# IMF Working Paper

# Modeling Stochastic Volatility with Application to Stock Returns

*Noureddine Krichene*

INTERNATIONAL MONETARY FUND

# IMF Working Paper

African Department

## Modeling Stochastic Volatility with Application to Stock Returns

Prepared by Noureddine Krichene[1]

Authorized for distribution by Menachem Katz

June 2003

## Abstract

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

A stochastic volatility model where volatility was driven solely by a latent variable called news was estimated for three stock indices. A Markov chain Monte Carlo algorithm was used for estimating Bayesian parameters and filtering volatilities. Volatility persistence being close to one was consistent with both volatility clustering and mean reversion. Filtering showed highly volatile markets, reflecting frequent pertinent news. Diagnostics showed no model failure, although specification improvements were always possible. The model corroborated stylized findings in volatility modeling and has potential value for market participants in asset pricing and risk management, as well as for policymakers in the design of macroeconomic policies conducive to less volatile financial markets.

Author's E-Mail address: Nkrichene@IMF.org

---

Contents                                                              Page

# I. INTRODUCTION

This paper estimates time-varying volatilities for three stock market indices: the Dow Jones Industrial Average, the FTSE 100, and the Nikkei 225 indices over the period January 1, 1999—September 24, 2002. Modeling of financial time series has focused on estimating the time-varying volatility. The latter is key for measuring risk, pricing asset derivatives, and hedging strategies (e.g., Hull and White, 1987; Chesney and Scott, 1989). Volatility, being a second moment, characterizes the uncertainty that the market assigns to the price of a particular asset. It provides market participants with up-to-date information on market sentiment and reactions to news, market beliefs about the future, and the changing attitude toward risk from bearish to bullish and vice versa.[2] As it turns out, dealers quote options' prices in terms of volatilities and strike prices in terms of deltas.[3] Besides market participants, the IMF and central banks have also a special interest in volatility as a source of information for gauging market expectations and reactions to news, assessing monetary conditions and monetary policy, and influencing the timing and effectiveness of monetary operations and interventions. In its surveillance role, the IMF monitors volatilities of asset prices and related derivatives' prices in both mature and emerging markets with a view to promoting macroeconomic policies conducive to financial stability, whereas monetary authorities routinely use the information that is embedded in financial assets to help in formulating and implementing monetary policy. Section II of the paper presents the stochastic volatility model (SVM) as proposed by Taylor (1986) and that later became the main framework for estimating stochastic volatility.[4] Section III describes the estimation methodology. A main feature of the SVM is the intractability of its likelihood function. In this section, a Markov chain Monte Carlo algorithm is used for estimating the parameters in a Bayesian context as well as the volatilities of the SVM. Section IV presents the empirical estimates of the parameters, the filtered volatilities, and the diagnostic checks of the model; and Section V concludes.

---

[2] Market participants may also be interested in the entire risk-neutral probability distribution for an asset price at a future date $T$ with a view to third moment (skewness) and fourth moment (kurtosis). First moment, given by forward and futures prices, measures only the expected level and does not capture uncertainty, which is accounted for by the volatility.

[3] The delta of an option measures the change in the option price in response to a change in the current price of the underlying asset.

[4] Surveys on the models of stochastic volatility can be found in Shephard (1996); and Ghysels, Harvey, and Renault (1996).

## II. MODELING VOLATILITY

Modeling volatility is key for pricing asset derivatives.[5] Time-varying volatility model was initially expressed by Engel (1982) as an autoregressive conditional heteroskedasticity (ARCH) model. A volatility model is defined by its first and second moment which can be referred to as the mean and variance equation.

Mean equation: $y_t = \sigma_t \varepsilon_t$

Variance equation: $\sigma_t^2 = \mu + \phi\, y_{t-1}^2$

$y_t$ = the returns on asset prices, defined as $y_t = 100*\text{Log}(X_t/X_{t-1})$, where $X_t$ is the observed market index or the asset price.

$\sigma_t^2$ = volatility of the asset returns.

Volatility is a measure of risk on returns. Each observed data point $y_t$ has a standard deviation $\sigma_t$. The disturbance $\varepsilon_t$ is Gaussian, i.e. $\varepsilon_t \sim$ iid $N(0, 1)$. In the ARCH model, volatility is a deterministic function of the squares of past return and therefore is an observed variable. This model was extended by Bollerslev (1986) to become a generalized autoregressive conditional heteroscedasticity (GARCH) model:

$$y_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \mu + \alpha\, y_{t-1}^2 + \beta\, \sigma_{t-1}^2$$

In this version of the model, the volatility is a deterministic function of the squares of past return as well as of past volatility. The GARCH formulation introduces terms analogous to moving average terms in an ARMA model, thereby making forecast a function of a distributed lag of past squared observations. Apart from volatility clustering, i.e. prolonged periods of high and low volatility, a GARCH model captures part of the excess kurtosis observed in financial time series. Subject to the parameter restrictions: $\mu > 0$, $\alpha \geq 0, \beta \geq 0$,

---

[5] A basic formula for pricing contingent claims is the Black-Scholes formula (1973). For instance the price of a European call option is given by
$C(S,K,T,t,r,\sigma) = S.N(d_1) - K.N(d_2)$, where $S$ is the current price of the underlying asset, $K$ = the strike price, $T$ = the maturity time, $t$ = the current time, $r$ = the risk free return, and $\sigma$ is the volatility of $S$. The symbol $N(.)$ is the cumulative normal distribution:
$$d_1 = \frac{\log(S) - \log(K) + (r + 0.5\sigma^2)(T - t)}{\sigma\sqrt{T-t}}$$ and $d_2 = d_1 - \sigma\sqrt{T-t}$ . In this formula, all the variables, except $\sigma$, are known. Therefore, pricing a contingent claim amounts to estimating $\sigma$.

and $\alpha + \beta < 1$, it can be shown that the fourth moment will exhibit excess kurtosis:

$$\kappa_y = \frac{\kappa_\varepsilon E(\sigma_t^4)}{E(\sigma_t^2)^2} = 3 + \frac{6\alpha^2}{1 - \beta^2 - 2\alpha\beta - 3\alpha^2}$$ and therefore $\kappa_y > \kappa_\varepsilon$, where $\kappa_\varepsilon$ is the kurtosis of

$\varepsilon_t$.[6] The GARCH model has been extended to become a stochastic GARCH

$$y_t = \sigma_t \varepsilon_t$$
$$\Delta \sigma_t^2 = \mu + \alpha \, y_{t-1}^2 + (\beta - 1) \, \sigma_{t-1}^2 + \sigma_\eta \, \eta_t$$

Where $\eta_t \sim$ iid N(0, 1) is called the volatility of volatility. The disturbances $\varepsilon_t$ and $\eta_t$ may be assumed to be independent or correlated. Correlation between the random disturbances introduces a leverage effect.[7] The mean equation is now affected by two error terms, $\varepsilon_t$ and $\eta_t$. In virtue of the Gaussianity of $\varepsilon_t$ and $\eta_t$, the ARCH and GARCH parameter vector $\theta$ = ($\mu, \alpha, \beta, \sigma_\eta$) can be estimated using the likelihood method. The SVM was proposed by Taylor (1986), to make volatility dependent on a latent, unobservable state variable $h_t$ called news, via the relation:[8]

$$\text{Log}\,\sigma_t^2 = h_t$$

$h_t$ is referred to as log volatility of $y_t$. Replacing $\sigma_t$ by $\exp(h_t/2)$, the SVM is a log-volatility model and is written in a state space form

(1) Observation equation: $y_t = \exp(h_t/2)\varepsilon_t$ , $\varepsilon_t \sim$ iid N(0,1), t=1,......, T

(2) Transition equation: $h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_\eta \eta_t$ , $\eta_t \sim$ iid N(0, $\sigma_\eta^2$), t=1,......, T

$h_0 \sim$ N($\mu, \sigma_\eta^2/(1-\phi^2)$)

---

[6] It is often necessary to use a non-Gaussian GARCH model, such us the Student-t distribution for $\varepsilon_t$, to capture the high kurtosis typically found in financial time series.

[7] The leverage effect, analyzed by Black (1976), suggests that stock price movements are negatively correlated with volatility. Because falling stock prices imply an increased leverage of firms, more uncertainty, and hence volatility, will arise.

[8] Contrary to ARCH-GARCH models where volatility is observed, in the SVM, volatility is latent. Asset prices move through a sequence of equilibria, where a move from one equilibrium to another is prompted by the arrival of new information to the market. While the agents realize the information and act upon it, the econometrician can only observe the asset prices, which result from the agents' actions and is forced to model the information as a latent variable.

The parameter vector is $\theta = (\mu, \phi, \sigma_\eta)$. The initial conditions for the state, $h_0$, are given by the unconditional mean and variance of $h_t$, i.e. $\mu$ and $\sigma_\eta^2/(1-\phi^2)$, respectively.

SVM is solely interested in market news and in modeling their effects on volatility. Contrary to ARCH-GARCH models, volatility in SVM is not affected by past returns. SVM arises as a discrete-time approximation to various diffusion processes analyzed in the continuous-time asset-pricing literature.[9] A state space model specifies in the observation equation (1) the conditional distribution of the observation $y_t$ given the unknown state $h_t$. The state variable $h_t$, appearing nonlinearly in the observation equation, determines the amount of volatility (variance) of each data point $y_t$. If the information flows are autocorrelated, then the state variable may be assumed to follow a Markov process given by the transition equation (2). The value of $\phi$, $-1 < \phi < 1$, measures the autocorrelation present in the logged squared data, i.e. $\log(y_t^2)$. Thus $\phi$ can be interpreted as the persistence in the volatility, with high $\phi$ indicating volatility clustering. The constant scaling factor $\beta = \exp(\mu/2)$ is seen as the modal volatility, and $\sigma_\eta$ as the volatility of the log-volatility. The stationarity of $h_t$, or mean reversion, is given by the restriction $-1 < \phi < 1$; it implies the stationarity of $y_t$. The variance of $y_t$ is related to the variance of $\varepsilon_t$ and $h_t$: $\text{Var}(y_t) = \text{Var}(\varepsilon_t)\exp(\text{Var}(h_t/2))$. If the fourth moment of $\varepsilon_t$ exists, the kurtosis of $y_t$, $\kappa_y$, is equal to $\kappa_\varepsilon \exp(\text{Var}(h_t/2))$, where $\kappa_\varepsilon$ is the kurtosis of $\varepsilon_t$ and therefore $\kappa_y > \kappa_\varepsilon$, allowing the SVM to capture part of the excess kurtosis observed in financial time series. Finally the odd moments are zero.

---

[9] The stochastic volatility model can be seen a discrete-time approximation to the Hull and White derivative pricing model(1987), where the stock price dynamics are governed by some unobservable state variables, such as the random volatility. In their model stock prices follow a diffusion process $dS_t / S_t = \alpha dt + \sigma(t)dW_1$, and the logarithm of $\sigma(t)$ follows a diffusion process given by the Ornstein-Uhlenbeck (O-U) process: $d(\ln \sigma) = \lambda(\xi - \ln \sigma)dt + \gamma dW_2$, where $S_t$ = stock price, $\sigma(t)$ is the instantaneous variance of $S_t$, $W_1$ and $W_2$ are two Wiener processes. The SVM is also a discrete-time approximation to the term structure of interest rates model $dr(t) = \kappa_1(\lambda - r(t))dt + \sigma(t)dW_1(t)$ (Vasicek, 1977), or

$dr(t) = \kappa_1(\lambda - r(t))dt + \sigma(t)\sqrt{r(t)}dW_1(t)$ (Cox, Ingersoll, and Ross, 1985),

where $r(t)$ = interest rate, and the stochastic volatility is given by the O-U process

$d\log\sigma^2(t) = \kappa_2(\alpha - \log\sigma^2(t))dt + \xi dW_2(t)$.

## III. ESTIMATION METHODOLOGY

### A. The Maximum Likelihood Dilemma

The objective of the SVM model is to estimate simultaneously the parameter vector $\theta = (\mu, \phi, \sigma_\eta)$ and the volatility vector $h = (h_1, \ldots, h_T)$ conditional on the observed data set $Y_T = (y_1, \ldots, y_T)$. Denoting conditional expectation by $E(h_n \mid Y_t, \theta)$, where $Y_t = (y_1, \ldots, y_t)$, then $E(h_n \mid Y_t, \theta)$ is called predicted volatility when $n > t$, filtered volatility when $t = n$, and smoothed volatility when $n < t$. Despite its parsimony and its appeal for modeling market news, the SVM given by equations (1) and (2) is not easy to fit by standard maximum likelihood methods. Because the observation equation is nonlinear in the state variable, direct application of the Kalman filter and the associated smoother will yield estimators of the state $h_t$ which are only optimal within the class of estimators based on linear combinations of $\log y_t^2$. The marginal likelihood over the parameters of the SVM is defined by a T-dimensional integral. Indeed, given the parameters $\theta = (\mu, \phi, \sigma_\eta)$, the likelihood of the SVM is the density of data $(y_1, \ldots, y_T)$:

$$L(\theta) = f(y \mid \theta) = \int f(y, h \mid \theta) dh \propto (\text{proportional}) \int f(y \mid h, \theta) f(h \mid \theta) dh \ ^{10}$$

Denoting $Y_{t-1} = (y_1, \ldots, y_{t-1})$, the density of the data can be expressed as a mixture over the $h_t$ distributions:

$$L(\theta) = f(y \mid \theta) = \prod_{t=1}^{T} f(y_t \mid Y_{t-1}, \theta) = \prod_{t=1}^{T} f(y_t \mid h_t, \theta) f(h_t \mid Y_{t-1}, \theta) dh_t$$

This likelihood function is intractable, i.e., it cannot be easily computed. The source of the problem is that the density $f(h_t \mid Y_{t-1}, \theta)$ cannot be expressed in closed form, and therefore $y_t \mid Y_{t-1}$ does not have an analytic expression. As a result, the likelihood function must be approximated or calculated by some kind of simulation method.

One example of approximation to the SVM uses the quasi-likelihood method by transforming the SVM into as a non-Gaussian but linear state space:

$$y_t^* = \log y_t^2 = h_t + \log \varepsilon_t^2$$

---

[10] This relation derives from Bayes formula. Let A and B be two events, then Bayes theorem states $P(A \mid B) = \dfrac{P(B \mid A)P(A)}{P(B)}$ where $P(A)$ and $P(B)$ are the marginal probabilities of the random events $A$ and $B$, respectively. The proportionality factor is $P(B)$.

Where $\log \varepsilon_t^2 \sim$ iid $\log \chi_1^2$ (log chi-square) with $E(\log \varepsilon_t^2) = -1.27$ and $Var$ $(\log \varepsilon_t^2) = 4.93$. Consequently, the Kalman filter can be used to provide the best linear unbiased estimator of $h_t$ given $Y_{t-1}^* = (y_1^*, \ldots, y_{t-1}^*)$. For instance, Harvey $et\ al.$, (1994) have employed Kalman filtering to estimate $\theta = (\mu, \phi, \sigma_\eta^2)$ by maximizing the quasi-likelihood:

$$\log L_Q(y^* \mid \theta) = -1/2 \sum_{t=1}^{T} \log F_t - 1/2 \sum_{t=1}^{T} v_t^2 / F_t$$

where $y^* = (y_1^*, \ldots, y_T^*)$, $v_t$ is the one-step-ahead prediction error for the best linear estimate of $y_t^* = \log y_t^2$, and $F_t$ is the corresponding mean square error. It turns out that $\log \varepsilon_t^2 \sim$ iid $\log \chi_1^2$ is poorly approximated by a normal distribution and filtered volatilities estimates have higher standard errors compared to estimates from simulation methods.

## B. Markov Chain Monte Carlo Methods

As mentioned above, a key issue is that the likelihood function $f(y \mid \theta) = \int f(y \mid h, \theta) f(h \mid \theta) dh$ is intractable. This precludes a direct analysis of the posterior density $f(\theta \mid y)$. As a solution, Jacquier, Polson, and Rossi (1994) advanced the method of the Markov chain Monte Carlo (MCMC) based on the general approach of Carlin, Polson, and Stoffer (1992) and on the method of data augmentation for computing posterior densities proposed by Tanner and Wong (1987).[11] Data augmentation is seen as an algorithm that allows the estimation of the parameters $\theta$ without computing the likelihood function. Basically, the parameters $\theta$ are augmented with the time series $h$ to form a large parameter vector $(\theta, h)$. The joint distribution of $(\theta, h)$ conditional on $y$, denoted by $f(\theta, h \mid y)$, is now the focus of attention. MCMC can be developed to sample this density without computing the likelihood function $f(y \mid \theta)$. Viewing the specification of the SVM as a hierarchical structure described by three conditional distributions: $y \mid h$, $h \mid \theta$, and $f(\theta)$, where $f(\theta)$ is the prior distribution of $\theta$, the joint posterior distribution of $(\theta, h)$ derives from the application of Bayes theorem. Hence $f(\theta, h \mid y)$ is proportional ($\propto$) to the product of these three densities: $f(\theta, h \mid y) \propto f(y \mid h) f(h \mid \theta) f(\theta)$.

---

[11] The idea of data augmentation was explored by Dempster $et\ al.$, (1977) as a way to overcoming complex maximum likelihood functions by developing the two step expectation-maximization EM algorithm. In the expectation step (E) and for a given value of the parameter vector, $\theta$, the observed data is augmented by a nonobserved (missing) data set, $h$. In the maximization step (M), the maximum likelihood function for the augmented data set $(y, h)$ turns out to be easier to compute and a new value for $\theta$ is derived. This new value is used in the E-step to compute a new set of latent variables, which form a new augmented data set. The EM algorithm is iterated till convergence obtains.

The MCMC method is simple and extremely general. In order to sample a given probability distribution that is referred to as the target distribution, here $f(\theta, h \mid y)$, an ergodic Markov chain (i.e., irreducible and aperiodic) is constructed with the property that its limiting invariant distribution is the target distribution. Once the Markov chain has been constructed, a sample of (correlated) draws from the target distribution can be obtained by simulating the Markov Chain a large number of times and recording its values. Let the marginal distributions arising from the joint posterior distribution $f(\theta, h \mid y)$ be $f(\theta \mid y)$ and $f(h \mid y)$, then by appealing to ergodic theorems for Markov chains, inference can be made about the posterior moments of $\theta$ and $h$ from these two marginal distributions, respectively. To construct the Markov chain with invariant distribution $f(\theta, h \mid y)$, Tanner and Wong (1987) proposed an iterative data augmentation algorithm: (i) first sample $\theta$ from $f(\theta \mid h, y)$; (ii) and then sample $h$ from $f(h \mid \theta, y)$, using the value of $\theta$ from step (i); (iii) iterate till convergence.

**Sampling Algorithms: The Gibbs and the Metropolis-Hastings Algorithms**

To produce sample variates from the joint posterior distribution of the volatilities and parameters given the data, Gibbs or the Metropolis-Hastings (M-H) sampling algorithms are used to construct Markov chains with equilibrium distribution the joint distribution. To illustrate each algorithm, let a multivariate distribution be defined as $\pi(\psi)$. In the Gibbs sampler, the variables are grouped into $p$ blocks $(\psi_1, \ldots, \psi_p)$, and each block is sampled according to the full conditional distribution of block $\psi_k$ denoted by $\pi(\psi_k \mid \psi_{-k})$, where $\psi_{-k}$ denotes all the blocks excluding $\psi_k$. Derivation of the full conditional distribution is usually quite simple since, by Bayes theorem, $\pi(\psi_k \mid \psi_{-k}) \propto \pi(\psi_k, \psi_{-k})$, the joint distribution of all blocks. In addition, the powerful device of data augmentation, due to Tanner and Wong (1987), in which latent or auxiliary variables are artificially introduced into sampling, is often used to simplify the derivation and sampling of the full conditional distribution. For some blocks, the conditional distribution may not be easily sampled directly. One can then use a Metropolis-Hastings (M-H) step whereby $\psi$ is sampled from a blanketing density $q(\psi, \psi')$, i.e. upper bounding density; $q(\psi, \psi')$ is also called a proposal density which provides a candidate value $\psi'$. The new draw $\psi'$ from $q(\psi, \psi')$ is accepted with probability

$$\alpha(\psi, \psi') = \min[\frac{\pi(\psi')q(\psi', \psi)}{\pi(\psi)q(\psi, \psi')}, 1]$$

Otherwise, the previous draw, i.e., $\psi$, is repeated, and one moves to the next block. The draws from the simulated distribution converge to draws from the stationary distribution

namely the required joint posterior.[12] It may be mentioned that if the full conditional densities are available, whether in the context of the Gibbs or the M-H sampler, then the MCMC output can be used to estimate posterior marginal density functions.[13]

**Application of the Sampling Algorithms to the SVM**

For the SVM given by equations (1) and (2), Jacquier *et al.*, (1994) showed that the Gibbs blocks can be written as $f(\mu,\phi,\sigma_\eta \mid y,h) \propto f(y,h \mid \mu,\phi,\sigma_\eta)f(\mu,\phi,\sigma_\eta)$ and T conditionals $f(h_t \mid h_{-t}, y_t, \mu, \phi, \sigma_\eta)$ where $h_{-t}$ means all elements of $h_1,\ldots,h_T$ except $h_t$. Since the model has a Markov random field structure, the conditional distribution $f(h_t \mid h_{-t}, y_t)$ can be written as $f(h_t \mid h_{-t}, y_t) \propto f(y_t \mid h_t)f(h_{t+1} \mid h_t)f(h_t \mid h_{t-1})$.

A Gibbs sampler for the SVM which produces variates from the joint posterior distribution of $(h,\theta)$ can be described as follows

1. Initialize $h$ and $\theta$
2. Sample $h_t$ from $h_t \mid h_{-t}, y, \theta$          t=1,.......,T
3. Sample $\sigma_\eta^2 \mid y, h, \phi, \mu$
4. Sample $\phi \mid y, h, \mu, \sigma_\eta^2$
5. Sample $\mu \mid y, h, \phi, \sigma_\eta^2$
6. Go to 2.

While sampling from the conditionals of the parameters $\theta$ is relatively easy, provided conjugancy of the prior distribution is appropriately chosen, sampling from $h_t \mid h_{-t}, y, \theta$ requires more efforts in spite its Markovian structure for essentially two reasons. First, the constant of proportionality is unknown in the approximation

$$f(h_t \mid h_{-t}, y_t) \propto f(y_t \mid h_t)f(h_{t+1} \mid h_t)f(h_t \mid h_{t-1}).$$

Second, $f(y_t \mid h_t)f(h_{t+1} \mid h_t)f(h_t \mid h_{t-1})$ yields a non standard kernel density which is the product of normal and log normal densities. As mentioned above, the solution is to find a

---

[12] Chib and Greenberg (1994) discuss a way of formulating proposal densities. They suggest matching the proposal density to the target density at the mode by a multivariate normal or a multivariate-t distribution with location given by the mode of the target density and the dispersion given by the inverse of the Hessian evaluated at the mode. Specifically, the parameters of the proposal density are taken to be $m = \arg\max \log \pi(\psi)$ and

$V = \tau\{-\dfrac{\partial^2 \log \pi(\psi)}{\partial \psi \partial \psi'}\}^{-1}_{\psi=\psi}$ . The proposal density is then specified as

$q(\psi') = f(\psi' \mid m, V)$ where $f$ is a multivariate normal or t-density.

[13] Marginalization of one variable over the MCMC output is called Rao-Blackwellisation.

blanketing, i.e. an upper bounding density function $g(h_t \mid h_{-t}, y_t)$, which has a standard form and from which it is easy to draw variates $\{h_t^{(1)}, \ldots\ldots, h_t^{(M)}\}$ for the estimation of the volatilities. If there exits a constant C such that $f(h_t \mid h_{-t}, y_t) \le Cg(h_t \mid h_{-t}, y_t) \; \forall h_t$, one can sample from $f(h_t \mid h_{-t}, y_t)$ by drawing from $g(h_t \mid h_{-t}, y_t)$ and accepting with probability $f(h_t \mid h_{-t}, y_t) \le Cg(h_t \mid h_{-t}, y_t)$. One can also sample from $g(h_t \mid h_{-t}, y_t)$ using an accept/reject Metropolis-Hastings algorithm.

The computation of the kernel of $f(y_t \mid h_t) f(h_{t+1} \mid h_t) f(h_t \mid h_{t-1})$ is carried out by taking the product of three kernels $f(y_t \mid h_t)$, $f(h_{t+1} \mid h_t)$, $f(h_t \mid h_{t-1})$. Note that the errors of the SVM, $\varepsilon_t$ and $\eta_t$, are iid N(0,1), implying from equations (1) and (2) that

$y_t \mid h_t \sim N(0, \exp(h_t))$, $h_{t+1} \mid h_t \sim N(\mu + \phi(h_t - \mu), \sigma_\eta^2)$, and $h_t \mid h_{t-1} \sim N(\mu + \phi(h_{t-1} - \mu), \sigma_\eta^2)$.

It follows that $f(y_t \mid h_t) \propto (\exp(h_t))^{-0.5} \exp(-0.5 y_t^2 / \exp(h_t))$,

$f(h_{t+1} \mid h_t) \propto \sigma_\eta^{-1} \exp(-0.5(h_{t+1} - \mu - \phi(h_t - \mu))^2 / \sigma_\eta^2)$.

$f(h_t \mid h_{t-1}) \propto \sigma_\eta^{-1} \exp(-0.5(h_t - \mu - \phi(h_{t-1} - \mu))^2 / \sigma_\eta^2)$.

Taking the logarithm of the product of the three above conditional densities, the following relation is obtained:

$\log f(h_t \mid h_{t+1}, h_t, y_t) \propto \log f(y_t \mid h_t) f(h_{t+1} \mid h_t) f(h_t \mid h_{t-1}) = const + \log f^*$, where

$\log f^* = -0.5 h_t - 0.5(h_t - h_t^*)^2 / \sigma^2 - 0.5 y_t^2 \exp(-h_t)$,

$h_t^* = [\mu(1-\phi) + \phi(h_{t+1} + h_{t-1})]/(\phi^2 + 1)$, and $\sigma^2 = \sigma_\eta^2 /(\phi^2 + 1)$.

Taking the first order Taylor expansion of $\exp(-h_t)$ around $h_t^*$, then:

$\log f^* \le -0.5 h_t - 0.5(h_t - h_t^*)^2 / \sigma^2 - 0.5 y_t^2 \{\exp(-h_t^*(1 + h_t^*) - h_t \exp(-h_t^*)\} = \log(g^*)$.

Accordingly, $\log(g^*)$ is a bounding function. The density $g^*$ has a normalized distribution, with mean $\delta_t = h_t^* + 0.5\sigma^2 [y_t^2 \exp(-h_t^*) - 1]$ and variance $\sigma^2 = \sigma_\eta^2 /(\phi^2 + 1)$. Therefore, in order to sample variates from $f(h_t \mid h_{t+1}, h_t, y_t)$, one can draw proposals from $h_t \sim N(\delta_t, \sigma^2)$ and accept with probability $f^* / g^*$.

## A Bayesian Approach: Priors of the Parameters $\theta = (\mu, \phi, \sigma_\eta^2)$

The SVM is estimated in a Bayesian approach. A prior distribution is formulated for each of the parameters $\theta = (\mu, \phi, \sigma_\eta^2)$. The same priors used in Kim et al., (1998) are considered here. It is assumed that each parameter is a prior independent. Conjugancy is fully addressed so that the posterior distribution of the parameter would be a standard one with easily computed moments. Thenceforth, for $\sigma_\eta^2$ a conjugate prior is assumed; namely $\sigma_\eta^2$ is believed to be distributed as an inverse-gamma (IG): $\sigma_\eta^2 \mid \phi, \mu \sim IG(\sigma_{IG}/2, S_\sigma/2)$. For $\phi$, the following relation is postulated: $\phi = 2\phi^* - 1$ where $\phi^*$ is distributed as a Beta with parameters $\phi^{(1)}, \phi^{(2)}$.

The prior $f(\phi)$ has support on the interval $(-1,1)$ with prior mean of $\{2\,\phi^{(1)}/(\phi^{(1)}+\phi^{(2)})-1\}$. For $\phi^{(1)}=20$, $\phi^{(2)}=1.5$, the prior mean of $\phi$ is 0.86. Finally a diffuse, i.e., non informative, prior is assumed for $\mu$.

## Multimove Sampling of the Volatilities: Approximation of the SVM by a Conditional Gaussian State Space

It is well established that single move Gibbs or Metropolis-Hastings samplers, which sample each volatility $h_t$ separately, t=1,...,T, produce a highly correlated sample and are slow to converge. Multimove samplers which sample all latent volatilities $h = (h_1,...,h_T)$ at once produce less correlated variates and are faster to converge. Carter and Kohn (1994), Shephard (1994), Kim et al., (1998), and Chib et al., (2002) suggested converting the nonlinear non Gaussian state space model into a linear and conditionally Gaussian state space model relying on an offset mixture of normal distributions as an approximation to the log chi-square distribution $\log \chi_1^2$. In the conditionally Gaussian model, the density of the data has an exact likelihood and all volatilities can be estimated at once using the Kalman filter. The approximating model is: $y_t^* = h_t + z_t$, where $y_t^* = \log(y_t^2 + c)$,[14]

$$\text{and } f(z_t) = \sum_{i=1}^{K} q_i\, f_N(z_t\,|\,m_i - 1.2704,\ v_i^2)$$

is a mixture of K normal densities $f_N$ with component probability $q_i$, means $m_i - 1.2704$, and variances $v_i^2$. The constants $\{q_i,\ m_i,\ v_i^2\}$ are selected to closely approximate the exact density of $\log \varepsilon_t^2$. It should be noted that the mixture density can also be written in terms of a component indicator variable $s_t$ such that

$$z_t\,|\,s_t = i \sim N(m_i - 1.2704,\ v_i^2),\ \Pr(s_t = i) = q_i$$

In the mixture model, the joint posterior density under focus is $f(s,h,\phi,\sigma_\eta^2,\mu\,|\,y^*)$, where $s = (s_1,...,s_T)$. Because $y^*\,|\,s,\ \phi,\sigma_\eta^2,\mu$ is a Gaussian time series model, it follows that all $h = (h_1,...,h_T)$ can be sampled at once from the entire Gaussian distribution $h\,|\,y^*,s,\phi,\sigma_\eta^2,\mu$, using the Gaussian simulation smoother (e.g., Kim et al., 1998; Chib et al., 2002 ). As for sampling $s$ from $s\,|\,y^*,h$, this is done by independently sampling each $s_t$ using the probability mass function
$$\Pr(s_t = i\,|\,y_t^*,h_t) \propto q_i\, f_N(y_t^*\,|\,h_t + m_i - 1.2704,\ v_i^2),\quad i \le K.$$

---

[14] The offset c was introduced into SVM literature by Fuller 1996, (pp. 494-497), in order to robustify the QML estimator of the SVM to $y_t^2$ being very small.

Exploiting the linearity of the conditionally Gaussian state space, Kim *et al.*, (1998) noted that correlation can still be reduced if the Gaussian structure of $y^* \mid s, \phi, \sigma_\eta^2, \mu$ is used.[15] It is possible to sample the joint distribution $f(h, \phi, \sigma_\eta^2, \mu \mid y^*, s)$ by sampling:

$(\phi, \sigma_\eta^2)$ from $f(\phi, \sigma_\eta^2 \mid y^*, s) \propto f(y^* \mid s, \phi, \sigma_\eta^2) f(\phi, \sigma_\eta^2)$ and then sampling $(h, \mu)$ from $f(h, \mu \mid y^*, s, \phi, \sigma_\eta^2)$. This improved version of the mixture sampler is called an integration sampler because $(h, \mu)$ are analytically integrated out of the kernel density of $y_t^*$.

The algorithm of the integration sampler is then generically:

1. Initialize ($\mu$, s, $\phi, \sigma_\eta$)

2. Sample $\phi, \sigma_\eta^2$ from $f(\phi, \sigma_\eta^2 \mid y^*, s)$, where $y^* = (y_1^*, ..., y_T^*)$ using a Metropolis-Hastings suggestion based on a proposal density $g(\phi, \sigma_\eta^2)$.[16]

3. Sample $h, \mu \mid y^*, s, \phi, \sigma_\eta^2$ using an augmented version of the Kalman filter.[17]

4. Sample $s \mid y^*, h$ using $\Pr(s_t = i \mid y_t^*, h_t) \propto q_i \, f_N(y_t^* \mid h_t + m_i - 1.2704, \, v_i^2)$.

5. Goto 2.

The integration sampler therefore draws $h$ and $\mu$ jointly. In addition, sampling the posterior joint density $(\phi, \sigma_\eta^2)$, i.e., $f(\phi, \sigma_\eta^2 \mid y^*, s) \propto f(y^* \mid s, \phi, \sigma_\eta^2) f(\phi, \sigma_\eta^2)$ is easy because $f(y^* \mid s, \phi, \sigma_\eta^2)$ can be evaluated using an augmented version of the Kalman filter.

---

[15] Consider $y_t^* = h_t + z_t$, and $h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_\eta \eta_t$. Substituting for $h_t$ yields $y_t^* = \mu + \phi(h_{t-1} - \mu) + \sigma_\eta \eta_t + z_t$. Substituting for $h_{t-1} = y_{t-1}^* - z_{t-1}$ yields

$y_t^* = \mu + \phi((y_{t-1}^* - z_{t-1}) - \mu) + \sigma_\eta \eta_t + z_t = \mu(1 - \phi) + \phi y_{t-1}^* + \sigma_\eta \eta_t + z_t - \phi z_{t-1}$. Since the density of $y_t^*$ is a function of the densities of the error terms $\eta_t$ and $z_t$, it is therefore affected only by the parameters $\phi$ and $\sigma_\eta^2$.

[16] $g(\phi, \sigma_\eta^2)$ could be a Gaussian or a Student-t distribution. For instance, one can make a proposal draw $\theta^{(i)}$ from a tailored multivariate-t density $f_T(\theta \mid m, V, \xi)$ with $\xi$ degrees of freedom, where $m$ is the value that maximizes the density log f($y_t^* \mid$s, $\theta$) and V is minus the inverse Hessian of log f($y_t^* \mid$ s, $\theta$) evaluated at $m$. This approach for specifying the proposal density was introduced by Chib and Greenberg (1994). The proposal value generated from this density is then accepted or rejected according to the Metropolis-Hastings algorithm.

[17] See Kim *et al.*, (1998) for details on the augmented Kalman filter.

As in any approximation sampler, reweighting the output is required to get closer to a sample drawn from the exact posterior density of the parameters and volatilities, i.e. $f(\theta, h \mid y)$. The reweighting coefficients are computed as a difference between the true and approximated density. Let the mixture approximating density be $k(\theta, h \mid y^*)$ and define:

$$w(\theta, h) = \log f(\theta, h \mid y) - \log k(\theta, h \mid y^*) = const + \log f(y \mid h) - \log k(y^* \mid h) \text{ where}$$

$$f(y \mid h) = \prod_{t=1}^{n} f_N\{y_t \mid \theta, \exp(h_t)\} \text{ and } k(y^* \mid h) = \prod_{t=1}^{T} \sum_{i=1}^{K} q_i \, f_N(y_t^* \mid h_t + m_i - 1.2704, \, v_i^2)$$

Both these functions involve Gaussian densities and are straightforward to evaluate for any given value of $h$. Then, the moments of the conditional distribution $f(\theta \mid y)$ can be defined

$$Eg(\theta \mid y) = \int g(\theta) f(\theta \mid y) d\theta =$$

$$\int \int g(\theta) \exp\{w(\theta, h)\} k(\theta, h \mid y^*) d\theta dh / \int \int \exp\{w(\theta, h)\} k(\theta, h \mid y^*) d\theta dh.$$

Thus we can estimate the moments of the posterior distribution by reweighting the MCMC draws according to $Eg(\theta \mid y) \approx \sum_{j=1}^{M} g(\theta^j) c^j$, where

$$c^j = \exp(w(\theta^j, h^j)) / \sum_{i=1}^{M} \exp(w(\theta^i, h^i)) \text{ and } \{(\phi^{(1)}, h^{(1)}), ...., (\theta^{(M)}, h^{(M)})\} \text{ is a sample of size}$$

$M$ of the MCMC output.

## IV. APPLICATION OF THE SVM TO STOCK RETURNS: EMPIRICAL ESTIMATES[18]

### A. Parameter Estimates

Estimated coefficients of the SVM for the Dow Jones Industrial Average, FTSE 100 and Nikkei 225 using daily data from 1/1/1999-9/24/2002 are reported in Table 1. The posterior means of $\phi$ is close to one, ranging between 0.94 and 0.97. These values imply high persistence in conditional variances, in accordance with typical estimates reported for the SVM models. Being strictly less than one, the estimated persistence coefficient tends to support the hypothesis of mean reverting volatility, i.e., random shocks to volatility will eventually taper off with a half-life of a few weeks. Estimates of $\sigma_n$ being relatively small, between 0.15 and 0.17, are similar to estimates reported for the SVM models and imply a good fit of the volatility process by equation (2). The mode volatility, $\beta$, is high, ranging between 1.15 and 1.38, implying that major daily news occur frequently with sizeable effect on volatility in stock markets. The distributions of the volatility parameters $(\mu, \phi, \sigma_n)$ are also concentrated around their means, implying thus statistically significant coefficients. In view of these results, the SVM tends to capture adequately the behavior of the stock indices.

---

[18] All estimations in this section were obtained using the SV package developed by Kim *et al.*, (1998). The results were based on their integration sampler.

| Table 1. Daily Returns for the Dow Jones Industrial Average, FTSE 100, and Nikkei 225: Posterior Parameter Estimates from the Reweighted Integration Sampler and Monte Carlo Standard Errors (MC S.E.) | | | | | |
|---|---|---|---|---|---|
| | Dow Jones Industrial Averge | | FTSE 100 | | Nikkei 225 | |
| | Mean | (MC S.E.) | Mean | (MC S.E.) | Mean | (MC S.E.) |
| $\phi \mid y$ | 0.961 | (0.001) | 0.974 | (0.001) | 0.944 | (0.002) |
| $\sigma_n \mid y$ | 0.165 | (0.004) | 0.157 | (0.003) | 0.178 | (0.004) |
| $\beta = \exp(\mu) \mid y$ | 1.154 | (0.009) | 1.182 | (0.007) | 1.387 | (0.008) |

## B. Filtering

The most important problem in the SVM is the estimation of the time-varying variance, i.e., the state $h_t$, from the observations $Y_t = \{y_1,...,y_t\}$ for purpose of pricing assets, hedging, and risk management. Based on estimates of $\theta$ in Table 1, the problem of state estimation can be handled as an evaluation of the conditional density $f(h_n \mid Y_t,\theta)$, which is a predictive density when $n > t$, filtering density when $n = t$, and smoothing density when $n < t$. Once we have a sample $(h_t^{(1)},.......,h_t^{(M)})$ from $f(h_t \mid Y_t,\theta)$ it is possible to produce many interesting quantities. For instance, the mean of the sample draws provides an estimate of $E(h_t \mid Y_t,\theta)$, which is the minimum mean square error estimate of the log-volatility. We can estimate the likelihood function and the predictive distribution function $u_{t+1} = \Pr(y_{t+1} \leq y_{t+1}^f \mid Y_t,\theta)$ for a predicted $y_{t+1}^f$, both useful for diagnostic checking. For the standard linear Gaussian state space model, each density $f(h_n \mid Y_t,\theta)$ can be expressed by a Gaussian density and its mean vector and the variance-covariance can be obtained by computationally efficient recursive formulas such as the Kalman filter. For non Gaussian nonlinear state space model, each density has to be approximated using various types of approximation methods; among these methods a Monte Carlo particle filtering has been suggested by Kitagawa (1996), Kim et al., (1998), and Shephard and Pitt (1999).

Filtering consists of updating the state $h_t$ given contemporaneously available data $Y_t = (Y_{t-1}, y_t)$. Therefore the updating is triggered by the availability of new contemporaneous data $y_t$. As in any forecast exercise, before updating the forecast, it is useful to know first what has been forecast one period earlier, i.e. at time $t-1$. The general principle of updating can be expressed as: posterior=(data)(prior). Let the predictive density be $f(h_t \mid Y_{t-1},\theta)$; assume a Monte Carlo sample is drawn from this predictive density $\{h_{t|t-1}^{(1)},.......,h_{t|t-1}^{(M)}\}$. Each element of this set is called a particle. There are therefore $M$ particles which could be thought of as the forecast made at time $t-1$. Filtering consists of

updating $\{h_{t|t-1}^{(1)},\ldots\ldots,h_{t|t-1}^{(M)}\}$ conditional on information $y_t$ into a new set of particles $\{h_t^{(1)},\ldots\ldots,h_t^{(M)}\}$ called particle filter. The knowledge of the relation between the filtering density $f(h_t \mid Y_{t-1},y_t,\theta)$ and the predictive density $f(h_t \mid Y_{t-1},\theta)$ is therefore crucial. Using

Bayes theorem: $f(h_t \mid Y_t,\theta) = f(h_t \mid Y_{t-1},y_t,\theta) = \dfrac{f(y_t \mid h_t,Y_{t-1},\theta)f(h_t \mid Y_{t-1},\theta)}{f(y_t \mid Y_{t-1},\theta)}$

This relation stipulates clearly that a set of particles from the predictive density could be transformed into a set of filtered particles if reweighted by the Bayes coefficient[19]

$\dfrac{f(y_t \mid h_t,Y_{t-1},\theta)}{f(y_t \mid Y_{t-1},\theta)}$. In view of this relation, filtering is a recursive procedure involving a prediction stage followed by a resampling (updating) stage.

(i) The predictive density

In view of equation (2) of the SVM, the predictive density can be expressed as

$f(h_t \mid Y_{t-1},\theta) = \int f(h_t,h_{t-1} \mid Y_{t-1},\theta)dh_{t-1} = \int f(h_t \mid h_{t-1},\theta)f(h_{t-1} \mid Y_{t-1},\theta)dh_{t-1}$

Given a set of particles $\{h_{t-1}^{(1)},\ldots,h_{t-1}^{(M)}\}$ from $f(h_{t-1} \mid Y_{t-1},\theta)$ and observing that

$f(h_t \mid h_{t-1},\theta) = f_N(h_t \mid \mu+\phi(h_{t-1}-\mu),\sigma_\eta^2)$ has the normal density, the predictive density can

thus be approximated by $f(h_t \mid Y_{t-1},\theta) \cong \dfrac{1}{M}\sum_{j=1}^{M} f_N(h_t \mid h_{t-1}^{(j)},\theta)$. Particle predictors can be

sampled as $h_{t|t-1}^{(j)} \sim N(\mu+\phi(h_{t-1}^{(j)}-\mu),\sigma_\eta^2)$, yielding a set $\{h_{t|t-1}^{(1)},\ldots\ldots,h_{t|t-1}^{(M)}\}$ which can be considered to be M realizations from the predictive density and which can be used to approximate this density.

(ii) The filtering density

The filtering density can be approximated as

$f(h_t \mid Y_t,\theta) \cong \dfrac{1}{M}\sum_{j=1}^{M} \dfrac{f(y_t \mid h_t,Y_{t-1},\theta)f(h_t \mid h_{t-1}^{(j)},\theta)}{f(y_t \mid Y_{t-1},\theta)}$

As filtering consists of updating the predictors' set based on new data $y_t$, it has necessarily to use the probability distribution of the data as given in equation (1). To estimate particle filters, Kitagawa (1996) suggested computing the likelihood of each predictor given by $\alpha_t^{(j)} = f(y_t \mid h_{t|t-1}^{(j)},Y_{t-1},\theta)$ and using importance resampling to draw $h_t^{(j)}$ from the importance

---

[19] Note that the likelihood function is $f(y_t \mid Y_{t-1},\theta) = \int f(y_t \mid h_t,Y_{t-1},\theta)f(h_t \mid Y_{t-1},\theta)dh_t$ .

Note also that $f(y_t \mid h_t,Y_{t-1},\theta)$ is called the likelihood of $h_t$ .

density $f(h_t \mid h_{t-1}^{(j)}, \theta) = N(\mu + \phi(h_{t-1}^{(j)} - \mu), \sigma_\eta^2)$, with importance weights given by

$$\pi_t^{(j)} = \frac{\alpha_t^{(j)}}{(\alpha_t^{(1)} + \ldots\ldots + \alpha_t^{(M)})}.$$ The set of particles $\{h_t^{(1)}, \ldots\ldots, h_t^{(M)}\}$ generated by this algorithm are considered to be M realizations from the filtering density $f(h_t \mid Y_t, \theta)$.

Kim *et al.*, (1998) proposed an accept/reject algorithm for sampling $\{h_t^{(1)}, \ldots\ldots, h_t^{(M)}\}$ using a blanketing density $g^*(y_t, h_t, h_{t|t-1}, \phi)$. Let $h_{t|t-1} = \mu + \phi((\sum_{j=1}^{M} h_{t-1}^{(j)} / M) - \mu)$ and note also that the log likelihood function can be expressed as: $\log f(y_t \mid h_t, \theta) = \text{const} + \log f^*(y_t, h_t, \theta)$, where $\log f^*(y_t, h_t, \theta) = -0.5 h_t - 0.5 y_t^2 \exp(-h_t)$. Now expand $\log f^*(y_t, h_t, \theta)$ in a Taylor series around the known value $h_{t|t-1}$ as:

$\log f^*(y_t, h_t, \theta) = -0.5 h_t - 0.5 y_t^2 \exp(-h_t) \leq 0.5 h_t - 0.5 y_t^2 \{\exp(-h_{t|t-1})(1 + h_{t|t-1}) - h_t \exp(-h_{t|t-1})\} = \log g^*(y_t, h_t, h_{t|t-1}, \phi)$.

Using this inequality, the filtering density could be shown to be bounded above as:

$$f(h_t \mid Y_t, \theta) \cong \frac{1}{M} \sum_{j=1}^{M} \frac{f(y_t \mid h_t, Y_{t-1}, \theta) f(h_t \mid h_{t-1}^{(j)}, \theta)}{f(y_t \mid Y_{t-1}, \theta)} \propto \frac{1}{M} \sum_{j=1}^{M} f^*(y_t, h_t, \theta) f(h_t \mid h_{t-1}^{(j)}, \theta) \leq$$

$$\frac{1}{M} \sum_{j=1}^{M} g^*(y_t, h_t, h_{t|t-1}, \theta) f(h_t \mid h_{t-1}^{(j)}, \theta).$$

Given that $f(h_t \mid h_{t-1}^{(j)}, \theta) = N(\mu + \phi(h_{t-1}^{(j)} - \mu), \sigma_\eta^2)$, Kim *et al.*, (1998) showed that the right hand side of the above inequality is a mixture of normal densities

$$\frac{1}{M} \sum_{j=1}^{M} g^*(y_t, h_t, h_{t|t-1}, \theta) f(h_t \mid h_{t-1}^{(j)}, \theta) = \sum_{j=1}^{M} \pi_t^{(j)} f_N(h_t \mid h_{t-1}^{(j)}, \theta)$$

Therefore, the availability of a blanketing normal density suggests a simple accept-reject procedure for drawing $h_t$. First, draw a proposal value from the normal mixture density $\sum_{j=1}^{M} \pi_t^{(j)} f(h_t \mid h_{t-1}^{(j)}, \theta)$. Second, accept this value with probability

$f^*(y_t, h_t, \theta) / g^*(y_t, h_t, h_{t|t-1}, \phi).$[20]

The filtered volatilities along with the absolute values of the returns are shown in Figure 1. Contrasting volatility and returns, Figure 1 shows that periods of high volatility cause large

---

[20] The filtered volatilities can be used to estimate the likelihood function. Note the one-step-ahead prediction density, $f(y_{t+1} \mid Y_t, \theta)$ can be approximated as $(1/M) \sum_{j=1}^{M} f(y_{t+1} \mid h_{t+1}^{(j)})$, with $h_{t+1}^{(j)} \mid h_t^{(j)} \sim N(\mu + \phi(h_t^{(j)} - \mu), \sigma_\eta^2)$ and $h_t^{(j)}$ drawn from the filtering simulator.

changes in the returns, whereas periods of low volatility cause small changes in the returns. The level of the filtered volatility at any point in time enables to infer about the state of the market and the effects of the news hitting the market at that point of time. Tracking volatility therefore enables to assess the uncertainty expected by investors at a point in time compared to uncertainty in the whole sample period. Figure 1 shows volatility clustering, where periods of high volatility tend to be followed by periods of even higher volatility, and vice versa for periods of low volatility. Figure 1 also shows that volatility tends to be mean reverting around its long-run unconditional mean. Accordingly, a shock to volatility will tend to taper off, with a half-life of a few weeks. Filtered volatility is also summarized by boxplots (Figure 2).[21] The filtered distributions exhibit a large number of outliers which reflect frequent sizeable changes in volatility. These outliers could be attributed to major market news that cause significant portfolio adjustment.

## C. Diagnostics

In econometric analysis, diagnostic checks are required to assess the adequacy of the model and how well the fitted model accords with the observed data. If the model is invalid, then it can yield false inference; for this reason model checking is crucial to statistical analysis. In ARMA(p,q) model, diagnostics calculate the autocorrelation function for the residuals of the estimated model and determine whether those residuals appear to be white noise. In Gaussian state space model, the disturbances $\varepsilon_t$ and $\eta_t$ are normally distributed and serially

independent with constant variances. The Kalman filter delivers innovations $v_t$, which should

be uncorrelated and have a mean square error $F_t$. If the assumptions on the disturbances hold,

then the standardized one-step-forecast errors $e_t = \dfrac{v_t}{\sqrt{F_t}}$, t=1,.....,T, are also normally

distributed and serially independent with unit variance. These properties could be verified by means of the following large-sample diagnostic tests:

(i) Normality: the first four moments of the standardized forecast errors are given by

$$m_1 = \frac{1}{T}\sum_{t=1}^{T} e_t , \quad m_q = \frac{1}{T}\sum_{t=1}^{T}(e_t - m_1)^q , \quad q = 2,3,4.$$

Skewness and kurtosis are denoted by S and K, respectively, and are defined as:

$$S = \frac{m_3}{\sqrt{m_2^3}} ; \quad K = \frac{m_4}{m_2^2}$$

---

[21] A box plot reports the distribution of a variable in a condensed form. It shows the total range of a variable and the concentration in a central region, using the quartiles of the distribution, and the inter-quartile range (IQR).

and it can be shown that when the model assumptions are valid they are asymptotically

normally distributed as $S \sim N(0, \frac{6}{T})$, $K \sim N(3, \frac{24}{T})$. They can be combined as

$$N = T\{\frac{S^2}{6} + \frac{(K-3)^2}{24}\}$$ which asymptotically has a chi-square with 2 degrees of freedom on

the null hypothesis that the normality assumption is valid. Normality is also checked by the
QQ plot which is a graphical display of ordered residuals against their theoretical quantiles.
The 45 degree line is a reference line, the closer the residual plot to this line, the better the fit.

(ii) Serial correlation: when the model fits the data, the standardized forecast errors are
serially uncorrelated. A standard test statistic for serial correlation is based on the Box-Ljung

statistic. This is given by $BL(k) = T(T+2)\sum_{j=1}^{k} \frac{r_j^2}{(T-j)}$ where

$r_j = \frac{1}{Tm_2}\sum_{t=j+1}^{T}(e_t - m_1)(e_{t-j} - m_1)$ is the $jth$ sample autocorrelation coefficient.

However, time series models, such as the SVM, which are nonlinear and non-normal, have
residuals which are definitely not normal. Diagnostics require a Monte Carlo estimation of
the forecast error distribution function and a transformation of this distribution into uniform
and independent distributed random variables.[22] Diagnostics on the forecast errors of the
model are conducted by diagnostics on the transformed uniform random variables. Consider
the one-step-ahead conditional predictive distribution function

$u_{t+1} = G(y_{t+1} \mid y_t, \theta) = \Pr(y_{t+1} \leq y_{t+1}^f \mid Y_t = y_t, \theta)$.

The random variables $\{u_{t+1}, t=1,2,3,..\}$ are called the forecast distribution transformed

residuals. Then, $y_t$ will have a forecast distribution function $G_t$, $t = 1,2,3....,T$ given by the

SVM equation (1). Rosenblatt (1952) showed that the joint distribution of the data

$F(y_1,....,y_T \mid \theta)$ can be transformed into a uniform and independently distributed random

variables $\{u_{t+1}\}$ given by the conditional forecast distribution functions $u_{t+1} = G(y_{t+1} \mid y_t, \theta)$.

He pointed out that testing the sample of data $y_1,....,y_T$ is drawn from a population with

distribution function $F(y_1,....,y_T \mid \theta)$ is equivalent to testing the sample $\{u_{t+1}\}$ is taken from a

population uniformly distributed on the T-dimensional hypercube. Based on Rosenblatt's
transformation, one can judge the adequacy of the fitted model by the serial correlation in the

$\{u_{t+1}\}$ and by the nature of the distributional shape of $\{u_{t+1}\}$. Let $N_{t+1} = \Phi^{-1}(u_{t+1})$, where $\Phi$

is the standard normal distribution function, $N_{t+1}$ are called the normal forecast transformed

---

[22] The transformation of a cumulative distribution function $F(x)$ into a uniform random
variable $U$ on $(0,1)$ is given by the inverse transformation method (See Ross, 2000, pp. 590).

residuals with zero mean and unit variance. Smith (1985) showed that if the model is correct, the series $\{u_{t+1}\}$ and $\{N_{t+1}\}$ will conform to respective series of independent uniform, U(0,1), and independent normal random variables, respectively.

To show that $u_{t+1} = G(y_{t+1} \mid y_t, \theta)$ is uniformly distributed on (0,1) under the correctness of the model, observe that $u_{t+1}$ lies in the interval (0,1). Furthermore,

$$\Pr(u_{t+1} \leq a \mid y_t, \theta) = \Pr(G(y_{t+1} \mid y_t, \theta) \leq a \mid y_t, \theta) = \Pr(y_{t+1} \leq G^{-1}(a) \mid y_t, \theta) = a$$

which shows that $u_{t+1}$ is uniform. Finally, $u_{t+1}$, being a forecast error independent of $y_t$, forms also an independent sequence.

Now consider a simulation-based approach for estimating $u_{t+1}$. Having designed a filtering algorithm, it is a simple matter to sample from the one-step-ahead prediction density function. By definition, the prediction density is:

$$f(y_{t+1} \mid Y_t, \theta) = \int f(y_{t+1} \mid Y_t, h_{t+1}, \theta) f(h_{t+1} \mid Y_t, h_t, \theta) f(h_t \mid Y_t, \theta) dh_{t+1} dh_t ,$$ which can be

sampled by the method of composition as follows. For each value $h_t^{(j)}$ (j=1,...,M) from the filter, one samples $h_{t+1}^{(j)}$ from $h_{t+1}^{(j)} \mid h_t^{(j)} \sim N(\mu + \phi(h_t^{(j)} - \mu), \sigma_\eta^2)$.[23] Based on these M draws on $h_{t+1}$ from the predictive density, one can estimate the probability that $y_{t+1}$ will be less than the forecast $y_{t+1}^f$

$$G(y_{t+1}^f \mid Y_t, \theta) = \Pr(y_{t+1} \leq y_{t+1}^f \mid Y_t, \theta) \cong u_{t+1}^M = \frac{1}{M} \sum_{j=1}^{M} \Pr(y_{t+1} \leq y_{t+1}^f \mid h_{t+1}^{(j)}, \theta)$$

where $\Pr(y_{t+1} \leq y_{t+1}^f \mid Y_t, \theta)$ is derived using the fact that $y_{t+1} \mid h_{t+1}^{(j)}, \theta$ is a normal distribution $N(0, \exp(h_{t+1}^{(j)}))$. For each t=1,...,T, under the null of a correctly specified model $u_t^M$ converges in distribution to an independent and identically distributed uniform random variable as $M \to \infty$. This provides a valid basis for diagnostic checking. These variables can be mapped into the normal distribution, by using the inverse of the normal distribution $N_t^M = \Phi^{-1}(u_t^M)$ to give a standard sequence of independent and identically distributed normal variables, which are then transformed into one-step-ahead forecasts normed by their corrected standard errors. Under the null hypothesis that both the model and the prior are correct, $N_t^M$ are independent and standard normal with $E(N_t^M) = 0$ and $\text{var}(N_t^M) = 1$. The $N_t^M$ transforms are used to carry out the normality and Box-Ljung tests.

---

[23] Gerlach, Carter, and Kohn (1999), noting that the predictive distribution function included unknown latent variables that needed to be integrated out, suggested to compute the sequence $\{u_{t+1}\}$ using a combination of Markov Chain Monte Carlo and importance sampling.

Table 2 shows the results from some standard diagnostic checks on $N_1^M, ...., N_T^M$ produced by the fitted model. Under the correctness of the model, the diagnostics should indicate that the variables are Gaussian white noise.

| Table 2. Diagnostics of the SVM Using M=2500: Skewness, Kurtosis, Normality, Box-Ljung Statistics on 30 Lags, and Log-Likelihood (Log-lik). | | | | | |
|---|---|---|---|---|---|
| | Skewness | Kurtosis | Normality | BL(30) | Log-lik |
| Dow Jones Industrial Average | -0.027 | 3.134 | 0.848 (p-val=0.654) | 31.809 (p-val=0.376) | -1555.0 |
| FTSE 100 | -0.077 | 2.790 | 2.735 (p-Val=0.255) | 45.789 (p-val=0.033) | -1554.3 |
| Nikkei 225 | 0.065 | 2.676 | 4.939 (p-val=0.085) | 35.142 (p-val=0.237) | -1749.5 |

The skewness statistic shows skewness to be small and kurtosis shows small positive kurtosis for the Dow Jones Industrial and small negative excess kurtosis for the FTSE 100 and the Nikkei 225. The joint normality test is highly significant indicating normally distributed transforms $N_t^M$. The Box-Ljung statistic shows absence of autocorrelation in the model's residuals at all lags up to 30 lags for each market index. These diagnostics show that the SVM is adequate to fit the data on the three stock indices considered in this paper. The computed forecast uniforms $\{u_{t+1}\}$ and resulting correlograms and QQ plots are given in Figure 3. The graphical diagnostics show that the model performs quite well, although they reveal some outliers, and that there are no failures in the way the model has been fitted.

## V. CONCLUSION

The stochastic volatility model has become a basic tool for pricing assets, hedging, and managing risks. SVM is a discrete-time approximation to diffusion asset pricing models. SVM focuses solely on market news and filters their instantaneous impact on time-varying variance. Contrary to ARCH-GARCH models, where volatility is influenced by past returns, SVM is interested only in market news and the information flows. While SVM's news aspect makes it very appealing, complexity of its likelihood function, which could not be expressed in a closed form, is problematic. Simulation was a solution for estimating a SVM. In this paper, an MCMC simulation method was chosen for estimating the SVM for three market indices: the Dow Jones Industrial Average, the FTSE 100, and the Nikkei 225. In accordance with results for SVM, parameter estimates showed high volatility persistence, a tendency toward volatility mean reversion, and a good approximation of the information flow by the state transition equation. Boxplots of the filtered volatilities showed that frequent, important news kept hitting the market causing, high volatility in stock markets. Diagnostics checks were conducted and provided some evidence of the goodness of fit of the data by the SVM.

The SVM has been shown, in the recent literature, to be flexible enough for modeling stylized features of financial time series. For instance, fat-tailed distributions could be

modeled in SVM by allowing for a Student-$t$ distribution in the model's disturbances. Leverage effects could be studied in the SVM via correlation in its disturbances. Different assumptions on the priors of the parameters could also be made. The observation equation could be specified to allow for covariates and for level effects. All these extensions indicate the wide-ranging flexibility of the SVM and its ability to adequately model time-varying volatility.

Besides market participants, the SVM could have a potential value for the IMF and central banks as a tool for gauging market expectations and reactions to news, assessing monetary conditions and monetary policy, and influencing the timing and effectiveness of monetary operations and interventions. A better understanding of volatility in equity markets would support the IMF's role in promoting international financial stability. Similarly, such understanding would enhance central banks' ability to formulate and implement monetary policies capable of reducing asset price and interest rate volatility and influencing market expectations toward less volatile financial markets.

Figure 1. Filtered Volatilities and Absolute Returns for
the Dow Jones Industrial Average (top),
FTSE 100 (middle), and Nikkei 225 (bottom)
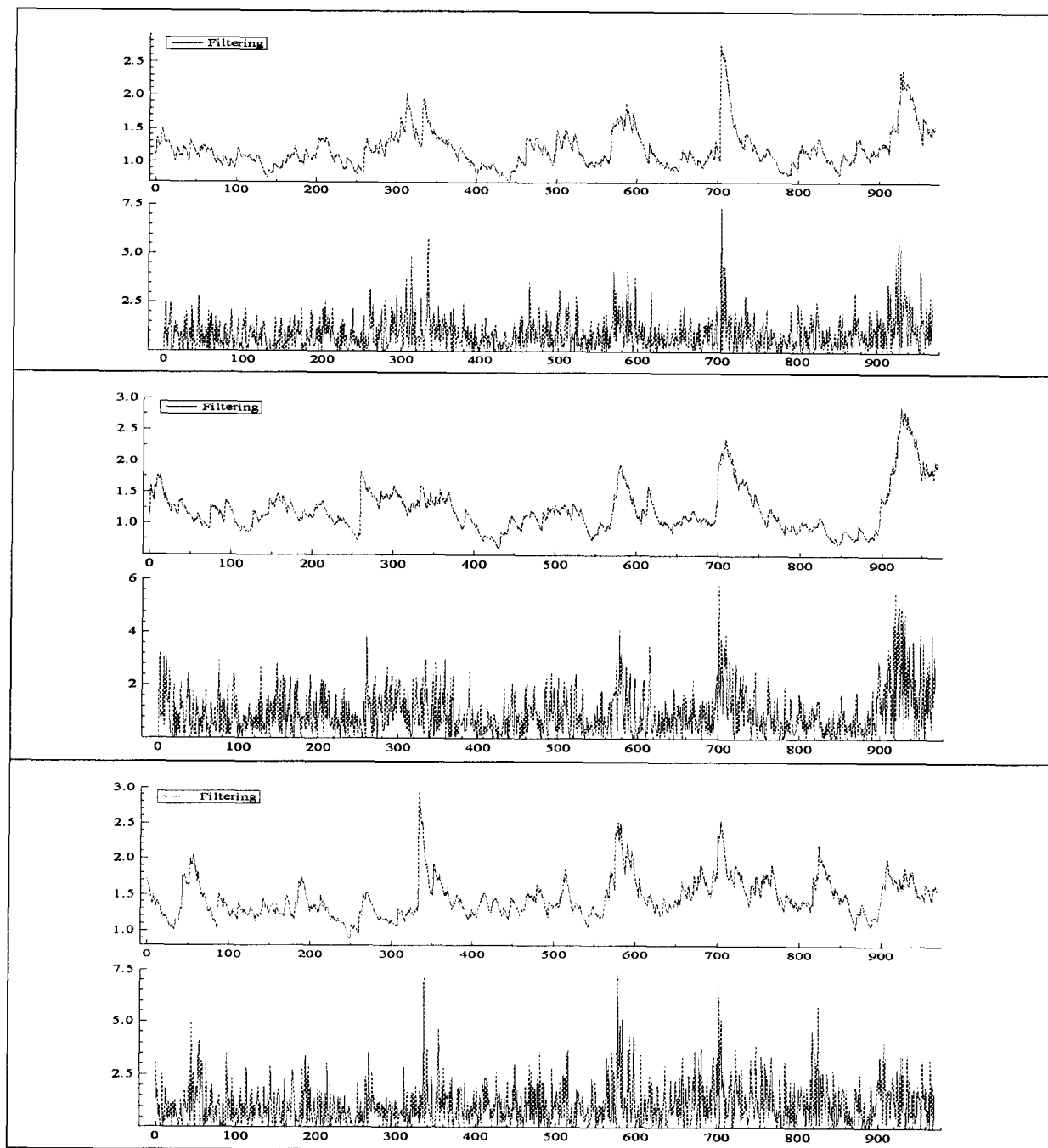
**Figure 2. Boxplots: Dow Jones Industrial Average (left),
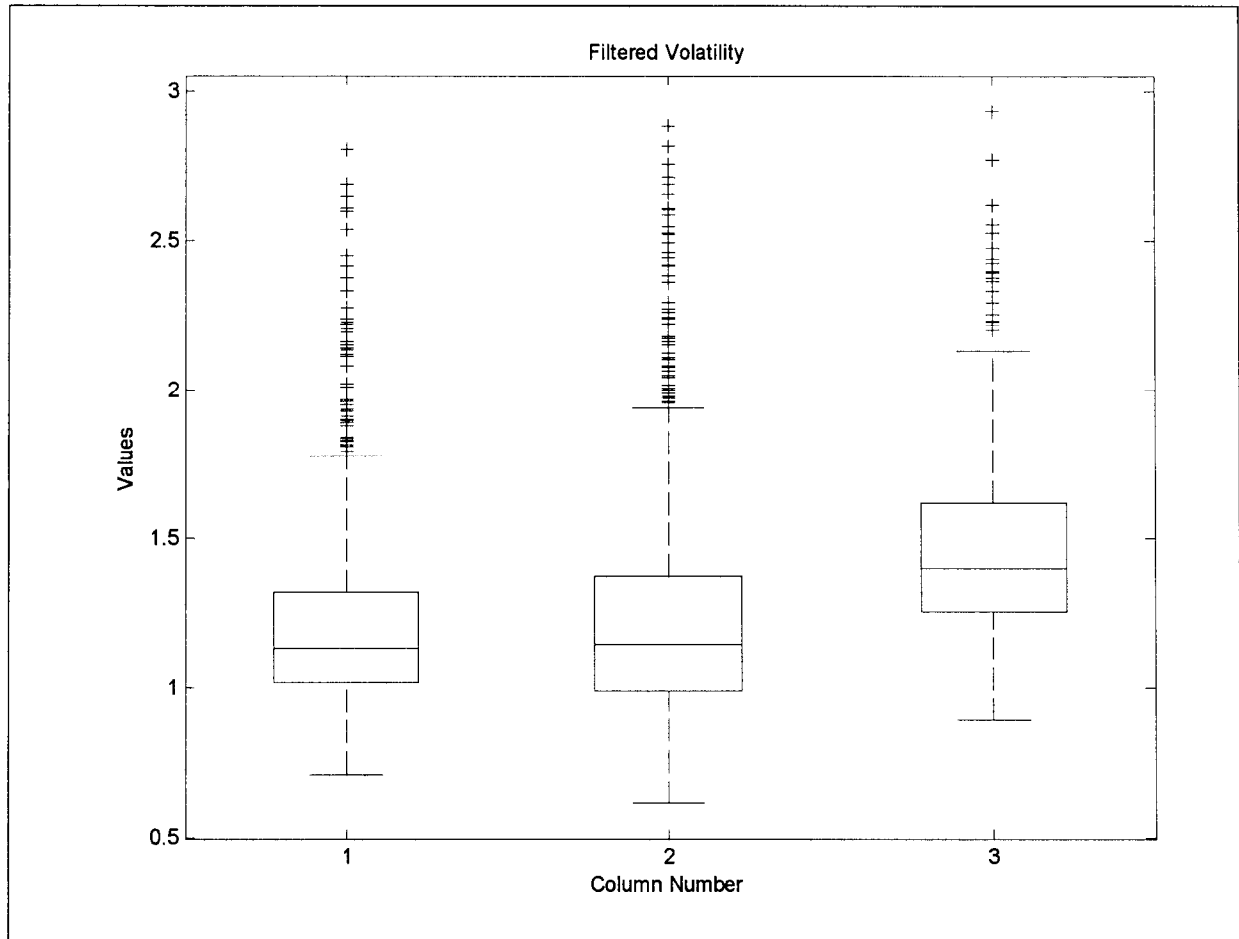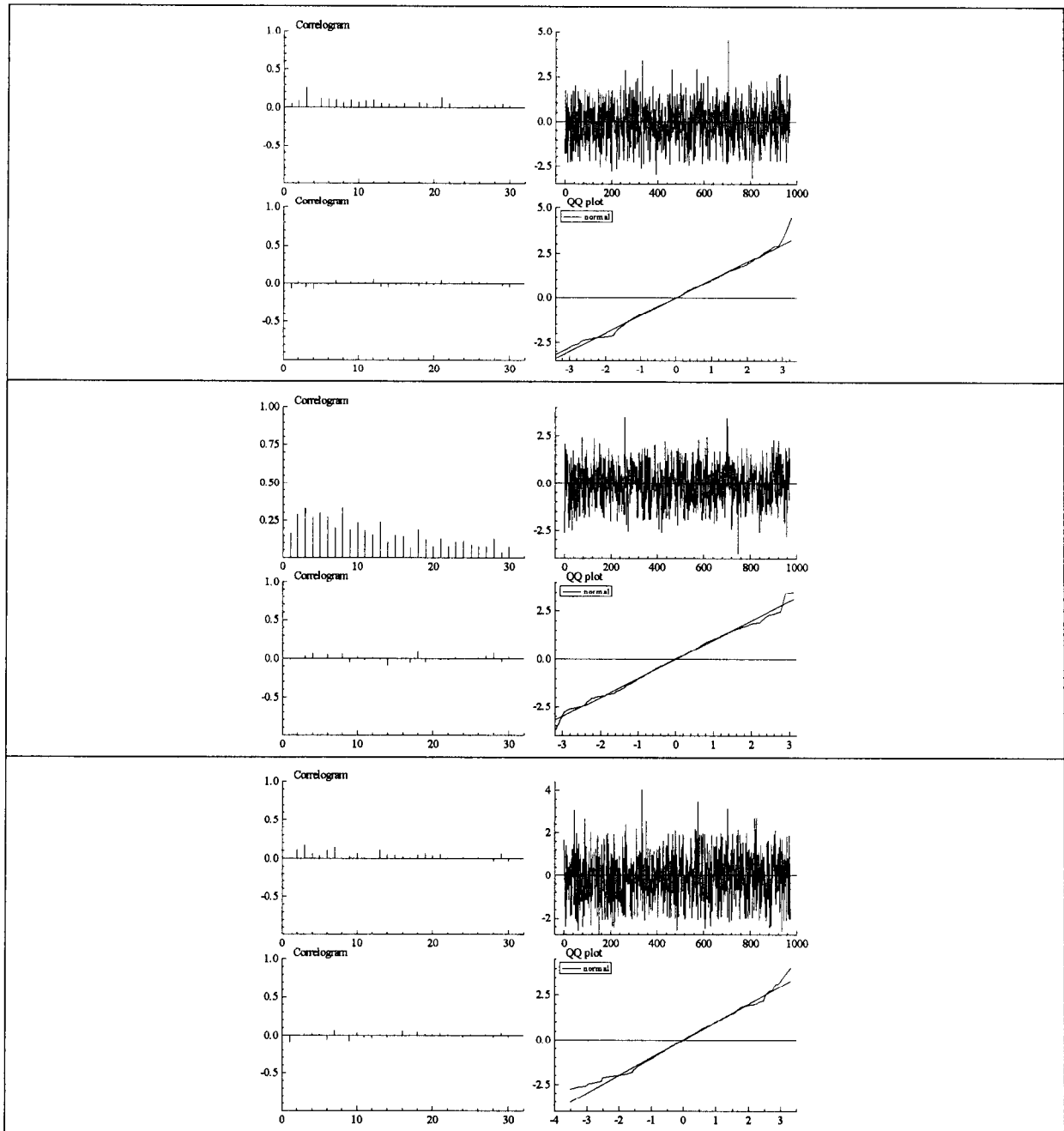FTSE 100 (middle), and Nikkei 225 (right).**

Figure 3: Diagnostics: Dow Jones Industrial (top),
FTSE 100 (middle), and Nikkei 225 (bottom).



Notes: In each panel, graphs are: (i) correlogram of $y_t^2$ (top left), (ii) the plot of the normalized innovations (top right), (iii) the correlogram of the normalized innovations (bottom left), and (iv) the QQ-plot (bottom right).

## References

Black, F., 1976. Studies of stock price volatility changes. Proceedings of the American Statistical Association, Business and Economic Statistics Section, 177-181.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy, 81, 637-654.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics, 51, 307-327.

Carlin, B. P., Polson, N.G., Stoffer, D.S., 1992. A Monte Carlo approach to nonnormal and nonlinear state spaced modeling. Journal of the American Statistical Association, 87, 493-500.

Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika, 81, 541-553.

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

Chib, S., 2001. Markovian chain Monte Carlo methods: computation and inference. In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics,Vol. 5, Elsevier Science B.V., pp 3569-3649.

Chib, S., and Greenberg, E., 1994. Bayes inference for regression models with ARMA(p,q) errors. Journal of Econometrics, 64, 183-206.

Chib, S., Nardari, F., Shephard, N., 2002. Markov Chain Monte Carlo methods for stochastic volatility models. Journal of Econometrics, 108, 281-316.

Cox, J.C., Ingersoll, J. E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica, 53, 385-407.

Dempster, A.P., Laird, N.M., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussions). Journal of the Royal Statistical Society, Series B, 39, 1-38.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. Econometrica, 50, 987-1007.

Fuller, W.A., 1996. Introduction to Time Series (2$^{nd}$ ed), John Wiley, New York.

Gerlach, R., Carter, C., Kohn, R., 1999. Diagnostics for time series analysis. Journal of Time Series Analysis, 20, 3, 309-330.

Ghysels, E., Harvey, A. C., Renault E., 1996. Stochastic volatility. In: Maddala, G. S., Rao, C. R. (Eds.),Handbook of Statistics, Vol. 14, Elsevier Science, pp 119-191.

Harvey, A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Review of Economic Studies, 61, 247-264.

Hull, J., White, A., 1987. The Pricing of options on assets with stochastic volatilities. The Journal of Finance, Vol. XLII, No 2, 281-300.

Jacquier, E., Polson, N.G., Rossi, P. E., 1994. Bayesian analysis of stochastic volatility models (with discussions). Journal of Business and Economic Statistics, 12, 371-417.

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

Kitagawa, G., 1996. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. Journal of Computational and Graphical Statistics, 5, 1-25.

Ross, S., 2000. Introduction to Probability Models, Seventh Edition, Harcourt, Academic Press.

Shephard, N., 1994. Partial non-Gaussian state space. Biometrika, 81, 115-131.

Shephard, N., 1996. Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinkley, D.V., Barndorff-Nielsen, O.E., (Eds.), Time Series Models in Econometrics, Finance, and Other Fields, Chapman & Hall, pp 1-67.

Shephard, N., and Pitt, M.K., 1999. Filtering via simulation: Auxiliary particle filters. Journal of the American Statistical Association, 94, 590-599.

Smith, J.Q., 1985. Diagnostic checks of non-standard time series models. Journal of Forecasting, Vol. 4, 283-291.

Rosenblatt, M., 1952. Remarks on a multivariate transformation. Annals of Mathematical Statistics, 23, 470-72.

Tanner, M. A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation (with discussions). Journal of the American Statistical Association, 82, 528-550.

Taylor, S.J., 1986. Modeling Financial Time Series. John Wiley, Chichester.

Vasicek O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics, 5, 177-188.