

Forecasting the Variability of Stock Index Returns with Stochastic Volatility Models and Implied Volatility

Eugenie Hol and Siem Jan Koopman*

Department of Accounting and Finance, University of Birmingham

Department of Econometrics, Free University Amsterdam

November 13, 2000

Abstract

In this paper we compare the predictive ability of Stochastic Volatility (SV) models to that of volatility forecasts implied by option prices. We develop an SV model with implied volatility as an exogenous variable in the variance equation which facilitates the use of statistical tests for nested models; we refer to this model as the SVX model. The SVX model is then extended to a volatility model with persistence adjustment term and this we call the SVX⁺ model. This class of SV models can be estimated by quasi maximum likelihood methods but the main emphasis will be on methods for exact maximum likelihood using Monte Carlo importance sampling methods. The performance of the models is evaluated, both within sample and out-of-sample, for daily returns on the Standard & Poor's 100 index. Similar studies have been undertaken with GARCH models where findings were initially mixed but recent research has indicated that implied volatility provides superior forecasts. We find that implied volatility outperforms historical returns in-sample but that the latter contains incremental information in the form of stochastic shocks incorporated in the SVX models. The out-of-sample volatility forecasts are evaluated against daily squared returns and intradaily squared returns for forecasting horizons ranging from 1 to 10 days. For the daily squared returns we obtain mixed results, but when we use intradaily squared returns as a measure of realised volatility we find that the SVX⁺ model produces the most accurate out-of-sample volatility forecasts and that the model that only utilises implied volatility performs the worst as its volatility forecasts are upwardly biased.

KEYWORDS: Forecasting, Implied Volatility, Monte Carlo likelihood method, Stochastic volatility, Stock indices.

1 Introduction

Forecasts of financial market volatility play a crucial role in financial decision making and the need for accurate forecasts is apparent in a number of areas, such as option pricing, hedging strategies, portfolio allocation and Value-at-Risk calculations. Unfortunately, it is notoriously difficult to accurately predict volatility and the problem is exacerbated by the fact that realised volatility has to be approximated as it is inherently unobservable. Due to its critical role the topic of volatility forecasting has however received much attention and the resulting literature is considerable.

One of the main sources of volatility forecasts are historical parameteric volatility models such as Generalised Autoregressive Conditional Heteroscedasticity (GARCH) and Stochastic Volatility (SV) models. The parameters in these models are estimated with historical data and subsequently used

*Corresponding author: Siem Jan Koopman, Department of Econometrics, Free University, De Boelelaan 1105, NL-1081 HV Amsterdam. Email s.j.koopman@econ.vu.nl.

to construct out-of-sample volatility forecasts. The high degree of intertemporal volatility persistence observed by these models suggests that the variability of stock index returns is highly predictable and that past observations contain valuable information for the prediction of future volatility. Studies comparing the forecasting abilities of the various volatility models have been undertaken for a number of stock indices and the general consensus appears to be that those models that attribute more weight to recent observations outperform others¹. Little effort has however been made to compare ex-ante volatility forecasts produced by GARCH models with those of SV models².

An alternative information source for volatility prediction is found in implied volatility which is calculated from option prices in combination with a certain option pricing model. Early empirical studies by Latané and Rendelman (1976), Chiras and Manaster (1978) and Beckers (1981) have indicated that implied volatility, when compared with historical standard deviations, can be regarded as a good predictor of future volatility. Implied volatility is also often referred to as the market's volatility forecast and is forward looking, as opposed to historical based methods which are by definition backward looking. Provided that the option market is efficient and that the option pricing model has been correctly specified, the information content of implied volatility should therefore subsume that of all other variables in the information set.

The question whether the most accurate volatility forecasts are produced by implied volatility, rather than by the historically based volatility models, was first addressed by Day and Lewis (1992) who developed a GARCH model with embedded implied volatility. Contrary to theory, their results indicated that GARCH models provided better volatility forecasts than implied volatility but that the latter might contain additional information as the best forecasts were obtained using both sources of information. Canina and Figlewski (1993) even found "*little or no correlation at all between implied volatility and subsequent realized volatility*" and favoured a simple historical volatility measure. Findings in the early nineties were therefore mixed and the assumed comprehensive information content of implied volatility appeared questionable as Lamoureux and Lastrapes (1993) were also unable to reject the hypothesis that predictions based on GARCH models contained incremental information about future volatility. Recent studies by Christensen and Prabhala (1998), Fleming (1998) and Blair, Poon and Taylor (2000) are however much more supportive as all present evidence that the most accurate volatility forecasts for returns on the Standard & Poor's 100 stock index are based on implied volatility. Moreover, their research strongly suggests that historical data contains little or no incremental information about future volatility³.

Thusfar the issue of comparative forecasting ability has however not been studied in the context of SV models. In recent years this class of volatility model has received considerable attention in the literature and it can now be regarded as a competitive alternative to GARCH models eventhough its empirical application has been limited. In this paper we examine the predictive ability of the SV model and compare its volatility forecasts with those of implied volatility. For this purpose we introduce an SV model which incorporates implied volatility as an exogeneous variable in the variance equation. This model, which we will refer to as the Stochastic Volatility with exogeneous variables (SVX) model, allows us to perform statistical tests for nested models. We evaluate the predictive performance for daily returns on the Standard & Poor's 100 index and as a measure of implied volatility we use the VIX index of the Chicago Board Options Exchange (CBOE). In addition, we compare the ex-ante forecasting ability of the different methods over a five year evaluation period for forecasting horizons ranging from 1 to 10 trading days. As measures of realised volatility we consider both daily squared returns and intradaily squared returns.

The SV class of models considered in this paper are estimated using exact maximum likelihood

¹See *e.g.* Akgiray (1989), Dimson and Marsh (1990) and Walsh and Tsou (1998) for an overview.

²An exception is Heynen (1995) who examined a variety of international stock indices and found that SV models produced more accurate volatility forecasts than GARCH models.

³It has been suggested, most notably by Blair *et.al* (2000), that the earlier findings are due to measurement errors in the calculated implied volatility measure.

methods which are based on Monte Carlo simulation techniques such as importance sampling and antithetics. More accurate estimates of the likelihood function are obtained when the number of simulations is increased. Therefore the estimates can be as accurate as desired at the cost of computer time.

The remainder of this paper is structured as follows. In the next section we discuss the various model specifications while in section 3 we present the relevant estimation methods. The data and in-sample estimation results are presented in section 4. In section 5 we give details of our forecasting methodology and the out-of-sample forecasting results are presented in section 6. In the final section we conclude and provide a summary.

2 Model Specifications

Generalised Autoregressive Conditional Heteroscedasticity (GARCH) models have thusfar been the most frequently applied class of time-varying volatility model. Since its introduction by Engle (1982) and subsequent generalisation by Bollerslev (1986) this model has been extended in numerous ways which usually involved alternative formulations for the volatility process⁴. Although the Stochastic Volatility (SV) model has been recognised as a viable alternative to the GARCH model, the latter is still the standard in empirical applications⁵. This is mainly due to the problems which arise as a consequence of the intractability of the likelihood function of the SV model which prohibits its direct evaluation. However, in recent years considerable progress has been made in this area which does not only encourage further empirical research but also enables the development of various extensions of the SV model. One of the possible extensions involves the inclusion of exogenous variables in the variance equation which we will discuss in this paper; the resulting model we refer to as the SVX model.

Volatility models are usually defined by their first two moments, the mean and the variance equation. The general notation for the mean equation of time-varying volatility models is given by

$$y_t = \mu_t + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, 1), \quad t = 1, \dots, T, \quad (1)$$

where y_t denotes the return series of interest and μ_t its conditional mean⁶. The disturbance term ε_t is assumed to be identically and independently distributed with zero mean and unit variance. In addition, the assumption of normality is added. A common notation for the variance equation of the SV class of volatility models is given by

$$\sigma_t^2 = \sigma^{*2} \exp(h_t), \quad (2)$$

and it is therefore defined as the product of a positive scaling factor σ^{*2} and the exponential of the stochastic process h_t . For the standard SV model this process is specified as

$$h_t = \phi h_{t-1} + \sigma_\eta \eta_t, \quad \eta_t \sim \text{NID}(0, 1), \quad (3)$$

where the degree of volatility persistence is measured by the ϕ parameter which is restricted to a positive value smaller than one in order to ensure the stationarity of the volatility process, so $0 < \phi < 1$. Further, it is assumed that the disturbance term η_t is mutually uncorrelated with the error term ε_t

⁴For surveys on GARCH models we refer to Bollerslev, Chou and Kroner (1992), Bera and Higgins (1993), Bollerslev, Engle and Nelson (1994) and Diebold and Lopez (1995).

⁵SV models are reviewed in, for example, Taylor (1994), Ghysels, Harvey and Renault (1996) and Shephard (1996).

⁶For SV models the conditional mean is usually assumed to be equal to zero or is modelled prior to estimation of the volatility process. Simultaneous estimation of the mean and variance equation has been undertaken in, for example, Koopman and Hol Uspensky (2000).

in the mean equation (1), both contemporaneously and at all lags. The SV model with embedded implied volatility is labelled the SVX model and we could specify its stochastic process as

$$h_t = \phi h_{t-1} + \gamma x_t + \sigma_\eta \eta_t, \quad \eta_t \sim \text{NID}(0, 1), \quad (4)$$

where x_t denotes the contemporaneous implied volatility measure in logarithmic squared form, so $x_t = \ln \sigma_{IV,t}^2$. The value for ϕ in the SVX model is restricted to be less than one in absolute values, *i.e.* $-1 < \phi < 1$. The problem with this specification is that it includes an entire lag structure of the implied volatility measure which becomes apparent when we rewrite the volatility process in logarithmic terms as

$$\begin{aligned} \ln \sigma_t^2 &= \ln \sigma^{*2} + h_t \\ &= \ln \sigma^{*2} + \phi h_{t-1} + \gamma x_t + \sigma_\eta \eta_t \\ &= (1 - \phi) \ln \sigma^{*2} + \phi \ln \sigma_{t-1}^2 + \gamma x_t + \sigma_\eta \eta_t, \end{aligned}$$

and if we repeatedly substitute for the lagged volatility process we observe that

$$\ln \sigma_t^2 = \ln \sigma^{*2} + \gamma x_t + \gamma \sum_{i=1}^{t-1} \phi^i x_{t-i} + \sigma_\eta \sum_{i=0}^{t-1} \phi^i \eta_{t-i}.$$

In comparison, the equivalent notation for the SV model, with h_t as defined in equation (3), can be written as

$$\ln \sigma_t^2 = \ln \sigma^{*2} + \sigma_\eta \sum_{i=0}^{t-1} \phi^i \eta_{t-i}.$$

Inclusion of these multiple lagged implied volatility measures lead to a downwardly biased value for γ when ϕ is positive. For a negative ϕ parameter the estimate for γ will be asymptotically upwardly biased, as $\sum_{i=1}^{\infty} \phi^i < 0$ for $-1 < \phi < 0$. Obviously, the size of this bias depends on the estimated value for the persistence parameter ϕ . If ϕ is close to zero and insignificant, *i.e.* if all volatility information is impounded in the implied volatility measure, γ will only be marginally biased. For the GARCH class of models the issue of a multiple lagged implied volatility structure was pointed out by Amin and Ng (1997), who suggested a persistence adjustment term. A similar structure can be implemented for the SVX model by defining h_t as follows

$$h_t = \phi h_{t-1} + \gamma(1 - \phi L)x_t + \sigma_\eta \eta_t, \quad \eta_t \sim \text{NID}(0, 1), \quad (5)$$

which by recursive substitution of the logarithmic variance equation leads to

$$\begin{aligned} \ln \sigma_t^2 &= (1 - \phi) \ln \sigma^{*2} + \phi \ln \sigma_{t-1}^2 + \gamma(1 - \phi L)x_t + \sigma_\eta \eta_t, \\ &= \ln \sigma^{*2} + \gamma x_t + \sigma_\eta \sum_{i=0}^{t-1} \phi^i \eta_{t-i}, \end{aligned}$$

and therefore omits the implied volatility lag structure. By defining h_t as in equation (5) we therefore obtain an alternative SVX model, which we will refer to hereafter as the SVX model with persistence adjustment, or the SVX⁺ model.

Finally, we also consider a deterministic volatility model that only utilises implied volatility as a source of volatility information. This model we obtain by imposing the restrictions $\phi = 0$ and $\sigma_\eta^2 = 0$ on equation (4) or (5), and therefore

$$h_t = \gamma x_t, \quad (6)$$

with

$$\ln \sigma_t^2 = \ln \sigma^{*2} + \gamma x_t.$$

This last of our four models we term the VX model as the volatility process does not have a separate error term and is solely determined by exogenous variables.

3 Model Estimation

In this section we show how the parameters of the SVX class of models can be estimated by simulated maximum likelihood using importance sampling. Further, we show how to compute the conditional mean and variance of the volatility process h_t . First we show that a quasi-maximum likelihood method can also be used.

3.1 Quasi-maximum likelihood of SV and SVX models

Consider model (1), (2) and (3) which we can transform by taking logs of squared y_t 's, that is $y_t^* = \ln y_t^2$, to obtain the model

$$y_t^* = \gamma^* + h_t + u_t, \quad \gamma^* = \ln \sigma^{*2}, \quad u_t \sim \ln \chi_1^2,$$

with h_t given by (3) for $t = 1, \dots, T$. We take μ_t as zero, implying that the y_t process remains unmodelled. The resulting model is within the class of linear state space models; for an introduction to state space models we refer to Harvey (1993, Chapter 4). Note that the disturbance term of the model for y_t^* is non-Gaussian. Nevertheless, the Kalman filter and the associated smoothing algorithm produce the minimum mean square linear estimator for h_t ; see Harvey (1993, section 4.3). By assuming that the disturbance u_t is normally distributed with mean and variance set equal to the mean and variance of a $\ln \chi_1^2$ variable, we obtain so-called *quasi maximum likelihood* estimates of the unknown parameters ϕ , σ_η and γ^* when maximising the Gaussian likelihood function with respect to these parameters. This estimation procedure is proposed by Harvey, Ruiz and Shephard (1994) and is implemented in the computer package STAMP of Koopman *et.al* (2000).

The inclusion of explanatory variables in the log-volatility process does not complicate matters further. The log-squared transformation does not affect the log-volatility processes of the SVX models as defined in equations (4) and (5). Therefore the Kalman filter can still be applied to the resulting linear model. However the regression coefficient γ need to be estimated additionally.

In the following we will develop exact maximum likelihood methods for the estimation of the parameters of SV and SVX models. Quasi-maximum likelihood methods will not be used in the empirical studies of sections 4 and 6 since we prefer to use exact likelihood methods.

3.2 Exact maximum likelihood of SVX models using importance sampling

Let $y = (y_1, \dots, y_T)'$ and $\theta = (\theta_1, \dots, \theta_T)'$ where observation y_t is modelled as (1) and its log-volatility is given by

$$\sigma_t^2 = \exp(\theta_t), \quad \theta_t = \gamma^* + h_t,$$

with signal h_t modelled as (3), (4) or (5), for $t = 1, \dots, T$. Note that $\sigma^{*2} = \exp(\gamma^*)$. Further we shall collect the parameters which are not included in the state vector below in the parameter vector ψ . The SVX model (1), (2) and (4) in state space form is given by

$$p(y|\theta, \psi) = \prod_{t=1}^T N(0, \sigma_t^2),$$

with $\mu_t = 0$, $\sigma_t^2 = \exp(\theta_t) = \exp(\gamma^* + h_t) = \sigma^{*2} \exp(h_t)$. The state vector collects the components of the log-volatility and is given by $\alpha_t = (\gamma^*, \gamma, h_t)'$. The so-called transition equation for the state vector is given by

$$\alpha_{t+1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & x_t & \phi \end{bmatrix} \alpha_t + \begin{pmatrix} 0 \\ 0 \\ \sigma_\eta \end{pmatrix} \eta_t,$$

where disturbances η_t are distributed as $\text{NID}(0, 1)$. The initial state vector α_1 is given by

$$\begin{pmatrix} \gamma^* \\ \gamma \\ h_1 \end{pmatrix} \sim \text{N} \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \kappa & 0 & 0 \\ 0 & \kappa & 0 \\ 0 & 0 & \sigma_\eta^2/(1 - \phi^2) \end{bmatrix} \right\},$$

for some arbitrary large value for κ . Finally, the parameter vector ψ is given by

$$\psi = \begin{pmatrix} \phi \\ \sigma_\eta \\ \sigma_\epsilon \end{pmatrix}.$$

This completes the model specification in state space form. It follows that

$$p(y|\theta, \psi) = \prod_{t=1}^T \text{N}(0, \exp \theta_t), \quad \theta_t = (1, 0, 1)\alpha_t.$$

This representation of SV and SVX models can be regarded as a nonlinear state space model.

The aim now is to estimate the parameter vector ψ by exact maximum likelihood. This requires the evaluation of the loglikelihood function. A convenient expression for the loglikelihood is developed below. In section 3.2.2 we provide some computational details required for estimation. The state vector α_t elements, which include the regression coefficients γ^* and γ and the stochastic log-volatility process h_t , are estimated using signal extraction methods which are briefly discussed in section 3.2.3.

3.2.1 Loglikelihood evaluation

The construction of the exact likelihood for the SV model using the Monte Carlo likelihood approach of Shephard and Pitt (1997) and Durbin and Koopman (1997) can be modified for the SVX model. The nonlinear relation between the log-volatility h_t and the observation equation is not altered in the SVX case; only the specification for h_t is different. Similar considerations are discussed by Chib, Nardari and Shephard (1998) in a Bayesian context. The same modification can be used for the SVX⁺ model since we merely replace the explanatory variable x_t by $(1 - \phi L)x_t$.

The loglikelihood function for the SVX model can be computed via the Monte Carlo technique of importance sampling. The likelihood function can be expressed as

$$L(\psi) = p(y|\psi) = \int p(y, \theta|\psi) d\theta = \int p(y|\theta, \psi) p(\theta|\psi) d\theta. \quad (7)$$

An efficient way of evaluating the likelihood is by using importance sampling; see Ripley (1987, Chapter 5). We require a simulation device to sample from an importance density $\tilde{p}(\theta|y, \psi)$ which we prefer to be as close as possible to the true density $p(\theta|y, \psi)$. An obvious choice for the importance density is the conditional Gaussian density since in this case it is relatively straightforward to sample from $\tilde{p}(\theta|y, \psi) = g(\theta|y, \psi)$. An approximating Gaussian model for the SVX model is developed in the appendix. The simulation smoother of de Jong and Shephard (1995) is used to sample from the approximating Gaussian model $g(\theta|y, \psi)$.

The likelihood function (7) can be obtained by writing

$$L(\psi) = \int p(y|\theta, \psi) \frac{p(\theta|\psi)}{g(\theta|y, \psi)} g(\theta|y, \psi) d\theta = \tilde{\text{E}} \left\{ p(y|\theta, \psi) \frac{p(\theta|\psi)}{g(\theta|y, \psi)} \right\}, \quad (8)$$

where $\tilde{\text{E}}$ denotes expectation with respect to the importance density $g(\theta|y, \psi)$. Expression (8) can be simplified following a suggestion of Durbin and Koopman (1997). The likelihood function of the approximating Gaussian model is given by

$$L_g(\psi) = g(y|\psi) = \frac{g(y, \theta|\psi)}{g(\theta|y, \psi)} = \frac{g(y|\theta, \psi) p(\theta|\psi)}{g(\theta|y, \psi)}, \quad (9)$$

and it follows that

$$\frac{p(\theta|\psi)}{g(\theta|y, \psi)} = \frac{L_g(\psi)}{g(y|\theta, \psi)}.$$

This ratio also appears in (8) and substitution leads to

$$L(\psi) = L_g(\psi) \tilde{\mathbb{E}}\left\{\frac{p(y|\theta, \psi)}{g(y|\theta, \psi)}\right\}, \quad (10)$$

which is the convenient expression we will use in our calculations. The likelihood function of the approximating Gaussian model $L_g(\psi)$ can be calculated via the Kalman filter. The conditional density functions $p(y|\theta, \psi)$ and $g(y|\theta, \psi)$ are easily computed given values for θ and ψ using (26). It follows that the likelihood function of the SVX model is equivalent to the likelihood function of an approximating Gaussian model, multiplied by a correction term. This correction term only needs to be evaluated via simulation.

An obvious estimator for the likelihood of the SVX model is

$$\hat{L}(\psi) = L_g(\psi)\bar{w}, \quad (11)$$

where

$$\bar{w} = \frac{1}{M} \sum_{i=1}^M w_i, \quad w_i = \frac{p(y|\theta^i, \psi)}{g(y|\theta^i, \psi)}, \quad (12)$$

and θ^i denotes a draw from the importance density $g(\theta|y, \psi)$. The accuracy of this estimator solely depends on M , that is the number of simulation samples. In practice, we usually work with the log of the likelihood function to manage the magnitude of density values. The log transformation of $\hat{L}(\psi)$ introduces bias for which we can correct up to order $O(M^{-3/2})$; see Shephard and Pitt (1997) and Durbin and Koopman (1997). We obtain

$$\ln \hat{L}(\psi) = \ln L_g(\psi) + \ln \bar{w} + \frac{s_w^2}{2M\bar{w}^2}, \quad (13)$$

with $s_w^2 = (M-1)^{-1} \sum_{i=1}^M (w_i - \bar{w})^2$.

3.2.2 Computational details

Given a particular vector for $\psi = (\phi, \sigma_\eta, \sigma_\varepsilon)'$, we evaluate the loglikelihood function (13) for which we use the approximating model (25) to generate simulation samples. To obtain a maximum likelihood estimate for ψ , which we denote by $\hat{\psi}$, the loglikelihood is numerically maximised with respect to ψ in a similar fashion as for Gaussian models; see Harvey (1989) and Koopman *et.al* (2000). The repeated evaluation of the loglikelihood for different ψ 's during the search for $\hat{\psi}$ will be based on the same set of random numbers used for simulation.

Although the approximating model is effective for simulation, we may wish to decrease the simulation variance further using standard simulation techniques based on antithetics and control variables; see Durbin and Koopman (1997). In our computations we have employed the standard antithetic variable as given by $\check{\theta}^i = 2\hat{\theta} - \theta^i$ where θ^i is a draw from the importance density $g(\theta|y, \psi)$ and where $\hat{\theta} = \tilde{\mathbb{E}}(\theta)$ can be obtained using the Kalman filter and smoother. Since $\check{\theta}^i - \hat{\theta} = -(\theta^i - \hat{\theta})$ and θ^i are normally distributed, the two vectors θ^i and $\check{\theta}^i$ are equi-probable.

The number of simulation samples M is set prior to the estimation procedure. The choice of M can be determined by computing the error variance due to simulation; see Durbin and Koopman (1997). It is shown by Sandmann and Koopman (1998) that M can be relatively small in the context of SV models. Therefore, in this study we have set M equal to 100 times two antithetic variables, that is $M = 200$.

3.2.3 Signal extraction

The Monte Carlo importance sampling techniques, which we have used for likelihood evaluation, can also be employed to compute the conditional mean and variance of the unobserved signal θ_t . The same approximating Gaussian model can be used for this purpose. The details are given by Durbin and Koopman (2000).

The conditional mean and variance of the signal θ_t are given by

$$E(\theta_t|y, \psi) = \bar{\theta}_t^{(1)}, \quad \text{Var}(\theta_t|y, \psi) = \bar{\theta}_t^{(2)} - [\bar{\alpha}_t^{(1)}]^2.$$

where we compute $\bar{\theta}_t^{(1)}$ and $\bar{\theta}_t^{(2)}$ by

$$\bar{\theta}_t^{(1)} = \frac{1}{M} \sum_{i=1}^M w_i \theta_t^i, \quad \bar{\theta}_t^{(2)} = \frac{1}{M} \sum_{i=1}^M w_i \theta_t^{i2},$$

with w_i as defined in (12) and θ_t^i as the t th element of θ^i which is the i th draw from the importance density $g(\theta|y, \psi)$. This device can be generalised to obtain the elements of conditional mean and variance of the state vector α_t .

In practice, the unknown parameter vector ψ is replaced by its Monte Carlo maximum likelihood estimate $\hat{\psi}$. The uncertainty related to the estimate $\hat{\psi}$ can be also taken into account by similar Monte Carlo simulation techniques; see Durbin and Koopman (2000). An alternative approach of signal extraction for SV and SVX models is provided by the Bayesian Markov chain Monte Carlo techniques; see, for example, Shephard and Pitt (1997) and Kim, Shephard and Chib (1998).

3.2.4 Numerical implementation of estimation procedure

The simulated Monte Carlo estimation procedure is implemented using the object-oriented matrix programming language Ox 2.2 of Doornik (1998) using the library SsfPack 2.2 of Koopman, Shephard and Doornik (1999). The relevant programs, including the one used for the empirical studies in this paper, can be downloaded from the Internet at www.econ.vu.nl/koopman/sv/ and the program documentation can be consulted on-line. The programs can be used in a more general context; they can be modified for other Monte Carlo studies and be applied to other data-sets.

3.3 Exact maximum likelihood of VX model

The VX model can be regarded as a simple regression model with heteroscedastic disturbances, that is $\sigma_t^2 = \exp(\gamma^* + \gamma x_t)$. The model can be estimated by maximum likelihood using standard methods; for example, see Johnston and DiNardo (1997, Chapter 6). The first and second derivatives of the log-likelihood with respect to the parameters γ^* and γ can be obtained analytically for the VX model and this allows straightforward application of Newton's method for numerically maximising the likelihood function.

4 Data Description and Empirical In-Sample Results

4.1 Data

The data we selected is the Standard & Poor's 100 stock index for the period 2 January 1986 to 31 December 1999 and was obtained from Datastream. After adjusting the series for holidays, our sample consists of 3532 daily observations. The continuously compounded returns on the stock index are expressed in percentage terms and are therefore given by $R_t = 100(\ln P_t - \ln P_{t-1})$ where P_t denotes the closing price of the Standard & Poor's 100 index at time t . The accompanying implied

volatility index we use is the Chicago Board Options Exchange Market Volatility Index (VIX) which was extracted from the CBOE on-line database⁷. The VIX is calculated as the weighted average of implied volatilities of eight near-the-money, nearby and second nearby call and put options on the Standard & Poor's 100 index and represents the implied volatility of a hypothetical at-the-money OEX option with twenty two trading days to expiry⁸. We use the daily closing level of the VIX index and from the annualised VIX, which is expressed in terms of standard deviations, we calculate the daily VIX variance at time t as $\sigma_{IV,t}^2 = VIX_t^2/252$. The main attraction of the VIX index is that it mitigates many of the problems which lead to biased implied volatility values.

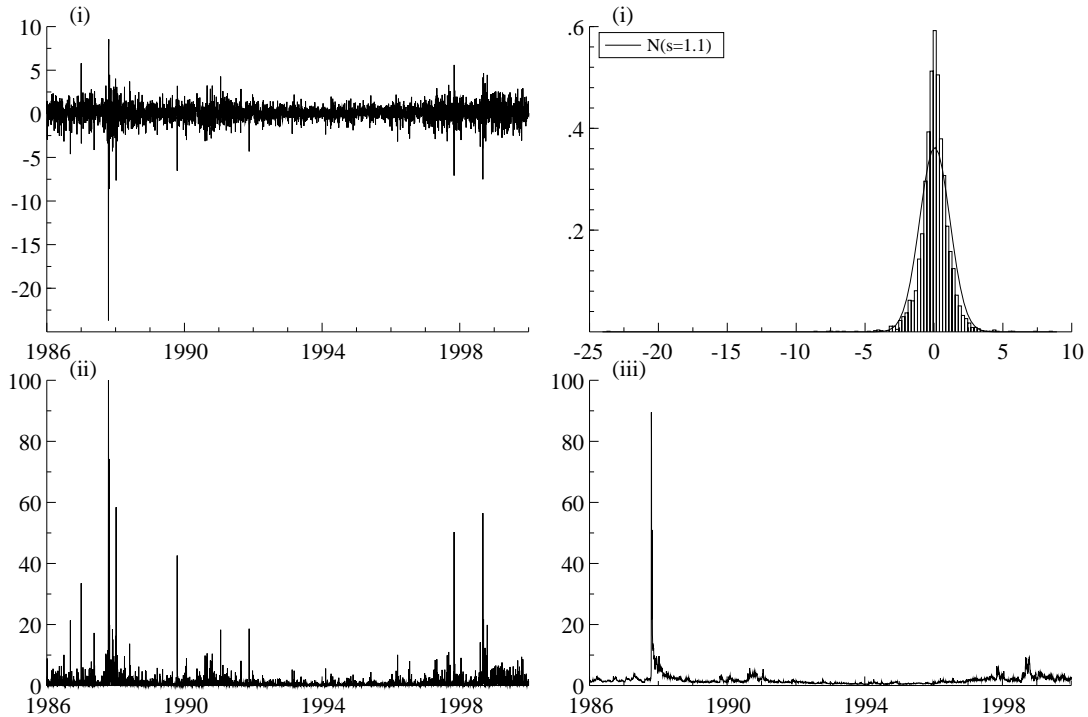


Figure 1: Daily (i) returns and (ii) squared returns (truncated at 100) on the Standard & Poor's 100 index and (iii) the VIX index between 02/01/86 and 31/12/99

In figure 1 we graph the daily return and the VIX series, together with the squared return series which can be regarded as an approximation of realised volatility⁹. In addition we show the histogram of the daily returns on the Standard & Poor's 100 index in the top right corner which shows that this series is negatively skewed and exhibits leptokurtosis. As our full sample includes observations relating to the October 1987 stock market crash, which might have a distorting influence on the estimation results, we also consider a subsample that starts on 1 January 1988. Summary statistics for both samples are given in table 1. The effects of the large outliers in the full sample are illustrated by the very high values for the skewness and excess kurtosis coefficients. When we compare the variances of the two return series we observe a value of 1.211 for the full sample against 0.905 for the shorter sample, which represent annual standard deviation values of 17.5% and 15.1%, respectively. The

⁷www.cboe.com/tools/historical/vix.htm

⁸The construction of the VIX index is described in detail by Whaley (1993) and by Fleming, Ostdiek and Whaley (1995) who regard it as a useful proxy for expected stock market volatility.

⁹Note that the graph of the squared return series is truncated at a value of 100 which only affects the observation of 19 October 1987 that has a value of 561.214.

Table 1: Summary statistics of daily returns and squared returns on the S&P 100 Index and the VIX index from 02/01/86 to 31/12/99 and from 04/01/88 to 31/12/99

Period	1986-1999			1988-1999		
No. of Obs. T	3531			3027		
Series	S&P100 R_t	R_t^2	VIX $\sigma_{IV,t}^2$	S&P100 R_t	R_t^2	VIX $\sigma_{IV,t}^2$
Mean	0.058	1.215	1.858	0.063	0.909	1.639
Variance	1.211	99.934	7.850	0.905	6.576	1.307
Skewness	-3.102	50.294	18.553	-0.471	12.791	2.226
Excess Kurtosis	66.847	2794.13	479.565	6.143	236.497	7.750
Maximum	8.539	561.214	89.521	5.606	58.438	9.668
Minimum	-23.690	0.000	0.324	-7.644	0.000	0.324
$\hat{\rho}_1$	-0.021	0.163	0.789	-0.019	0.195	0.960
$\hat{\rho}_2$	-0.053	0.138	0.623	-0.019	0.087	0.933
$\hat{\rho}_3$	-0.039	0.071	0.622	-0.048	0.051	0.914
$\hat{\rho}_4$	-0.042	0.026	0.598	-0.019	0.114	0.899
$\hat{\rho}_5$	0.024	0.128	0.568	-0.034	0.139	0.875
$Q(12)$	29.644	253.45	10760	34.147	341.68	27065

$\hat{\rho}_\ell$ is the sample autocorrelation coefficient at lag ℓ with asymptotic standard error $1/\sqrt{T}$ and $Q(\ell)$ is the Box-Ljung portmanteau statistic based on ℓ squared autocorrelations. The critical value at the 1% significance level for the $Q(12)$ statistic is 26.22.

decrease in volatility is also reflected in the VIX series which has however much larger mean values than the squared return series. When these VIX values are translated into annual standard deviations these amount to 21.6% for the full sample and 15.8% for the post crash period which indicates that the implied volatility measure tends to overestimate actual volatility. The graphs in figure 1 show that the VIX and the squared return series follow a very similar pattern although the VIX series is much smoother, *i.e.* less volatile, which is confirmed by their respective variance statistics in table 1. Further, the degree of autocorrelation is much higher for the VIX series than for the squared return series, especially for the 1988–1999 sample where autocorrelation coefficients for the VIX series are comparable with the persistence parameter values generally found for SV and GARCH models. The $Q(12)$ statistics for the return series indicates that the null hypothesis of zero values for the first twelve autocorrelation coefficients has to be rejected at the 1% level for both samples. The first order autocorrelation coefficients are however not significantly different from zero eventhough this is frequently observed for stock index series¹⁰. We therefore leave the conditional mean μ_t unmodelled, *i.e.* we assume that in equation (1) $\mu_t = 0$.

¹⁰See, *e.g.*: Campbell, Lo and MacKinlay (1997, Chapter 2).

4.2 Empirical in-sample results

In this section we present the results obtained with the four models introduced in section 2, which are the SV, the SVX, the SVX⁺ and the VX model. The general mean and variance equations are defined in equations (1), with $y_t = R_t$ and $\mu_t = 0$, and (2), respectively. The log processes for these models are given by

$$\begin{aligned}
 \text{SV Model:} & & h_t &= \phi h_{t-1} + \sigma_\eta \eta_t \\
 \text{SVX Model:} & & h_t &= \phi h_{t-1} + \gamma x_t + \sigma_\eta \eta_t \\
 \text{SVX}^+ \text{ Model:} & & h_t &= \phi h_{t-1} + \gamma(1 - \phi L)x_t + \sigma_\eta \eta_t \\
 \text{VX Model:} & & h_t &= \gamma x_t.
 \end{aligned}$$

The SV model can be obtained by imposing the parameter restriction $\gamma = 0$ in the h_t definition of either the SVX or the SVX⁺ model and when we impose $\phi = \sigma_\eta = 0$ for the SVX or the SVX⁺ model we obtain the VX model. For these combinations of models we can therefore apply statistical tests for nested models.

In table 2 we present the parameter estimates and results for the tests of the presumed comprehensive informational content of the VIX index relative to the SV model for daily returns on the Standard & Poor's 100 index over the periods 1986–1999 and 1988–1999.

For the SV model we find that the estimated coefficients for the persistence parameter ϕ are close to unity and statistically significant for both samples. The fact that the shorter 1988–1999 sample does not contain the large outlier of the longer sample is reflected not only in the size of the ϕ estimate, but also in the estimated values for the scaling parameter σ^{*2} and the variance of η_t , as these are both larger for the more volatile 1986–1999 sample. Of further interest are the statistics for ε_t , the error term in the mean equation, which indicate that the assumption of zero serial correlation is not violated by the SV model and that the SV model is capable of absorbing excess kurtosis found in the underlying series.

The normality statistics are considerably worse for the deterministic VX model which is mainly attributable to the fact that this model does not have an error term in the variance equation. The values for γ in this model are significantly larger than unity for both samples and as they are combined with relatively low estimates for σ^{*2} this means that the volatility process of the VX model exhibits more movement than the VIX index while at the same time it results in lower variances than the implied volatility measure. This reflects our observation in section 4.1 that the VIX index tends to overstate the volatility process while at the same time it underestimates its degree of variation¹¹. On the basis of the maximum likelihood statistics we would have to favour the VX model to the SV model although the former clearly violates the model assumptions with regard to ε_t .

The best maximum likelihood statistics for both periods are obtained with the SVX and the SVX⁺ model which combine the two sources of volatility information. We find that the γ parameters in these models are always statistically significant which confirms the earlier findings in the GARCH literature that implied volatility contains crucial information about the volatility process. The estimates for the ϕ parameters in the SVX models are negative and statistically significant implying that the γ parameter is upwardly biased. When we include the persistence adjustment term in the SVX⁺ model ϕ is still found to be negative but no longer significant and the values for γ move towards unity and are close to be statistically equal to one. The values for γ are now smaller than those of the VX model. The extra movement in the volatility process is compensated by η_t which has highly significant values

¹¹See Fleming (1998) who also observes that implied volatility, on average, exceeds realised volatility and is therefore upwardly biased.

Table 2: Estimation results

Period	1986–1999				1988–1999			
	3531				3027			
Model	SV	SVX	SVX ⁺	VX	SV	SVX	SVX ⁺	VX
σ^{*2}	0.726 0.571 0.923	0.426 0.398 0.456	0.425 0.397 0.455	0.499 0.473 0.527	0.688 0.514 0.921	0.421 0.393 0.452	0.420 0.392 0.451	0.483 0.457 0.511
ϕ	0.975 0.959 0.984	-0.213 -0.373 -0.035	-0.008 -0.247 0.233		0.982 0.966 0.991	-0.229 -0.402 -0.031	-0.136 -0.420 0.184	
γ		1.310 1.224 1.395	1.081 1.009 1.154	1.156 1.085 1.123		1.314 1.213 1.416	1.072 0.987 1.156	1.150 1.070 1.229
σ_η^2	0.033 0.022 0.051	0.285 0.210 0.387	0.301 0.228 0.398		0.020 0.012 0.036	0.274 0.195 0.384	0.290 0.211 0.399	
$\ln L$	-4633.02	-4481.46	-4484.44	-4530.95	-3813.83	-3696.53	-3698.82	-3717.78
$LR(\gamma = 0)^a$	303.12				234.60			
$LR(\gamma = 0)^b$	297.16				230.02			
$LR(\phi = \sigma_\eta^2 = 0)^a$								42.50
$LR(\phi = \sigma_\eta^2 = 0)^b$								37.92
AIC	9272.04	8970.92	8976.88	9065.90	7633.66	7401.06	7405.64	7439.56
$Q(12)$	19.633	20.588	20.241	20.693	19.348	21.856	21.213	22.776
N	18.969	21.127	22.796	325.822	18.364	24.461	22.925	80.302

Parameter estimates are reported together with the asymptotic 95% confidence intervals which are asymmetric for σ^{*2} , ϕ and σ_η^2 ; $LR(\gamma = 0)$ and $LR(\phi = \sigma_\eta^2 = 0)$ are the likelihood ratio statistics for the hypotheses $\gamma = 0$ and $\phi = \sigma_\eta^2 = 0$, respectively, as measured against ^a the SVX and ^b the SVX⁺ model. AIC is the Akaike Information Criterion which is calculated as $-2(\ln L) + 2p$; $Q(\ell)$ is the Box-Ljung portmanteau statistic for the estimated observation errors which is asymptotically χ^2 distributed with $\ell - p$ degrees of freedom where p is the total number of estimated parameters; N is the χ^2 normality test statistic with 2 degrees of freedom.

for its variance where the increase in σ_η^2 relative to that of the SV model is due to the fact that the level of the h_t process is higher. The likelihood ratio statistics then indicate that we can never reject the joint hypothesis that both ϕ and σ_η^2 are statistically insignificant. The volatility process is therefore best described by the SVX⁺ model, rather than the SVX model with its higher maximum likelihood values as this model overestimates the value for γ and therefore provides a biased estimate of the volatility process. The SVX⁺ model is also to be preferred to the VX model which appears to be mainly attributable to the omission of the stochastic component η_t in the deterministic VX model. Our overall conclusion is therefore that the in-sample volatility process is best described by a volatility model which includes not only implied volatility but also a stochastic element, as shocks to the volatility process are not sufficiently captured by the implied volatility measure alone.

5 Volatility Forecasting Methodology

We develop forecasts based on the rolling window principle where we initially estimate the parameters over the period 2 January 1986 to 31 December 1994. This sample therefore spans a period of 9 years and consists of 2270 observations. This leaves an evaluation period of 1261 observations covering five years of data, *i.e.* from 3 January 1995 to 31 December 1999. Having calculated the volatility forecasts based on the parameters of this initial sample we roll it forward by one trading day, thus keeping the sample size constant at 2270 observations. We also construct volatility forecasts for 2, 3, 4, 5 and 10 day horizons. We obtain non-overlapping forecasts because we roll the estimation sample forward by N observations, where N denotes the forecasting horizon in terms of trading days. This means that we re-estimate the model parameters 1261 times for the one day ahead forecasts and 630, 420, 315, 252 and 126 times for the 2, 3, 4, 5 and 10 day forecasts, respectively.

5.1 Stochastic Volatility model forecasts

The one step ahead volatility forecast for the SV model, as defined in equations (1), (2) and (3), is calculated as

$$E(\sigma_{T+1|T}^2) = \exp(\ln \hat{\sigma}^{*2} + h_{T+1|T} + 0.5p_{T+1|T}), \quad (14)$$

and the N step ahead volatility forecast is defined as

$$\begin{aligned} E(\sigma_{T+1,T+N|T}^2) &= \sum_{j=1}^N \exp(\ln \hat{\sigma}^{*2} + h_{T+j|T} + 0.5p_{T+j|T}) \\ &= \hat{\sigma}^{*2} \exp(h_{T+1|T} + 0.5p_{T+1|T}) + \\ &\quad \hat{\sigma}^{*2} \sum_{j=2}^N \exp \left[\hat{\phi}^{j-1} h_{T+1|T} + 0.5 \left(\hat{\phi}^{2(j-1)} p_{T+1|T} + \sum_{i=0}^{N-2} \hat{\phi}^{2i} \hat{\sigma}_\eta^2 \right) \right], \end{aligned} \quad (15)$$

where $\hat{\sigma}^{*2}$, $\hat{\phi}$ and $\hat{\sigma}_\eta^2$ are the maximum likelihood estimates of σ^{*2} , ϕ and σ_η^2 , respectively. The estimator of h_{T+1} using all observations available at time T is denoted by $h_{T+1|T}$ with variance $p_{T+1|T}$ and both are computed by simulation methods which are discussed in section 3. The quantities $h_{T+1|T}$ and $p_{T+1|T}$ are computed using the methods of section 3.2.3. The multi-step forecasts are defined as a summation of the N individual forecasts conditional on the information available at time T . As N increases, the individual forecast $E(\sigma_{T+N|T}^2)$ will converge to a constant which we call the individual long-term volatility forecast and which is identical to the unconditional variance as given by

$$\hat{\sigma}^{*2} \exp \left(0.5 \frac{\hat{\sigma}_\eta^2}{1 - \hat{\phi}^2} \right).$$

It is evident from equation (15) that the rate at which the individual forecasts move towards this value is determined by the size of $\hat{\phi}$; the smaller the volatility persistence estimate, the faster the individual forecasts converge to the individual long-term volatility forecast value. In empirical applications for daily stock returns the volatility persistence estimates are invariably found to be close to unity, which means that for shorter forecasting horizons individual forecasts are almost solely determined by the size of the short-term volatility, denoted by $\hat{\sigma}^{*2} \exp(h_{T+1|T} + 0.5p_{T+1|T})$.

5.2 SVX⁺ model forecasts

The one step ahead forecasts for the SVX⁺ model, as defined in equations (1), (2) and (5) are obtained in a similar manner as for the SV model in equation (14) and using the same methods. However, for the SVX⁺ model one period ahead volatility forecast we require x_{T+1} and x_T in order to calculate the values for $h_{T+1|T}$ and $p_{T+1|T}$. As x_{T+1} is not known at time T we choose to replace it by x_T , the last available implied volatility measure in the information set, but only for the purpose of calculating $h_{T+1|T}$ and $p_{T+1|T}$. The same problem occurs for volatility forecasts further into the future and therefore we choose to define the N step ahead volatility forecasts of the SVX⁺ model as an N multiple of the one step ahead volatility forecast, so we define these as

$$E(\sigma_{T+1, T+N|T}^2) = N\hat{\sigma}^{*2} \exp(h_{T+1|T} + 0.5p_{T+1|T}). \quad (16)$$

5.3 VX model forecasts

The implied volatility forecasts are based on the VX model which we defined in equations (1), (2) and (6), so

$$h_t = \gamma x_t.$$

The one step ahead VX volatility forecast we calculate as

$$E(\sigma_{T+1|T}^2) = \hat{\sigma}^{*2} \exp(\hat{\gamma}x_T + 0.5\hat{\sigma}_h^2), \quad (17)$$

where $\hat{\sigma}^{*2}$ and $\hat{\gamma}$ are the maximum likelihood estimates of σ^{*2} and γ . The sample prediction error variance is denoted by $\hat{\sigma}_h^2$. Again we replace x_{T+1} with x_T in the one period ahead forecasting equation as the former is not yet known; the N step ahead forecasts are then given by

$$E(\sigma_{T+1, T+N|T}^2) = N\hat{\sigma}^{*2} \exp(\hat{\gamma}x_T + 0.5\hat{\sigma}_h^2), \quad (18)$$

and these are therefore also defined as a multiple of the one day ahead VX volatility forecast.

5.4 Measuring predictive forecasting ability

To evaluate the accuracy of variance forecasts they have to be compared with realised volatility, which can not be observed. It is common practice in the literature to define the actual or realised variance as the squared observed returns, which for one day ahead volatility is equal to

$$R_{T+1}^2 = \sigma_{T+1}^2 \varepsilon_{T+1}^2, \quad (19)$$

However, the squared error ε_{T+1}^2 will vary widely which implies that only a relatively small part is attributable to σ_{T+1}^2 .

An alternative approach which addresses this problem has been suggested; see, for example, Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold and Labys (1999), Barndorff-Nielsen and Shephard (2001) and Andersen, Bollerslev, Diebold and Ebens (2000). In these studies intradaily return data is used to approximate ex-post volatility more accurately. Following these studies we define

intradaily squared returns as the sum of the squared five minute returns between 9.30 a.m. and 4.00 p.m. EST during the relevant trading day and to this we then add the overnight return, so

$$\tilde{\sigma}_{T+1}^2 = \left\{ \sum_{k=1}^m [100(\ln P_{T+1,(k+1/m)} - \ln P_{T+1,(k/m)})]^2 \right\} + [100(\ln P_{T+2,(1/m)} - \ln P_{T+1,(m/m)})]^2, \quad (20)$$

where $P_{T+1,(1/m)}$ denotes the first price of the Standard & Poor's Index on day $T + 1$ at 9.30 a.m. with m representing the number of observations during day $T + 1$, which on a full trading day amount to 79 observations. The price of the Standard & Poor's Index at 9.30 a.m. on the subsequent trading day is then denoted by $P_{T+2,(1/m)}$.

The multiple-day values of the daily squared returns and the intradaily squared returns are obtained by summing the realised volatility measures of equations (19) and (20) over the relevant forecasting interval, so

$$R_{T+1,T+N}^2 = \sum_{i=1}^N R_{T+i}^2 \quad (21)$$

with

$$R_{T+i}^2 = [100(\ln P_{T+i} - \ln P_{T+i-1})]^2,$$

where P_{T+i} denotes the closing price of the Standard & Poor's Index at time $T + i$, and

$$\tilde{\sigma}_{T+1,T+N}^2 = \sum_{i=1}^N \tilde{\sigma}_{T+i}^2, \quad (22)$$

with

$$\tilde{\sigma}_{T+i}^2 = \left\{ \sum_{k=1}^m [100(\ln P_{T+i,(k+1/m)} - \ln P_{T+i,(k/m)})]^2 \right\} + [100(\ln P_{T+i+1,(1/m)} - \ln P_{T+i,(m/m)})]^2.$$

In order to assess the predictive accuracy of the volatility forecasts we compare the goodness-of-fit R^2 statistics, which are calculated from the regressions

$$R_{T+1,T+N}^2 = a + b E(\sigma_{T+1,T+N|T}^2) + \epsilon, \quad (23)$$

and

$$\tilde{\sigma}_{T+1,T+N}^2 = a + b E(\sigma_{T+1,T+N|T}^2) + \epsilon, \quad (24)$$

for the squared and the intradaily squared returns, respectively. If the relevant volatility forecast is unbiased, then $a = 0$ and $b = 1$.

In addition to the regression based evaluation method we also report on a number of error statistics, which are the mean squared error (MSE), the median squared error (MedSE) and the mean absolute error (MAE) as these criteria are also frequently applied in the volatility forecasting literature.

6 Out-of-Sample Results

In this section we report on the out-of-sample forecasting results of the SV, the SVX⁺ and the VX model over the evaluation period 1995 to 1999. Before we discuss these results in section 6.2 we will examine the relationships between the various parameters in the SV model.

6.1 The parameters estimates of the SV model

For the one day ahead SV volatility forecasts we had to estimate the SV model 1261 times which resulted in an equal number of estimates for all the parameters in the model. This now allows us to examine the dynamics of the SV model as the forecasting sample rolls forward by one observation at the time. For this purpose we plot in figure 2 the estimates of the persistence parameter ϕ , the error variance of the volatility process $\hat{\sigma}_\eta^2$, and the scaling parameter σ^{*2} . These estimates are based on the previous 9 years of data and the sample variance of each of these data series is plotted in the same figure alongside those of the SV parameter estimates.

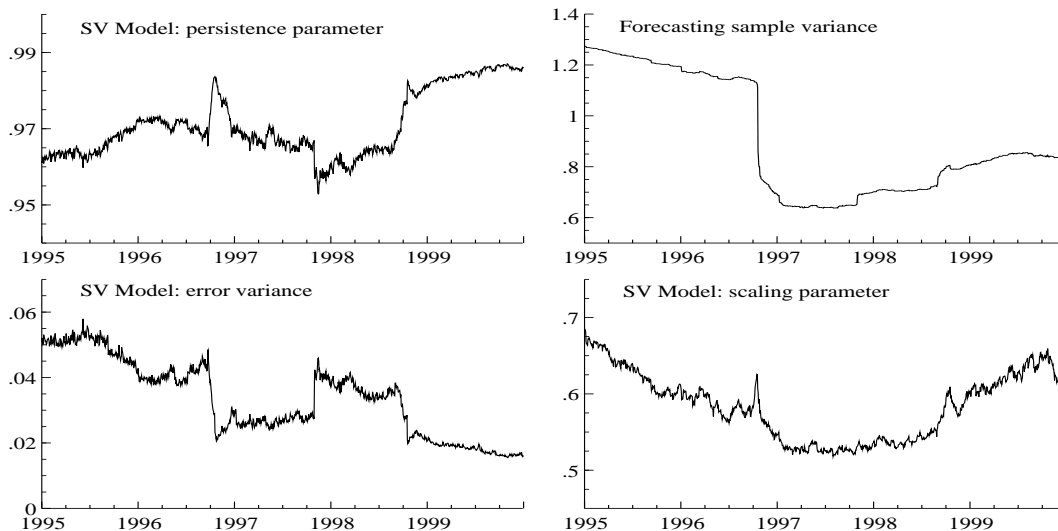


Figure 2: Parameter estimates of the SV model and forecasting sample variance based on the previous 9 years of data.

The graph for the $\hat{\phi}$ parameter shows that volatility is highly persistent for all 1261 samples and when we compare this time series with that of $\hat{\sigma}_\eta^2$ we observe that these series move in opposite directions during the entire period. The negative relationship between these two parameter estimates indicates that large (unexpected) shocks to the volatility process have a downward effect on the ϕ estimate and that volatility persistence is higher when these shocks are more moderate in size, *i.e.* when values for $\hat{\sigma}_\eta^2$ are smaller. For the forecasting sample variance, depicted in the top right corner of figure 2, we see a sharp decrease after approximately two years when the observations relating to the 1987 stock market crash drop out of the forecasting sample. This decrease in volatility is also reflected in $\hat{\sigma}_\eta^2$ and in the estimated value of the scaling parameter σ^{*2} which displays a positive correlation with the variance series of the forecasting samples over the full period. The relationship between the scaling parameter and the other two SV parameters differs however across the sample which is to be expected as they measure different aspects of the volatility process: the estimate for σ^{*2} reflects the level of volatility, $\hat{\phi}$ measures the degree of volatility persistence and the value for $\hat{\sigma}_\eta^2$ indicates the amount of variation in the volatility process. This means that during times of persistent high volatility we will observe high values for $\hat{\sigma}^{*2}$ and $\hat{\phi}$ but low values for $\hat{\sigma}_\eta^2$ as there is relatively little movement in the volatility process itself. However, the estimated value for the scaling parameter will still be large when a high level of volatility is due to a few outliers, but the variation in the volatility process as measured by $\hat{\sigma}_\eta^2$ is going to be higher and will be accompanied by a smaller value for the volatility persistence estimate $\hat{\phi}$.

6.2 Empirical out-of-sample forecasting results

The empirical out-of-sample forecasting results based on the methodology described in section 5 are presented in tables 3 and 4.

In table 3 we evaluate the volatility forecasts obtained with the SV, SVX⁺ and VX model against the squared returns over the full five year evaluation period. The goodness-of-fit R^2 statistic for the SVX⁺ and the VX model forecasts are higher than those of the SV model. However, the hypothesis that $a = 0$ or $b = 1$ is least violated for the SV model volatility forecasts indicating that the SV model forecasts exhibit the smallest degree of bias. When we evaluate the accuracy of the volatility forecasts on the basis of error statistics a different picture emerges. Although the SVX⁺ and VX model still perform very well when we consider the MSE error statistic, we observe that the VX model fares much worse in terms of the MedSE and the MAE statistics. For these error statistics the VX model is not only outperformed by the SVX⁺ model but also frequently by the SV model. The results of this evaluation are therefore mixed: the SV model produces volatility forecasts that have the smallest degree of bias, the SVX⁺ and VX model have very similar goodness-of-fit and MSE values and the SVX⁺ model produces the most accurate out-of-sample volatility forecasts in terms of the MedSE and MAE error statistics.

In table 4 we present the forecasting evaluation results with realised returns defined in terms of five minute squared returns. As this high frequency data series does not start until 6 January 1997, our evaluation period consists of 3 years of data resulting in forecasting samples of 754, 377, 251, 188 and 75 observations for $N = 1, 2, 3, 4, 5$ and 10, respectively. We observe that the values for the R^2 statistic increase considerably when we define realised volatility in terms of five minute squared returns and this degree of increase is conform earlier findings in the high frequency return literature. The highest values for the goodness-of-fit statistic are almost always those of the VX model; its volatility forecasts are however severely upwardly biased as the hypothesis that $b = 1$ has to be rejected at very low significance levels for values of b smaller than unity. The SVX⁺ model which has comparable values for R^2 produces volatility forecasts that are far less biased. The error statistics also favour this model as the forecasts of the SVX⁺ model are, with the exception of two statistics for $N = 10$, always the lowest of the three models. In addition, we observe that the VX model is almost consistently the worst performing forecasting model in terms of error statistics. We therefore conclude that the worst performing forecasting model is the VX model eventhough it has the highest R^2 statistics. The most accurate volatility forecasts are obtained with the SVX⁺ model which has goodness-of-fit values similar to those of the VX model but combines this with the best error statistics and forecasts that are less biased.

In figure 3 we graph the one period ahead volatility forecasts of the SV, the SVX⁺ and the VX model together with the two measures of realised volatility where the scale of the daily squared return series differs from those of the other four series. However, it is obvious from the graphs that all five series follow a very similar pattern and we can clearly distinguish two periods of increased volatility which occur towards the end of 1997 and 1998. The SVX⁺ and VX forecasting series are near perfectly correlated but forecasts of the VX model are on average 17% higher than those of the SVX⁺ model which is favourable for the VX model when volatility is very high but leads to overestimation during times of relative tranquility. In terms of sample moments the SVX⁺ model produces forecasts that are very much alike those of the SV model eventhough it correlation coefficient with the SV forecasting series is lower at 83%. Although difficult to discern from figure 3 all forecasting series, on average, overestimate realised volatility both in terms of daily squared and intradaily squared returns; this problem is then most severe for the VX volatility forecasts.

Table 3: Out-of-sample forecasting results evaluated against daily squared returns for the (i) SV model, (ii) SVX⁺ model and the (iii) VX model based on the 1986–1999 sample and for the evaluation period 3 January 1995 to 31 December 1999

Forecasting Model		1986–1999					
		$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 10$
SV Model	a	0.0365 (0.268)	0.2881 (0.908)	0.6118 (1.207)	0.8098 (1.161)	1.3221 (1.410)	2.4698 (1.025)
	b	1.1356 (1.157)	1.0221 (0.161)	0.9601 (0.272)	0.9671 (0.215)	0.9024 (0.589)	0.9519 (0.210)
	R^2	0.0694	0.0817	0.0929	0.1131	0.1061	0.1222
	MSE	8.1406	21.462	35.785	49.878	70.705	190.76
	MedSE	0.3517	0.9406	1.4160	2.0022	2.5819	7.3689
	MAE	1.1171	1.8056	2.2633	2.8325	3.1066	5.4334
SVX ⁺ Model	a	-0.1675 (1.313)	-0.0345 (0.119)	0.3600 (0.787)	0.3726 (0.609)	1.2079 (1.392)	2.9023 (1.464)
	b	1.3959 (3.517)	1.2330 (1.871)	1.0758 (0.582)	1.1219 (0.921)	0.9451 (0.362)	0.8992 (0.578)
	R^2	0.1130	0.1350	0.1404	<u>0.1869</u>	<u>0.1347</u>	<u>0.1766</u>
	MSE	7.8423	20.378	34.023	46.074	<u>68.578</u>	<u>179.39</u>
	MedSE	<u>0.2991</u>	<u>0.7753</u>	<u>1.2804</u>	<u>1.6861</u>	<u>2.2644</u>	<u>6.5155</u>
	MAE	<u>1.0682</u>	<u>1.6943</u>	<u>2.1192</u>	<u>2.5732</u>	<u>3.0286</u>	<u>5.2690</u>
VX Model	a	-0.1357 (1.100)	0.0651 (0.232)	0.5011 (1.130)	0.6776 (1.147)	1.4608 (1.729)	3.3400 (1.720)
	b	1.1222 (1.416)	0.9702 (0.307)	0.8447 (1.524)	0.8526 (1.440)	0.7326 (2.241)	0.7031 (2.134)
	R^2	<u>0.1185</u>	<u>0.1369</u>	<u>0.1411</u>	0.1815	0.1309	0.1708
	MSE	<u>7.6897</u>	<u>20.073</u>	<u>33.829</u>	<u>45.895</u>	69.284	182.69
	MedSE	0.4436	1.0658	1.7170	2.1253	2.8655	8.2397
	MAE	1.1430	1.8008	2.2585	2.7704	3.1914	5.7354

Parameter estimates and goodness-of-fit R^2 statistics for the OLS regressions as defined in equation (23) with t-statistics in parentheses testing for the null hypotheses $a = 0$ and $b = 1$. The highest values for R^2 and the lowest error statistic values are underlined.

Table 4: Out-of-sample forecasting results evaluated against intradaily squared returns for the (i) SV model, (ii) SVX⁺ model and the (iii) VX model based on the 1988–1999 sample and for the evaluation period 6 January 1997 to 31 December 1999

Forecasting Model		1988–1999					
		$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 10$
SV Model	a	0.0585 (0.807)	0.2670 (1.485)	0.3966 (1.275)	0.6137 (1.210)	0.9790 (1.402)	3.0864 (1.628)
	b	0.9240 (1.509)	0.8712 (2.033)	0.8714 (1.761)	0.8757 (1.330)	0.8334 (1.640)	0.7428 (1.836)
	R^2	0.3091	0.3350	0.3639	0.3207	0.3126	0.2781
	MSE	0.9803	2.8944	5.7255	9.6701	14.951	50.127
	MedSE	0.1549	0.4704	1.0704	1.5317	2.6939	<u>5.2282</u>
	MAE	0.5944	1.0395	1.4697	1.8761	2.3689	<u>4.1301</u>
SVX ⁺ Model	a	-0.1130 (1.661)	-0.0036 (0.022)	0.3572 (1.231)	0.8304 (1.978)	1.0334 (1.772)	3.2963 (2.153)
	b	1.0472 (1.011)	0.9595 (0.703)	0.8579 (2.168)	0.7999 (2.790)	0.8021 (2.475)	0.6934 (2.944)
	R^2	0.4006	0.4258	0.4079	<u>0.4007</u>	0.4046	0.3780
	MSE	<u>0.8508</u>	<u>2.4842</u>	<u>5.3923</u>	<u>8.8366</u>	<u>13.289</u>	<u>46.585</u>
	MedSE	<u>0.1369</u>	<u>0.4193</u>	<u>0.8846</u>	<u>1.1947</u>	<u>2.1332</u>	8.1839
	MAE	<u>0.5548</u>	<u>0.9981</u>	<u>1.4528</u>	<u>1.8032</u>	<u>2.2530</u>	4.3358
VX Model	a	-0.0496 (0.781)	0.1182 (0.754)	0.5301 (1.957)	1.0793 (2.684)	1.2621 (2.272)	3.6432 (2.462)
	b	0.8204 (5.097)	0.7507 (5.691)	0.6686 (6.716)	0.6182 (6.860)	0.6303 (5.985)	0.5503 (5.469)
	R^2	<u>0.4189</u>	<u>0.4392</u>	<u>0.4244</u>	0.3988	<u>0.4130</u>	<u>0.3802</u>
	MSE	0.9557	3.0408	7.0630	12.206	18.034	69.094
	MedSE	0.2613	0.6934	1.8086	3.0596	3.6090	16.905
	MAE	0.6681	1.2213	1.8188	2.2921	2.8458	5.9025

Parameter estimates and goodness-of-fit R^2 statistics for the OLS regressions as defined in equation (24) with t-statistics in parentheses testing for the null hypotheses $a = 0$ and $b = 1$. The highest values for R^2 and the lowest error statistic values are underlined.

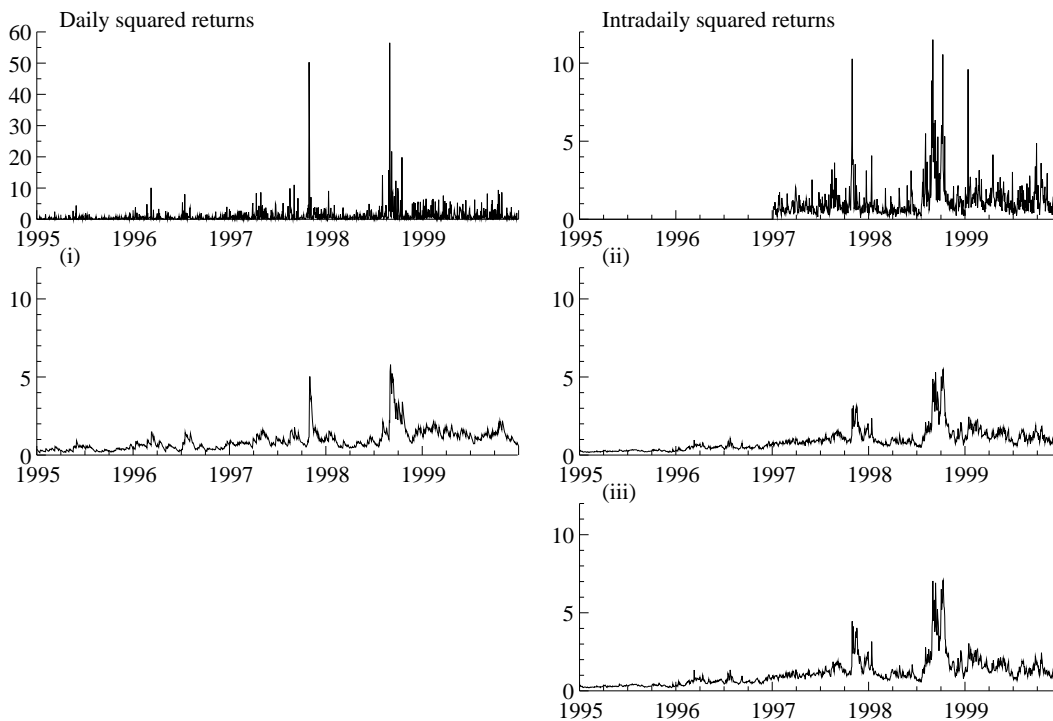


Figure 3: Daily squared and intradaily squared returns together with the one day ahead volatility forecasts of the (i) SV, (ii) SVX⁺ and (iii) VX model for the Standard & Poor's 100 index over the period 03/01/95 to 31/12/99 based on a 9 year rolling window sample.

7 Summary and Conclusions

In this paper we examine the predictive ability of Stochastic Volatility (SV) models and compare its forecasts with the volatility forecasts implied by option prices for daily returns on the Standard & Poor's 100 Stock Index. For this purpose we extend the SV model to a volatility model which allows for the inclusion of implied volatility as an exogeneous variable in the variance equation. As the resulting SVX model includes an entire lag structure of the implied volatility measure we extend the SVX model further to adjust for this with a persistence adjustment term and thus obtain a second model which we call the SVX⁺ model. In addition we define a volatility model which only utilises implied volatility and refer to it as the VX model. We have estimated the SV, SVX and SVX⁺ models successfully by exact maximum likelihood using Monte Carlo importance sampling methods. Our in-sample results indicate that historical returns are outperformed by implied volatility but that the former contains additional information about the volatility process in the form of stochastic shocks that are incorporated in the variance equation of the SVX type models. Our results do not contradict earlier findings in the GARCH literature, where recent research has indicated that implied volatility provides the most accurate volatility forecasts, as the GARCH class of model is by definition a deterministic model which does not allow for a stochastic element in the variance equation. The out-of-sample volatility forecasts are constructed for forecasting horizons ranging from 1 to 10 trading days and we approximate realised volatility as daily squared returns and intradaily squared returns following research by, for example, Andersen and Bollerslev (1998). The relative forecasting accuracy of the various volatility models is evaluated using both regression based evaluation methods and error statistics. We obtain mixed results when we define realised volatility in terms of daily squared returns but when we use intradaily squared

returns for our forecasting evaluation we find that the most accurate ex-ante volatility forecasts are obtained with the SVX^+ model. Although the R^2 statistics are the highest for the VX model we find on closer examination that this model produces forecasts that are severely upwardly biased and therefore conclude that this model is outperformed not only by the SVX^+ but also by the SV model.

Appendix: approximating model used for simulation

Consider the single density component $p_t = N(0, \sigma_t^2)$ with $\sigma_t^2 = \exp(\theta_t)$ where θ_t is given in section 3.2. Here we develop the approximating model based on a linear Gaussian model with mean $E(y_t) = \theta_t + c_t$ and variance $V(y_t) = H_t$, that is

$$y_t = \theta_t + u_t, \quad u_t \sim N(c_t, H_t), \quad t = 1, \dots, T, \quad (25)$$

where c_t and H_t are determined in such a way that the mean and variance of y_t implied by the approximating model (25) and by the true model (1) and (4) are as close as possible¹².

We achieve this by equalising the first and second derivatives of $p(y|\theta, \psi)$ and $g(y|\theta, \psi)$ with respect to θ at $\hat{\theta} = \tilde{E}(\theta) = \int \theta g(\theta|y, \psi)$. Note that $p(\cdot)$ refers to a density for the true model and $g(\cdot)$ refers to a density for the approximating Gaussian model. Further, it follows that $\hat{\theta}$ can simply be obtained via the Kalman filter and smoother applied to the approximating model (25). The conditional densities are given by

$$p(y|\theta, \psi) = \prod_{t=1}^T p_t, \quad g(y|\theta, \psi) = \prod_{t=1}^T g_t, \quad (26)$$

with

$$\begin{aligned} p_t &= N(0, \sigma_t) = p(y_t|\theta_t, \psi) = -0.5[\ln 2\pi + \theta_t + \exp(-\theta_t)y_t^2], \\ g_t &= N(c_t + \theta_t, H_t) = g(y_t|\theta_t, \psi) = -0.5[\ln 2\pi + \ln H_t + H_t^{-1}(y_t - c_t - \theta_t)^2]. \end{aligned}$$

Differentiating both densities twice with respect to θ_t gives

$$\begin{aligned} \dot{p}_t &= -0.5[1 - \exp(-\theta_t)y_t^2], \\ \ddot{p}_t &= -0.5 \exp(-\theta_t)y_t^2, \\ \dot{g}_t &= H_t^{-1}(y_t - c_t - \theta_t), \\ \ddot{g}_t &= -H_t^{-1}. \end{aligned}$$

Equalising the first and second derivatives, that is $\dot{p}_t = \dot{g}_t$ and $\ddot{p}_t = \ddot{g}_t$ for $t = 1, \dots, T$, leads to

$$\begin{aligned} c_t &= y_t - \theta_t + 0.5H_t - 1, \\ H_t &= 2 \exp(\theta_t)/y_t^2. \end{aligned}$$

For given values of $\theta_t = \tilde{\theta}_t$, the resulting model for $\tilde{y}_t = y_t - c_t$ is equivalent to

$$\tilde{y}_t = \theta_t + \tilde{u}_t, \quad \tilde{u}_t \sim N(0, \tilde{H}_t), \quad t = 1, \dots, T,$$

with

$$\tilde{y}_t = \tilde{\theta}_t - 0.5\tilde{H}_t + 1, \quad \tilde{H}_t = 2 \exp(\tilde{\theta}_t)/y_t^2.$$

It should be noted that $H_t > 0$ for any value of θ_t . We cannot solve out for \tilde{y}_t and \tilde{H}_t at $\hat{\theta}_t = \tilde{E}(\theta_t)$ because \tilde{E} refers to expectation with respect to the approximating model which depend on θ_t . However, such complicated but linear system of equations is usually solved iteratively by starting with a trial value $\theta_t = \tilde{\theta}_t$. Computing \tilde{y}_t and \tilde{H}_t and applying the Kalman filter smoother to model (25) leads to a smoothed estimate for θ_t which can be used as a new trial value for θ_t . Recomputing \tilde{y}_t and \tilde{H}_t based on this new trial value leads to an iterative procedure which converges to $\hat{\theta}_t$. Note that the first and second derivatives of the true and approximating densities are equal at $\theta_t = \hat{\theta}_t$. More details are given by Durbin and Koopman (1997). It is worth mentioning that $\hat{\theta}_t$ is equal to the mode of $p(\theta_t|y, \psi)$ which can be of interest.

¹²Note that the true model implies a nonlinear relationship between y_t and θ_t ; the approximating (linear) model is effectively a second-order Taylor expansion of the true model around θ_t . Further, the multivariate Gaussian density $g(\theta|y, \psi)$ can be regarded as a Laplace approximation to the true density $p(\theta|y, \psi)$; see Shephard and Pitt (1997).

References

- Akgiray, V. (1989), Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts, *Journal of Business* 62, 55-80.
- Amin, K.I. and V.K. Ng (1997), Inferring Future Volatility from the Information in Implied Volatility in Eurodollar Options: A New Approach, *Review of Financial Studies* 2, 333-367.
- Andersen, T.G. and T. Bollerslev (1998), Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts, *International Economic Review* 39, 885-905.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and H. Ebens (2000), The Distribution of Stock Return Volatility, Working Paper, Northwestern University, Duke University, New York University and John Hopkins University.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (1999), The Distribution of Exchange Rate Volatility, Working Paper, Northwestern University, Duke University, New York University and University of Pennsylvania, Revised version of NBER Working Paper No. 6961 (1998).
- Barndorff-Nielsen, O.E. and N. Shephard (2001), Non-Gaussian OU Based Models and Some of Their Uses in Financial Economics (with discussion), *Journal of the Royal Statistical Society, Series B*, forthcoming.
- Beckers, S. (1981), Standard Deviations Implied in Options Prices as Predictors of Future Stock Price Variability, *Journal of Banking and Finance* 5, 363-381.
- Bera, A.K. and M.L. Higgins (1993), ARCH Models: Properties, Estimation and Testing, *Journal of Economic Surveys* 7, 305-366.
- Blair, B., S. Poon and S.J. Taylor (2000), Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High Frequency Returns, Working Paper, Lancaster University.
- Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31, 307-327.
- Bollerslev, T., R.Y. Chou and K.F. Kroner (1992), ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence, *Journal of Econometrics* 52, 5-59.
- Bollerslev, T., R.F. Engle and D.B. Nelson (1994), ARCH Models, in: *Handbook of Econometrics*, Vol. 4, eds. R.F. Engle and D.L. McFadden, Elsevier Science, Amsterdam, 2959-3038.
- Campbell, J.Y., A.W. Lo and A.C. MacKinlay (1997), *The Econometrics of Financial Markets*, New Jersey: Princeton University Press.
- Canina, L. and S. Figlewski (1993), The Informational Content of Implied Volatility, *Review of Financial Studies* 6, 659-693.
- Chib, S., F. Nardari and N. Shephard (1998), Markov Chain Monte Carlo Methods for Generalized Stochastic Volatility Models, Discussion Paper: Nuffield College, Oxford.
- Chiras, D.P. and S. Manaster (1978), The Information Content of Options Prices and a Test of Market

Efficiency, *Journal of Financial Economics* 6, 213-234.

Christensen, B.J. and N.R. Prabhala (1998), The Relation between Implied and Realized Volatility, *Journal of Financial Economics* 50, 125-150.

Day, T.E. and C.M. Lewis (1992), Stock Market Volatility and the Information Content of Stock Index Options, *Journal of Econometrics* 52, 267-287.

de Jong, P. and N. Shephard (1995), The Simulation Smoother for Time Series Models, *Biometrika* 82, 339-350.

Diebold, F.X. and J.A. Lopez (1995), Modeling Volatility Dynamics, in: *Macroeconometrics: Developments, Tensions and Prospects*, eds. K. Hoover, Kluwer Academic Publishers, Amsterdam, 427-472.

Dimson, E. and P. Marsh (1990), Volatility Forecasting without Data-Snooping, *Journal of Banking and Finance* 14, 399-421.

Doornik, J.A. (1998), *Object-Oriented Matrix Programming using Ox 2.0*, Timberlake Consultants Press, London. www.nuff.ox.ac.uk/Users/Doornik/

Durbin, J. and S.J. Koopman (1997), Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models, *Biometrika* 84, 669-84.

Durbin, J. and S.J. Koopman (2000), Time Series Analysis of Non-Gaussian Observations based on State Space Models from both Classical and Bayesian Perspectives (with discussion), *Journal of the Royal Statistical Society, Series B*, 62, 3-56.

Engle, R.F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50, 987-1006.

Fleming, J., B. Ostdiek and R.E. Whaley (1995), Predicting Stock Market Volatility: A New Measure, *Journal of Futures Markets* 15, 265-302.

Fleming, J. (1998), The Quality of Market Volatility Forecast Implied by S&P 100 Index Option Prices, *Journal of Empirical Finance* 5, 317-345.

Ghysels, E., A.C. Harvey and E. Renault (1996), Stochastic Volatility, in: *Handbook of Statistics*, Vol. 14, Statistical Methods in Finance, eds. G.S. Maddala and C.R. Rao, North-Holland, Amsterdam, 128-198.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey, A.C. (1993), *Time Series Models*, 2nd edition, Harvester Wheatsheaf, Hemel Hempstead.

Harvey, A.C., E. Ruiz and N. Shephard (1994), Multivariate Stochastic Variance Models, *Review of Economic Studies* 61, 247-264.

Heynen, R.C. (1995), *Essays on Derivatives Pricing Theory*, PhD Dissertation, Erasmus University Rotterdam, Thesis Publishers, Amsterdam.

- Johnston, J. and J. DiNardo (1997), *Econometric Methods*, 4th edition, McGraw-Hill, International editions.
- Kim, S., N. Shephard and S. Chib (1998), Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies* 65, 361-393.
- Koopman, S.J., Harvey, A.C., Doornik, J.A., and Shephard, N. (2000), *STAMP 6.0 Structural Time Series Analyser, Modeller and Predictor*, Timberlake Consultants, London. www.stamp-software.com
- Koopman, S.J. and E. Hol Uspensky (2000), The Stochastic Volatility in Mean Model: Empirical Evidence from International Stock Markets, Discussion Paper, Tinbergen Institute, The Netherlands. www.tinbergen.nl
- Koopman, S.J., N. Shephard and J. Doornik (1999), Statistical Algorithms for Models in State Space using SsfPack 2.2, *Econometrics Journal*, 2, 113-166. www.ssfpack.com
- Lamoureux, C.G. and W.D. Lastrapes (1993), Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatility, *Review of Financial Studies* 6, 293-326
- Latané, H.A. and R.J. Rendleman (1976), Standard Deviations of Stock Price Ratios Implied in Option Prices, *Journal of Finance* 31, 369-381.
- Sandmann, G. and S.J. Koopman (1998), Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood, *Journal of Econometrics* 87, 271-301.
- Shephard, N.G. (1996), Statistical Aspects of ARCH and Stochastic Volatility, in: *Time Series Models in Econometrics, Finance and Other Fields*, Monographs on Statistics and Applied Probability 65, eds. D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, Chapman and Hall, London, 1-67.
- Shephard, N. and M.K. Pitt (1997), Likelihood Analysis of Non-Gaussian Measurement Time Series. *Biometrika* 84, 653-667.
- Taylor, S.J. (1994), Modeling Stochastic Volatility: A Review and Comparative Study, *Mathematical Finance* 4, 183-204.
- Walsh D.M. and G.Y. Tsou (1998), Forecasting Index Volatility: Sampling Interval and Non-Trading Effects, *Applied Financial Economics* 8, 477-485.
- Whaley, R.E. (1993), Derivatives on Market Volatility: Hedging Tools Long Overdue, *Journal of Derivatives* 1, 71-84.