Linear state models for volatility estimation and prediction

Richard Hawkes Paresh Date

June 16, 2006

Abstract

This report covers the important topic of stochastic volatility modelling with an emphasis on linear state models. The approach taken focuses on comparing models based on their ability to fit the data and their forecasting performance. To this end several parsimonious stochastic volatility models are estimated using realised volatility, a volatility proxy from high frequency stock price data. The results indicate that a hidden state space model performs the best among the realised volatility-based models under consideration. For the state space model different sampling intervals are compared based on in-sample prediction performance. The comparisons are partly based on the multi-period prediction results that are derived in this report.

Keywords: Stochastic Volatility, high frequency data, linear models, multi-period prediction, forecasting performance.

Contents

Chapter 1

Introduction

In finance and business it is clear that uncertainty is ever present. The introduction of preventive measures generally does not obliterate this uncertainty although it may help to mitigate it. What is left must at least be quantified if possible. Quantified uncertainty is denominated risk. Modelling risk can take on two distinct forms: parametric and non-parametric modelling. In the first a model is assumed in terms of an equation or set of equations that are assumed to describe the quantity being modelled in terms of a parameter set. In the second the quantity of interest is estimated from the data using a variety of statistical techniques that do not assume a parameterised form.

The main thrust of this report is that of modelling financial risk in general and of volatility in particular. Financial risk is defined in (Watsham, 1998) as the undesirable change in the value of a financial commitment. Given the generality of this definition we see that this includes a whole range of different types of risk: market risk, credit or default risk, operational risk, legal risk, liquidity risk and model risk among other types. It is beyond the scope of this report to expand on this. The above can be combined into either micro risk or macro risk. The former is also called unsystematic or specific risk since this is the type of risk that is specific to the financial instrument (such as legal, operational, model, volatility and correlation risk). On the other hand, macro, or systematic, risk is common to all types of financial instruments. This includes market, currency and interest rate risk just to mention a few types. Although these kinds of risk are often not independent from each they are often modelled separately.

To value many financial commitments we need to determine the amount of risk of the underlying factors affecting them. In the case of many types of options on stocks this will involve modelling the riskiness of the underlying stock. By 'riskiness' here we mean a measure of change in value over some time-interval and it is usually measured by the volatility of the stock. In that the option is to be exercised some time in the future this will involve forecasting the volatility of the stock. One way of doing this will be considered in Chapter 3. First some background to financial modelling will be given.

1.1 Financial Modelling

We can trace the beginning of financial theory back to the (Bachelier, 1900) dissertation on speculation. This work marks both the origin of the continuous-time mathematics of stochastic problems and the continuous-time economics of option pricing. With respect to the latter Bachelier presented two different derivations of the Fourier partial differential equation as the equation for the probability distribution of what we now call Brownian motion. Modern financial theory began in the late 1950's. Before then the focus was mainly on the time value of money. The theory that was presented in (Markowitz, 1952) was ground-breaking. The topic of this work, mean - variance analysis, has since been investigated in depth and has become the standard way of approaching portfolio optimization by practitioners. The issue of the trade off between profit and risk is seminal but it is the latter that is most often modelled. One reason for this is that the risk factor dominates the expected returns. Another is that the variance of returns is highly predictable whereas the returns themselves are not .

Building on Markowitz, (Sharpe, 1964) and (Linter, 1965) introduced the *Capital Asset Pricing Model* (CAPM) which later became so key in measuring the performance of investments. The idea behind this model is that the components of a portfolio of assets are associated with a value of non-diversifiable risk. This value is denoted the β of the asset. During the same decade one of the major building blocks of economic theory - the efficient market theory - was introduced by (Samuelson, 1965) and (Fama, 1970). This hypothesis along with

empirical evidence presented by (Kendall, 1953) indicate that future asset returns are not 'forecastable'. The late 1960's and the 1970's saw an advance in the development of financial models involving dynamic asset allocation and choice under uncertainty. CAPM was extended to inter-temporal valuation. Under one period CAPM an asset risk measure is given by a single value, the β of the asset, but with this extension we are dealing with a multi-dimensional measure. For the kinds of models being developed during this period the partial and stochastic differential equations and integral equations governing these models were much more complex than had been worked with before in this field. (Ross, 1976) introduced the Arbitrage-Pricing Theory (APT) which can be viewed as a generalised competitor to CAPM.

The well known Black and Scholes (BS) model was introduced by (Black and Scholes, 1973) and (Merton, 1973) and revolutionised the financial research and practice of the time. The reason for this being that this model makes precise in a straightforward manner the way in which to price European options. For a given stock a dynamic trading strategy can be found which will replicate the returns of an option to that stock. Hence the fair price for the option is the value of the replicating strategy. The model is straightforward in the sense that there is just one input which is not directly observable: the volatility of the stock. Estimating the volatility then became a key issue in finance and many sophisticated models have since been developed to this end. (Cox *et al.*, 1979) remodelled the BS pricing derivation to a simple binomial stochastic process formulation.

The 1980's brought unification and extension of existing theories. In particular the BS model was generalised using the concept of stochastic integrals, (Harrison and Pliska, 1981), (Duffie and Huang, 1985) and (Duffie, 1986), a definition of which will follow in the preliminaries. (Cox *et al.*, 1985*b*) and (Cox *et al.*, 1985*a*) then extended the general pricing framework to allow for stochastic interest rates. A final work of particular interest is that of (Heath *et al.*, 1992). These authors showed how hedging could be carried out on derivative securities associated with bonds. The Heath-Jarrow-Morton model is now key in the world of option pricing.

This concludes a brief overview of the history of financial modelling up the beginning of the 1990's. A more detailed description of outstanding relevant contributions of more recent years will be left for the presentation of specific areas of financial modelling such as estimation techniques, dynamic volatility models and implied volatility in subsequent sections. The rest of this report is organised as follows. In Chapter 2 background theory to the applications of the subsequent chapters is presented. In Chapter 3 timescales in forecasting volatility will be considered using the Kalman Filter. In Chapter 4 filtering for high frequency data will be presented for linear state models. The conclusion includes suggestions for further work as well as a summary of the main results from the empirical and theoretical work.

Chapter 2

Modelling and statistical preliminaries

A series of modelling and statistical preliminaries will now be presented as an introduction and motivation for the work that will be carried out in subsequent chapters.

2.1 System Identification

Before we can model anything that is of interest to us we must first identify a system that is representative of the entities we are seeking to model in terms of their evolution. This has been done for a long time in some form or another. When the outcome of the entities we are seeking to model is completely random we say we have a stochastic system. In either case the first step to system identification is to choose an appropriate modelling structure. There are many issues that determine the choice of such a structure. In a deterministic setting discretising the differential equation(s) that describe the process(es) we are seeking to model will involve considering stability and convergence criteria. In a stochastic setting we will be concerned with incorporating all relevant information from observed data so as to predict as best as we can future phenomena conditioned on this information. If we are modelling several processes simultaneously we need to determine the relationship between these and to incorporate this into our modelling structure. Having determined such a structure we seek representative values of its parameters. In some situations we may be able to measure the corresponding physical phenomena and determine these parameters to arbitrary precision.

Often, however, due to physical uncertainty, noisy measurements and unobservability issues we may have to make do with estimating such quantities.

The process of system identification can in practice be broken down into several steps. The first step is to identify a model structure as described above. The next two steps involve calibrating and then validating our model. For this we need to choose a data set from which the values of the entities we are seeking to model can be extracted. In many situations part of this data set will be used to estimate the parameters of the model - i.e. calibration - and the rest will be used to back-test the estimated model - i.e. validation. Provided the results of the validation are satisfactory to some degree we can then claim that a system has been identified. For a more in depth exposure to a parametric approach to identification for which we have sought to summarize above see (Ljung, 1987).

2.2 Model Validation

Having estimated a model the validation of it can be carried out using two main approaches which we denote as *internal* and *external* validation.
The first consists of testing the model's performance as a stand alone problem. In this way we may be testing such things as the model's correct specification, the goodness of fit and the optimality of the parameter set.The second consists of comparing the model with competing and/or benchmark models. The values of the criteria for choosing between the models in consideration are likely to mean little on their own. In the context of comparison however these values can be very significant.

Firstly the internal approach will be presented. There are two main issues in testing for misspecification. Firstly the reliance of the model on the correct specification is a determining factor in its validation. Secondly the tests carried out should be powerful enough to check for any misspecification. Currently there is a large array of tests to choose from in any major area of statistical testing. Normality tests are often carried out as many model specifications assume this property. This assumption is mostly for practical purposes but this does not

necessarily imply it is unrealistic. The goodness of fit of a model is not such a clear cut matter. The main focus of this dissertation is linear state volatility models. In this context for example the popular R^2 measure is not necessarily an adequate indicator of the goodness of fit for volatility models, see (Andersen, 2000). Instead an internal measure such as the set of prediction errors may be more realistic and useful. Moreover it is important to bear in mind that fitting the data is not the main issue. A whole range of methods can be employed to get a model that fits the data but the real question is whether this model is relevant. On the other hand goodness of fit plays an important role in model validation and it would be unwise to disregard it completely. The covariance matrix of the parameter set (see appendix) gives a measure of the optimality of the parameter set in two contexts. In the context of the particular model that has been selected the diagonal entries of covariance matrix of the parameter set are the asymptotic standard errors on the parameters. However these values give us no assurance of the quality of the estimates over a set of (competing) models. In the context of the complexity of the model the off-diagonals are considered. In that these entries give the correlation between parameters these will show whether or not the model is over parameterised. Hence if the parameters are strongly correlated there is some redundancy in the parameters and we may want to simplify the model. Statistical tests have been developed to determine whether a subset of a larger model set, i.e. nested models, is adequate to describe data. The f-test for example gives a criterion for deciding between models in this context whereas the AIC criterion is used more generally to decide between competing models.

For the external validation approach we compare our model with a known model in terms of which gives better performance or fits the data the best. These known models are often called benchmark models. They may be known to perform reasonably well or are simply popular due to their tractability. In the context of this report these will be constant volatility and GARCH models as shall be seen. If we find our model outperforms a benchmark model we have some guarantee of the validity of our model. More will be said about this when we come to numerical results and the introduction of the relevant models.

2.3 Information and Probabilistic modelling

The type of information used determines the methodology that is used in modelling. It is of interest to consider the modelling performance from the forecasting angle as this is the main approach taken to modelling in this report. As is so often the case there is a tradeoff between parsimony and using all relevant information in a model for forecasting. Parsimony is not just for the sake of simplicity. Having many factors in a model can lead to collinearity, i.e. the factors are correlated which in turn means there is redundancy. On the other hand the model should take advantage of all the relevant information to get as much accuracy in the forecast. In quantitative models for forecasting the variables or factors that might influence the quantity that is being forecast constitute this 'information', which we denote the information set. These variables are often called the independent variables and the quantity being forecast is called the dependent variable. For autoregressive models, which shall be considered further on in this work, independent variables are previous values of the same time series. There are also forecasting methods that are not model-based. In finance the foremost example of this is implied volatility. The implied volatility of an financial instrument is an estimate of the volatility as implied by the prices of certain derivatives on this instrument. The idea behind using this kind of estimate is to let the market expectation alone determine the volatility of an instrument as opposed to making use of a model-based estimate[1] based on assumptions that may at times be hard to verify. From the point of view that the market aggregates all possible information, backing out information for estimation from derivative prices makes sense.

Now the scene has been set a more formal approach to forecasting in terms of conditioning will be presented. This will be initiated with a series of definitions.

Let us denote a probability space by $(\Omega, \mathcal{F}, P)$ where Ω is the outcome space, P is the probability measure and $\mathcal{F}$ a σ-field or σ-algebra of subsets of Ω, i.e. the set of all subsets of Ω

[1] although implied volatility estimates may be combined with a modelling structure to infer parameter values as shall be seen.

A probability measure P on $(\Omega, \mathcal{F})$ is a function mapping $\mathcal{F}$ onto $(0, 1)$ such that

- $P(\Omega) = 1$

- if $A \in \mathcal{F}$ then $P(A^c) = 1 - P(A)$

- if $A_1, A_2, ..., A_n, ... \in \mathcal{F}$ then $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.

A σ-field $\mathcal{F}$ is a collection of subsets of Ω such that

- $\Omega \in \mathcal{F}$

- if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$

- if $A_1, A_2, ..., A_n, ... \in \mathcal{F}$ then $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Let X be a random variable on $(\Omega, \mathcal{F})$ be defined as

$$X : \begin{cases} \Omega \longrightarrow \mathbb{R} \\ \omega \longrightarrow X(\omega) \end{cases}$$

so that a random variable is a function mapping an outcome to a real number.

The $\sigma-$field, σ, generated by X is defined to be the collection of all sets of the form $\{\omega \in \Omega : X(\omega) \in A\}$ where A is a subset of $\mathbb{R}$. Let $\mathcal{G}$ be a a sub-σ-algebra of $\mathcal{F}$. We say that X is $\mathcal{G}$-measurable if every set in $\sigma(X)$ is also in $\mathcal{G}$. We can also say that X is adapted to $\mathcal{G}$. The intuition behind the above is that the content of the $\sigma-$field is exactly the information obtained by observing X.

The unconditional expectation of X is defined to be

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) dP(\omega) \tag{2.1}$$

The above is equivalent to the mean value of the random variable over the entire outcome space. Unconditional refers to there not being any conditions to provide information on the set of outcomes. Once an event has been realised the outcome space is reduced to a subset of Ω. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. The conditional expectation of $\mathbb{E}(X \mid \mathcal{G})$ is defined to be any random variable that satisfies

1. $Y = \mathbb{E}(X \mid \mathcal{G})$ is $\mathcal{G}$-measurable
2. For every set $A \in \mathcal{G}$, we have that $\mathbb{E}(X \mid \mathcal{G}) = \dfrac{1}{P(A)} \displaystyle\int_A X(\omega) dP(\omega)$ (2.2)

In practice it is of interest to consider not just random variables but random, or stochastic processes, i.e. a sequences of random variable $\{X_t : t \in T\}$ where if the process is of discrete-time $T = 0, 1, 2, ...$ and if it is of continuous-time $T = [0, \infty)$.

Within this context the concept of observing the random process can be introduced. A filtration, or information flow, $\mathcal{F}_t$, $t \geq 0$, is defined to be the sequence of σ-fields such that

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset ... \subset \mathcal{F}_t \tag{2.3}$$

A random process is adapted to the filtration $\mathcal{F}_t$ if the process X_t is $\mathcal{F}_t$-measurable. In other words X_t does not carry more information than $\mathcal{F}_t$. A final point is that the stochastic process X_t is always adapted to the natural filtration generated by X_t:

$$\mathcal{F}_t = \sigma(X_s, s \leq t) \tag{2.4}$$

This is then in essence all the past and present information associated with the stochastic process.

We are now in a position to introduce a stochastic process called Brownian motion.

2.4 Brownian motion and Stochastic Integration

Brownian motion is central to probability theory and has far-reaching applications. It was named after the biologist Robert Brown at the beginning of the 19th century and was developed further at the beginning of the 20th century by Louis Bachelier, Albert Einstein and

Norbert Wiener.

Standard Brownian motion is a continuous-time stochastic process $B(.)$ such that

- [B(0)]=0.
- For any times $0 \leq t_1 < t_2 < ... < t_k$ the changes $[B(t_2)-B(t_1), B(t_3)-B(t_2), ..., B(t_k)-B(t_{k-1})]$ are independent Gaussian with $[B(s) - B(t)] \sim N(0, s-t)$.
- For any given realisation, $B(t)$ is continuous in t with probability 1.

Brownian motion is a specific type of the more general Wiener-Lévy process which also allows for non-normal increments and discontinuous trajectories, i.e. the process can jump randomly. More precisely it is composed of both a Gaussian component and a pure jump component. The Wiener term is most often associated with the Gaussian component and the Lévy term with the jump component[2]. Since Brownian motion has a Gaussian component but no jump component it can be simply denoted a Wiener process. The more general Wiener-Lévy process will be considered in more detail in the section on jump-diffusion models. It may be of interest to consider special processes derived from Brownian motion. One of these is called a Brownian Bridge which is a process within a given interval that has a fixed end point at zero but evolves as a Brownian motion in between. Stochastic interpolation using a shifted Brownian Bridge involves a skewed Brownian Bridge since the interval start and end points can take any values and need not coincide.

A key feature of Brownian motion is that it is no where differentiable as the trajectories are not of bounded variation. Standard calculus cannot therefore be applied being replaced by stochastic calculus. This theory pioneered by Itô is vast and is a major building block in financial theory. We will limit the overview of this theory to the introduction of the Itô formula and the Itô Stochastic Integral.

The theory of stochastic processes begins at formulating the derivation of functions of a Wiener process. Let $X_t = f(W_t)$ for some given f and the Wiener process W_t. The usual

[2]there seems to be some ambiguity in nomenclature but the general consensus appears to be in this fashion.

chain rule does not apply for this equation but if f is sufficiently smooth Taylor's theorem can be applied to give

$$X_{t+\delta t} - X_t = f'(W_t)(\delta W_t) + \frac{1}{2}f''(W_t)(\delta W_t)^2 + ... \tag{2.5}$$

where $\delta W_t = W_{t+\delta t} - W_t$. Note that $(\delta W_t)^2$ can be approximated by its mean δt and higher order terms are insignificant as $\delta t \to 0$. The Itô formula is the limit of (2.5) with higher order terms ignored,

$$dX_t = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt \tag{2.6}$$

which is a shorthand form for (the integrated form)

$$X_t - X_0 = \int_0^t f'(W_s)dW_s + \frac{1}{2}\int_0^t f''(W_s)ds \tag{2.7}$$

(2.6) can be generalised for time as an independent variable in the function $Y = f(t, W_t)$. This formula then gives

$$dY = f_w(W)dW + [f_t(t, W) + \frac{1}{2}f_{ww}(t, W)]dt \tag{2.8}$$

The Itô formula above is for a Wiener-Lévy process without a jump component. For this formula for processes with a jump component cf. (Cont and Tankov, 2004, p. 276).

Since $f'(W)$ is stochastic and the integrator is the limiting difference of a stochastic process that although continuous is not differentiable the first term on the right hand side of (2.7) must be treated differently from the normal Riemann integral. This integral is known as the Itô Stochastic Integral and will be defined in what follows.

For some finite time T let $(X_t)_{0\leq t<T}$ be a stochastic process adapted to $(\mathcal{F}_t)_{0\leq t<T}$ the natural filtration of the Brownian motion such that

$$\mathbb{E}\int_0^T (X_t)^2 dt < +\infty \tag{2.9}$$

The stochastic integral of (X_t) w.r.t. the Brownian motion W_t is defined as a limit in the mean-square sense

$$\int_0^t X_s dW_s = \lim_{n\to\infty} \sum_{i=1}^{n} X_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}) \tag{2.10}$$

A simple statement of the definition above begs explanation. The background theory for the construction of this integral is not so straightforward. For a rigorous treatment of the steps leading up to this definition readers are referred to (Mikosch, 1998, section 2.2.).

2.5 Maximum Likelihood Estimation

Maximum Likelihood provides an estimator that maximises the probability of an observed event. It was introduced by (Fisher, 1912). Let X_n be some observed scalar-valued i.i.d. Gaussian random process. At each time n the probability density of X_n is given as

$$p(X_n) = \frac{1}{\sqrt{2\pi\mathbb{E}(X_n - \mathbb{E}(X_n))^2}} \exp\left(-\frac{(X_n - \mathbb{E}(X_n))^2}{2\mathbb{E}(X_n - \mathbb{E}(X_n))^2}\right) \tag{2.11}$$

The joint probability of the set of T observations occurring in the order in which they are observed is

$$P(X_n) = \prod_{n=1}^{T}\left[\frac{1}{\sqrt{2\pi\mathbb{E}(X_n - \mathbb{E}(X_n))^2}} \exp\left(-\frac{(X_n - \mathbb{E}(X_n))^2}{2\mathbb{E}(X_n - \mathbb{E}(X_n))^2}\right)\right] \tag{2.12}$$

Maximising the joint probability (2.12) is denoted maximising the likelihood of observations. For this reason the joint probability function P is often substituted by L to represent likelihood. Often $\mathbb{E}(X_n)$ and/or $\mathbb{E}(X_n - \mathbb{E}(X_n))^2$ are not known but can be estimated conditional on parameterised past information. Indeed even the process X_n can be dependent on a set of parameters. A parameter set $\underline{\theta}$ is sought that maximises the likelihood at each time step n based on past information $\mathcal{F}_{n-1}$,

$$\max\, L(X_n \mid \underline{\theta}, \mathcal{F}_{n-1}) = \max\left\{\prod_{n=1}^{T}\left[\frac{1}{\sqrt{2\pi\mathbb{E}(X_n - \mathbb{E}(X_n))^2}} \exp\left(-\frac{(X_n - \mathbb{E}(X_n))^2}{2\mathbb{E}(X_n - \mathbb{E}(X_n))^2}\right)\right]\right\} \tag{2.13}$$

Maximising L is equivalent, to all extent and purposes, to maximising the log of L. This transformation is carried out purely for computational ease. The transformed function of (2.13) becomes

$$L_{\log}(X_n \mid \underline{\theta}, \mathcal{F}_{n-1}) = -\sum_{n=1}^{T} \log(\mathbb{E}(X_n - \mathbb{E}(X_n))^2 - \sum_{n=1}^{T} \left(-\frac{(X_n - \mathbb{E}(X_n))^2}{\mathbb{E}(X_n - \mathbb{E}(X_n))^2} \right) \tag{2.14}$$

when constant terms are ignored. The likelihood of a vector-valued i.i.d. Gaussian process can be defined in a similar way.

In certain financial applications it is the asset return or variance process that constitute the observable process. More will be said about this in the section on GARCH and state-space models.

2.6 Dynamic volatility models

In this section dynamic volatility models will be introduced. Although these are arguably the foremost way of modelling volatility we should note that static volatility models are still widely used, especially in industry and when modelling the correlation in volatility. Modelling volatility dynamically has played a central part in finance since a phenomenon was observed (Mandelbrot, 1963) in the variances of returns called clustering, i.e. that these variances cluster around some level for a certain period of time before returning to a mean level. This clustering phenomenon implies serial correlation in the return variance which in turn means that they can be predicted to some degree. Many methods have since been proposed for modelling the above phenomenon. These fall roughly into three categories[3]: GARCH-type[4] models, "pure" Stochastic Volatility (SV) models, often denoted Stochastic Variance[5] models and Jump-Diffusion models. SV models assume the volatility follows an Itô process satisfying a SDE driven by Brownian motion or some other stocastic process. In this way the dynamics of the variance is given by a function of "past" variance plus a noise term. Using SDE's is sensible from the point of view that volatility is known to be random. However, as there is already randomness in the stock price process, having an extra source of randomness means that the market will no longer be complete. For an overview of the

[3]although they can be unified, see for example, (Albanese and Kuznetsov, 2003).

[4]an acronym for Generalized Autoregressive Conditional Heteroscedasticity with Conditional Heteroscedasticity referring to the variance of returns being serially correlated over time.

[5]the variance of returns is a proxy for the volatility of returns which is unobservable. As most volatility models use this proxy we may at times simply refer to these as Stochastic Variance models.

implications of this fact see (Bjork, 2004)). ARCH and GARCH-type models on the other hand imply deterministic, time-varying volatility and were first introduced by (Engle, 1982) and (Bollerslev, 1986). Perhaps partly due to their simplicity and flexibility they have since become very popular principally in industry. Jump-diffusion models have been around for some time for a variety of purposes. In general these models concern the dynamics of the asset price/return. Modelling volatility with random jumps is a special case of jump-diffusion models that tie in jumps in the volatility with jumps in the asset price. Although they in theory reproduce the statistical features often present in the data these models are often hard to implement.

2.6.1 Stochastic Volatility models

A short overview of stochastic volatility models follows. The basic stochastic volatility model is a two state discrete time model describing the dynamics of the asset price return, r_t, and the volatility, σ_t, of the form

$$\begin{aligned} r_t &= \kappa\sigma_t\epsilon_t \\ \ln\sigma_{t+1}^2 &= \phi\ln\sigma_t^2 + \eta_t. \end{aligned} \tag{2.15}$$

where ϵ_t and η_t have mean zero and variances equal to one and β^2 respectively. Here it is of interest to generalise the above model to any model with a time-varying stochastic representation for the asset price return volatility, or some function of the volatility. Moreover the return dynamics will be allowed to take forms other than (2.15) but our interest lies in the volatility dynamics so the former will be put to one side in this brief presentation. This goes against the bivariate form for stochastic volatility models as often given in the literature but allows for a more focused exposure. We will begin with the general continuous time SV model which is given by

$$dY(t) = \alpha(Y,t)dt + \beta(Y,t)dW(t) \tag{2.16}$$

where α and β are given functions, usually continuous in (Y,t) and $W(t)$ is a Wiener process, i.e. $dW(t)$ is white noise. The volatility, σ, is some positive function, f say, of Y. As special cases of the general model above we have: the CIR or Feller model, (Cox *et al.*, 1985*b*), the

lognormal model, and the Ornstein Ulenbeck (OU) model to cite the most common ones. The Feller model is given by

$$dY(t) = \alpha(\kappa - Y(t))dt + \beta\sqrt{Y(t)}dW(t) \tag{2.17}$$

When the Wiener process above is correlated with the underlying stock price's Wiener process and $f(Y) = \sqrt{Y}$ we have the Heston model, (Heston, 1993). The lognormal model is given by

$$dY(t) = C_1 Y(t)dt + C_2 Y(t)dW(t) \tag{2.18}$$

The well known Hull-White model, (Hull and White, 1987), is (2.18) with $f(Y) = \sqrt{Y}$.
Finally the OU model is given by

$$dY(t) = \alpha(\kappa - Y(t))dt + \beta dW(t) \tag{2.19}$$

(Scott, 1987) works with the above model for $f(Y) = e^y$.
The OU and Feller models are *mean reverting* ones, i.e. the volatility frequently leaves an average level but then reverts back to it at a certain rate. One of the SV model that will be considered in this report is of the OU-type. Whatever the model structure, the main issue in SV modelling is how to estimate the model parameters given that volatility is unobservable. A common way of doing so is to model the volatility as a hidden state. This approach often involves a set of linear space-space equations, with the hidden state being estimated using the output of the Kalman filter and the parameters by a likelihood function. Another approach is to assume the volatility takes values according to the state of a hidden Markov chain, cf. (Elliott *et al.*, 2003). In this setting there are a finite number of states that are estimated using a noisy observation process.

2.6.2 GARCH-type models

Let us consider the residuals, $\varepsilon(k)$, obtained from subtracting the mean return from the actual returns $r(k)$, and the variance, $\sigma^2(k)$ of these residuals[6]. A ARCH/GARCH model stipulates that these residuals are conditionally normal $\varepsilon(k) \mid \mathcal{F}(k-1) \sim N(0, \sigma^2(k))$. In a GARCH-type model the variance terms are given in terms of past residuals and past variance terms

[6] in certain applications the residuals come from a regression of the returns on several explanatory variables.

$$\begin{aligned}\sigma^2(k) &= \gamma + \beta_1\sigma^2(k-1) + \beta_2\sigma^2(k-2) + ... + \beta_p\sigma^2(k-p) + \\ &\quad \alpha_1\varepsilon(k-1)^2 + \alpha_2\varepsilon(k-2)^2 + ... + \alpha_q\varepsilon(k-p)^2 \end{aligned} \tag{2.20}$$

(2.20) is known as a GARCH(p,q) model.

Considering these residuals as the observable process as given in the section on Maximum Likelihood it is not difficult to verify that the log-likelihood (2.14) for a GARCH model is

$$L_{\log}(\varepsilon(k) \mid \mathcal{F}_{k-1}, \underline{\theta}) = -\sum_{k=1}^{T} \log(\sigma^2(k)) - \sum_{k=1}^{T} \frac{(\varepsilon(k))^2}{\sigma^2(k)} \tag{2.21}$$

Considering (2.20) for p=q=1, and with a slight simplification of notation, we have,

$$\sigma^2(k) = \gamma + \beta\sigma^2(k-1) + \alpha\varepsilon(k-1)^2 \tag{2.22}$$

Thus we find values of γ, α and β that maximise (2.21) for $k = 1, ..., T$. Since the log-likelihood function has a closed form, estimation and calibration via maximum likelihood is straightforward. GARCH(1,1) with $\gamma = 0$ is known as the exponential weighted moving average (EWMA) model. This model in a similar way to GARCH-type models is very popular in industry. Common variance estimates are given as weighted averages of past squared returns. To keep these estimates relevant the weights will decrease as we move back through time. It turns out that a exponential decrease leads to the parsimonious EWMA model formulation.

Since they were proposed in the 80's, ARCH and GARCH-type models have since been built upon to incorporate modelling features that better describe, as empirical evidence would suggest, the properties of the entities that are sought to be modelled. Two of the foremost of these features are the 'leverage effect' and excess kurtosis. It has been observed that negative returns tend to increase the volatility more than positive ones of the same magnitude. This form of asymmetry is denoted the leverage effect. Let us note that the standard GARCH model does not allow for this feature. Excess kurtosis means that returns distributions tend to have 'fatter' tails than the Gaussian distribution. Examples of models that allow for some of these features are (exponential) EGARCH, (Nelson, 1991), and t-GARCH, (Bollerslev, 1987). The former allows for negative parameters while guaranteeing that the

volatility remains positive that and can also incorporate the leverage effect. The latter uses the student t-distribution in the calibration of the model parameters. The literature appears to indicate however that in many applications a standard parsimonious representation such as GARCH(1,1) suffices.

Clearly correlations between stocks is also an important issue to be considered. Multivariate GARCH models have indeed been considered for examples see (Engle and Kroner, 1995) and (Bollerslev *et al.*, 1988). Finally it is worth mentioning the existence of GARCH models for option valuation, see for example (Duan, 1995) and (Heston and Nandi, 2000).

Finally although GARCH models appear to be quite distinct from SV models they have been shown to be limiting approximations of these. There are certain SV models where the relation between these and GARCH models has been demonstrated. This is the case for Stochastic Autoregressive Volatility models, see (Fleming, 2003) and (Meddahi and Renault, 1997), and for Heston's square root model, see (Heston and Nandi, 2000). The pivotal work of (Nelson, 1990) provided the framework for this by interpreting the continuous time limit of a discrete time GARCH processes.

2.6.3 Jump-diffusion models

A short overview of jump-diffusion[7] models follows[8]. For a comprehensive survey readers are referred to (Cont and Tankov, 2004). A general jump-diffusion model for the asset price S is given by

$$dS(t) = \alpha(S,t)dt + \beta(S,t)dW(t) + dZ(t) \tag{2.23}$$

where α and β are given functions, usually continuous in (S,t) and the second term is a diffusion. $Z(t)$ is a specific type of Wiener-Lévy process, namely it is a process with independent and stationary increments with jumps, no Gaussian component and no drift. A process with

[7] although we will restrict our attention to the jump part of these models to remain in a general framework this nomenclature is preferred.

[8] although we are within the framework of dynamic volatility models jump-diffusion models will be introduced more generally with volatility models with random jumps presented as a special case.

jumps is defined to be one where the instantaneous variance can be singularly large at a finite number of points. (Bates, 1996) extended the Heston model to include jumps. In (Barndorff-Nielsen and Shephard, 2001) the Lévy process was assumed to follow a generalised inverse Gaussian law whilst in (Merton, 1976) the timing of the jumps was assumed to follow a Poisson distribution. Jumps in the stock price are related to discrepancies in business time and calendar time which has led to work on *time-changed* Lévy processes, (Geman *et al.*, 2001) and (Carr and Wu, 2004). Time-changed processes are ones for which the underlying time of a time-dependent process is allowed to have random jumps. The motivation behind this approach is to make a transformation to a world which has nice properties for valuation and/or which allows empirically observed properties to be characterised. In the context here business time will follow a stochastic process with jumps. In (Carr *et al.*, 2003) the rate of the jumps and time change is tied in with the volatility which produces the well documented leverage effect. Another approach for incorporating leverage, introduced by (Barndorff-Nielsen and Shephard, 2001), is to use the same jumps in the volatility and in the price. As well as leverage another property of non-normality, excess kurtosis, can arise from a substantial jump component. It is important to verify how well Jump-diffusion models reproduce the profiles of empirically observed implied volatility surfaces and smiles, cf. (Bakshi *et al.*, 1997) and (Skiadopoulos *et al.*, 2000). For valuing contingent claims allowing for jumps in the volatility see for example (Naik, 1993).

It should be pointed out that models with jumps go beyond stock price models for typically daily frequencies. More generally these have been introduced to reproduce/model the statistical features often found in financial time-series data right across the ball. Of particular interest here is their use in modelling ultra high frequency data. Since tick-by-tick prices remain at some level until a transaction causes these to jump to a new level the dynamics of ultra high frequency data follow a non-Markovian process with jumps, there is no diffusion component. Theory from fractional Brownian motion provides opens up work in this area that was pioneered by the Olsen group, cf. (Muller *et al.*, 1993). For further research in this direction see (Scalas *et al.*, 2000), (Woerner, 2003) and (Woerner, 2005).

2.7 State-space formulation and the Kalman filter

In many dynamical systems the entity that is sought to be modelled is not directly observable, i.e. this entity, known as the (hidden) state is measured in noise. However if the noise is assumed to be known in distribution this state can often be estimated in a particularly efficient way. Such a estimation procedure delivers pointwise estimates. If less is known about the measurement noise one may have to make do with set membership estimation. A special case of the former situation is when the unobservable variable is a linear function of observable variables. There is a particular class of SDE's that lend themselves to an unobservable, yet tractable and efficient, model formulation - the OU-type process (2.19). For simplicity a first order Euler discretisation of this SDE will be considered,

$$x_{n+1} = ax_n + b + w_{n+1}, \tag{2.24}$$

where $a = 1-\alpha\Delta$, $b = \alpha\kappa\Delta$, and $w_{n+1} = \beta(W(t+\Delta)-W(t))$, where Δ represents a typically small time interval. Consider the situation where the state variable x_n is unobservable yet there is a variable, y_n say, that is observable and is an affine function of x_n of the form,

$$y_n = cx_n + f_n, \tag{2.25}$$

When $f_n = d + u_n$. where u_n is typically white noise, a 'State-Space' system can be set up of the form,

$$\begin{aligned} x_{n+1} &= ax_n + b + w_{n+1} \\ y_n &= cx_n + d + u_n, \end{aligned} \tag{2.26}$$

where $\mathbb{E}(w_n) = \mathbb{E}(u_n) = 0$, $\mathbb{E}(w_n^2) = q^2$ and $\mathbb{E}(u_n^2) = r^2$. $\mathbb{E}(u_n)$ and $\mathbb{E}(w_n)$ are known as error terms. The first and second equations of the above State-Space system are known as the transition equation and the measurement equation respectively. It should be noted that (2.25) as a model in its own right is called a 'General Linear Model'.

It is of interest to generalise (2.26) to multiple states and multiple observable variables,

$$\begin{aligned} \underline{x}_{n+1} &= A\underline{x}_n + \underline{b} + Q\underline{\varepsilon}_{n+1} \\ \underline{y}_n &= C\underline{x}_n + \underline{d} + R\underline{\eta}_n \end{aligned} \tag{2.27}$$

where $\underline{x}_n, \underline{b}$ and $\underline{\varepsilon}_n$ are vectors of length N and $\underline{y}_n, \underline{d}$ and $\underline{\eta}_n$ are vectors of length M. C is a $M \times N$ matrix, R is a $M \times M$ matrix and A, B and Q are $N \times N$ matrices. These parameters could be specified to be time-dependent. This would involve introducing evolution equations for the unknown parameters as extra states. The main issue that limits this approach is the curse of high dimensionality. For this reason only time-invariant systems are considered here although further in this work one of the parameters will be introduced as time dependent.

The above state-space formulation became an increasingly popular modelling procedure since (Kalman, 1960) and (Kalman and Bucy, 1961) developed what is now known as the Kalman-Bucy Filter, or simply the Kalman Filter (KF). Under a state space specification such as the one above the KF procedure is a predictor-corrector scheme in which the covariance of estimation error is minimised. In this way the state estimates that are delivered are optimal[9] among all other one-step predictor schemes if the disturbances are Gaussian. If this is not the case and the model has been misspecified the filter still delivers estimates that are optimal in regards to all other linear predictors.

Let us denote the KF conditional one-step-ahead estimate of the hidden state, $\hat{x}_{n+1|n}$, and the covariance of this estimate, $P_{n+1|n}$. The innovation, or $\underline{v}_n$, is defined as the difference between the observation at time n and an affine function of the previous step's state prediction, $\underline{y}_n - C\underline{\hat{x}}_{n|n-1} - \underline{d}$. The correction is based on the innovation itself, its variance, F_n and the state estimate variance. Related to this correction is the Kalman gain, K_n, defined below. It is usual to combine the prediction and correction equations into one set of equations. The Kalman Filter equations as given in (Harvey, 1989, p. 100-106), and, with slightly different

[9] in the MSE sense.

notation, are reproduced here for convenience

$$\begin{aligned}
\underline{v}_n &= \underline{y}_n - C\underline{\hat{x}}_{n|n-1} - \underline{d} \\
F_n &= CP_{n|n-1}C^T + RR^T \\
K_n &= AP_{n|n-1}C^T F_n^{-1} \\
\underline{\hat{x}}_{n+1|n} &= A\underline{\hat{x}}_{n|n-1} + \underline{b} + K_n\underline{v}_n \\
P_{n+1|n} &= A(P_{n|n-1} - P_{n|n-1}C^T F_n^{-1} CP_{n|n-1})A^T + QQ^T
\end{aligned} \tag{2.28}$$

The KF can be considered a weighted recursive least squares problem although for time-invariant systems such as the one considered here there is convergence to equal weighting. The KF algorithm is recursive as the state is updated for every measurement based on (an affine function of) the previous state. In many cases the system will be stationary, i.e. the mean and covariance of the state do not depend on time. For time-invariant systems, as the one above, when the roots of A are inside the unit circle this will be the case.

To initialise the KF estimates for the mean and variance of the initial state, $\underline{\hat{x}}_0$ and P_0 respectively, are needed. If the model is stationary these values can consistently be set equal to the expected stationary state and the variance of the stationary state respectively, cf. (Anderson and Moore, 1979, p. 64-70). As we see from the following equations

$$\underline{\hat{x}}_0 = \underline{b}/(I_N - A), P_0 = QQ^T/(I_N - AA^T) \tag{2.29}$$

these moments are just given in terms of the unknown parameters. If the model is not stationary the model must be initialised in some other way, often using a diffuse of proper prior for the covariance. A diffuse prior in some special cases takes the form $P_0 = kI_N$ for some large k. In general the use of a diffuse prior calls for extending the KF and correcting the likelihood function. A proper prior generally only applies to observable models in which the first p set of observations is used for constructing priors.

The setup above assures optimality of the state estimates for given parameters. However these may not be known. The optimal parameter set is defined to be the one which minimises the difference between predicted values of y_n and the actual observations, i.e. the prediction

error. The minimisation is carried out under a certain weighted average procedure derived from better known as Maximum Likelihood.

In the case of the univariate State-Space model (2.26) the parameters can be estimated as described above in a straightforward manner if we assume the observed variables are Gaussian. Thus (2.14) can be written as

$$L_{\log}(y_n \mid \underline{\theta}, \mathcal{F}_{n-1}) = -\sum_{n=1}^{T} \log(\mathbb{E}(y_n - \mathbb{E}(y_n))^2 - \sum_{n=1}^{T} \left(-\frac{(y_n - \mathbb{E}(y_n))^2}{\mathbb{E}(y_n - \mathbb{E}(y_n))^2} \right) \tag{2.30}$$

We see that the above is in prediction error form but it is of interest to view it in terms of the KF output. Thus when substituting $\mathbb{E}(y_n)$ by $C\underline{\hat{x}}_{n|n-1} + \underline{d}$ along with some other substitutions (2.30) becomes

$$L_{\log}(y_n \mid \underline{\theta}, \mathcal{F}_{n-1}) = -\sum_{n=1}^{T} \log F_n - \sum_{n=1}^{T} v_n^2 F_n^{-1} \tag{2.31}$$

In the context of maximising the log-likelihood we see from (2.31) that the innovations with a smaller variance are given more weight in the optimisation. The parameter vector which maximises the likelihood of the observations is called the maximum likelihood estimate. It is worth pointing out that if the state-space is multivariate the expression (2.31) suffers minor modifications.

A Gaussian Filter being applied to a model which is not necessarily Gaussian implies that the state estimates may be biased and thus the estimation will be suboptimal. In *Quasi-Maximum Likelihood Estimation* (QMLE), (Bollerslev and Wooldridge, 1992), these biases are ignored in the actual estimation. These are however accounted for when calculating standard errors on the estimates. Details of this for a univariate State-Space model estimated from the output of the Kalman Filter are given in Appendix A.

2.8 Multi-step prediction

Multistep prediction can also be considered using the parameters of the one-step-ahead state space equations. The Kalman filter equations, (2.28), are a combination of prediction and updating equations. The derivation of the multi-step prediction equations comes from repeatedly applying the one-step prediction equations. Updating equations are not considered

as there is no way to update the state or error estimates. For m, say, step-ahead prediction the state estimate and variance equations for the univariate model with serially uncorrelated error terms are

$$\begin{aligned} \hat{x}_{n+m|n} &= a^m \hat{x}_n + \sum_{j=0}^{m-1} a^j b \\ P_{n+m|n} &= a^{2m} P_n + \sum_{j=0}^{m-1} a^{2j} q^2 \end{aligned} \tag{2.32}$$

or in terms of the previous step predictions

$$\begin{aligned} \hat{x}_{n+m|n} &= a\hat{x}_{n+m-1|n} + b \\ P_{n+m|n} &= a^2 P_{n+m-1|n} + q^2 \\ F_n^* &= c^2 P_{n|n-m} + r^2, \end{aligned} \tag{2.33}$$

The total[10] variance of state estimates would then be the sum of $P_{n+i|n}, i = 1, ..., l$. We also have an expression for the innovations variance based on m-step ahead prediction: $F_n^* = c^2 P_{n|n-m} + r^2$. It is important to point out that the innovations variance is a model-based estimate that is used in the Kalman Filter to contribute towards 'on-line' best linear state prediction. Although it is principally the observed actual deviation in the innovations that will determine the quality of a multi-step forecast a model-based estimate as the expression for the innovations variance above may still be indicative of how good a forecast is.

As pointed out in (Johnston and Harrison, 1986), some care is needed if dealing with flow variables for which predictions involve the cumulative effect of some variable several steps ahead. Clearly this is the case in forecasting volatility, the volatility of a future period up to lead time m is given by the sum of the future sub-periods. For flow variables (2.33) does not apply.

Assume the error terms in the univariate model (2.26) are serially uncorrelated and uncorrelated with each other. Under these assumption we have derived the following two propositions.

[10] in the sense of accumulating the variance estimates from one- and two-step prediction

Proposition 1. *The total variance of the state estimates, $\bar{P}_n$ for the cumulative state prediction, $\hat{\bar{x}}$, for m steps is,*

$$\bar{P}_n = \mathbb{E}(\bar{x} - \hat{\bar{x}})^2 = \left(\sum_{j=0}^{m-1} a^j\right)^2 P_{n+1|n} + \sum_{i=0}^{m-2}\left(\sum_{j=0}^{i} a^j\right)^2 q^2, \tag{2.34}$$

where,

$$\bar{x} = \sum_{j=1}^{m} x_{n+j}. \tag{2.35}$$

Proof: See appendix.

Proposition 2. *The total innovations variance, $\bar{F}_n$ for the cumulative state prediction, $\hat{\bar{x}}$, for m steps is,*

$$\bar{F}_n = \mathbb{E}(\bar{y} - c\hat{\bar{x}} - md)^2 = \left(\sum_{j=0}^{m-1} a^j\right)^2 c^2 P_{n-m+1|n-m} + \sum_{i=0}^{m-2}\left(\sum_{j=0}^{i} a^j\right)^2 c^2 q^2 + mr^2, \tag{2.36}$$

where

$$\bar{y} = \sum_{j=0}^{m-1} y_{n-j} \tag{2.37}$$

and

$$\hat{\bar{x}} = \sum_{j=1}^{m} \hat{x}_{n-m+j|n-m}. \tag{2.38}$$

Proof: See appendix.

Finally it is worth pointing out that there is a simple multi-step expression for the GARCH(1,1) model:

$$\sigma^2(n+k) = \beta^k \sigma^2(n) + \gamma \sum_{i=0}^{k-1} \beta^i + \alpha \sum_{i=1}^{k} \beta^{i-1} \varepsilon^2(n+k-i). \tag{2.39}$$

Now that the scene has been set some applications of the theory that has been presented will be given.

Chapter 3

A Stochastic Volatility Model from high-frequency stock price data: calibration, forecasting and the choice of timescales

3.1 Introduction

For a decade or so high frequency financial data has been readily available and for this reason modelling volatility in particular has reached an altogether higher level. Volatility estimated using a historic sample of high-frequency data with equal weights is called *realised volatility* (RV). For an empirical study on the properties of RV as an estimator see (Andersen *et al.*, 2001). It has been shown that if the asset path is sampled sufficiently frequently integrated volatility, a natural measure for the variation over a given interval, can in theory be estimated from RV with arbitrary precision, see (Andersen and Bollerslev, 1998) and (Nelson, 1992). For certain very liquid assets data for every minute, or even more frequent than this, is available. RV as a volatility proxy plays a pivotal role in volatility estimation and SV models have been formulated based on the properties of RV. The building blocks for the use of realised volatility as a proxy for the variance in returns stems from the theory of quadratic variation introduced in this context by (Andersen and Bollerslev, 1998) and (Barndorff-Nielsen and

Shephard, 2001). More generally we have the theory of power variation that has also been developed by the latter authors as well as by (Woerner, 2003) and (Woerner, 2005) and is a powerful tool.

In this chapter we will follow to some extent the work of (Barndorff-Nielsen and Shephard, 2002) using RV in estimating SV models. The SV model will be estimated and calibrated based on a Gaussian OU process for instantaneous volatility. This approach differs from (Barndorff-Nielsen and Shephard, 2002) since these authors consider combining non-Gaussian Ornstein-Uhlenbeck (OU) type processes where as we consider a single Gaussian OU process. The overall objective is to complement the body of empirical research on forecasting horizons and sampling procedures. In particular we wish to provide some empirical evidence as to the choice of sampling interval for short and long run forecasts. Work that most resembles this contribution is to be found in (Andersen *et al.*, 1999). These authors study the choice of sampling frequency for different forecast horizons for exchange-rate data. They work with a GARCH model and employ the diffusion approximations of (Nelson, 1990). In contrast we work with a fixed sampling frequency and vary the sampling interval[1] according to the forecast horizon. Furthermore we work with an unobservable SV model. We take advantage of the fact that the model specification we employ lends itself to multi-period prediction. As such this is carried out for shorter sampling intervals and compared with one-step-ahead prediction for longer ones but for the same forecast horizon. Although this study is fairly small scale and the prediction comparisons are only ex-ante, this paper can be considered as motivational work in the increasingly important field of empirical forecasting evaluation.

The rest of the chapter is organised as follows. In Section 2 the model used for forecasting volatility will be presented. In Section 3 the calibration procedure for the state and parameter estimates of this model will be described. In Section 4 the numerical results of the model validation and choice of time scales will be given.

[1]In GARCH models the notion of a sampling interval does not exist as it coincides with the sampling frequency. A de facto exception to this is when the variance at a relatively low frequency is implied from the parameters of a model estimated at a higher frequency, see (Drost and Nijman, 1993)

3.2 Linear State-Space formulation

In this section the model for forecasting realised volatility will be presented.

First of all consider the following continuous-time model for log-stock price returns,

$$dS^l(t) = \mu dt + \sigma(t) dB(t), \tag{3.1}$$

where $S^l(t)$ is the log-stock price, for some $0 < \delta \ll 1$, $dS^l(t) = S^l(t+\delta) - S^l(t)$ is the log stock price return and μ is the drift. $dB(t)$ is the differential of Brownian motion and is $N(0, dt)$-distributed. Furthermore it is uncorrelated with $\sigma^2(t)$ [2]. When σ is constant $dS^l(t) \sim N(\mu dt, \sigma^2(t)dt)$, which is known as the lognormal model for the stock price continuously compounded returns, $S(t+\delta)/S(t)$ where $S(t)$ is the price level. In economic terms μ is the nominal dt-growth rate when $\sigma^2(t) = 0$ [3] , i.e. the no-arbitrage nominal dt-growth rate. The diffusion typically dominates the drift and although modelling the drift dynamically is considered most work is concentrated on estimating the volatility. We are also assuming that there is no risk premium, or at least we do not price market risk.

$\sigma^2(t)$ is called the *spot volatility*. The spot volatility is unobservable as is the actual volatility σ_n^2, which is given by,

$$\sigma_n^2 = \int_0^{n\Delta} \sigma^2(u)du - \int_0^{(n-1)\Delta} \sigma^2(u)du, \tag{3.2}$$

where Δ is typically a small time interval. In this way actual volatility is a piecewise constant process representing the total return variation in the interval Δ. Note that σ_n^2 is not an approximation. Since volatility is a flow variable integrated volatility is a natural measure for the volatility over some interval.

Realised volatility defined as the sum of M squared intraday changes over a fixed interval,

[2] In practice a small negative correlation is observed between these two terms. Many authors incorporate correlation in the model, cf. (Heston, 1993). However for an unobservable one such as the one that will be considered in this paper this assumption is needed for the model to be tractable

[3] as $\mathbb{E}(S(t+\delta)/S(t)) = \exp^{(\mu + 1/2\sigma^2(t))dt}$.

$$z_n = \sum_{j=1}^{M} \left[S^l \left\{ (n-1)\Delta + \frac{\Delta(j+1)}{M} \right\} - S^l \left\{ (n-1)\Delta + \frac{\Delta j}{M} \right\} \right]^2, \tag{3.3}$$

is an estimate of σ_n^2. Each squared return is an approximation to the spot volatility. If the objective is to model the spot volatility alone a rolling sample of intraday returns, (Andreou and Ghysels, 2002), would typically be used. Realised volatility is a consistent estimate as $M \to \infty$ while it is unbiased when $\mu = 0$. For this reason it is perhaps the most popular volatility proxy and is used as a benchmark in assessing out of sample performance for a whole range of models' forecasts. For the sampling interval that are used in this paper μ is indeed very close to zero and thus the bias is small enough to be ignored. In practice, due to discontinuity in the stock price path, the returns are typically only sampled every couple of minutes inducing a restriction on the size of M. This implies that z_n will be a noisy estimate of the actual volatility. Thus we have the measurement equation,

$$z_n = \sigma_n^2 + u_n, \tag{3.4}$$

where $\mathbb{E}(u_n \mid \sigma_n) = 0$. Since realised volatility is the sum of squared returns measured in noise the measurement noise will be asymptotically normal. A state space formulation can then be set up to estimate and predict actual volatility. Employing the Kalman filter provides state estimates that are optimal, in a MSE sense, if the noise is Gaussian and are best linear estimates even if the noise is non-Gaussian. We will use a mean reverting Gaussian Ornstein-Uhlenbeck (OU) type stochastic process to model the actual volatility. The reason for using a mean reverting process is that it has been observed that volatility stays close to a mean level for a period of time and when volatility leaves this mean level pressures of supply and demand cause it to revert to it at a certain rate; see (Fouque *et al.*, 2000) for a more in depth explanation. Mean reversion and autoregression are terms for the clustering phenomenon that has already been referred to. A first order discretisation of the OU-process, $d\sigma^2(t) = a(b - \sigma^2(t))dt + \beta dW(t)$, together with the measurement equation give the state-space model,

$$\begin{aligned} \sigma_{n+1}^2 &= \phi\sigma_n^2 + \Gamma + q\eta_{n+1} \\ z_n &= \sigma_n^2 + r\epsilon_n, \end{aligned} \tag{3.5}$$

where $r\epsilon_n = u_n$, $\phi = 1 - a\Delta, \Gamma = ab\Delta, q = \beta\sqrt{\Delta}$ and where the disturbances η_n, or innovations, and , or measurement noise, ϵ_n, have a unit variance. The state equation innovations ϵ_n are uncorrelated with the return innovations. (Barndorff-Nielsen and Shephard, 2002) show that $\mathbb{E}(u_n^2) = r^2$ is a function of Δ as well as the mean, variance and autocorrelation of the continuous time OU process. Although these authors give the relation between these and the mean, variance and covariance of the discrete time model practical implementation remains an issue. Instead we decide to estimate the model without incorporating this dependence on Δ, i.e. we take r as a parameter to be estimated.

In the last few years substantial theoretical work has been carried out designed at formulating models which reproduce the statistical features often found in financial time-series data. Of note is the work by (Barndorff-Nielsen and Shephard, 2002) and (Barndorff-Nielsen and Shephard, 2001) on Lévy processes and integrated volatility. These authors show how non-normality and leverage can be incorporated in SV models for high frequency data. A non-Gaussian OU process of the form $d\sigma^2(t) = -\lambda\sigma^2(t)dt + dZ(\lambda t)$ will typically lead to an ARMA(1,1) representation. The latter implies that $\sigma^2_{n+1} = \phi\sigma^2_n + \Gamma + q\eta_{n+1} + q\theta\eta_n$ [4]. An alternative formulation, or one that can be used in addition, is to use a mixture (linear combination) of OU processes, (Barndorff-Nielsen and Shephard, 2002). In this way σ^2_{n+1} is composed of several OU processes or 'components'. These authors report positive results for this specification over using a single process. The model (3.5) can be called a Realised Volatility Unobserved Components (RV-UC) where there is only one component. The curse of dimensionality hinders using many components although from the literature it appears that two or three components suffices.

In the next section the calibration of the above model will be considered.

[4] A model of this form is tractable but from the implementation carried out appears to offer little improvement in fitting the data we consider

3.3 The calibration procedure

For the calibration of the above model the hidden state will be estimated using the Kalman Filter and the parameters by a Maximum Likelihood method.

3.3.1 Kalman Filter

The Kalman Filter is a recursive algorithm which gives a conditional one-step-ahead estimate of the hidden state, $\hat{\sigma}^2_{n+1|n}$, and the variance of this estimate, $P_{n+1|n}$. The former is principally given as some function K_n of the previous step's state estimation variance and noise variance as well as the previous step's *innovation*. The innovation in this setup is nothing more than the difference between the observations and the previous step's state prediction, $z_n - \hat{\sigma}^2_{n|n-1}$. The state estimation variance is principally given as a function of K_n and the variance of innovations, F_n, as well as noise variance. The Kalman Filter equations (2.28) for the present setup are

$$\begin{aligned} v_n &= z_n - \hat{\sigma}^2_{n|n-1}, \\ F_n &= P_{n|n-1} + r^2, \\ K_n &= \phi P_{n|n-1} F_n^{-1}, \\ \hat{\sigma}^2_{n+1|n} &= \phi \hat{\sigma}^2_{n|n-1} + \Gamma + K_n v_n, \\ P_{n+1|n} &= \phi^2 P_{n|n-1} + q^2 - K_n^2 F_n. \end{aligned} \tag{3.6}$$

From the above it can be seen that the states are being estimated by one-step-ahead prediction. To initialise the Kalman Filter estimates for the mean and variance of the initial state are needed. Due to the simplicity and stationarity of the model these values can be set equal to the expected stationary volatility and the variance of the stationary volatility respectively. As we see from the following equations, $\hat{\sigma}^2_0 = \Gamma/(1-\phi), P_0 = q^2/(1-\phi^2)$, these moments are just given in terms of the unknown parameters.

To compare forecasting two-step ahead and four-step ahead with a shorter sampling interval with one-step with a longer one some of the equations of (3.6) will have to be extended to

two-step and four-step. In particular state variance and innovation variance equations are required. The equations for these were introduced in the section on multi-step prediction. When $c = 1$, as is the case in model (3.5), for two steps these are,

$$\bar{P}_n = \mathbb{E}(\bar{\sigma}^2 - \hat{\bar{\sigma}}^2)^2 = (\phi + 1)^2 P_{n+1|n} + q^2, \tag{3.7}$$

where,

$$\bar{\sigma}^2 = \sigma^2_{n+2} + \sigma^2_{n+1}. \tag{3.8}$$

The total variance of innovations is given by,

$$\bar{F}_n = \mathbb{E}(\bar{v})^2 = (\phi + 1)^2 P_{n-1|n-2} + q^2 + 2r^2, \tag{3.9}$$

where,

$$\bar{v} = (z_n + z_{n-1}) - (\hat{\sigma}^2_{n|n-2} + \hat{\sigma}^2_{n-1|n-2}). \tag{3.10}$$

and similar expressions for four steps.

In the next section the estimation of the model parameters will be considered using the output of the Kalman Filter.

3.3.2 Maximum Likelihood Estimation

As was seen in the preliminaries the general idea of Maximum Likelihood estimation is that the model parameters are chosen in such a way as to maximise the joint probability function of the observations. This will involve minimising the innovations and their variance. Consider the joint probability function of the observations conditional on past observations,

$$P(z_n \mid \underline{\theta}, Z_{n-1}) = \prod_{n=1}^{T} p(z_n \mid Z_{n-1}, \underline{\theta}), \tag{3.11}$$

where $\underline{\theta}$ is the vector of parameters, $Z_n = \{z_n, ..., z_1\}$ and T is the sample size. Once the observations have been realised $P(z_n \mid \underline{\theta}, Z_{n-1})$ is reinterpreted as the likelihood function

$L(.)$. Assuming that the probability densities are normal, maximising the likelihood of the observations is equivalent to maximising the log-likelihood,

$$L_{\log}(z_n \mid \underline{\theta}, Z_{n-1}) = -\sum_{n=1}^{T} \left(\log F_n + \frac{v_n^2}{F_n} \right), \tag{3.12}$$

when the constant term is ignored. We see from (3.12) that the innovations with a smaller variance are given more weight in the optimisation. The parameter vector which maximises the likelihood of the observations is called the maximum likelihood estimate.

The assumption that the noise terms in (3.5) are Gaussian means that a standard Gaussian Filter, such as the Kalman Filter described earlier can be applied to obtain the innovations and their conditional variances. A Gaussian Filter being applied to a model which is not necessarily Gaussian implies that the state estimates may be biased and thus the estimation will be suboptimal. In *Quasi-Maximum Likelihood Estimation* (QMLE), (Bollerslev and Wooldridge, 1992), these biases are ignored in the actual estimation. These are however accounted for when calculating standard errors on the estimates. The Kalman Filter can be extended to deal with models with non-linearity in the state although one has to make do with approximations. The Euler discretisation of the Feller model (2.17) for $f(Y) = \sqrt{Y}$, $\sigma_{n+1}^2 = \phi\sigma_n^2 + \Gamma + \beta\sqrt{\sigma_n^2}\,\eta_{n+1}$, leads to a particularly simple extension and approximation. The only difference in (3.6) is the last equation which is extended to $P_{n+1|n} = \phi^2 P_{n|n-1} + \hat{\sigma}_n^2 q^2 - K_n^2 F_n$. The $\hat{\sigma}_n^2$ term has been substituted in for the unknown σ_n^2.

3.3.3 Choice of sampling interval

Having set up the model and the calibration procedure the objective in this paper is to look at choosing a suitable sampling interval over which to calculate the volatility and to forecast. In this way the sampling interval that best captures the underlying mean reverting dynamics will be examined. A sampling interval with a better fit to the data will imply a better forecasting performance as shall be seen. So that the comparisons can be made on an equal par the forecast horizon should be equal for the different sampling intervals. To carry out such a procedure both one and two-step-ahead forecasts in succession will be needed for the shorter sampling interval. In this way the same forecasting horizon can be considered as

for the one-step-ahead forecast for a longer sampling interval. Thus for a shorter sampling interval the cumulative predicted volatility from the one and two steps will be taken as the volatility over the relevant forecast horizon.

3.4 Numerical Experiments

A relevant data set of two very liquid assets was used as input to the state space model. *PriceData* provided 10 minute intraday data for *IBM* from April 1997 to January 2005 and *Citigroup* from May 1997 to March 2005 and giving over 76300 observations. From this we took the first 76128 returns corresponding to $N = 1952$ days of data given that there were 39 bars per day. Days that contained no data whatsoever, such as holidays and weekends, were ignored. For simplicity overnight and over-weekend/holiday returns were not treated any differently to the intraday ones. (Koopman and Hol, 2002) suggest alternative treatments.

The data were preprocessed by stochastically interpolating any missing values in the data within a one day period. The standard deviation value used in the interpolation was an average of standard deviation values from a group of sub-intervals proceeding the missing price range. In calculating the standard deviation of these sub-intervals the stock prices were scaled by the stock price value in the middle of the interval. The reason for using an average of sub-interval values was to lessen the dependence of the standard deviation on the drift, the impact of which was seen to be otherwise quite significant. The motivation for scaling was due to the fact that the standard deviation of unscaled stock prices were seen to be proportional to the stock price level. Scaling the stock prices nullified this dependence.

It should be noted that preprocessing the data had little impact on the total variation of the data, i.e. the variance of the raw data, with missing prices taking the value zero, and the variance of the data after the intervention, were very similar. The fact that after carrying out the preprocessing every trading day contained a full set of stock prices meant that we could proceed to estimating a model for sampling intervals of one day (or multiples of one day).

The stock price path was sampled every 10 minutes corresponding to $M = 39, 78$ and 156 in (3.3) for $\Delta = 1, 2$ and 4 respectively. Although the choice of sampling frequency was is a little ad hoc the emphasis here is on the choice of sampling interval. Moreover for the purposes of comparison of different sampling interval procedures what is most important is that the sampling frequency is fixed. The value of μ in (3.1) was close to zero so the bias in using realised volatility as an estimate of actual volatility was ignored. Realised volatility outliers corresponding to values in the upper 0.5% percentile of the sample were effectively removed by setting the values of these equal to the mean of the observations.

The optimisation of (3.12) was carried out using *Matlab*'s *fmincon* algorithm. *fmincon* is an algorithm tuned to finding the optimum parameter set of a non-linear function of several variables with constraints. It uses a sequential quadratic programming method. In this method, the function solves a quadratic programming problem at each iteration. An estimate of the Hessian of the Lagrangian is updated using the BFGS formula. In the model in question there are natural constraints on some of the variables such as $\phi < 1$ [5] and $0 < q, r < \infty$. Starting off with these it became apparent that a local optimum was being found. Having some idea of the magnitude of the parameters some of the bounds were then tightened. After carrying this out a better (and probably global) optimum was arrived at.

The estimates for ϕ, Γ, q and r and are given in the following table:

Table 1. Parameter values for IBM and Citigroup stocks

	IBM stock			Citigroup stock		
	$\Delta = 1$	$\Delta = 2$	$\Delta = 4$	$\Delta = 1$	$\Delta = 2$	$\Delta = 4$
ϕ	$9.44 * 10^{-1}$	$9.20 * 10^{-1}$	$8.21 * 10^{-1}$	$9.71 * 10^{-1}$	$9.46 * 10^{-1}$	$9.22 * 10^{-1}$
Γ	$2.39 * 10^{-5}$	$7.05 * 10^{-5}$	$3.25 * 10^{-4}$	$1.54 * 10^{-5}$	$5.95 * 10^{-5}$	$1.73 * 10^{-4}$
q	$1.08 * 10^{-4}$	$2.66 * 10^{-4}$	$8.47 * 10^{-4}$	$1.10 * 10^{-4}$	$3.12 * 10^{-4}$	$6.84 * 10^{-4}$
r	$4.31 * 10^{-4}$	$7.68 * 10^{-4}$	$1.26 * 10^{-3}$	$4.83 * 10^{-4}$	$8.77 * 10^{-4}$	$1.65 * 10^{-3}$

[5] this constraint assures some degree of stationarity and is a consequence of the assumed underlying dynamics.

Let us note that ϕ, the serial correlation, decreases as the sampling interval increases whereas the magnitude of the disturbances increases. This is to be expected as it demonstrates that less predictive power implies more noisy predictions. Along with comparing the forecasting performance of different time scales the model should be validated. To this end standard errors on the parameters will be found. Besides this, a comparison will be made with the forecasting performance of a constant volatility model.

A Bayesian approach to parameter estimation would be to find the whole distribution of the parameter estimates. This is carried out using prior information on the parameters' distributions together with a likelihood function. Here the classical approach is taken where only the first two moments of the parameter estimates' distribution are considered. The first moment corresponds to the values given in the table above. The 'standard errors' correspond to the second moment, i.e. the diagonal entries of the parameter covariance matrix. This matrix was estimated taking into account the non-normality of disturbances and as such we will refer to the errors being QMLE standard errors. These, having been normalised w.r.t. their nominal values, are given in the following table:

Table 2. QMLE standard errors for the IBM and Citigroup stocks

	IBM stock			Citigroup stock		
	$\Delta = 1$	$\Delta = 2$	$\Delta = 4$	$\Delta = 1$	$\Delta = 2$	$\Delta = 4$
ϕ	$5.56 * 10^{-7}$	$5.04 * 10^{-5}$	$1.69 * 10^{-5}$	$7.94 * 10^{-8}$	$2.06 * 10^{-6}$	$2.30 * 10^{-4}$
Γ	$1.10 * 10^{-4}$	$1.14 * 10^{-3}$	$6.63 * 10^{-5}$	$6.76 * 10^{-6}$	$4.76 * 10^{-5}$	$1.10 * 10^{-3}$
q	$8.40 * 10^{-7}$	$5.89 * 10^{-6}$	$5.92 * 10^{-6}$	$8.32 * 10^{-7}$	$3.93 * 10^{-6}$	$1.43 * 10^{-4}$
r	$5.36 * 10^{-6}$	$1.04 * 10^{-4}$	$1.38 * 10^{-5}$	$8.47 * 10^{-7}$	$1.32 * 10^{-5}$	$3.71 * 10^{-5}$

The standard errors on the estimates as shown in the table above are seen to be small for all three sampling intervals under consideration and for both stocks.

The motivation for the QMLE approach taken here was that the literature suggests that disturbances in a model structure such as the one chosen here are non-normal. This was confirmed by a Jacque-Berra test. There will thus be estimation biases induced that we have ignored. However the focus here is on actual forecasting performance rather than the

statistical validity of the underlying model.

A comparison will now be made between forecasting two steps and four steps ahead for a shorter sampling interval against forecasting only one step ahead respectively for a longer sampling interval. This will be done first for $\Delta = 1$ against $\Delta = 2$, then for $\Delta = 2$ against $\Delta = 4$ and then finally for $\Delta = 1$ against $\Delta = 4$. In the following tables there is a summary of the forecasting performance results for the two stocks. The performance is based on the average conditional state variance, the average innovation and the average variance of innovations.

Table 3. Comparison of forecasts for the IBM and Citigroup stocks

	IBM stock		Citigroup stock	
	$\Delta = 1$, 2-step	$\Delta = 2$, 1-step	$\Delta = 1$, 2-step	$\Delta = 2$, 1-step
$1/T \sum_1^T \bar{P}_{n+1\|n}$	$1.71 * 10^{-7}$	$1.95 * 10^{-7}$	$2.17 * 10^{-7}$	$2.82 * 10^{-7}$
$1/T \sum_1^T \mid \bar{v}_n \mid$	$4.35 * 10^{-4}$	$4.85 * 10^{-4}$	$4.43 * 10^{-4}$	$5.11 * 10^{-4}$
$1/T \sum_1^T \bar{F}_n$	$5.42 * 10^{-7}$	$7.84 * 10^{-7}$	$6.83 * 10^{-7}$	$1.05 * 10^{-6}$
	$\Delta = 2$, 2-step	$\Delta = 4$, 1-step	$\Delta = 2$, 2-step	$\Delta = 4$, 1-step
$1/T \sum_1^T \bar{P}_n$	$7.89 * 10^{-7}$	$1.17 * 10^{-6}$	$1.17 * 10^{-6}$	$1.16 * 10^{-6}$
$1/T \sum_1^T \mid \bar{v}_n \mid$	$8.89 * 10^{-4}$	$9.67 * 10^{-4}$	$9.27 * 10^{-4}$	$1.02 * 10^{-3}$
$1/T \sum_1^T \bar{F}_n$	$1.97 * 10^{-6}$	$2.77 * 10^{-6}$	$2.70 * 10^{-6}$	$3.89 * 10^{-6}$
	$\Delta = 1$, 4-step	$\Delta = 4$, 1-step	$\Delta = 1$, 4-step	$\Delta = 4$, 1-step
$1/T \sum_1^T \bar{P}_n$	$7.20 * 10^{-7}$	$1.17 * 10^{-6}$	$9.36 * 10^{-7}$	$1.16 * 10^{-6}$
$1/T \sum_1^T \mid \bar{v}_n \mid$	$7.99 * 10^{-4}$	$9.67 * 10^{-4}$	$8.37 * 10^{-4}$	$1.02 * 10^{-3}$
$1/T \sum_1^T \bar{F}_n$	$1.46 * 10^{-6}$	$2.77 * 10^{-6}$	$1.87 * 10^{-6}$	$3.89 * 10^{-6}$

where $T = N/\Delta$ and where for one-step $\bar{P}_n = P_{n|n-1}$ and $\bar{F}_n = F_n$ and $\bar{v}_n = v_n$ as given in (3.6). For purposes of comparison the states themselves are of the orders 10^{-3} to 10^{-4}. As can be seen from the above tables the prediction performance is generally better for the multi-step prediction. Although the difference in the multi-step versus one-step results is not very significant the consistency of these strongly favours forecasting using a shorter sampling interval. This may appear to be counterintuitive due to the fact that for the longer intervals there are twice or four times as many points in the sample. However this demonstrates that

the mean reversion dynamics are not being captured so well for longer sampling intervals. Although the focus here is on actual forecasting performance rather than the statistical validity of the model, as has been pointed out, a statistical test was carried out. The Durbin-Watson test confirmed that the innovations were not serially correlated.

Finally a validation of the model for $\Delta = 1$ will now be made by means of comparison with a constant volatility (CV) model and a daily GARCH(1,1) model, i.e. the model $\sigma^2(k) = \gamma + \beta\sigma^2(k-1) + \alpha\varepsilon^2(k-1)$ where $\varepsilon(k-1)$ are daily demeaned returns and $\varepsilon(k) \mid F(k-1) \sim N(0, \sigma^2(k))$. $F(k-1)$ is the information set including all the information available up to time $k-1$. This will be done in terms of the out-of-sample innovations of the two models for different sample sizes. These will be given by the difference between the measurements and the predicted out-of-sample volatility values, for 10 points, i.e stepping ten points ahead using the out-of-sample predictions to step ahead, one step at a time. Table 4. gives a summary of the results for $\Delta = 1$, N= 800, 1150 and 1500 days, each entry being an average of the absolute values of the innovations over the ten data points, $\sum_{i=1}^{10} \mid \hat{\sigma}^2_{N+i|N} - z_{N+i} \mid$:

Table 4. Out-of-sample innovations for the IBM and Citigroup stocks

	$N = 800$	$N = 1150$	$N = 1500$
CV: IBM stock	$1.95 * 10^{-4}$	$3.01 * 10^{-4}$	$2.35 * 10^{-4}$
SV: IBM stock	$1.67 * 10^{-4}$	$1.20 * 10^{-4}$	$1.16 * 10^{-4}$
GARCH: IBM stock	$1.45 * 10^{-4}$	$9.62 * 10^{-5}$	$1.19 * 10^{-4}$
CV: Citigroup stock	$3.81 * 10^{-4}$	$4.09 * 10^{-4}$	$4.56 * 10^{-4}$
SV: Citigroup stock	$1.45 * 10^{-4}$	$1.39 * 10^{-4}$	$1.20 * 10^{-4}$
GARCH: Citigroup stock	$7.49 * 10^{-5}$	$6.45 * 10^{-5}$	$1.48 * 10^{-4}$

Constant volatility and GARCH models are commonplace in industry and can be used as benchmark models. It can be seen from the table that for all three sample sizes SV outperforms constant volatility considerably, despite the simplicity of the SV model that was used and that the estimation procedure was relatively straightforward. The SV model's superior out-of-sample performance gives us a fair degree of confidence in it and in this way also in the results previously given on the choice of sampling intervals. It can be seen however that the GARCH model generally performs better than the SV model. (Andersen and Boller-

slev, 1998) have shown that GARCH models perform well out of sample when using realised volatility as a proxy for the ex-post volatility as confirmed also here. However rejecting the SV model is not warranted because of the lack of consistency in the out-performance of the GARCH model over the SV model.

3.5 Conclusion

The contributions of this chapter are three-fold. First we have illustrated calibration and (in-sample as well as out-of-sample) forecasting performance of a stochastic volatility model using high frequency asset price data. Secondly we have compared different sampling intervals in terms of forecasting performance. Our study suggests that the serial correlation in volatility is best modelled over a shorter sampling interval. With these results in hand we can apply the model to the estimation of the volatility term structure. Finally the superiority of out-of-sample prediction performance using a SV model over using a constant volatility model is demonstrated.

Chapter 4

Linear models for high frequency data with and without filtering

It is known however high frequency price path data are subject to microstructure noise and so estimates of the volatility from price path differentials are also noisy, cf. (Zumbach *et al.*, 2002), (Andersen *et al.*, 1999), (Bai *et al.*, 2003) and (Bollerslev and Zhou, 2002). Beyond this there are other sources of measurement noise the impact of which depends on the sampling frequency. A whole range of statistical methods that had previously been applied mainly in problems in engineering and physics have been proposed and implemented to deal with noise in volatility. Filtering the noise using a Gaussian Filter such as the Kalman Filter for linear systems and the Extended Kalman Filter for non-linear systems is the obvious way of doing this; see (Anderson and Moore, 1979) for the theory and (Barndorff-Nielsen and Shephard, 2002), (Owens and Steigerwald, 2006) and (Alizadeh *et al.*, 2002) for applications. Some other less standard approaches to filtering are given in (Bandi and Russell, 2004), (Zumbach *et al.*, 2002). Hidden Markov Chain models have also been applied, (Turner *et al.*, 1989), (Hamilton and Susmel, 1994) and (Elliott *et al.*, 2003), although these estimation methods are perhaps generally more well suited to interest rate and commodity modelling. Simulation based methods are quite popular, cf. (Chib *et al.*, 2002). In a multivariate setting noise has been dealt with using certain shrinkage techniques stemming from physics-based research, cf. (Plerou *et al.*, 2002) and (Andersson *et al.*, 2005). Last but not least GARCH models deliver filtered estimates of the volatility, cf. (Nelson, 1992). More details on GARCH

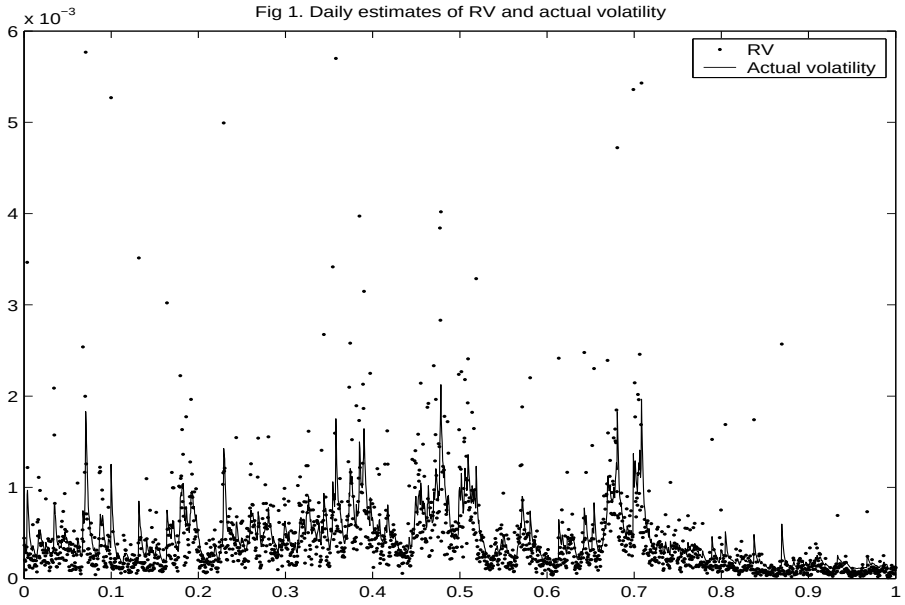
models as filters will be given later on in this chapter. Work has been carried out on estimation without filtering by defining an optimal sampling frequency in which minimum noise contamination is weighed against consistency. See (Ait-Sahalia *et al.*, 2005), (Oomen, 2002) and (Bandi and Russell, 2004).

As has been alluded to, when the sampling frequency is high, RV is subject to microstructure noise that contaminates the measurement of the true volatility. Each high frequency squared return in (3.3) is an approximation to the spot volatility as defined earlier. Let us note a dichotomy in regard to sampling. As the sampling frequency increases the measurement noise moves towards normality but on the other hand the microstructure noise also increases[1]. Numerous studies have been carried out on the impact of microstructure noise. It appears that above a sampling frequency of around five minutes microstructure noise contributes significantly to price path measurements. It may be shown that RV underestimates the actual volatility in the presence of this kind of noise. At lower frequencies measurement noise results in particular from using $(S_l(t+\delta) - S_l(t))^2$ as an approximation to $\sigma^2(t)$. Although high frequency squared returns are unbiased estimators, the variance of the estimation error increases as M decreases. This estimation error is called idiosyncratic noise. The objective here is to assess the importance of filtering for a sampling frequency in which it unlikely that microstructure noise is present and in which at the same time idiosyncratic noise is relatively small. To this end filtered volatility estimation and unfiltered model calibration will be carried out comparing in- and out-of-sample forecasting performance. An issue that makes a study of this kind particularly relevant is that a hidden state estimation limits the introduction of explanatory variables other than past (filtered) estimates of the same series. If the empirical results indicate little difference in the filtered and unfiltered approaches direct estimation of observable models will have some justification.

[1]clearly increasing the sampling interval resolves this issue but the emphasis here is on short duration volatility measures.

4.1 Introduction to the models

The question arises of how persistent volatility really is and how well it can be forecast. In the previous chapter we had established that filtered RV was highly persistent and that out-of-sample prediction performed fairly well. In the case of unfiltered estimates we shall see that for a RV model with a noise term there is not much persistence and performs poorly in comparison to the hidden state model. By contrast for the same RV model with a non-standard estimation procedure a fair degree of persistence is imputed but similarly to the previous model both in- and out-of-sample average innovations are larger than for the hidden state model. This is not surprising if we consider Fig. 1. where we see just how noisy RV is.



The poor performance of the two observable models is at odds with the results for the observable GARCH models, cf. (Andersen and Bollerslev, 1998), but the difference is that our estimates[2] are non-filtered. GARCH models implicitly filter out measurement noise. As shall be seen our results point to filtering in the context of RV even when working at a relatively low sampling frequency.

Throughout this chapter we shall be considering three estimation methods in the context of RV: Quasi Maximum likelihood estimation for a latent SV model, Quasi-GARCH Maximum Likelihood and Non-linear Least Squares (NLLS) estimation for a SV one-factor model.

[2]from 10 minute returns over a day

4.1.1 Latent SV model

The first model to be considered is the state-space model of the previous chapter,

$$\begin{aligned} \sigma_{n+1}^2 &= \phi\sigma_n^2 + \gamma + q\eta_{n+1}, \\ z_n &= \sigma_n^2 + r\epsilon_n. \end{aligned} \tag{4.1}$$

Since the model and its estimation procedure are identical to those in the previous chapter the next model will be introduced.

4.1.2 Quasi-GARCH Maximum Likelihood SV model

The second model to be considered is a Stochastic Volatility-type model in the sense that it has an extra source of randomness other than the asset price one but its structure and calibration procedure is akin to that of a GARCH model. Standard GARCH models owe their popularity to being simple to estimate and versatile. Traditionally daily or weekly GARCH processes were considered but the advent of high frequency data and the temporal aggregation results of (Drost and Nijman, 1993) brought a lot of consideration to GARCH at high frequency. These authors showed how certain GARCH models at one frequency could be inferred from models at another frequency. Thus for certain specifications one could for example estimate a model at a frequency of five minutes say and infer the daily GARCH model. Unfortunately there appear to be two issues that may impede using temporal aggregation. The first is that at a high frequency estimation is computationally burdensome. The second is that strong intraday periodic patterns have been observed as first pointed out by (Andersen and Bollerslev, 1997).

We may still be able to make use of high frequency data using a parsimonious model akin to the GARCH(1,1) specification and calibrated in a similar way. For these reasons we keep the nomenclature, despite the deviation from a standard GARCH model as explained above and shall be seen, and denote the estimation as 'quasi-GARCH'. We substitute the variance terms of the GARCH(1,1) model by RV. In this way the model takes the form

$$z_n = \gamma + \phi z_{n-1} + \alpha\varepsilon_{n-1}^2 + \eta_n \tag{4.2}$$

where η_n is zero mean and z_n is RV and ε_n are daily demeaned returns. When using RV together with daily returns information is lost in the latter relation to the former. Having said this we still feel that it is sensible approach and if α turns out to be significant in the above model including r_{n-1} as an additional factor makes sense. Moreover the simplicity of the above model makes it a good candidate for widespread implementation by practitioners. Unlike the SV model in the next section the distribution of noise in (4.2) is left unspecified, although we assume normality of returns.

To calibrate the above model we consider the probability density of the daily demeaned returns, with mean zero and (the realised volatility estimate of the) variance z_n,

$$p(\varepsilon_n) = \frac{1}{\sqrt{2\pi z_n}} \exp\left(-\frac{\varepsilon_n^2}{2z_n}\right) \tag{4.3}$$

then maximise the joint log-likelihood

$$L_{\log}((\varepsilon_n) = -\sum_{k=1}^{T} \log z_n - \sum_{k=1}^{T} \frac{\varepsilon_n^2}{z_n} \tag{4.4}$$

ignoring constants. We then substitute the variance by its prediction from (4.2) and maximise the pseudo-log-likelihood

$$L_{\log}(\varepsilon_n \mid \mathcal{F}_{n-1}, \underline{\theta}) = -\sum_{k=1}^{T} \log(\gamma + \beta z_{n-1} + \alpha\varepsilon_{n-1}^2) - \sum_{k=1}^{T} \frac{\varepsilon_n^2}{\gamma + \phi z_{n-1} + \alpha\varepsilon_{n-1}^2} \tag{4.5}$$

Thus in the above model the variance process will be discontinuous and in a sense a random process though not the true random process for volatility by any stretch of the imagination. The model above allows variance to vary randomly but is driven by prediction errors which are not minimised leading to lack of efficiency. We seek to see however if there is some improvement for in- and out-of-sample forecasting as compared to standard (non-linear) least squares (NLLS) estimation. GARCH models assume a smooth process whereas the true data generating process for variance is a random process. GARCH models are therefore misspecified to some degree since the assumption is that variance is only time-varying while the true data generating process is used for the model calibration. However these are known to perform very well as demonstrated in the previous chapter.

4.1.3 NLLS estimation of a one factor SV model

The third model is has the structure of the previous model except that we assume the noise is Gaussian

$$z_n = \gamma + \phi z_{n-1} + \alpha \varepsilon_{n-1}^2 + \eta_n \tag{4.6}$$

where $\eta_n \sim N(0, C)$. The estimation procedure is simply carried out by minimising the sum of the squares of prediction errors but to keep close to the estimation procedures of the previous two models we will approach the estimation by maximising the likelihood. Let us denote η_n as the prediction error process and let us consider the joint log-likelihood of this process

$$L_{\log}(\eta_n \mid \mathcal{F}_{n-1}, \underline{\theta}) = -\sum_{n=1}^{T} \log(C) - \sum_{n=1}^{T} \frac{(z_n - \gamma - \phi z_{n-1} - \alpha \varepsilon_{n-1}^2)^2}{C} \tag{4.7}$$

ignoring constants. Maximising the above is equivalent to NLLS as the variance is constant. Via the above likelihood formulation we see that we have a handle on other specifications for the noise such as a GARCH specification but there seems to be no particular reason why it should be anything else but constant in variance.

In the next section parameter estimates and forecasting results will be presented.

4.2 Numerical results

The data set

The data set used was the same as that of the previous chapter and the preprocessing was the same as well. Similarly to the last chapter the stock price path was sampled every 10 minutes corresponding to $M = 39$ in (2.17) for $\Delta = 1$. This choice of sampling frequency was not entirely ad hoc. It was of interest to see at relatively high frequency (but not ultra high) what the impact of the measurement noise would be while at the same time the sample would be large enough for reliable estimation.

The focus here is on in-sample and out-of-sample prediction. Before displaying these results it is of interest to show the parameter estimates of the different models. The value of the

parameter α for the observable models was not significant so only the other two parameters are displayed. The estimates for ϕ and γ are given in the following table:

Table 5. Parameter values for IBM and Citigroup stocks

	IBM stock			Citigroup stock		
	RV-UC	Quasi-Garch	NLLS	RV-UC	Quasi-Garch	NLLS
ϕ	$9.44 * 10^{-1}$	$9.00 * 10^{-1}$	$3.88 * 10^{-1}$	$9.71 * 10^{-1}$	$9.00 * 10^{-1}$	$3.63 * 10^{-1}$
γ	$2.39 * 10^{-5}$	$9.32 * 10^{-5}$	$2.51 * 10^{-4}$	$1.54 * 10^{-5}$	$6.90 * 10^{-5}$	$3.07 * 10^{-4}$

For the estimation of the above models normal and independent observations were assumed. However for each of the models these assumptions were consistently rejected by statistical tests. Thus we find ourselves in the QML estimation setup. QML standard errors on the estimates have been produced and we find that these were large for the observable models but small for the hidden model.

A comparison of in-sample prediction for the three models will now be presented. This will be based on the average of the absolute values of the innovations over the whole sample. It can be seen from the table below that the hidden SV model performs the best followed by the RV model and then Quasi-GARCH model.

Table 6. In-sample innovations for the IBM and Citigroup stocks

	$1/T \sum_1^T \mid v_n \mid$
NLLS RV: IBM stock	$2.46 * 10^{-4}$
Quasi-GARCH RV: IBM stock	$2.75 * 10^{-4}$
SV: IBM stock	$2.43 * 10^{-4}$
NNLS RV: Citigroup stock	$3.09 * 10^{-4}$
Quasi-GARCH RV: Citigroup stock	$3.29 * 10^{-4}$
SV: Citigroup stock	$2.84 * 10^{-4}$

Table 7 gives a summary of the results for samples of sizes N= $800, 1150$ and 1500, each entry being an average of the absolute values of the innovations over the ten data points out of the sample:

Table 7. Out-of-sample innovations for the IBM and Citigroup stocks

	$1/10\sum_1^{10} \mid v_n \mid$		
	$N = 800$	$N = 1150$	$N = 1500$
NLLS RV: IBM stock	$1.71 * 10^{-4}$	$2.85 * 10^{-4}$	$2.45 * 10^{-4}$
Quasi-GARCH RV: IBM stock	$3.06 * 10^{-4}$	$3.54 * 10^{-4}$	$3.13 * 10^{-4}$
SV: IBM stock	$1.67 * 10^{-4}$	$1.20 * 10^{-4}$	$1.16 * 10^{-4}$
NNLS RV: IBM stock	$2.97 * 10^{-4}$	$3.73 * 10^{-4}$	$4.29 * 10^{-4}$
Quasi-GARCH RV: IBM stock	$1.83 * 10^{-4}$	$2.37 * 10^{-4}$	$2.52 * 10^{-4}$
SV: Citigroup stock	$1.45 * 10^{-4}$	$1.39 * 10^{-4}$	$1.20 * 10^{-4}$

Out-of-sample performance also favours the hidden SV model as can be seen from the table above.

4.3 Conclusion

Both the in-sample and out-of-sample results indicate that filtering is paramount and the gain obtained from this is quite significant. Furthermore the fact that the additional regressor in the observable model was not significant means that for the setup and data set used there is no advantage in using observable models over a latent one.

Conclusion

This report has covered some background theory related to linear state model estimation and prediction. This theory has successfully been applied in the estimation of models using high frequency stock price data. One of the conclusions from the empirical work is that filtering measurement noise improves volatility estimation and prediction for the data set used in this study. Different sampling intervals for the same forecasting horizon were compared for a realised volatility model with filtering, namely a latent state space model. The results indicate that the serial correlation in volatility is best modelled over a shorter sampling interval.

Work is presently being carried out in the estimation of two dimensional state space models. One of these models incorporates implied volatility measurements. The objective is to compare one and two state models in terms of forecasting performance and how well these fit the data.

Notation

$\mathbb{E}(X)$	expectation of a random variable X
$\sigma^2(t)$	spot volatility
σ_n^2	actual volatility
I_n	the identity matrix of size N
$i.i.d.$	independently and identically distributed
A^c	The complement of the set A

Appendix A: Maximum Likelihood Estimation

Here we seek to show the details behind calculating the covariance error matrix, and from this the standard errors, of a misspecified univariate hidden state model.

Consider the log-likelihood function,

$$L_{\log}(\underline{\theta}, \mathcal{F}_{k-1}) = -\sum_{k=1}^{n} \log F_k - \sum_{k=1}^{n} v_k^2 F_k^{-1}, \tag{1}$$

where v_k are the innovations, and F_k are their variance, $\underline{\theta}$ is the vector of unknown parameters and $\mathcal{F}_{k-1}$ is the filtration generated by the observation process up to time $k-1$. The innovations are defined as,

$$v_k = y_k - c\hat{x}_{k|k-1} - d,$$

where z_k are the observations and $\hat{x}_{k|k-1}$ are the estimates of the hidden state as given in (2.28). The minimisation of (1) is equivalent to maximising the probability of the outcome of the set of observations.

We see from (1) that the innovations with a smaller variance are given more weight in the optimisation. The parameter vector which maximises the likelihood of the observations is called the maximum likelihood estimate, $\hat{\underline{\theta}}$. If the sample size is sufficiently large and under certain regularity conditions $\hat{\underline{\theta}}$ can be approximated by the density,

$$\hat{\underline{\theta}} \approx N(\underline{\theta}_0, n^{-1} I^{-1}(\underline{\theta})), \tag{2}$$

where $\underline{\theta}_0$ denotes the true parameter vector. The matrix $I(\underline{\theta})$ is denoted the information matrix and is based on derivatives of the likelihood function w.r.t. the parameter vector. From the above we see that $\hat{\underline{\theta}}$ is an asymptotically unbiased estimator of $\underline{\theta}_0$. $I^{-1}(\underline{\theta}))$ is a minimum variance bound. In large samples we would expect the variance of a estimator to reach this bound; otherwise it would not be an efficient estimator (Harvey, 1981). There are two common estimators of $I(\underline{\theta})$. The *second derivative estimator* is given by

$$I(\hat{\underline{\theta}})_{2D} = -n^{-1} \left. \frac{\delta^2 L_{\log}(\underline{\theta}, \mathcal{F}_{k-1})}{\delta \underline{\theta} \delta \underline{\theta}^T} \right|_{\underline{\theta}=\hat{\underline{\theta}}} \tag{3}$$

The *outer product estimate* is given by

$$I(\hat{\underline{\theta}})_{OP} = n^{-1} \sum_{k=1}^{n} [h(\hat{\underline{\theta}}).h(\hat{\underline{\theta}})'] \tag{4}$$

where

$$h(\hat{\underline{\theta}}) = \left. \frac{\delta \log l_k(\underline{\theta}, \mathcal{F}_{k-1})}{\delta \underline{\theta}} \right|_{\underline{\theta}=\hat{\underline{\theta}}}$$

$\log l_k$ being the individual kth-term of (1). A model is said to be misspecified if, for example, the errors are not normal even though these have been assumed to be so in the filtering and estimation process. If this is the case $I(\hat{\underline{\theta}})_{OP}$ and $I(\hat{\underline{\theta}})_{2D}$ may diverge significantly from each other. An approximate covariance matrix for $\hat{\underline{\theta}}$ was given by (White, 1982),

$$\mathbb{E}(\hat{\underline{\theta}} - \underline{\theta}_0)(\hat{\underline{\theta}} - \underline{\theta}_0) \cong n^{-1}(I_{2D} I_{OP}^{-1} I_{2D})^{-1} \tag{5}$$

This approximation may be valid if the model is misspecified. This approach is known as quasi-maximum likelihood estimation.

To derive the actual expressions for (3) and (4) let us consider the k stage likelihood value,[3]

$$\log l_k = -\log F_k - v_k^2 F_k^{-1} \tag{6}$$

Differentiating $\log l_k$ with respect to the ith element of $\underline{\theta}$ gives the gradient,

[3] ignoring constants which do not affect the optimization.

$$\frac{\delta(\log l_k)}{\delta\theta_i} = -\left(F_k^{-1}\frac{\delta F_k}{\delta\theta_i}\right)\left(1 - F_k^{-1}v_k^2\right) - 2\frac{\delta v_k}{\delta\theta_i}F_k^{-1}v_k = h(\theta_i, y_k) \tag{7}$$

Differentiating the above with respect to the jth element of $\underline{\theta}$ gives,

$$\begin{aligned}\frac{\delta^2(\log l_k)}{\delta\theta_i\delta\theta_j} = - & \left(F_k^{-1}\frac{\delta^2 F_k}{\delta\theta_i\delta\theta_j} - \frac{\delta F_k}{\delta\theta_i}F_k^{-2}\frac{\delta F_k}{\delta\theta_j}\right)\left(1 - F_k^{-1}v_k^2\right) \\ & -\frac{\delta F_k}{\delta\theta_i}F_k^{-3}\frac{\delta F_k}{\delta\theta_j}v_k^2 + 2\frac{\delta F_k}{\delta\theta_i}F_k^{-2}\frac{\delta v_k}{\delta\theta_j}v_k \\ & -2\frac{\delta^2 v_k}{\delta\theta_i\delta\theta_j}F_k^{-1}v_k + 2\frac{\delta v_k}{\delta\theta_i}\frac{\delta F_k}{\delta\theta_j}F_k^{-2}v_k - 2\frac{\delta v_k}{\delta\theta_i}F_k^{-1}\frac{\delta v_k}{\delta\theta_j}\end{aligned} \tag{8}$$

Summing over k in (8) and dividing by n we have the ij-th element of the second derivative estimate of the information matrix as in (3). If the model is correctly specified the above simplifies considerably. The derivatives of F_k and v_k can be found using a set of recursions that run in parallel with the Kalman Filter. See (Harvey, 1989, p. 140-143), for example, for details. The presentation of the results for the off-diagonals of covariance matrix in (5) using (3) and (4) and the information matrix (8) are given in the numerical results section of Chapter 2.

For GARCH models a similar derivation to the one above can be carried out. The expressions however are greatly simplified as the $\epsilon(k)$ terms, corresponding to the v_k above, are not functions of the parameters[4] Multivariate state space models on the other hand involve very extensive derivations even for two state models such as the ones considered in this thesis. For this reason no multivariate derivations have been presented.

[4]unless the model is a GARCH regression which has not been considered in this thesis.

Appendix B: Multi-step Prediction

Proof of Proposition 1.

$$
\begin{aligned}
\mathbb{E}[\bar{x}-\hat{\bar{x}}]^2 &= \mathbb{E}[x_{n+m}-\hat{x}_{n+m|n}+x_{n+m-1}-\hat{x}_{n+m-1|n}+...+x_{n+2}-\hat{x}_{n+2|n}+x_{n+1}-\hat{x}_{n+1|n}]^2 \\
&= \mathbb{E}[ax_{n+m-1}+b+w_{n+m}-a\hat{x}_{n+m-1|n}-b+ax_{n+m-2}+b+w_{n+m-1}- \\
&\quad a\hat{x}_{n+m-2|n}-b+...+ax_{n+1}+b+w_{n+2}-a\hat{x}_{n+1|n}-b+x_{n+1}-\hat{x}_{n+1|n}]^2 \\
&= \mathbb{E}[a^2(x_{n+m-2}-\hat{x}_{n+m-2|n})+aw_{n+m-1}+w_{n+m}+a^2(x_{n+m-3}-\hat{x}_{n+m-3|n})+ \\
&\quad aw_{n+m-2}+w_{n+m-1}+...+a(x_{n+1}-\hat{x}_{n+1|n})+w_{n+2}+x_{n+1}-\hat{x}_{n+1|n}]^2 \\
&= ... \\
&= \mathbb{E}[a^{m-1}(x_{n+1}-\hat{x}_{n+1|n})+a^{m-2}w_{n+2}+a^{m-3}w_{n+3}+...+aw_{n+m-1}+w_{n+m}+ \\
&\quad a^{m-2}(x_{n+1}-\hat{x}_{n+1|n})+a^{m-3}w_{n+2}+a^{m-4}w_{n+3}+...+aw_{n+m-2}+w_{n+m-1}+... \\
&\quad a(x_{n+1}-\hat{x}_{n+1|n})+w_{n+2}+(x_{n+1}-\hat{x}_{n+1|n})]^2 \\
&= \mathbb{E}[(a^{m-1}+a^{m-2}+...+a+1)(x_{n+1}-\hat{x}_{n+1|n})+ \\
&\quad (a^{m-2}+a^{m-3}+...+a+1)w_{n+2}+(a^{m-3}+a^{m-4}+...+a+1)w_{n+3}+... \\
&\quad +(a+1)w_{n+m-1}+w_{n+m}]^2 \\
&= (a^{m-1}+a^{m-2}+...+a+1)^2P_{n+1|n}+ \\
&\quad (a^{m-2}+a^{m-3}+...+a+1)^2q^2+(a^{m-3}+a^{m-4}+...+a+1)^2q^2+...+(a+1)^2q^2+q^2 \\
&= \left(\sum_{j=0}^{m-1}a^j\right)^2P_{n+1|n}+\sum_{i=0}^{m-2}\left(\sum_{j=0}^{i}a^j\right)^2q^2 \qquad \blacksquare \qquad (9)
\end{aligned}
$$

Proof of Proposition 2.

$$
\begin{aligned}
\mathbb{E}[\bar{y} - c\hat{\bar{x}} - md]^2 &= \mathbb{E}[y_n - c\hat{x}_{n|n-m} - d + y_{n-1} - c\hat{x}_{n-1|n-m} - d + ... \\
&\quad + y_{n-m+2} - c\hat{x}_{n-m+2|n-m} - d + y_{n-m+1} - c\hat{x}_{n-m+1|n-m} - d]^2 \\
&= \mathbb{E}[y_n - cx_n + cx_n - c\hat{x}_{n|n-m} - d + y_{n-1} - cx_{n-1} + cx_{n-1} - c\hat{x}_{n-1|n-m} - d + ... \\
&\quad + y_{n-m+2} - cx_{n-m+2} + cx_{n-m+2} - c\hat{x}_{n-m+2|n-m} - d + \\
&\quad y_{n-m+1} - cx_{n-m+1} + cx_{n-m+1} - c\hat{x}_{n-m+1|n-m} - d]^2 \\
&= \mathbb{E}[ca(x_{n-1} - \hat{x}_{n-1|n-m}) + cw_n + u_n + ca(x_{n-2} - \hat{x}_{n-2|n-m}) + \\
&\quad cw_{n-1} + u_{n-1} + ... + ca(x_{n-m+1|n-m} - \hat{x}_{n-m+1|n-m}) + cw_{n-m+2} + u_{n-m+2} + \\
&\quad c(x_{n+1} - \hat{x}_{n+1|n}) + u_{n-m+1}]^2 \\
&= ... \\
&= \mathbb{E}[ca^{m-1}(x_{n-m+1} - \hat{x}_{n-m+1|n-m}) + ca^{m-2}w_{n-m+2} + \\
&\quad ca^{m-3}w_{n-m+3} + ... + caw_{n-1} + cw_n + ca^{m-2}(x_{n-m+1} - \hat{x}_{n-m+1|n-m}) + \\
&\quad ca^{m-3}w_{n-m+2} + ca^{m-4}w_{n-m+3} + ... + caw_{n-2} + cw_{n-1} + ... + \\
&\quad ca(x_{n-m+1} - \hat{x}_{n-m+1|n-m}) + cw_{n-m+2} + c(x_{n-m+1} - \hat{x}_{n-m+1|n-m}) + \\
&\quad u_n + u_{n-1} + ... + u_{n-m+1}]^2 \\
&= \mathbb{E}[c(a^{m-1} + a^{m-2} + ... + a + 1)(x_{n-m+1} - \hat{x}_{n-m+1|n-m}) + \\
&\quad c(a^{m-2} + a^{m-3} + ... + a + 1)w_{n-m+2} + c(a^{m-3} + a^{m-4} + ... + a + 1)w_{n-m+3} + ... \\
&\quad + c(a+1)w_{n-1} + cw_n + u_n + u_{n-1} + ... + u_{n-m+1}]^2 \\
&= c^2(a^{m-1} + a^{m-2} + ... + a + 1)^2 P_{n+1|n} + \\
&\quad c^2(a^{m-2} + a^{m-3} + ... + a + 1)^2 q^2 + c^2(a^{m-3} + a^{m-4} + ... + a + 1)^2 q^2 + ... \\
&\quad + c^2(a+1)^2 q^2 + q^2 + + mr^2 \\
&= \left(\sum_{j=0}^{m-1} a^j\right)^2 c^2 P_{n-m+1|n-m} + \sum_{i=0}^{m-2}\left(\sum_{j=0}^{i} a^j\right)^2 c^2 q^2 + mr^2 \qquad \blacksquare \qquad (10)
\end{aligned}
$$

Bibliography

Ait-Sahalia, Y., P. A. Mykland and L. Zhang (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* **18**(2), 351–416.

Albanese, C. and A. Kuznetsov (2003). Unifying the three stochastic volatility models. *Risk Magazine*.

Alizadeh, S., M. W. Brandt and F. X. Diebold (2002). Range-based estimation of stochastic volatility models. *Journal of Finance* **57**(3), 1047–1089.

Andersen, T. G. (2000). Some reflections on analysis of high-frequency data. *Journal of Business and Economic Statistics* **18**(2), 146.

Andersen, T. G. and T. Bollerslev (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* **4**(2-3), 115–158.

Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* **39**(4), 885–905.

Andersen, T. G., T. Bollerslev and S. Lange (1999). Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance* **6**(5), 457–477.

Andersen, T. G., T. Bollerslev, F. X. Diebold and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics* **61**(1), 43–76.

Anderson, B. D. O. and J. B. Moore (1979). *Optimal filtering*. Prentice-Hall. Englewood Cliffs; London.

Andersson, P. J., A. Oberg and T. Guhr (2005). Power mapping and noise reduction for financial correlations. *Acta Physica Polonica B* **36**(9), 2611–2619.

Andreou, E. and E. Ghysels (2002). Rolling-sample volatility estimators: Some new theoretical, simulation, and empirical results. *Journal of Business and Economic Statistics* **20**(3), 363.

Bachelier, L. (1900). Theorie de la Speculation. *Annales Scientifiques de l'École Normale Supérieure Sér. 3* **17**, 21–86.

Bai, X. Z., J. R. Russell and G. C. Tiao (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics* **114**(2), 349–360.

Bakshi, Gurdip, Charles Cao and Zhiwu Chen (1997). Empirical performance of alternative option pricing models. *Journal of Finance* **52**(5), 2003–2049.

Bandi, F. M. and J. R. Russell (2004). Separating microstructure noise from volatility. *Journal of Financial Economics*.

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **63**, 167–207.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **64**, 253–280.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of Financial Studies* **9**(1), 69–107.

Bjork, T. (2004). *Arbitrage theory in continuous time*. Oxford University Press. Oxford.

Black, F. and M. Scholes (1973). Pricing of options and corporate liabilities. *Journal of Political Economy* **81**(3), 637–654.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3), 307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economic Statistics* **69**(3), 542–547. GR: Notes.

Bollerslev, T. and H. Zhou (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics* **109**(1), 33–65.

Bollerslev, T. and J. M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* **11**(2), 143–172.

Bollerslev, T., R. F. Engle and J. M. Wooldridge (1988). A capital-asset pricing model with time-varying covariances. *Journal of Political Economy* **96**(1), 116–131.

Carr, P. and L. R. Wu (2004). Time-changed levy processes and option pricing. *Journal of Financial Economics* **71**(1), 113–141.

Carr, Peter, Helyette Geman, Dilip B. Madan and Marc Yor (2003). Stochastic volatility for lvy processes. *Mathematical Finance* **13**(3), 345–382.

Chib, S., F. Nardari and N. Shephard (2002). Markov chain monte carlo methods for stochastic volatility models. *Journal of Econometrics* **108**(2), 281–316.

Cont, Rama and Peter Tankov (2004). *Financial modelling with jump processes*. Chapman & Hall/CRC. Boca Raton, Fla.

Cox, J. C., J. E. Ingersoll and S. A. Ross (1985*a*). An Intertemporal General Equilibrium Model of Asset Prices. *Econometrica* **53**(2), 363–384.

Cox, J. C., J. E. Ingersoll and S. A. Ross (1985*b*). A theory of the term structure of interest-rates. *Econometrica* **53**(2), 385–407.

Cox, J.C., S.A. Ross and M. Rubinstein (1979). Option Pricing: A Simplified Approach. *Journal of Financial Economics* **7**, 229–263.

Drost, F. C. and T. E. Nijman (1993). Temporal aggregation of GARCH processes. *Econometrica: Journal of the Econometric Society* **61**(4), 909–927.

Duan, J.C. (1995). The GARCH option pricing model. *Mathematical Finance* **5**, 13–32.

Duffie, D. (1986). Stochastic Equilibria: Existence, Spanning Number, and the "No Expected Financial Gain from Trade" Hypothesis. *Econometrica* **54**(5), 1161–1184.

Duffie, D. and C.F. Huang (1985). Implementing Arrow-Debreu Equilibria by Continuous Trading of Few Long-Lived Securities. *Econometrica* **53**(6), 1337–1356.

Elliott, R. J., W. P. Malcolm and A. H. Tsoi (2003). Robust parameter estimation for asset price models with markov modulated volatilities. *Journal of Economic Dynamics and Control* **27**(8), 1391–1409.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* **50**(4), 987–1007.

Engle, R. F. and K. Kroner (1995). Multivariate simulataneous GARCH. *Econometric Theory* **11**, 122–150.

Fama, E.F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance* **25**(2), 383–417.

Fisher, R.A. (1912). On an absolute criterion for fitting frequency curves. 41 155 160. *Messenger of Mathematics* **1**, 155–160.

Fleming, J. (2003). A closer look at the relation between GARCH and stochastic autoregressive volatility. *Journal of Financial Econometrics* **1**(3), 365–419.

Fouque, J. P., G. Papanicolaou and K. R. Sircar (2000). *Derivatives in financial markets with stochastic volatility*. Cambridge University Press. Cambridge.

Geman, Helyette, Dilip B. Madan and Marc Yor (2001). Time changes for lvy processes. *Mathematical Finance* **11**(1), 79–96.

Hamilton, J. D. and R. Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* **64**(1-2), 307–333.

Harrison, J.M. and S.R. Pliska (1981). Martingales and Stochastic Integrals in the Theory of Continous Trading. *Stochastic Processes and their Applications* **11**, 215–260.

Harvey, A. C. (1981). *The Econometric Analysis of Time-series*. Philip Allan. Oxford.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press. Cambridge.

Heath, D., R. Jarrow and A. Morton (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica* **60**(1), 77–105.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6**(2), 327–343.

Heston, Steven L. and Saikat Nandi (2000). A closed-form GARCH option valuation model. *Review of Financial Studies* **13**(3), 585–625.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance* **42**(2), 281–300.

Johnston, F. R. and P. J. Harrison (1986). The variance of lead-time demand. *Journal of the Operational Research Society* **37**(3), 303–308.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions AMSE* **Series D 82**, 35–45.

Kalman, R. E. and R. S. Bucy (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering, Transactions AMSE* **Series D 83**, 95–108.

Kendall, M.G. (1953). The Analysis of Time Series, Part I: Prices. *Journal of the Royal Statistical Society* **96**, 11–25.

Koopman, S. J. and E. Hol (2002). Stock index volatility forecasting with high frequency data. Working paper available at http://www.tinbergen.nl/discussionpapers/02068.pdf.

Linter, J (1965). Security prices, risk and maximal gains from diversification. *Journal of Finance* **20**, 585–615.

Ljung, L (1987). *System identification : theory for the user*. Prentice-Hall. Englewood Cliffs; London.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Journal of Business* **36**, 394–419.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance* **7**(1), 77–91.

Meddahi, N. and E. Renault (1997). *Aggregations and Marginalization of GARCH and Stochastic Volatility Models*. Université de Montréal, Centre de recherche et développement en économique.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* **3**(1-2), 125–144.

Merton, R.C. (1973). Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science* **4**(1), 141–183.

Mikosch, T. (1998). *Elementary stochastic calculus with finance in view*. Vol. 6. World Scientific.

Muller, U.A., M.M. Dacorogna, R.D. Dave, O.V. Pictet, R. B. Olsen and J.R. Ward (1993). Fractals and intrinsic time - a challenge to econometricians. Olsen, internal document.

Naik, V. (1993). Option valuation and hedging strategies with jumps in the volatility of asset returns. *Journal of Finance* **48**(5), 1969–1984.

Nelson, D. B. (1990). ARCH models as diffusion approximations. *Journal of Econometrics* **45**(1-2), 7–38.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns - a new approach. *Econometrica* **59**(2), 347–370.

Nelson, D. B. (1992). Filtering and forecasting with misspecified ARCH models 1. getting the right variance with the wrong model. *Journal of Econometrics* **52**(1-2), 61–90.

Oomen, R. (2002). Modelling realized variance when returns are serially correlated. Working paper http://www.warwick.ac.uk/staff/R.C.A.Oomen/.

Owens, J. P. and D. G. Steigerwald (2006). *Noise reduced realized volatility: a Kalman Filter approach*. Vol. Volume 20, Part 1 of *Advances in Econometrics*. JAI.

Plerou, V., P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr and H. E. Stanley (2002). Random matrix approach to cross correlations in financial data. *Physical Review E*.

Ross, S.A. (1976). The Arbitrage Pricing Theory of Capital Asset Pricing. *Journal of Economic Theory* **13**(3), 341–360.

Samuelson, P.A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review* **6**(2), 41–49.

Scalas, E., R. Gorenflo and F. Mainardi (2000). Fractional calculus and continuous-time finance. *Physica A* **284**(1-4), 376–384.

Scott, Louis O. (1987). Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative Analysis* **22**(4), 419–438.

Sharpe, W.F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance* **19**(3), 425–442.

Skiadopoulos, George, Stewart Hodges and Les Clewlow (2000). The dynamics of the s&p 500 implied volatility surface. *Review of Derivatives Research* **3**(3), 263–282.

Turner, C. M., R. Startz and C. R. Nelson (1989). A Markov Model of heteroskedasticity, risk, and learning in the stock-market. *Journal of Financial Economics* **25**(1), 3–22.

Watsham, T.J. (1998). *Futures and options in risk management.* International Thomson Business Press Boston.

White, H. (1982). Maximum-likelihood estimation of mis-specified models. *Econometrica* **50**(1), 1–25.

Woerner, J. H. C. (2003). Purely discontinuous levy processes and power variation: inference for integrated volatility and the scale parameter.. Working Paper Series in Mathematical Finance, University of Oxford, 2003. Available at http://www.maths.ox.ac.uk/mfg/mfg/mfghome.htm.

Woerner, J. H. C. (2005). Estimation of integrated volatility in stochastic volatility models. *Applied Stochastic Models in Business and Industry* **21**(1), 27–44.

Zumbach, G., F. Corsi and A. Trapetti (2002). Effcient estimation of volatility using high frequency data. Working paper, available at http://www.olsen.ch/research.