JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **17**: 667–689 (2002)
Published online 13 November 2002 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.652

THE STOCHASTIC VOLATILITY IN MEAN MODEL: EMPIRICAL EVIDENCE FROM INTERNATIONAL STOCK MARKETS

SIEM JAN KOOPMAN[a]* AND EUGENIE HOL USPENSKY[b]

[a] *Department of Econometrics, Free University Amsterdam, The Netherlands*
[b] *Department of Accounting and Finance, University of Birmingham, UK*

SUMMARY

In this paper we present an exact maximum likelihood treatment for the estimation of a Stochastic Volatility in Mean (SVM) model based on Monte Carlo simulation methods. The SVM model incorporates the unobserved volatility as an explanatory variable in the mean equation. The same extension is developed elsewhere for Autoregressive Conditional Heteroscedastic (ARCH) models, known as the ARCH in Mean (ARCH-M) model. The estimation of ARCH models is relatively easy compared with that of the Stochastic Volatility (SV) model. However, efficient Monte Carlo simulation methods for SV models have been developed to overcome some of these problems. The details of modifications required for estimating the volatility-in-mean effect are presented in this paper together with a Monte Carlo study to investigate the finite sample properties of the SVM estimators. Taking these developments of estimation methods into account, we regard SV and SVM models as practical alternatives to their ARCH counterparts and therefore it is of interest to study and compare the two classes of volatility models. We present an empirical study of the intertemporal relationship between stock index returns and their volatility for the United Kingdom, the United States and Japan. This phenomenon has been discussed in the financial economic literature but has proved hard to find empirically. We provide evidence of a negative but weak relationship between returns and contemporaneous volatility which is indirect evidence of a positive relation between the expected components of the return and the volatility process.

1. INTRODUCTION

It is generally acknowledged that the volatility of many financial return series is not constant over time and that these series exhibit prolonged periods of high and low volatility, often referred to as volatility clustering. Over the past two decades two prominent classes of models have been developed which capture this time-varying autocorrelated volatility process: the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and the Stochastic Volatility (SV) model. GARCH models define the time-varying variance as a deterministic function of past squared innovations and lagged conditional variances whereas the variance in the SV model is modelled as an unobserved component that follows some stochastic process.[1] The most popular version of the SV model defines volatility as a logarithmic first-order autoregressive process, which is a

* Correspondence to: Professor Siem Jan Koopman, Department of Econometrics, Free University, de Boelelaan 1105, NL-1081, HV, Amsterdam, The Netherlands. E-mail: s.j.koopman@feweb.vu.nl
Contract/grant sponsor: Royal Netherlands Academy of Arts and Sciences.

[1] For surveys on the extensive GARCH literature we refer to Bollerslev *et al.* (1992), Bera and Higgins (1993), Bollerslev *et al.* (1994) and Diebold and Lopez (1995). SV models are reviewed in, for example, Taylor (1994), Ghysels *et al.* (1996) and Shephard (1996).

Received 22 November 1999
Revised 15 August 2001

discrete-time approximation of the continuous-time Ornstein–Uhlenbeck diffusion process used in the option-pricing literature.[2]

Although SV models are seen as a competitive alternative to GARCH models their empirical application has been limited. This can mainly be attributed to the difficulties that arise as a result of the intractability of the likelihood function which prohibits its direct evaluation. However, in recent years considerable advances have been made in this area. The estimation techniques that have been proposed for SV models can be divided into two groups: those that seek to construct the full likelihood function and those that approximate it or avoid the issue altogether. The methods originally suggested by Taylor (1986) and Harvey *et al.* (1994) belong to the latter category. Recently attention has moved towards the development of techniques that attempt to evaluate the full likelihood function.[3] For recent reviews on these full likelihood methods we refer to Fridman and Harris (1998), Sandmann and Koopman (1998) and Pitt and Shephard (1999), among others. The estimation method we adopt here is based on the Monte Carlo likelihood approach developed by Shephard and Pitt (1997) and Durbin and Koopman (1997) where the likelihood function is evaluated using importance sampling. These new techniques enable us to include explanatory variables in the mean equation and estimate their coefficients simultaneously with the parameters of the volatility process.[4] One of the explanatory variables in our model is the variance process itself, hence its name: Stochastic Volatility in Mean (SVM). The estimation of such an intricate model is not straightforward since volatility now appears in both the mean and the variance equation. This requires modification of the simulation maximum likelihood estimation method, details of which are given in Section 3.

The SV models we present are a practical alternative to the GARCH type models that have been used so widely in empirical financial research and which have relied on simultaneous modelling of the first and second moment. For certain financial time series such as stock index returns, which have been shown to display high positive first-order autocorrelations, this constitutes an improvement in terms of efficiency; see Campbell *et al.* (1997, Chapter 2). The volatility of daily stock index returns has been estimated with SV models but usually results have relied on extensive pre-modelling of these series, thus avoiding the problem of simultaneous estimation of the mean and variance.[5] The fact that we are able to estimate an SV model that includes volatility as one of the determinants of the mean makes our model suitable for empirical applications in which returns are partially dependent on volatility, such as studies that investigate the relationship between the mean and variance of stock returns. The SVM model can therefore be viewed as the SV counterpart of the ARCH-M model of Engle *et al.* (1987). The main difference between the two classes of models is that the ARCH-M model intends to estimate the relationship between expected returns and expected volatility, whereas the aim of the SVM model is to simultaneously estimate the *ex ante* relation between returns and volatility and the volatility feedback effect. This is further discussed in Section 4.1.

The remainder of this paper is organized as follows. The specification of time-varying variance models in general and the SVM model in particular are discussed in Section 2. In Section 3

---

[2] See Hull and White (1987), Scott (1987), Wiggins (1987) and Chesney and Scott (1989).

[3] See, for example, Jacquier *et al.* (1994), Kim, Shephard and Chib (1998), Sandmann and Koopman (1998) and Fridman and Harris (1998).

[4] Also see Fridman and Harris (1998) and Chib *et al.* (1998).

[5] The same seasonally adjusted S&P Composite stock index series (Gallant, Rossi and Tauchen, 1992) has been used in a number of studies; see, for example, Jacquier *et al.* (1994), Danielsson (1994), Sandmann and Koopman (1998), Fridman and Harris (1998) and Chib *et al.* (1998).

we develop the simulated maximum likelihood estimation method for the SVM model. Further, a Monte Carlo study is carried out to investigate the finite sample properties of the estimated parameters. Section 4 discusses the financial economic theory, describes the stock index data and reports on parameter estimation results. In the final section we present a summary and some conclusions.

2. MODELLING VOLATILITY

2.1. Basic Model

The aim is to simultaneously model the mean and variance of a series of returns on an asset denoted by y_t. Both the SV and GARCH model are defined by their first and second moments which can be referred to as the mean and variance equations. The most general form of the mean equation for both models is then defined as

$$y_t = \mu_t + \sigma_t \varepsilon_t \qquad \varepsilon_t \sim \text{NID}(0, 1) \tag{1}$$

$$\mu_t = a + \sum_{i=1}^{k} b_i x_i, t \tag{2}$$

where the mean μ_t depends on a constant a and regression coefficients $b_1, \ldots, b_k$. The explanatory variables $x_{1,t}, \ldots, x_{k,t}$ may also contain lagged exogeneous and dependent variables. The disturbance term ε_t is independently and identically distributed with zero mean and unit variance. Usually, the assumption of a normal distribution for ε_t is added. The positive volatility process is denoted by σ_t which remains to be specified in Section 2.2 for GARCH and Section 2.3 for SV models. The mean adjusted series is therefore defined as white noise with unit variance multiplied by the volatility process σ_t.

2.2. GARCH Model

The general form of the GARCH(p, q) model is

$$\begin{aligned}\sigma_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i (y_{t-i} - \mu_{t-i})^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 \\ &= \omega + \sum_{i=1}^{p} \alpha_i (\sigma_{t-i} \varepsilon_{t-i})^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 \end{aligned} \tag{3}$$

where the parameters to be estimated are $\omega, \alpha_1, \ldots, \alpha_p$ and $\beta_1, \ldots \beta_q$. An unanticipated shock to the return process at time t is therefore not incorporated into the volatility process until time $t + 1$.

The most commonly used model in applied financial studies is the GARCH(1,1) model which is given by

$$\sigma_t^2 = \omega + \alpha (y_{t-1} - \mu_{t-1})^2 + \beta \sigma_{t-1}^2 \tag{4}$$

with parameter values restricted to $\omega > 0, \alpha \geq 0$ and $\beta \geq 0$. Provided that the sum of α and β is less than one, the unconditional expectation of the conditional variance is constant and finite and given by

$$\frac{\omega}{1 - \alpha - \beta}$$

In empirical financial research with high-frequency data, $\alpha + \beta$ is often estimated as being close to unity, which implies a high degree of volatility persistence. Apart from volatility clustering GARCH models also capture part of the excess kurtosis observed in financial time series. Under the assumption of normality, existence of the fourth-order moment for the GARCH(1,1) model is ensured if $\beta^2 + 2\alpha\beta + 3\alpha^2 < 1$. Subject to this restriction it can be shown that the fourth moment will exhibit excess kurtosis

$$\kappa_y = \frac{\kappa_\varepsilon \mathrm{E}(\sigma_t^4)}{\mathrm{E}(\sigma_t^2)^2} = 3 + \frac{6\alpha^2}{1 - \beta^2 - 2\alpha\beta - 3\alpha^2}$$

and therefore $\kappa_y > \kappa_\varepsilon$; see Bollerslev (1986). For a further discussion on the features of GARCH models we refer to a number of surveys such as the ones given in footnote 1 and to the comprehensive selection of influential (G)ARCH papers in Engle (1995).

2.3. SV Model

In the case of the SV model the variance equation is specified in logarithmic form, that is,

$$\sigma_t^2 = \sigma^{*2} \exp(h_t) \tag{5}$$

with positive scaling factor σ^*. It follows that $h_t = \ln(\sigma_t^2/\sigma^{*2})$ where the stochastic process for h_t is

$$h_t = \phi h_{t-1} + \sigma_\eta \eta_t \qquad \eta_t \sim \mathrm{NID}(0, 1) \tag{6}$$

with persistence parameter ϕ which is restricted to a positive value less than one to ensure stationarity. The disturbances ε_t and η_t are mutually uncorrelated, contemporaneously and at all lags. The unconditional variance implied by the SV model is given by

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

and it can be shown that this model also captures part of the excess kurtosis as

$$\kappa_y = \frac{\kappa_\varepsilon \mathrm{E}(\sigma_t^4)}{\mathrm{E}(\sigma_t^2)^2} = 3 \exp\left(\frac{\sigma_\eta^2}{1 - \phi^2}\right)$$

which also implies that $\kappa_y > \kappa_\varepsilon$. Alternative specifications for the SV model can be deduced from

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

The main distinction between GARCH and SV models is that the latter has separate disturbance terms in the mean and variance equation, ε_t and η_t, respectively, which precludes direct observation of the variance process σ_t^2. GARCH models are deterministic in the sense that only the mean equation has a disturbance term and that its variance is modelled conditionally on the information up to and including time $t-1$. Therefore, the variance can be observed at time t. For the SV model, the deviation of y_t from the mean is captured by a function of the two disturbance terms whereas in the GARCH model this deviation is accounted for by a single disturbance term. For the GARCH model this point is evident but to clarify this for the SV model, we rewrite the model as follows:

$$\begin{aligned} y_t &= \mu_t + \sigma_t \varepsilon_t \\ &= \mu_t + \sigma^* \exp(0.5h_t)\varepsilon_t \\ &= \mu_t + \sigma^* \exp(0.5\phi h_{t-1}) \exp(0.5\eta_t)\varepsilon_t \end{aligned}$$

The overall innovation term of the SV model is the error term $\exp(0.5\eta_t)\varepsilon_t$ with a zero mean but with a non-Gaussian density. References to SV models can be found in footnote 1.

2.4. Volatility in Mean

The SV model with volatility included in the mean is given by equations (1) and (5) where the mean equation (2) is rewritten as

$$\mu_t = a + \sum_{i=1}^{k} b_i x_{i,t} + d\sigma^{*2} \exp(h_t) \tag{7}$$

with d as the regression coefficient measuring the volatility-in-mean effect. In particular, we will use the mean specification

$$\mu_t = a + b y_{t-1} + d\sigma^{*2} \exp(h_t) \tag{8}$$

This SVM model has six parameters which are to be estimated simultaneously using simulation methods discussed in the next section. Inclusion of the variance as one of the determinants of the mean facilitates the examination of the relationship between returns and volatility. It enables us to perform studies in the vein of French *et al.* (1987) but now in the context of SV models. The relative ease with which they were able to conduct their research, i.e. without prior manipulation of the original data series, is now also feasible for SV models.

The equivalent in-mean specification for the GARCH model is

$$\mu_t = a + b y_{t-1} + d\{\omega + \alpha(\sigma_{t-1}\varepsilon_{t-1})^2 + \beta\sigma_{t-1}^2\} \tag{9}$$

3. ESTIMATION OF THE SVM MODEL

In this section we show how the parameters of the SVM model are estimated by simulated maximum likelihood. Further, we show how to compute the conditional mean and variance of the volatility process h_t.

3.1. Model

To simplify the exposition we initially consider the model

$$\begin{aligned} y_t &= d\sigma^{*2}\exp(h_t) + \sigma^*\exp(0.5h_t)\varepsilon_t \\ h_t &= \phi h_{t-1} + \sigma_\eta \eta_t \end{aligned} \tag{10}$$

where y_t denotes the underlying series of interest, in our case these are stock index returns. The disturbances ε_t and η_t are standard normally distributed and they are mutually and serially uncorrelated. The latent variable h_t is modelled as a stationary Gaussian autoregressive process of order 1 and with $0 < \phi < 1$. The unknown parameters are collected in the vector

$$\psi = (\phi, \sigma_\eta, \sigma^{*2}, d)'$$

The nature of the model is conditionally Gaussian but we deal with a non-linear model since the variance of the overall disturbance term in y_t is given by σ^{*2} exp (h_t) which is stochastic. The Gaussian density for ε_t can be replaced by other continuous distributions. We note that the conditional Gaussian density function $p(y|\theta, \psi)$ of the SVM model with

$$\theta = (h_1, \ldots, h_T)'$$

is log-concave in h_t. This property is useful when employing importance sampling.

The techniques presented in the subsections below for model (10) can be adjusted straightforwardly to deal with the full model (1), where σ_t is given by equations (5) and (6) and μ_t is given by (8), since the extensions do not interact directly with the stochastic variance σ_t^2. However, it will be indicated in the following sections where details are different for the full model.

3.2. Likelihood Evaluation Using Importance Sampling

The construction of the likelihood for the SVM model is complicated because the latent variable h_t appears in both the mean and the variance of the SVM model. We adopt the Monte Carlo likelihood approach developed by Shephard and Pitt (1997) and Durbin and Koopman (1997). This simulation method of computing the loglikelihood function can be derived as follows.

Define the likelihood as

$$L(\psi) = p(y|\psi) = \int p(y, \theta|\psi)\mathrm{d}\theta = \int p(y|\theta, \psi)p(\theta|\psi)\mathrm{d}\theta \tag{11}$$

An efficient way of evaluating this likelihood is by using Monte Carlo integration and, in particular, importance sampling; see Ripley (1987, Chapter 5). We require a simulation device for sampling from an importance density $\tilde{p}(\theta|y, \psi)$ which relates to the true density $p(\theta|y, \psi)$. We follow standard practice in developing a Gaussian approximating model for y with the assumption of given values for θ and ψ and with the density denoted by $g(y|\theta, \psi)$. The approximating model is constructed in such a way that the first two moments of $p(y|\theta, \psi)$ and $g(y|\theta, \psi)$ are equal. Since $g(y|\theta, \psi)$ is Gaussian, it will be relatively straightforward to sample from $\tilde{p}(\theta|y, \psi) = g(\theta|y, \psi)$. The conditions under which this approximation converges almost surely to the true value as the number of simulations from the importance density increases are given by Geweke (1989, Theorem 1). An approximating Gaussian model for the SVM model is developed in Section 3.3.

Simulation smoothers such as the ones of de Jong and Shephard (1995) and Durbin and Koopman (2001b) can be used to sample from the importance density $g(\theta|y, \psi)$ in the case of SV models.

The likelihood function (11) is rewritten as

$$L(\psi) = \int p(y|\theta, \psi)\frac{p(\theta|\psi)}{g(\theta|y, \psi)}g(\theta|y, \psi)\,\mathrm{d}\theta = \tilde{\mathrm{E}}\left\{p(y|\theta, \psi)\frac{p(\theta|\psi)}{g(\theta|y, \psi)}\right\} \tag{12}$$

where $\tilde{\mathrm{E}}$ denotes expectation with respect to the importance density $g(\theta|y, \psi)$. Expression (12) can be simplified using a suggestion of Durbin and Koopman (1997). The likelihood function associated with the importance density is given by

$$L_g(\psi) = g(y|\psi) = \frac{g(y, \theta|\psi)}{g(\theta|y, \psi)} = \frac{g(y|\theta, \psi)p(\theta|\psi)}{g(\theta|y, \psi)} \tag{13}$$

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 (12) and substitution leads to

$$L(\psi) = L_g(\psi)\tilde{\mathrm{E}}\left\{\frac{p(y|\theta, \psi)}{g(y|\theta, \psi)}\right\} \tag{14}$$

which is the convenient expression we will use in our calculations. The likelihood function of the approximating Gaussian model can be calculated via the Kalman filter and the two conditional densities are easy to compute given a value for θ. It follows that the likelihood function of the SVM 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 SVM model is

$$\hat{L}(\psi) = L_g(\psi)\overline{w} \tag{15}$$

where

$$\overline{w} = \frac{1}{M}\sum_{i=1}^{M} w_i \qquad w_i = \frac{p(y|\theta^i, \psi)}{g(y|\theta^i, \psi)} \tag{16}$$

and θ^i denotes a draw from the importance density $g(\theta|y, \psi)$. The accuracy of this estimator depends on the properties of the so-called weights w_i; see Geweke (1989). Since the simulation samples are independent of each other, it follows immediately that the variance due to simulation decreases as M increases. 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

$$\ln \hat{L}(\psi) = \ln L_g(\psi) + \ln \overline{w} + \frac{s_w^2}{2M\overline{w}^2} \tag{17}$$

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

3.3. Approximating Gaussian Model Used for Importance Sampling

The approximating model is based on a linear Gaussian model with mean $\mathrm{E}(y_t) = h_t + c_t$ and variance $\mathrm{V}(y_t) = H_t$, that is,

$$y_t = h_t + u_t \qquad u_t \sim \mathrm{N}(c_t, H_t) \qquad t = 1, \ldots, n \tag{18}$$

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 (18) and by the true model, under consideration, (10) are as close as possible.[6]

We achieve this by equalizing the first and second derivatives of $p(y|\theta, \psi)$ and $g(y|\theta, \psi)$ with respect to θ at $\hat{\theta} = \tilde{\mathrm{E}}(\theta) = \int \theta g(\theta|y, \psi)\,\mathrm{d}\theta$. 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 (18). The conditional densities are given by

$$p(y|\theta, \psi) = \Pi_{t=1}^n p_t \qquad g(y|\theta, \psi) = \Pi_{t=1}^n g_t \tag{19}$$

with

$$\begin{aligned} p_t &= p(y_t|h_t, \psi) = -0.5[\ln 2\pi\sigma^2 + h_t + \exp(-h_t)\sigma^{-2}\{y_t - d\exp(h_t)\}^2] \\ g_t &= g(y_t|h_t, \psi) = -0.5\{\ln 2\pi + \ln H_t + H_t^{-1}(y_t - c_t - h_t)^2\} \end{aligned} \tag{20}$$

Differentiating both densities twice with respect to h_t gives

$$\begin{aligned} \dot{p}_t &= -0.5[1 + \sigma^{-2}\{d^2\sigma^{*2}\exp(h_t) - y_t^2\exp(-h_t)\}] \\ \ddot{p}_t &= -0.5\sigma^{-2}[d^2\sigma^{*2}\exp(h_t) + y_t^2\exp(-h_t)] \\ \dot{g}_t &= H_t^{-1}(y_t - c_t - h_t) \\ \ddot{g}_t &= -H_t^{-1} \end{aligned}$$

Equalizing the first and second derivatives, that is, $\dot{p}_t = \dot{g}_t$ and $\ddot{p}_t = \ddot{g}_t$ for $t = 1, \ldots, n$, leads to

$$\begin{aligned} c_t &= y_t - h_t + 0.5H_t[1 + \sigma^{-2}\{d^2\sigma^{*2}\exp(h_t) - y_t^2\exp(-h_t)\}] \\ H_t &= 2\sigma^{*2}/[d^2\sigma^{*2}\exp(h_t) + y_t^2\sigma^{*-2}\exp(-h_t)] \end{aligned}$$

The resulting model for $\tilde{y}_t = y_t - c_t$ is equivalent to

$$\tilde{y}_t = h_t + \tilde{u}_t \qquad \tilde{u}_t \sim \mathrm{N}(0, H_t) \qquad t = 1, \ldots, n$$

with

$$\tilde{y}_t = h_t - \frac{\sigma^{*2} + d^2\sigma^{*2}\exp(h_t) - y_t^2\sigma^{*-2}\exp(-h_t)}{d^2\sigma^{*2}\exp(h_t) + y_t^2\sigma^{*-2}\exp(-h_t)} \quad H_t = \frac{2\sigma^2}{d^2\sigma^{*2}\exp(h_t) + y_t^2\sigma^{*-2}\exp(-h_t)}$$

[6] Note that the true model describes a non-linear relationship between y_t and h_t; the approximating (linear) model is effectively a second-order Taylor expansion of the true model around h_t. Further, the multivariate Gaussian density $g(\theta|y, \psi)$ can be regarded as a Laplace approximation to the true density $p(\theta|y, \psi)$.

It should be noted that $H_t > 0$ for any value of h_t. We cannot solve out for $\tilde{y}_t$ and H_t at $\hat{h}_t = \tilde{\mathrm{E}}(h_t)$ because $\tilde{\mathrm{E}}$ refers to expectation with respect to the approximating model which depend on h_t. However, such a complicated but linear system of equations is usually solved iteratively by starting with a trial value $h_t = h_t^*$. Computing $\tilde{y}_t$ and H_t based on h_t^* and applying the Kalman filter smoother to model (18) leads to a smoothed estimate for h_t which can be used as a new trial value for h_t. Recomputing $\tilde{y}_t$ and H_t based on this new trial value leads to an iterative procedure which converges to $\hat{h}_t$. Note that the first and second derivatives of the true and approximating densities are equal at $h_t = \hat{h}_t$. More details are given by Durbin and Koopman (1997). It is worth mentioning that $\hat{h}_t$ is equal to the mode of $p(h_t|y, \psi)$ which can be of interest.

When we consider the full model (1), with σ_t given by equations (5) and (6) and μ_t by (8), the likelihood function is only affected in the squared error term. The last term of the definition of p_t in (20) is replaced by the term

$$\exp(-h_t)\sigma^{-2}\{y_t - a - by_{t-1} - d\exp(h_t)\}^2$$

We observe that the extensions do not change the stochastic process for h_t. Therefore, the simulation scheme for computing the Monte Carlo likelihood remains the same. However, the approximating model changes slightly; that is, $\dot{p}_t$ changes but $\ddot{p}_t$ does not change. In other words, the definition for c_t changes but the definition for H_t does not. Finally, numerical maximisation of the Monte Carlo likelihood is now also with respect to the parameters a and b.

3.4. Monte Carlo Evidence of Estimation Procedure

In this section we present some results of a Monte Carlo study which is carried out to investigate the small sample performance of the estimation procedure presented in Section 3.2. In short, we generate K simulated SVM series for the model presented in Section 3.1 and for some given 'true' parameter vector ψ. Subsequently, we treat ψ as unknown and estimate it for each series using the maximum likelihood method described in Section 3.2. We compute the sample mean and standard deviation together with a histogram for each element in ψ and compare it with the 'true' parameter value.

The details of the likelihood estimation procedure are as follows. For a given parameter vector ψ, we obtain the approximating Gaussian model as described in Section 3.3. The Gaussian loglikelihood function of the approximating model $\ln L_g(\psi)$ can be computed using the standard Kalman filter; see, for example, Durbin and Koopman (2001a). We have used the simulation smoothing method of de Jong and Shephard (1995) to generate the importance samples $\theta^i (i = 1, \ldots, M)$ but recently a simpler method has been developed by Durbin and Koopman (2001b) which can also be used for this purpose. The polar method as implemented in Ox by Doornik (1998) is used for generating standard normal deviates which are required as input for the simulation smoother to obtain samples θ^i from $g(\theta|y, \psi)$. The random number generator needs to be initialized by some fixed value. The importance samples θ^i allow us to compute weight w_i using (16) for $i = 1, \ldots, M$ where M is set equal to 200 for all our calculations. For each sample θ^i, three antithetic variables are computed in the way described by Durbin and Koopman (1997). This leads to a total of 800 weights for each likelihood evaluation. The mean $\overline{w}$ and variance s_w^2 of these weights are then used to compute the importance sampling estimate of the loglikelihood value (17). For the purpose of estimating the parameter vector ψ, this likelihood estimation procedure is used repeatedly for different values of vector ψ. It is noted that the random number generator is initialized by the same fixed value for each loglikelihood computation so that the numerical search procedure for the maximum likelihood is not affected by randomness. The BFGS maximization method is used to maximize the simulated likelihood function with respect to ψ; the BFGS method is documented by, among others, Fletcher (1987) and it is implemented in Ox by Doornik (1998). This estimation procedure worked satisfactorily in our simulation study below as well as in our empirical study of Section 4. In our implementation, a single likelihood evaluation of an SVM model with 5000 observations took 0.7 second on a Pentium III 800 Mhz when employing $M = 200$ importance samples. Maximum likelihood estimation of parameters was, on average, completed in 20 seconds.[7]

Parameter estimation is not with respect to vector ψ as defined in Section 3.1, but with respect to the transformed parameter vector ψ^*. The autoregressive parameter ϕ is restricted to have a value between zero and one; therefore we estimate ψ_1^* where

$$\phi = \psi_1 = \frac{\exp(\psi_1^*)}{1 + \exp(\psi_1^*)} \qquad \psi_1^* = \ln \frac{\phi}{1 - \phi}$$

Further, we estimate the log variance σ^{*2} and the log standard deviation σ_η. The mean parameter d is estimated without transformation.

We start by considering the standard SV model, that is, the SVM model of Section 3.1 with $d = 0$. Therefore, the last element of ψ is omitted. For generating Monte Carlo samples, the 'true' parameter values are set to

	ψ	ψ^*
$\psi_1 = \phi$	0.97	3.5
$\psi_2 = \sigma_\eta$	0.135	−2
$\psi_3 = \sigma^{*2}$	0.549	−0.3

which are typical values found in our empirical study or Section 4.

The Monte Carlo results for the basic SV model are similar but slightly better compared to results presented in similar studies of Jacquier *et al.* (1994) and Sandmann and Koopman (1998). Note that in these studies the parameter values were not transformed and that the estimation procedures used were different from ours. The results given in Figure 1 are for the typical sample size $n = 5000$ with the number of iterations set to $K = 500$. The graphical output includes a histogram of the estimated parameter values and an estimated density function which is computed using a standard non-parametric Gaussian kernel method.

The sample mean and standard deviation of the K estimated coefficients are given by

	'True'	Mean	Stand.dev	Mean asym.stand.err
ψ_1^*	3.5	3.490	0.215	0.225
ψ_2^*	−2.0	−2.020	0.112	0.111
ψ_3^*	−0.3	−0.301	0.0338	0.0337

[7] The estimation procedure is implemented using the object-oriented matrix programming language Ox 2.1 of Doornik (1998, `www.nuff.ox.ac.uk/Users/Doornik/`) and the SsfPack 2.2 procedures of Koopman *et al.* (1999, `www.ssfpack.com`). Relevant programs for the estimation of the SVM model can be downloaded from `www.feweb.vu.nl/koopman/sv/`

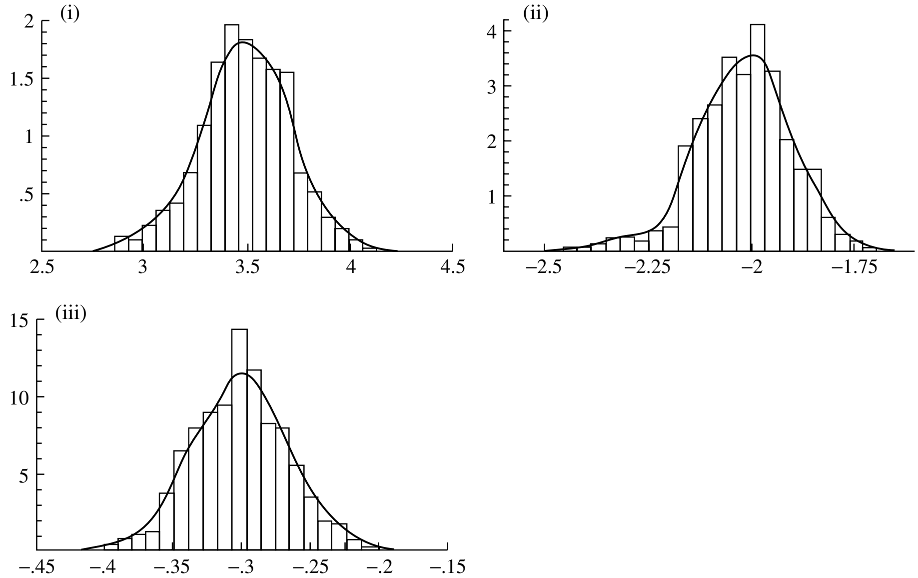

Figure 1. Monte Carlo results for standard SV model. Histograms and estimated densities (solid lines) of the maximum likelihood parameter estimates are presented for the SV model with (i) $\psi_1^* = 3.5$, (ii) $\psi_2^* = -2$ and (iii) $\psi_3^* = -0.3$. The Monte Carlo experiment is based on $K = 500$ iterations and sample size $n = 5000$

The last column contains the averages of the asymptotic standard errors of the estimates and they are conveniently close to the sample standard deviations of the estimates. The Monte Carlo results can also be presented in terms of vector ψ; we note that the resulting confidence intervals are asymmetric due to the nonlinear transformations. We obtain

	Mean	LHS '95% CI'	RHS '95% CI'
$\psi_1 = \phi = 0.97$	0.970	0.955	0.981
$\psi_2 = \sigma_\eta = 0.135$	0.133	0.107	0.165
$\psi_3 = \sigma^{*2} = 0.549$	0.547	0.480	0.625

where LHS is the left-hand side border and RHS is the right-hand side border of the 95% confidence interval. These results will be used as a benchmark for the Monte Carlo results of the SVM model.

We now turn our attention to the Monte Carlo evidence for the SVM model. We keep the 'true' parameters of the SV model and look at the results for a typical value of d, that is, $d = 0.1$. The Monte Carlo experiments are again based on $n = 5000$ and $K = 500$. The results for the SVM model with 'true' $d = 0.1$ are given in Figure 2 and the sample statistics are given by

	'True'	Mean	Stand.dev	Mean asym.stand.err
ψ_1^*	3.5	3.537	0.210	0.213
ψ_2^*	−2.0	−2.059	0.108	0.103
ψ_3^*	−0.3	−0.299	0.0341	0.0338
ψ_4^*	0.1	0.0960	0.0113	0.0115

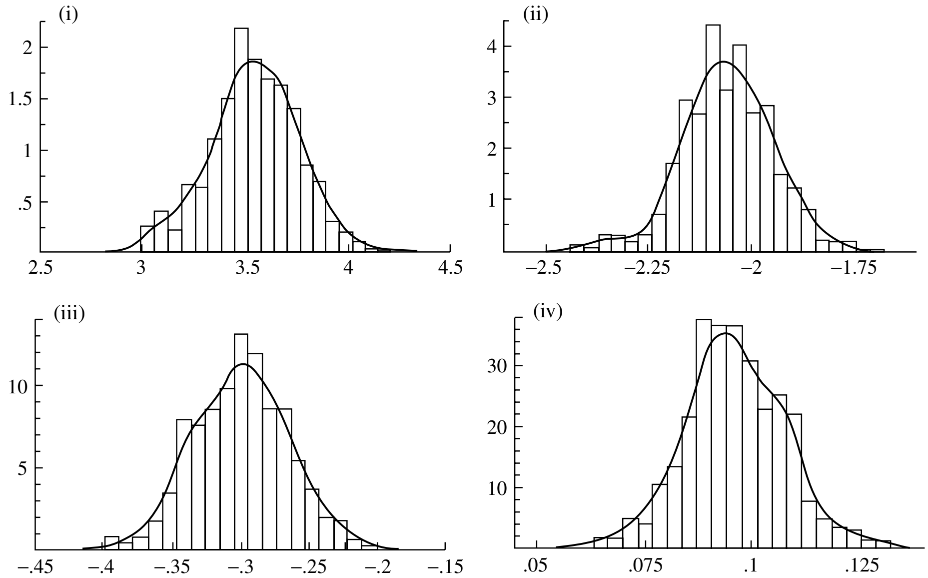

Figure 2. Monte Carlo results for the SVM model. Histograms and estimated densities (solid lines) of the maximum likelihood parameter estimates are presented for the SVM model with (i) $\psi_1^* = 3.5$, (ii) $\psi_2^* = -2$, (iii) $\psi_3^* = -0.3$ and (iv) $\psi_4^* = 0.1$. The Monte Carlo experiment is based on $K = 500$ iterations and sample size $n = 5000$

The standard deviation of the estimates for ψ^* obtained from the Monte Carlo samples are very close to the averages of the asymptotic standard errors of the estimates. This indicates that asymptotic standard errors can be used for estimates obtained from the methods of Section 3.2. The results in terms of ψ are given by

	Mean	LHS '95% CI'	RHS '95% CI'
$\psi_1 = \phi = 0.97$	0.972	0.958	0.981
$\psi_2 = \sigma_\eta = 0.135$	0.128	0.104	0.156
$\psi_3 = \sigma^{*2} = 0.549$	0.550	0.481	0.628
$\psi_4 = d = 0.1$	0.0960	0.0733	0.119

Comparing the results for the standard SV model, we conclude that the confidence intervals for ψ_1, ψ_2 and ψ_3 are very similar and that the in-mean parameter d can be accurately estimated with a relatively small standard deviation.[8]

Finally, we have assessed the variability of the importance sampling estimator due to the choice of different random numbers. This variability depends on the number of simulations M in (16) and by setting this value to 200 we obtained results which only varied marginally when different random numbers were used. By repeating the Monte Carlo experiment for different values of n we found that the variability of the estimator increases when n gets larger but not dramatically.

[8] We repeated this experiment with various values for d and the outcome only changed marginally. Similar overall results were obtained for simulations based on the smaller sample size $n = 500$; these can be found in Koopman and Hol Uspensky (2000).

4. EMPIRICAL EVIDENCE FROM INTERNATIONAL STOCK MARKETS

4.1. Some Theory on the Relationship between Returns and Volatility

The relation between expected stock index returns and conditional volatility has received much attention in the financial economics literature. Although a positive relationship between expected returns and expected volatility is consistent with the Capital Asset Pricing Model (CAPM) and intuitively appealing, as rational risk-averse investors require higher expected returns during more volatile periods, empirical research has been unable to establish a convincing positive relationship between the expected risk premium and conditional volatility using GARCH-M models.[9] Instead, there appears to be stronger evidence of a negative relationship between unexpected returns and innovations to the volatility process which French *et al.* (1987) interpret as indirect evidence of a positive correlation between the expected risk premium and *ex ante* volatility. They reason that unanticipated large shocks to the return process, which can be caused by either good or bad news, induce higher expected volatility for future time periods. If expected volatility and expected returns are positively related and future cash flows are unaffected, the current stock index price should fall. Conversely, small shocks to the return process lead to an increase in contemporaneous stock index prices. This theory, known as the volatility feedback theory, therefore hinges on two assumptions: first, the existence of a positive relation between the expected components of the return and the volatility process and second, volatility persistence. An alternative explanation for asymmetric volatility where causality runs in the opposite direction is the leverage effect put forward by Black (1976) who asserts that a negative (positive) return shock leads to an increase (decrease) in the firm's financial leverage ratio which has an upward (downward) effect on the volatility of the stock returns. However, it has been argued by Black (1976), Christie (1982), French *et al.* (1987) and Schwert (1989) that leverage alone cannot account for the magnitude of the negative relationship.[10] If both volatility feedback and leverage effects are present then large pieces of bad news are associated with an increase in volatility, whereas the net impact of large pieces of good news is unclear. The reverse then holds for small shocks to the return process as small positive shocks are associated with a decrease in volatility and the net effect of small negative shocks is unknown.

In the GARCH literature the leverage effect has been empirically established with the EGARCH model of Nelson (1991) and the GJR-GARCH model developed by Glosten *et al.* (1993). In these models the conditional volatility at time $t + 1$ is allowed to respond asymmetrically to unanticipated rises and falls in the stock price that occurred at time t. SV models, unlike GARCH models, define volatility as truly contemporaneous and hence its volatility measure includes not only expected but also unexpected volatility. The correlation between unexpected shocks to the return and the volatility process, which can both be observed at time t, could then be measured by estimating $corr(\varepsilon_t, \eta_t)$ as an additional parameter in the SV model.[11]

---

[9] See e.g. for the US stock market French *et al.* (1987) and Campbell and Hentschel (1992) who observe a positive relation, whereas Glosten *et al.* (1993) who develop a much richer asymmetric GARCH-M model present evidence of a negative relation, as does Nelson (1991) with his EGARCH model. Poon and Taylor (1992) who study the UK stock market report a weak positive relationship.

[10] Campbell and Hentschel (1992) find evidence of both volatility feedback and leverage effects, whereas Bekaert and Wu (2000) present results which strongly favour the volatility feedback hypothesis.

[11] Jacquier *et al.* (2001) estimate $corr(\varepsilon_t, \eta_t)$ and report a negative relationship between contemporaneous unexpected stock index returns and unexpected volatility. Harvey and Shephard (1996), on the other hand, estimate $corr(\varepsilon_t, \eta_{t+1})$ and Watanabe (1999) develops an SV model which includes the lagged shock to the return process as an explanatory variable

The general mean equation with time-varying variance we consider for estimation is

$$y_t = a + by_{t-1} + d\sigma_t^2 + \sigma_t \varepsilon_t \tag{21}$$

where y_t denotes the excess returns on the stock index at time t and σ_t^2 is the variance which is contemporaneous for the SV model and conditional for the GARCH model. For GARCH models the d parameter therefore measures the relation between expected returns and expected volatility, whereas in the case of SV models the d parameter estimate contains information about the relation between returns, on the one hand, and expected and unexpected, or *ex post*, volatility, on the other. For the GARCH(1,1)-M model the conditional mean is defined as

$$\mathrm{E}_{t-1}(y_t) = a + by_{t-1} + d\{w + \alpha(\sigma_{t-1}\varepsilon_{t-1})^2 + \beta\sigma_{t-1}^2\} \tag{22}$$

and for the SVM model the equivalent notation is given by

$$\begin{aligned}\mathrm{E}_{t-1}(y_t) &= a + by_{t-1} + d\{\sigma^{*2}\exp(\phi h_{t-1})\exp(\mathrm{E}_{t-1}(\eta_t))\} \\ &= a + by_{t-1} + d\{\sigma^{*2}\exp(\phi h_{t-1})\}\end{aligned} \tag{23}$$

It is evident from equations (22) and (23) that all the explanatory variables in the conditional mean of the GARCH-M(1,1) are known at time $t-1$, whereas the shock to the volatility process at time t, denoted by η_t, prevents observation of the σ_t^2 term in the SVM model at time $t-1$. The mean equation for the SVM model can then be written as

$$\begin{aligned}y_t &= \mathrm{E}_{t-1}(y_t) + d\{\sigma^{*2}\exp(\phi h_{t-1})[\exp(\eta_t) - \exp(\mathrm{E}_{t-1}(\eta_t))]\} + \sigma_t\varepsilon_t \\ &= \mathrm{E}_{t-1}(y_t) + d\{\sigma^{*2}\exp(\phi h_{t-1})[\exp(\eta_t) - 1]\} + \sigma_t\varepsilon_t\end{aligned} \tag{24}$$

where the second term on the right-hand side is d multiplied by the *unexpected* volatility at time t. The d coefficient in the SVM model therefore measures the relation not only between the expected components of the return and the volatility process, but also between the unexpected components as equations (23) and (24) can be combined and written as

$$y_t = a + by_{t-1} + d\sigma_{t|t-1}^2 + d\{\sigma_t^2 - \sigma_{t|t-1}^2\} + \sigma_t\varepsilon_t \tag{25}$$

where $\sigma_{t|t-1}^2$ denotes the conditional variance at time t given the information available at time $t-1$, or the expected volatility, and σ_t^2 the contemporaneous volatility measure at time t. The $\{\sigma_t^2 - \sigma_{t|t-1}^2\}$ term denotes the unexpected shock to the volatility process which should not be related to the predictable components. The volatility feedback effect is then measured by the d parameter preceeding the $\{\sigma_t^2 - \sigma_{t|t-1}^2\}$ term which is expected to be negative as large (small) shocks to the return process raise (lower) contemporaneous volatility through η_t, presumably irrespective of the sign of ε_t, inducing a drop (increase) in the current stock index price in the case of initial large negative (small positive) return shocks which are amplified, whereas initial large positive and small negative shocks are dampened. As deterministic GARCH models do not contain an unexpected volatility component, that is, $\sigma_t^2 = \sigma_{t|t-1}^2$, the issue does not arise for this class of volatility models and the d parameter only measures the relation between the expected returns and the expected volatility.

in the variance equation allowing for an asymmetric response. Both studies report negative coefficients for the relation between current unexpected returns and *future* volatility.

4.2. Data

The data we analyse includes daily stock index returns from three international stock markets: the United Kingdom, the United States and Japan. The UK Financial Times All Share Index and the US Standard and Poor's Composite stock index series cover the period 1 January 1975 to 31 December 1998 whereas the Japanese Topix series starts on 1 January 1988 and ends at 31 December 1998. The stock data was obtained from Datastream. From the same data source we also collected daily UK and Japanese 1-month Treasury bill rates; the US 3-month Treasury bill rate data was extracted from the on-line Federal Reserve Bank of Chicago Statistical Release H.15 database. These interest rate series are used as proxies for the risk free rate of return. The stock index prices are in local currencies and not adjusted for dividends following studies of French *et al.* (1987) and Poon and Taylor (1992) who found that inclusion of dividends affected estimation results only marginally. Returns are calculated on a continuously compounded basis and expressed in percentages, they are therefore calculated as $R_t = 100(\ln P_t - \ln P_{t-1})$ where P_t is the price of the stock market index at time t. From these returns we subtract the daily risk free rate multiplied by 100, denoted by Rf_t, in order to obtain the excess returns which are therefore defined as $y_t = R_t - Rf_t$.

In this section we model the behaviour of five series: we consider daily excess return series on the UK and US index that cover a period of 24 years ending in 1998, as well as 11 year sub-samples of these two series together with excess returns on the Japanese stock market index. These shorter

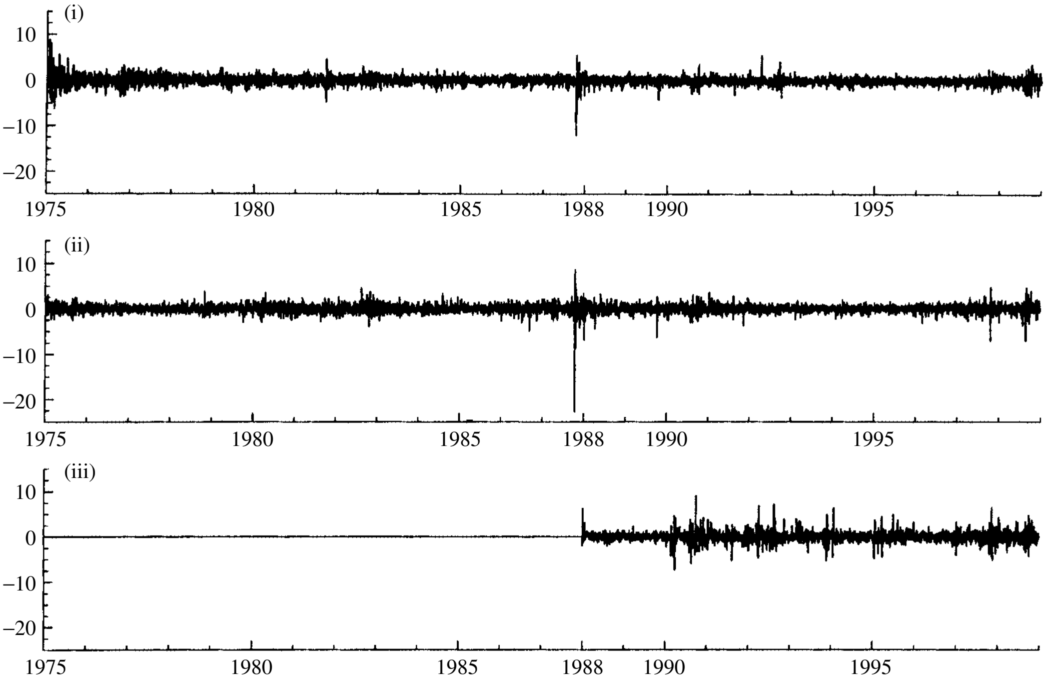

Figure 3. Excess returns for the (i) FT All Share Index (UK) and (ii) S&P Composite Stock Index (US) between 02/01/75 and 31/12/98 and for the (iii) Topix Stock Index (Japan) between 4 January 1988 and 31 December 1988

series start in 1988 and therefore exclude the extreme negative observations relating to the 1987 stock market crash. Figure 3 contains graphs of the excess return series and the accompanying summary statistics are presented in Table I.

We observe that the effects of the October 1987 crash are especially pronounced for the US stock market where excess returns on the Standard & Poor's Composite index fell by nearly 23% on one single trading day. This one observation contributes to a great extent to the large excess kurtosis value of 62.758 and the high negative skewness coefficient of −2.562. The most volatile series of the five is the Topix series which can not be attributed to one extreme movement, as can be seen in Figure 3, but to several prolonged periods of market turbulence initiated in the early 1990s by the collapse of the Japanese asset market. The Topix series is further characterized by a negative mean and is positively skewed, which are features not typically found in a stock index (excess) return series. We further observe that the UK excess returns and squared excess returns for the period starting in 1975 are highly autocorrelated at lag 1 but that these values are much lower and comparable with those of the Topix stock index for the subsample period 1988–1998. First-order serial correlation coefficients for the Standard & Poor's Composite Index excess returns, on the other hand, are relatively low for both the full and the sub-sample period. In the case of excess returns high first-order autocorrelation reflects the effects of non-synchronous or thin trading, whereas highly correlated squared returns can be seen as an indication of volatility clustering. The $Q(12)$ test statistic, which is a joint test for the hypothesis that the first twelve autocorrelation coefficients are equal to zero, indicates that this hypothesis has to be rejected at the 1% significance level for all excess return and squared excess return series.

Table I. Summary statistics of daily excess returns

Period	1975–1998		1988–1998		
Number of observations T	6261		2869		
Stock index	FT All	S&P	FT All	S&P	Topix
Mean	0.033	0.028	0.017	0.042	−0.025
Variance	0.943	0.874	0.584	0.747	1.357
Skewness	−0.194	−2.562	−0.022	−0.664	0.343
Excess kurtosis	11.828	62.758	3.491	7.954	6.107
N	36536	1034312	1458	7774	4516
Excess returns					
$\hat{\rho}_1$	0.167	0.054	0.115	0.004	0.100
$\hat{\rho}_2$	0.008	−0.024	−0.002	−0.013	−0.062
$\hat{\rho}_3$	0.037	−0.021	−0.005	−0.041	−0.009
$\hat{\rho}_4$	0.046	−0.024	0.041	−0.016	0.027
$\hat{\rho}_5$	0.019	0.032	0.009	0.006	−0.030
$Q(12)$	262.08	41.94	61.32	29.86	64.86
Squared excess returns					
$\hat{\rho}_1$	0.478	0.112	0.163	0.176	0.163
$\hat{\rho}_2$	0.281	0.149	0.155	0.087	0.161
$\hat{\rho}_3$	0.238	0.077	0.136	0.049	0.118
$\hat{\rho}_4$	0.290	0.020	0.111	0.087	0.173
$\hat{\rho}_5$	0.202	0.137	0.109	0.097	0.179
$Q(12)$	4543.79	404.14	560.52	286.52	527.91

Note: N is the χ^2 normality test statistic with 2 degrees of freedom; $\hat{\rho}_l$ is the sample autocorrelation coefficient at lag l with asymptotic standard error $1/\sqrt{T}$ and $Q(l)$ is the Box–Ljung portmanteau statistic based on l-squared autocorrelations.

4.3. Estimation Results for the SVM Model and Some Diagnostics

Our main objective in this empirical section is to estimate the contemporaneous relationship between excess returns on stock market indices and their volatility with our SVM model, which we already defined in equations (1), (5), (6) and (8) as

$$\begin{aligned} y_t &= a + by_{t-1} + d\sigma_t^2 + \sigma_t\varepsilon_t && \varepsilon_t \sim \mathrm{NID}(0, 1) \\ \sigma_t^2 &= \sigma^{*2}\exp(h_t) \\ h_t &= \phi h_{t-1} + \sigma_\eta \eta_t && \eta_t \sim \mathrm{NID}(0, 1) \end{aligned}$$

and from equation (25) we recall that $y_t = a + by_{t-1} + d\sigma^2_{t|t-1} + d\{\sigma_t^2 - \sigma^2_{t|t-1}\} + \sigma_t\varepsilon_t$.

Table II reports the SVM model estimation results for the stock index series over the full sample period 1975–1998 and the subsample period 1988–1998. Volatility persistence estimates for the five series are all highly significant and quite similar with values for ϕ ranging from 0.966 for the post-crash Topix to 0.984 for the full sample Financial Times All Share Index, the series which displayed the highest degree of sample autocorrelation in the squared returns. This near-unity volatility persistence for high-frequency data is consistent with findings from both the SV and the GARCH literature. The two remaining volatility process parameters cover a much wider range. The highest values for the scaling parameter and the parameter which measures the variation in the volatility process are observed for the Topix series with $\sigma^{*2} = 0.832$ and $\sigma_\eta^2 = 0.058$. The SVM

Table II. Estimation results for the SVM model

Period	1975–1998				1988–1998					
No. of obs.	6261				2869					
Stock Index	FT All		S&P		FT All		S&P		Topix	
a	0.038		0.045		0.061		0.074		0.019	
	0.014	0.063	0.017	0.073	0.016	0.106	0.038	0.111	−0.014	0.052
b	0.146		0.074		0.100		0.024		0.099	
	0.123	0.172	0.054	0.101	0.068	0.144	0.009	0.062	0.067	0.143
d	−0.011		−0.023		−0.085		−0.046		−0.031	
	−0.049	0.028	−0.065	0.019	−0.176	0.006	−0.103	0.011	−0.066	0.005
σ^{*2}	0.615		0.597		0.458		0.539		0.832	
	0.498	0.758	0.508	0.701	0.358	0.587	0.427	0.682	0.642	1.077
ϕ	0.984		0.979		0.976		0.970		0.966	
	0.977	0.990	0.969	0.986	0.958	0.986	0.954	0.981	0.947	0.978
σ_η^2	0.018		0.021		0.019		0.035		0.058	
	0.014	0.025	0.015	0.029	0.012	0.032	0.025	0.050	0.041	0.082
AIC	15075.6		15005.6		6090.7		6610.7		7988.7	
$Q(12)$	24.45		20.06		7.44		21.79		10.30	
N	16.265		18.578		2.326		9.969		11.643	
σ_∞^2	0.825		0.760		0.560		0.727		1.277	

Notes: Parameter estimates are reported together with the asymptotic 95% confidence interval which are asymmetric for b σ^{*2}, ϕ and σ_η^2; AIC is the Akaike Information Criterion which is calculated as $-2(\ln L) + 2\mathrm{p}$ and $Q(\ell)$ 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 for ε_t with 2 degrees of freedom; σ_∞^2 denotes the unconditional variance as implied by the volatility process.

model therefore captures the more erratic behaviour of the Topix quite well through a combination of parameters: the high scaling parameter indicates a higher level of volatility whereas the relative small value for ϕ and the large value for σ_η^2 imply that its volatility process is less predictable than that of the other four series.

The three parameters which govern the mean process are reported in the first three rows together with their 95% confidence intervals. We observe that the mean parameter a is always positive and statistically significant for all series with the exception of the Topix series which has a negative sample mean; see Table I. We note, however, that we have simultaneously estimated the constant a and the in-mean coefficient d where the latter is associated with a regressor which is strictly positive at all times. The estimates of b are statistically significant for all series and very similar to the first-order autocorrelation coefficients reported in Table I. The d parameter which measures both the *ex ante* relationship between returns and volatility and the volatility feedback effect is negative for all series, although the hypothesis of d equal to zero can never be rejected at the conventional 5% significance level. For some of the series d is, however, very close to being statistically significant. A similar SVM model was estimated by Fridman and Harris (1998), who studied daily returns on the Standard & Poor's index over the period 1980 to 1987, and Watanabe (1999), who examined the daily Topix series over the eight-year period 1990–1997. Both studies reported significant positive values for the contemporaneous relationship.[12] Our findings here are more in line with those of French *et al*. (1987) who regress monthly excess returns of a US stock portfolio against expected and unexpected volatility obtained with ARIMA models based on daily data. For the regression which excludes unexpected volatility they observe a weak positive relation between expected returns and volatility. Inclusion of both volatility measures, however, not only results in a highly significant negative relation between the unexpected components but also turns the sign for the *ex ante* relationship which becomes weakly negative. The negative relation between the unexpected components therefore dominates the weaker, presumably positive, relation between the expected components. We further observe that the largest negative values for d are found for the shorter samples where they are also closest to being statistically significant. A possible explanation is that the positive *ex ante* relation is more pronounced for the period following the stock market crash inducing a stronger volatility feedback effect which forces the d parameter further downward. Such an interpretation is also supported by the findings of Campbell and Hentschell (1992) who incorporate volatility feedback into a GARCH-M model where the volatility feedback parameter is restricted so that a large positive value for the *ex ante* relation implies a large volatility feedback effect.[13] In their study of the US stock market they find that estimates of the unrestricted volatility feedback parameter are very similar to those of the restricted version. French *et al*. (1987) further point out that the volatility feedback effect would be especially strong and dominant when volatilities are highly autocorrelated. As a consequence, rational risk-averse investors require even higher expected returns when unanticipated increases in future volatility are highly persistent. This would be consistent with our findings for the post-crash samples where higher values for ϕ are combined with larger negative values for the in-mean parameter. The relatively large negative estimates for the d parameter we therefore interpret as evidence of the existence of convincing negative feedback effects which appear especially strong

[12] Fridman and Harris (1998) did not, however, allow for a constant in the mean and the likelihood ratio test for $d = 0$ amounted to a value of 0.14. We also estimated the SVM model with constant a restricted to zero and found that estimates for d were forced upward to a positive value except for the Topix series which has a negative sample mean; see Table I.
[13] Also see Campbell *et al*. (1997, pp. 497–498).

when volatility is persistent. This then provides indirect evidence of a positive intertemporal relation between expected excess market returns and its volatility as this is one of the assumptions underlying the volatility feedback hypothesis. Therefore, in the SVM model a negative in-mean parameter indicates that when investors expect higher persistent levels of volatility in the future they require compensation for this in the form of higher expected returns.

With regard to the distributional assumptions we observe that the standardized error term ε_t abides the normality assumption only for the Financial Times post-crash period and the $Q(12)$ statistic indicates that little serial correlation remains in the standardized error term. The hypothesis that the first twelve autocorrelation coefficients of ε_t are equal to zero can also not be rejected for the Topix series as the critical value at the 5% significance level is 12.6. Eventhough the remaining values for the $Q(12)$ and normality statistics exceed their critical values, they are much smaller than those observed in Table I, especially the values for the normality statistic are substantially reduced.

In addition to the SVM model we also estimated a number of restricted versions of our model which were obtained by restricting one or more of the mean parameters to be equal zero.[14] The likelihood ratio tests for the hypothesis $H_0 : d = 0$ never exceeded the critical χ_1^2 5% significance value of 3.84, confirming the insignificance of the d parameter. The estimates of the remaining parameters for this restricted model only changed marginally with the exception of the estimates of the a parameter which decreased in value for all series. More importantly, however, the likelihood ratio tests for the null hypothesis $H_0 : a = b = d = 0$ were all significant at the 1% significance level. This was confirmed by the values of the AIC statistic, which is a goodness-of-fit statistic that allows for comparison between different models with different numbers of parameters. Therefore we conclude that the SVM model is a useful empirical model for the modelling of the relation between current returns and returns of the previous period together with the feedback effect of current volatility. An additional advantage of modelling the mean is that ε_t behaved better, especially in terms of the assumption of zero autocorrelation.

4.4. Some Comparisons with GARCH-M Estimation Results

The GARCH model we estimate is the GARCH-M(1,1) model defined in equations (1), (4) and (9) as

$$\begin{aligned} y_t &= a + by_{t-1} + d\sigma_t^2 + \sigma_t\varepsilon_t \qquad \varepsilon_t \sim \mathrm{NID}(0, 1) \\ \sigma_t^2 &= \omega + \alpha(\sigma_{t-1}\varepsilon_{t-1})^2 + \beta\sigma_{t-1}^2 \end{aligned}$$

As research in the empirical GARCH literature has shown that the assumption of normally distributed error terms is often violated we estimate the GARCH-M(1,1) model in this section with ε_t following a Student-t instead of a Gaussian distribution.[15] Given our results in the previous section the need to impose an alternative error distribution for the SV class of volatility models is less evident and attributable to the fact that SV models are by definition better suited to incorporate extreme values.

[14] For the estimation results of $d = 0$ and $a = b = d = 0$ we refer to the JAE Data Archive http:www.econ.queensu.ca/jae or to the original discussion paper of Koopman and Hol Uspensky (2000).

[15] In our original discussion paper we estimated the GARCH-M model with normally distributed error terms and observed relatively high values for the χ^2 normality test statistic N compared to the ones reported in Table II.

The estimation results for the GARCH-M model are given in Table III where we observe estimates for the b parameter that are very similar to those obtained with the SVM model and that near-zero estimates for a are combined with positive values for the d parameter. The null hypothesis of a zero *ex ante* relationship between excess returns and volatility can, however, never be rejected at the 5% significance level, although in some cases only by a small margin. With regard to our estimation results we further observe that for the post-crash samples the relationship is strongest. In fact, the magnitude of the estimates for d are quite similar to those obtained with the SVM model in absolute terms, so a large negative contemporaneous relationship in the SVM model is accompanied by a large positive *ex ante* relationship in the GARCH-M model. This could be interpreted as confirmation of our hypothesis that a stronger *ex ante* relationship between the return and volatility process induces a more convincing volatility feedback effect which dominates the relation between expected returns and expected volatility.

The volatility persistence parameters are comparable to those found for the SVM model with near-unity values for the sum of α and β, although we observe that the persistence values for the Topix and the Standard & Poor's sample starting in 1988 are considerably higher when modelled with the GARCH-M model. They are in fact so high that they exceed those of the 1975–1998 Financial Times All Share Index series which exhibits very high autocorrelated squared returns as shown in Table I. We find that the GARCH-M model also captures the volatility in the variance equation of the Topix series as the α parameter, which measures the extent to which volatility at time t is influenced by a shock to the return process at time $t-1$, is highest for this series.

Table III. Estimation results for the GARCH-M model with t-distribution

Period	1975–1998		1988–1998		
No. of obs.	6261		2869		
Stock Index	FT All	S&P	FT All	S&P	Topix
a	0.023	0.007	0.000	0.024	−0.032
	−0.006 0.053	−0.026 0.041	−0.052 0.051	−0.021 0.070	−0.079 0.015
b	0.145	0.061	0.094	0.021	0.085
	0.120 0.170	0.037 0.086	0.057 0.132	−0.032 0.037	0.047 0.122
d	0.017	0.038	0.062	0.049	0.028
	−0.024 0.058	−0.010 0.086	−0.039 0.162	−0.021 0.120	−0.017 0.072
ω	0.014	0.008	0.009	0.004	0.028
	0.009 0.019	0.005 0.011	0.003 0.015	0.001 0.007	0.015 0.041
α	0.084	0.043	0.055	0.034	0.108
	0.070 0.099	0.034 0.052	0.037 0.074	0.022 0.046	0.080 0.135
β	0.897	0.947	0.928	0.962	0.876
	0.880 0.914	0.936 0.958	0.903 0.953	0.950 0.975	0.848 0.904
ν	11.914	6.046	8.704	4.572	4.955
	10.342 14.049	5.360 6.933	6.987 11.538	3.919 5.487	4.152 6.142
$\alpha+\beta$	0.982	0.990	0.984	0.996	0.983
$Q(12)$	38.53	16.16	11.81	19.53	14.42
σ^2_∞	0.792	0.775	0.561	0.985	1.689

Notes: Parameter estimates are reported together with the asymptotic 95% confidence interval which are all symmetric with the exception of those of the ν parameter; $Q(\ell)$ 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; σ^2_∞ denotes the unconditional variance as implied by the volatility process.

Imposing restrictions of the form $d = 0$ and $a = b = d = 0$ were also carried out for the GARCH-M model.[16] Again the estimates of the remaining parameters only changed marginally, with the exception of the a parameter which increased when we set d equal to zero. The likelihood ratio statistic indicated that the in-mean effect had little explanatory value as the hypothesis of $d = 0$ could not be rejected at the 5% significance level for any of the five samples. The hypothesis $a = b = d = 0$, on the other hand, had to be rejected for all series, which was consistent with our findings for the SVM model.

5. SUMMARY AND CONCLUSIONS

In this paper we have presented a Stochastic Volatility model where the mean is modelled simultaneously with the variance equation. When one of the variables in the mean is the volatility process itself, we obtain the Stochastic Volatility in Mean (SVM) model with which we are able to investigate the contemporaneous relationship between excess returns on a stock market index and its time-varying volatility. We estimate the parameters in our model using a special simulation-based maximum likelihood method and we also present results of a Monte Carlo experiment to show that if such an interdependence is present the SVM model is capable of detecting it.

For the empirical application we examined stock indices from the United Kingdom, the United States and Japan over two time periods. The results were then compared with the estimation results obtained for their GARCH counterparts. The conclusions of our empirical study can be summarized as follows. First, with our SVM model we find evidence of a weak negative relationship for all stock index series, whereas estimation with the GARCH-M model produces statistically insignificant positive estimates for the in-mean parameter. The difference in the sign for d is directly attributable to the fact that the SVM model, unlike the GARCH-M model, measures not only the relation between expected returns and expected volatility but also the volatility feedback effect where the latter appears to dominate the former. The largest negative values are then observed for those series which have a larger positive d estimate in the GARCH-M models and we assert that a strong *ex ante* positive relation induces a more convincing volatility feedback effect which then provides indirect evidence of the positive *ex ante* relationship between the return and the volatility process as this is one of the main assumptions underlying the volatility feedback hypothesis. Second, we find that simultaneous modelling of the mean and the variance equations lead to a better fit of the volatility series. The first-order autoregressive term b in the mean equation appears robust across model specifications and classes of volatility models. Although it is possible to model the original series prior to estimation with a volatility model, simultaneous estimation is more efficient. Finally, we observe that the volatility persistence parameter ϕ in the SV models, which is an indication of volatility clustering, is comparable with the persistence measure $(\alpha + \beta)$ of GARCH models. An advantage of SV models over GARCH models is that the distributional assumptions of the error term in the mean ε_t are much less violated. This makes the case for alternative error distributions and hence the estimation of an additional parameter less strong for the SV class of volatility models. On the basis of the above we therefore feel that SV models can be regarded as a competitive alternative to GARCH models, not only in theoretical terms but also in empirical research.

[16] These estimation results can be downloaded from the JAE Data Archive `http:www.econ.queensu.ca/jae`.

ACKNOWLEDGEMENTS

Most of this work was conducted while the first author was at CentER of Tilburg University and the second author was visiting CentER. In particular, we like to thank Theo Nijman for his support in this project. Further, we are indebted to the co-editor, three anonymous referees and the participants of the Financial Markets Group meeting 'Limits to Forecasting in Financial Markets' at the London School of Economics in June 1999 for helpful comments and suggestions. Also, for most of this project, the first author was a Research Fellow of the Royal Netherlands Academy of Arts and Sciences and its financial support is gratefully acknowledged. The second author would especially like to thank John Cadle (Birmingham University).

REFERENCES

Bekaert G, Wu G. 2000. Asymmetric volatility and risk in equity markets. *Review of Financial Studies* **13**: 1–42.

Bera AK, Higgins ML. 1993. ARCH models: properties, estimation and testing. *Journal of Economic Surveys* **7**: 305–366.

Black F. 1976. Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statistics Section* 177–181.

Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307–327.

Bollerslev T, Engle RF, Nelson DB. 1994. ARCH Models. In *Handbook of Econometrics* (Vol. 4), Engle RF, McFadden DL (eds). Elsevier Science: Amsterdam; 2959–3038.

Bollerslev T, Chou RY, Kroner KF. 1992. ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics* **52**: 5–59.

Campbell JY, Hentschel L. 1992. No news is good News, an asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* **31**: 281–318.

Campbell JY, Lo AW, MacKinlay AC. 1997. *The Econometrics of Financial Markets*, Princeton University Press: New Jersey.

Chesney M, Scott L. 1989. Pricing European currency options: a comparison of the modified Black–Scholes model and a random variance model. *Journal of Financial and Quantitative Analysis* **24**: 267–284.

Chib S, Nardari F, Shephard N. 1998. Markov chain Monte Carlo methods for generalized stochastic volatility models. Discussion Paper, Nuffield College, Oxford.

Christie AA. 1982. The stochastic behavior of common stock variances: value, leverage and interest rate effects. *Journal of Financial Economics* **10**: 407–432.

Danielsson J. 1994. Stochastic volatility in asset prices, estimation with simulated maximum likelihood. *Journal of Econometrics* **64**: 375–400.

de Jong P, Shephard N. 1995. The simulation smoother for time series models. *Biometrika* **82**: 339–350.

Diebold FX, Lopez JA. 1995. Modeling volatility dynamics. In *Macroeconometrics: Developments, Tensions and Prospects*, Hoover K (ed.). Kluwer Academic Publishers: Amsterdam; 427–472.

Doornik JA. 1998. *Object-Oriented Matrix Programming using Ox 2.0*, Timberlake Consultants Ltd: London. `http://www.nuff.ox.ac.uk/Users/Doornik/`

Durbin J, Koopman SJ. 1997. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* **84**: 669–684.

Durbin J, Koopman SJ. 2001a. *Time Series Analysis by State Space Methods*, Oxford University Press: Oxford.

Durbin J, Koopman SJ. 2001b. A simple and efficient simulation smoother for state space time series analysis. Discussion Paper, London School of Economics, Department of Statistics.

Engle RF. 1995. *ARCH: Selected Readings*. Oxford University Press: Oxford.

Engle RF, Lilien DM, Robins RP. 1987. Estimating time-varying risk premia in the term structure: the ARCH-M model. *Econometrica* **55**: 391–407.

Fletcher R. 1987. *Practical Methods of Optimisation*, 2nd edn, John Wiley: New York.

French KR, Schwert GW, Stambaugh RF. 1987. Expected stock returns and volatility. *Journal of Financial Economics* **19**: 3–29.

Fridman M, Harris L. 1998. A maximum likelihood approach for non-Gaussian stochastic volatility models. *Journal of Business and Economic Statistics* **16**: 284–291.

Gallant AR, Rossi PE, Tauchen G. 1992. Stock prices and volume. *Review of Financial Studies* **5**: 199–242.

Geweke J. 1989. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* **57**: 1317–1339.

Ghysels E, Harvey AC, Renault E. 1996. Stochastic volatility. In *Handbook of Statistics* (Vol. 14), Statistical Methods in Finance, Maddala GS, Rao CR (eds). North-Holland: Amsterdam; 128–198.

Glosten LR, Jagannathan R, Runkle DE. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**: 1779–1801.

Harvey AC. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press: Cambridge.

Harvey AC, Ruiz E, Shephard N. 1994. Multivariate stochastic variance models. *Review of Economic Studies* **61**: 247–264.

Harvey AC, Shephard N. 1996. Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics* **14**: 429–434.

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

Jacquier E, Polson NG, Rossi PE. 1994. Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics* **12**: 371–417.

Jacquier E, Polson NG, Rossi PE. 2001. An extended stochastic volatility model with fat-tails and leverage effect: Bayesian analysis. Working Paper, Boston College, CIRANO and University of Chicago.

Kim S, Shephard N, Chib S. 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* **65**: 361–393.

Koopman SJ, Hol Uspensky E. 2000. The stochastic volatility in mean model: empirical evidence from international stock markets. Discussion Paper, Tinbergen Institute, Amsterdam-Rotterdam, The Netherlands. `http://www.tinbergen.nl`

Koopman SJ, Shephard N, Doornik J. 1999. Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal* **2**: 113–166. `http://www.ssfpack.com`

Nelson DB. 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**: 347–370.

Pitt MK, Shephard N. 1999. Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association* **94**: 590–599.

Poon S, Taylor SJ. 1992. Stock returns and volatility: an empirical study of the UK stock market. *Journal of Banking and Finance* **16**: 37–59.

Ripley BD. 1987. *Stochastic Simulation*. John Wiley: New York.

Sandmann G, Koopman SJ. 1998. Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics* **87**: 271–301.

Schwert GW. 1989. Why does stock market volatility change over time? *Journal of Finance* **44**: 1115–1153.

Scott LO. 1987. Option pricing when the variance changes randomly: theory, estimation and application. *Journal of Financial and Quantitative Analysis* **22**: 419–438.

Shephard NG. 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, Cox DR, Hinkley DV, Barndorff-Nielsen OE (eds). Chapman and Hall: London; 1–67.

Shephard N, Pitt MK. 1997. Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**: 653–667.

Taylor SJ. 1986. *Modelling Financial Time Series*. John Wiley: Chichester.

Taylor SJ. 1994. Modeling stochastic volatility: a review and comparative study. *Mathematical Finance* **4**: 183–204.

Watanabe T. 1999. A non-linear filtering approach to stochastic volatility models with an application to daily stock returns. *Journal of Applied Econometrics* **14**: 101–121.

Wiggins JB. 1987. Option values under stochastic volatility: theory and empirical estimates. *Journal of Financial Economics* **19**: 351–372.

