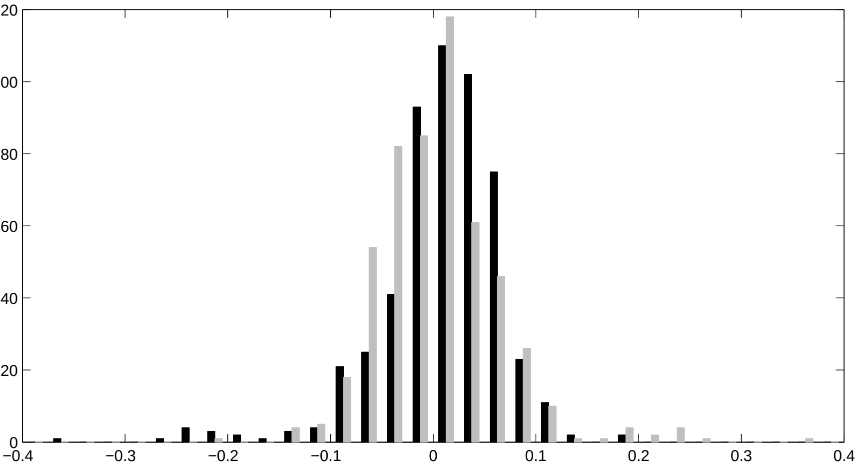


Figure B.6: Distribution of out-of sample returns of the S&P 500 index (black bars) and of our trading strategy (grey bars) over a horizon of 5 months.

B.5 Additional Tables

Panel A1: $MAIVE$ for SVJ_{20} model, in-sample

	$\tau < 30$	$30 \leq \tau < 75$	$75 \leq \tau < 180$	$\tau \geq 180$	all
$\lvert\Delta\rvert < 0.2$	0.938	0.607	0.655	0.593	0.664
$0.2 \leq \lvert\Delta\rvert < 0.4$	0.786	0.480	0.637	0.634	0.610
$0.4 \leq \lvert\Delta\rvert < 0.6$	0.699	0.510	0.537	0.517	0.539
$0.6 \leq \lvert\Delta\rvert < 0.8$	0.818	0.654	0.482	0.548	0.589
$0.8 \leq \lvert\Delta\rvert$	1.284	1.049	0.809	1.023	1.002
all	0.894	0.662	0.606	0.652	0.670

Panel A2: $MAIVE$ for SVJ_{31} model, in-sample

	$\tau < 30$	$30 \leq \tau < 75$	$75 \leq \tau < 180$	$\tau \geq 180$	all
$\lvert\Delta\rvert < 0.2$	0.897	0.552	0.496	0.506	0.576
$0.2 \leq \lvert\Delta\rvert < 0.4$	0.842	0.419	0.480	0.468	0.502
$0.4 \leq \lvert\Delta\rvert < 0.6$	0.714	0.465	0.554	0.387	0.486
$0.6 \leq \lvert\Delta\rvert < 0.8$	0.735	0.497	0.486	0.378	0.476
$0.8 \leq \lvert\Delta\rvert$	0.887	0.567	0.628	0.871	0.721
all	0.804	0.495	0.531	0.503	0.542

Panel B1: $MAIVE$ for SVJ_{20} model, out of sample

	$\tau < 30$	$30 \leq \tau < 75$	$75 \leq \tau < 180$	$\tau \geq 180$	all
$\lvert\Delta\rvert < 0.2$	0.804	0.976	1.067	1.393	1.089
$0.2 \leq \lvert\Delta\rvert < 0.4$	0.675	0.531	0.719	1.018	0.730
$0.4 \leq \lvert\Delta\rvert < 0.6$	0.955	0.501	0.469	0.759	0.605
$0.6 \leq \lvert\Delta\rvert < 0.8$	1.228	0.780	0.471	0.846	0.757
$0.8 \leq \lvert\Delta\rvert$	1.403	1.207	0.753	1.026	1.062
all	1.061	0.797	0.645	0.961	0.826

Panel B2: $MAIVE$ for SVJ_{31} model, out of sample

	$\tau < 30$	$30 \leq \tau < 75$	$75 \leq \tau < 180$	$\tau \geq 180$	all
$\lvert\Delta\rvert < 0.2$	0.819	0.690	0.655	0.957	0.765
$0.2 \leq \lvert\Delta\rvert < 0.4$	0.753	0.430	0.506	0.566	0.522
$0.4 \leq \lvert\Delta\rvert < 0.6$	0.871	0.483	0.549	0.414	0.523
$0.6 \leq \lvert\Delta\rvert < 0.8$	0.954	0.519	0.503	0.544	0.566
$0.8 \leq \lvert\Delta\rvert$	0.964	0.653	0.487	0.881	0.704
all	0.886	0.546	0.528	0.639	0.602

Table B.1: $MAIVE$ stratified by maturity and moneyness. We report for models SVJ_{20} and SVJ_{31} the mean absolute implied volatility error across maturity and moneyness bins.

Panel A1: Fraction of prices within bid/ask spread SVJ_{20} model, in-sample

	$\tau < 30$	$30 \leq \tau < 75$	$75 \leq \tau < 180$	$\tau \geq 180$	all
$\lvert\Delta\rvert < 0.2$	0.363	0.405	0.322	0.371	0.368
$0.2 \leq \lvert\Delta\rvert < 0.4$	0.439	0.535	0.348	0.316	0.400
$0.4 \leq \lvert\Delta\rvert < 0.6$	0.602	0.574	0.413	0.318	0.445
$0.6 \leq \lvert\Delta\rvert < 0.8$	0.722	0.607	0.554	0.360	0.516
$0.8 \leq \lvert\Delta\rvert$	0.709	0.555	0.492	0.222	0.443
all	0.588	0.555	0.449	0.317	0.449

Panel A2: Fraction of prices within bid/ask spread SVJ_{31} model, in-sample

	$\tau < 30$	$30 \leq \tau < 75$	$75 \leq \tau < 180$	$\tau \geq 180$	all
$\lvert\Delta\rvert < 0.2$	0.372	0.441	0.453	0.409	0.423
$0.2 \leq \lvert\Delta\rvert < 0.4$	0.379	0.617	0.451	0.404	0.472
$0.4 \leq \lvert\Delta\rvert < 0.6$	0.604	0.596	0.315	0.396	0.453
$0.6 \leq \lvert\Delta\rvert < 0.8$	0.766	0.713	0.489	0.488	0.580
$0.8 \leq \lvert\Delta\rvert$	0.844	0.852	0.598	0.276	0.589
all	0.616	0.668	0.461	0.403	0.517

Panel B1: Fraction of prices within bid/ask spread SVJ_{20} model, out of sample

	$\tau < 30$	$30 \leq \tau < 75$	$75 \leq \tau < 180$	$\tau \geq 180$	all
$\lvert\Delta\rvert < 0.2$	0.471	0.351	0.273	0.176	0.298
$0.2 \leq \lvert\Delta\rvert < 0.4$	0.574	0.615	0.395	0.302	0.464
$0.4 \leq \lvert\Delta\rvert < 0.6$	0.590	0.727	0.607	0.383	0.586
$0.6 \leq \lvert\Delta\rvert < 0.8$	0.684	0.717	0.742	0.374	0.624
$0.8 \leq \lvert\Delta\rvert$	0.815	0.693	0.742	0.474	0.665
all	0.649	0.650	0.593	0.360	0.556

Panel B2: Fraction of prices within bid/ask spread SVJ_{31} model, out of sample

	$\tau < 30$	$30 \leq \tau < 75$	$75 \leq \tau < 180$	$\tau \geq 180$	all
$\lvert\Delta\rvert < 0.2$	0.452	0.498	0.512	0.253	0.433
$0.2 \leq \lvert\Delta\rvert < 0.4$	0.521	0.757	0.595	0.487	0.613
$0.4 \leq \lvert\Delta\rvert < 0.6$	0.625	0.683	0.503	0.557	0.590
$0.6 \leq \lvert\Delta\rvert < 0.8$	0.803	0.819	0.668	0.518	0.690
$0.8 \leq \lvert\Delta\rvert$	0.905	0.915	0.875	0.489	0.794
all	0.696	0.763	0.643	0.483	0.647

Table B.2: Fraction of model-implied option prices within bid-ask spread for models SVJ_{20} and SVJ_{31}, across maturity and moneyness bins.

τ	$(E^{\mathbb{P}} - E_t^{\mathbb{Q}})\frac{1}{\tau}\left[\int_t^{t+\tau} X_{11s}ds\right]$			$(E^{\mathbb{P}} - E_t^{\mathbb{Q}})\frac{1}{\tau}\left[\int_t^{t+\tau} X_{12s}ds\right]$			$(E^{\mathbb{P}} - E_t^{\mathbb{Q}})\frac{1}{\tau}\left[\int_t^{t+\tau} X_{22s}ds\right]$		
months	X_{11}	X_{12}	X_{22}	X_{11}	X_{12}	X_{22}	X_{11}	X_{12}	X_{22}
1	-0.043	0.000	0.000	-0.025	-0.037	0.000	-0.002	-0.043	-0.000
2	-0.083	0.000	0.000	-0.048	-0.064	0.000	-0.006	-0.071	-0.000
3	-0.120	0.000	0.000	-0.069	-0.082	0.000	-0.012	-0.087	-0.000
4	-0.156	0.000	0.000	-0.088	-0.094	0.000	-0.018	-0.096	-0.000
5	-0.189	0.000	0.000	-0.106	-0.101	0.000	-0.025	-0.101	-0.000
6	-0.221	0.000	0.000	-0.121	-0.105	0.000	-0.032	-0.102	-0.000
7	-0.251	0.000	0.000	-0.136	-0.107	0.000	-0.038	-0.101	-0.000
8	-0.279	0.000	0.000	-0.149	-0.106	0.000	-0.044	-0.098	-0.000
9	-0.306	0.000	0.000	-0.161	-0.105	0.000	-0.050	-0.095	-0.000
10	-0.331	0.000	0.000	-0.172	-0.103	0.000	-0.055	-0.091	-0.000
11	-0.355	0.000	0.000	-0.182	-0.100	0.000	-0.060	-0.087	-0.000
12	-0.377	0.000	0.000	-0.191	-0.096	0.000	-0.065	-0.083	-0.000
12-1	-0.335	0.000	0.000	-0.166	-0.059	0.000	-0.063	-0.040	-0.000
12-3	-0.257	0.000	0.000	-0.122	-0.015	0.000	-0.053	0.004	0.000
3-1	-0.078	0.000	0.000	-0.044	-0.044	0.000	-0.010	-0.044	-0.000

Table B.3: Loadings of option-implied components X_{11}, X_{12}, X_{22} on the market price of the smile $(E^{\mathbb{P}} - E_t^{\mathbb{Q}})\frac{1}{\tau}\left[\int_t^{t+\tau} X_{ijs}ds\right]$, $i, j = 1, 2$, for horizons τ from 1 to 12 months, (columns 2 to 4). The last three rows compute the contribution of each option-implied component to three proxies for the slope of the term structure of the market price of the smile, measured as $(E^{\mathbb{P}} - E_t^{\mathbb{Q}})\left[\int_t^{t+12} X_{ijs}ds\right] - (E^{\mathbb{P}} - E_t^{\mathbb{Q}})\frac{1}{\tau}\left[\int_t^{t+\tau} X_{ijs}ds\right]$, for $\tau = 1, 3$ months, respectively.

τ	$VRP_t^c(\tau)$			$VRP_t^{dc}(\tau)$			$VRP_t^{dj}(\tau)$			$VRP_t(\tau)$		
months	X_{11}	X_{12}	X_{22}	X_{11}	X_{12}	X_{22}	X_{11}	X_{12}	X_{22}	X_{11}	X_{12}	X_{22}
1	-0.044	-0.043	-0.000	-0.068	-0.045	-0.000	-0.537	-0.762	-0.272	-0.649	-0.851	-0.272
2	-0.089	-0.071	-0.000	-0.131	-0.076	-0.000	-0.525	-0.679	-0.223	-0.745	-0.825	-0.223
3	-0.132	-0.087	-0.000	-0.191	-0.095	-0.000	-0.511	-0.607	-0.186	-0.835	-0.790	-0.186
4	-0.174	-0.096	-0.000	-0.248	-0.108	-0.000	-0.497	-0.546	-0.158	-0.919	-0.750	-0.158
5	-0.214	-0.101	-0.000	-0.300	-0.114	-0.000	-0.483	-0.492	-0.135	-0.998	-0.708	-0.135
6	-0.253	-0.102	-0.000	-0.349	-0.117	-0.000	-0.469	-0.446	-0.118	-1.071	-0.665	-0.118
7	-0.289	-0.101	-0.000	-0.395	-0.118	-0.000	-0.455	-0.406	-0.103	-1.139	-0.625	-0.103
8	-0.323	-0.098	-0.000	-0.437	-0.117	-0.000	-0.441	-0.371	-0.092	-1.202	-0.586	-0.092
9	-0.356	-0.095	-0.000	-0.477	-0.114	-0.000	-0.427	-0.341	-0.083	-1.260	-0.550	-0.083
10	-0.386	-0.091	-0.000	-0.514	-0.111	-0.000	-0.414	-0.315	-0.075	-1.314	-0.517	-0.075
11	-0.415	-0.087	-0.000	-0.548	-0.107	-0.000	-0.401	-0.291	-0.068	-1.364	-0.486	-0.068
12	-0.442	-0.083	-0.000	-0.581	-0.103	-0.000	-0.388	-0.271	-0.063	-1.411	-0.457	-0.063
12-1	-0.398	-0.040	-0.000	-0.513	-0.058	-0.000	0.149	0.492	0.210	-0.762	0.394	0.210
12-3	-0.310	0.004	0.000	-0.389	-0.007	0.000	0.123	0.337	0.123	-0.576	0.333	0.123
3-1	-0.088	-0.044	-0.000	-0.124	-0.050	-0.000	0.026	0.155	0.086	-0.186	0.061	0.086

Table B.4: Loadings of option-implied components X_{11}, X_{12}, X_{22} on diffusive, intensity and pure jump variance risk premia $VRP_t^c(\tau)$, $VRP_t^{dc}(\tau)$ and $VRP_t^{dj}(\tau)$, respectively. For horizons τ from 1 to 12 months, we compute the model implied loading of state variables X_{11}, X_{12} and X_{22} in diffusive, intensity and pure-jump variance risk premia (columns 2 to 4). The last column reports the state variables loadings in the total model-implied variance risk premium $VRP_t(\tau)$. The last three rows compute the contribution of each option-implied component to three proxies for the slope of the term structures of variance risk premia, measured as $VRP_t^u(12) - VRP_t^u(\tau)$, for $u = c, dc, dj$ and $\tau = 1, 3$ months, respectively.

Panel A: 1 month horizon

	Index			Strategy		
	mean	SR	skewness	mean	SR	skewness
Out of sample	0.031	0.184	-1.597	0.046	0.229	-1.064
Conundrum	0.056	0.511	-0.669	0.022	0.133	-0.398
Crisis	-0.117	-0.438	-1.167	-0.042	-0.134	-1.067
Post crisis	0.088	0.584	-1.040	0.145	0.960	0.022

Panel B: 2 month horizon

	Index			Strategy		
	mean	SR	skewness	mean	SR	skewness
Out of sample	0.033	0.190	-1.807	0.064	0.323	-1.828
Conundrum	0.060	0.556	-0.197	0.061	0.388	0.116
Crisis	-0.118	-0.401	-1.187	-0.108	-0.333	-1.570
Post crisis	0.091	0.654	-0.704	0.189	1.533	-0.481

Panel C: 3 month horizon

	Index			Strategy		
	mean	SR	skewness	mean	SR	skewness
Out of sample	0.032	0.181	-1.593	0.069	0.357	-0.343
Conundrum	0.058	0.542	-0.124	0.015	0.099	0.037
Crisis	-0.104	-0.347	-0.892	0.170	0.549	-0.775
Post crisis	0.083	0.609	-0.555	0.092	0.672	-0.141

Panel D: 4 month horizon

	Index			Strategy		
	mean	SR	skewness	mean	SR	skewness
Out of sample	0.031	0.173	-1.531	0.046	0.230	0.081
Conundrum	0.056	0.535	-0.157	0.003	0.022	0.162
Crisis	-0.098	-0.313	-0.760	0.167	0.512	-0.450
Post crisis	0.079	0.589	-0.314	0.034	0.237	-0.137

Table B.5: Out-of sample statistics of the trading strategy discussed in the main text. We compare mean return, Sharpe ratio (SR) and skewness of the returns of the S&P 500 index to our trading strategy over horizons from 1 to 4 months. Returns and Sharpe ratios are annualized. We break down the out of sample period (2003-Jan 2013) into three sub periods: the Conundrum (2003-2007), the Financial Crisis (2008-2009) and the Post-Crisis period (2010-Jan 2013).

	X_{11}	X_{12}	X_{22}
Min	0.0000	-0.0096	0.0001
Max	0.0516	0.0893	0.2610
Mean	0.0102	0.0044	0.0233
Median	0.0091	0.0014	0.0171
Positive	1.0000	0.6659	1.0000
Stdv	0.0091	0.0104	0.0259
Skewness	1.3460	3.6063	4.6427
Kurtosis	5.6376	20.9171	35.1886
AR(1)	0.9896	0.9529	0.8842
Half life	1.2753	0.2766	0.1083

Table B.6: Summary statistics of filtered option-implied components X_{11}, X_{12} and X_{22} for model SVJ_{31}. “Positive” denotes the fraction of positive realizations of each option-implied component in our sample period. Half lifes are given in years.

Chapter 4

Eliciting a Smile

Numerical Methods for Option Pricing with Matrix Affine Jump Diffusions

The matrix affine jump diffusion (MAJD) process, introduced in Leippold and Trojani (2008), represents a promising class of multivariate stochastic volatility models. It allows for mutually-exciting risk factors, unspanned skewness effects and a realistic modelling of variance risk, all without compromising on tractability. So far, the literature has focused on the theoretical properties of these processes (Muhle-Karb, Pfaffel and Stelzer (2010), Meyerhofer (2014)), on applications in option pricing (da Fonseca et al. (2008), Gruber et al. (2010)), and on approximations of the resulting volatility smile (Benabid et al. (2009)). Numerical aspects of estimation and evaluation of these models in the context of transform methods have not yet been considered. This paper aims at filling this gap.

In doing so, I build on a vast literature on the numerical aspects of option pricing. The use of Fourier integrals for option pricing has been pioneered by Chen and Scott (1992), Heston (1993) and Bates (1996). Carr and Madan (1999) introduce the Fast Fourier Transform (FFT) method to finance, which allows for the pricing of an entire option chain with a fixed number (typically 2^{12}) of evaluations of the Laplace transform. Fang and Oosterlee (2008) improve on the work of Carr and Madan by choosing a new set of basis functions and by introducing an efficient truncation of the probability density function. Their Fourier-Cosine method (COS) is typically 20 times faster than the FFT, without compromising precision.

Numerical problems related to the implementation of transform methods, especially with a complex logarithm, were first emphasized by Schöbel and Zhu (1999). Kahl and Jäckel (2005) take this topic up and formulate the rotation count algorithm for scalar stochastic volatility models. Lord and Kahl (2010) build on their insights, but advocate a re-formulation of the Laplace transform to avoid rather than remedy the problem of the complex logarithm. Cont and Hamida (2005) use evolutionary algorithms to estimate a low-dimensional volatility model.

The matrix nature of the MAJD class introduces several new challenges in the fields of computational efficiency, numerical stability, parameter identification and optimization:

First, matrix models are by definition high-dimensional models. In particular, the large number of state components increases the computational complexity of jointly estimating model parameters and the latent state. The smallest MAJD model, which is based on a symmetric 2×2 matrix and serves as an illustration throughout this paper, has three state components and 16 risk-neutral parameters.

Second, the Laplace transform has several matrix components, some of which double the dimensionality of the problem. At its core, it contains a complex matrix exponential of double dimension, i.e. of a 4×4 matrix in the smallest case. Matrix exponentials have a computational cost which is roughly two orders of magnitude larger than scalar exponentials.

Third, the Laplace transform contains a complex matrix logarithm, which is a multivalued function. Using the principal branch of the complex logarithm, as implemented by default in nearly every programming language, causes potentially large biases in the option price.

Fourth, the Laplace transform contains an inverse matrix, which is numerically unstable if the matrix is close to rank deficiency. This is the case for large arguments of the Laplace transform, i. e. for long horizons and/or if the integration limits are set conservatively (i. e. very widely).

Fifth, little is ex-ante known about the role of individual state components. We know for example that the state in a one-factor volatility model basically corresponds to short term variance, which makes the square of the short term implied volatility a natural candidate as starting value in a convex optimization. The high dimension of MAJD models makes it possible that different state components pick up distinct second-order phenomena.[1] It is therefore very difficult to deduce starting values for the state from economic considerations. The same applies to parameter estimation.

Sixth, although speed is nowadays rarely an issue in estimating an option pricing model, MAJD models are an exception. On one side, the dimensionality of the problem requires an unusually high number of evaluations of the Laplace transform, on the other side, these evaluations are unusually costly. Pricing one option with standard numerical integration takes about 9 seconds. Estimating the benchmark 2×2 model on 7 years of weekly data using nonlinear least squares requires the calculation of some 1 billion

---

[1] See Gruber et al. (2010) for a detailed analysis of the role of the state components in the benchmark case.

option prices and would take 285 years.[2]

This paper proposes several independent speed improvements that total to a factor of 10^6, reducing the estimation time to a few hours and thus rendering the model estimation feasible. The biggest contribution to these improvements comes from the matrix rotation count algorithm, with a factor of about 70. It enables the use of the FFT method, which prices an entire chain of about 20 options in less than 3 seconds. An additional factor of almost 20 is achieved by using the COS method, which requires fewer evaluations of the Laplace transform. A further factor of 50 obtains from a novel evaluation scheme of the likelihood function, that separates state- and parameter-dependent parts of the Laplace transform. This evaluation scheme also improves the scaling of the computational cost with respect to the size of the data set, especially the length of the time series. It requires a delicate choice of the integration bounds. I present an approximation that limits the error in implied volatility to 10^{-3} implied volatility percentage points (i.e. less than 0.1 implied volatility basis points). A final factor of 20 obtains from parallelization.

Beyond speed improvements, this paper identifies a numerical instability of the Laplace transform for large values of the imaginary part of its argument. This requires a precise choice of the integration bounds and truncation parameters for the FFT and COS series.

I furthermore present two useful reparametrizations of the state and parameter matrices, which facilitate the constrained optimization with respect to admissibility conditions of the stochastic process.

The rest of the paper is organized as follows. Section 2 gives a brief overview of the MAJD process and the estimation strategies under consideration. Section 3 discusses issues arising from the complex logarithm and introduces the matrix rotation count algorithm. Section 4 presents the fast evaluation scheme of the likelihood function based on the separation of state- and parameter-dependent parts of the Laplace transform and an approximation of the integration limits. Section 5 discusses the numerical instability of a matrix inverse in the Laplace transform. Section 6 discusses problems of optimization and proposes alternative matrix representations. Section 7 concludes. The appendix specifies the 2×2 matrix process which is used as an illustration throughout this paper.

---

[2] See Table 3.1 for the specification of the reference data set, and Table 3.3 for the benchmark model. The reference computer system is a 2.8 GHz Intel Xeon 540 running MATLAB 2014b on one core.

4.1 The MAJD Process for Option Pricing

4.1.1 The Process

Variance states follow the affine diffusion process:[3]

$$dX_t = [\beta Q'Q + MX_t + X_t M']dt + \sqrt{X_t}dB_t Q + Q'dB_t'\sqrt{X_t} \ , \tag{4.1}$$

where the state matrix X_t is an $n \times n$ symmetric, positive definite matrix, the parameters M and Q are $n \times n$ matrices and B is an $n \times n$ standard Brownian motion under the risk-neutral martingale measure $\mathbb{Q}$. X is the Wishart process introduced by Bru (1991). The process is positive definite (semi-definite) if $\beta > n+1$ ($\beta > n-1$), such that the diffusive variance $tr(X_t)$ cannot reach (cross) the zero bound.

Remark. The symmetric state matrix X_t has $n(n+1)/2$ distinct components. The n diagonal elements X_{ii} are always positive (non-negative), if the process is positive definite (positive semi-definite). The $n(n-1)/2$ out-of diagonal elements X_{ij} with $i \neq j$ can take positive or negative values within the bounds of the (semi)definiteness property. When M or Q are not diagonal matrices, all state components are dynamically interconnected. With both M and Q diagonal, the process collapses to n independent variance processes of the Heston-type. This model therefore naturally nests multivariate stochastic volatility models with independent factors such as Heston (1993), Bates (2000) or Christoffersen et al. (2009).

Given the matrix state dynamics (4.1), the returns process under the risk neutral probability measure $\mathbb{Q}$ is specified as

$$\frac{dS_t}{S_{t-}} = (r - q - \lambda_t \bar{k})dt + tr(\sqrt{X_t}dZ_t) + kdN_t \ , \tag{4.2}$$

where r and q denote interest rate and dividend yield, $tr(\cdot)$ denotes the trace of a matrix and $Z_t = B_t R + W_t\sqrt{I_n - RR'}$. Matrix W is another $n \times n$ standard Brownian motion, independent of B, and the correlation between variance and return shocks R is an $n \times n$ matrix such that $I_n - RR'$ is positive semi-definite.

Jumps in returns follow a compound Poisson process kdN_t with jump intensity $\lambda_t = \lambda_0 + tr(\Lambda X_t)$, where $\lambda_0 \geq 0$ and Λ is a positive definite $n \times n$ matrix. The distribution of the return jump size k is specified by its Laplace transform Θ^Y.

---

[3] Leippold and Trojani allow for jumps in variance and denote the Laplace transform of the respective jump size distribution Θ^X. Jumps in variance are omitted here, as their presence has no impact on the results presented in this paper.

4.1.2 Estimation Strategies and Computational Cost

Estimating a latent factor model such as the MAJD class usually encompasses the joint estimation of the parameters and the latent state. Before we can estimate the parameter matrices, we need to rewrite them in form of a vector θ, as all major optimization algorithms work only for vectors, see Section 4.5.2. We can formulate the optimization problem in a maximum likelihood context as:

$$(\widehat{\theta}, \{\widehat{X}_t\}) = \arg\max_{\theta, \{X_t\}} \mathcal{L}(\theta, \{X_t\}; \mathcal{O}) \tag{4.3}$$

where $\mathcal{L}(\cdot)$ denotes the likelihood function, $\{X_t\}$ the whole time series of X_t and $\mathcal{O}$ stands for the panel of option data. This is a high-dimensional problem, which does not lend itself well to optimization. For a time series of length T, we need to estimate $n_\theta + T \cdot n(n+1)/2$ quantities.[4]

In order to break the dimensionality of the joint problem, I opt for a nested estimation:[5]

$$\text{Inner problem:} \qquad \{\widehat{X}_t|\theta\} = \arg\max_{\{X_t\}} \mathcal{L}(\theta, \{X_t\}; \mathcal{O}) \tag{4.4}$$

$$\text{Outer problem:} \qquad \widehat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta, \{\widehat{X}_t|\theta\}; \mathcal{O}) \tag{4.5}$$

i.e. for each parameter vector θ I find the optimal time series of the state $\widehat{X}_t$ (inner optimization) and the I optimize over θ, evaluating the likelihood function at $\widehat{X}_t|\theta$. The point estimator for the parameter vector θ can be written as

$$\widehat{\theta} = \arg\max_{\theta} \mathcal{L}\Big(\theta, \arg\max_{\{X_t\}} \mathcal{L}(\theta, \{X_t\}; \mathcal{O}); \mathcal{O}\Big) \ . \tag{4.6}$$

This reduces the dimension of the optimization problem, as we now have one problem of dimension n_θ and T problems of dimension $n(n+1)/2$. The form of the likelihood function (4.5) and the number of option prices required for one evaluation depends on the estimation strategy:

---

[4] In the benchmark case of the minimal 2×2 model, $n_\theta = 16$ and $n(n+1)/2 = 3$. In the benchmark data set, $T = 359$, thus the overall dimension is 1093.

[5] In the context of NLLS estimation, Huang and Wu (2004) successfully iterate between an optimization over the state given a parameter vector and an optimization over the parameters, given the (previous) state estimate. This approach does not produce stable results for the MAJD class and is furthermore not compatible with stochastic optimization algorithms. Furthermore, extended Kalman Filter (EKF) estimation is only conceivable as nested optimization.

Nonlinear least squares (NLLS). This optimization approach follows Bates (2000) and Huang and Wu (2004) and minimizes the (weighted) relative pricing errors. The first (inner) step computes the optimal state conditional on the parameter vector. This is defined independently for each trading day t

$$\widehat{X}_t|\theta = \arg\min_{X_t} \sum_{i=1}^{N_t} \left[\left(\widehat{O}_{it}(\theta, X_t) - O_{it} \right) / F_t \right]^2 \ , \tag{4.7}$$

where O_{it} and $\widehat{O}_{it}(\theta; X_t)$ denote the observed and model-implied prices of option i, with $1 \leq i \leq N_t$. We can now define the vector $\mathbf{e}_t$ of relative option pricing errors at the conditionally optimal state with elements $e_{i,t} = \big(\widehat{O}_{it}(\theta, \widehat{X}_t|\theta) - O_{it}\big)/F_t$. Note that N_t, the length of vector $\mathbf{e}_t$, varies from trading day to trading day.

Our point estimate for parameter θ is given by the following pseudo Maximum Likelihood estimator:

$$\widehat{\theta} = \arg\max_{\theta} \ -\frac{1}{2} \sum_{t=1}^{T} \left(\ln|\Omega_t| + \mathbf{e}_t' \, \Omega_t^{-1} \, \mathbf{e}_t \right) \ . \tag{4.8}$$

where the $N_t \times N_t$ matrix Ω_t is the conditional covariance matrix of these errors which is obtained from calculating group-specific error covariances in three maturity and three moneyness groups, similar to Bates (2000).

In this optimization strategy, the outer optimization problem has a dimension of n_θ = 16. The T inner optimizations of dimension $n(n+1)/2$ are independent and can therefore be parallelized. We furthermore note that θ does not change during the inner optimization step, a fact that will be used in Section 4.3. A drawback of the nested NLLS approach is that $M_{NLLS}^{X\,opt}$, the number of optimization steps for (each) inner optimization, is quite large.[6] The total number of option prices that need to be calculated for one evaluation of the NLLS-likelihood function (4.8) is

$$M_{NLLS}^{\mathcal{L}} = T \, n_\tau \, n_k \, M_{NLLS}^{X\,opt} \tag{4.9}$$

where n_τ denotes the average number of maturities per trading day and n_k the average number of strikes per option chain. In the benchmark example, $M_{NLLS}^{\mathcal{L}}$ is roughly 2 million.[7]

Extended Kalman Filter (EKF). This optimization approach follows, among others, Carr and Wu (2007) and Carr and Wu (2009b). The inner optimization is replaced by an extended Kalman Filter, which produces $\widehat{X}_t|\theta$. Apart from obvious advantages such

---

[6] For the 2 × 2 benchmark model, $M_{NLLS}^{X\,opt} istypicallyabout40$.

[7] For the benchmark data set $T = 359$, $n_\tau = 6.09$ and n_k=23.

as consistency of the estimate with the assumed form of the process and the possibility to estimate the physical dynamics of the latent state, the EKF strategy usually requires the calculation of fewer option prices: one evaluation of the option price plus $n(n+1)/2+1$ evaluations to numerically calculate the Jacobian matrix, i. e. $M^J_{EKF} = 2 + n(n+1)/2$.

The point estimate for the parameter vector θ is now

$$\widehat{\theta} = \arg\max_{\theta} \; -\sum_{t=1}^{T} \mathbf{e}_t' \mathbf{e}_t \; . \tag{4.10}$$

where we define $\mathbf{e}_t$ analogously to the NLLS case. The total number of option prices that need to be calculated for one evaluation of the likelihood function (4.10) is

$$M^{\mathcal{L}}_{EKF} = T \, n_\tau \, n_k \, M^J_{EKF} \tag{4.11}$$

which is ca. 200 000 for the benchmark example, i.e. ten times fewer than the NLLS approach. We note that during one evaluation of the filter, θ does not change, a fact that we will use in Section 4.3. A drawback of the EKF approach is that, unlike the NLLS case, the calculation of the state cannot be parallelized, as the estimate of X_t depends on $\widehat{X}_{t-1}|\theta$.

4.1.3 The Laplace Transform

In the canonical form, the Laplace transform of the log-return process is written as (see Leippold and Trojani (2008))[8]:

$$\Psi_{\gamma,t}(\tau) = \exp\{\gamma Y_t + Tr\left[A(\tau;\gamma) X_t\right] + B(\tau;\gamma)\} \tag{4.12}$$

Where the $n \times n$ matrix $A(\tau)$ and the scalar $B(\tau)$ are solutions to the following Riccati differential equations:

$$\begin{aligned} \frac{\partial A(\tau)}{\partial \tau} &= A(\tau)M + M'A(\tau) + 2\gamma R Q A(\tau) + 2A(\tau)Q'QA(\tau) + \\ &\quad + \frac{\gamma(\gamma-1)}{2} Id_n + \Lambda\Big[\Theta^Y(\gamma) - 1 + \gamma\big(\Theta^Y(1) - 1\big)\Big] \qquad (4.13) \\ \frac{\partial B(\tau)}{\partial \tau} &= tr[\Omega\Omega' A(\tau)] + (\gamma - 1)r + \lambda_0\Big[\Theta^Y(\gamma) - 1 + \gamma\big(\Theta^Y(1) - 1\big)\Big] \qquad (4.14) \end{aligned}$$

where $A(0) = B(0) = 0$. Note: $tr[\Omega\Omega' A] = tr[A\Omega\Omega'] = tr[AQ'Q\beta]$.

---

[8] We follow the notation of Leippold and Trojani (2008), where X denotes the variance factors and Y the returns. Jumps in the volatility factors are omitted, as they do not impact the results in this paper.

Equations (4.13) and (4.14) have the following solutions:

$$\begin{aligned} A(\tau) &= C_{22}^{-1}(\tau)C_{21}(\tau) && (4.15)\\ B(\tau) &= \int_0^\tau tr[\Omega\Omega' A(s)] + \gamma r + \lambda_0\Big[\Theta^Y(\gamma) - 1 + \gamma\big(\Theta^Y(1) - 1\big)\Big]ds \\ &= \tau\Big\{(\gamma-1)r + \lambda_0\Big[\Theta^Y(\gamma) - 1 + \gamma\big(\Theta^Y(1) - 1\big)\Big]\Big\} \\ &\quad + \int_0^\tau tr[\Omega\Omega' A(s)]ds && (4.16)\\ &= \tau\Big\{(\gamma-1)r + \lambda_0\big[\Theta^Y(\gamma) - 1 - \gamma\Theta^Y(1)\big]\Big\} \\ &\quad - \frac{\beta}{2} tr[\log(C_{22}(\tau)) + \tau(M' + \gamma R'Q)] && (4.17) \end{aligned}$$

with the $2n \times 2n$ matrix

$$\begin{pmatrix} C_{11}(\tau) & C_{12}(\tau) \\ C_{21}(\tau) & C_{22}(\tau) \end{pmatrix} = \exp\left[\tau\begin{pmatrix} M + \gamma Q'R & -2Q'Q \\ C_0(\gamma) & -(M' + \gamma R'Q) \end{pmatrix}\right] := \exp(E) \qquad (4.18)$$

and the $n \times n$ matrix C_0:

$$C_0(\gamma) = \frac{\gamma(\gamma-1)}{2} Id_n + \Lambda\Big[\Theta^Y(\gamma) - 1 + \gamma\big(\Theta^Y(1) - 1\big)\Big] \ . \qquad (4.19)$$

The univariate Laplace transform of the return jump size distribution is for the lognormal case

$$\Theta_{LN}^Y(\gamma) = (1 + \overline{k})^\gamma \exp\left(\gamma(\gamma-1)\frac{\delta^2}{2}\right) \qquad (4.20)$$

and for the double exponential case

$$\Theta_{DX}^Y(\gamma) = \frac{\lambda^+\lambda^-}{\lambda^+\lambda^- + \gamma(\lambda^+ - \lambda^-) - \gamma^2}. \qquad (4.21)$$

Option prices obtain via the transform methods, i.e. the FFT of Carr and Madan (1999), or the COS method of Fang and Oosterlee (2008) or the quadrature method of Attari (2004).

4.2 The Matrix Rotation Count Algorithm

4.2.1 Multivalued Complex Logarithms

Ambiguity is a rare phenomenon in computational finance, but one such example is the complex logarithm embedded in the Laplace transform of most affine volatility models, such as the models of Heston (1993), Bates (2000), Duffie et al. (2000) or Leippold and

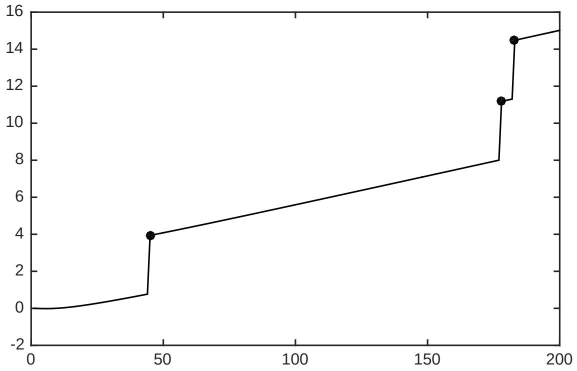


Figure 4.1: Imaginary part of $B(\tau, \gamma_k)$ in the benchmark model as a function of k in γ_k in the COS inversion for $\tau = 1yr$, when directly evaluating (4.17). Typical values for γ_k are given in Table 4.1.

Trojani (2008). To illustrate the problem, consider a complex scalar $z = a + ib$ with exponential $e^z = e^a \cdot e^{ib}$. The complex exponential $e^{ib} = \cos(b) + i\sin(b)$ is a periodic function of b, therefore for any integer k

$$e^{ib} = e^{i(b+k\,2\pi)} \ ,$$

and we cannot know from e^z the value of k in argument $z = a + i(b + k\,2\pi)$. As $\log(e^z)$ is multivalued, it is a convention to restrict b to the interval $(-\pi, \pi]$. This interval is called the “principal branch”. If the imaginary part of z passes $(2k+1)\pi$, the logarithm “rotates”, resulting in a discontinuity of the principal branch of the imaginary part of the logarithm. As Lord and Kahl (2010) observe, this may be “*leading to completely wrong option prices if options are priced by Fourier inversion.*” Figure 4.4 provides an illustration for the magnitude of the problem.

As the characteristic function in formulation (4.17) involves a matrix logarithm in $B(\gamma, \tau)$, we indeed observe discontinuities illustrated by Figure 4.1. More precisely, the jumps in $B(\gamma, \tau)$ are all of the order of π, with more jumps for larger durations τ. If we apply the (computationally much slower) formulation (4.16), no jumps occur. In order

τ	γ_0	γ_1	γ_{199}	τ	γ_0	γ_1	γ_{199}
1 month	0	1.06	212.4	12 months	0	0.35	69.9
3 months	0	0.70	139.8	24 months	0	0.25	49.4
9 months	0	0.41	80.7	60 months	0	0.16	31.3

Table 4.1: Typical values for $\gamma_k(\tau)$

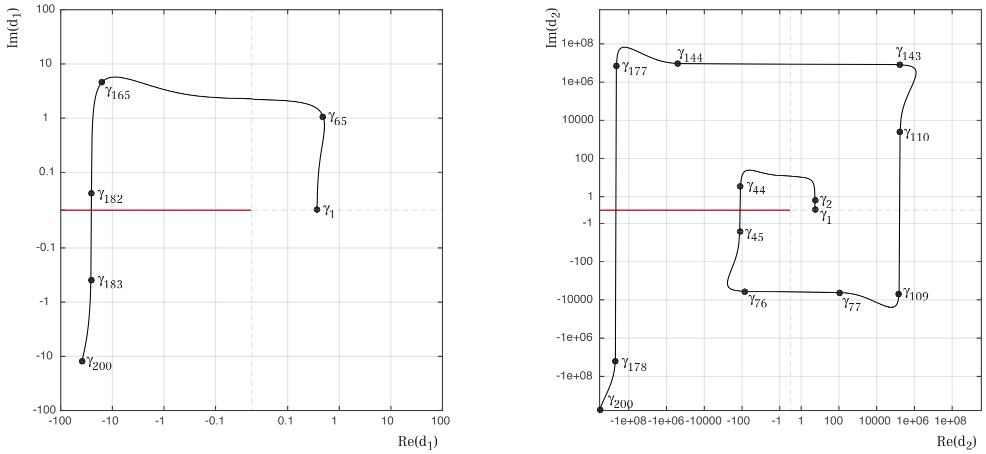


Figure 4.2: Complex eigenvalues d_1, d_2 of the matrix C_{22} of the benchmark model as a function of γ_k in the COS method for $\tau = 1yr$. The horizontal red lines at the transition from the second to the third quadrant highlight the discontinuity in the principal branch of the complex logarithm occurs. Left: first eigenvalue d_1, right: second eigenvalue d_2. The dashed grid lines illustrate the fact that each plot is composed of four log-log plots, omitting very small positive and negative values along both axes. Typical values for γ_k are given in Table 4.1.

to understand the phenomenon in more detail, we rewrite the matrix logarithm in (4.17) in terms of a PDP decomposition.

$$\begin{aligned} PDP^{-1} &= C_{22} \\ \log(C_{22}) &= P \log D P^{-1} \ . \end{aligned} \tag{4.22}$$

D is a diagonal matrix containing the eigenvalues of C_{22}, which means that the matrix logarithm of D is simply the diagonal matrix of the logarithms of the diagonal elements: $(\log(D)_{ii} = \log(D_{ii})$. Figure 4.2 illustrates the two eigenvalues of the 2×2 benchmark model as functions of γ. Both eigenvalues start in the first quadrant and rotate counter-clockwise. When an eigenvalue passes from the second to the third quadrant, a “rotation” occurs and the imaginary part of the logarithm jumps. The fact that the two eigenvalues rotate independently is the reason why we did not see a regular pattern in Figure 4.1.

4.2.2 The Algorithm

A practical solution to the problem above is to count the number of rotations, preferably separately for each eigenvalue, and to correct for them. This is the goal of the following algorithm:

Algorithm 1 (Matrix Rotation Count Algorithm)**.** The following algorithm solves the problem of the discontinuous matrix logarithm of C_{22} in equation (4.17) and produces a continuous B-function. The algorithm must be run separately for every τ.

Define $k = 1 \dots N$ as counter of the arguments of the Laplace transform γ_k and $j = 1 \dots n$ as counter of the eigenvalues of C_{22} (n is dimension of the Wishart process). The symbols $\Re(x)$ and $\Im(x)$ denote the real and imaginary parts of x.

```
1. For every eigenvalue $d_j$ of $C_{22}$ with $1 \leq j \leq n$, initialize the number of rotations at $r_{1,j} := 0$
2. For every $\gamma_k$, with $1 \leq k \leq N$ and $\gamma_{k-1} < \gamma_k$ do
    3. Calculate $C_{22}(\tau, \gamma_k)$
    4. Perform a decomposition $P_k D_k P_k^{-1} = C_{22}(\tau, \gamma_k)$, where $D_k$ is a diagonal matrix containing the eigenvalues of $C_{22}(\tau, \gamma_k)$.
    5. For every eigenvalue $(D_k)_{jj}$ with $1 \leq j \leq n$ do
        6. Calculate the complex logarithm $d_{k,j} = \log\big((D_k)_{jj}\big)$.
        7. Produce the sawtooth-like function $m_{k,j} = \Im(d_{k,j}) \mod \pi$.
        8. Verify whether a rotation has occurred. Increase $r_{k,j}$ by 1 for every positive rotation, i.e. when $m_{k,j} - m_{k-1,j} > \pi/2$ and decreased by 1 for every negative rotation, i.e. when $m_{k,j} - m_{k-1,j} < -\pi/2$.
        9. The correct branch of the imaginary part of $\log\big((D_k)_{ii}\big)$ obtains as $\Im(d_{k,j}) := m_{k,j} + \pi \cdot r_{k,j}$. The real part $\Re(d_{k,j})$ is not changed. We obtain the elements of the diagonal matrix $(D_k^{\log})_{ii} = \Re(d_{k,j}) + j\,(m_{k,j} + \pi r_{k,j}) = \log(D_k)$.
    enddo
    10. Use $\log(C_{22}(\tau, \gamma_k)) = P_k D_k^{log} P_k^{-1}$ to calculate $B$ in equation (4.17).
enddo
```

Figure 4.3 illustrates steps 7 to 9 of this algorithm.

Remark 1. The rotations of the individual eigenvalues illustrated in Figure 4.2 are counted in step 8 by calculating $m_{k,j} - m_{k-1,j}$. The algorithm therefore requires the sequential evaluation of the characteristic function for increasing values of γ_k, starting form $\gamma_1 = 0$. Thus it is only suitable for inversion methods that sequentially evaluate the characteristic function on a uniform grid like FFT, COS or Attari's method. A straight forward application to adaptive algorithms such as Gauss-Lobatto integration is not possible.

Remark 2. Some authors, e.g. Benabid et al. (2009) reformulate equation (4.17) using the identity $\log(\det(C_{22})) = tr(\log(C_{22}))$. In this formulation, knowledge of the

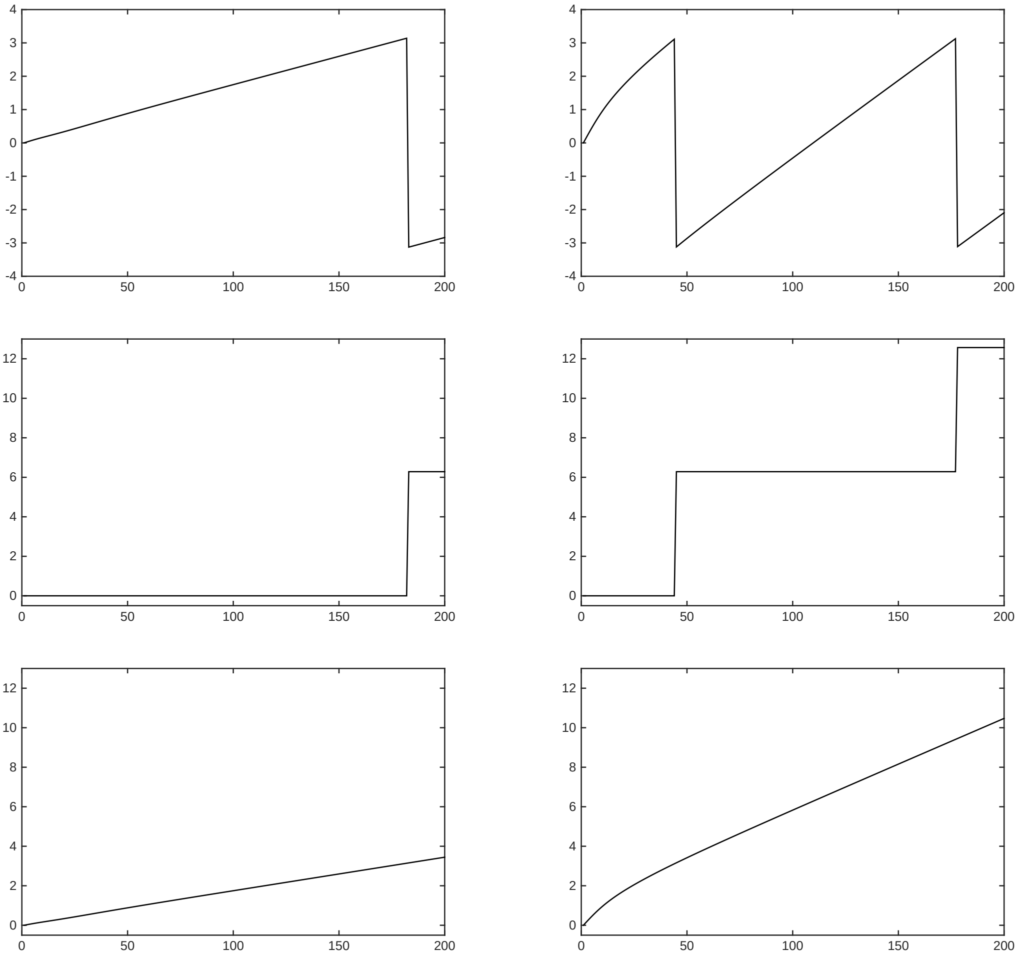


Figure 4.3: Illustration of Algorithm 1 (Matrix Rotation Count or MRC). All panels depict the complex part of the eigenvalues d_i of matrix C_{22} as a function of k in γ_k. Top panels: sawtooth-function from step 7. Middle panels: step function from step 8. Bottom panels: continuous eigenvalues from step 9. The left panels depict the first eigenvalue d_1, the right panels the second eigenvalue d_2. Parameters are taken from the benchmark model, $\tau = 1\,yr$.

individual eigenvalues $(D_k)_{jj}$ is lost, which makes it more difficult to count the individual rotations.

Remark 3. If the first derivative of the imaginary part of the eigenvalues of $C_{22}(\gamma)$ with respect to γ is too large or if the grid width $\gamma_k - \gamma_{k-1}$ is too wide, the algorithm will fail, as it would identify a rotation at every k. This can be avoided by choosing the following modified algorithm, that has a more sophisticated identification of the rotations:

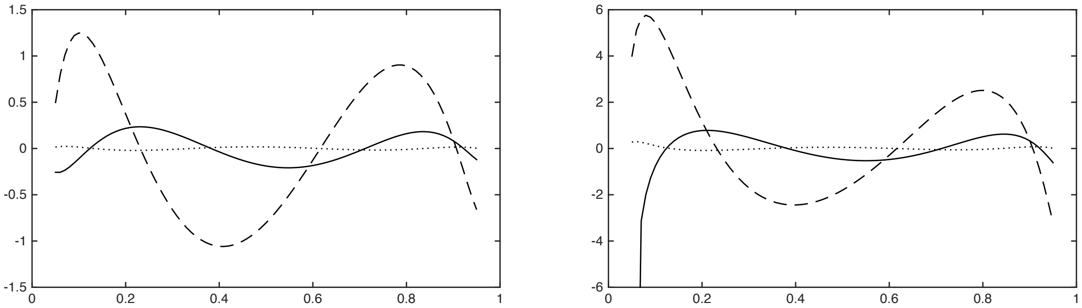


Figure 4.4: Typical corrections generated by the Matrix Rotation Count algorithm for call options in COS-inversions, as a function of the Black-Scholes delta. The values are shown here are for calls, for puts simply set $\Delta_{put} = 1 - \Delta_{call}$. Left panel: dollar pricing correction in percent of the price of the underlying. Right panel: implied volatility correction in volatility percentage points. Dotted lines denote a maturity of 0.75 years, full lines 1 year and dashed lines 1.25 years.

Algorithm 2 (Modified Matrix Rotation Count Algorithm)**.** This algorithm can be used if $\Im(d_{k,i}) - \Im(d_{k,i-1})$ is typically larger than π.

1–7. Perform steps $1-7$ like in algorithm 1.

8. Verify whether a rotation has occurred. The value of $r_{k,i}$ is increased (decreased) by 1 for every positive (negative) rotation, i.e. when $(m_{k,i}-m_{k-1,i})-(m_{k-1,i}-m_{k-2,i}) > (<)\,\pi/2$ *and* $(m_{k,i+1}-m_{k,i})-(m_{k,i}-m_{k1,i}) < (>)\,\pi/2$.

9–10. Perform steps $9-10$ like in algorithm 1.

Lord and Kahl (2010) propose an algorithm for the scalar case, which calculates the number of rotations for γ_k without knowledge of γ_{k-1}. In the matrix case, their approach is not feasible, as the following example for the 2×2 case shows. From (4.18), $C = \exp(E(\tau,\gamma))$, therefore we can calculate $PDP^{-1} = \tau E$ and anticipate any rotation in $C = P\exp(D)P^{-1}$. However, the argument of the matrix logarithm is now

$$C_{22} = \begin{pmatrix} \sum_{j=1}^{4} \exp(D_{jj})\bar{P}_{j3}P_{3j} & \sum_{j=1}^{4} \exp(D_{jj})\bar{P}_{j4}P_{3j} \\ \sum_{j=1}^{4} \exp(D_{jj})\bar{P}_{j3}P_{4j} & \sum_{j=1}^{4} \exp(D_{jj})\bar{P}_{j4}P_{4j} \end{pmatrix}$$

with $\bar{P} = P^{-1}$, i.e. the elements of C_{22} are now weighted sums of the eigenvalues of $E(\tau,\gamma)$. Thus the eigenvalues of C_{22} may pass a 2π-threshold even when no eigenvalue of E does, implying that the number of rotations for a single value of γ cannot be known.

4.2.3 Impact of the Matrix Rotation Count Algorithm

Wrongly choosing the principal branch of the complex matrix logarithm causes a bias in the option price that is potentially large. Using the benchmark model, the maximal dollar error for a horizon of one year is 0.3% of the price of the underlying. An error of this magnitude exceeds the typical option premium for a call with $\Delta = 0.05$ (viz. a put with $\Delta = -0.95$), making it impossible to calculate the implied volatility. A more detailed numerical analysis reveals that the bias is larger for longer maturities, for higher volatilities and for extreme structures of the state matrix, with the largest error obtained for singular state matrices.[9] The error is present in all Fourier inversion methods.

Figure 4.4 provides an illustration of the magnitude of the correction for the COS inversion method as a function of the Black-Scholes delta. The sine-shaped form of the error derives from the central term in the COS inversion formula (C.7), which reads $Re\left\{\phi_{Levy}\left(\frac{k\pi}{b-a}\right)\exp\left(ik\pi\frac{y-a}{b-a}\right)\right\}$. The bias in $B(\tau,\gamma)$ causes a shift of the imaginary part of $\phi_{Levy}(\cdot)$, which is multiplied by the periodic function $\exp(ik\pi\frac{y-a}{b-a}) = \cos(ik\pi\frac{y-a}{b-a}) + i\sin(ik\pi\frac{y-a}{b-a})$.

After the application of the rotation count algorithm to the analytical expression (4.17), the results for $B(\tau,\gamma)$ are identical to the ones obtained through numerical integration in (4.16) up to precision of the numerical integration.

4.3 Fast Evaluation Scheme of the Likelihood Function in the COS Inversion

4.3.1 Execution Speed of the Likelihood Function

With modern computer hardware, speed is normally not an issue in estimating option pricing models. Estimating the MAJD model is an exception. The dimension of the model necessitates a nested likelihood function. Its complexity requires the use stochastic optimization, based on at least 10 000 evaluations[10] of the likelihood function. Furthermore, matrix exponentials and logarithms are computationally more costly by a factor of 30 to 40 compared to their scalar counterparts.

After the 70-fold speed gain afforded by the Matrix Rotation Count Algorithm and the 20-fold speed gain from the use of the COS method of Fang and Oosterlee (2008),

---

[9] In the benchmark example, the state matrix is close to singular in 3.5% of the observations in the sample, with $\frac{X_{11}}{X_{22}}$ and $\frac{|X_{12}|}{X_{22}}$ both smaller than 0.02. As an extreme example, the filtered state for March 7, 2007, is $X = 0.0158\begin{pmatrix}1.7\times10^{-6} & -1.3\times10^{-3}\\ -1.3\times10^{-3} & 1\end{pmatrix}$.

[10] For example, a minimal optimization setup using differential evolution has a population size of 100 and an iteration count of 100 generations.

		COS		FFT	
		time	fraction	time	fraction
(1)	Parameter-dependent part of Ψ	144.1 ms	98.8%	2854 ms	99.4%
	Evaluation of $\exp(E)$ in (4.18)	77.9 ms	53.0%	1560 ms	54.3%
	Matrix Rotation Count Algorithm	46.1 ms	31.4%	860 ms	30.0%
(2)	State-dependent part of Ψ	0.8 ms	0.5%	5 ms	0.2%
(3)	Fourier inversion	1.1 ms	0.7%	11 ms	0.4%

Table 4.2: Breakdown of the execution time[11] for the pricing of one option chain.

pricing one option chain takes 0.146 seconds in the benchmark case. Still, one evaluation of the likelihood function takes 3.5 hrs (0.4 hrs) in the NLLS (EKF) scheme. Thus estimating the model would take almost 4 (0.4) years.

In order to find additional speed improvements, I profile the pricing of one option chain. There are three major program blocks: (1) Evaluation of the parameter-dependent part of the Laplace transform, i.e. of matrices $A(\tau)$ and $B(\tau)$ in (4.15) and (4.17), based on the evaluation of the matrix exponential in (4.18) and the application of the Matrix Rotation Count Algorithm. (2) Evaluation of the state-dependent part of the Laplace transform (4.12). (3) Fourier Inversion for the option prices. in the 2×2 benchmark case, Table 4.2 shows that 99% of the total computational cost is caused by the parameter-dependent part of the Laplace transform. The evaluation of the matrix exponential (4.18) alone takes half of the execution time. It is therefore evident that any further improvement requires a reduction of the number of evaluations of $A(\tau)$ and $B(\tau)$.

Reconsider the Levy-part of the Laplace transform (C.6):

$$\Psi_{Levy}(\gamma;\tau,X_t) = \exp\left\{Tr\left[A(\tau,\gamma)X_t\right] + B(\tau,\gamma)\right\}$$

Only expression X_t is time-varying. The costly expressions $A(\tau,\gamma)$ and $B(\tau,\gamma)$ depend on τ and γ only. We can therefore envisage an evaluation scheme of the likelihood function where we re-use any value of $A(\tau,\gamma), B(\tau,\gamma)$ that has already been calculated. The cost of such a scheme only depends on the number of different combinations of τ and γ.

---

[11] Omitting program overhead. See the appendix for the specification of the reference data set, the benchmark model and the reference computer system. One option chain contains $n_k = 23$ strikes. Timing in milliseconds (ms), with 1000 ms = 1 second.

4.3.2 The Evaluation Scheme

The first step towards a fast evaluation scheme is to reduce the number of different option maturities. Assuming a maximum maturity of 1 year, this number can be trimmed to just 52 if we use weekly data sampled on a fixed day of the week, e.g., Wednesday.[12] The shortest maturity is 3 days (Wednesday until expiry on Saturday), the next one 10 days and so on. Two limitations have to be accepted: if there is no data for a given Wednesday, a whole trading week has to be omitted,[13] as replacing missing Wednesdays by the preceding Tuesdays potentially doubles the number of different maturities in the sample. Quarterly options also have to be discarded, as they usually do not expire on a Saturday.

The second step is more delicate. In the COS expansion, γ_k is a function on the integration limits $[a, b]$, which in turn depend on the second and fourth cumulants of the risk neutral distribution, see (C.8) and (C.9):

$$\gamma_k = \frac{k}{b_t - a_t} \quad \text{with} \quad -a_t = b_t = L\sqrt{\kappa_{2,t}(\theta, \tau, X_t) + \sqrt{\kappa_{4,t}(\theta, \tau, X_t)}}$$

Thus $\underline{\gamma} = \underline{\gamma}(\theta, \tau, X_t)$, which makes re-using a previously calculated $A(\cdot), B(\cdot)$ impossible. The solution is a "one size fits all" approximation for a and b as a function of τ alone. We start from the observation that, for sufficiently long durations, κ_2 can be approximated by the variance level V_t times the duration:

$$\kappa_{2,t} \approx \tau \, V_t$$

Next, define $\overline{\kappa_2} = \tau \, \overline{V}$, with $\overline{V}$ the sample average of the short term, at the money variance. (In the benchmark example, $\overline{V} = 0.2$.) To accommodate the kurtosis created by jumps at short horizons, we correct τ for durations below 2 months: $\tau^c = \tau + 0.3(2/12 - \tau)$. We can now approximate the integration limits as function of τ exclusively

$$-a(\tau) = b(\tau) = L \cdot \sqrt{\overline{\kappa_2}} = L \cdot \sqrt{\tau^c \, \overline{V}} \tag{4.23}$$

This gives rise to the following algorithm:

Algorithm 3 (Fast Evaluation Scheme of the Likelihood Function)**.** Using approximation (4.23), the likelihood function (4.8) can be evaluated for a given parameter set θ and time series $\{X_t\}$ of the state.

---

[12] Further reducing the frequency to monthly reduces the number of possible maturities only to 26, due to the fact that some months have four and some have five Wednesdays.

[13] For the S&P 500, there are four such occasions in the seven-year period 1996-2002: christmas 1996 and 2002, 9/11/2001 and 7/4/2001.

	Quadrature	FFT	COS
Pricing one option chain	207 s	2.88 s	0.147 s
One evaluation of the likelihood function			
NLLS standard scheme	4954 hrs	69.5 hrs	3.5 hrs
NLLS advanced scheme	*n/a*	562 s	138 s
EKF standard scheme	495 hrs	7 hrs	0.4 hrs
EKF advanced scheme	*n/a*	191 s	26 s
Marginal cost for adding . . .			
One maturity	207 s	2.87 s	0.146 s
One state × maturity	207 s	4.76 ms	0.71 ms
One strike	9 s	1.25 μs	35 μs

Table 4.3: Execution time for one evaluation of the likelihood function and marginal computational costs for additional observations. FFT and COS methods use formulation (4.17) for $B(\tau)$ and apply the matrix rotation count algorithm.

```
1. Make a list of all option maturities $\tau_i$ in the data set.
2. For every $\tau_i$, with $1 \leq i \leq n_\tau$ do
      3. Calculate vector $\underline{\gamma_i}(\tau_i) = \frac{k}{b(\tau_i)-a(\tau_i)}$ based on (4.23)
      4. Calculate $A(\tau_i, \underline{\gamma_i}), B(\tau_i, \underline{\gamma_i})$ using Algorithm 1, save in list $L_1$
   enddo
4. For every trading day $t$ with $1 \leq t \leq T$ do
      5. For every duration $\tau_j$ on day $t$ with $1 \leq j \leq n_{\tau,t}$ do
            6. Retrieve elements $A(\tau_j), B(\tau_j)$ from list $L_1$
            7. Calculate $\Psi_{Levy} = \exp(tr(A(\tau_j)X_t + B(\tau_j))$
            8. Perform the Fourier inversion
         enddo
   enddo
```

Remark 1. To accommodate the NLLS or EKF estimation strategy, simply augment steps $5.-8.$ As the parameter-dependent part of the Laplace transform is evaluated in steps $1.-3.$, the additional cost of the nested optimization is negligible. In the case of the NLLS scheme, the objective function of the inner optimization obtains from evaluating steps $5.-8.$ The state $\widehat{X}_t$ is estimated from an optimization over this objective function, separately day-by-day. In the case of the Kalman filter, the Jacobian matrix is calculated via numeric differentiation. For this, steps $7.-8.$ are repeated with $X_t + hE_j$, where E_j is a matrix that is one at the location of state component j and zero otherwise.

Remark 2. Table 4.3 presents a detailed timing analysis of the speed increase enabled

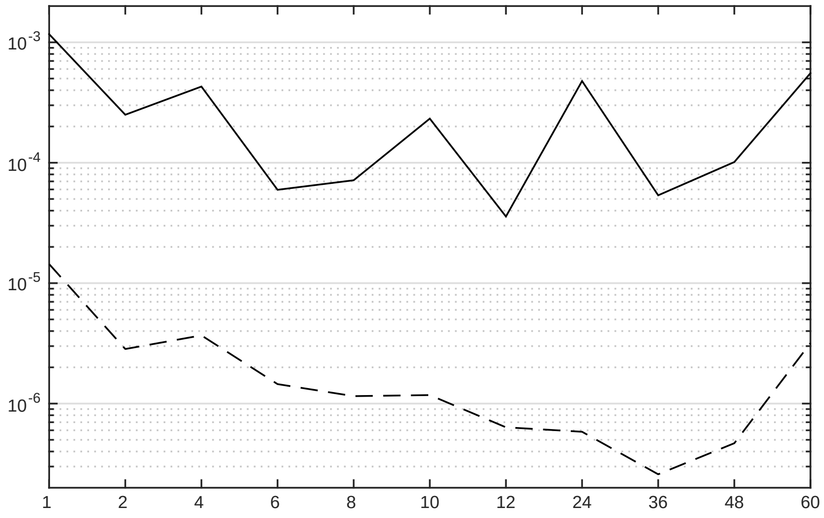


Figure 4.5: Error analysis of the fast evaluation scheme for the benchmark 2 × 2 model. Average (dashed line) and the maximal (solid line) absolute implied volatility approximation error as function of maturity τ in months. Volatilities in the range from 10% to 63% and strikes in the range $0.05 \leq |\Delta| \leq 0.95$ were considered for each τ. Furthermore, all admissible volatility structures are scanned using the polar coordinate representation (4.30). See Tab. C.1 in the Appendix for the exact parameters used to create this figure.

by Algorithm 3. Beyond a vast cut in execution time, we observe a vast improvement of the scaling properties of the likelihood function with respect to the dimension of the process, n. For the Kalman Filter approach, the number of evaluations of the parameter-dependent part of the Laplace transform is reduced from $M^{\mathcal{L}}_{EKF} = T\, n_\tau\ (2 + n(n+1)/2)$ to 52. Already in the 2 × 2 benchmark case, this is a reduction by a factor of 168. The improvement for the NLLS approach is even larger, as the cost for the nested (inner) optimization is now greatly reduced. As a consequence, the difference in computational cost between NLLS and Kalman Filter shrinks from 10:1 to 2:1.

4.3.3 Error Analysis

Fixing the integration limits regardless of the variance level involves a trade-off between two types of error. In the case of high volatility, the integration limits are too tight, which results in an analytical error from the truncation of tails of the risk-neutral distribution. In the case of low volatility, the integration limits are too wide, which yields to a numerical error. To assess the error induced by fixing the integration limits to the values in (4.23), I perform a detailed of the pricing behavior for the benchmark model.

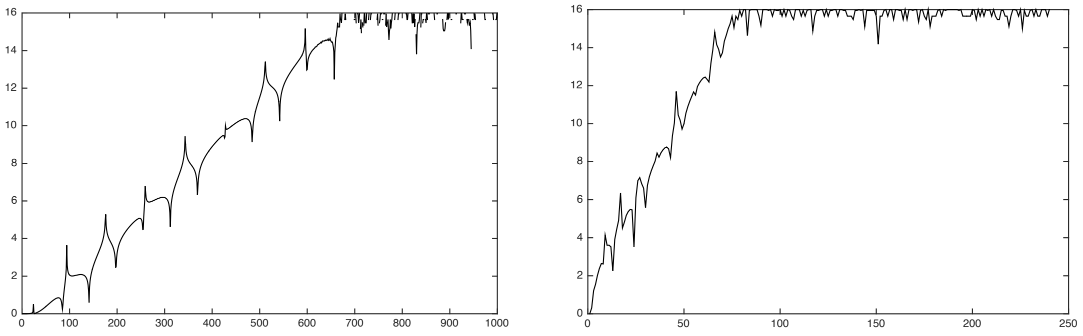


Figure 4.6: Loss of precision in calculating $tr[\log\, C_{22}(\gamma_k)]$ for the COS inversion. The graph depicts the number of significant digits (out of 16) lost when calculating $tr[\log\, C_{22}]$ as a function of k in the COS inversion. Left: $\tau = 1yr$ with integration limits as defined in (4.23). Right: $\tau = 5yr$ using sub-optimal integration limits (4.23) $a(\tau), b(\tau)$ with $\tau = 1\, yr$.

Using the polar coordinate representation (4.30), I create state matrices with variances between 0.01 and 0.4 (volatilities from 10% to 63%) and all admissible volatility structures, see Tab. C.1 for the exact quantities used. I test the pricing for strikes in the range $0.05 \leq |\Delta| \leq 0.95$ (puts and calls) and durations from 1 month to 5 years.

The absolute approximation error of using integration limits (4.23) in terms of implied volatility as a function of τ is quantified in Figure 4.5. The maximum value is always below 10^{-3} volatility percentage points, i.e. below one tenth of a volatility basis point, and the mean is typically below one-thousandth of a volatility basis point. Given that the best model in Gruber et al. (2010) has an mean absolute implied volatility error ($MAIV$) of 69 volatility basis points, the numerical approximation error is typically four to five orders of magnitude smaller than the model error.

4.4 Loss of Precision in $tr[\log(C_{22})]$

A numerical instability in the calculation of $tr[\log(C_{22})]$ arises, if the Laplace transform is evaluated for unsuitable (usually too large) values of γ. The problem is best analyzed for the 2×2 case using the identity $\log(\det(C_{22})) = tr[\log(C_{22})]$, though it does not, depend on the way how $tr[\log(C_{22})]$ is calculated. The determinant of C_{22}

$$\det(C_{22}) = C_{22}^{11} C_{22}^{22} - C_{22}^{12} C_{22}^{21} \tag{4.24}$$

contains a subtraction. For large values of γ, the two terms in (4.24) have very similar values. The numerical error induced by this subtraction can be assessed using the loss of precision theorem, which states that in a subtraction $x - y$ with $0 < y < x$, the

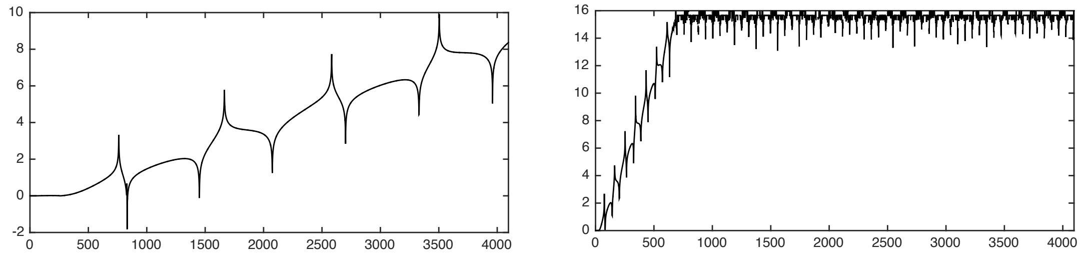


Figure 4.7: Loss of precision in calculating $tr[\log C_{22}(\gamma)]$ for the FFT inversion. The graph depicts the number of significant digits lost (out of 16) when calculating $tr[\log C_{22}]$ as a function of k in the FFT inversion for $\tau = 1\,yr$. Left: upper integration limit $a = 100$. Right: upper integration limit $a = 1024$, the original value of Carr and Madan (1999).

number of significant digits lost is

$$q \approx -log_{10}(1 - \frac{y}{x}) \tag{4.25}$$

To avoid loss of precision in (4.24), too large values of γ should be avoided in the evaluation of the Laplace transform. Figure 4.6 shows the number of significant digits lost (out of 16 for double precision numerics), for the COS inversion. The left panel is evaluated at a duration of 1 year, using the approximated integration limits (4.23). Up to six digits are lost, if we truncate the COS series (C.7) at a typical value of $N = 250$.

This affords a sufficient precision for option prices.[14] However, we observe the phenomenon that adding more terms to the COS series (C.7) can lead to worse results: for $N > 600$, some elements of the COS expansion loose all significant digits, resulting in a logarithm of zero and option prices that cannot be calculated. The right panel of Figure 4.6 also highlights the importance of the correct choice of integration limits. If we perform the COS inversion for a maturity of 5 years using the (incorrect) integration limits (4.23) for $\tau = 1\,yr$, the argument γ_k of the Laplace transform becomes too large by a factor ≈ 2.3 (see Table 4.1), and all significant digits are lost for $N > 80$.

Worse results obtain for the FFT method, where the upper limit of the integration $a = N \cdot \eta$ (in the original notation of Carr and Madan) needs to be comparatively high to dampen the effects of the oscillating summands. Carr and Madan propose $N = 4096$ and $\eta = 0.25$, i.e. $a = 1024$. Figure 4.7 shows that for $\tau = 1\,yr$, an upper limit of 1024 leads to a loss of all significant digits for $N > 600$. Numerical experiments show that a reasonable choice of a for the FFT is $100/\tau$, leading to a loss of up to 8 digits.

---

[14] Numerical experiments show that a similar number of digits is lost for durations from 1 month to 5 years, when using integration limits (4.23).

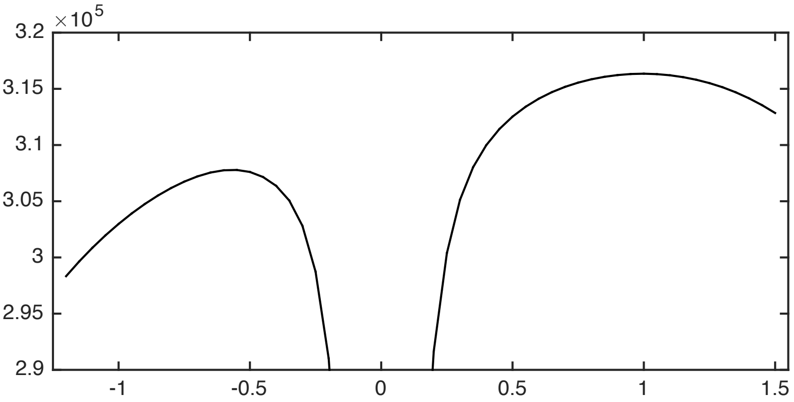


Figure 4.8: Example for the non-convexity of the objective function: In-sample likelihood of the benchmark model as function of Q_{11}, expressed in multiples of the estimated value $\widehat{Q}_{11}$. All other parameters are kept at their point estimates.

4.5 Numerical Problems in the Model Estimation

4.5.1 Stochastic Optimization

The estimation of the MAJD model is formulated in (4.6) as a nested optimization problem:

$$\widehat{\theta} = \arg\max_{\theta} \mathcal{L}\Big(\theta, \arg\max_{\{X_t\}} \mathcal{L}(\theta, \{X_t\}; \mathcal{O}); \mathcal{O}\Big) \ .$$

Following the convention in the optimization literature, I implement the estimator $\hat{\theta}$ by minimizing the negative likelihood. There is no theoretical evidence that this optimization problem is convex, on the contrary. As a simple counter-example, Figure 4.8 depicts the likelihood of the benchmark model as a function of parameter Q_{11}, expressed in multiples of the point estimate $\widehat{Q}_{11}$. Besides the obvious maximum at 1 there is a strong local maximum around -0.55. The presence of this local maximum precludes the use of convex optimization algorithms.

Furthermore, little is ex ante known about the role and interpretation of the parameter matrices. In independent component models of the Bates (2000)-type, the signs of almost all parameters obtain from admissibility criteria, which helps guessing good starting values for a (locally) convex optimization. Such criteria do not exist for the out-of-diagonal elements of the parameter matrices in the model SVJ_{31}.

Both problems – lack of good starting values and non-convexity – can be attacked by using population-based stochastic optimization algorithms such as differential evolution of Storn and Price (1997). This non-convex optimization algorithm only requires an interval of plausible starting values for each element of the parameter vector. The elevated computational cost of this algorithm – Storn and Price propose a population size of five to ten times the number of parameters and 100 generations – is accommodated thanks to the speed increases afforded by the Matrix Rotation Count algorithm and the Fast Evaluation Scheme of the likelihood function. Moreover, differential evolution can be parallelized up to the order of the population size.

To increase the precision of the point estimate, the result of the stochastic optimization is used as starting value for a simplex optimization using the algorithm of Nelder and Mead (1965), assuming local convexity of the likelihood function around the optimum.

4.5.2 Matrix Representations and Optimization Constraints

The standard convention in the optimization literature is optimization over a parameter vector. As the MAJD model features several parameter matrices and a state matrix, we encode the argument of the objective function as a vector. In the unconstrained case, this can be done in a straight-forward manner using the vectorization function *vec* viz. the half-vectorization function *vech* for symmetric matrices. The parameter vector is then $\theta = \{vec(M), vec(R),\ vec(Q),\ \beta, \lambda_0, vec(\Lambda), \theta_\xi; vech(X_t)\}$ where θ_ξ denotes the parameters of the return jump distribution.

Considerations of identification[15] and admissibility require the following constraints:

- Positive definiteness and symmetry of the state X_t,
- Negative definiteness and lower triangularity of M,
- Positive definiteness of $\mathrm{tr}(\Lambda X_t)$ for all X_t,
- Existence of $\sqrt{Id_n - R'R}$

Some constraints are tedious to implement in cartesian coordinates. For example the positive definiteness of a $n \times n$ state matrix implies $n(n+1)/2$ constraints.[16]

The following two matrix reparametrizations allow for simpler constraints and support a better understanding of the volatility structure. The first one transforms a constrained optimization problem into an unconstrained one. The second one transforms some unbounded variables to bounded ones.

Choleski Decomposition

The Choleski decomposition of a symmetric, positive definite matrix X is the unique triangular matrix D such that

$$X = DD' \tag{4.26}$$

---

[15] See (Gruber et al. 2010) for a detailed discussion of model identification.

[16] In the 2×2 case, these are: $X_{11} > 0$; $X_{22} > 0$ and $X_{11}X_{22} > (X_{21})^2$.

Conversely, the product DD' of any matrix D will be a positive definite, symmetric matrix. The Choleski decomposition is therefore a useful representation of the state matrix. It allows for an unconstrained optimization over $vech(D)$ using definition (4.26) to recover the state X.

Eigendecomposition and Polar Coordinate Representation of the State

Any square matrix A with linearly independent eigenvectors can be written as

$$A = PDP^{-1} \tag{4.27}$$

where D is a diagonal matrix of the eigenvalues of A and P is a matrix containing the normalized eigenvectors of A as column vectors.

The PDP decomposition makes it straight-forward to implement definiteness constraints on non-symmetric matrices by constraining the sign of the diagonal elements of D.

The eigendecomposition can be taken further by representing the eigenvectors in terms of polar coordinates, as any n-dimensional normalized eigenvector can be expressed in polar coordinates using $n-1$ angles. For a general 2×2 matrix, $A = PDP^{-1}$ can be written as:

$$A = \begin{pmatrix} \sin\alpha_1 & \sin\alpha_2 \\ \cos\alpha_1 & \cos\alpha_2 \end{pmatrix} \begin{pmatrix} D_{11} & 0 \\ 0 & D_{22} \end{pmatrix} \begin{pmatrix} \sin\alpha_1 & \sin\alpha_2 \\ \cos\alpha_1 & \cos\alpha_2 \end{pmatrix}^{-1} \tag{4.28}$$

Expression (4.30) can be simplified for symmetric matrices like the state X, as $\cos\alpha_1 = \sin\alpha_2$ implies $\alpha_2 = \alpha_1 + \pi/2$. Using the symbol

$$\xi := \frac{D_{11}}{tr(X)} = \frac{D_{11}}{D_{11} + D_{22}} \tag{4.29}$$

we define the following mapping the 2×2 state matrix X_t

$$(X_{11,t}, X_{12,t}, X_{22,t}) \to (V_t, \xi_t, \alpha_{1,t}) \tag{4.30}$$

where $V := tr(X)$ is the (diffusive) variance. The remaining two variables describing the volatility structure are bounded: $0 \leq \xi \leq 1$ and $0 \leq \alpha_1 \leq \pi$. This boundedness of ξ, α_1 makes it possible to scan the space of admissible state matrices for a given volatility level.

Representation (4.30) can be used to separate variance level V_t from variance structure $(\xi_t, \alpha_{1,t})$.

4.6 Conclusion

The useful properties of matrix affine jump diffusions (MAJD), introduced by Leippold and Trojani (2008), come at the price that models of this class are notoriously difficult to estimate. MAJD models have higher computational cost for three reasons. Firstly, the Laplace transform contains matrix exponentials and logarithms, which increase the computational cost by two orders of magnitude. Secondly, the matrix nature of the parameter matrices makes it impossible to guess starting parameters of a (locally) convex optimization, requiring stochastic optimization. Thirdly, the comparatively high dimension of the state matrix makes it impossible to alternate between state and parameter estimation, requiring a nested joint estimation instead.

This paper addresses the challenges in the estimation and evaluation of MAJD models and reduces the computational cost in several steps. First I formulate the Matrix Rotation Count algorithm, which makes it possible to use of transform methods, resulting in a 1400-fold speed increase. Next, I propose a fast evaluation scheme of the likelihood function based on an approximation of the integration limits in the COS inversion method. This evaluation scheme reduces the number of required evaluations of the Laplace transform, resulting in a 50- to 500-fold speed increase, depending on the estimation procedure. Combined with moderate use of parallelization in the differential evolution optimization scheme, I achieve a reduction in the estimation time of a factor of approximately 10^6.

I also highlight a numerical instability of the Laplace transform, that requires a deliberate choice of integration bounds, especially when applying the FFT method of Carr and Madan (1999). This instability has the counter-intuitive consequence that fewer elements of the Fourier sum result in a more precise result.

I finally propose a reformulation of the Matrix state space in terms of polar coordinates that allows for an elegant implementation of identification and admissibility constraints for the matrix optimization.

Appendix C

Appendix for “Eliciting a Smile”

C.1 The COS Method

The Cosine-Fast Fourier inversion (COS) method, introduced by Fang and Oosterlee (2008), is an efficient algorithm to approximate option prices given the Laplace transform. It reduces the number of required evaluations from 2^{12} for a standard FFT, e.g. Carr and Madan (1999), to typically 200. A second useful property of the Cosine-FFT method is the fact that it does not involve an interpolation between strikes, which allows for a more precise calculation of skewness measures like $\mathcal{S}_t = \lim_{T \to t} \frac{\partial IV(T,S_t)}{\partial K}$.

The COS algorithm approximates a (density) function $f(y;\tau)$ on a finite support $[a,b]$ via a truncated cosine-series expansion:

$$f(y;\tau) = {\sum_{k=0}^{\infty}}' A_k(\tau) \cos\left(k\pi \frac{y-a}{b-a}\right) \approx {\sum_{k=0}^{N-1}}' A_k(\tau) \cos\left(k\pi \frac{y-a}{b-a}\right) \tag{C.1}$$

with coefficients

$$\begin{aligned} A_k(\tau) &= \frac{2}{b-a} \int_a^b f(y;\tau) \cos\left(k\pi \frac{y-a}{b-a}\right) dy \\ &\approx \frac{2}{b-a} Re\left[\int_{\mathbb{R}} f(y;\tau) \exp\left(ik\pi \frac{y-a}{b-a}\right) dy\right] \\ &\approx \frac{2}{b-a} Re\left[\phi\left(\frac{k\pi}{b-a};\tau\right) \exp\left(-i\frac{k\pi a}{b-a}\right)\right] \end{aligned} \tag{C.2}$$

where $\phi(\cdot)$ denotes the Laplace transform of density $f(\cdot)$.

The price $O(X_t; K, \tau)$ of a contingent claim with payoff $v_t(y; K)$ and time to maturity $\tau = T - t$ is:

$$\begin{aligned} O(X_t; K, \tau) &= e^{-r\tau} \int_{-\infty}^{\infty} v(y; K) f(y|X_t; \tau) dy && \text{(C.3)} \\ &\approx e^{-r\tau} \int_a^b v(y; K) \sum_{k=0}^{N-1}{}' A_k(\tau) \cos\left(k\pi \frac{y-a}{b-a}\right) dy \\ &= e^{-r\tau} \sum_{k=0}^{N-1}{}' \frac{1}{2}(b-a) A_k(\tau) \int_a^b \frac{2}{b-a} v(y; K) \cos\left(k\pi \frac{y-a}{b-a}\right) && \text{(C.4)} \end{aligned}$$

where $y = \ln(S_T/K)$ and we assume a, b have been chosen such that $f(y; \tau) \approx 0$ outside $[a, b]$.

For a plain vanilla call viz. put with payoffs $v^c(y; K) = [K(e^y - 1)]^+$ and $v^p(y) = [K(1 - e^y)]^+$ the integral $U_k := \frac{1}{K} \int_a^b \frac{2}{b-a} v(y; K) \cos\left(k\pi \frac{y-a}{b-a}\right) dy$ evaluates as

$$U_k^c = \frac{2}{b-a} \left(\chi_k(0, b) - \psi_k(0, b)\right) \quad \text{and} \quad U_k^p = \frac{2}{b-a} \left(-\chi_k(a, 0) + \psi_k(a, 0)\right)$$

with

$$\begin{aligned} \chi_k(c, d) &= \int_c^d e^y \cos\left(k\pi \frac{y-a}{b-a}\right) dy \\ &= \frac{1}{1 + \left(\frac{k\pi}{b-a}\right)^2} \Bigg[\cos\left(k\pi \frac{d-a}{b-a}\right) e^d - \cos\left(k\pi \frac{c-a}{b-a}\right) e^c \\ &\qquad + \frac{k\pi}{b-a} \sin\left(k\pi \frac{d-a}{b-a}\right) e^d - \frac{k\pi}{b-a} \sin\left(k\pi \frac{c-a}{b-a}\right) e^c\Bigg] \\ \psi_k(c, d) &= \int_c^d \cos\left(k\pi \frac{y-a}{b-a}\right) dy = \begin{cases} \frac{b-a}{k\pi} \left[\sin\left(k\pi \frac{d-a}{b-a}\right) - \sin\left(k\pi \frac{c-a}{b-a}\right)\right] & k \neq 0 \\ d - c & k = 0 \end{cases} \end{aligned}$$

This leaves us with

$$O(X_t; K, \tau) = e^{-r\tau} K \sum_{k=0}^{N-1}{}' \frac{1}{2}(b-a) A_k(\tau) U_k \tag{C.5}$$

This expression is not very efficient: We would have to recalculate A_k for every strike, as it depends on $y = \ln(S_T/K)$ via the characteristic function. We therefore need to separate the contract-dependent part of the Laplace transform from the model-dependent part. In our case, this Laplace transform (4.12) can be separated as

$$\Psi(\gamma; \tau, X_t) = \exp\left(\gamma y + tr[A(\gamma; \tau) X_t] + B(\gamma; \tau)\right) = \exp(\gamma y)\, \Psi_{Levy}(\gamma; \tau) \tag{C.6}$$

with the according characteristic function $\phi(u) = \exp(iuy)\phi_{Levy}(u)$. We now insert (C.2) into (C.5) to obtain the COS pricing formula

$$\begin{aligned}
O(X_t; K, \tau) &= e^{-r\tau} K {\sum_{k=0}^{\infty}}' \frac{1}{2}(b-a)\frac{2}{b-a} Re\left\{\phi\left(\frac{k\pi}{b-a};\tau\right)\exp\left(-i\frac{k\pi a}{b-a}\right)\right\} U_k \\
&= e^{-r\tau} K {\sum_{k=0}^{N-1}}' Re\left\{\phi_{Levy}\left(\frac{k\pi}{b-a};\tau\right)\exp\left(i\frac{k\pi y}{b-a}\right)\exp\left(-i\frac{k\pi a}{b-a}\right)\right\} U_k \\
&= e^{-r\tau} K {\sum_{k=0}^{N-1}}' Re\left\{\phi_{Levy}\left(\frac{k\pi}{b-a};\tau\right)\exp\left(ik\pi\frac{y-a}{b-a}\right)\right\} U_k \qquad \text{(C.7)}
\end{aligned}$$

Expression (C.7) can be evaluated with arbitrary precision for any strike K without further interpolation. Note that the evaluation of $\phi_{Levy}(u;\tau)$ and U_k is independent of y and therefore need not be repeated for different strikes.

The choice of the integration limits a, b is governed by a trade-off between two errors: too tight limits introduce an analytical error from cutting off the tails of the distribution. Too wide limits cause a numerical error as the elements of the COS expansion are not employed efficiently. Fang and Oosterlee (2008) propose as integration limits

$$[a, b] = \left[\kappa_1 - L\sqrt{\kappa_2 + \sqrt{\kappa_4}},\ \kappa_1 + L\sqrt{\kappa_2 + \sqrt{\kappa_4}}\right] \qquad \text{(C.8)}$$

with κ_i denoting cumulant i of the risk-neutral distribution and the scaling factor L chosen to be 10. All cumulants of the returns distribution and therefore the integration limits depend on the parameter vector θ, the duration τ and on the state X_t.

The values of γ_k also depend on a, b and are

$$\gamma_k = \frac{k}{b(\theta,\tau,X_t) - a(\theta,\tau,X_t)}, \qquad 0 \le k \le N-1. \qquad \text{(C.9)}$$

C.2 Additional Table

Quantity	Values
Variance	0.01, 0.02, 0.05, 0.1, 0.3
Strike [Δ]	0.05, 0.1 $\ldots$ 0.95
Time to maturity [months]	1, 2, 4, 6, 8, 10, 12, 24, 36, 48, 60
Eigenvalue ratio ξ	0, 0.25, 0.5
Eigenvector angle α_1	0, $\pi/8$, $\ldots$, π

Table C.1: The error analysis reported in Fig. 4.5 is based on the evaluation of 32’076 option prices, namely all combinations of the above quantities. The eigenvalue ratio ξ and the eigenvector angle α_1 are defined in (4.29) and (4.28).

Bibliography

Adrian, Tobias and Hyun Song Shin, "Liquidity and Leverage," *Journal of Financial Intermediation*, 2010, *19* (3), 418–437.

——— **and Joshua Rosenberg**, "Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk," *The Journal of Finance*, 2008, *63* (6), 2997–3030.

———**, Emanuel Moench, and Hyun Song Shin**, "Leverage asset pricing," *Staff Reports 625, Federal Reserve Bank of New York*, 2013.

Ait-Sahalia, Y., M. Karaman, and L. Mancini, "The Term Structure of Variance Swaps, Risk Premia and the Expectation Hypothesis," *SSRN Working Paper*, 2012.

Andersen, T.G., N. Fusari, and V. Todorov, "Parametric inference and dynamic state recovery from option panels," *Econometrica, forthcoming*, 2015.

Andersen, Torben, Nicola Fusari, and Viktor Todorov, "The Risk Premia Embedded in Index Options," *Journal of Financial Economics*, forthcoming.

Attari, Mukarram, "Option Pricing Using Fourier Transforms: A Numerically Efficient Simplification," *SSRN Working Paper 55042*, 2004.

Backus, David, Silverio Foresi, Kai Li, and Liuren Wu, "Accounting for Biases in Black-Scholes," *CRIF Working Paper, November 1997*, 1997.

Bakshi, Gurdip and Nikunj Kapadia, "Delta-Hedged Gains and the Negative Market Volatility Risk Premium," *Review of Financial Studies*, 2003, *16(2)*, 527–566.

———**, Charles Cao, and Zhiwu Chen**, "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance*, 1997, *52 (5)*, 2003–2049.

Bates, David, "Jumps and stochastic volatility: Exchange rate processes implicit in Deutschemark options," *Review of Financial Studies*, 1996, *9*, 69–108.

——, "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics*, 2000, *94* (1), 181–238.

Bekaert, Geert and Marie Hoerova, "The VIX, the Variance Premium and Stock Market Volatility," *Journal of Econometrics*, 2014, *183* (2), 181–192.

Benabid, Anas, Harry Bensusan, and Nicole El Karoui, "Wishart stochastic volatility : Asymptotic smile and numerical framework," *Working paper March 2009*, 2009.

Bollerslev, Tim and Viktor Todorov, "Tails, Fears, and Risk Premia," *The Journal of Finance*, 2011, *66* (6), 2165–2211.

——, George Tauchen, and Hao Zhou, "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 2009, *22* (11), 4465–4492.

——, ——, and ——, "Expected Stock Returns and Variance Risk Premia," *The Review of Financial Studies*, 2009, *22* (11), 4463–4492.

——, James Marrone, Lai Xu, and Hao Zhou, "Stock Return Predictability and Variance Risk Premia: Statistical Inference and International Evidence," *Journal of Financial and Quantitative Analysis*, 2014, *49* (3), 633–661.

Bondarenko, O., "Variance Trading and Market Price of Variance Risk," *Journal of Econometrics*, 2014.

Britten-Jones, Mark and Anthony Neuberger, "Option Prices, Implied Price Processes, and Stochastic Volatility," *Journal of Finance*, 2000, *55*, 839–866.

Broadie, Mark, Mikhail Chernov, and Michael Johannes, "Model Specification and Risk Premia: Evidence from Futures Options," *Journal of Finance*, 2007, *62 (3)*, 1453–1490.

Bru, M.-F., "Wishart Processes," *Journal of Theoretical Probability*, 1991, *4*, 725–751.

Buraschi, A., P. Porchia, and F. Trojani, "Correlation Risk and Optimal Portfolio Choice," *Journal of Finance*, 2010, *65* (1), 393–420.

Carr, P. and D. Madan, "Option valuation using the fast Fourier transform," *Journal of Computational Finance*, 1999, *2*, 61–73.

—— and L. Wu, "Variance Risk Premiums," *Review of Financial Studies*, 2009, *22* (3), 1311–1341.

Carr, Peter and D. Madan, "Towards a theory of volatility trading," *in R. Jarrow (ed) Risk Book on Volatility. New York: Risk, pp 417-27*, 1998.

____ **and Liuren Wu**, "Time-Changed Levy Processes and Option Pricing," *Journal of Financial Economics*, 2004, *17*, 113–141.

____ **and** ____ , "A tale of two indices," *The Journal of Derivatives*, 2006, *13*, 13–29.

____ **and** ____ , "Stochastic Skew in Currency Options," *Journal of Financial Economics*, 2007, *86 (1)*, 213–247.

____ **and** ____ , "Leverage Effect, Volatility Feedback, and Self-Exciting Market Disruptions: Disentangling the Multi-dimensional Variations in S&P500 Index Options," *Bloomberg Portfolio Research Paper No. 2009-03*, 2009.

____ **and Roger Lee**, "Volatility Derivatives," *Annual Review of Financial Economics*, 2009, *1*, 1–21.

Chen, Ren-Raw and Louis Scott, "Pricing Interest Rate Options in a Two-Factor Cox-Ingersoll-Ross Model of the Term Structure," *Review of Financial Studies*, 1992, *5* (613-636).

Cheridito, P., D. Filipovic, and R. L. Kimmel, "Market Price of Risk Specifications for Affine Models: Theory and Evidence," *Journal of Financial Economics*, 2007, *83*, 123–170.

Chernov, M. and E. Ghysels, "A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of option valuation," *Journal of Financial Economics*, 2000, *56*, 407–458.

Christoffersen, Peter, Kris Jacobs, and Karim Mimouni, "Models for S&P500 Dynamics: Evidence from Realized Volatility, Daily Returns, and Option Prices," *Review of Financial Studies*, 2010, *23* (8), 3141–3189.

____ **, Steve Heston, and Kris Jacobs**, "The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well," *Management Science*, 2009, *55*, 1914–1932.

Chung, S.-L., W.-C. Tsai, Y.-H. Wang, and P.-S. Weng, "The information content of the S&P 500 index and VIX options on the dynamics of the S&P 500 indeX," *Journal of Futures Markets*, 2011, *31(12)*, 11701201.

Cont, Rama and Sana Ben Hamida, "Recovering volatility from option prices by evolutionary optimization," *Journal of Computational Finance, Vol. 8, No. 4, Summer 2005*, 2005, *8 (4).*

Cox, J. C., J. E. Ingersoll, and S. A. Ross, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 1985, *53*, 385–407.

da Fonseca, J., M. Grasselli, and C. Tebaldi, “A Multifactor Volatility Heston Model,” *Quantitative Finance*, 2008, *8 (6)*, 591–604.

Dai, Q. and K. Singleton, “Specification analysis of affine term structure modes,” *Journal of Finance*, 2000, *50*, 1943–1978.

Demeterfi, Kresimir, Emanuel Derman, Michael Kamal, and Joseph Zou, “A Guide to Volatility and Variance Swaps,” *The Journal of Derivatives*, 1999, *6(4)*, 9–32.

Duffie, Darrell, Jun Pan, and Kenneth Singleton, “Transform analysis and asset pricing for affine jump-diffusions,” *Econometrica*, 2000, *68 (6)*, 1343–1376.

Dumas, Bernard, Jeff Fleming, and Robert E. Whaley, “Implied Volatility Functions: Empirical Tests,” *Journal of Finance*, 1998, *53 (6)*, 2059–2106.

Dupire, Bruno, “Model art,” *Risk*, 1993, *September*, 118–120.

Durrleman, Valdo, “From Implied to Spot Volatilities,” *Finance and Stochastics*, 2010, *14* (2), 157–177.

____ **and Nicole El Karoui**, “Coupling Smiles,” *Quantitative Finance*, 2008, *8* (6), 573–590.

Eraker, Bjørn, “Do Stock Prices and Volatility Jump? Reconciling eidence from Spot and Option Prices,” *Journal of Finance*, 2004, *59* (3), 1367–1404.

____ **, Michael Johannes, and Nicholas Polson**, “The Impact of Jumps in Equity Index Volatility and Returns,” *Journal of Finance*, 2003, *58* (3), 1269–1300.

Fang, Fang and Kees Oosterlee, “A novel pricing method for European Options based on Fourier-Cosine Expansions,” *SIAM Journal on Scientific Computing*, 2008, *31(2)*, 826–848.

Filipovic, Damir, Elise Gourier, and Loriano Mancini, “Quadratic Variance Swap Models,” *Journal of Financial Economics, forthcoming*, 2015.

Gourieroux, Christian, “Continuous Time Wishart Process for Stochastic Risk,” *Econometric Reviews*, 2006, *25*, 177–217.

Gruber, P., C. Tebaldi, and F. Trojani, “Three Make a Smile Dynamic Volatility, Skewness and Term Structure Components in Option Valuation,” 2010. Working Paper. University of Lugano and Bocconi University.

Gruber, Peter H., "Eliciting a Smile: Numerical Methods for Option Pricing with Matrix Affine Jump Diffusions," *Working paper, Universita della Svizzera Italiana, Lugano*, 2015.

Heston, S., "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 1993, *6*, 327–343.

Huang, Jing-Zhi and Liuren Wu, "Specification analysis of option pricing models based on time-changed Levy processes," *Journal of Finance*, 2004, *59*, 1405–39.

Hull, J. and A. White, "An Analysis of the Bias in Option Pricing Caused by a Stochastic Volatility,," *Advances in Futures and Options Research*, 1988, *3*, 29–61.

Jiang, George and Yisong S. Tian, "Extracting Model-Free Volatility from Option Prices: An Examination of the Vix Index," *Journal of Derivatives*, 2007, *14*, 1–26.

Johannes, Michael, Nicholas Polson, and Jon Stroud, "Learning about Jumps and Stochastic Volatility: Filtering Stochastic Differential Equations with Jumps," *Review of Financial Studies*, 2010, *forthcoming.*

Johnson, Travis L., "Equity Risk Premia and the VIX Term Structure," *Working Paper Jan 2012*, 2012.

Jones, Christopher S., "The dynamics of stochastic volatility: evidence from underlying and options markets," *Journal of Econometrics*, 2003, *116*, 181–224.

Kahl, Christian and Peter Jäckel, "Not-so-complex logarithms in the Heston model," *Wilmott*, September 2005, pp. 94–103.

Kozhan, Roman, Anthony Neuberger, and Paul Schneider, "Understanding Risk Premia in Index Option Prices," *SSRN Working paper August 12, 2010*, 2010.

Leippold, Markus and Fabio Trojani, "Asset Pricing with Matrix Jump Diffusions," *SSRN eLibrary*, 2008.

Li, Minqiang and Neil D. Pearson, "Price Deviations of S&P 500 Index Options from the Black-Scholes Formula Follow a Simple Pattern," *MPRA Resarch Paper 11530, University Library of Munich*, 2008.

Lord, Roger and Christian Kahl, "Complex logarithms in Heston-like models," *Mathematical Finance*, 2010, *20* (4), 671–694.

Martin, Ian, "Simple Variance Swaps," Working Paper 16884, National Bureau of Economic Research March 2012.

Melino, A. and S. Turnbull, "Pricing foreign Curency options with Stochastic Volatility," *Journal of Econometrics*, 1990, *45*, 239–265.

Merton, Robert C., "Option pricing when underlying stock returns are discontinuous," *Journal of Financial Economics*, 1976, *3*, 125?143.

Meyerhofer, E., "Wishart processes and Wishart Distributios: An affine Process Point of View," *Lecture Notes*, 2014.

Muhle-Karb, Johannes, Oliver Pfaffel, and Robert Stelzer, "Option pricing in multivariate stochastic volatility models of OU type," *Working paper*, 2010.

Muir, Tyler, "Financial Crises, Risk Premia, and the Term Structure of Risky Assets," *Job Market Paper*, 2013.

Nandi, S., "How important is the correlation between returns and volatility in a stochastic volatility model? Empirical evidence form pricing and hedging in the S&P 500 Index," *Journal of Banking and Finance*, 1998, *22*, 589–610.

Nelder, J.A. and R. Mead, "A simplex method for function minimization," *The Computer Journal*, 1965, *7*, 308–313.

Neuberger, A., "The Log Contract," *Journal of Protfolio Management*, 1994, *20* (2), 74–80.

Neuberger, Anthony, "Realized Skewness," *SSRN eLibrary*, 2012.

Pan, Jun, "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study," *Journal of Financial Economics*, 2002, *63* (1), 3–50.

Schneider, Paul and Fabio Trojani, "Divergence And The Price Of Uncertainty," *Working Paper*, 2014.

___ **and** ___ , "Fear Trading," *Working Paper*, 2014.

Schöbel, Rainer and Jianwei Zhu, "Stochastic Volatility With an Ornstein-Uhlenbeck Process: An Extension," *European Finance Review*, 1999, *3*, 23–46.

Storn, Rainer and Kenneth Price, "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous," *Journal of Global Optimization*, 1997, *11*, 341–359.

Todorov, Viktor, "Variance Risk-Premium Dynamics: The Role of Jumps," *Review of Financial Studies*, 2010, *23* (1), 345–383.

Wu, Liuren, "Variance dynamics: Joint evidence from options and high-frequency returns," *Journal of Econometrics*, 2011, *160* (1), 280–287.