A Review of Stochastic Volatility: univariate and multivariate models.

K. Platanioti, E. J. McCoy, D. A. Stephens
Imperial College London

March 16, 2005

Abstract

We review a number of methods proposed in the literature for fitting a stochastic volatility (SV) model; namely, the Generalised Method-of-Moments, the Quasi-Maximum Likelihood method, the simulated EM algorithm and we concentrate on Bayesian approaches which utilize Markov chain Monte Carlo methods. In addition, we discuss the advantages and drawbacks associated with each of these procedures. Although univariate SV models have undergone a lot of research, the literature on multivariate models is limited. We review the Constant Conditional Correlation and the Factor SV models. Since both models impose a deterministic relation between variances and covariances, we consider them to be too restrictive. As possible alternatives, we suggest different models based on the Wishart distribution, which allow for a stochastic evolution of the covariance matrix and we discuss their properties.

1 Introduction

Volatility modelling plays an important role in the analysis of financial time series. One of the most well-known effects exhibited by the variance of financial products, such as stock or exchange rate returns, is the volatility clustering phenomenon. Essentially, this means that the variance of the returns exhibits high serial autocorrelation, which becomes evident by observing in the markets periods of high volatility, with large changes in the returns of an asset being followed by large ones, and periods of low volatility where small changes are followed by small ones. As this observation obviously is of great importance, researchers concentrated in developing and studying models which capture this effect. This is the reason why univariate stochastic volatility (SV) and Autoregressive Conditional Heteroscedasticity (ARCH) type models have been introduced and undergone a lot of research during the last two decades.

Apart from the volatility clustering phenomenon, there are other important stylized facts associated with financial returns series. Since the early sixties, both Mandelbrot (1963) and Fama (1965) have noted the fat-tailed nature of returns and it is now well known that the returns from an asset exhibit excess kurtosis and have heavier tails than the normal distribution. In addition, for stock-returns we often observe the so-called 'leverage' effect. This causes the conditional variance to respond assymetrically to rises and falls of the asset price, with negative returns producing a rise in volatility and vice versa. A final stylized fact, which cannot be captured by univariate descriptions is the covariation effect. Volatilities of different financial series are often seen to move together and large changes in one asset are matched by large movements in another. This is the result of either a formal linkage between a set of

assets, eg. currency exchange rates, or due to the assets being subject to the same overall environment and as a result, having factors in common which influence their behavior.

The ultimate goal of this report is the study of multivariate SV models, which have undergone far less research than the univariate case. Apart from the well-established constant conditional correlation (see Harvey et al. (1994)) and multivariate factor SV models (see Pitt & Shephard (1999)), other possible alternatives seem also suitable to describe the time evolution of the joint distribution of different assets. However, a thorough understanding of the properties and estimation procedures developed for the univariate SV model is imperative, before we can extend the ideas to a multi-dimensional equivalent. For this reason, the rest of this section discusses the univariate SV model and its properties, section 2 reviews some of the methods proposed in the literature for the estimation of the model and discusses advantages and drawbacks associated with each of these. Section 3 reviews the properties of the constant conditional correlation and factor SV models and discusses other potential options built on the Wishart distribution. Finally, section 4 concludes and offers guidelines for future research.

1.1 The univariate SV model

Although ARCH type models are now used routinely and are easy to estimate, there are also a number of drawbacks associated with most of these models. Nelson (1991) notes two of these. Firstly, the constraints imposed on the parameters of these models, to ensure the positivity of the conditional variance, are often violated during the process of estimation. Secondly, a random oscillatory behavior of the conditional variance process is ruled out.

The basic alternative to ARCH type models is the stochastic volatility (SV) model due to Taylor (1986) which allows the variance of the returns to be an unobserved random process. SV models overcome the drawbacks encountered with ARCH models and fit more naturally into the theoretical framework within which much of modern finance theory has been developed. In particular, as shown for example in Taylor (1994), SV models can easily be seen to have simple continuous-time analogues used for option pricing. Moreover, their form nests them into the class of state-space models, studied extensively by Harvey (1989) and West & Harrison (1997). If we let y_t denote the return from an asset at time t, the most popular SV parameterization sets

$$y_t = \exp(h_t/2)\,\varepsilon_t \tag{1.1a}$$

$$h_t = \gamma + \phi h_{t-1} + \eta_t \tag{1.1b}$$

where the ε_t's and η_t's are i.i.d. $N(0,1)$ and $N\left(0,\sigma_\eta^2\right)$ random variables, respectively. Because the SV model, allows the log of the volatility to evolve, it is ensured that the variance of the process always remains positive without the need of further constraints.

The latent process h_t in (1.1) can be interpreted as the random and uneven flow of new information in financial markets, while ϕ as the persistence in the volatility. There are also alternative parameterizations of the SV model. For example, Jacquier et al. (1994) model the log of h_t as an AR(1) process so that $y_t = \sqrt{h_t}\varepsilon_t$ and $\log h_t = \gamma + \phi \log h_{t-1} + \eta_t$, which is clearly equivalent to (1.1). Alternatively, we can write $y_t = \beta \exp(h_t/2)\,\varepsilon_t$, where β is a constant scaling factor and can be thought of as the

modal instantaneous volatility. The inclusion of the term β in (1.1a) removes the need of including the constant term γ in the AR(1) process (1.1b) so that $h_t = \phi h_{t-1} + \eta_t$.

The advantage of using SV models lies in the fact that they provide greater flexibility in describing stylized facts. Many financial series are often found to exhibit a kurtosis which is greater than the one resulting by incorporating conditional heteroscedasticity into a normal process. This effect can be incorporated in SV models by allowing ε_t in (1.1a) to have a Student t-distribution. This generalization has been studied, amongst others, by Ruiz (1994), Harvey et al. (1994), Chib et al. (2002) and Jacquier et al. (2004). In addition, although most ARCH type models fail to accommodate for the 'leverage' effect, the latter phenomenon can also be described by the SV model by introducing a negative contemporaneous correlation between the errors η_t and ε_t in (1.1), an extension studied by Harvey & Shephard (1996) and Jacquier et al. (2004).

Before we proceed to describe how the SV model can be estimated, we discuss its basic properties. Throughout we assume that the errors, ε_t and η_t, in (1.1) are independent and Gaussian. The process h_t in (1.1b) is a standard Gaussian AR(1) process and if $|\phi| < 1$, h_t is strictly stationary with unconditional mean and variance given respectively by

$$\mu_h = E(h_t) = \frac{\gamma}{1-\phi} \quad \text{and} \quad \sigma_h^2 = Var(h_t) = \frac{\sigma_\eta^2}{1-\phi^2}. \tag{1.2}$$

Since y_t is the product of two processes, white noise and $\exp(h_t/2)$, y_t is stationary if and only if h_t is stationary. Thus, the restrictions needed to ensure stationarity of y_t are just the standard restrictions that ensure stationarity of an AR(1) process. Moreover, since the ε_t's are, by definition, zero mean i.i.d., $E(y_t) = E(y_t \mid \Psi_{t-1}) = 0$. Thus, y_t is zero mean and in addition, given the independence of ε_t and η_t, the autocovariance function (ACVF) of y_t is

$$s_{y_t}(\tau) = E(y_t y_{t-\tau}) = E\left[\exp\left(\frac{h_t}{2} + \frac{h_{t-\tau}}{2}\right)\right] E[\varepsilon_t \varepsilon_{t-\tau}] = 0, \qquad |\tau| > 0.$$

Therefore, the series y_t is a martingale difference (see Grimmett & Stirzaker (2001) for the definition). Furthermore, since the distribution of ε_t is symmetric, as is the standard normal or even the Student-t, all the odd moments of y_t are zero.

We now investigate the even moments of y_t. Assuming $|\phi| < 1$, $h_t \sim N\left(\mu_h, \sigma_h^2\right)$ and hence, $\exp(h_t)$ is log-normal distributed. In addition, from standard properties of the log-normal distribution, we have $E\left[\exp(h_t)^j\right] = \exp\left\{j\mu_h + \frac{1}{2}j^2\sigma_h^2\right\}$, so that, if r is even and h_t is stationary, all the even moments of y_t exist and are given by

$$E(y_t^r) = E\left[(\exp h_t)^{r/2}\right] E[(\varepsilon_t)^r] = \exp\left\{\frac{r}{2}\mu_h + \frac{r^2}{8}\sigma_h^2\right\} \frac{r!}{2^{r/2}\left(\frac{r}{2}\right)!}. \tag{1.3}$$

In particular, from (1.3) we deduce that

$$Var(y_t) = E\left[(y_t)^2\right] = \exp\left\{\mu_h + \frac{1}{2}\sigma_h^2\right\}$$

and hence, that, if h_t is stationary, y_t is a white-noise process. The fourth moment is $E\left[(y_t)^4\right] = 3\exp\left\{2\mu_h + 2\sigma_h^2\right\}$ and so the coefficient of kurtosis for y_t is $-3 + E\left(y_t^4\right)/E^2\left(y_t^2\right) = 3\left(\exp\left(\sigma_h^2\right) - 1\right) > 0$. Thus, y_t has a leptokurtic, symmetric distribution.

The dynamic properties of the SV model appear most clearly if we square y_t and take logarithms, so that

$$\log y_t^2 = h_t + \log \varepsilon_t^2. \tag{1.4}$$

If ε_t has a standard normal distribution, then (see Abramowitz & Stegun (1970)) $\log \varepsilon_t^2$ has a log-chi-square distribution with mean $\psi(1) - \log 2 \simeq -1.2704$ and variance $\pi^2/2 \simeq 4.9348$, where $\psi(\cdot)$ is the digamma function. Thus, if we define $\xi_t = \log \varepsilon_t^2 + 1.2704$, then clearly ξ_t is zero mean i.i.d. with variance $\pi^2/2$ and we may rewrite (1.4) as

$$\log y_t^2 = -1.2704 + h_t + \xi_t. \tag{1.5}$$

Therefore, it follows that $\log y_t^2$ is a linear process which is the sum of the AR(1) process h_t and white noise. Hence, it behaves approximately like an ARMA(1, 1) process, with its autocorrelation function (ACF) being equivalent to that of an ARMA(1, 1) process and given by (see Ghysels et al. (1996))

$$\rho_{\log y_t^2}(\tau) = \frac{\phi^\tau}{1 + \left(\pi^2/2\sigma_h^2\right)}, \quad \tau = 1, 2, \ldots.$$

A detailed discussion on the properties of the SV model and the corresponding moments for the case with fat-tailed errors appears in Ghysels et al. (1996). Next, we proceed with the estimation of the model.

2 Estimation of the SV Model

One of the most important limitations of SV models is that, unlike ARCH-type models, the distribution of y_t conditional on Ψ_{t-1} does not possess an analytic expression. As a result, the likelihood function is hard to evaluate. One way of deriving the likelihood is by integrating the latent log-volatilities out of the joint probability distribution. Let us denote $\boldsymbol{y} = (y_1, \ldots, y_T)'$ the vector of T consecutive observations, $\boldsymbol{h} = (h_1, \ldots, h_T)'$ the vector of corresponding log-volatilities and $\boldsymbol{\theta} = \left(\gamma, \phi, \sigma_\eta^2\right)$ the vector of hyperparameters. Then, the likelihood is given by

$$\mathcal{L}(\boldsymbol{y}; \boldsymbol{\theta}) = \int p(\boldsymbol{y}, \boldsymbol{h}|\boldsymbol{\theta})\, d\boldsymbol{h} = \int p(\boldsymbol{y} \mid \boldsymbol{h}, \boldsymbol{\theta})\, p(\boldsymbol{h} \mid \boldsymbol{\theta})\, d\boldsymbol{h}. \tag{2.1}$$

This last integral is of dimension equal to the sample size, T, its evaluation requires the use of numerical procedures and this makes the estimation of the hyperparameters, $\boldsymbol{\theta}$, via Maximum Likelihood quite involved. Various methods have been proposed during the last decade for the estimation of SV models. Most involve the estimation or approximation of the integral in (2.1) and others make use of the method-of-moments procedures. Surveys on the estimation of SV models can be found in Shephard (1996), Ghysels et al. (1996) and Broto & Ruiz (2004).

One of the early methods proposed for the estimation of general non-linear, non-Gaussian parameter-

driven models and probably the one that is conceptually more straightforward is due to Kitagawa (1987). He suggested using numerical integration techniques to approximate the integral in (2.1). This method can be considered as an extended Kalman filter (cf Anderson & Moore (1979)). However, as Shephard (1993) and Tanizaki & Mariano (1998) point out, numerical integration suffers from a number of drawbacks such as accumulation of computational errors, long computational time and tedious programming especially in higher-dimensional cases. Therefore, the estimation of multivariate SV models for example, via this method is ruled out.

Another approach that is in a similar spirit to the latter procedure uses stochastic integration. The method was initially proposed by Geweke (1989) for general econometric models and was later refined and applied to the SV model by Sandmann & Koopman (1998) and Danielsson (1994). It is well known that Monte Carlo integration deals better with higher dimensions and is easier to implement compared to deterministic numerical integration techniques. Nonetheless, the procedure suffers too from the accumulation of computational errors on top of which stochastic errors are also added. In general, numerical integration techniques, which try to approximate the integral in (2.1) and then maximize the estimate with respect to the parameters, have not met great success in the SV literature. In what follows, we review some popular procedures adapted for the estimation of the SV model.

2.1 Generalized Method-of-Moments (GMM)

The simplest estimation procedure of SV models is the Method-of-Moments suggested by Taylor (1986). Later, Melino & Turnbull (1990) proposed to estimate the unknown parameters via the Generalized Method-of-Moments procedure, due to Hansen (1982). The key advantage of GMM is that it does not require the specification of the likelihood function, but only certain moment conditions are needed. The method is based on the fact that the theoretical population moments should approximately match the corresponding sample values.

More precisely, given a sample $\boldsymbol{y} = (y_1, \ldots, y_T)$, the GMM procedure requires the minimization of the criterion function $Q = \boldsymbol{g}'\boldsymbol{W}\boldsymbol{g}$ with respect to $\boldsymbol{\theta} = \left(\gamma, \phi, \sigma_\eta^2\right)$, where

$$\boldsymbol{g}' = \left[\frac{1}{T}\sum y_t^2 - E\left(y_t^2\right), \frac{1}{T}\sum y_t^4 - E\left(y_t^4\right), \frac{1}{T}\sum y_t^2 y_{t-1}^2 - E\left(y_t^2 y_{t-1}^2\right), \ldots, \frac{1}{T}\sum y_t^2 y_{t-\tau}^2 - E\left(y_t^2 y_{t-\tau}^2\right)\right],$$

(the theoretical values of $E\left(y_t^2 y_{t-\tau}^2\right)$, for $\tau \geq 1$, can be found in Taylor (1986) or Ghysels et al. (1996)) and $\boldsymbol{W}$ is a $(\tau+2) \times (\tau+2)$ positive definite, symmetric weighting matrix reflecting the importance given to matching each of the moments. Thus, the larger the $(1,1)$ element of $\boldsymbol{W}$, the greater is the importance of the unconditional variance of the observations being as close as possible to the unconditional sample variance. Normally Q will be minimized using numerical optimization procedures. A discussion on the GMM procedure and possible choices for the weighting matrix can be found in Hamilton (1994).

Empirical applications of the GMM approach in the context of SV models can be found in Melino & Turnbull (1990), Ruiz (1994), Jacquier et al. (1994) and Andersen & Sorensen (1996). The great advantage realised by using the GMM method is that the likelihood function in (2.1) does not need to be computed. However, this can also be a drawback in that GMM often does not make efficient use of

all the information contained in the sample.

It can be shown (cf Hansen (1982)) that the GMM estimator is consistent and asymptotically normal. On the other hand, GMM is asymptotically consistent if the observations y_t are stationary. When the persistence in the latent process, h_t, is high, i.e. ϕ is close to unity, as is usually the case in practice, the GMM estimator works poorly. In their simulation studies Jacquier et al. (1994) find substantial bias in their estimates, especially for σ_η^2, and report sampling variability, which results in large root mean squared errors, when there is high persistence and low coefficient of variation, defined as $C.V. = Var\left[\exp\left(h_t\right)\right] / \left\{E\left[\exp\left(h_t\right)\right]\right\}^2$. Another drawback to the GMM procedure is that the parameter estimates are not invariant, meaning that if the model is reparameterized so that $\boldsymbol{\psi} = f\left(\boldsymbol{\theta}\right)$, then $\widehat{\boldsymbol{\psi}} \neq f\left(\widehat{\boldsymbol{\theta}}\right)$. This is important because it means that estimation based on different parameterizations but on the same dataset will yield inconsistent results.

Finally, another disadvantage of the GMM estimation is that it does not deliver filtered or smoothed estimates of h_t and therefore, a second estimation procedure needs to be used, eg. the Kalman filter, in order to obtain these. Nonetheless, it should be pointed out that the simplicity of the GMM procedure and the fact that it can deal with both standard normal and t-distributed observation errors makes it an attractive estimation method.

2.2 Quasi-Maximum Likelihood (QML) estimation

A simpler approach that yields both estimates of the hyperparameters as well as filtered and smoothed estimates of h_t was proposed by Harvey et al. (1994). The method is based on the linearized form of the SV model. Assuming that the errors $\varepsilon_t \sim N\left(0,1\right)$ and denoting $w_t = \log y_t^2$, as already seen, the SV model can be written as

$$w_t = -1.2704 + h_t + \xi_t, \tag{2.2a}$$

$$h_t = \gamma + \phi h_{t-1} + \eta_t, \tag{2.2b}$$

where $\xi_t = \log \varepsilon_t^2 - E\left(\log \varepsilon_t^2\right)$, with $\sigma_\xi^2 = Var\left(\xi_t\right) = \pi^2/2$. Clearly, the model in (2.2) is a linear state-space model, as defined by Harvey (1989), but non-normal since the errors, ξ_t, in the observation equation (2.2a) do not have a normal distribution. In addition, it should be noted that even if ε_t and η_t had been contemporaneously correlated in the initial formulation (1.1) of the model, then (see Harvey et al. (1994)) it can be proved that ξ_t and η_t are uncorrelated. This is because the original observations are squared and thus all the information on the dependence between ε_t and η_t is lost.

The idea of the QML estimation method is to treat the observation errors, ξ_t, as though they were i.i.d $N\left(0, \pi^2/2\right)$ and apply the standard Kalman filter equations to the state-space model in (2.2). The Kalman filter (see Appendix A) will produce one-step-ahead forecasts of the observations, w_t, and the log-volatilities, h_t, as well as filtered estimates of the latter. Given a set of observations, the recursions can also be used to construct the Gaussian likelihood of the data via the prediction error decomposition, as shown in Harvey (1989). This Gaussian form of the likelihood can then be maximized with respect to the parameters of the model, typically using numerical procedures, to yield QML estimates of the unknown parameters. Before proceeding with a more detailed description of the method, we make one

more simplifying transformation of the model in (2.2), following Ruiz (1994).

Let us consider $E(w_t)$, assuming that $|\phi| < 1$ in (2.2b), so that h_t is stationary. Clearly, $E\{w_t\} = \gamma^* = -1.2704 + \mu_h = -1.2704 + \gamma/(1-\phi)$. Moreover, if we denote $w_t^* = w_t - \gamma^*$ and $\alpha_t = h_t - \mu_h$ to be new mean-centered observations and states respectively, then the model (2.2) can be rewritten as

$$w_t^* = \alpha_t + \xi_t, \tag{2.3a}$$

$$\alpha_t = \phi\alpha_{t-1} + \eta_t. \tag{2.3b}$$

The latter state-space model does not explicitly contain the constant term γ of the state-transition equation (2.2b). A consistent estimator of γ^* is given by the sample mean of $w_t = \log y_t^2$ and is also the QML estimator of γ^*. We can now apply the Kalman filter to the model in (2.3) and obtain the QML estimates of $\boldsymbol{\theta} = \left(\phi, \sigma_\eta^2\right)'$. Once the estimates ϕ and σ_η^2 are available, the QML estimator of γ will be given by $\gamma = (1-\phi)\left(1.2704 + \frac{1}{T}\sum \log y_t^2\right)$.

We now concentrate on the estimation of $\boldsymbol{\theta} = \left(\phi, \sigma_\eta^2\right)'$. The Kalman filter recursions compute the one-step-ahead prediction, $a_{t|t-1}$, and the filtered estimates, $a_{t|t}$, of the unobserved states α_t assuming that the observations sequentially become available. In particular, since $|\phi| < 1$, the states α_t are stationary and the recursions are started off with the unconditional mean and variance of α_t. Thus, starting with $a_{0|0} = E(\alpha_t) = 0$ and $P_{0|0} = Var(\alpha_t) = \sigma_\eta^2/\left(1-\phi^2\right)$, the one-step-ahead prediction estimates of α_t and their mean square errors (MSE's) are respectively given by (cf Harvey (1989))

$$a_{t|t-1} = \phi a_{t-1|t-1}, \quad P_{t|t-1} = \phi^2 P_{t-1|t-1} + \sigma_\eta^2, \quad t = 1, \ldots, T, \tag{2.4}$$

while the filtered estimates, $a_{t|t}$, and their MSE's, $P_{t|t}$, are respectively given by

$$a_{t|t} = a_{t|t-1} + P_{t|t-1} f_t^{-1}\left(w_t^* - a_{t|t-1}\right), \quad P_{t|t} = P_{t|t-1} - P_{t|t-1}^2 f_t^{-1}, \quad t = 1, \ldots, T, \tag{2.5}$$

where the terms $w_t^* - a_{t|t-1}$ are the innovations in predicting w_t^* given the past information $\Psi_t^* = \left\{w_{t-1}^*, \ldots, w_1^*\right\}$ and $f_t = P_{t|t-1} + \sigma_\xi^2$ are the MSE's of the one-step-ahead prediction estimates of w_t^*. It should be noted that, since the errors ξ_t are not truly Gaussian, the filtered and smoothed estimators $a_{t|t-1}$ and $a_{t|t}$ are only minimum mean square linear estimators (MMSLE's) of the unobserved variable α_t, given the observations up to time $t-1$ and t, respectively. By this we mean that the estimators $a_{t|t-1}$ and $a_{t|t}$ are only optimal in the sense that they yield the minimum MSE within the class of all linear estimators.

In addition, if we treat the observations w_t^* as though they were indeed Gaussian, then their mean and variance conditional on Ψ_t^* are respectively given by $a_{t|t-1}$ and f_t. Consequently, we can construct the Gaussian log-likelihood, called the quasi-log-likelihood function:

$$l_q(\boldsymbol{\theta}; w^*) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log f_t - \frac{1}{2}\sum_{t=1}^{T}\frac{\left(w_t^* - a_{t|t-1}\right)^2}{f_t}. \tag{2.6}$$

Maximizing (2.6) with respect to $\boldsymbol{\theta}$, typically using a numerical optimization procedure, yields the corresponding QML estimates. The QML estimators of $\boldsymbol{\theta}$ are consistent with asymptotically normal

distribution; see Ruiz (1994). Moreover, once the QML estimates of $\boldsymbol{\theta}$ have been computed, starting with the final filtered estimate $a_{T|T}$ and its MSE $P_{T|T}$, the following backward recursions (see Anderson & Moore (1979)) produce the smoothed estimates $a_{t|T}$ of α_t along with their MSE $P_{t|T}$

$$a_{t|T} = a_{t|t} + \phi P_{t|t} P_{t+1|t}^{-1} \left(a_{t+1|T} - \phi a_{t|t} \right) \tag{2.7a}$$

$$P_{t|T} = P_{t|t} + \phi^2 P_{t|t}^2 P_{t+1|t}^{-2} \left(P_{t+1|T} - P_{t+1|t} \right), \quad t = T-1, \ldots, 1. \tag{2.7b}$$

Also note that the QML procedure can be applied to the SV model when ϕ is set equal to one and the log-volatilities follow a random walk. The procedure is the same as before and only slight modifications need to be made. The fact that the Kalman filter can be applied even when the log-volatilities are allowed to follow a random walk is an obvious advantage of the QML estimation over the GMM estimation procedure. Secondly, it makes the QML procedure quite general, in the sense that the dynamics of the latent process can take many different forms. Furthermore, the QML estimation is not restricted only to the case when $\varepsilon_t \sim N(0,1)$, but with minor modifications can also be used to estimate a SV model with $\varepsilon_t \sim$Student-t; see Harvey et al. (1994). Finally, it should be noted that an asymmetric SV model, where the errors ε_t and η_t are correlated, can still be estimated via the QML procedure, despite the fact that when squaring the observations, the relevant information on the dependence between ε_t and η_t is lost; see Harvey & Shephard (1996).

To sum up, let us discuss some of the advantages and drawbacks to the QML estimation. From the discussion so far, it has been obvious that the QML method is simple to implement and computationally efficient. Furthermore, since the QML estimation procedure depends only on first and second order moment properties, it can be applied without modification if the distribution of the error terms changes and allows the dynamics of the unobserved components to take many different forms. Harvey et al. (1994) show also how the method can be generalized to multivariate SV models. The method produces filtered and smoothed estimates of h_t and in addition, missing or irregularly spaced observations can also be handled via the Kalman filter. Finally, the QML estimators are consistent and in most practical applications as in Ruiz (1994), Jacquier et al. (1994), it can be seen that they outperform in terms of efficiency the corresponding GMM estimators.

On the other hand, the QML procedure can be termed inefficient because it does not rely on the exact log-likelihood, but approximates the density of a $\log\left(\chi_1^2\right)$ variable by a normal density. A comparison of these densities, as in Figure 1, illustrates that this approximation is rather inappropriate. In particular, the adequacy of the approximation depends critically on the true parameter values. For large values of σ_η^2, the AR(1) process, h_t, dominates ξ_t, the non-Gaussian error term in the observation equation, so that the normal approximation may be adequate and the QML approach is close to optimal. Nonetheless, as σ_η^2 decreases, the approximation worsens and for small values of σ_η^2, usually found in practice, the QML estimates can be extremely biased and have high root mean square error.

As an illustrative example of the precision of the QML method, we have generated a sample of 1,000 values from a SV model with parameters $\gamma = 0$, so that $\mu_h = \gamma/(1-\phi) = 0$, $\phi = 0.9$ and $\sigma_\eta^2 = 0.1$. The size of the sample is typical for financial data, as are the chosen parameter values. We use the simulated annealing algorithm to find an approximate maximum of the quasi-likelihood function (2.6).

Figure 2 shows plots of the generated samples of ϕ and σ_η^2 once the algorithm has converged, with

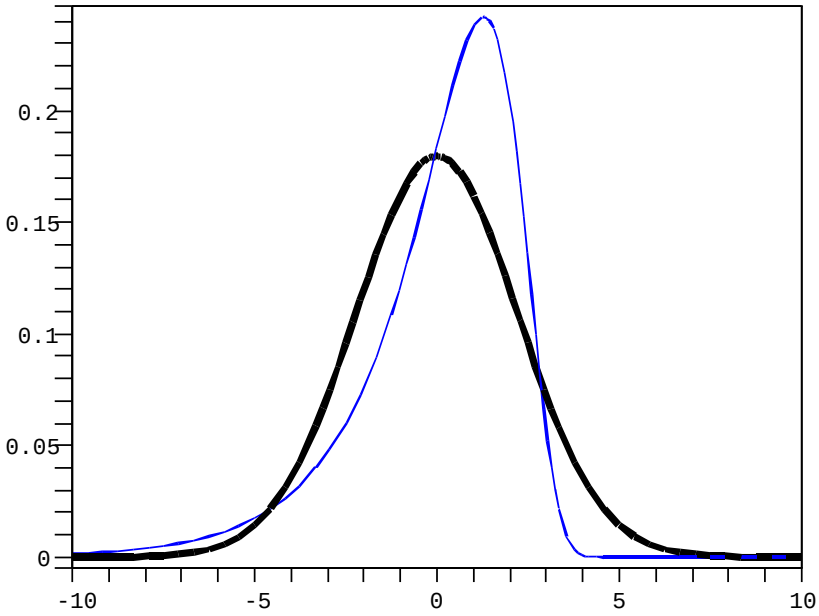

Figure 1: Comparison of the $\log\left(\chi_1^2\right)$ density (thin solid line) with the $N\left(0, \pi^2/2\right)$ density (thick solid line).

the straight horizontal lines indicating the true value of the parameters. It is readily seen that the QML estimate of ϕ exhibits a small positive bias while the corresponding estimate of σ_η^2 is quite accurate. In fact, the mean of the sample obtained for ϕ is equal to 0.9162 while the corresponding sample mean for σ_η^2 is equal to 0.0967, an estimate which does not exhibit any significant bias. Finally, the estimate for μ_h results in a value equal to 0.1133 while the estimate for γ is obtained by the transformation $\gamma = (1-\phi)\,\mu_h$ and turns out to be equal to 0.0095, which is quite accurate.

As already noted, apart from estimating the parameters of a SV model, it is often desired to obtain also estimates of the unobserved volatilities. Since our series of data points has been simulated, this means that we actually do have access to the latent volatility process and this in turn gives us the opportunity to gain some insight in the precision of the QML method in terms of smoothing. Using the estimates of ϕ and σ_η^2 and the recursions (2.7), we produce smoothed estimates of the volatilities which are plotted along with the corresponding true values in Figure 3. As noted earlier, the fixed interval smoothing algorithm yields only MMSLE's of the undelying volatility process and therefore, we should not expect great accuracy from our estimates. This is readily seen in Figure 3, where the estimated volatilities do trace the true values in some regions but in general, when compared to the true volatility process they do not exhibit the same variability.

Finally, a drawback to the QML procedure worth noting is the so-called 'inlier' problems encountered by taking logarithms of very small numbers. In particular, when the asset returns, y_t, are close to zero $\log y_t^2$ is a large negative number and in the extreme case where $y_t = 0$, $\log y_t^2$ is not defined. For the cases where the inlier cannot be assumed to be an irregular observation, researchers have proposed two

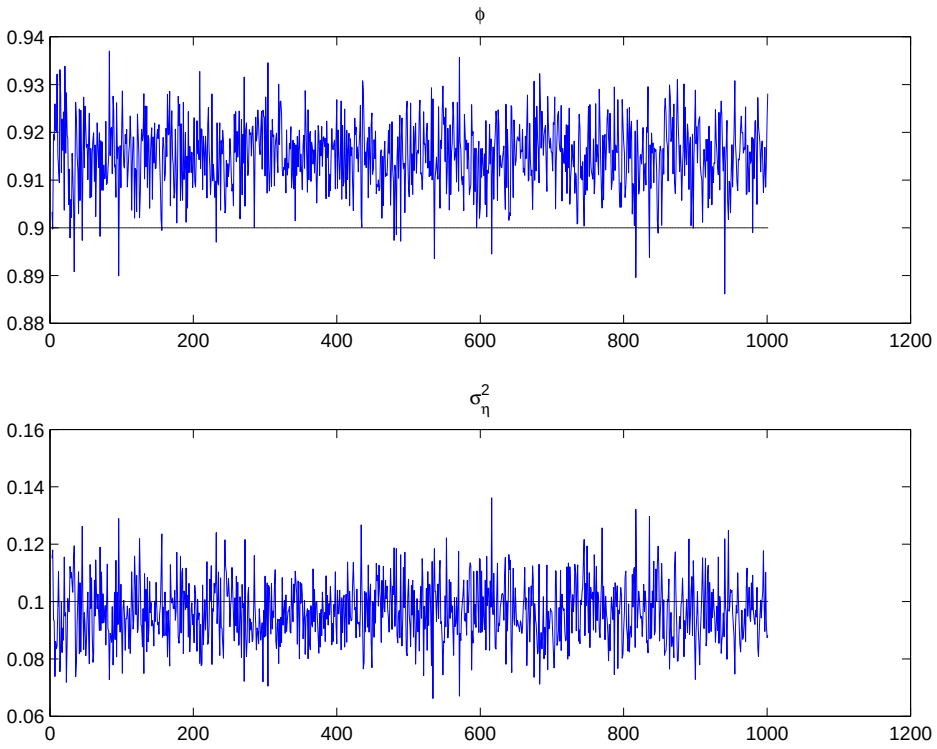

Figure 2: QML estimates of ϕ and σ_η^2 for a simulated series of size 1,000.

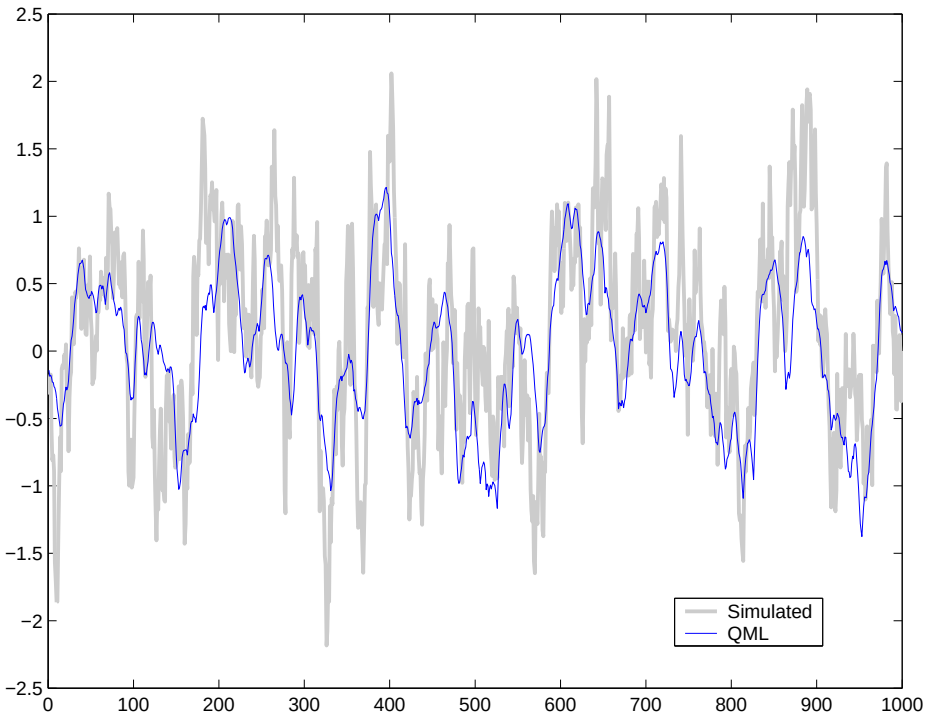

Figure 3: Simulated underlying volatility process (thick grey line) and estimated smoothed volatilities via the QML method (thin black line).

possible solutions to the problem. Kim et al. (1998) suggested adding a small constant c to the squared observations and work with $\log\left(y_t^2 + c\right)$. Fuller (1996), on the other hand, suggested working with the series

$$\omega_t = \log\left(y_t^2 + \delta s^2\right) - \frac{\delta s^2}{y_t^2 + \delta s^2},$$

where s^2 is the sample variance of y_t and δ is a small constant specified by the researcher. Following the latter suggestion, Breidt & Carriquiry (1996) investigated the modified transformation. Their results showed that the improved QML estimator performs better in terms of bias but still has a relatively high root mean square error.

2.3 Bayesian Approaches of estimation

An alternative approach that has become attractive for the estimation of SV models is the Bayesian one. Adopting a Bayesian view to the problem of estimation of the unknown parameters, $\boldsymbol{\theta}$, in a model, effectively means that we initially treat $\boldsymbol{\theta}$ as a random variable with a distribution, known as the prior distribution, over the parameter space, Θ. The prior distribution, $\pi(\boldsymbol{\theta})$, reflects the a priori degree of belief about $\boldsymbol{\theta}$ without having observed any data, $\boldsymbol{Y}$. After we have collected a set of data, $\boldsymbol{Y}$, we update our degree of belief about $\boldsymbol{\theta}$, using Bayes' Theorem, and obtain the posterior distribution of the parameters, denoted $p(\boldsymbol{\theta}|\boldsymbol{Y})$ and given by

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})}{\int_\Theta p(\boldsymbol{Y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}} \propto p(\boldsymbol{Y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}). \tag{2.8}$$

Using the posterior distribution of $\boldsymbol{\theta}$, we can draw inferences regarding the model that is assumed to have generated the data. For more details on Bayesian statistics, we refer the reader to O'Hagan (1994) and Bernardo & Smith (1994). Before we proceed with describing the Bayesian analysis of the stochastic volatility model, we should note that the main drawback of the Bayesian approach is the computational complexity.

The posterior density does not necessarily need to be of standard form. One might often face difficulties in the calculation of the normalizing constant, which is usually a high dimensional integral, or in the computation of posterior marginal densities of a particular element of the parameter vector. A closed form solution to these integrals might not exist and consequently, inference becomes hard. One way to overcome the problem of computational complexity is via sampling methods, such as rejection sampling, Markov Chain Monte Carlo (MCMC) and Monte Carlo integration; see Robert & Casella (1999).

2.3.1 Single-Move MCMC Samplers for the SV Model

The estimation of the SV model via MCMC methods was initially proposed by Shephard (1993) and Jacquier et al. (1994) and has since been extensively used. The main idea is to treat the SV model as a hierarchical structure of conditional distributions. Letting $\boldsymbol{\theta} = \left(\gamma, \phi, \sigma_\eta^2\right)'$ denote the vector of hyperparameters, $\boldsymbol{h} = (h_1, \ldots, h_T)'$ the vector of log-volatilities and $\boldsymbol{y} = (y_1, \ldots, y_T)'$ the vector of observations, the hierarchy is specified by the sequence of three conditional distributions. The first

is the distribution of the observations conditional on the log-volatilities, $p(\boldsymbol{y}|\boldsymbol{h})$, the second is the distribution of the log-volatilities conditional on the hyperparameters, $p(\boldsymbol{h}|\boldsymbol{\theta})$, and the third is the prior distribution of the hyperparameters, $p(\boldsymbol{\theta})$. In this sense, using Bayes' theorem, the joint posterior distribution of $\boldsymbol{h}$ and $\boldsymbol{\theta}$ is $p(\boldsymbol{h},\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{h})\, p(\boldsymbol{h}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$.

The intuition behind this idea is that, in a Bayesian framework, the posterior distribution of $\boldsymbol{\theta}$ is given by the product of the likelihood of the observations $\boldsymbol{y}$ and the prior distribution of $\boldsymbol{\theta}$. However, as already noted, the likelihood function (2.1) for the stochastic volatility model is intractable and consequently, the direct analysis of the posterior density $p(\boldsymbol{\theta}|\boldsymbol{y})$ via MCMC methods is ruled out. To overcome this problem, the vector of hyperparameters $\boldsymbol{\theta}$ is augmented with the vector $\boldsymbol{h}$ of the latent log-volatilities. Therefore, if we focus instead on the posterior density $p(\boldsymbol{\theta},\boldsymbol{h}|\boldsymbol{y})$, we can employ MCMC techniques to obtain samples from this density without having to compute the likelihood function (2.1).

There are several procedures that one could use in order to sample from $p(\boldsymbol{\theta},\boldsymbol{h}|\boldsymbol{y})$. The most popular and probably the most suitable of these is the Gibbs sampler. The posterior density $p(\boldsymbol{\theta},\boldsymbol{h}|\boldsymbol{y})$ is highly multivariate and the Gibbs sampler provides the flexibility to split a high-dimensional problem into lower dimensional ones. The idea therefore is to produce samples alternating between the posterior densities $p(\boldsymbol{h}|\boldsymbol{y},\boldsymbol{\theta})$ and $p(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{h})$ so that these will converge to samples approximately generated from $p(\boldsymbol{\theta},\boldsymbol{h}|\boldsymbol{y})$. The general form of the Gibbs sampler for the stochastic volatility model proceeds as follows.

Algorithm 2.1 (Gibbs sampler for the SV model) *Choose arbitrary starting values* $\boldsymbol{h}^{(0)}$, $\boldsymbol{\theta}^{(0)}$ *and let* $i = 0$.

1. *Sample* $\boldsymbol{h}^{(i+1)} \sim p\left(\boldsymbol{h}|\boldsymbol{y},\boldsymbol{\theta}^{(i)}\right)$.
2. *Sample* $\boldsymbol{\theta}^{(i+1)} \sim p\left(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{h}^{(i+1)}\right)$.
3. *Set* $i = i+1$ *and goto* 1.

Step (2) of algorithm 2.1 is relatively simple to compute. On the other hand, sampling from $p\left(\boldsymbol{h}|\boldsymbol{y},\boldsymbol{\theta}^{(i)}\right)$ is not that straightforward. Single-move algorithms circumvent this difficult part of the procedure by decomposing further the density $p\left(\boldsymbol{h}|\boldsymbol{y},\boldsymbol{\theta}^{(i)}\right)$ into the conditionals $p\left(h_t|\boldsymbol{h}_{\setminus t}^{(i)},\boldsymbol{y},\boldsymbol{\theta}^{(i)}\right)$, where $\boldsymbol{h}_{\setminus t}^{(i)} = \left(h_1^{(i+1)},\ldots,h_{t-1}^{(i+1)},h_{t+1}^{(i)},\ldots,h_T^{(i)}\right)$. Thus, step (1) of algorithm 2.1 becomes: for $t = 1,\ldots,T$, sample $h_t^{(i+1)} \sim p\left(h_t|\boldsymbol{h}_{\setminus t}^{(i)},\boldsymbol{y},\boldsymbol{\theta}^{(i)}\right)$. Several suggestions exist in the literature as to which sampling method should be employed in order to generate $h_t^{(i+1)} \sim p\left(h_t|\boldsymbol{h}_{\setminus t}^{(i)},\boldsymbol{y},\boldsymbol{\theta}^{(i)}\right)$. The common feature of all single-move algorithms is that they exploit the Markovian structure of the log-volatilities process, which allows one to write (for ease of notation we drop the superscripts)

$$p\left(h_t|\boldsymbol{h}_{\setminus t},\boldsymbol{y},\boldsymbol{\theta}\right) = p\left(h_t|h_{t-1},h_{t+1},y_t,\boldsymbol{\theta}\right) \tag{2.9a}$$

$$\propto p\left(y_t|h_t\right)\, p\left(h_{t+1}|h_t,\boldsymbol{\theta}\right)\, p\left(h_t|h_{t-1},\boldsymbol{\theta}\right). \tag{2.9b}$$

Suggestions as to how sampling from (2.9) should proceed include those by Shephard (1993), Jacquier et al. (1994), Shephard & Kim (1994), Geweke (1994), Shephard & Pitt (1997) and Kim et al. (1998). We further detail three of these methods due to their increased efficiency and ability to generalize to

other cases. We only discuss the case of the SV model with uncorrelated Gaussian errors; the reader is referred to Jacquier et al. (2004) for the asymmetric SV model.

Rejection Metropolis-Hastings A first approach to the estimation of the SV model via MCMC is that of Jacquier et al. (1994), who built on the work of Carlin et al. (1992). Jacquier et al. (1994) concentrated on the following parameterization of the SV model:

$$y_t = \sqrt{h_t}\varepsilon_t, \tag{2.10a}$$

$$\log h_t = \gamma + \phi \log h_{t-1} + \eta_t, \quad t = 1, \ldots, T, \tag{2.10b}$$

where ε_t and η_t are contemporaneously and serially independent random variables with distributions $N(0,1)$ and $N\left(0, \sigma_\eta^2\right)$, respectively. In addition, it is assumed that $|\phi| < 1$, so that the model is stationary. In (2.10b) the logarithm of the latent volatilities follows a stationary, Gaussian AR(1) process, so that $\log h_t | h_{t-1}, \boldsymbol{\theta} \sim N\left(\gamma + \phi \log h_{t-1}, \sigma_\eta^2\right)$, and hence

$$p\left(h_t | h_{t-1}, \boldsymbol{\theta}\right) \propto \frac{1}{h_t} \exp\left\{-\frac{\left(\log h_t - \gamma - \phi \log h_{t-1}\right)^2}{2\sigma_\eta^2}\right\}.$$

In addition, noting that from (2.10a) $y_t | h_t \sim N(0, h_t)$, it is easy to deduce that

$$\begin{aligned} & p\left(h_t | \boldsymbol{h}_{\setminus t}, \boldsymbol{y}, \boldsymbol{\theta}\right) \\ \propto \; & \frac{1}{h_t^{1/2}} \exp\left\{-\frac{y_t^2}{2h_t}\right\} \times \frac{1}{h_t} \exp\left\{-\frac{\left(\log h_{t+1} - \gamma - \phi \log h_t\right)^2 + \left(\log h_t - \gamma - \phi \log h_{t-1}\right)^2}{2\sigma_\eta^2}\right\}. \end{aligned} \tag{2.11}$$

After some algebra, (2.11) can be rewritten as

$$p\left(h_t | \boldsymbol{h}_{\setminus t}, \boldsymbol{y}, \boldsymbol{\theta}\right) \propto \frac{1}{h_t^{1/2}} \exp\left\{-\frac{y_t^2}{2h_t}\right\} \times \frac{1}{h_t} \exp\left\{-\frac{\left(\log h_t - m_t\right)^2}{2\sigma_*^2}\right\} = f\left(h_t\right), \tag{2.12}$$

$$\text{where} \qquad m_t = \frac{\gamma\left(1-\phi\right) + \phi\left(\log h_{t+1} + \log h_{t-1}\right)}{\left(1+\phi^2\right)} \quad \text{and} \quad \sigma_*^2 = \frac{\sigma_\eta^2}{1+\phi^2}.$$

Obviously, the posterior distribution of h_t in (2.12) is not a standard, well-known density and hence we cannot sample from it directly. One solution to this problem is to use rejection sampling. However, this method requires a density $g(\cdot)$ and a constant c such that $f(x) \leq c g(x)$ for all x. Jacquier et al. (1994) argued that it was difficult to find a density g that satisfies this last condition for all h_t and also does not reject a large number of draws. In addition, they maintained that, even if it were possible to find such a density g, the algorithm would be computationally slow, because the constant c would depend on the conditioning arguments, which vary as the sampler progresses and hence it would be required to solve an optimization problem for each time t in the sample.

The solution proposed by the authors was to place rejection sampling within an independence M-H algorithm; see Tierney (1994) or Chib & Greenberg (1995). The idea is to find a density g and a constant c such that $p\left(x | \boldsymbol{h}_{\setminus t}, \boldsymbol{y}, \boldsymbol{\theta}\right) \leq c g(x)$, but not necessarily for all x, so that g is a pseudo-dominating density.

For each time t, proposals x_t are generated from the density g, until one of this proposals is accepted with probability $\min\left\{1, p\left(x_t|\boldsymbol{h}_{\backslash t}^{(i)}, \boldsymbol{y}, \boldsymbol{\theta}\right)/cg\left(x_t\right)\right\}$. The accepted x_t then enters an independence M-H accept-reject step and we set $h_t^{(i+1)} = x_t$ with probability

$$\min\left[1, \frac{p\left(x_t|\boldsymbol{h}_{\backslash t}^{(i)}, \boldsymbol{y}, \boldsymbol{\theta}\right)\min\left\{p\left(h_t^{(i)}|\boldsymbol{h}_{\backslash t}^{(i)}, \boldsymbol{y}, \boldsymbol{\theta}\right), cg\left(h_t^{(i)}\right)\right\}}{p\left(h_t^{(i)}|\boldsymbol{h}_{\backslash t}^{(i)}, \boldsymbol{y}, \boldsymbol{\theta}\right)\min\left\{p\left(x_t|\boldsymbol{h}_{\backslash t}^{(i)}, \boldsymbol{y}, \boldsymbol{\theta}\right), cg\left(x_t\right)\right\}}\right].$$

If x_t is not accepted, then $h_t^{(i+1)} = h_t^{(i)}$.

Thus, the problem of sampling from (2.12) is reduced to finding the pseudo-dominating density g and choosing accordingly the constant c. Jacquier et al. (1994) observed that the posterior density of h_t in (2.12) can be seen as the product of two densities, an improper inverse-gamma density, $IG\left(-0.5, 0.5y_t^2\right)$, and a log-normal one. Their suggestion was to approximate the log-normal part of (2.12) by an inverse-gamma density, $IG\left(\alpha, \beta_t\right)$, with the same mean and variance. This amounts to setting

$$\alpha = \frac{1 - 2\exp\left(\sigma_*^2\right)}{1 - \exp\left(\sigma_*^2\right)} \quad \text{and} \quad \beta_t = (\alpha - 1)\exp\left(m_t + 0.5\sigma_*^2\right).$$

Moreover, the product of the two inverse-gamma densities gives an inverse-gamma pseudo-dominating density g with

$$g\left(x_t\right) \propto x_t^{-(\alpha+0.5+1)}\exp\left\{-\frac{\beta_t + 0.5y_t^2}{x_t}\right\}.$$

Finally, Jacquier et al. (1994) proposed an 'ad hoc' rule for selecting the constant c. It should be noted that the choice of c has a dual effect on the algorithm, in the sense that it governs how often proposals x_t are accepted in the rejection sampling step and also how often repeats are produced in the M-H step, i.e. how often we set $h_t^{(i+1)} = h_t^{(i)}$. Jacquier et al. (1994) suggested that a more informative chain of log-volatilities is produced if repeats are avoided, which implies that more effort will be required for a draw from g to be retained. According to the authors' practical experience, a good choice is $c = 1.1 \times p\left(x_t^*|\boldsymbol{h}_{\backslash t}^{(i)}, \boldsymbol{y}, \boldsymbol{\theta}\right)/g\left(x_t^*\right)$, where $x_t^* = \lambda_t/\left(\nu + 1\right)$ is the point at which the inverse-gamma density g attains its mode. Jacquier et al. (1994) claimed that this c results in an acceptance rate between 70% and 80% in the M-H step. For details on how the procedure can be extended to cover the cases of correlated and t-distributed errors, the reader is referred to Jacquier et al. (2004).

A similar Bayesian approach for the estimation of the SV model is due to Shephard & Pitt (1997). They maintained that the method proposed by Jacquier et al. (1994) was inefficient for the following reason. The posterior density $p\left(h_t|\boldsymbol{h}_{\backslash t}, \boldsymbol{y}, \boldsymbol{\theta}\right)$ can clearly be written as

$$p\left(h_t|\boldsymbol{h}_{\backslash t}, \boldsymbol{y}, \boldsymbol{\theta}\right) \propto p\left(y_t|h_t\right)p\left(h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta}\right). \tag{2.13}$$

In essence, the method of Jacquier et al. (1994) attempts to approximate the density $p\left(h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta}\right)$ by the density $p\left(y_t|h_t\right)$. However, Shephard & Pitt (1997) argued that in (2.13) the prior density $p\left(h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta}\right)$ dominates the likelihood $p\left(y_t|h_t\right)$ and therefore one should focus in approximating $p\left(h_t|\boldsymbol{h}_{\backslash t}, \boldsymbol{y}, \boldsymbol{\theta}\right)$ by a density of the same form as $p\left(h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta}\right)$. In order to describe

the method suggested by the authors, we use the following form of the SV model:

$$y_t = \exp\left(\frac{h_t}{2}\right)\varepsilon_t, \tag{2.14a}$$

$$h_t = \mu_h + \phi\left(h_{t-1} - \mu_h\right) + \eta_t, \quad t = 1, \ldots, T, \tag{2.14b}$$

where ε_t and η_t are as in (2.10). The log-volatilities in (2.14b) have been centered around their unconditional mean and clearly $h_t|h_{t-1}, \boldsymbol{\theta} \sim N\left(\mu_h + \phi\left(h_{t-1} - \mu_h\right), \sigma_\eta^2\right)$, while $y_t|h_t \sim N\left(0, \exp\left(h_t\right)\right)$. Moreover, it is easy to see that

$$p\left(h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta}\right) \propto \exp\left\{-\frac{\left[\left(h_{t+1} - \mu_h\right) - \phi\left(h_t - \mu_h\right)\right]^2 + \left[\left(h_t - \mu_h\right) - \phi\left(h_{t-1} - \mu_h\right)\right]^2}{2\sigma_\eta^2}\right\}.$$

By combining the terms in the exponential term and completing the square it is easy to see after some algebra that $h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta} \sim N\left(m_t, \sigma_*^2\right)$, where

$$m_t = \mu_h + \frac{\phi\left[\left(h_{t+1} - \mu_h\right) + \left(h_{t-1} - \mu_h\right)\right]}{1 + \phi^2} \quad \text{and} \quad \sigma_*^2 = \frac{\sigma_\eta^2}{1 + \phi^2}. \tag{2.15}$$

Thus, combining the latter result with (2.13), we have:

$$\begin{aligned}
\log p\left(h_t|\boldsymbol{h}_{\setminus t}, \boldsymbol{y}, \boldsymbol{\theta}\right) &= \text{const} + \log p\left(h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta}\right) + \log p\left(y_t|h_t\right) \\
&= \text{const} - \frac{\left(h_t - m_t\right)^2}{2\sigma_*^2} - \frac{h_t}{2} - \frac{y_t^2}{2}\exp\left(-h_t\right) \\
&\simeq \text{const} - \frac{\left(h_t - m_t\right)^2}{2\sigma_*^2} - \frac{h_t}{2} - \frac{y_t^2}{2}\exp\left(-m_t\right)\left[1 - \left(h_t - m_t\right) + \frac{\left(h_t - m_t\right)^2}{2}\right] \\
&= \log g\left(h_t\right),
\end{aligned} \tag{2.16}$$

where the third line comes from a second order Taylor expansion of $\exp\left(-h_t\right)$ about m_t. In addition, the terms in $\log g$ can be combined and after some algebra, it is easily seen that $g\left(h_t\right) \propto f_N\left(h_t|\mu_t, \sigma_t^2\right)$, where $f_N\left(x|\alpha, \beta^2\right)$ denotes the Gaussian density with mean α and variance β^2, and

$$\mu_t = \frac{\sigma_t^2}{\sigma_*^2}m_t + \frac{\sigma_t^2}{2}\left[y_t^2\exp\left(-m_t\right)\left(1 + m_t\right) - 1\right] \quad \text{and} \quad \sigma_t^2 = \left[\frac{1}{\sigma_*^2} + \frac{y_t^2}{2}\exp\left(-m_t\right)\right]^{-1}.$$

A point to be stressed is that, as Shephard & Pitt (1997) explain, since $\log g$ contains a quadratic term, this implies that $\log g$ is an approximation that does not bound $\log p\left(h_t|\boldsymbol{h}_{\setminus t}, \boldsymbol{y}, \boldsymbol{\theta}\right)$. Therefore, simple rejection sampling cannot be applied. Nonetheless, we can consider the Gaussian density with mean μ_t and variance σ_t^2 as a pseudo-dominating density of $p\left(h_t|\boldsymbol{h}_{\setminus t}, \boldsymbol{y}, \boldsymbol{\theta}\right)$ and then use the rejection M-H algorithm as in Jacquier et al. (1994).

Both algorithms of Jacquier et al. (1994) and Shephard & Pitt (1997) are important in the sense that their procedures are general and can be extended to handle fat-tailed observation errors and the asymmetric stochastic volatility model. The idea of using a M-H algorithm within the Gibbs sampler can also be extended to multivariate stochastic volatility models. However, the simplest procedure for

updating the log-volatilities in a single-move Gibbs sampler was proposed by Kim et al. (1998). Their algorithm, though not very general, is simpler and less computationally consuming because it does not involve a M-H accept-reject step. Kim et al. (1998) found a density g that bounds $p\left(h_t|\boldsymbol{h}_{\setminus t}, \boldsymbol{y}, \boldsymbol{\theta}\right)$ so that rejection sampling can be used to update the log-volatilities.

Rejection Sampling Kim et al. (1998) used the same parameterization of the SV model as Shephard & Pitt (1997). Their reasoning is exactly the same, but they do not use a Taylor expansion of $\exp\left(-h_t\right)$. Instead, Kim et al. (1998) recognised that the function $\exp\left(-h_t\right)$ is convex and hence, it can be bounded below by its tangent at any point. Thus, if we use the latter result in (2.16), we obtain

$$\begin{aligned}\log p\left(h_t|\boldsymbol{h}_{\setminus t}, \boldsymbol{y}, \boldsymbol{\theta}\right) &\leq \text{const} - \frac{\left(h_t - m_t\right)^2}{2\sigma_*^2} - \frac{h_t}{2} - \frac{y_t^2}{2}\exp\left(-m_t\right)\left(1 + m_t - h_t\right) \\ &= \log g\left(h_t\right).\end{aligned}$$

If the terms in $\log g$ are combined, it can be shown that $g\left(h_t\right) \propto f_N\left(h_t|\mu_t, \sigma_*^2\right)$, where σ_*^2 is as in (2.15) and

$$\mu_t = m_t + \frac{\sigma_*^2}{2}\left[y_t^2\exp\left(-m_t\right) - 1\right].$$

Most importantly, the density $f_N\left(\cdot|\mu_t, \sigma_*^2\right)$ does indeed bound $p\left(\cdot|\boldsymbol{h}_{\setminus t}, \boldsymbol{y}, \boldsymbol{\theta}\right)$ and so simple rejection sampling can be used. Thus, to update the log-volatilities, for each time t, proposals x_t are generated from a $N\left(\mu_t, \sigma_*^2\right)$ density until one of these is accepted with probability $\min\left\{1, p\left(x_t|\boldsymbol{h}_{\setminus t}^{(i)}, \boldsymbol{y}, \boldsymbol{\theta}\right)/f_N\left(x_t|\mu_t, \sigma_*^2\right)\right\}$, we then set $h_t^{(i+1)} = x_t$ and proceed to update $h_{t+1}^{(i)}$.

The algorithm proposed by Kim et al. (1998) is efficient, because it does not use a M-H accept-reject step, as the previous methods did, and hence it does not allow for repeats in the sequence of log-volatilities. Thus, a more informative chain is produced, which converges more rapidly. In addition, Kim et al. (1998) state that draws from the $N\left(\mu_t, \sigma_*^2\right)$ are accepted at an average rate of over 99%.

Sampling the hyperparameters We now focus attention on the updating of the parameters, μ_h, ϕ, σ_η^2. Jacquier et al. (1994) used a standard Bayesian approach for linear models, i.e. they proposed using a conjugate normal-inverse gamma prior for the parameters. However, most authors argue that it is better to restrict ϕ to the stationarity region beforehand, by choosing a prior distribution for ϕ that has support on the interval $(-1, 1)$. An uninformative prior for ϕ can cause problems, in practice, when the data are close to being non-stationary. If $\phi = 1$, the parameters μ_h cancel in (2.14b) and as a result, μ_h cannot be identified from the data.

In what follows, we pursue the approach by Kim et al. (1998), who restrict $|\phi| < 1$. Shephard (1996) and Shephard & Pitt (1997) employ a similar approach on this subject. The parameterization we consider is that in (2.14). Step (2) of algorithm 2.1 is split into three parts:

2a Sample $\sigma_\eta^{2\,(i+1)} \sim p\left(\sigma_\eta^2|\boldsymbol{h}^{(i+1)}, \mu_h^{(i)}, \phi^{(i)}\right)$.

2b Sample $\phi^{(i+1)} \sim p\left(\phi|\boldsymbol{h}^{(i+1)}, \mu_h^{(i)}, \sigma_\eta^{2\,(i+1)}\right)$.

2c Sample $\mu_h^{(i+1)} \sim p\left(\mu_h|\boldsymbol{h}^{(i+1)}, \phi^{(i+1)}, \sigma_\eta^{2\,(i+1)}\right)$.

In order to specify the prior and posterior densities of each of the parameters we note the following result. Using one-step-ahead prediction densities for the log-volatilities, the likelihood function of the log-volatilities can evidently be expressed as

$$\begin{aligned} p\left(\boldsymbol{h}|\boldsymbol{\theta}\right) &= p\left(h_1|\boldsymbol{\theta}\right)\prod_{t=1}^{T-1} p\left(h_{t+1}|h_t,\boldsymbol{\theta}\right) \qquad (2.17)\\ &= \left(\frac{1}{2\pi\sigma_\eta^2}\right)^{\frac{T}{2}}\left(1-\phi^2\right)^{\frac{1}{2}}\exp\left\{-\frac{\left(h_1-\mu_h\right)^2\left(1-\phi^2\right)+\sum_{t=1}^{T-1}\left[\left(h_{t+1}-\phi h_t\right)-\left(1-\phi\right)\mu_h\right]^2}{2\sigma_\eta^2}\right\}. \end{aligned}$$

Sampling σ_η^2. For σ_η^2 it is natural to choose a conjugate prior such as the inverse-gamma, so that $\sigma_\eta^2 \sim IG\left(\alpha_\sigma/2, \beta_\sigma/2\right)$. Combining this prior with the likelihood of the log-volatilities (2.17), it can be shown that the posterior density $p\left(\sigma_\eta^2|\boldsymbol{h},\mu_h,\phi\right)$ is an $IG\left(\nu_\sigma,\lambda_\sigma\right)$, with

$$\nu_\sigma = \frac{\alpha_\sigma+T}{2} \quad \text{and} \quad \lambda_\sigma = \frac{\beta_\sigma+\left(h_1-\mu_h\right)^2\left(1-\phi^2\right)+\sum_{t=1}^{T-1}\left[\left(h_{t+1}-\phi h_t\right)-\left(1-\phi\right)\mu_h\right]^2}{2}.$$

As already noted, estimates of σ_η^2 for financial time series are usually found in practice to be very small. For this reason, a popular choice in the stochastic volatility literature is to set $\alpha_\sigma = 5$ and $\beta_\sigma = 0.05$, so that the mean of this prior is equal to 0.0167 approximately.

Sampling ϕ. To restrict ϕ in the stationary region, a transformed Beta prior is used, by setting $\left(\phi+1\right)/2 \sim Beta\left(\alpha_\phi,\beta_\phi\right)$ so that the prior has support on $(-1,1)$ and the prior density is

$$p\left(\phi\right) \propto \left(\frac{\phi+1}{2}\right)^{\alpha_\phi-1}\left(\frac{1-\phi}{2}\right)^{\beta_\phi-1}.$$

If this prior is multiplied by the likelihood of the log-volatilities the full conditional density for ϕ becomes

$$\begin{aligned} p\left(\phi|\boldsymbol{h},\mu_h,\sigma_\eta^2\right) &\propto p\left(\phi\right)\left(1-\phi^2\right)^{1/2}\exp\left\{-\frac{\left(h_1-\mu_h\right)^2\left(1-\phi^2\right)}{2\sigma_\eta^2}\right\} \qquad (2.18a)\\ &\quad\times\exp\left\{-\frac{\sum_{t=1}^{T-1}\left[\left(h_{t+1}-\mu_h\right)-\phi\left(h_t-\mu\right)\right]^2}{2\sigma_\eta^2}\right\}. \qquad (2.18b) \end{aligned}$$

One way to sample from this density is to recognise that it is log-concave in ϕ and hence, use the adaptive rejection sampling procedure by Gilks & Wild (1992). An alternative and simpler approach is to use the independence M-H algorithm. Expanding the exponent of the term in (2.18b), it is easy to see that

$$\exp\left\{-\frac{\sum_{t=1}^{T-1}\left[\left(h_{t+1}-\mu_h\right)-\phi\left(h_t-\mu\right)\right]^2}{2\sigma_\eta^2}\right\} \propto f_N\left(\phi|\mu_\phi,\sigma_\phi^2\right),$$

where $f_N\left(\cdot|\mu_\phi,\sigma_\phi^2\right)$ is the density of a Normal random variable with

$$\mu_\phi = \frac{\sum_{t=1}^{T-1}(h_{t+1}-\mu_h)(h_t-\mu_h)}{\sum_{t=1}^{T-1}(h_t-\mu_h)^2} \quad \text{and} \quad \sigma_\phi^2 = \frac{\sigma_\eta^2}{\sum_{t=1}^{T-1}(h_t-\mu_h)^2}.$$

In conclusion, the posterior density of ϕ is:

$$p\left(\phi|\boldsymbol{h},\mu_h,\sigma_\eta^2\right) \propto p(\phi)\left(1-\phi^2\right)^{1/2}\exp\left\{-\frac{(h_1-\mu_h)^2\left(1-\phi^2\right)}{2\sigma_\eta^2}\right\} f_N\left(\phi|\mu_\phi,\sigma_\phi^2\right).$$

Therefore, within a M-H setting, we choose an initial value $\left|\phi^{(0)}\right| < 1$. We sample draws $\phi_* \sim N\left(\mu_\phi,\sigma_\phi^2\right)$ and then provided $|\phi_*| < 1$, the proposed ϕ_* enters a M-H accept-reject step, where the acceptance probability is given by

$$\alpha\left(\phi^{(i)},\phi_*\right) = \frac{p(\phi_*)\left(1-\phi_*^2\right)^{1/2}\exp\left\{-(h_1-\mu_h)^2\left(1-\phi_*^2\right)/2\sigma_\eta^2\right\}}{p(\phi^{(i)})(1-(\phi^{(i)})^2)^{1/2}\exp\left\{-(h_1-\mu_h)^2(1-(\phi^{(i)})^2)/2\sigma_\eta^2\right\}}.$$

Popular choices for the parameters α_ϕ and β_ϕ of the prior density are $\alpha_\phi = 20$ and $\beta_\phi = 3/2$ so that the prior mean for ϕ is approximately equal to 0.8605. Hence, values of ϕ close to one are favoured more in order to model the persistence in the volatility of asset prices.

Sampling μ_h. Since $\mu_h \in \mathbb{R}$, we can specify a proper but relatively diffuse prior distribution for μ_h, such as a $N\left(0,\sigma_\mu^2\right)$, with σ_μ^2 a large number, e.g. 10. It is easily seen that the posterior density satisfies

$$p\left(\mu_h|\boldsymbol{h},\phi,\sigma_\eta^2\right) \propto \exp\left\{-\frac{(h_1-\mu_h)^2\left(1-\phi^2\right)+\sum_{t=1}^{T-1}\left[(h_{t+1}-\phi h_t)-(1-\phi)\mu_h\right]^2}{2\sigma_\eta^2} - \frac{\mu_h^2}{2\sigma_\mu^2}\right\}. \quad (2.19)$$

Expanding the terms in the exponent in (2.19), after some algebra, it is seen that the posterior is a $N\left(m_h,s_h^2\right)$, where

$$m_h = \frac{(1+\phi)h_1+\sum_{t=1}^{T-1}(h_{t+1}-\phi h_t)}{(1+\phi)+(T-1)(1-\phi)+\dfrac{\sigma_\eta^2}{(1-\phi)\sigma_\mu^2}} \quad \text{and} \quad s_h^2 = \left[\frac{\left(1-\phi^2\right)+(T-1)(1-\phi)^2}{\sigma_\eta^2}+\frac{1}{\sigma_\mu^2}\right]^{-1}.$$

Discussion Let us summarize some of the basic features of the single-move MCMC approach. Firstly, as we have seen, it is fairly straightforward to sample the parameters $\boldsymbol{\theta}$ from their full conditional densities. Once the Gibbs sampler has converged, we can use the draws from this posterior distribution to obtain density estimates and the posterior mean of the parameters. As Broto & Ruiz (2004) point out, an important advantage of the MCMC procedure is that, since the estimation is based on finite sample distributions, there is no need to use asymptotic theory. Moreover, the sampling experiments of Jacquier et al. (1994) compared the method of moments and QML estimation procedures to the MCMC approach and showed that the latter outperforms the other two approaches.

To gain some impression of the precision of the single-move MCMC method, we shall use the same sampling experiment as for the QML method. Using the same sample of size 1,000 generated from the

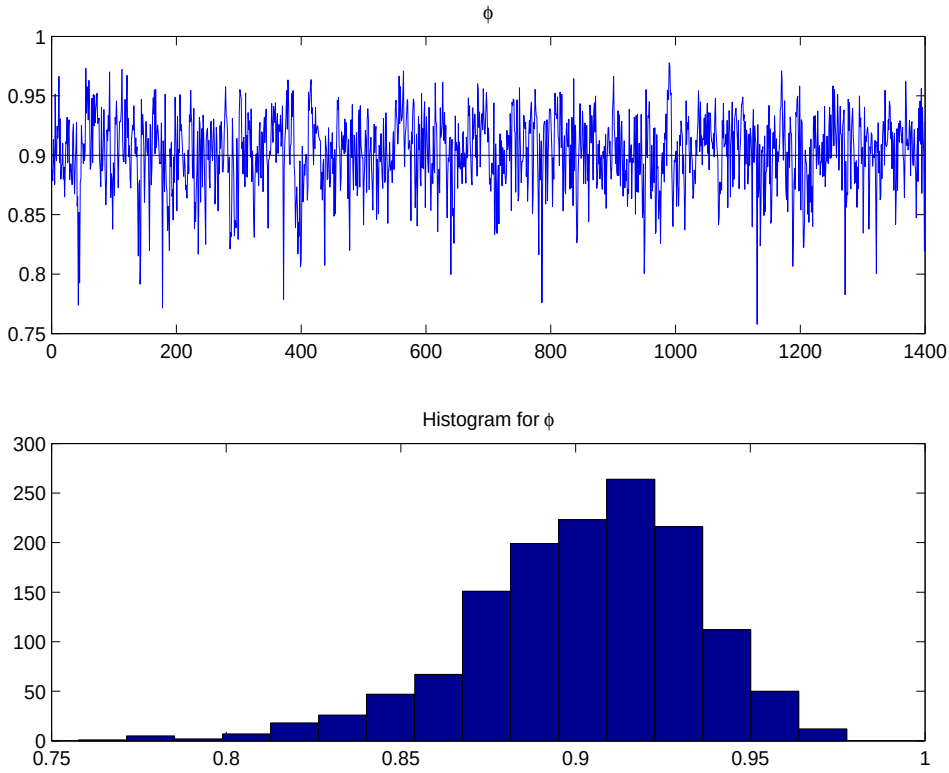

Figure 4: Trace plot for MCMC estimate of ϕ and histogram of the obtained posterior sample.

SV model with $\gamma = 0$, i.e. $\mu_h = 0$, $\phi = 0.9$ and $\sigma_\eta^2 = 0.1$, we use the rejection sampling method of Kim et al. (1998) to sample the log-volatilities while the sampling of the hyperparameters proceeds as above with the prior distribution for μ being $N(0, 10)$.

Posterior summaries of the three parameters ϕ, σ_η^2 and μ_h are presented through trace and histogram plots in Figures 4, 5 and 6 respectively. The histogram plots represent the posterior marginal distributions, while the trace plots indicate whether or not convergence can be safely diagnosed and also how well the output samples behave. All three trace plots in Figures 4, 5 and 6 indicate that the single-move MCMC algorithm has indeed converged. It can be readily seen that the posterior samples for the parameter ϕ do not exhibit any significant bias when compared to the corresponding estimate obtained from the QML method. In fact, the mean of the sample obtained from the Bayesian approach is equal to 0.9026. The output samples for the parameter σ_η^2 behave well and with the sample mean being equal to 0.0940 there is no evidence of considerable bias.

On the other hand, the output samples for the parameter μ_h do exhibit a slight positive bias with the sample mean being equal to 0.1241. The Bayes' estimate for γ is obtained after transformation and turns out to be equal to 0.0121. The small bias found is in accordance with the results of Broto & Ruiz (2004), whose extensive sampling experiments indicated that the best estimate for the parameters μ_h and γ are based on using the corresponding sample moment of the observations.

Finally, it should be noted that for this type of estimation method we hope that there is little autocorrelation within the output series so that the samples obtained can be as close as possible to

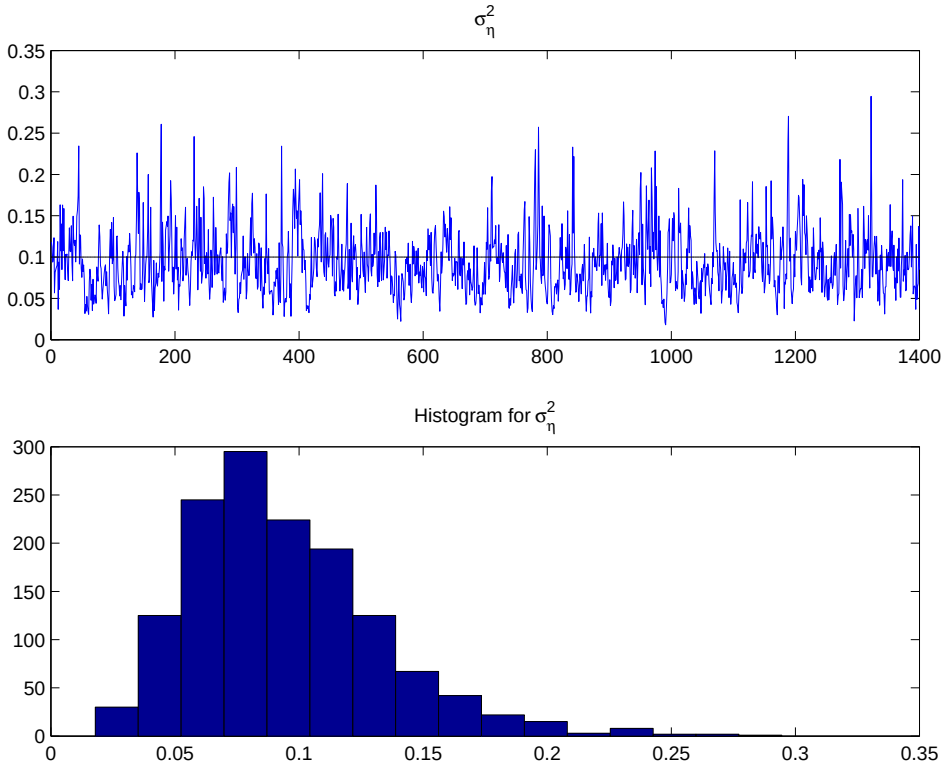

Figure 5: Trace plot for MCMC estimate of σ_η^2 and histogram of the obtained posterior sample.

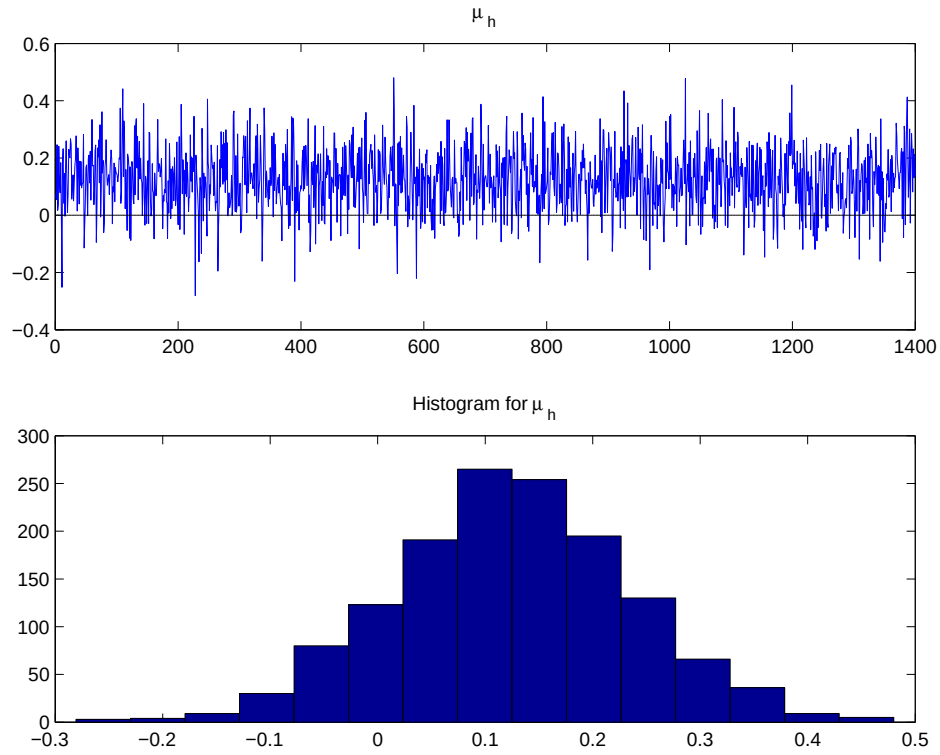

Figure 6: Trace plot for MCMC estimate of μ_h and histogram of the obtained posterior sample.

being serially independent. This can be assessed using correlograms of the samples of ϕ, σ_η^2 and μ, which are displayed in Figure 7. It can be seen that both the samples of ϕ and σ_η^2, even after thinning, do exhibit some autocorrelation, which dies out exponentially fast however. On the other hand, the autocorrelation for the output series of μ_h is insignificant at all lags and thus, the posterior sample obtained for μ_h can be deemed to be serially independent.

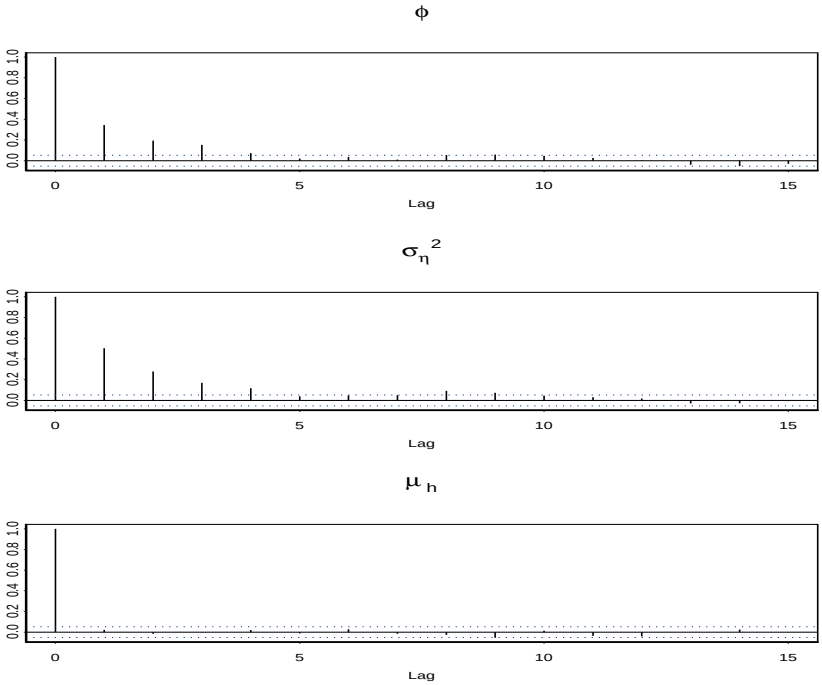

Figure 7: Correlograms for the samples of ϕ, σ_η^2 and μ_h obtained via the single-move MCMC approach of estimation.

An important advantage of the MCMC approach is that it produces samples of the log-volatilities conditional on the observations $\boldsymbol{y}$. The draws can then be used to approximate $E(h_t|\boldsymbol{y})$ by means of a Monte Carlo estimate and hence, obtain a natural solution to the smoothing problem. Moreover, as Jacquier et al. (1994) demonstrate, with a minor modification of the algorithm, it is also possible to obtain multi-step-ahead forecasts of future log-volatilities. The procedure can also be adapted to handle fat-tailed distributions and correlated errors, as in Jacquier et al. (2004), as well as multivariate SV models.

Shephard & Kim (1994) note an issue that might arise when using the single-move MCMC algorithms outlined above. If the persistence in the volatility is high, i.e. $\phi \simeq 1$, and σ_η^2 is small, the log-volatility process is highly correlated. As a result, the draws $h_t^{(i)}$ sampled from the full conditional densities exhibit high levels of correlation for long lags and consequently, there is little movement in the chain. Thus, the single-move MCMC procedure converges slowly in this case and incurs a large computational overhead. In order to overcome this problem, another approach has been proposed, which attempts to sample the whole vector of the log-volatilities with a single draw from the density $p(\boldsymbol{h}|\boldsymbol{y},\boldsymbol{\theta})$. This approach is usually referred to as multimove MCMC sampler and is described in the next section.

2.3.2 Multimove MCMC Samplers

One of the multimove MCMC samplers that has had a great impact on other approaches was suggested by Kim et al. (1998). Their procedure is based on the linear state-space representation (1.4) of the SV model. In particular, the authors suggest to write the SV model as

$$w_t = \log y_t^2 = h_t + z_t, \tag{2.20a}$$
$$h_t = \mu_h + \phi\left(h_{t-1} - \mu_h\right) + \eta_t, \quad t = 1, \ldots, T. \tag{2.20b}$$

However, instead of approximating the $\log \chi^2$ distribution of $z_t = \log \varepsilon_t^2$ with a $N\left(-1.2704, \pi^2/2\right)$, which is suboptimal and results in poor sample properties, a suggestion by Shephard (1994) is followed. Kim et al. (1998) approximated the $\log \chi^2$ distribution by a mixture of seven normals, so that

$$f_{z_t}\left(z_t\right) \simeq \sum_{j=1}^{7} q_j f_{z_t|s_t=j}\left(z_t|s_t = j\right),$$

where $z_t|s_t = j \sim N\left(m_j - 1.2704, v_j^2\right)$, s_t is the indicator variable at time t, and the q_j, $j = 1, \ldots, 7$ are the weights attached to each component, so that $p\left(s_t = j\right) = q_j$ and $\sum_{j=1}^{7} q_j = 1$. Most importantly, given the indicator variable at each time t, the state-space model in (2.20) is Gaussian, that is

$$w_t|s_t = j, h_t \sim N\left(h_t + m_j - 1.2704, v_j^2\right)$$

and thus, the model (2.20) can be thought of as a time-inhomogeneous, Gaussian state-space model. This allows the algorithm of Carter & Kohn (1994), or deJong & Shephard (1995) (see also Appendix A) to be used in order to sample all of $h_1, \ldots, h_T$ at once from $p\left(\boldsymbol{h}|\boldsymbol{w}, \boldsymbol{s}\right)$, where $\boldsymbol{w} = \left(\log y_1^2, \ldots, \log y_T^2\right)$ and $\boldsymbol{s} = \left(s_1, \ldots, s_T\right)$.

In essence, this approach augments further the vector of parameters and log-volatilities, $(\boldsymbol{\theta}, \boldsymbol{h})$, with the vector $\boldsymbol{s}$ of indicator variables and constructs a Gibbs sampler which will produce samples from the density $p\left(\boldsymbol{\theta}, \boldsymbol{h}, \boldsymbol{s}|\boldsymbol{w}\right)$. The general form of this Gibbs sampler is as follows.

Algorithm 2.2 (Multimove Gibbs Sampler) *Choose arbitrary starting values* $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{h}^{(0)}$ *and let* $i = 0$.

1. *Sample* $\boldsymbol{s}^{(i+1)} \sim p\left(\boldsymbol{s}|\boldsymbol{w}, \boldsymbol{h}^{(i+1)}\right)$.
2. *Sample* $\boldsymbol{h}^{(i+1)} \sim p\left(\boldsymbol{h}|\boldsymbol{w}, \boldsymbol{s}^{(i)}, \boldsymbol{\theta}^{(i)}\right)$.
3. *Sample* $\boldsymbol{\theta}^{(i+1)} \sim p\left(\boldsymbol{\theta}|\boldsymbol{h}^{(i+1)}\right)$.
4. *Set* $i = i + 1$ *and goto* 1.

Step (2) of the algorithm is best conducted via the simulation smoother of deJong & Shephard (1995), as shown in Appendix A. To sample the vector $\boldsymbol{s}$ from its full conditional density, we sample

each s_t independently. In particular, we have

$$\begin{aligned} p(\boldsymbol{s}|\boldsymbol{w},\boldsymbol{h}) &= \prod_{t=1}^{T} p(s_t|w_t,h_t) \propto \prod_{t=1}^{T} p(w_t|s_t,h_t)\, p(s_t) \quad \text{and} \\ p(s_t=j|w_t,h_t) &\propto q_j\, f_N\left(w_t|h_t+m_j-1.2704,\upsilon_j^2\right), \quad j=1,\ldots,7. \end{aligned}$$

This a 7-point mass function whose normalizing constant is straightforwardly computed and hence, we can easily produce draws for each s_t. Finally, to update the parameters $\boldsymbol{\theta}$, exactly the same approach as with the single-move Gibbs sampler can be used.

As Kim et al. (1998) illustrate, the above procedure greatly reduces the correlation between successive draws. The method has also been extended by Chib et al. (2002) to handle fat-tailed observation errors, while Omori et al. (2004) generalise to accommodate the leverage effect. However, as is the case with any procedure using the linearized SV model, this method suffers too from the 'inlier' problem. To overcome this problem, Kim et al. (1998) suggest working instead with the transformed observations $w_t^* = \log\left(y_t^2+c\right)$, where $c=10^{-6}$. Furthermore, as Shephard (1996) explains, although the method improves the correlation of the draws and allows for inference, one could argue that this is only an approximation. To correct for the approximation error Kim et al. (1998) propose a procedure that reweighs the draws produced by their sampler in a similar spirit to importance sampling.

2.4 Simulated EM (SIEM) algorithm

The SIEM algorithm for the estimation of SV models is a non-Bayesian approach. It is of special importance because it can be thought of as a refined version of the QML method and in addition, it provides a link between the latter procedure and the MCMC methods discussed above. A first version of the SIEM algorithm for the estimation of the SV model was proposed by Shephard (1993). The approach was later improved through the work of Shephard (1994), Kim et al. (1998) and Mahieu & Schotman (1998). For general details on the EM and SIEM algorithms see Robert & Casella (1999).

The simulated EM algorithm uses the linear state-space form of the SV model (2.20) and a mixture of normals to approximate the density of $z_t = \log\varepsilon_t^2$ as in section 2.3.2. The 'missing' data are assumed to be the indicator variables $\boldsymbol{s}=(s_1,\ldots,s_T)$, while the 'observed' data are the transformed observations $\boldsymbol{w}=\left(\log y_1^2,\ldots,\log y_T^2\right)$. If we let $f(\boldsymbol{w};\boldsymbol{\theta})$ denote the likelihood of the observations, then we may write

$$\begin{aligned} f(\boldsymbol{w};\boldsymbol{\theta}) &= \frac{\dfrac{p(\boldsymbol{w},\boldsymbol{s};\boldsymbol{\theta})}{p(\boldsymbol{s})}p(\boldsymbol{s})}{\dfrac{p(\boldsymbol{s},\boldsymbol{w};\boldsymbol{\theta})}{f(\boldsymbol{w};\boldsymbol{\theta})}} = \frac{p(\boldsymbol{w}|\boldsymbol{s};\boldsymbol{\theta})\,p(\boldsymbol{s})}{p(\boldsymbol{s}|\boldsymbol{w};\boldsymbol{\theta})} \Rightarrow \\ \log f(\boldsymbol{w};\boldsymbol{\theta}) &= \log p(\boldsymbol{w}|\boldsymbol{s};\boldsymbol{\theta}) + \log p(\boldsymbol{s}) - \log p(\boldsymbol{s}|\boldsymbol{w};\boldsymbol{\theta}). \end{aligned} \tag{2.21}$$

If we know apply the 'E-step' of the standard EM algorithm and take expectations on both sides of (2.21) with respect to the density $p(\boldsymbol{s}|\boldsymbol{w};\boldsymbol{\theta})$ conditional on the observed data and a preliminary estimate

$\boldsymbol{\theta}^{(0)}$ of $\boldsymbol{\theta}$, we can see that maximizing $f(\boldsymbol{w};\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ is equivalent to maximizing the function

$$\sum_{\boldsymbol{s}} \log p(\boldsymbol{w}|\boldsymbol{s};\boldsymbol{\theta})\, p\left(\boldsymbol{s}|\boldsymbol{w};\boldsymbol{\theta}^{(0)}\right). \tag{2.22}$$

The maximization step of the EM algorithm then produces an improved estimate $\boldsymbol{\theta}^{(1)}$, according to

$$\boldsymbol{\theta}^{(1)} = \arg\max_{\boldsymbol{\theta}} \sum_{\boldsymbol{s}} \log p(\boldsymbol{w}|\boldsymbol{s};\boldsymbol{\theta})\, p\left(\boldsymbol{s}|\boldsymbol{w};\boldsymbol{\theta}^{(0)}\right). \tag{2.23}$$

In general for $i = 1, 2, \ldots,$ once $\boldsymbol{\theta}^{(i)}$ has been found, we iterate through the expectation and the maximization step again with $\boldsymbol{\theta}^{(i-1)}$ replaced by $\boldsymbol{\theta}^{(i)}$, to obtain a better estimate $\boldsymbol{\theta}^{(i+1)}$. Under some mild regularity conditions the sequence $\left\{\boldsymbol{\theta}^{(i)}\right\}$ converges to a stationary point $\widehat{\boldsymbol{\theta}}$ of $f(\boldsymbol{w};\boldsymbol{\theta})$.

However, the problem with the EM procedure in the case of the SV model is that the density $p(\boldsymbol{s}|\boldsymbol{w};\boldsymbol{\theta})$ is unknown. This is circumverted by cycling through steps 1 and 2 of the multimove Gibbs algorithm 2.2 and obtaining M draws $\boldsymbol{s}^{(1)}, \ldots, \boldsymbol{s}^{(M)}$ from $p(\boldsymbol{s}|\boldsymbol{w};\boldsymbol{\theta})$. Thus, the function to be maximized in (2.23) is replaced by its Monte Carlo estimate $(1/M)\sum_{j=1}^{M} \log p\left(\boldsymbol{w}|\boldsymbol{s}^{(j)};\boldsymbol{\theta}\right)$. Most importantly, given a draw $\boldsymbol{s}^{(j)}$ of the indicator variables vector, the state-space $\boldsymbol{w}|\boldsymbol{s}^{(j)}$ is Gaussian and hence the log-likelihood $\log p\left(\boldsymbol{w}|\boldsymbol{s}^{(j)};\boldsymbol{\theta}\right)$ can be evaluated from the Kalman filter as shown in Appendix A.

In conclusion, given the M draws of the indicator variables vector the function to be maximized is the sum of M Gaussian log-likelihoods. Clearly, as $M \to \infty$, the Monte Carlo estimate converges to (2.22) and hence the above procedure converges to the standard EM algorithm. Shephard (1996) reports that for the SV model in particular, there is little improvement in the procedure if M is increased above the value of 10 and this is because the 'missing' data $\boldsymbol{s}$ contain little information about the parameter vector $\boldsymbol{\theta}$.

In a simulation study Shephard (1993) demonstrates that the SIEM algorithm outperforms the QML method especially when the sample size is small and/or the value of the variance σ_η^2 is small. Furthermore, Mahieu & Schotman (1998) propose a different mixture representation to accommodate for a wide range of shapes of the distribution of ε_t and suggest using as an initial estimate of the parameters the estimates produced by the QML method. Both studies report that the SIEM algorithm finds the neighbourhood of the maximum quite quickly. On the other hand, once the neighbourhood of the maximum is reached, the EM algorithm tends to move slowly on the neighbourhood itself. For this reason, Shephard (1993) suggests switching to some other optimization algorithm when the SIEM slows down. Finally, it should also be noted that, although the use of mixtures makes the procedure more robust than QML to the inlier problem, the problem is not fully resolved.

3 Multivariate Stochastic Volatility Models

In this section we shall be concerned with multivariate models of stochastic volatility, which simultaneously model the movements in the volatility of a number of assets. The extension of the univariate model to the multivariate case is important because of the covariation effect. As Aydemir (1998) explains, we often observe related movements between markets, or sectors, or stocks in a sector, or exchange rates.

This phenomenon, is usually due to a set of assets being formally linked together or due to them being influenced by common unobserved factors.

In particular, currency exchange rates offer a good example of financial time series that are linked together. Diebold & Nerlove (1989) study the behavior of seven dollar exchange rates for a period of 12 years. Their results provide substantial evidence that the seven rates show similarities in the volatility movements in response to a number of actions taken by the U.S. government through the years. For example, in 1978 intervention efforts by the U.S. government to strengthen the dollar cause an increase in the volatility of all seven exchange rates.

On the other hand, another reason that we might observe commonality in the volatility behavior is the flow of 'news' in the market. New information, such as trading volume, quote arrivals, dividend announcements or a change in a government's policy, are likely to affect the volatility of all the assets in the market or at least, a sector of the market. It is usually observed that the flow of new information produces the same qualitative change, though often of different magnitude, in the volatility of the assets. Ghysels et al. (1996) mention also that the globalization of equity markets has caused an increase in the correlation of stock returns.

The aforementioned phenomena suggest that a multivariate volatility analysis can be considered as being more comprehensive than univariate, since it takes account of more information; the cross-correlations in the changes of the asset returns. As multivariate models allow for possible interactions between the volatilities of different financial time series, they can lead to a better fit in data and improve forecast estimates. However, multivariate analysis is also important for most practical applications.

Pitt & Shephard (1999) link multivariate analysis with the theory of financial economics. Arbitrage pricing theory suggests that there is a relationship between the expected return of a stock and the covariance of the returns. The Capital Asset Pricing Model, portfolio optimization and Value-at-Risk (VaR) theory, which studies the extreme downside of a portfolio of assets, all require knowledge of the joint distribution of the asset returns. Hence, there are clearly also practical reasons inducing interest in multivariate analysis.

Multivariate ARCH models started to develop almost at the same period as the univariate ARCH, with the most popular ones being the diagonal vech model by Bollerslev et al. (1988), the constant conditional correlation model by Bollerslev (1990), the factor-ARCH model introduced by Engle et al. (1990), the BEKK model proposed by Engle & Kroner (1995) and the latent factor ARCH model due to Diebold & Nerlove (1989). Most of these models are known to suffer from the curse of dimensionality, while the constraints on the parameters needed to ensure symmetry and positivity are complicated and hard to interpret. The dynamic conditional correlation GARCH model introduced more recently by Engle (2002), the flexible multivariate GARCH model due to Ledoit et al. (2003) and the orthogonal GARCH advocated by Alexander (2001) show more promise, although the latter has been criticized by Engle & Sheppard (2001) for poor fit to the data.

On the other hand, far fewer multivariate models have been proposed in the stochastic volatility literature. The main reason for this is that stochastic volatility models are expressed in terms of normal log-volatilities, which in turn are hard to extend to a multivariate counterpart. The most well known multivariate SV model, due to Harvey et al. (1994) attempts to overcome this difficulty by defining vectors of log-volatilities which interact through a constant correlation structure. Another approach is

to use factor structures, as in Pitt & Shephard (1999) and Aguilar & West (2000), which are more flexible and allow for a large number of assets to be modelled simultaneously. More recently, the multivariate random walk model due to Uhlig (1997) has received significant attention, and many researchers look at modeling the covariance matrix of asset returns using the Wishart distribution. In what follows we describe a number of multivariate stochastic volatility models and their properties.

3.1 Constant Conditional Correlation Model

The first multivariate SV model, which is also the simplest one, was proposed by Harvey et al. (1994). The most important feature of this model, as will become evident, is that it does not allow the covariances between different time series to evolve independently of the variances and thus, it can be considered as the stochastic volatility analogue to the constant conditional correlation model of Bollerslev (1990). In what follows, we denote $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{N\,t})'$ the $(N \times 1)$ vector of one-period returns from N assets at time t. It is assumed that for all $t = 1, \ldots, T$ and for each $i = 1, \ldots, N$ the following relation holds

$$\begin{aligned} y_{it} &= \exp\{h_{it}/2\}\,\varepsilon_{it}, \\ h_{it} &= \gamma_i + \phi_i h_{i\,t-1} + \eta_{it}, \end{aligned}$$

where the observation errors $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{N\,t})'$ and $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{N\,t})'$ are serially and mutually independent, both normal distributed with zero means, $Var(\boldsymbol{\eta}_t) = \boldsymbol{\Sigma}_\eta$ and

$$Var\{\boldsymbol{\varepsilon}_t\} = \boldsymbol{\Sigma}_\varepsilon = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1N} \\ \rho_{12} & 1 & & \rho_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1N} & \rho_{2N} & & 1 \end{bmatrix}, \tag{3.1}$$

with $|\rho_{ij}| < 1$, so that $\boldsymbol{\Sigma}_\varepsilon$ is a correlation matrix. More concisely the above model can be written as: $\boldsymbol{y}_t|\boldsymbol{h}_t \sim N\left(\boldsymbol{0}, \boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}_\varepsilon\boldsymbol{H}_t^{1/2}\right)$, where $\boldsymbol{H}_t$ =diag$(\exp(h_{1t}), \ldots, \exp(h_{N\,t}))$, $\boldsymbol{\Sigma}_\varepsilon$ is given by (3.1) and with the vector of log-volatilities $\boldsymbol{h}_t = (h_{1t}, \ldots, h_{N\,t})'$ following a vector-AR(1) (VAR(1)) process so that

$$\boldsymbol{h}_t = \boldsymbol{\gamma} + \boldsymbol{\Phi}\boldsymbol{h}_{t-1} + \boldsymbol{\eta}_t, \tag{3.2}$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_N)'$, $\boldsymbol{\Phi}$ =diag$(\phi_1, \ldots, \phi_N)$ and $\boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta)$. In Harvey et al. (1994), the matrix $\boldsymbol{\Phi}$ is taken to be the identity matrix, and $\boldsymbol{\gamma}$ is equal to the N-dimensional zero vector, so that the log-volatilities follow a random walk and hence, the fact that volatility movements are persistent is stressed.

This last representation of the model makes also more evident the connection with the model by Bollerslev (1990) and the fact that the covariances between the assets are not allowed to evolve independently of the variances. Clearly, for $i \neq j$,

$$Cov\{y_{it}, y_{jt}|\boldsymbol{h}_t\} = E\{y_{it}^2 y_{jt}^2|\boldsymbol{h}_t\} = \rho_{ij}\exp\left\{\frac{h_{it}}{2}\right\}\exp\left\{\frac{h_{jt}}{2}\right\}$$

and so, since $Var\{y_{it}|\boldsymbol{h}_t\} = \exp\{h_{it}\}$, $i = 1, \ldots, N$, it is obvious that the model has constant correlations. Bollerslev (1990) and Harvey et al. (1994) state that the assumption of constant correlations is empirically reasonable. However, Diebold & Nerlove (1989) argue that the correlations may not be constant and may behave like the variances, exhibiting, for example, temporal persistence. Shephard (1996) concludes that the model above is of limited interest because it cannot capture all of the theoretical features of the data.

Another important feature of this model is that it can be written in a linear, non-Gaussian state space form by taking the logarithms of the squared observations. Thus, it allows for estimation via the QML procedure, as originally suggested by Harvey et al. (1994). Clearly, if we denote $\boldsymbol{w}_t = \left(\log y_{1t}^2, \ldots, \log y_{N\,t}^2\right)'$ and by $\mathbf{1}$ the N-dimensional vector of ones, it is straightforward to see that

$$\boldsymbol{w}_t = -1.2704 \times \mathbf{1} + \boldsymbol{h}_t + \boldsymbol{\xi}_t, \tag{3.3}$$

where $\boldsymbol{\xi}_t = \left(\log \varepsilon_{1t}^2, \ldots, \log \varepsilon_{N\,t}^2\right)' + 1.2704 \times \mathbf{1}$ is a zero-mean, serially independent vector process whose covariance matrix is denoted $\boldsymbol{\Sigma}_\xi$. Clearly, as $\boldsymbol{\xi}_t$ is non-Gaussian, equation (3.3) combined with (3.2) gives a linear but non-Gaussian state space model. Moreover, since $\boldsymbol{\Phi}$ is a diagonal matrix, it is clear that the elements of the state vector, $\boldsymbol{h}_t$, and the observation vector, $\boldsymbol{w}_t$, are linked together only through the off-diagonal elements of the covariance matrices $\boldsymbol{\Sigma}_\eta$ and $\boldsymbol{\Sigma}_\xi$. Therefore, as Harvey et al. (1994) explain, the linearized model resembles a Seemingly Unrelated Time Series (SUTSE) model (see Harvey (1989)), where the N time series, $w_{1t}, \ldots, w_{N\,t}$, are modelled together not because they interact with each other, but because they are all subject to the same economic environment and thus, are contemporaneously correlated.

Furthermore, it can be shown (see Harvey et al. (1994)) that the covariance matrix, $\boldsymbol{\Sigma}_\xi$, of $\boldsymbol{\xi}_t$ can be expressed in a functional form in terms of $\boldsymbol{\Sigma}_\varepsilon$. In particular, let c_{ij} denote the ij-th element of $\boldsymbol{\Sigma}_\xi$. Clearly, since for $i = 1, \ldots, N$, $\xi_{it} = \log \varepsilon_{it}^2 + 1.2704$, the diagonal elements are all equal to $Var\{\xi_{it}\}$, so that $c_{ii} = \pi^2/2$, while the off-diagonal elements are given by

$$c_{ij} = \sum_{n=1}^{\infty} \frac{(n-1)!}{n\,(1/2)_n} \rho_{ij}^{2n}, \quad i \neq j \tag{3.4}$$

where $(\alpha)_n = \alpha\,(\alpha+1)\cdots(\alpha+n-1)$. However, it should be noted that in this case the transformation of the model incurs loss of information. In particular, since in (3.4) the correlation coefficients ρ_{ij} have been squared, all the relevant information on their sign has been lost. Thus, if the transformed observations (3.3) are used to estimate the model, then we can recover the information on the magnitude of the original correlations ρ_{ij}, using (3.4), but not their sign. To overcome, this problem Harvey et al. (1994) suggest using the signs of the untransformed observations to obtain the sign of the correlation coefficients.

Finally, another feature of the model by Harvey et al. (1994) is that it can incorporate common factors and hence allow common trends and cycles in the volatility. In a common factor model (Harvey (1989)) each of the N elements of a multivariate time series are related to $K \leq N$ common factors via a linear function. It can be considered as a generalization of a SUTSE model where the N elements have certain common components. For the case where the log-volatilities follow a random walk as in Harvey

et al. (1994), the model can incorporate common factors, as follows

$$\boldsymbol{w}_t = -1.2704 \times \mathbf{1} + \boldsymbol{B}\boldsymbol{h}_t^* + \bar{\boldsymbol{h}} + \boldsymbol{\xi}_t \tag{3.5a}$$

$$\boldsymbol{h}_t^* = \boldsymbol{h}_{t-1}^* + \boldsymbol{\eta}_t^*. \tag{3.5b}$$

In (3.5) $\boldsymbol{w}_t$ and $\boldsymbol{\xi}_t$ are defined as before, $\boldsymbol{B}$ is an $(N \times K)$ matrix of factor loadings, $\bar{\boldsymbol{h}}$ is an N-dimensional vector, while (3.5b) is a K-dimensional random walk model with $\boldsymbol{\eta}_t^* \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_\eta^*\right)$. It should be noted that unless restrictions are placed on the elements β_{ij} of $\boldsymbol{B}$ and on the vector $\bar{\boldsymbol{h}}$, the model in (3.5) is not identifiable. The usual restrictions are to set $\beta_{ij} = 0$ for $j > i$ and $\boldsymbol{\Sigma}_\eta^*$ equal to the identity matrix, while the first K elements of the vector $\bar{\boldsymbol{h}}$ are restricted to be zero and the rest are unconstrained. An appealing feature implied by these restrictions is that the common factors $\boldsymbol{h}_t^*$ are uncorrelated with each other and this simplifies estimation and the theoretical properties of the model. On the other hand, the restrictions placed on the matrix $\boldsymbol{B}$ are displeasing, in the sense that they imply that the first observation w_{1t} depends only on the first factor, w_{2t} depends only on the first two factors and only the last K observations, $w_{kt}, \ldots, w_{Nt}$ are related to all of the common factors. Clearly depending on the ordering that we choose for the N series in order to model them, we will get different relations between the N observations and the K factors. The unappealing feature about these constraints is that we may often be led to a senseless interpretation of the factors. In the latter case Harvey et al. (1994) suggest, once estimation of the model has been completed, using an orthogonal matrix $\boldsymbol{R}$ to give a factor rotation and a more useful interpretation to the factors.

Finally, it should be noted that the idea of common factors can also be extended to stationary models of the log-volatilities. For a stationary model, we allow $\boldsymbol{h}_t^*$ in (3.5b) to follow a stationary VAR(1) process and this also removes the need to include $\bar{\boldsymbol{h}}$ in the observation equation (3.5a). The non-stationary model with two common factors was estimated using the QML method by Harvey et al. (1994). Another strategy proposed by Jacquier et al. (1995) is to use a Bayesian approach with MCMC methods.

3.2 Factor SV models

Factor models have been introduced in the ARCH literature by Engle et al. (1990) and their basic advantage is that they reduce the dimensionality of the parameter space. Another factor model, which is in similar spirit, but with unobserved latent ARCH factors was proposed by Diebold & Nerlove (1989). The multivariate factor SV models are closer to this latter suggestion, in the sense that the returns $\boldsymbol{y}_t$ from N series are linear functions of K unobserved latent factors, where each one is driven by a SV model. The difference with the common factor model of Harvey et al. (1994) is that in this case the common factors are incorporated for the untransformed observations. The first and simplest factor model in the SV literature was proposed by Jacquier et al. (1999). Denoting by $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{Kt})'$ the vector of K common factors, at time t, Jacquier et al. (1999) set

$$\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{\omega}_t, \tag{3.6}$$

$$f_{it} = \exp(h_{it}/2)\,\varepsilon_{it}, \tag{3.7a}$$

$$h_{it} = \mu_i + \phi_i(h_{it-1} - \mu_i) + \eta_{it}, \quad i = 1, \ldots, K. \tag{3.7b}$$

In this factor model, the matrix $\boldsymbol{B}$ is a constant $(N \times K)$ matrix of factor loadings with $K < N$, the errors $\boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{\Omega})$ and are serially and mutually independent of all the other terms in (3.7a) and (3.7b). The errors ε_{it} and η_{it}, $i = 1, \ldots, K$, are serially and mutually independent $N(0,1)$ and $N\left(0, \sigma^2_{\eta i}\right)$, respectively. We also have $|\phi_i| < 1$, so that the factor log-volatility processes in (3.7b) are stationary.

Clearly, since $\boldsymbol{f}_t$ and $\boldsymbol{\omega}_t$ are independent, zero mean processes, we have $E(\boldsymbol{y}_t) = 0$ and $Var\{\boldsymbol{y}_t|\boldsymbol{h}_t\} = \boldsymbol{B}\boldsymbol{H}_t\boldsymbol{B}' + \boldsymbol{\Omega}$, where $\boldsymbol{H}_t = \text{diag}(\exp(h_{1t}), \ldots, \exp(h_{Kt}))$. Jacquier et al. (1999) proposed to use MCMC methods to estimate the model, as the QML method is not effective in this case. However, the method has not been implemented and another more complicated multivariate factor SV model is usually preferred.

The second factor SV model was independently proposed by Aguilar & West (2000) and Pitt & Shephard (1999) and allows the errors $\boldsymbol{\omega}_t = (\omega_{1t}, \ldots, \omega_{Nt})$ in (3.6) to be driven by independent SV processes also. That is,

$$\omega_{jt} = \exp(\alpha_{jt}/2)\,e_{jt}, \tag{3.8a}$$

$$\alpha_{jt} = m_j + \delta_j(\alpha_{jt-1} - m_j) + u_{jt}, \quad j = 1, \ldots, N. \tag{3.8b}$$

As before, the e_{jt} are i.i.d $N(0,1)$ random variables with no serial dependence and the u_{jt} are i.i.d $N\left(0, \sigma^2_{uj}\right)$ serially and mutually independent of the e_{jt}. As usual we place the constraint $|\delta_j| < 1$ to ensure the stationarity of the log-volatility processes.

The intuition behind this extension of the model is that the K factors can account for the off-diagonal elements of the covariance matrix of the returns but cannot account for all the marginal persistence in the volatility. Loosely speaking the volatility resulting from the fact that the N series are subject to the same economic environment is explained by the volatility of the common factors $\boldsymbol{f}_t$, while each of the idiosyncratic errors, or series-specific shocks $\boldsymbol{\omega}_t$ account for the volatility associated with a particular series only. As Pitt & Shephard (1999) explain, the inclusion of SV effects in the errors $\boldsymbol{\omega}_t$ makes the model more robust, in the sense that if an unusual return is observed for a specific series, then this can be attributed to the idiosyncratic error associated with that particular series.

Before we proceed with describing the properties of this model, we should firstly discuss some identifiability issues. As with the common factors model of Harvey et al. (1994), the parameters of the model are not identifiable unless constraints are placed on the elements β_{ij} of the matrix $\boldsymbol{B}$. There are two ways that this can be done. The solution suggested by Aguilar & West (2000) is to set $\beta_{ii} = 1$ for $i = 1, \ldots, K$ and $\beta_{ij} = 0$ for $j > i$. These are the same restrictions placed by Harvey et al. (1994). The approach suggested by Shephard & Pitt (1997) is somewhat different. They also set $\beta_{ii} = 1$ for $i = 1, \ldots, K$, while $\beta_{i(i+1)} = 0$ for $i = 1, \ldots, K-1$. In this case the first series of returns depends in all but the second common factor, the second series is influenced by all the factors apart from the third

and so on, while only the last $N-K+1$ series depend on all the factors. Both set of constraints are somewhat arbitrary and once the model has been estimated, a factor rotation can be used to give a more meaningful interpretation to the factors.

To discuss the properties of the factor SV model given by (3.7) and (3.8), we adopt the following notation. We denote by $\boldsymbol{h}_t$ the vector $(h_{1t},\ldots,h_{Kt})'$ of log-volatilities of the factors, while $\boldsymbol{\alpha}_t = (\alpha_{1t},\ldots,\alpha_{Nt})'$. Clearly, since the errors ε_{it} $(i=1,\ldots,K)$and e_{jt} $(j=1,\ldots,N)$ are zero mean, we have $E(\boldsymbol{f}_t)=E(\boldsymbol{\omega}_t)=\mathbf{0}$ and $E\{\boldsymbol{f}_t|\boldsymbol{h}_t\}=E\{\boldsymbol{\omega}_t|\boldsymbol{\alpha}_t\}=\mathbf{0}$. Since the model for the returns is $\boldsymbol{y}_t=\boldsymbol{B}\boldsymbol{f}_t+\boldsymbol{\omega}_t$, clearly both the unconditional and conditional mean of $\boldsymbol{y}_t$ are equal to the zero vector. Moreover, since each of the elements of $\boldsymbol{f}_t$ and $\boldsymbol{\omega}_t$ follow mutually independent SV processes, it is easy to see that

$$\left(\begin{array}{c}\boldsymbol{f}_t\\ \boldsymbol{\omega}_t\end{array}\right)\Bigg|\,\boldsymbol{h}_t,\boldsymbol{\alpha}_t \sim NID\left(\mathbf{0},\left[\begin{array}{cc}\boldsymbol{H}_t & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Omega}_t\end{array}\right]\right),$$

where $\boldsymbol{H}_t$ =diag$(\exp(h_{1t}),\ldots,\exp(h_{Kt}))$ and $\boldsymbol{\Omega}_t$ =diag$(\exp(\alpha_{1t}),\ldots,\exp(\alpha_{Nt}))$, while NID denotes normally and serially independently distributed. Since $\boldsymbol{f}_t$ is independent of $\boldsymbol{\omega}_t$ and both are normally distributed, it is implied that

$$\boldsymbol{y}_t|\boldsymbol{h}_t,\boldsymbol{\omega}_t \sim NID\left(\mathbf{0},\boldsymbol{B}\boldsymbol{H}_t\boldsymbol{B}'+\boldsymbol{\Omega}_t\right). \tag{3.9}$$

Moreover, expanding the matrix multiplications in (3.9) it is easily seen that

$$\begin{aligned} Var\{y_{it}|\boldsymbol{h}_t,\boldsymbol{\omega}_t\} &= \beta_{i1}^2\exp(h_{1t})+\cdots+\beta_{iK}^2\exp(h_{Kt})+\exp(\alpha_{it}) && \text{(3.10a)}\\ Cov\{y_{it},y_{jt}|\boldsymbol{h}_t,\boldsymbol{\omega}_t\} &= \beta_{i1}\beta_{j1}\exp(h_{1t})+\cdots+\beta_{iK}\beta_{jK}\exp(h_{Kt}). && \text{(3.10b)}\end{aligned}$$

Clearly, from (3.10) it is readily seen that the evolution of the covariances in the factor SV model does depend on the variances in a complicated, but still deterministic way.

As regards the unconditional covariance matrix of $\boldsymbol{y}_t$, this is deduced using the properties of log-normal distributions. From the properties of the AR processes, we know that $h_{it}\sim N\left(\mu_i,\sigma_{\eta i}^2/\left(1-\phi_i^2\right)\right)$ $(i=1,\ldots K)$, while $\alpha_{jt}\sim N\left(m_j,\sigma_{uj}^2/\left(1-\delta_i^2\right)\right)$ $(j=1,\ldots,N)$. Clearly, $E(f_{it})=E(\omega_{jt})=0$ while

$$\sigma_{f_i}^2=Var(f_{it})=\exp\left(\mu_i+\frac{\sigma_{\eta i}^2}{2\left(1-\phi_i^2\right)}\right)\quad,\quad \sigma_{\omega_j}^2=Var\left(m_j+\frac{\sigma_{uj}^2}{2\left(1-\delta_j^2\right)}\right).$$

Therefore, $Var(\boldsymbol{y}_t)=\boldsymbol{B}\boldsymbol{H}\boldsymbol{B}'+\boldsymbol{\Omega}$, where $\boldsymbol{H}$ =diag$\left(\sigma_{f_1}^2,\ldots,\sigma_{f_K}^2\right)$ and $\boldsymbol{\Omega}$ =diag$\left(\sigma_{\omega_1}^2,\ldots,\sigma_{\omega_N}^2\right)$.

At this point, it should also be noted that more complicated dynamics could be introduced in the latent SV processes. For example, Pitt & Shephard (1999) suggest allowing the errors $\boldsymbol{e}_t$ in the idiosyncratic shocks to be contemporaneously correlated, so that $\boldsymbol{e}_t\sim NID(\mathbf{0},\boldsymbol{\Sigma}_e)$. In this case the multivariate SV model for the idiosyncratic errors $\boldsymbol{\omega}_t$ resembles the multivariate model by Harvey et al. (1994). On the other hand, Aguilar & West (2000) consider a factor SV model where the errors in the latent AR processes of the log-volatilities $\boldsymbol{h}_t$ are contemporaneously correlated, so that

$$\boldsymbol{h}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}\left(\boldsymbol{h}_{t-1} - \boldsymbol{\mu}\right) + \boldsymbol{\eta}_t,$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$, $\boldsymbol{\Phi}$ =diag$(\phi_1, \ldots, \phi_K)$ and $\boldsymbol{\eta}_t \sim NID\left(\mathbf{0}, \boldsymbol{\Sigma}_\eta\right)$.

Other work on this model has been conducted by Han (2002), who exploits the fact that the number of unknown parameters in the model increase linearly with N and fits a multivariate factor SV model on the returns of 36 stocks. Chib et al. (2005) look at extending the model in two ways: by incorporating marginally t-distributed errors $\boldsymbol{\omega}_t$ with stochastic volatility and allowing the series of returns to have jumps. Thus, their model captures the observed fat-tailed nature of financial returns and can be considered as a discretization of the Lévy process, which is widely used in continuous-time finance; see Barndorff-Nielsen & Shephard (2001).

An open issue regarding this model is the number of common factors that are used. The estimation of this model is performed using MCMC methods, but the number of factors is user controlled and is not estimated from the data. On this matter, Aguilar & West (2000) suggest that, according to their experience, if the number of factors K is too large, then the MCMC behaves poorly and does not converge. On the other hand, if the factors are too few, this is identified by detecting co-movements in the estimated variances of the idiosyncratic errors.

3.3 Multivariate SV models based on Wishart processes

One of the first genuinely multivariate SV models was proposed by Uhlig (1997). The model exploits the conjugacy between the Wishart distribution and the matrix variate beta distribution; see Appendix B. It allows the variance of the observations to evolve according to a matrix random walk. Its attractive feature is that it leads to closed form prediction and update equations that can be considered as a generalization of the standard Kalman filter recursions. The results derived by Uhlig (1997) have become popular in the literature of matrix normal dynamic linear models; see West & Harrison (1997, Chapter 16), but has not met great success in a finacial context. The reason for this is that the conjugacy between the Wishart and the matrix variate beta distribution cannot be sustained if we try to move away from the matrix random walk formulation and impose stationarity to the process generating the volatility matrix. Allowing non-stationarity on the volatility matrix is sometimes regarded as an unattractive assumption. However, stationarity constraints are often neglected in econometrics and the model may be valuable in terms of explaining volatility, so is worth pursuing.

In what follows we shall briefly describe the model and we shall refer the interested reader to Uhlig (1997) and West & Harrison (1997) for more details. Theorem B.10 in Appendix B, presents the conjugacy between the Wishart and the beta distributions. In particular, if $U \sim B_n\left(\nu/2, m/2\right)$ with $\nu, m > n-1$ and $V = R'R \sim W_n\left(\nu + m, \Sigma\right)$ for some positive definite symmetric $(n \times n)$ matrix Σ and for an upper-triangular matrix R with positive diagonal elements, then $S = R'UR \sim W_n\left(\nu, \Sigma\right)$. Note also that the matrix beta distribution and its density in Definition B.5 are only defined for $\nu, m > n-1$ and is singular otherwise. Uhlig (1994) extended the definition of the matrix variate beta distribution for positive integers $m > 0$ and showed that the conjugacy result still holds for $\nu > n-1$ and integer-valued $m > 0$. In particular, the density of a singular $B_n\left(\nu/2, 1/2\right)$-distributed random matrix can be found in Uhlig (1994).

As before, we denote by $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})$ the vector of one-period returns from N assets at time t and by $\boldsymbol{\Sigma}_t$ the $(N \times N)$ covariance matrix of $\boldsymbol{y}_t$. In addition $\boldsymbol{T}_t$ denotes the Cholesky factor of $\boldsymbol{\Sigma}_t$, i.e. the upper-triangular matrix with positive diagonal elements such that $\boldsymbol{T}_t' \boldsymbol{T}_t = \boldsymbol{\Sigma}_t$. Then, Uhlig (1997) set

$$\boldsymbol{y}_t = \boldsymbol{T}_t^{-1} \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim NID(\boldsymbol{0}, \boldsymbol{I}_N) \tag{3.11a}$$

$$\boldsymbol{\Sigma}_t = \frac{\nu+1}{\nu} \boldsymbol{T}_{t-1}' \boldsymbol{Q}_t \boldsymbol{T}_{t-1}, \quad \boldsymbol{Q}_t \sim B_N\left(\frac{\nu}{2}, \frac{1}{2}\right), \quad t = 1, \ldots, T. \tag{3.11b}$$

It is assumed that the random variables $\boldsymbol{Q}_t$ in (3.11b) are serially and mutually independent of all other variables in the model as well as the initial condition $\boldsymbol{\Sigma}_0 \sim W_N\left(\nu+1, (\nu+1)^{-1} \boldsymbol{S}_0^{-1}\right)$ for some $(N \times N)$ matrix $\boldsymbol{S}_0 > 0$ (i.e. $\boldsymbol{S}_0$ is positive definite and symmetric).

Using the conjugacy between the Wishart and matrix beta distributions, the model suggests that, at each time t, before observing $\boldsymbol{y}_t$, the prior for $\boldsymbol{\Sigma}_t$ conditional on $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}$ is a $W_N\left(\nu, \nu^{-1} \boldsymbol{S}_{t-1}^{-1}\right)$ distribution with $E(\boldsymbol{\Sigma}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_T) = \boldsymbol{S}_{t-1}^{-1}$. Once $\boldsymbol{y}_t$ becomes available, it can be shown that the posterior for $\boldsymbol{\Sigma}_t$ is a Wishart $W_N\left(\nu+1, (\nu+1)^{-1} \boldsymbol{S}_t^{-1}\right)$, where $\boldsymbol{S}_t = \left(\boldsymbol{y}_t \boldsymbol{y}_t' + \nu \boldsymbol{S}_{t-1}\right) / (\nu+1)$ and with expected value $\boldsymbol{S}_t^{-1}$.

This filtering algorithm delivers the one-step-ahead prediction densities for $\boldsymbol{y}_t$ and hence, the estimation of the model by maximum likelihood is is principle straightforward. In fact, the original formulation of the model by Uhlig (1997) allows $\boldsymbol{y}_t$ to be given by a VAR process whose errors follow the stochastic volatility model in (3.11). Then the estimation of the coefficient matrices in the VAR process requires numerical procedures and Uhlig (1997) proposes an importance-sampling based approach for this.

Recently Philipov & Glickman (2004) introduced another multivariate SV model, which is in similar spirit to the model by Uhlig (1997). It can be considered as a generalized version of the model by Uhlig (1997), which contains the matrix-variate random walk as a special case. Using the same notation as before, the model by Philipov & Glickman (2004) sets for $t = 1, \ldots, T$

$$\boldsymbol{y}_t = \boldsymbol{T}_t' \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim NID(\boldsymbol{0}, \boldsymbol{I}_N)$$

so that $\boldsymbol{y}_t | \boldsymbol{\Sigma}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_t)$ and hence the density of the returns at time t given the covariance structure is

$$p(\boldsymbol{y}_t | \boldsymbol{\Sigma}_t) = \frac{1}{(2\pi)^{\frac{N}{2}} \det(\boldsymbol{\Sigma}_t)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \boldsymbol{y}_t' \boldsymbol{\Sigma}_t^{-1} \boldsymbol{y}_t\right\}. \tag{3.12}$$

The evolution of the covariance matrices $\boldsymbol{\Sigma}_t$ is specified through one-step-ahead prediction densities and more precisely through the density of the inverse of the current covariance structure $\boldsymbol{\Sigma}_t$ conditional on $\boldsymbol{\Sigma}_{t-1}$. In particular, we have

$$\boldsymbol{\Sigma}_t^{-1} | \boldsymbol{\Sigma}_{t-1} \sim W_N(\nu, \boldsymbol{S}_{t-1}),$$

where the degrees of freedom of the Wishart distribution satisfy $\nu \geq N$. To define the scale matrix $\boldsymbol{S}_{t-1}$ of the Wishart distribution, we introduce a positive definite symmetric $(N \times N)$ parameter matrix $\boldsymbol{A}$, with Cholesky factor $\boldsymbol{R}$, i.e. $\boldsymbol{A} = \boldsymbol{R}'\boldsymbol{R}$ and a scalar parameter d. Then, for $t = 0, \ldots, T-1$, $\boldsymbol{S}_t$ is

defined as follows

$$\boldsymbol{S}_t = \frac{1}{\nu}\boldsymbol{R}'\left(\boldsymbol{\Sigma}_t^{-1}\right)^d\boldsymbol{R},$$

where we assume a known initial condition $\boldsymbol{\Sigma}_0 > 0$ for the covariance matrix. Thus, $\boldsymbol{\Sigma}_t^{-1}|\boldsymbol{\Sigma}_{t-1} \sim W_N\left(\nu, \frac{1}{\nu}\boldsymbol{R}'\left(\boldsymbol{\Sigma}_{t-1}^{-1}\right)^d\boldsymbol{R}\right)$ and hence the conditional density of $\boldsymbol{\Sigma}_t^{-1}$ is (see Definition B.2 in Appendix B)

$$p\left(\boldsymbol{\Sigma}_t^{-1}|\boldsymbol{\Sigma}_{t-1}\right) = \frac{1}{2^{\frac{\nu N}{2}}\Gamma_N\left(\frac{\nu}{2}\right)}\frac{\det\left(\boldsymbol{\Sigma}_t\right)^{-\frac{(\nu-N-1)}{2}}}{\left(\nu^{-N}\det\left(\boldsymbol{A}\right)\det\left(\boldsymbol{\Sigma}_{t-1}^{-d}\right)\right)^{\frac{\nu}{2}}}\exp\left\{-\frac{1}{2}tr\left[\nu\left(\boldsymbol{R}\right)^{-1}\left(\boldsymbol{\Sigma}_{t-1}\right)^d\left(\boldsymbol{R}'\right)^{-1}\boldsymbol{\Sigma}_t^{-1}\right]\right\},$$

where $\Gamma_N\left(a\right)$ is the multivariate Gamma function (see Appendix B). The latter distribution defines the evolution of the covariance matrices. Alternatively, using the relation between the Wishart and the inverse Wishart distribution (see Theorem B.7), we can express the time-variation of the covariances in terms of $\boldsymbol{\Sigma}_t$. Since $\boldsymbol{\Sigma}_t^{-1}|\boldsymbol{\Sigma}_{t-1}$ is $W_N\left(\nu, \boldsymbol{S}_{t-1}\right)$-distributed, it follows that

$$\boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1} \sim IW_N\left(\nu+N+1, \boldsymbol{S}_{t-1}^{-1}\right) \equiv IW_N\left(\nu+N+1, \nu\left(\boldsymbol{R}\right)^{-1}\left(\boldsymbol{\Sigma}_{t-1}\right)^d\left(\boldsymbol{R}'\right)^{-1}\right).$$

Moreover, using the properties of the Wishart and inverse Wishart distributions (see Theorems B.5 and B.6), it is easy to see that

$$E\left(\boldsymbol{\Sigma}_t^{-1}|\boldsymbol{\Sigma}_{t-1}\right) = \nu\boldsymbol{S}_{t-1} = \boldsymbol{R}'\left(\boldsymbol{\Sigma}_{t-1}^{-1}\right)^d\boldsymbol{R} \quad \text{and} \tag{3.13a}$$

$$E\left(\boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1}\right) = \frac{1}{\nu-N-1}\boldsymbol{S}_{t-1}^{-1} = \frac{\nu}{\nu-N-1}\left(\boldsymbol{R}\right)^{-1}\left(\boldsymbol{\Sigma}_{t-1}\right)^d\left(\boldsymbol{R}'\right)^{-1}. \tag{3.13b}$$

Before we proceed any further, at this point, we should give an explanation to the role of the parameters $\boldsymbol{A}$ and d. Clearly, both of them have their own role in the evolution of the covariance matrices as they form part of the scale matrix of the Wishart distribution.

The restriction $\boldsymbol{A} > 0$ ensures that the scale matrix $\boldsymbol{S}_t$, $t = 0, \ldots, T-1$, remains positive definite and symmetric and hence the Wishart distribution governing the time-variation of the covariance matrices is properly defined. Most importantly, as Philipov & Glickman (2004) explain, it is a measure of intertemporal sensitivity. From the conditional moment in (3.13b), we can see that the elements of $\boldsymbol{A}^{-1}$ reveal how the current variances and covariances of the returns depend on the previous period's corresponding values. For example, high magnitude values of the diagonal elements of $\boldsymbol{A}^{-1}$ and low magnitude values in the off-diagonal elements imply that the current variances of the assets are primarily affected by the previous period's variance and less by the covariances/correlations with other assets and vice versa. Thus, the off-diagonal elements of $\boldsymbol{A}$ define the degree by which each asset's current volatility is affected by the previous period's correlations with other assets.

On the other hand, the scalar d defines the strength of the relationship between the current values of the volatilities and covariances of the assets and the previous period's. It can be thought of as the persistence in the volatility and the multivariate analogue of the parameter ϕ in the univariate case. In fact, as we shall see, d needs to be restricted in the region $(-1, 1)$ for the process generating the conditional variances to be stationary. In fact, Philipov & Glickman (2004) suggest restricting d between

zero and one so that an inverse relationship between the current covariance matrix and the previous period's one is ruled out.

To see the role played by the scalar d, it is best to use the conditional expectations in (3.13). As in the univariate case, a value of d close to zero implies small persistence in the covariance structure. Consider for example, the extreme case resulting by setting $d = 0$ in (3.13). Then,

$$E\left(\boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1}\right) = \frac{\nu}{\nu - N - 1}\left(\boldsymbol{R}'\boldsymbol{R}\right)^{-1} = \frac{\nu}{\nu - N - 1}\boldsymbol{A}^{-1} \quad \text{and} \quad E\left(\boldsymbol{\Sigma}_t^{-1}|\boldsymbol{\Sigma}_{t-1}\right) = \boldsymbol{A}.$$

This implies that the conditional expectation of the current covariance structure does not depend on the previous period's one and it is constant. Thus, at each time period the covariance matrix of returns in this case can be thought of as taking a value around some constant level irrespective of the value of the previous period's covariance matrix. On the other hand, consider the special case $d = 1$ and $\boldsymbol{A} = \boldsymbol{I}_N$. Substituting this values in (3.13) gives

$$E\left(\boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1}\right) = \frac{\nu}{\nu - N - 1}\boldsymbol{\Sigma}_{t-1} \quad \text{and} \quad E\left(\boldsymbol{\Sigma}_t^{-1}|\boldsymbol{\Sigma}_{t-1}\right) = \boldsymbol{\Sigma}_{t-1}^{-1}.$$

Clearly, this results in the covariance matrix being governed by a matrix variate random walk and can be considered as an analogue to the model by Uhlig (1997). In this latter case, there is high persistence in the volatility and the effect of a shock in the returns at a particular period is present in subsequent periods' returns while it gradually dies out.

To see that d needs tp lien in the interval $(-1, 1)$ for the process generating the conditional covariance matrix to be stationary, we follow Philipov & Glickman (2004) and we use a linear decomposition of the process. Let us first denote by $\boldsymbol{L}_t$ the Cholesky decomposition for $t = 0, \ldots, T-1$ of $\left(\boldsymbol{\Sigma}_t^{-1}\right)^d$, so that $\left(\boldsymbol{\Sigma}_t^{-1}\right)^d = \boldsymbol{L}_t'\boldsymbol{L}_t$. Then, using the result on the transformation of the Wishart distribution (see Theorem B.2), we obtain an autoregressive representation for the process generating the conditional variance.

$$\begin{aligned}
\boldsymbol{\Sigma}_t^{-1}|\boldsymbol{\Sigma}_{t-1} &\sim W_N\left(\nu, \frac{1}{\nu}\boldsymbol{R}'\boldsymbol{L}_{t-1}'\boldsymbol{L}_{t-1}\boldsymbol{R}\right) \Leftrightarrow \boldsymbol{\Sigma}_t^{-1}|\boldsymbol{\Sigma}_{t-1} = \frac{1}{\nu}\boldsymbol{R}'\boldsymbol{L}_{t-1}'W_N\left(\nu, \boldsymbol{I}_N\right)\boldsymbol{L}_{t-1}\boldsymbol{R} \Leftrightarrow \\
\boldsymbol{\Sigma}_t^{-1} &= \frac{1}{\nu}\boldsymbol{R}'\boldsymbol{L}_{t-1}'\boldsymbol{V}\boldsymbol{L}_{t-1}\boldsymbol{R}, \qquad (3.14)
\end{aligned}$$

where now $\boldsymbol{V}$ denotes a random variable with $\boldsymbol{V} \sim W_N\left(\nu, \boldsymbol{I}_N\right)$. If we now take log-determinants on both sides of the conditional distribution of $\boldsymbol{\Sigma}_t^{-1}$, we obtain a AR(1) process for the log of $\det\left(\boldsymbol{\Sigma}_t^{-1}\right)$ conditional on $\boldsymbol{\Sigma}_{t-1}$. In particular, we have

$$\begin{aligned}
\log\left[\det\left(\boldsymbol{\Sigma}_t^{-1}\right)\right] &= \log\left[\det\left(\frac{1}{\nu}\boldsymbol{R}'\boldsymbol{L}_{t-1}'\boldsymbol{V}\boldsymbol{R}\boldsymbol{L}_{t-1}\right)\right] \Rightarrow \\
\log\left[\det\left(\boldsymbol{\Sigma}_t^{-1}\right)\right] &= \log\left[\nu^{-N}\det\left(\boldsymbol{R}'\boldsymbol{R}\right)\det\left(\boldsymbol{L}_{t-1}'\boldsymbol{L}_{t-1}\right)\det\left(\boldsymbol{V}\right)\right] \Rightarrow \\
\log\left[\det\left(\boldsymbol{\Sigma}_t^{-1}\right)\right] &= -N\log\nu + \log\left[\det\left(\boldsymbol{A}\right)\right] + \log\left[\det\left(\boldsymbol{\Sigma}_{t-1}^{-1}\right)^d\right] + \log\left[\det\left(\boldsymbol{V}\right)\right] \Rightarrow \\
\log\left[\det\left(\boldsymbol{\Sigma}_t^{-1}\right)\right] &= -N\log\nu + \log\left[\det\left(\boldsymbol{A}\right)\right] + d\log\left[\det\left(\boldsymbol{\Sigma}_{t-1}^{-1}\right)\right] + \log\left[\det\left(\boldsymbol{V}\right)\right]. \qquad (3.15)
\end{aligned}$$

Equation (3.15) is an AR(1) process for the log-determinant of the covariance matrix of returns. Writing $\Delta_t = \log\left[\det\left(\boldsymbol{\Sigma}_t^{-1}\right)\right]$, this AR(1) process can also be written as

$$\Delta_t = \gamma + d\Delta_{t-1} + \epsilon_t, \tag{3.16}$$

where $\gamma = -N \log \nu + \log [\det (\boldsymbol{A})]$ and ϵ_t is a random variable that follows the distribution of the log-determinant of a $W_N (\nu, \boldsymbol{I}_N)$ random variable. Moreover, from standard properties of autoregressive processes, we know that the process for the log-determinant Δ_t of the covariance matrix $\boldsymbol{\Sigma}_t$ is (weakly) stationary if $|d| < 1$. Thus, we may conclude that in order to have a stationary process generating the covariance matrix we need to set $|d| < 1$.

Clearly, even when the restriction $|d| < 1$ is placed on the model by Philipov & Glickman (2004) the process generating the covariance matrices is not strictly stationary as in the constant correlation model or the factor models. On the other hand, the advantage of this model over the factor SV and the constant correlation models is that it allows the covariance matrix to evolve stochastically over time. Finally, as regards estimation of the model, since there is no obvious transformation of the model that will result in a linear state-space form, Philipov & Glickman (2004) suggest using an MCMC approach which is in similar spirit to the single-move samplers.

3.3.1 The Wishart Autoregressive process

Another model for multivariate SV can be constructed using the Wishart autoregressive process of order one (WAR(1)), introduced by Gourieroux et al. (2004). The continuous time analogue of the model has been applied for the pricing of fixed-income products and credit derivatives (see Gourieroux & Sufana (2003) and Gourieroux & Sufana (2004)), while Gourieroux et al. (2004) concentrate on the case where it is assumed that the covariance matrices are observed. The WAR(1) is also plausible as a driving process for the unobserved volatility matrices of returns from N assets.

The idea underlying the WAR(1) processes relies on the fact (see Theorem B.1 in the Appendix) that if a $(p \times n)$ matrix $\boldsymbol{X}$ is formed by setting its columns to be $n \geq p$ random draws from a multivariate $N(\boldsymbol{0}, \boldsymbol{\Sigma})$ distribution, where the $(p \times p)$ matrix $\boldsymbol{\Sigma} > 0$, then the random variable $\boldsymbol{X}\boldsymbol{X}'$ is $W_p(n, \boldsymbol{\Sigma})$-distributed. To introduce intertemporal dependence between the covariance matrices formed in this way, Gourieroux et al. (2004) proposed defining the columns of the matrix $\boldsymbol{X}$ to be observations from a latent Gaussian VAR(1) process. To describe the multivariate SV model arising by allowing the covariance matrices of N returns to follow a WAR(1), we shall start by describing the process generating the covariance structure.

As before let $\boldsymbol{\Sigma}_t$ be the covariance structure of the returns-vector $\boldsymbol{y}_t$ from N assets. Consider now, $\nu > N - 1$ identical, N-dimensional, Gaussian VAR(1) processes $\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{\nu t}$, where each $\boldsymbol{x}_{it}$ $(i = 1, \ldots, \nu)$ is generated for $t = 1, \ldots, T$ from

$$\boldsymbol{x}_{it} = \boldsymbol{\Phi}\boldsymbol{x}_{it-1} + \boldsymbol{\eta}_{it}, \quad \boldsymbol{\eta}_{it} \sim NID(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta), \tag{3.17}$$

where $\boldsymbol{\Phi}$ is a $(N \times N)$ matrix. It is well known (see Hamilton (1994) and Harvey (1989) for general properties of VAR processes) that the VAR(1) process in (3.17) is stationary if and only if all the eigenvalues λ_i $(i = 1, \ldots, N)$ of the matrix $\boldsymbol{\Phi}$ lie inside the unit circle. In what follows, we assume that this condition holds and that consequently all of the ν VAR(1) processes $\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{\nu t}$ are stationary. Since

the processes are Gaussian weak stationarity implies strong stationarity. In particular, the unconditional distribution of all of the $\boldsymbol{x}_{it}$ $(i = 1, \ldots, \nu)$ is multivariate normal with unconditional mean the zero vector and unconditional variance $Var(\boldsymbol{x}_{it}) = \boldsymbol{S}$, satisfying the equation

$$\boldsymbol{S} = \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{\Phi}' + \boldsymbol{\Sigma}_\eta. \tag{3.18}$$

More precisely, by applying the $vec(\cdot)$ operator on both sides of equation (3.18), it is seen that $vec(\boldsymbol{S}) = (\boldsymbol{I}_{N^2} - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi})\, vec(\boldsymbol{\Sigma}_\eta)$. We can also represent the ν processes together as a matrix normal linear model (see West & Harrison (1997)) by writing

$$\boldsymbol{X}_t = \boldsymbol{\Phi}\boldsymbol{X}_{t-1} + \boldsymbol{H}_t, \tag{3.19}$$

where the $(N \times \nu)$ matrices $\boldsymbol{X}_t$ and $\boldsymbol{H}_t$ are defined as $\boldsymbol{X}_t = [\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{\nu t}]$ and $\boldsymbol{H}_t = [\boldsymbol{\eta}_{1t}, \ldots, \boldsymbol{\eta}_{\nu t}]$. Clearly since the vector $\boldsymbol{\eta}_{it}$ $(i = 1, \ldots, \nu)$ are i.i.d $N(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta)$ we have $\boldsymbol{H}_t \sim N_{N,\nu}(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta \otimes \boldsymbol{I}_\nu)$; see Definition B.1. Most importantly, using the unconditional distributions of the ν VAR(1) processes, we also have that $\boldsymbol{X}_t \sim N_{N,\nu}(\boldsymbol{0}, \boldsymbol{S} \otimes \boldsymbol{I}_\nu)$. Therefore, if we set $\boldsymbol{\Sigma}_t = \boldsymbol{X}_t\boldsymbol{X}_t'$ for all $t = 1, \ldots, T$, then $\boldsymbol{\Sigma}_t > 0$ almost surely and the marginal distribution of the covariance matrix is (see Theorem B.1) $\boldsymbol{\Sigma}_t \sim W_N(\nu, \boldsymbol{S})$, $t = 1, \ldots, T$ with density:

$$p(\boldsymbol{\Sigma}_t) = \frac{1}{2^{\frac{\nu N}{2}}\Gamma_N\left(\frac{\nu}{2}\right)} \frac{\det(\boldsymbol{\Sigma}_t)^{\frac{(\nu-N-1)}{2}}}{\det(\boldsymbol{S})^{\frac{\nu}{2}}} \exp\left\{-\frac{1}{2}tr\left(\boldsymbol{S}^{-1}\boldsymbol{\Sigma}_t\right)\right\}. \tag{3.20}$$

Moreover, using the properties of the Wishart distribution (see Theorem B.5), we have $E(\boldsymbol{\Sigma}_t) = \nu\boldsymbol{S}$.

Let us now consider the distribution of $\boldsymbol{\Sigma}_t$ conditional on $\boldsymbol{\Sigma}_{t-1}$. We start by considering the conditional distribution of the latent VAR(1) processes contained in $\boldsymbol{X}_t$ conditional on $\boldsymbol{X}_{t-1}$. Since $\boldsymbol{H}_t \sim N_{N,\nu}(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta \otimes \boldsymbol{I}_\nu)$ using the dynamic linear model representation, it is straightforward to see that $\boldsymbol{X}_t|\boldsymbol{X}_{t-1} \sim N_{N,\nu}(\boldsymbol{\Phi}\boldsymbol{X}_{t-1}, \boldsymbol{\Sigma}_\eta \otimes \boldsymbol{I}_\nu)$. From Theorem B.8 in Appendix B, we have that the distribution of $\boldsymbol{\Sigma}_t$ conditional on $\boldsymbol{X}_{t-1}$ is a non-central Wishart distribution (see Definition B.4) with ν degrees of freedom, scale matrix $\boldsymbol{\Sigma}_\eta$ and non-centrality parameter $\boldsymbol{\Theta} = \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Phi}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\boldsymbol{\Phi}' = \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Phi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Phi}'$, i.e.

$$\boldsymbol{\Sigma}_t|\boldsymbol{X}_{t-1} \sim W_N\left(\nu, \boldsymbol{\Sigma}_\eta, \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Phi}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\boldsymbol{\Phi}'\right) \equiv W_N\left(\nu, \boldsymbol{\Sigma}_\eta, \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Phi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Phi}'\right). \tag{3.21}$$

At this point we need to stress the fact that there are some important implications resulting from (3.21). The conditional distribution of $\boldsymbol{\Sigma}_t$ depends on $\boldsymbol{X}_{t-1}$ only through the product $\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}' = \boldsymbol{\Sigma}_{t-1}$, which implies that the process is Markov. That is, the current covariance matrix $\boldsymbol{\Sigma}_t$ depends only on the previous period's covariance matrix $\boldsymbol{\Sigma}_{t-1}$ and not on the latent VAR(1) processes contained in $\boldsymbol{X}_{t-1}$. In particular, the conditional density of $\boldsymbol{\Sigma}_t$ is given by

$$\begin{aligned} p(\boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1}) = {} & \frac{1}{2^{\frac{\nu N}{2}}\Gamma_N\left(\frac{\nu}{2}\right)} \frac{\det(\boldsymbol{\Sigma}_t)^{\frac{(\nu-N-1)}{2}}}{\det(\boldsymbol{\Sigma}_\eta)^{\frac{\nu}{2}}} \times \\ & \exp\left\{-\frac{1}{2}tr\left[\boldsymbol{\Sigma}_\eta^{-1}\left(\boldsymbol{\Sigma}_t + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Phi}'\right)\right]\right\} {}_0F_1\left(\frac{\nu}{2}; \frac{1}{4}\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Phi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Phi}'\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_t\right), \end{aligned}$$

where ${}_0F_1$ is the hypergeometric function of matrix argument (see Appendix B for the definition) and can be efficiently computed using the algorithm by Koev & Edelman (2004). Moreover, using the properties of the non-central Wishart distribution (see Theorem B.9), we have

$$E\left(\boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1}\right) = \nu\boldsymbol{\Sigma}_\eta + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Phi}'. \tag{3.22}$$

The fact that the WAR(1) process is Markov is important because it means that we can ignore the underlying VAR(1) processes and we can define the process generating the covariance structures through the unconditional and conditional distributions of $\boldsymbol{\Sigma}_t$, just as in the model by Philipov & Glickman (2004). That is, we may specify

$$\begin{aligned} \boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1} &\sim W_N\left(\nu, \boldsymbol{\Sigma}_\eta, \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Phi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Phi}'\right) \quad \text{with} \\ \boldsymbol{\Sigma}_t &\sim W_N\left(\nu, \boldsymbol{S}\right), \quad \text{where} \quad \boldsymbol{S} = \boldsymbol{\Sigma}_\eta + \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{\Phi}'. \end{aligned}$$

This allows us also to consider non-integer degrees of freedom, since both the Wishart and the non-central Wishart distributions can be defined for any $\nu > N-1$. Of course, as Gourieroux et al. (2004) point out, in this case the WAR(1) process which generates the covariance structures loses its interpretation in terms of the VAR(1) processes. Nonetheless, in the context of a multivariate SV model, the latent VAR(1) processes are meaningful and therefore, there is no reason why we should be interested in them.

Let us now specify how the returns $\boldsymbol{y}_t$ depend on the covariance $\boldsymbol{\Sigma}_t$ generated by the WAR(1) process. There are two ways that we can define this. One way is to set $\boldsymbol{y}_t|\boldsymbol{\Sigma}_t \sim N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_t\right)$ as in the model by Philipov & Glickman (2004), in which case the density of $\boldsymbol{y}_t$ conditional on the covariance structure is given by (3.12). A more interesting case arises if we allow the returns to depend on the inverse $\boldsymbol{\Sigma}_t^{-1}$, that is $\boldsymbol{y}_t|\boldsymbol{\Sigma}_t \sim N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_t^{-1}\right)$. Since the unconditional distribution of $\boldsymbol{\Sigma}_t$ is $W_N\left(\nu, \boldsymbol{S}\right)$ it follows, using the relation between the Wishart and the inverse Wishart distributions (see Theorem B.7) that $\boldsymbol{\Sigma}_t^{-1} \sim IW_N\left(\nu+N+1, \boldsymbol{S}^{-1}\right)$. Therefore, since $\boldsymbol{y}_t|\boldsymbol{\Sigma}_t \sim N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_t^{-1}\right)$ and marginally $\boldsymbol{\Sigma}_t^{-1} \sim IW_N\left(\nu+N+1, \boldsymbol{S}^{-1}\right)$, Theorem B.12 in Appendix B shows that the marginal distribution of $\boldsymbol{y}_t$ is a multivariate t-distribution, denoted $t_N\left(\nu-N+1, 1, \boldsymbol{0}, \boldsymbol{S}^{-1}\right)$ with density

$$p\left(\boldsymbol{y}_t\right) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\pi^{\frac{N}{2}}\Gamma\left(\frac{\nu}{2}\right)} \det\left(\boldsymbol{S}\right)^{\frac{1}{2}} \left(1+\boldsymbol{y}_t'\boldsymbol{S}\boldsymbol{y}_t\right)^{-\frac{1}{2}(\nu+1)}.$$

This latter result is interesting from a financial point of view. It is known that the distribution of the returns $\boldsymbol{y}_t$ has fatter tails than the normal distribution and that the kurtosis arising by accounting for heteroscedasticity only cannot account for all of the excess kurtosis. This is the reason why most researchers consider extending the standard SV model by allowing the observation errors to have fat-tailed distributions. There have recently been attempts to construct stationary time series which account for heteroscedasticity and have specified marginal distributions with heavy tails; see Pitt & Walker (1999) and Pitt et al. (2002). Therefore, it is possible that using the above distribution for the returns vector, we might be able to get a better fit to actual financial returns data.

4 Conclusions and Future Work

In recent years multivariate SV models have attracted considerable attention both by researchers and practitioners. The reason for this is that multivariate analysis seems more appropriate for modeling returns data. In contrast to univariate models, which model the returns of each asset independently of other assets' returns, multivariate SV models take also account of the covariances and correlations between returns. It is therefore believed that a multivariate analysis takes account of additional factors that affect the dynamics of financial returns and in effect, predictions of future asset movements are improved. Moreover, multivariate analysis considers the joint distribution of returns and thus, provides a solution to practical financial problems such as portfolio optimization and VaR issues.

There exists a vast literature on multivariate ARCH models. Nonetheless, most of these models suffer from the lack of parsimony, while the constraints placed on the parameters of these models, in order to ensure that the covariance matrix is positive definite and symmetric are quite complicated. Recent advances by Engle (2002), Ledoit et al. (2003) and Alexander (2001) in the ARCH literature show more promise. On the other hand, SV models seem difficult to extend to a multivariate analogue due to the fact that they are expressed in terms of log-volatilities. The most well known approach to multivariate SV models is the one by Harvey et al. (1994), which restricts the assets returns to interact through a constant correlation structure, a constraint which is not appealing. SV factor models by Pitt & Shephard (1999) and Aguilar & West (2000) allow richer intuition and a large number of assets to be jointly modelled (see Chib et al. (2005)), although open issues still exist regarding the specification of the number of latent factors.

The Wishart distribution, which by definition satisfies the constraints of covariance matrices, looks like a natural extension of the univariate SV case. Most importantly, the use of a Wishart process in order to generate the covariance matrix allows both variances and covariances to evolve stochastically over time. Moreover, a Wishart process allows the current volatility of an asset to depend both on its past realized volatility as well as the correlations with other assets' returns.

The estimation of these models seems challenging and computationally intensive, especially because there is no obvious linearizing transformation and this rules out the use of some methods described in Chapter 2. However, the extension of MCMC methods to a multivariate analogue, where it is vectors and matrices that are sampled instead of scalar parameters, provides an obvious solution to the problem of estimation. Moreover, as the inference via MCMC methods is based on finite sample distributions there is no need to use asymptotic approximations and a natural solution to the smoothing problem is offered.

It would be interesting to compare the multivariate models described in Chapter 3 in terms of fit to the data as well as in terms of their ability to improve prediction on assets' movements. An application of these models to the problem of the optimal asset allocation or that of pricing complex financial products would be of interest not only from a statistical point of view but from a practical aspect as well.

A Appendix: Filtering and smoothing algorithms

The state-space form of time series models has received considerable attention during the last years (see Harvey (1989) and West & Harrison (1997)). State-space representations allow a better study and understanding of the dynamics of a model. Let y_t denote the observation from a (possibly multivariate) time series at time t. In most time series models y_t is usually related to a vector α_t, called the state vector, which is possibly unobserved and whose dimension, m say, is independent of the dimension, n, of y_t. The general form of a linear state-space model, is given by the following two equations

$$y_t = Z_t\alpha_t + d_t + G_t\varepsilon_t, \quad t = 1, \ldots, T \tag{A.1a}$$

$$\alpha_{t+1} = T_t\alpha_t + c_t + H_t\varepsilon_t. \tag{A.1b}$$

Equation (A.1a) is known as the observation or measurement equation, while (A.1b) is called the state transition equation. In (A.1) the parameters Z_t, d_t, G_t, T_t, c_t and H_t are referred to as the system matrices and are assumed to be non-stochastic. In addition, ε_t is a vector of serially independent, identically distributed disturbances with $E(\varepsilon_t) = \mathbf{0}$ and $Var(\varepsilon_t) = \boldsymbol{I}$, the identity matrix. Finally, the formulation of the state-space model (A.1) is completed by the assumption that the initial state vector α_1 is independent of ε_t at all times t and has unconditional mean and variance $a_{1|0}$ and $P_{1|0}$ respectively.

The Kalman filter is used primarily for prediction, filtering and smoothing. In what follows, we assume that the state-space model in (A.1) is Gaussian and that $G_tH_t' = \mathbf{0}$, as is usually the case. Moreover, we drop the terms d_t and c_t from the observation and transition equations for simplicity and write $G_tG_t' = \Sigma_t$ and $H_tH_t' = \Omega_t$. The Kalman filter computes recursively the quantities $a_{t|t} = E(\alpha_t|\Psi_t)$ and $a_{t+1|t} = E(\alpha_{t+1}|\Psi_t)$ along with $P_{t|t} = MSE(\alpha_t|\Psi_{t-1})$ and $P_{t+1|t} = MSE(\alpha_{t+1}|\Psi_t)$, where $\Psi_t = (y_1, \ldots, y_T)$. Then, (Harvey (1989)) starting with $a_{1|0}$ and $P_{1|0}$, $a_{t|t}$, $a_{t+1|t}$ and their MSE's are obtained by running for $t = 1, \ldots, T$, the recursions

$$v_t = y_t - Z_ta_{t|t-1} \quad , \quad F_t = Z_tP_{t|t-1}Z_t' + \Sigma_t \tag{A.2a}$$

$$a_{t|t} = a_{t|t-1} + P_{t|t-1}Z_t'F_t^{-1}v_t, \tag{A.2b}$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}Z_t'F_t^{-1}Z_tP_{t|t-1}, \tag{A.2c}$$

$$a_{t+1|t} = T_ta_{t|t}, \tag{A.2d}$$

$$P_{t+1|t} = T_tP_{t|t}T_t' + \Omega_t. \tag{A.2e}$$

The quantities v_t and F_t in (A.2a) denote respectively the one-step-ahead error in forecasting y_t and its MSE conditional on Ψ_{t-1}. The quantities $a_{t|t}$ and $a_{t|t-1}$ are optimal estimators of α_t conditional on the available information, in the sense that they are minimum mean square estimators. This last property holds only if the state-space model is Gaussian. If the normality assumption is dropped the quantities $a_{t|t}$ and $a_{t|t-1}$ are optimal only within the class of linear estimators. In other words, they are the minimum mean square linear estimators of α_t conditional on Ψ_t and Ψ_{t-1} (see also Brockwell & Davis (1991)).

Equations (A.2b) and (A.2d) can be combined together to yield recursions that only compute the one-step-ahead prediction estimates of α_{t+1} given Ψ_t. Similarly, by taking together (A.2c) and (A.2e), a

single set of recursions for the MSE is obtained, which goes directly from $P_{t|t-1}$ to $P_{t+1|t}$. The resulting recursions, along with (A.2a), for $t = 1, \ldots, T-1$, are as follows:

$$a_{t+1|t} = T_t a_{t|t-1} + K_t v_t \quad , \quad K_t = T_t P_{t|t-1} Z_t' F_t^{-1} \tag{A.3a}$$

$$P_{t+1|t} = T_t P_{t|t-1} L_t' + \Omega_t \quad , \quad L_t = T_t - K_t Z_t. \tag{A.3b}$$

Another application of the Kalman filter is the estimation of any unknown parameters θ that might be contained in the system matrices. Suppose that we have a set of T observations $\boldsymbol{y} = (y_1, \ldots, y_T)$ from the time series of interest. Because the T observations are not i.i.d. the likelihood needs to be constructed by the one-step-ahead prediction densities. In particular, the joint probability density function of the T observations, can be written as

$$p(y_1, \ldots, y_T) = p(y_T|\Psi_{T-1}) \cdots p(y_2|\Psi_1)\, p(y_1) = \prod_{t=1}^{T} p(y_t|\Psi_{t-1}).$$

The one-step-ahead prediction density $p(y_t|\Psi_{t-1})$ is the density of a multivariate normal random variable with mean $Z_t a_{t|t-1}$ and covariance matrix F_t. Thus, the likelihood function is given by

$$\mathcal{L}(\boldsymbol{y};\theta) = \frac{1}{(2\pi)^{nT/2}} \prod_{t=1}^{T} \frac{1}{(\det F_t)^{1/2}} \exp\left\{ -\frac{1}{2} \left(y_t - Z_t a_{t|t-1}\right)' F_t^{-1} \left(y_t - Z_t a_{t|t-1}\right) \right\} \Rightarrow$$

$$l(\boldsymbol{y};\theta) = \log \mathcal{L} = -\frac{nT}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \log(\det F_t) - \frac{1}{2} \sum_{t=1}^{T} v_t' F_t^{-1} v_t. \tag{A.4}$$

It should be noted that, under certain regularity conditions, the maximum likelihood estimators are consistent and asymptotically normal (Harvey (1989)) and hence asymptotic standard errors of the estimates can be reported. In addition, if the state-space model is non-Gaussian, the likelihood can still be constructed in the same way using the minimum mean square linear estimators of the state vector. However, the estimators $\widehat{\theta}$ that maximize (A.4) are the quasi-maximum likelihood estimators of the parameters. It can be shown that the QML estimators are also consistent and asymptotically normal (Hamilton (1994)).

Another algorithm that can be applied to a state-space model is the fixed interval smoothing algorithm. In mathematical terms, the fixed interval smoothing algorithm computes $a_{t|T} = E(\alpha_t|\Psi_T)$ along with its MSE, $P_{t|T}$, for all $t = 1, \ldots, T-1$. These quantities are computed via a set of backward recursions and use the quantities $a_{t+1|t}$, $a_{t|t}$ and the MSE matrices $P_{t|t}$, $P_{t+1|t}$, which are obtained from (A.2). In particular, to obtain $a_{t|T}$ and $P_{t|T}$, we start with $a_{T|T}$ and $P_{T|T}$ and we run backwards for $t = T-1, \ldots, 0$

$$a_{t|T} = a_{t|t} + P_t^* \left(a_{t+1|T} - a_{t+1|t}\right)$$

$$P_{t|T} = P_{t|t} + P_t^* \left(P_{t+1|T} - P_{t+1|t}\right) P_t^* \quad , \quad P_t^* = P_t T_t' P_{t+1|t}.$$

During the last decade, the extensive use of the Gibbs sampler, has given rise to another smoothing algorithm called the simulation smoother and is also tightly related to the Kalman filter. In contrast,

to the fixed interval smoother, which computes the conditional mean and variance of the state vector at each time t in the sample, a simulation smoother is used for drawing samples from the density $p(\alpha_0, \ldots, \alpha_T|Y_T)$. There exist nowadays a number of smoothing algorithms in the literature proposed by Fruhwirth-Schnatter (1994), Carter & Kohn (1994), Carter & Kohn (1996), deJong & Shephard (1995) and Durbin & Koopman (2002).

The most popular simulation smoothing algorithm in the SV literature is probably that of deJong & Shephard (1995). The main idea of the algorithm is to obtain a sample from $p(G_0\varepsilon_0, \ldots, G_T\varepsilon_T|\Psi_T)$, by sampling $\widehat{\eta}_t \sim p(G_t\varepsilon_t|\Psi_T, G_{t+1}\varepsilon_{t+1}, \ldots, G_T\varepsilon_T)$. This allows the state vector to be reconstructed, using the observation equation (A.1a). In particular, the simulation smoother by deJong & Shephard (1995) requires only the prediction equations (A.3) to be initially run and the quantities v_t, F_t, K_t and L_t to be stored. Then, starting with $r_T = 0$ and $U_T = 0$, the following backward recursions are run for $t = T, T-1, \ldots, 1$.

$$\begin{aligned} C_t &= \Sigma_t - \Sigma_t\left(F_t^{-1} + K_t' U_t K_t\right)\Sigma_t \quad , \quad \kappa_t \sim N(0, C_t), \\ r_{t-1} &= Z_t' F_t^{-1} v_t + L_t' r_t - \left(Z_t' F_t^{-1} - L_t' U_t K_t\right)\Sigma_t C_t^{-1}\kappa_t, \\ U_{t-1} &= Z_t' F_t^{-1} Z_t + L_t' U_t L_t + \left(Z_t' F_t^{-1} - L_t' U_t K_t\right)\Sigma_t C_t^{-1}\Sigma_t\left(F_t^{-1} Z_t - K_t' U_t L_t\right). \end{aligned}$$

The quantities $\widehat{\eta}_t = \Sigma_t\left(F_t^{-1} v_t - K_t' r_t\right) + \kappa_t$ are draws from the density $p\left(G_t\varepsilon_t|\Psi_T, \widehat{\eta}_{t+1}, \ldots, \widehat{\eta}_T\right)$ and hence, the vectors $\widehat{\alpha}_t = y_t - c_t - \Sigma_t\left(F_t^{-1} v_t - K_t' r_t\right) - \kappa_t$ represent draws from the filtered distribution of the state vector.

B Appendix: Matrix Variate Distributions

In this section we present some definitions and fundamental results on matrix variate distributions. References for this section are Gupta & Nagar (2000), Muirhead (1982), Press (1972) and Dawid (1981). The following notation is used throughout. We denote by $tr(A)$ the trace of a square matrix A, $\otimes$ the Kronecker product of two matrices and by *vec* the matrix operator which stacks the columns of a matrix one under the other (see Harville (1997) for the properties of the operators). Moreover, a symmetric positive definite symmetric matrix A is denoted $A > 0$.

Definition B.1 *The $(p \times n)$ random matrix $X \in \mathbb{R}^{p\times n}$ is said to have a matrix variate normal distribution with mean matrix $(p \times n)$ $M \in \mathbb{R}^{p\times n}$ and covariance matrix $\Sigma \otimes \Psi$, where Σ is $(p \times p)$ and Ψ is $(n \times n)$ with $\Sigma, \Psi > 0$, denoted $X \sim N_{p,n}(M, \Sigma \otimes \Psi)$, if $vec(X') \sim N_{pn}(vec(M'), \Sigma \otimes \Psi)$. The p.d.f. of X is given by*

$$\frac{1}{\sqrt{(2\pi)^{np}\det(\Sigma)^n \det(\Psi)^p}} \exp\left\{-\frac{1}{2} tr\left[\Sigma^{-1}(X - M)\Psi^{-1}(X - M)'\right]\right\}.$$

A special case of the matrix normal distribution arises when we have a sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ of size n from a p-dimensional multivariate normal distribution $N_p(\boldsymbol{\mu}, \Sigma)$. The matrix $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, formed by placing the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ one next to the other, has a matrix variate normal distribution with mean $\boldsymbol{\mu}\mathbf{1}'$, where $\mathbf{1} = (1, \ldots, 1)'$ is the n-dimensional vector of ones, and variance $\Sigma \otimes I_n$, where I_n is the

n-dimensional identity matrix, i.e. $X \sim N_{p,n}(\boldsymbol{\mu}\mathbf{1}', \Sigma \otimes I_N)$. Next, we define the Wishart distribution.

Definition B.2 *A $(p \times p)$ random, symmetric, positive definite matrix S is said to have a Wishart distribution with $(p \times p)$ scale matrix $\Sigma > 0$ and $n \geq p$ degrees of freedom, denoted $S \sim W_p(n, \Sigma)$ if its p.d.f. is given by*

$$\frac{1}{\sqrt{2^{np}}\Gamma_p\left(\frac{n}{2}\right)} \frac{\det(S)^{\frac{1}{2}(n-p-1)}}{\det(\Sigma)^{\frac{1}{2}n}} \exp\left\{-\frac{1}{2}tr\left(\Sigma^{-1}S\right)\right\}. \tag{B.1}$$

In the above definition, $\Gamma_p(a)$ denotes the multivariate gamma function, which is defined for $\mathrm{Re}(a) > (p-1)/2$ by

$$\Gamma_p(a) = \int_{A>0} \exp\{-tr(A)\} \det(A)^{a-\frac{1}{2}(p+1)} dA$$

where the integral is over the space of $(p \times p)$ symmetric positive definite matrices A. Moreover, it can be shown (see Muirhead (1982) and Gupta & Nagar (2000)) that for $\mathrm{Re}(a) > (p-1)/2$,

$$\Gamma_p(a) = \pi^{\frac{1}{4}p(p-1)} \prod_{i=1}^{p} \Gamma\left(a - \frac{1}{2}(i-1)\right).$$

The density (B.1) of $S \sim W_p(n, \Sigma)$ exists only for $n \geq p$. If $n < p$, the distribution is singular (see Muirhead (1982) p. 82) and $S \sim W_p(n, \Sigma)$ does not have a density function. Finally, it should be noted that the Wishart distribution plays a very important role in multivariate analysis. It is the distribution of the covariance matrix of a sample of size N from a multivariate $N_p(\boldsymbol{\mu}, \Sigma)$ distribution. More precisely, the following result holds.

Theorem B.1 *Let the p-dimensional vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a random sample of size $n \geq p$ from a multivariate $N_p(0, \Sigma)$ distribution, where $\Sigma > 0$ is $(p \times p)$ and define the $(p \times n)$ matrix $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Define the $(p \times p)$ matrix $S = XX'$. Then $S > 0$ with probability one and $S \sim W_p(n, \Sigma)$.*

Proof See Press (1972) p.101, or Gupta & Nagar (2000) p.88. ■

In Theorem B.1, if $n < p$, the density of $V = XX'$ is singular and the distribution of V is often referred to as a pseudo-Wishart. In what follows, we give some properties of the Wishart distribution.

Theorem B.2 (Transformation of Wishart Matrices) *Let the $(p \times p)$ random matrix $S \sim W_p(n, \Sigma)$ and let A be a $(q \times p)$ constant matrix with $q \leq p$. Define the $(q \times q)$ matrix $V = A\Sigma A'$, then $V \sim W_q(n, A\Sigma A')$.*

Proof See Press (1972) p.106. ■

Corollary B.3 *Let $S \sim W_p(n, \Sigma)$ and suppose that $\Sigma^{-1} = AA'$. Then, $ASA' \sim W_p(n, I_p)$.*

Theorem B.4 (Additive Property) *Let the $(p \times p)$ matrices $S_1, \ldots, S_k$ be independent with $S_i \sim W_p(n_i, \Sigma)$, $i = 1, \ldots, k$, for some $(p \times p)$ matrix $\Sigma > 0$. Then $\sum_{i=1}^{k} S_i \sim W_p\left(\sum_{i=1}^{k} n_i, \Sigma\right)$.*

Proof See Gupta & Nagar (2000) p.94. ■

Theorem B.5 (Moments of the Wishart distribution) *Let $S \sim W_p(n, \Sigma)$ with $(p \times p)$ matrix $\Sigma > 0$. An let s_{ij} and σ_{ij} $(1 \leq i, j \leq p)$ be the elements of S and Σ respectively. Then,*

1. $E(S) = n\Sigma$.
2. $Var(s_{ij}) = n\left(\sigma_{ij}^2 + \sigma_{ii}\sigma_{jj}\right)$.
3. $Cov(s_{ij}, s_{k\ell}) = n(\sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk})$.
4. $E\{\det(S)^r\} = 2^{pr}\det(\Sigma)^r \dfrac{\Gamma_p\left(\frac{1}{2}n + r\right)}{\Gamma_p\left(\frac{1}{2}r\right)}$, *for* $\mathrm{Re}(r) > -\frac{1}{2}(n - p + 1)$.

Proof See Press (1972) p.107 and Gupta & Nagar (2000) p.105. ■

Two important distributions that are related to the Wishart distribution are the inverse Wishart distribution and the non-central Wishart distribution. The former is the multivariate generalization of the inverse Gamma distribution and is the distribution of the inverse of a random $(p \times p)$ matrix $S \sim W(n, \Sigma)$. It is usually used in Bayesian analysis as a conjugate prior for the covariance matrix of a normal distribution. The non-central Wishart distribution plays an important role in applications in econometrics and is the multivariate generalization of the non-central chi-squared distribution.

Definition B.3 *A random $(p \times p)$ matrix $V > 0$ is said to have an inverse Wishart distribution with $m > 2p$ degrees of freedom and $(p \times p)$ scale matrix $\Psi > 0$, denoted $V \sim IW_p(m, \Psi)$ if its pdf is given by*

$$\frac{1}{\sqrt{2^{(m-p-1)p}}\Gamma_p\left(\frac{1}{2}(m - p - 1)\right)} \frac{\det(\Psi)^{\frac{1}{2}(m-p-1)}}{\det(V)^{\frac{1}{2}m}} \exp\left\{-\frac{1}{2}tr\left(V^{-1}\Psi\right)\right\}.$$

Theorem B.6 (Moments of the inverse Wishart distribution) *Let $V \sim IW_p(m, \Psi)$ with m and Ψ as in definition B.3 and let v_{ij} and ψ_{ij} $(1 \leq i, j \leq p)$ denote the elements of V and Ψ, respectively. Then,*

1. $E(V) = \dfrac{1}{m - 2p - 2}\Psi$.
2. $Var(v_{ii}) = \dfrac{2\psi_{ii}^2}{(m - 2p - 2)^2(m - 2p - 4)}, \quad m - 2p > 4$.
3. $Var(v_{ij}) = \dfrac{(m - 2p - 2)\psi_{ii}\psi_{jj} + (m - 2p)\psi_{ij}^2}{(m - 2p - 1)(m - 2p - 2)^2(m - 2p - 4)}, \quad m - 2p > 4$ *and* $i \neq j$.
4. $Cov(v_{ij}, v_{k\ell}) = \dfrac{2\psi_{ij}\psi_{k\ell} + (m - 2p - 2)\left(\psi_{ik}\psi_{j\ell} + \psi_{i\ell}\psi_{kj}\right)}{(m - 2p - 1)(m - 2p - 2)^2(m - 2p - 4)}, \quad m - 2p > 4$ *and* $\forall$ i, j, k, ℓ.

Proof See Press (1972) p. 112 and the references therein. ■

From definition B.3 we can also derive the relationship between the inverse Wishart distribution and the Wishart distribution.

Theorem B.7 *If $S \sim W_p(n, \Sigma)$ with $(p \times p)$ scale matrix $\Sigma > 0$ and $n > p - 1$ degrees of freedom, then $S^{-1} = V \sim IW_p\left(n + p + 1, \Sigma^{-1}\right)$. Conversely, if $V \sim IW_p(m, \Psi)$ with $(p \times p)$ scale matrix $\Psi > 0$ and $m > 2p$ degrees of freedom, then $V^{-1} = S \sim W_p\left(m - p - 1, \Psi^{-1}\right)$.*

Proof See Press (1972) p.110 and Gupta & Nagar (2000) p.111. ■

The non-central Wishart distribution, though of quite complicated form, proves useful in various applications. It is a generalized version of the Wishart distribution and arises as the distribution of the random variable $S = XX'$ where $X \sim N_{p,n}(M, \Sigma \otimes I_n)$. In particular, it is defined as follows.

Definition B.4 *A $(p \times p)$ random matrix $S > 0$ is said to have a non-central Wishart distribution with $(p \times p)$ scale matrix $\Sigma > 0$, $n \geq p$ degrees of freedom and non-centrality parameter matrix Θ, denoted $S \sim W_p(n, \Sigma, \Theta)$, if its p.d.f. is given by*

$$\frac{1}{\sqrt{2^{np}}\Gamma_p\left(\frac{n}{2}\right)} \frac{\det(S)^{\frac{1}{2}(n-p-1)}}{\det(\Sigma)^{\frac{1}{2}n}} \exp\left\{-\frac{1}{2}tr\left(\Sigma^{-1}S\right)\right\} \exp\left\{-\frac{1}{2}tr(\Theta)\right\} {}_0F_1\left(\frac{n}{2}; \frac{1}{4}\Theta\Sigma^{-1}S\right). \tag{B.2}$$

In (B.2) ${}_0F_1\left(n/2; (1/4)\Theta\Sigma^{-1}S\right)$ is the hypergeometric function (Bessel function) of matrix argument. Its general form is given by (see Muirhead (1982) p.258)

$${}_0F_1(a; A) = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{\kappa} \frac{C_\kappa(A)}{(a)_\kappa}, \tag{B.3}$$

where $\sum_\kappa$ denotes summation over all partitions $\kappa = (k_1, \ldots, k_m)$ of k ($\sum_i k_i = k$, $k_1 \geq \cdots \geq k_m \geq 0$), $C_\kappa(A)$ is called the zonal polynomial of the matrix A associated with partition κ. Generally speaking, the role of the zonal polynomials, which are symmetric polynomials in the eigenvalues of a matrix A, is to generalize the powers x^κ of scalar x when x is replaced by a matrix A. The term $(a)_\kappa$ is called the generalized hypergeometric coefficient corresponding to partition κ and is given by

$$(a)_\kappa = \prod_{i=1}^{m} \prod_{j=1}^{k_i} \left(a + j - 1 - \frac{i-1}{2}\right).$$

For more details on the hypergeometric functions of matrix argument and the zonal polynomials the interested reader is referred to Muirhead (1982), Chapter 7. We should note that the hypergeometric function (B.3) does not admit a closed form expression, however it can be shown (see Muirhead (1982)) that the series converges for all A. Muirhead (1982) gives a simple algorithm that recursively computes the value of the hypergeometric function. More recently, Koev & Edelman (2004) devised a very accurate and computationally efficient algorithm that computes the value of the hypergeometric function.

Finally, it should be noted that if in (B.2) $\Theta = 0$, then the non-central Wishart distribution reduces to the Wishart distribution. The following theorem, illustrates how the non-central Wishart distribution arises from the matrix normal distribution.

Theorem B.8 *Suppose that the random $(p \times n)$ matrix $X \sim N_{p,n}(M, \Sigma \otimes I_n)$, with $n \geq p$ and $(p \times p)$ covariance matrix $\Sigma > 0$. Then, the random variable $S = XX'$ has a non-central Wishart distribution with n degrees of freedom, scale matrix $\Sigma > 0$ and non-centrality parameter $\Theta = \Sigma^{-1}MM'$, i.e. $S \sim W_p\left(n, \Sigma, \Sigma^{-1}MM'\right)$.*

Proof See Gupta & Nagar (2000) p.114. ■

Theorem B.9 *Suppose that the random $(p \times n)$ matrix $X \sim N_{p,n}(M, \Sigma \otimes I_n)$, with $n \geq p$ and $(p \times p)$ covariance matrix $\Sigma > 0$. Define the random variable $S = XX'$. Then, $E(S) = n\Sigma + MM'$.*

Proof See Gupta & Nagar (2000) p.119. ■

Another matrix variate distribution that plays an important role in the matrix variate beta distribution. It is the multivariate generalization of the univariate beta distribution. Its importance lies in the fact that there is a conjugacy result between the Wishart and the matrix beta distribution.

Definition B.5 *A $(p \times p)$ random matrix $U > 0$, such that $(I_p - U) > 0$ also holds, is said to have a matrix variate Beta distribution with parameters $a, b > (p-1)/2$, denoted $U \sim B_p(a, b)$ if its p.d.f. is given by*

$$\frac{\Gamma_p(a+b)}{\Gamma_p(a)\Gamma_p(b)} \det(U)^{a-\frac{1}{2}(p+1)} \det(I_p - U)^{b-\frac{1}{2}(p+1)}.$$

Clearly, if $U \sim B_p(a, b)$, then the random variable $(I_p - U) \sim B_p(b, a)$.

The conjugacy relating the multivariate Beta to the Wishart distribution is presented in the following theorem.

Theorem B.10 *Let the $(p \times p)$ matrices $S_1 \sim W_p(n_1, \Sigma)$ and $S_2 \sim W_p(n_2, \Sigma)$ be independent, with $n_1, n_2 > p-1$ and $\Sigma > 0$. Let R be the upper-triangular $(p \times p)$ matrix with positive diagonal elements such that $S_1 + S_2 = R'R$ and define U to be the $(p \times p)$ symmetric matrix such that $S_1 = R'UR$. Then, $S_1 + S_2$ and U are independent and $U \sim B_p\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$.*

Proof See Muirhead (1982) p.109. ■

Using theorem B.10 we can give an alternative, though less general, definition of the matrix variate beta distribution. In particular, using the same notation as in theorem B.10, we may say that a random matrix U is $B_p(n_1/2, n_2/2)$-distributed, if it can be written as $U = (R')^{-1} S_1 (R)^{-1}$. In addition, it is clear that if we define the random variable $V = (R')^{-1} S_2 (R)^{-1}$, then $V \sim B_p(n_2/2, n_1/2)$ (see also Dawid (1981) p.267 and Gupta & Nagar (2000) p.186).

Finally, another matrix variate distribution, which plays an important role in multivariate Bayesian analysis, is the matrix variate t-distribution. In what follows, we give the definition and some properties of this distribution.

Definition B.6 *The random $(p \times m)$ matrix $T \in \mathbb{R}^{p \times m}$ is said to have a matrix variate t-distribution with parameters, $M \in \mathbb{R}^{p \times m}$, $(p \times p)$ spread matrix $\Sigma > 0$, $(m \times m)$ spread matrix $\Psi > 0$ and $n > 0$ degrees of freedom, denoted $T_{p,m}(n, M, \Sigma, \Psi)$, if its p.d.f. is given by*

$$\frac{\Gamma_p\left(\frac{n+m+p-1}{2}\right)}{\sqrt{\pi^{mp}}\Gamma_p\left(\frac{n+p-1}{2}\right)} \det(\Sigma)^{-\frac{1}{2}m} \det(\Psi)^{-\frac{1}{2}p} \det\left(I_p + \Sigma^{-1}(T-M)\Psi^{-1}(T-M)'\right)^{-\frac{1}{2}(n+m+p-1)}. \quad \text{(B.4)}$$

Theorem B.11 *Let $T \sim T_{p,m}(n, M, \Sigma, \Psi)$ with n, M, Σ and Ψ as in definition B.6, then*

1. *$E(T) = M$ and*

2. $Var\left(vec\left(T'\right)\right) = \dfrac{1}{n-2}\Sigma \otimes \Psi$, $n > 2$.

Proof See Gupta & Nagar (2000) p.135. ■

The following theorem shows that the matrix variate t-distribution can be derived in a similar manner to the univariate theory.

Theorem B.12 *Let* $S \sim IW_p\left(n+2p, \Sigma\right)$ *for some* $n > 0$ *and* $(p \times p)$ *matrix* $\Sigma > 0$. *Suppose that conditional on* S, *the* $(p \times m)$ *matrix* $T \in \mathbb{R}^{p \times m}$ *has distribution* $T|S \sim N_{p,m}\left(0, S \otimes \Psi\right)$ *for some* $(m \times m)$ *matrix* $\Psi > 0$. *Then, marginally* $T \sim T_{p,m}\left(n, 0, \Sigma, \Psi\right)$.

Proof See Dawid (1981) p.266 and Gupta & Nagar (2000) p.134. ■

Finally, it should be noted (see Gupta & Nagar (2000) and the references therein) that when $m = 1$, $T = \boldsymbol{t}$ and $M = \boldsymbol{\mu}$ are $(p \times 1)$ and $\Psi = \psi$ is scalar, the matrix t-distribution reduces to a multivariate t-distribution, denoted $t_p\left(n, \psi, \boldsymbol{\mu}, \Sigma\right)$ and (B.4) becomes

$$\frac{\Gamma\left(\frac{n+p}{2}\right)}{\sqrt{\pi^p}\Gamma\left(\frac{n}{2}\right)} \det\left(\Sigma\right)^{-\frac{1}{2}} \psi^{-\frac{1}{2}p} \left[1 + \frac{1}{\psi}\left(\boldsymbol{t}-\boldsymbol{\mu}\right)' \Sigma^{-1}\left(\boldsymbol{t}-\boldsymbol{\mu}\right)\right]^{-\frac{1}{2}(n+p)}.$$

References

Abramowitz, M. & Stegun, N. C. (1970), *Handbook of Mathematical Functions*, Dover Publications, New York.

Aguilar, O. & West, M. (2000), 'Bayesian Dynamic Factor Models and Portfolio Allocation', *Journal of Business and Economic Statistics* **18**, 338–357.

Alexander, C. O. (2001), Orthogonal GARCH, *in* C. O. Alexander, ed., 'Mastering Risk', Vol. II, FT-Prentice Hall, London, pp. 21–38.

Andersen, T. G. & Sorensen, B. E. (1996), 'GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study', *Journal of Business and Economic Statistics* **14**, 328–352.

Anderson, B. D. O. & Moore, J. B. (1979), *Optimal Filtering*, Prentice Hall, N.J.

Aydemir, A. B. (1998), Volatility Modelling in Finance, *in* J. Knight & S. Satchell, eds, 'Forecasting Volatility in the Financial Markets', Butterworth-Heinemann, Oxford, pp. 1–46.

Barndorff-Nielsen, O. E. & Shephard, N. (2001), 'Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics', *Journal of the Royal Statistical Society, Series B* **63**, 167–241.

Bernardo, J. M. & Smith, A. F. M. (1994), *Bayesian Theory*, John Wiley and Sons, Chichester.

Bollerslev, T. (1990), 'Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model', *The Review of Economics and Statistics* **72**, 498–505.

Bollerslev, T., Engle, R. F. & Wooldridge, J. M. (1988), 'A Capital Asset Pricing Model with Time-Varying Covariances', *The Journal of Political Economy* **96**, 116–131.

Breidt, F. J. & Carriquiry, A. L. (1996), Improved Quasi-Maximum Likelihood Estimation for Stochastic Volatility Models, *in* J. C. Lee, W. O. Johnson & A. Zellner, eds, 'Modelling and Prediction: Honoring Seymour Geisser', Springer-Verlag, New York, pp. 228–247.

Brockwell, P. J. & Davis, R. A. (1991), *Time Series: theory and methods*, Springer-Verlag, New York.

Broto, C. & Ruiz, E. (2004), 'Estimation Methods for Stochastic Volatility Models: A Survey', *Journal of Economic Surveys* **18**, 613–742.

Carlin, B. P., Polson, N. G. & Stoffer, D. S. (1992), 'A Monte Carlo Approach to Nonnormal and Nonlinear State-Space 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.

Carter, C. K. & Kohn, R. (1996), 'Markov Chain Monte Carlo in Conditionally Gaussian State Space Models', *Biometrika* **83**, 589–601.

Chib, S. & Greenberg, E. (1995), 'Understanding the Metropolis-Hastings Algorithm', *The American Statistician* **49**, 327–335.

Chib, S., Nardari, F. & Shephard, N. (2002), 'Markov Chain Monte Carlo Methods for Stochastic Volatility Models', *Journal of Econometrics* **108**, 281–316.

Chib, S., Nardari, F. & Shephard, N. (2005), 'Analysis of High Dimensional Multivariate Stochastic Volatility Models', *Washington University* . Working Paper.

Danielsson, J. (1994), 'Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood', *Journal of Econometrics* **64**, 375–400.

Danielsson, J. & Richard, J. F. (1993), 'Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models', *Journal of Applied Econometrics* **8**, 153–173.

Dawid, A. P. (1981), 'Some Matrix-Variate Distribution Theory: Notational Considerations and a Bayesian Application', *Biometrika* **68**, 265–274.

deJong, P. & Shephard, N. (1995), 'The Simulation Smoother for Time Series Models', *Biometrika* **82**, 339–350.

Diebold, F. X. & Nerlove, M. (1989), 'The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor Arch Model', *Journal of Applied Econometrics* **4**, 1–21.

Durbin, J. & Koopman, S. J. (1997), 'Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models', *Biometrica* **84**, 669–684.

Durbin, J. & Koopman, S. J. (2002), 'A Simple and Efficient Simulation Smoother for State Space Time Series Analysis', *Biometrika* **89**(3), 603–616.

Engle, R. F. (2002), 'Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroscedasticity Models', *Journal of Business and Economic Statistics* **20**, 339–350.

Engle, R. F. & Kroner, K. F. (1995), 'Multivariate Simultaneous Generalized ARCH', *Econometric Theory* **11**, 122–150.

Engle, R. F., Ng, V. K. & Rothschild, M. (1990), 'Asset Pricing with a Factor-ARCH Covariance Structure', *Journal of Econometrics* **45**, 213–237.

Engle, R. F. & Sheppard, K. (2001), 'Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH', *University of California, San Diego* . Working Paper.

Fama, E. F. (1965), 'The Behavior of Stock-Market Prices', *Journal of Business* **38**, 34–105.

Fruhwirth-Schnatter, S. (1994), 'Data Augmentation and Dynamic Linear Models', *Journal of Time Series Analysis* **15**, 183–202.

Fuller, W. A. (1996), *Introduction to Statistical Time Series*, Wiley Series in Probability and Statistics, second edn, Wiley, New York.

Geweke, J. (1989), 'Bayesian Inference in Econometric Models Using Monte Carlo Integration', *Econometrica* **57**, 1317–1339.

Geweke, J. (1994), 'Comment on 'Bayesian Analysis of Stochastic Volatility Models' by Jacquier, Polson and Rossi', *Journal of Business and Economic Statistics* **12**, 389–392.

Ghysels, E., Harvey, A. C. & Renault, E. (1996), Stochastic Volatility, *in* G. S. Maddala & C. R. Rao, eds, 'Statistical Methods in Finance', North Holland, Amsterdam, pp. 119–191.

Gilks, W. R. & Wild, P. (1992), 'Adaptive Rejection Sampling for Gibbs Sampling', *Applied Statistics* **41**, 337–348.

Gourieroux, C., Jasiak, J. & Sufana, R. (2004), 'The Wishart Autoregressive Process of Multivariate Stochastic Volatility', *CREST* . Working Paper.

Gourieroux, C. & Sufana, R. (2003), 'Wishart Quadratic Term Structure Models'. Working Paper 03-10.

Gourieroux, C. & Sufana, R. (2004), 'Derivative Pricing with Multivariate Stochastic Volatility: Application to Credit Risk'. Working Paper.

Grimmett, G. R. & Stirzaker, D. R. (2001), *Probability and Random Processes*, third edn, Oxford University Press, New York.

Grunwald, G. K., Hyndman, R. J. & Tedesco, L. M. (1997), 'A Unified View of Linear AR(1) Models', *University of Melbourne* . Working Paper.

Gupta, A. K. & Nagar, D. K. (2000), *Matrix Variate Distributions*, Chapman and Hall, Florida.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, N.J.

Han, Y. (2002), 'The Economic value of Volatility Modelling: Asset Allocation with a High Dimensional Dynamic Latent Factor Multivariate Stochastic Volatility Model', *Washington University* . Working Paper.

Hansen, L. P. (1982), 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica* **50**, 1029–1054.

Harvey, A. C. (1981), *Time Series Models*, Philip Allan, Hertfordshire.

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

Harvey, A. C. (1993), *Time Series Models*, second edn, Pearson Education, Essex.

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

Harvey, A. C. & Shephard, N. (1996), 'Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns', *Journal of Business and Economic Statistics* **14**, 429–434.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, Springer Verlag, New York.

Jacquier, E., Polson, N. G. & Rossi, P. (1995), 'Models and Priors for Multivariate Stochastic Volatility', *CIRANO* . Working Paper: 95s-18.

Jacquier, E., Polson, N. G. & Rossi, P. (1999), 'Stochastic Volatility: Univariate and Multivariate Extensions', *CIRANO* . Working Paper: 99s-26.

Jacquier, E., Polson, N. G. & Rossi, P. E. (1994), 'Bayesian Analysis of Stochastic Volatility Models', *Journal of Business and Economic Statistics* **12**, 371–389.

Jacquier, E., Polson, N. G. & Rossi, P. E. (2004), 'Bayesian analysis of stochastic volatility models with fat-tails and correlated errors', *Journal of Econometrics* **122**, 185–212.

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

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models', *The Review of Economic Studies* **65**, 361–393.

Kitagawa, G. (1987), 'Non-Gaussian State-Space Modelling of Nonstationary Time Series', *Journal of the American Statistical Association* **82**, 1032–1041.

Koev, P. & Edelman, A. (2004), 'The Efficient Evaluation of the Hypergeometric Function of Matrix Argument', *Massachusetts Institute of Technology* . Working Paper.

Ledoit, O., Santa-Clara, P. & Wolf, M. (2003), 'Flexible Multivariate GARCH Modeling With an Application to International Stock Markets', *The Review of Economics and Statistics* **85**, 735–747.

Mahieu, R. J. & Schotman, P. C. (1998), 'An Empirical Application of Stochastic Volatility Models', *Journal of Applied Econometrics* **13**, 333–360.

Mandelbrot, B. (1963), 'The Variation of Certain Speculative Prices', *Journal of Business* **36**, 394–419.

Melino, A. & Turnbull, S. M. (1990), 'Pricing Foreign Currency Options with Stochastic Volatility', *Journal of Econometrics* **45**, 239–265.

Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory*, John Wiley and Sons, New York.

Nelson, D. B. (1991), 'Conditional Heteroscedasticity in Asset Returns: A New Approach', *Econometrica* **59**, 347–370.

O'Hagan, A. (1994), *Bayesian Inference*, Vol. 2B of *Kendall's Advanced Theory of Statistics*, Edward Arnold, Chichester.

Omori, Y., Chib, S., Shephard, N. & Nakajima, J. (2004), 'Stochastic Volatility with Leverage Effect: Fast Likelihood Inference', *University of Tokyo* . Working Paper.

Philipov, A. & Glickman, M. E. (2004), 'Multivariate Stochastic Volatility via Wishart Processes', *Boston University* . Working Paper.

Pitt, M. K., Chatfield, C. & Walker, S. G. (2002), 'Constructing First Order Stationary Autoregressive Models via Latent Processes', *Scandinavian Journal of Statistics* **29**, 657–663.

Pitt, M. K. & Shephard, N. (1999), Analysis of Time Varying Covariances: A Factor Stochastic Volatility Approach, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds, 'Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting', Oxford: Oxford University Press, pp. 547–570.

Pitt, M. K. & Walker, S. G. (1999), 'Construction of Stationary Time Series via the Gibbs Sampler with Application to Volatility Models', *Imperial College* . Working Paper.

Press, S. J. (1972), *Applied Multivariate Analysis*, Holt, Rinehart and Winston, New York.

Robert, C. P. & Casella, G. (1999), *Monte Carlo Statistical Methods*, Springer-Verlag, New York.

Ruiz, E. (1994), 'Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models', *Journal of Econometrics* **63**, 289–306.

Sandmann, G. & Koopman, S. J. (1998), 'Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood', *Journal of Econometrics* **87**, 271–301.

Shephard, N. (1993), 'Fitting Nonlinear Time-Series Models with Applications to Stochastic Variance Models', *Journal of Applied Econometrics* **8**, S135–S152. Supplement: Special Issue on Econometric Inference Using Simulation Techniques.

Shephard, N. (1994), 'Partial Non-Gaussian State Space', *Biometrika* **81**, 115–131.

Shephard, N. (1996), Statistical aspects of ARCH and stochastic volatility, *in* D. R. Cox, D. V. Hinkley & O. E. Barndorff-Nielsen, eds, 'Time Series Models: In econometrics, finance and other fields', Chapman and Hall, London, pp. 1–67.

Shephard, N. & Kim, S. (1994), 'Comment on 'Bayesian Analysis of Stochastic Volatility Models' by Jacquier, Polson and Rossi', *Journal of Business and Economic Statistics* **12**, 406–410.

Shephard, N. & Pitt, M. K. (1997), 'Likelihood Analysis of Non-Gaussian Measurment Time Series', *Biometrika* **84**, 653–667.

Tanizaki, H. (2000), 'Nonlinear and non-Gaussian state-space modelling with Monte Carlo Techniques: A Survey and Comparative Study', *Kobe University* . Working Paper.

Tanizaki, H. & Mariano, R. S. (1994), 'Prediction, Filtering and Smoothing in Non-Linear and Non-Normal Cases Using Monte Carlo Integration', *Journal of Applied Econometrics* **9**, 163–179.

Tanizaki, H. & Mariano, R. S. (1998), 'Nonlinear and non-Gaussian state-space modelling with Monte Carlo simulations', *Journal of Econometrics* **83**, 263–290.

Taylor, S. J. (1986), *Modelling Financial Time Series*, John Wiley, Chichester.

Taylor, S. J. (1994), 'Modelling Stochastic Volatility: A Review and Comparative Study', *Mathematical Finance* **4**, 183–204.

Tierney, L. (1994), 'Markov Chains for Exploring Posterior Distributions', *The Annals of Statistics* **22**, 1701–1728.

Uhlig, H. (1994), 'On Singular Wishart and Singular Multivariate Beta Distributions', *The Annals of Statistics* **22**, 395–405.

Uhlig, H. (1997), 'Bayesian Vector Autoregressions with Stochastic Volatility', *Econometrica* **65**, 59–73.

West, M. & Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, Second edn, Springer-Verlag, New York.