BAYESIAN STATISTICS 6, pp. 547–570
J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (Eds.)

Time-Varying Covariances: A Factor Stochastic Volatility Approach

MICHAEL K. PITT and NEIL SHEPHARD
Imperial College, London, UK, and *Nuffield College, Oxford, UK*

SUMMARY

We propose a factor model which allows a parsimonious representation of the time series evolution of covariances when the number of series being modelled becomes very large. The factors arise from a standard stochastic volatility model as does the idiosyncratic noise associated with each series. We use an efficient method for deriving the posterior distribution of the parameters of this model. In addition we propose an effective method of Bayesian model selection for this class of models. Finally, we consider diagnostic measures for specific models.

Keywords: EXCHANGE RATES; FILTERING; MARKOV CHAIN MONTE CARLO; SIMULATION; SIR; STATE SPACE; VOLATILITY.

1. INTRODUCTION

Many financial time series exhibit changing variance and this can have important consequences in formulating economic or financial decisions. In this paper we will suggest some very simple multivariate volatility models in an attempt to capture the changing cross-covariance patterns of time series. Our aim is to produce models which can eventually be used on time series of many 10s or 100s of asset returns.

There are two types of univariate volatility model for asset returns; the autoregressive conditional heteroskedastic (ARCH) and stochastic volatility (SV) families. Our focus will be on the latter. The stochastic volatility class builds a time varying variance process by allowing the variance to be a latent process. The simplest univariate SV model, due to Taylor (1982) in this context, can be expressed as

$$y_t = \varepsilon_t \sigma \exp(\alpha_t/2), \quad \alpha_{t+1} = \phi\alpha_t + \eta_t, \quad \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim NID\left\{0, \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}\right\}. \qquad (1)$$

Here σ is the modal volatility of the model, while σ_η is the volatility of the log-volatility. One interpretation of the latent variable α_t is that it represents the random and uneven flow of new information into the market; this follows the work of Clark (1973) [1].

Stochastic volatility models are a variance extension of the Gaussian 'Bayesian dynamic linear models' reviewed in West and Harrison (1997) [2]. In recent years many estimation

[1] The model also represents a Euler discretisation of the continuous time model for a log asset price $y^*(t)$, where $w(t)$ and $b(t)$ are independent Brownian motions, and $dy^*(t) = \sigma \exp\{\alpha(t)/2\}\, dw(t)$ where $d\alpha(t) = -\phi\alpha(t)dt + \tau db(t)$. This model was proposed by Hull and White (1987) for their generalisation of the Black-Scholes option pricing scheme. Throughout the paper we will work in discrete time, however our proposed multivariate model has an obvious continuous time version.

[2] A conjugate time series model for time varying variances was put forward by Shephard (1994a) and generalized to covariances by Uhlig (1997). Although these models have some attractions, they impose non-stationarity on the volatility process which is not attractive from a financial economics viewpoint.

procedures have been suggested for SV models. Markov chain Monte Carlo (MCMC) methods are commonly used in this context following papers by Shephard (1993) and Jacquier, Polson and Rossi (1994) which have been greatly refined and simplified by Kim, Shephard and Chib (1998) and Shephard and Pitt (1997). Some of the early literature on SV models is discussed in Shephard (1996) and Ghysels, Harvey and Renault (1996).

The focus of this paper will be on building multivariate SV models for asset returns in financial economics. In order to do this we will need some notation. We refer to (1) as "uncentered" as the states have an unconditional mean of 0. We generally work with the "centered" version of (1) with $y_t = \varepsilon_t \exp(\alpha_t/2)$ and $\alpha_{t+1} = \mu + \phi(\alpha_t - \mu) + \eta_t$, for reasons of computational efficiency in MCMC estimation, see Pitt and Shephard (1998). We write this as $y \sim \mathcal{I}SV_n(\phi; \sigma_\eta; \mu)$, that is the series $y = (y_1, ..., y_n)'$ arises from a stochastic volatility model, conditionally independent of any other series.

1.1. *Economic Motivation*

Multivariate models of asset returns are very important in financial economics. In this subsection we will discuss three reasons for studying multivariate models.

Asset pricing theory (APT). This links the expected return on holding an individual stock to the covariance of the returns. A simple exposition of APT, developed by Ross (1976), is given in Campbell, Lo and MacKinlay (1997, pp. 233–240). The main flavour of this can be gleaned from a parametric version of the basic model where we assume that arithmetic returns follow a classic factor analysis structure (Bartholomew (1987)) for an N dimensional time series $y_t = \alpha + Bf_t + \varepsilon_t$ where $(\varepsilon_t', f_t')' \sim NID\,(0, D)$, where D is diagonal, B is a matrix of factor loadings and f_t is a K dimensional vector of factors. The APT says that as the dimension of y_t increases to such an extent that y_t well approximates the market then so $\alpha \simeq \iota r + B\lambda$, where r is the riskless interest rate, ι is a vector of ones and λ is a vector representing the factor risk premium associated with the factors f_t. Typically applied workers take the factor risk premiums as the variances of the factors. Important Bayesian work to estimate and test the above restrictions imposed by the theory has included Geweke and Zhou (1996) and McCulloch and Rossi (1991). Unfortunately unless very low frequency data is used, such as monthly returns, the NID assumption is massively rejected by the data which displays statistically significant volatility clustering and fat tails and so the methods they develop need to be extended.

Asset allocation. Suppose an investor is allocating resources between assets which have a one period (say a month) arithmetic return of $y_t \sim NID$. A classic solution to this (Ingersoll (1987 Ch. 4)) is to design a portfolio which minimises its variance for a given level of expected wealth. Interesting Bayesian work in this context includes Quintana (1992). For high frequency data we need to extend the above argument by writing that $E(y_t|\mathcal{F}_{t-1}) = a_t$ and $Var(y_t|\mathcal{F}_{t-1}) = \Sigma_t$ where $\mathcal{F}_{t-1}$ is the information available at the time of investment.

Value at Risk (VaR). VaR studies the extreme behaviour of a portfolio of assets (see, for example, Dave and Stahl (1997)). In the simplest case the interest is in the tails of the density of $\omega' y_t|\mathcal{F}_{t-1}$.

1.2. *Empirically Reasonable Models*

Although factor models give one way of tackling the APT, portfolio analysis problems and VaR, the standard NID assumptions used above cannot be maintained. Instead, in Section 2 we propose replacing the $(\varepsilon_t', f_t') \sim NID\,(0, D)$ assumption by specifying a model which allows each element of this vector to follow an ISV process.

Diebold and Nerlove (1989) and King, Sentana and Wadhwani (1994) have used a similar type of model where the factors and idiosyncratic errors follow their own ARCH based process with the conditional variance of a particular factor being a function of lagged values of that factor [3]. Unfortunately the rigorous econometric analysis of such models is very difficult from a likelihood viewpoint (see Shephard (1996, pp 16–18)). Jacquier, Polson and Rossi (1995) have briefly proposed putting a SV structure on the factors and allowing the ε_t to be NID. However, they have not applied the model or the methodology they propose, nor have they consider the identification issues which arise with this type of factor structure. Their proposed estimation method is based upon MCMC for Bayesian inference.

Kim, Shephard and Chib (1998) put forward the basic model structure we suggest in this paper. They allow the ε_t to follow independent SV processes — although this model was not fitted in practice. In a recent paper, Aguilar and West (1998) have implemented this model using the Kim, Shephard and Chib (1998) mixture MCMC approach. The work we report here was conducted independently of the Aguilar and West (1998) paper. We use different MCMC techniques which we believe are easier to extend to other interesting volatility problems. Further we design a simulation based filtering algorithm to validate the fit of the model, as well as to estimate volatility using contemporaneous data.

1.3. *Data*

Although our modelling approach is based around an economic theory for stock returns, in our applied work we will employ exchange rates, with 4290 observations on daily closing prices of five exchange rates quoted in US dollars (USD) from $2/1/81$ to $30/1/98$ [4].

We write the underlying exchange rates as $\{R_{it}\}$ and then construct the continually compounded rates $y_{it} = 100 \times (\log R_{it} - \log R_{it-1})$, for $i = 1, 2, 3, 4, 5$ and $t = 2, 3, ..., 4290$. The five currencies we use are the Pound (P), Deutschemark (DM), Yen (Yen), Swiss Franc (SF) and French Franc (FF). From an economic theory view it would be better to alter the returns to take into account available domestic riskless interest rates (see, for example, McCurdy and Morgan, 1991) as well as some other possible explanatory variables. However, neglecting these additional variables does not make a substantial difference to our volatility analysis as the movements in exchange rates dominate the typically small changes in daily interest rate differentials and other variables. Hence we will relegate consideration of these second order effects to later work.

The time series of the five returns are shown in Figure 1 together with the correlograms of the returns and their absolute values. The correlograms indicate no great autocorrelation in the returns. The changing volatility of the returns is clearly indicated by the correlogram of the absolute values. It is clear that there is positive but small autocorrelation at high lags for each of the returns. The sample mean and covariance (correlations in upper triangle) of the 5 returns

[3] We note in passing that there are other classes of multivariate models which have been developed in the econometrics literature. The literature on multivariate ARCH models has been cursed by the problem of parsimony as their most general models, discussed in Engle and Kroner (1995), have enormous numbers of parameters. Hence much of this literature is concerned with appropriately paring down the structure in order to get estimable models. The focus is, as before, on allowing the one step ahead covariance matrix $Var(y_t|\mathcal{F}_{t-1})$ to depend on lagged data. As we will not be using this style of model we refer the interested reader to Bollerslev, Engle and Nelson (1994 pp. 3002–3010) for a detailed discussion of this literature.

[4] We use the 'Noon buying rates in New York City certified by the Federal Reserve Bank of New York for customs purposes...' Extensive exchange rate data is made available by the Chicago Federal Reserve Bank at `www.frbchi.org/econinfo/finance/for-exchange/welcome.html`

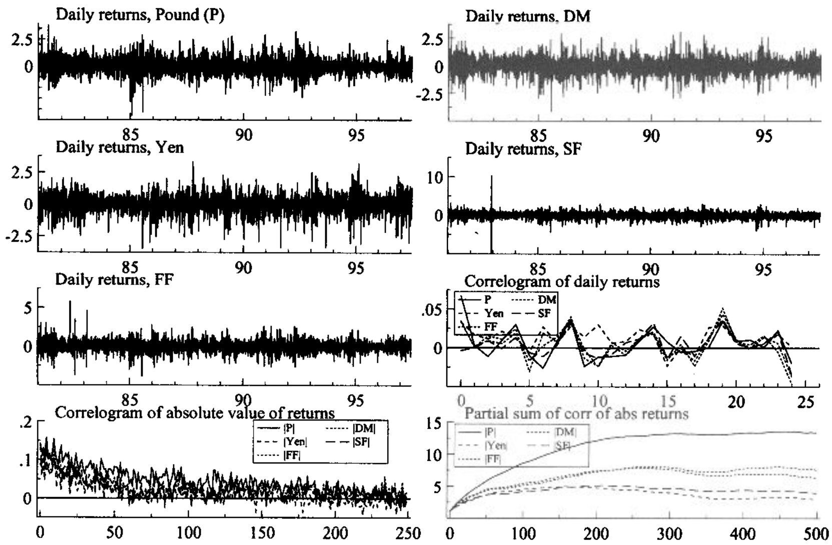

Figure 1. *Daily returns for P (top left), DM (top right), Yen (second row, left), SF (second row, right) and FF (third row, left). Correlograms for returns (third row, right) and for the absolute values of returns (fourth row, left) and the corresponding partial sum of the correlograms (fourth row, right).*

(US dollar versus P, DM, Yen, SF, FF in order) are

$$\overline{y} = \begin{pmatrix} 0.00881 \\ -0.00175 \\ -0.01086 \\ -0.00440 \\ 0.00690 \end{pmatrix}, \Sigma = \begin{pmatrix} 0.4669 & 0.7518 & 0.5068 & 0.6917 & 0.7280 \\ 0.3598 & 0.4906 & 0.6448 & 0.8915 & 0.9454 \\ 0.2263 & 0.2951 & 0.4269 & 0.6235 & 0.6119 \\ 0.3779 & 0.4993 & 0.3257 & 0.6393 & 0.8463 \\ 0.3427 & 0.4563 & 0.2754 & 0.4662 & 0.4747 \end{pmatrix}.$$

The mean return is close to 0 for all the series. The returns are all strongly positively correlated, with the SF, DM and FF being particularly correlated. In our applied work we will typically subtract the sample mean before fitting volatility models, in order to simplify the analysis.

1.4. *Outline of the Paper*

The structure of the paper is as follows. In Section 2 we consider the multivariate factor SV model. We go on to consider MCMC issues in Section 3. In Section 3.1 we discuss the univariate SV model together with the MCMC methods used in its estimation. We apply the univariate methods to the individual exchange rates in our dataset. The MCMC methods for Bayesian inference applied to the factor SV (FSV) model are discussed in Section 3.2.

In Section 4, we consider the estimation and testing of the factor model for our dataset of 5 exchange rate returns. In Section 4.1 we estimate the parameters of the model under consideration. We consider one-step-ahead estimates as diagnostics for assessing the model's fit in Section 4.2. These diagnostics are obtained by use of filtering (via simulation) algorithm, the method being detailed in Section 4.4.

2. MULTIVARIATE FACTOR SV MODEL

2.1. *Specification*

In this paper we consider the following factor SV (FSV) model,

$$\begin{aligned} y_t &= \beta f_t + \omega_t, t = 1, ..., n \\ \omega_j &\sim ISV_n(\phi^{\omega_j}; \sigma_\eta^{\omega_j}; \mu^{\omega_j}),\ j = 1, .., N \\ f_i &\sim ISV_n(\phi^{f_i}; \sigma_\eta^{f_i}; 0), i = 1, .., K, \end{aligned} \tag{2}$$

where N represents the number of separate series, K $(< N)$ represents the number of factors [5]. β represents a $N \times K$ matrix of factor loadings, whilst f_t is a $K \times 1$ vector, the unobserved factor at time t. For the moment we shall assume that β is unrestricted. The necessary restrictions will be outlined presently. Jacquier, Polson and Rossi (1995) have briefly discussed a similar model, but they set $\omega_t \sim NID$ rather than allowing each of the N idiosyncratic error terms of ω_t to follow an independent SV process. Our hope is that this will allow us to fit the data with K being much smaller than N as we regard the factor structure as sufficient (particularly if K is reasonably large) to account for the non-diagonal elements of the variance matrix of the returns, but not sufficient to explain all of the marginal persistence in volatility.

Our choice of model naturally leads to a parsimonious structure as the number of unknown parameters is now linear in N when the number of factors is fixed. For exchange rates this model appears extremely plausible. If we consider the returns on various currencies against the USD, for example, then a single factor model may be sensible. In this case, a large part of common factor term, f_t, may account for the part of the return resulting from changes in the American economy. The idiosyncratic terms could explain the part of the returns which results from the independent country-specific shocks.

2.2. *Identification and Priors*

For identifiability, restrictions need to be imposed upon the factor weighting matrix. Sentana and Fiorentini (1997) indicate that the identifiability restrictions, for the conditionally heteroskedastic factor models, are less severe than in static (non-time series) factor analysis (Bartholomew (1987) and Geweke and Zhou (1996)). However, we have decided to impose the traditional structure in order to allow the parameters to be easily estimated. Following for example Geweke and Zhou (1996), we set $\beta_{ij} = 0$, and $\beta_{ii} = 1$ for $i = 1, .., K$ and $j > i$.

Our model has three sets of parameters: idiosycratic SV parameters $\left\{\phi^{\omega_j}; \sigma_\eta^{\omega_j}; \mu^{\omega_j}\right\}$, factor SV parameters $\left\{\phi^{f_i}; \sigma_\eta^{f_i}; \mu^{f_i}\right\}$ and the factor loading matrix β. We take priors for all the SV parameters which are independent, with the same distribution across the factors and idiosycratics. We do this as we have little experience of how the data will split the variation into the factor and idiosycratic components. We adopt proper priors for each of the $\left\{\phi^{\omega_j}; \sigma_\eta^{\omega_j}; \mu^{\omega_j}\right\}$ and $\left\{\phi^{f_i}; \sigma_\eta^{f_i}; \mu^{f_i}\right\}$ parameters that have previously been successfully used on daily exchange rate data by Shephard and Pitt (1997) and Kim, Shephard and Chib (1998). In particular we let $\phi = 2\phi^* - 1$ where ϕ^* is distributed as Beta with parameters $(18, 1)$, imposing stationarity

[5] The first multivariate SV model proposed in the literature was due to Harvey, Ruiz and Shephard (1994) who allowed the variances of multivariate returns to vary over time but constrained the correlations to be constant. This is an unsatisfactory model from an economic viewpoint. There is a predating literature on informal methods for allowing covariance matrices to evolve over time in order to introduce a measure of discounting into filtering equations. Important work includes Quintana and West (1987). These techniques can be rationalised by the non-stationary variance and covariance models of Shephard (1994a) and Uhlig (1997).

on the process, while setting $\mu \sim N(-1,9)$. Further we set $\sigma_\eta^2|\phi,\mu \sim \mathcal{IG}(\frac{\sigma_r}{2},\frac{S_\sigma}{2})$, where $\mathcal{IG}$ denotes the inverse-gamma distribution and $\sigma_r = 10$ and $S_\sigma = 0.01 \times \sigma_r$. The conjugate Gaussian updating of μ and conjugate $\mathcal{IG}$ updating of σ_η^2, in each case conditional upon the corresponding states, is described in Pitt and Shephard (1998) whilst the more intricate (but very efficient) rejection method used to update ϕ is used in Shephard and Pitt (1997) and more fully outlined in Kim, Shephard and Chib(1998).

For each element of β we assume $\beta_{ij} \sim N(1,25)$, reflecting the large prior uncertainty we have regarding these parameters. The updating strategy for β is detailed in Section 3.2.

3. MARKOV CHAIN MONTE CARLO ISSUES

3.1. *Univariate Models*

Before proceeding with multivariate extensions we first estimate the univariate SV model (1) using the MCMC methods designed by Shephard and Pitt (1997). Extending to the multivariate case is then largely trivial as the univariate code can be included to take care of all the difficult parts of the sampling. Computationally efficient single-move MCMC methods (which move a single state α_t conditional upon all other states $\alpha_1,...,\alpha_{t-1},\alpha_{t+1},...,\alpha_n$ and the parameters) have been used on this model by Shephard and Pitt (1997) and Kim, Shephard and Chib (1998).

Table 1. *Parameter of univariate models for the 5 currencies from 1981 to 1998. Summaries of Figure 2. 20,000 replications of the multi-move sampler, using 40 stochastic knots. M-C S.E. denotes Monte Carlo standard error and is computed using 1000 lags (except for beta for which 200 lags are used). Ineff denotes the estimated integrated autocorrelation.*

	Mean	M-C S.E.	Ineff	Covariance & *Correlation* of Posterior		
			British Pound			
$\sigma\|y$	0.5992	0.000333	2.4	0.000917	*-0.0982*	*0.0698*
$\sigma_\eta\|y$	0.1780	0.00251	285	-0.0000625	0.000442	*-0.796*
$\phi\|y$	0.9702	0.000672	186	0.0000148	-0.000117	0.0000487
			German Deutschemark			
$\sigma\|y$	0.6325	0.000282	2.3	0.000694	*-0.105*	*0.0868*
$\sigma_\eta\|y$	0.1714	0.00153	153	-0.000048	0.000307	*-0.766*
$\phi\|y$	0.9652	0.000503	94	0.0000168	-0.0000982	0.0000536
			Japanese Yen			
$\sigma\|y$	0.5544	0.000388	9.5	0.000316	*-0.453*	*0.389*
$\sigma_\eta\|y$	0.4470	0.00584	203	-0.000467	0.00336	*-0.916*
$\phi\|y$	0.8412	0.00322	192	0.000227	-0.00175	0.00108
			Swiss Franc			
$\sigma\|y$	0.7087	0.00029	2.8	0.000594	*-0.115*	*0.0959*
$\sigma_\eta\|y$	0.1911	0.00199	181	-0.0000587	0.000437	*-0.820*
$\phi\|y$	0.9531	0.000787	124	0.0000234	-0.000171	0.00010
			French Franc			
$\sigma\|y$	0.6042	0.000240	2.2	0.000522	*-0.151*	*0.108*
$\sigma_\eta\|y$	0.2342	0.00207	159	-0.0000799	0.000539	*-0.802*
$\phi\|y$	0.9472	0.00076	113	0.0000248	-0.000188	0.000102

This sampler is then combined with an algorithm which samples the parameters conditional upon the states and measurements, i.e. from $f(\theta|\alpha, y)$, where $\theta = (\mu, \phi, \sigma_\eta^2)'$. However, the high posterior correlation which arises between states for typical financial time series means that the integrated autocorrelation time can be very high. To combat this a method of proposing moves of blocks of states simultaneously for the density

$$\log f(\alpha_t, ..., \alpha_{t+k}|\alpha_{t-1}, \alpha_{t+k+1}, y_t, ..., y_{t+k}, \theta)$$

via a Metropolis method was introduced by Shephard and Pitt (1997). An important feature of this method is that k is chosen randomly for each proposal, meaning sometimes the blocks are small and other times they are very large. This ensures the method does not become stuck by excessive amounts of rejection. This is the method which we shall adopt in this paper. An additional advantage is that the method is extremely general and extendable.

The univariate SV model is estimated, using 20,000 iterations of the above method, for each of the exchange rates. The simulated parameters and corresponding correlograms are given in Figure 2. Here, as later in the paper, we report the σ parameter, for ease of interpretation, associated with the uncentred SV model of (1) rather than the unconditional mean of the log-volatilities in the $ISV_n(\phi; \sigma_\eta; \mu)$ parameterisation. The corresponding Table 1 show the posterior estimates of the mean, standard error (of the sample mean), covariance and correlation for the three parameters for each of the series under examination. The standard errors (estimated using a Parzen based spectral estimator) have been calculated taking into account the variance inflation (which we call inefficiency) due to the autocorrelation in the MCMC samples. We set the expected number of blocks, which we call knots, in the sampling mechanism to 40 and use the centered parameterisation in the computations. Every 10 iterations the single move state sampler detailed in Shephard and Pitt (1977) has been employed [6]. The entire dataset of 4290 returns on daily closing prices of the five exchange rates from 2/1/81 to 30/1/98 has been used.

The USD/Yen return has the least persistence in volatility changes, as we can see by the low posterior mean for ϕ and the high posterior mean for σ_η. This indicates that there is relatively little predictive power for the variance of this return in comparison with the other series. The USD/P return is the most persistent of the series, closely followed by the USD/DM. The USD/SF and USD/FF returns exhibit similar medium persistence.

The parameter plots on the left of Figure 2 have been thinned out (taking every 20th iteration) for visibility. The correlograms (for all the sampled parameters) indicate that the MCMC method works well as the correlograms (over all iterations) die down at or before lags of 500.

In the following section, we shall examine how the univariate SV methodology outlined aids in estimating the FSV model. In addition, we will see how the estimated volatilities change under the factor model.

3.2. *MCMC Issues for Factor Models*

In this section we consider MCMC issues for the FSV model. The key additional feature of the approach is that we will augment the posterior to simulate from all of $\{\omega, f, \theta, \alpha, \beta \mid y\}$ (where θ includes all the fixed parameters in the model except β) for this allows the univariate code to be bootstrapped in order to tackle the multivariate problem. This key insight appeared first in Jacquier, Polson and Rossi (1995). Most of the new types of draws are straightforward as the $\{\omega_t, f_t \mid \theta, \alpha, y, \beta\}$ are conditionally independent and Gaussian (although degenerate).

[6] This ensures that even in the presence of very large returns or low state persistence, each of the states will be sampled with probability close to 1.

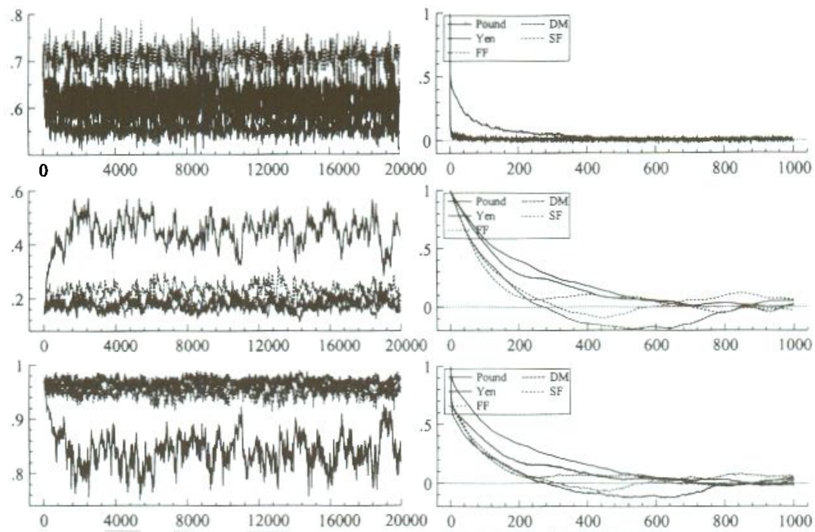

Figure 2. *Parameters for univariate SV model. The simulated parameters (20000 iterations) shown on left; σ (top), σ_η and ϕ (bottom) together with corresponding acfs on right. See Table 1.*

The only new issue which arises is updating samples from $\{\beta] \mid \omega, f, \alpha, y, \theta\}$. Let us now consider column i represented by β_i, $i = 1, ..., K$ and the remaining columns by $\beta_{\backslash i}$. Then, assuming a Gaussian prior $N(\mu_i, \Sigma_i)$ on each column β_i we find that $\beta_i|\beta_{\backslash i}, y_t, \alpha_t^\omega, f$ is Gaussian and can easily be drawn imposing the identification constraints $\beta_{ij} = 0$ for $j > i$ and $\beta_{ii} = 1$, as suggested in Section 2. We iterate through the columns for $i = 1, .., K$.

4. SINGLE FACTOR MODEL FOR FIVE SERIES

4.1. *MCMC Analysis*

We now concentrate on the fit of a single factor ($K = 1$) FSV model to the 5 series already considered in our univariate SV analysis. We used 4018 returns by discarding the last year of data for later model checking purposes. We apply the above MCMC approach to the data. We used 80 knots (average block size about 50) for the block sampler for both the states of the factor and the five sets of idiosyncratic states. However, after an initial short run we introduced an additional sweep (for each overall MCMC iteration) for the parameters and states associated with the DM and FF idiosyncratic errors. For this additional sweep we increased the knot size to 160. For all our states, we also performed the single-move method of Shephard and Pitt (1997) every 4 iterations (of the overall sampler) to ensure that our sampler made local moves with high probability. We ran our sampler for $100,000$ iterations.

The results for the three parameters of the factor f and the four unrestricted elements of β are given in Table 3. The corresponding plots are given in Figure 3. As for the univariate analysis the plots of the samples have been thinned out, only displaying every 100^{th} iteration. The correlograms are calculated using the entire sample. It is clear that our MCMC method is reasonably efficient as the correlograms for the elements of β (from unlikely initial values) become negligible before lags of 1000 in each case. Similarly, the correlogram for the factor parameters dies down rapidly. Given the multivariate and high time dimension of our model

Table 2. *Parameters for idiosyncratic multivariate SV processes. Summaries of Figure 4, 100000 replications of the multi-move sampler, using 80 stochastic knots (discarding first 1000). Ineff are the integrated autocorrelation estimates. M-C S.E. denotes Monte Carlo standard error, using 2000 lags for all parameters except σ where it is 1000.*

	Mean	M-C S.E.	Ineff	Covariance & *Correlation* of Posterior		
			British Pound, ω_1			
$\sigma\|y$	0.3508	0.000143	8.6	0.00245	*-0.019*	*-0.089*
$\sigma_\eta\|y$	0.3369	0.00169	312	-0.0000286	0.000921	*-0.815*
$\phi\|y$	0.9358	0.000541	224	-0.0000506	-0.000284	0.000132
			German Deutschemark, ω_2			
$\sigma\|y$	0.0666	0.000178	53	0.00555	*0.679*	*-0.907*
$\sigma_\eta\|y$	0.1248	0.00164	563	0.00154	0.000926	*-0.800*
$\phi\|y$	0.9907	0.000141	238	-0.000497	-0.000179	0.0000541
			Japanese Yen, ω_3			
$\sigma\|y$	0.4083	0.000162	14	0.00192	*-0.164*	*0.167*
$\sigma_\eta\|y$	0.3840	0.00198	275	-0.000276	0.00147	*-0.866*
$\phi\|y$	0.8988	0.000810	210	0.000130	-0.000587	0.000318
			Swiss Franc, ω_4			
$\sigma\|y$	0.2490	0.000130	22	0.00331	*-0.000418*	*0.041*
$\sigma_\eta\|y$	0.3342	0.00210	449	-0.000000760	0.000998	*-0.838*
$\phi\|y$	0.9180	0.000858	341	-0.0000344	-0.000390	0.000216
			French Franc, ω_5			
$\sigma\|y$	0.0848	0.000217	128	0.00527	*-0.563*	*0.0952*
$\sigma_\eta\|y$	0.7479	0.00441	564	-0.00291	0.00510	*-0.656*
$\phi\|y$	0.9075	0.000923	421	0.0000920	-0.000672	0.000206

Table 3. *Factor parameters and elements of β. Summaries of Figure 3, 100,000 replications of the multi-move sampler, using 80 stochastic knots (discarding first 1000). M-C S.E. denotes Monte Carlo standard error, computed using 1000 lags.*

		Mean	M-C S.E.	Ineff	Covariance & *Correlation* of Posterior			
				Factor parameters				
	$\sigma\|y$	0.5045	0.000347	20	0.000584	*-0.0723*	*0.0591*	
	$\sigma_\eta\|y$	0.1674	0.000850	221	-0.0000313	0.000320	*-0.756*	
	$\phi\|y$	0.9696	0.000248	131	0.00000971	-0.0000919	0.0000462	
				Beta elements				
DM	$\beta_2\|y$	1.240	0.000780	359	0.000166	*0.455*	*0.851*	*0.971*
Yen	$\beta_3\|y$	0.710	0.000503	97	0.0000937	0.000255	*0.400*	*0.452*
SF	$\beta_4\|y$	1.298	0.000813	275	0.000168	0.0000979	0.000235	*0.851*
FF	$\beta_5\|y$	1.190	0.000761	364	0.000156	0.0000903	0.000162	0.000156

this is reassuring, particularly as the factor parameters and β may well be regarded as the most interesting part of the model.

The posterior covariance matrix for the parameters $\{\sigma, \sigma_\eta, \phi\}$ of the factor f in Table 3 is similar in magnitude to that for the univariate parameters for the P and DM of Table 1. The posterior correlation between these parameters is also similar. As we would expect for the factor parameters, σ is not highly correlated with ϕ or σ_η. This is due to our centred parameterisation. The elements of β are all tightly estimated and are positively correlated. β_2, β_3 and β_5 (representing the factor of DM, SF and FF respectively) are all particularly strongly correlated. This is not surprising as the correlation between the returns is reflected in the posterior correlation of the factor weights. However, it emphases the importance of sampling all the elements of each column of β (in this case there is only one) simultaneously.

Table 2 shows the results of the MCMC analysis for each of the 5 idiosyncratic errors. The samples (thinned out) together with the correlograms for the three parameters associated with each idiosyncratic error are given in Figure 4. The correlograms do not die down as quickly as for the factor parameters but still indicate reasonable efficiency in our MCMC method. The correlograms for the parameters of the DM error are the slowest to decay. Apart from the DM error, the parameters of the remaining errors indicate far less persistence than the factor component of Table 3 and than their univariate counterparts of Table 1. For all but the DM, the factor part of our model is isolating the persistent volatility movements whilst the idiosyncratic error terms pick up the more temporal volatility features.

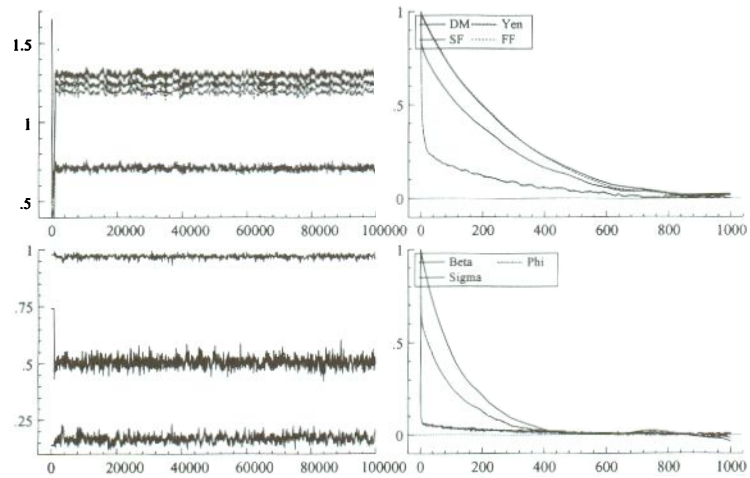

Figure 3. *Elements of β and factor parameters. The simulated parameters (100000 iterations) shown on left; 4 unrestricted elements of β (top) and factor parameters (bottom) together with corresponding acfs on right. See Table 3.*

The relative importance of the factor for each of the returns considered can be shown by considering the unconditional variance estimated from the model. This may be compared with the corresponding sample variance given in Section 1.3. The Bayesian mean of the unconditional variance from our model is

$$\Sigma = E\left\{\beta\beta'\sigma_f^2 + diag(\sigma_{\omega_1}^2, ..., \sigma_{\omega_N}^2)\right\} = \Sigma_f + \Sigma_\omega,$$

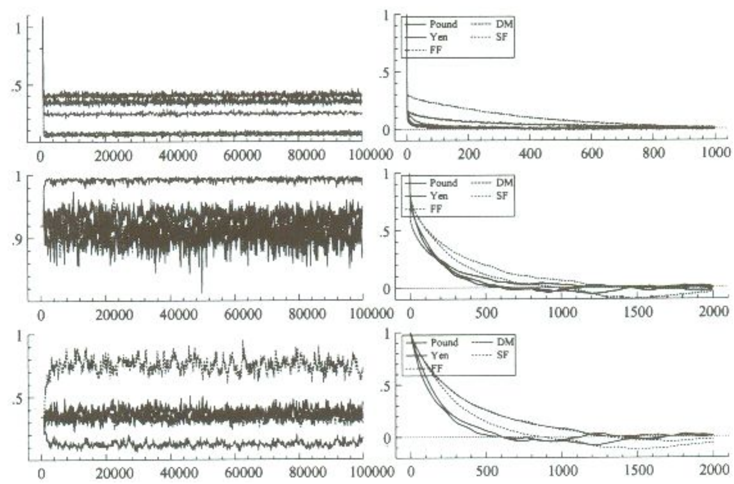

Figure 4. *Parameters for ω. The simulated parameters (100000 iterations) shown on left; σ (top), ϕ and σ_η (bottom) together with corresponding acfs on right. See Table 2.*

where $E(.)$ is with respect to the posterior density and

$$\sigma_f^2 = \exp\left\{\mu^f + \frac{1}{2}\frac{\sigma_\eta^{f2}}{(1-\phi^{f2})}\right\} \quad \text{and} \quad \sigma_{\omega_i}^2 = \exp\left\{\mu^{\omega_i} + \frac{1}{2}\frac{\sigma_\eta^{\omega_i 2}}{(1-\phi^{\omega_i 2})}\right\}.$$

Hence we can easily unbiasedly estimate using our MCMC samples. We estimate Σ_f and Σ_ω as

$$\Sigma_f = \begin{pmatrix} 0.2570 & \cdots & & \cdots & \cdots \\ 0.3186 & 0.3949 & & \cdots & \cdots \\ 0.1824 & 0.2261 & 0.1296 & & \cdots \\ 0.3335 & 0.4134 & 0.2367 & 0.4328 & \\ 0.3059 & 0.3792 & 0.2172 & 0.3970 & 0.3642 \end{pmatrix} \quad \text{and} \quad \Sigma_\omega = diag\begin{pmatrix} 0.1913 \\ 0.0058 \\ 0.2414 \\ 0.0846 \\ 0.0310 \end{pmatrix}$$

It is clear that the unconditional variances associated with the idiosyncratic terms are generally small relative to the corresponding marginals of the factor part. This is particularly the case for the DM, SF and FF where the contribution of the idiosycratic is term tiny. This interpretation suggests the factor is basically a DM, SF, FF effect, while the P and Yen are influenced but not wholly determined by this factor.

The addition of these two matrices gives us (with the correlations in italics),

$$\Sigma = \begin{pmatrix} 0.4481 & 0.7514 & 0.4474 & 0.6920 & 0.7266 \\ 0.3185 & 0.4008 & 0.5868 & 0.9075 & 0.9530 \\ 0.1824 & 0.2263 & 0.3711 & 0.5404 & 0.5674 \\ 0.3334 & 0.4135 & 0.2369 & 0.5179 & 0.8775 \\ 0.3058 & 0.3793 & 0.2173 & 0.3970 & 0.3952 \end{pmatrix}.$$

The corresponding sample variance and correlations for the data (4018 returns) is given below,

$$S = \begin{pmatrix} 0.4781 & 0.7688 & 0.5275 & 0.7069 & 0.7445 \\ 0.3752 & 0.4981 & 0.6604 & 0.8925 & 0.9434 \\ 0.2360 & 0.3016 & 0.4188 & 0.6375 & 0.6254 \\ 0.3950 & 0.5090 & 0.3333 & 0.6529 & 0.8450 \\ 0.3574 & 0.4622 & 0.2810 & 0.4740 & 0.4820 \end{pmatrix}.$$

The two matrices are similar. However, the diagonal elements from our model are smaller in each case than those of the sample variance. This may indicate that there is more volatility in the data than the model accounts for (for instance, heavy tailed measurement densities). The unconditional correlations are very similar to those of the data. It is therefore clear our parsimonious model is nevertheless rich enough to model the unconditional properties of the model. The factor part of our model accounts for 57%, 99%, 35%, 84% and 92% of the marginal variance of the P, DM, Yen, SF and FF respectively. This is what we might expect as the factor appears to explain European movements whereas the Yen may move more independently against the USD, being influenced by other factors (which also affect other Asian countries).

Table 4. *Posterior means of the factor parameters and idiosyncratic terms for 2 factor model. 100,000 replications of the multi-move sampler, using 80 stochastic knots (discarding first 2000).*

Idiosyncratic and factor means							
	P	DM	Yen	SF	FF	FACT1	FACT2
$\sigma\|y$	0.2067	0.06503	0.4087	0.2502	0.08845	0.5068	0.2386
$\sigma_\eta\|y$	0.5994	0.1242	0.3844	0.3316	0.7363	0.1478	0.2013
$\phi\|y$	0.9283	0.9902	0.8994	0.9188	0.9077	0.9839	0.9824
Beta results							
	Column 1	Mean	Variance	Column 2	Mean	Variance	
P	$\beta_{11}\|y$	1	0	$\beta_{12}\|y$	0	0	
DM	$\beta_{21}\|y$	1.055	0.000323	$\beta_{22}\|y$	1	0	
Yen	$\beta_{31}\|y$	0.617	0.000253	$\beta_{32}\|y$	0.527	0.00145	
SF	$\beta_{41}\|y$	1.108	0.000339	$\beta_{42}\|y$	1.020	0.000521	
FF	$\beta_{51}\|y$	1.0149	0.000301	$\beta_{52}\|y$	0.948	0.0000565	

We estimated a two factor model on the same dataset. The results of this analysis are summarized in Table 4. The factor and idiosyncratic components of the unconditional variance of y_t for the two factor model are given below. It is clear that the results do not alter very much with the inclusion of an additional factor. This suggests a certain robustness in these models generally.

$$\Sigma_f = \begin{pmatrix} 0.2605 & \cdots & & & \cdots \\ 0.2747 & 0.3490 & & & \cdots \\ 0.1607 & 0.2005 & 0.1155 & & \cdots \\ 0.2885 & 0.3647 & 0.2096 & 0.3812 & \\ 0.2643 & 0.3350 & 0.1925 & 0.3502 & 0.3217 \end{pmatrix} \quad \text{and} \quad \Sigma_\omega = diag \begin{pmatrix} 0.1565 \\ 0.0051 \\ 0.2416 \\ 0.0850 \\ 0.0315 \end{pmatrix}$$

4.2. *One-Step-Ahead Testing*

We are going to use filtering to examine the model residuals and to assessing the overall fit. To motivate and simplify our discussion we shall delay the outline of our filtering method until Section 4.4. We shall regard our time-invariant parameters θ as fixed and known for the moment. We shall assume we can evaluate and simulate from the density $f(y_t|\alpha_t;\theta)$ for $t = 1, ..., n$. These assumptions clearly hold for our FSV model for which $\alpha_t = (\alpha^{\omega\prime}, \alpha_t^{f\prime})'$. Let us also assume that we can easily obtain samples from $f(\alpha_{t+1}|\mathcal{F}_t;\theta)$, the prediction density, where as usual $\mathcal{F}_t = (y_1, ..., y_t)'$. This last assumption results from our filtering method of Section 4.4. It is clear that with these assumptions in place a whole army of residuals can be constructed.

However, we focus only on four for assessing overall model fit, outliers and observations which have substantial influence on the fitted model.
Log likelihood. $l_{t+1} = \log f(y_{t+1}|\mathcal{F}_t;\theta)$. We have

$$f(y_{t+1}|\mathcal{F}_t;\theta) = \int f(y_{t+1}|\alpha_{t+1};\theta)dF(\alpha_{t+1}|\mathcal{F}_t;\theta).$$

Hence we use Monte Carlo integration as

$$\widehat{f(y_{t+1}|\mathcal{F}_t;\theta)} = \frac{1}{M}\sum_{i=1}^{M} f(y_{t+1}|\alpha_{t+1}^i;\theta),$$

where $\alpha_{t+1}^i \sim f(\alpha_{t+1}|\mathcal{F}_t;\theta)$. Since we can evaluate the density $f(y_{t+1}|\mathcal{F}_t;\theta)$ we can calculate the likelihood of the model M, say, at the Bayesian mean $\overline{\theta}_M$ via the prediction decomposition. Evaluating the likelihood allows model comparison.

Normalised log likelihood. l_t^n. We take S (100 are used in the next section) samples of z^j, $j = 1,..,S$, where $z^j \sim f(y_{t+1}|\mathcal{F}_t;\theta)$ evaluating for each sample l_{t+1}^j using the above method. We then construct μ_{t+1}^l and σ_{t+1}^l as the sample mean and standard deviation of these quantities, respectively. The normalised log likelihood at time t is therefore computed as $l_{t+1}^n = \left(l_{t+1} - \mu_{t+1}^l\right)/\sigma_{t+1}^l$. If the model (and parameters) are correct then this statistic should have mean 0 and variance 1. Large negative values of course, indicate that an observation is less likely than we would expect. Under the WLLN we expect $\sum_{t=1}^{T} l_t^n/T \to 0$ as $T \to \infty$.

Uniform residuals. $u_{t+1} = F(l_{t+1}|\mathcal{F}_t;\theta)$. This quantity is estimated as $\widehat{u}_{t+1} = \widehat{F}(l_{t+1}) = \frac{1}{S}\sum_{j=1}^{S} I(l_{t+1}^j < l_{t+1})$ where the l_{t+1}^j's are constructed as above. If we assume that we know the parameter vector θ, then under the null hypothesis that we have the correct model $\widehat{u}_{t+1} \sim UID(0,1)$. In addition, the reflected residuals (Kim, Shephard and Chib (1998)) $2\left|\widehat{u}_t - 0.5\right| \sim UID(0,1)$, $t = 1,...,n$. The former has been used by, amongst others, Smith (1985) and Shephard (1994b) to see if their fitted models were well calibrated.

Distance measure d_t. We can compute $\Sigma_{t+1} = Var(y_{t+1}|\mathcal{F}_t;\theta) \doteq \frac{1}{M}\sum_{i=1}^{M} Var(y_{t+1}|\alpha_{t+1}^i)$ where $\alpha_{t+1}^i \sim \alpha_{t+1}|\mathcal{F}_t;\theta$. Then at each time step t we compute $d_t = y_t'\Sigma_t^{-1}y_t = a_t'a_t$, where $a_t = \Sigma_t^{-1/2}y_t$ consisting of N independent elements each with mean 0 and variance 1. It is therefore the case, if the model and parameters are correct, that $d_t \overset{iid}{\sim} \chi_N^2$, so $\sum_{t=1}^{n} d_t \sim \chi_{nN}^2$.

We can now identify outlying data and can also form overall tests of fit easily. The difficulty is that in practise we do not know θ but the posterior density becomes tighter around the true value of course. We therefore simply use $\overline{\theta}$, the Bayesian mean, in our calculations.

4.3. *One-Step-Ahead Testing and Filtering Results*

We ran the auxiliary filter, see Section 4.4, over the entire data setting $M = 10,000$. For evaluating u_t and l_t^n, S the number of simulations from the prediction density for y_t, is set to 100 at each time step.

In Figure 5, the residuals together with the corresponding average returns over the period of interest are plotted against date. The two large values of d_t, occur at around the end of 1981 and the beginning of 1983. These two outliers appear in the plots of l_t and l_t^n. Whilst the abnormal returns at the beginning 1983 are clear from the plot of returns, the outlier at the end of 1981 is not. In addition it appears, from the plot of l_t and l_t^n that there is an unlikely return around the

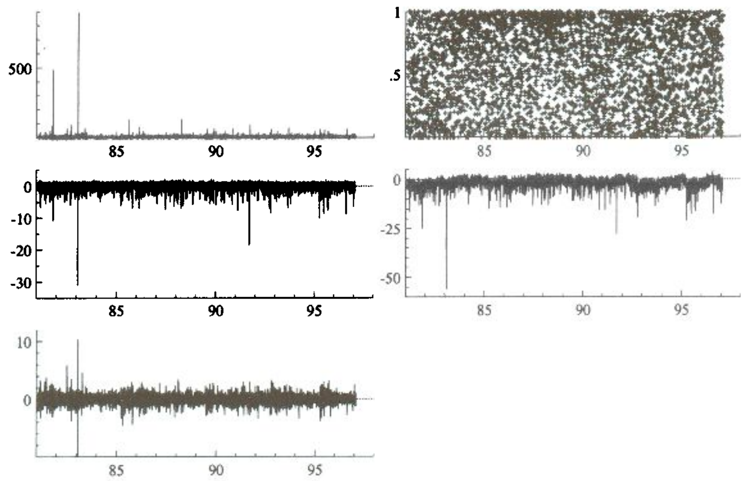

Figure 5. *One step ahead residuals against date. Top row: $d(t)$ (left), $u(t)$ (right). Second row: standardised $l^n(t)$ (left), $l(t)$ (right). Last row: returns for five series.*

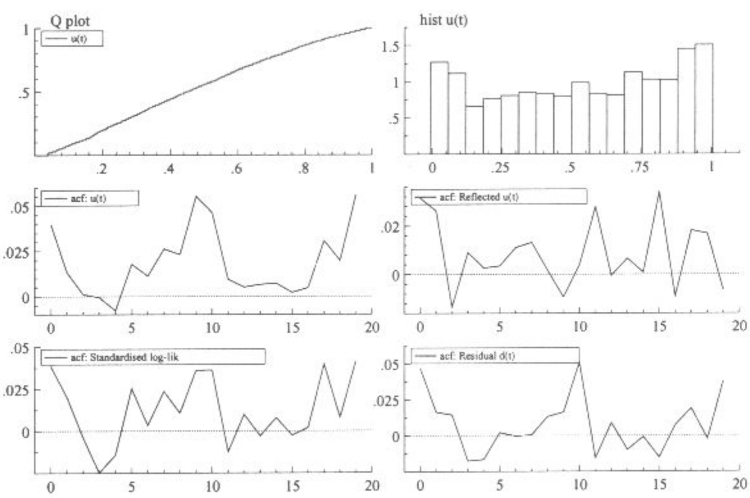

Figure 6. *Residual analysis. Top row: quantile plot for u_t (right), histogram for u_t (left). Second row: correlogram for residual u_t (left), for reflected u_t (right). Bottom row: correlogram for residual $(l_t)^n$ (left), for distance measure d_t (right).*

middle of 1991. Again this is not obvious simply by examining the returns. The log-likelihood was computed as $-6,206.9$ for the overall single factor model computing using the posterior mean of the parameters.

From each of the univariate $\mathcal{ISV}$ models estimated we obtain log-likelihoods of $-3,863$ (P), $-4,042$ (DM), $-3,663$ (Yen), $-4,539$ (SF) and $-4,050$ (FF). The overall log-likelihood for all the series is $-20,157$. Clearly the log-likelihood is far smaller than for our FSV model since the correlation between returns is not accounted for by this model. Further, the mean of resulting d_t was 5.1911, indicating that the distance is not much greater than we would expect were the model to be operating. The mean of the l_t^n is 0.00255, close to zero (not significantly different) as we would expect under the model. The variance of l_t^n is 1.6678, larger than we would expect indicating that there are a lot of either very likely or very unlikely observations (but less in between) than expected.

From Figure 6 it is clear that the residuals u_t are not quite uniform but are overdispersed. This again suggests using a heavy tailed SV model. The autocorrelations of all the residuals displayed are not significantly different from zero. This is particularly reassuring as it indicates we have accounted for the persistence in volatility.

The filter we apply delivers samples from $\alpha_t|\mathcal{F}_t$ which we can compare to the draws from the MCMC smoothing algorithm $\alpha_t|\mathcal{F}_n$. The average (over time) of the difference is -0.000487 whilst its variance is 0.0665. For Figure 7, we have transformed the samples to give the smoothed mean and filtered mean factor standard deviation. It is clear that the two mean standard deviations move together, the filtered mean delivering a coarser plot than the smoothed mean. The difference between the two is also displayed together, and varies around 0, as we would expect. Finally the filtered mean standard deviations for the idiosyncratic terms are shown in Figure 8.

4.4. *A Simulation Filter*

The methodology outlined above presupposes that we can simulate from the one-step ahead density $f(\alpha_{t+1}|\mathcal{F}_t;\theta)$. We employ the auxiliary sampling-importance resampling (ASIR) particle filtering method of Pitt and Shephard (1997) to carry out this non-trivial filtering task. We use the notation $f(\alpha_{t+1}|\alpha_t)$ to denote the evolution of the unobserved log-volatilities over time.

The particle filter has the following basic structure. The density of $\alpha_t|\mathcal{F}_t = (y_1, ..., y_t)'$ is approximated by a distribution with discrete support at the points $\alpha_t^1, ..., \alpha_t^M$. Then we try to produce a sample of size M from

$$\widehat{f}(\alpha_{t+1}|\mathcal{F}_{t+1}) \propto f(y_{t+1}|\alpha_{t+1}) \sum_{k=1}^{M} f(\alpha_{t+1}|\alpha_t^k). \tag{3}$$

This provides the update step of the ASIR filter. This is carried out by sampling k^j with probability proportional to $f(y_{t+1}|\mu_{t+1}^k)$, where $\mu_{t+1}^k = E(\alpha_{t+1}|\alpha_t^k)$, and then drawing from $\alpha_{t+1}^j \sim \alpha_{t+1}|\alpha_t^{k^j}$. This is carried out R times. The resulting population of particles are given weights proportional to

$$w_j = \frac{f(y_{t+1}|\alpha_{t+1}^j)}{f(y_{t+1}|\mu_{t+1}^{k^j})}, \qquad \pi_j = \frac{w_j}{\sum_{i=1}^R w_i}, \qquad j = 1, ..., R.$$

We resample this population with probabilities $\{\pi_j\}$ to produce a sample of size M. In this way we update at each time step. The efficiency of this method is analysed in Pitt and Shephard (1997).

In practice when we applied the auxiliary SIR particle filter in this paper we have taken $M = 10,000$. At each time step we set $R^* = 200$ and went forward a single time step computing our resample probabilities w. We then went back and set the value of R (the

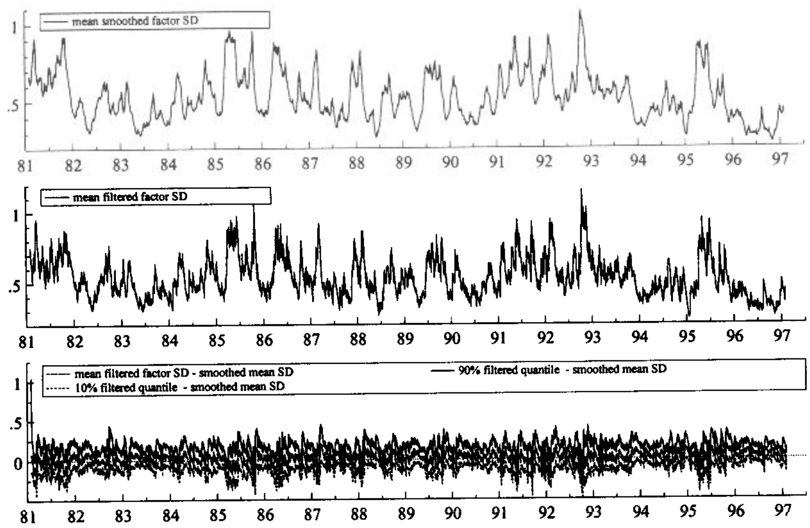

Figure 7. *Factor log volatilities. Top row: smoothed mean of factor standard deviation. Second row: filtered mean of factor standard deviation. Last row: filtered mean, filtered 90% quantile, filtered 10% quantle - smoothed mean.*

number of prior sample) to be $\min(10 \times M, \text{INEFF} \times M)$ where at each step we computed the INEFF$=1/\{1 + Var(R^{*}w)\}$, using an approximate result of Liu (1996).

5. OPEN ISSUES

Risk premium. The use of a factor structure for our model suggests that we should add a risk premium to the mean of the returns. In a simple one factor model the structure would be that

$$y_t = r\iota + \beta Cov(f_t|\alpha_t)\pi + \beta f_t + \omega_t,$$

where r is a riskless interest rate, π is some (very small) unknown parameter vector. Such a model predicts higher expected returns in periods of high volatility and is in keeping with the APT.

The presence of quite a sophisticated mean term in the returns model does not change our MCMC calculations very much. As the information is quite small we propose ignoring it in our proposal density and adding the implied density from the above residual to the Metropolis acceptance rate.

Leverage effects. Unlike exchange rate data, stock price falls are often associated with increases in volatility (Nelson, 1991). In the context of SV models this can be achieved by allowing ϵ_t and η_t to be negatively correlated. The presence of this correlation does not make the multivariate model anymore complicated, but it does mean the analysis of the univariate models has to become slightly more sophisticated. However, the method of Shephard and Pitt (1997) goes through in that case.

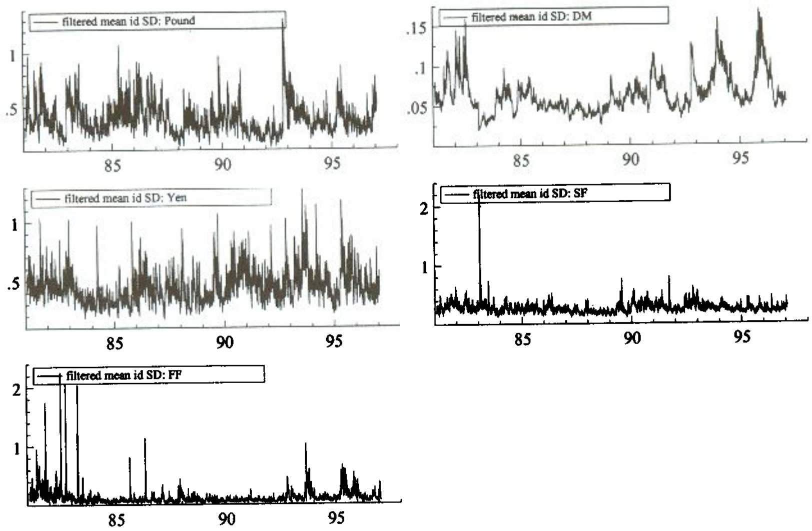

Figure 8. *Idiosyncratic volatilities. Filtered mean standard deviations for Pound, DM, Yen, SF and FF.*

More general dynamics. In this paper we have assumed a very simple AR(1) dynamic structures for the volatility process. However, our analysis would allow these processes to be generalized to be any Gaussian process.

Heavy tailed densities. An empirically important generalisation of the model is to allow for heavier tails. In particular each of the basic SV models can be generalised to allow

$$\varepsilon_t = \frac{\varsigma_t}{\sqrt{\kappa_t}}\sqrt{p-2}, \quad \text{where} \quad \varsigma_t \sim NID(0,1) \quad \text{and} \quad \kappa_t \sim \mathcal{IG}\left(\frac{p}{2}, \frac{1}{2}\right).$$

This has generalised $\{\varepsilon_t\}$ from being *iid* normal to scaled *iid* Student's t with p degrees of freedom but still a unit variance. This style of model also requires us to specify a proper prior for p constrained so that $p > 2$.

6. CONCLUSION

The factor model attempts to model both the correlation and the time varying variances of returns. It is an appealing model from an economic perspective, its roots being in finance theory. Simple multivariate factor models for SV processes have been suggested, but not applied, by Jacquier, Polson and Rossi (1995) and extended into an empirically reasonable form by Kim, Shephard and Chib (1998). As the number of asset returns considered becomes large, our preferred factor SV model allows the possibility of a fairly parsimonious model with a small number of factors. The residuals for the one factor model suggest that the volatility process of the returns considered is captured by the model.

There is a great deal of work to be carried out in this area. Applying these methods to very large datasets, with many tens or hundreds of assets, is theoretically possible but computationally challenging. Using the fitted models in terms of testing APT and carrying out optimal portfolio choice should be interesting. Further, exploiting the models in order to accurately measure VaR is a useful topic.

ACKNOWLEDGEMENTS

We thank Enrique Sentana and the conference participants for their comments on the first draft of our paper and the ESRC for their financial help on this project.

REFERENCES

Aguilar, O. and West, M. (1998). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Tech. Rep.* **98-03**, ISDS, Duke University..

Bartholomew, D. J. (1987). *Latent Variable Models and Factor Analysis*. Oxford: Oxford University Press

Bollerslev, T., Engle, R.F. and Nelson, D. B. (1994). ARCH models. *The Handbook of Econometrics* **4** (D. McFadden, ed.). Amsterdam: North-Holland, 2959–3038.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Clark, P. K. (1973). A subordinated stochastic process model with variance for speculative prices. *Econometrica* **41**, 135–156.

Dave, R. D. and Stahl, G. (1997). On the accuracy of VaR estimates based on the variance-covariance approach. *Tech. Rep.*, Olsen and Associates, Zurich.

Diebold, F. X. and Nerlove, M. (1998). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *J. Applied Econometrics* **4**, 1–21.

Engle, R. F. and Kroner, F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* **11**, 122–150.

Geweke, J. F. and Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *Rev. Financial Studies* **9**, 557–587.

Ghysels, E., Harvey, A. C. and Renault, E. (1996). Stochastic volatility. *Statistical Methods in Finance* (C. R. Rao and Maddala, G. S.). Amsterdam: North-Holland, 119–191.

Harvey, A. C., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models. *Rev. Economic Studies* **61**, 247–264.

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

Ingersoll, J. E. (1987). *Theory of Financial Decision Making*. Maryland: RowmanLittlefield.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *J. Business and Economic Statist.* **12**, 371–417, (with discussion).

Jacquier, E., Polson, N. G. and Rossi, P. E. (1995). Models and prior distributions for multivariate stochastic volatility. *Tech. Rep.*, Graduate School of Business, University of Chicago..

Kim, S., Shepard, N. and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Rev. Economic Studies* **65**, 361–393.

King, M., Sentana, E. and Wadhwani, S. (1994). Volatility and links between national stock markets. *Econometrica* **62**, 901–933.

Liu, J. (1996). Metropolized independent sampling with comparison to rejection sampling and importance sampling. *Statistics and Computing* **6**, 113–119.

McCulloch, R. and Rossi, P. E. (1991). A Bayesian approach to testing the arbitrage pricing theory. *J. Econometrics* **49**, 141–168.

McCurdy, T. H. and Morgan, I. G. (1991). Test for a systematic risk component in deviations from uncovered interest rate parity. *Rev. Economic Studies* **58**, 587–602.

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

Pitt, M. K. and Shephard, N. (1997). Filtering via simulation based on auxiliary particle filters. *J. Amer. Statist. Assoc.* (to appear).

Pitt, M. K. and Shephard, N. (1998). Analytic convergence rates and parameterisation issues for the Gibbs sampler applied to state space models. *J. Time Series Analysis* **19**. (to appear).

Quintana, J. M. (1992). Optimal portfolios of forward currency contracts. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press.

Quintana, J. M. and West, M. (1987). An analysis of international exchange rates using multivariate DLM's. *Statistican* **36**, 275–281.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *J. Economic Theory* **13**, 341–360.

Sentana, E. and Fiorentini, G. (1997). Identification, estimation and testing of multivariate conditionally heteroskedastic factor models. *Tech. Rep.* **9709**, CEMFI.

Shephard, N. (1993). Fitting non-linear time series models, with applications to stochastic variance models. *J. Applied Econometrics* **8**, 135–152.

Shephard, N. (1994a). Local scale model: state space alternative to integrated GARCH processes. *J. Econometrics* **60**, 181–202.

Shephard, N. (1994b). Partial non-Gaussian state space. *Biometrika* **81**. 115–131.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. *Time Series Models in Econometrics, Finance and Other Fields* (D. R. Cox, Hinkley, D. V. and Barndorff-Nielson, O. E., eds.). London: Chapman and Hall, 1–67.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**, 653–667.

Smith, J. Q. (1985). Diagnostic checks of non-standard time series models. *J. Forecasting* **4**, 283–291.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes —a study of the daily sugar prices 1961–1975. *Time Series Analysis: Theory and Practice, 1* (O. D. Anderson, ed.). Amsterdam: North-Holland.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica* **65**, 59–73.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models* (2nd ed.). New York: Springer-Verlag.

DISCUSSION

SIDDHARTHA CHIB *(Washington University at St. Louis, USA)*

It is a pleasure to congratulate the authors on an authoritative and wide ranging analysis of a class of factor stochastic volatility models for multivariate high frequency data. On the basis of this paper one can expect that the huge potential of these hitherto intractable models will now be realized.

The model discussed in this paper has two key components. One component is concerned with the modeling of the contemporaneous correlation amongst the N time series. This correlation is achieved by writing the observation model as $y_t = \beta f_t + \omega_t$, where f_t is a K vector of time varying factors and ω_t is an error vector. The second component is concerned with the modeling of the one-step ahead (*i.e.*, conditional) variance of f_t and ω_t. Letting α_t denote the state vector at time t one assumes that independently

$$f_{it} \sim N(0, \exp(\alpha_{it}^f))\,;\; \omega_{jt} \sim N(0, \exp(\alpha_{jt}^\omega))$$

where the (unobserved) conditional variances are allowed to evolve according to the stationary stochastic volatility processes

$$\alpha_{it}^f = \mu_i^f + \phi_i^f(\alpha_{it-1}^f - \mu_i^f) + \sigma_{\eta i}^f \eta_{it}^f$$

$$\alpha_{jt}^\omega = \mu_j^\omega + \phi_j^\omega(\alpha_{jt-1}^\omega - \mu_j) + \sigma_{\eta j}^\omega \eta_{jt}^\omega,$$

where the η's are standard normal.

The first contribution of the paper is in the development of a practical MCMC scheme for estimating the model. The paper deals with an example involving five time series with at most two factors. The authors in their usual fine style report all the key steps in the algorithm and summarize the MCMC output in considerable detail. The output analysis, in particular the autocorrelation plots and the inefficiency factors, reveals that the algorithm requires retuning. First, one can adopt a different scheme for sampling the conditional variances $\{\alpha_{it}^f\}$ and $\{\alpha_{jt}^\omega\}$.

One question is why one should not use the method of Kim, Shephard and Chib (1998) which can be applied easily to the class of models discussed in this paper. Second, Pitt and Shephard could have considered alternative blocking schemes, specifically in the sampling of β and $\{f_t\}$. Some initial work has revealed that dramatic reductions in the serial correlation are possible by sampling β marginalized over $\{f_t\}$.

The second contribution of the paper is the nice work on model diagnostics and model fit issues. This is an important area in general and the ideas described here are likely to prove useful. There is, however, the open question of determining the number of factors in the model. This issue needs to be addressed in future work. It is possible that the computation of the marginal likelihood of the model may be feasible using one of the methods that have appeared in recent years (for example, see Chib (1995)). Typically in finance applications a considerable amount of data is available and some of that data can be used to build the prior distribution. Marginal likelihoods also penalize complexity which is important in evaluating large multivariate models.

DANI GAMERMAN *(Instituto de Matemática - UFRJ, Brazil)* and
AJAX R. B. MOREIRA *(IPEA-RJ, Brazil)*

Introduction. This is a very nice and well written paper that discusses many interesting points concerning modeling of covariance structure for multivariate time series. In addition to the novel models proposed, it goes into details about the MCMC implementation, applies the results to multivariate exchange rate data and discusses some model fit issues. The main thrust driving the paper is the application to financial data but the ideas can be usefully applied in other contexts.

This discussion will thus be concentrated on some comments about possible extensions and/or generalizations rather than criticism about the content of the paper. We will therefore structure the discussion on a few general comments about: (i) Relation with GARCH models, (ii) Stationarity, (iii) Hierarchical priors, (iv) Doubly dynamic models, (v) Updating and sampling and (vi) Data analysis

Relation with GARCH models. The basic model used for returns assumes that $y_t \sim N(\mu_t, \sigma_t^2)$ and, usually, $\mu_t = 0$. In stochastic volatility (SV) models, $\log \sigma_t^2 = \phi_0 + \phi_1 \log \sigma_{t-1}^2 + u_t$. Another important class of volatility model is given by the GARCH models where $\sigma_t^2 = \phi_0 + \sum_{i=1}^p \phi_i \sigma_{t-i}^2 + \sum_{j=1}^q \psi_j (y_{t-j} - \mu_{t-j})^2$. We think it is important for a paper presented at a general Bayesian meeting (rather than a specialized finantial time series meeting) that some comparative comments about SV $\times$ GARCH are provided to make a broader audience aware of the choices available in this area. In particular, Ng, Engle and Rothschild (1992) use a similar approach with factor models to multivariate time series but model the volatilities through the GARCH route. They also used a single factor, which they refer to as the market and is observable. More recently, Aguilar and West (1998) used a similar model with unobserved factors and SV.

Stationarity. The transformed beta prior for ϕ ensures stationarity for the log volatilities of factors and error terms. It is important that this assumption is justified if it is to be used. Jacquier, Polson and Rossi (1994) used a different prior in the normal form and did not restrict the range of values for ϕ. This seems to tie in more naturally with hierarchical priors discussed below. In any case, we wonder how crucial the stationarity assumption is. For the analysis reported, in particular, the value of ϕ for the DM seems very close to 1. It would be nice in this case at least to have the posterior histogram.

Hierarchical priors. The analysis of multivariate time series inevitably leads to a profusion of parameters. The factor approach is specifically designed to reduce the parameter dimensionality in a very elegant and potentially meaningful fashion. Even then, many parameters are still left in the models and they may represent similar aspects across the series. A typical example concerns the ϕ's across the idiosyncratic terms. In the paper, a reasonably strong assumption with independent Beta(18,1) priors was made. An alternative prior assumption in line with comments above could take $\phi_i \sim N(\phi, \sigma_\phi^2)$, $i = 1, ..., N$ and the prior can be completed with a (possibly vague) normal prior for ϕ. The same idea could also be used for parameters associated with factors.

Doubly dynamic models. These are models with dynamic structure on mean and variance given in the case of normal observations by

$$\begin{aligned} \text{Observation equation: } & y_t \sim N(\mu_t, \sigma_t^2) \\ \text{Mean link: } & g_1(\mu_t) = \eta_t = X_t'\beta_t \\ \text{Variance link: } & g_2(\sigma_t^2) = \xi_t = Z_t'\gamma_t \\ \text{Mean system equation: } & \beta_t = G_{1t}\beta_{t-1} + w_{1t} \\ \text{Variance system equation: } & \gamma_t = G_{2t}\gamma_{t-1} + w_{2t}. \end{aligned}$$

In SV models, $\mu_t = 0$, $g_2 = \log$ and the system evolution is in AR(1) form with G_{2t} containing unknown hyperparameters.

This idea has been partially used in similar models with GARCH (static) variance evolution by Valle and Migon (1998) and Harvey, Ruiz and Sentana (1992). The first one used a Bayesian approach and solved the required integrals with Gaussian quadrature while the second one used a maximum quasi-likelihood approach. Aguilar et al. (1998) proposed some models with a dynamic structure on the mean and AR(1) variance evolution.

Dynamic modelling allows *both* mean and variance to be described by dynamic components such as: trend (eg in AR(1) form), seasonality (with free form or harmonics), cycles and explanatory variables (lagged or not). As an example consider the presence of seasonal variation on volatilities. In this case, one possible model for this component could be

$$\begin{aligned} \log \sigma_t^2 = \xi_t &= \xi_{Lt} + \xi_{St} \\ \xi_{Lt} &= \mu + \phi\xi_{L,t-1} + u_{Lt} \\ \xi_{St} &= -(\xi_{S,t-1} + \dots + \xi_{S,t-p+1}) + u_{St}, \end{aligned}$$

where p is number of trading days in the week, for daily data. The Brazilian stock market is sometimes affected by the *black thursday* effect where unconfirmed alarmist news are *leaked* to the operators usually on this day of the week causing an excess volatility. This is just an example and many other possibilities for seasonal modelling and for more general model components can be accommodated into this flexible structure.

Updating and sampling. The marginal likelihood for any given model M, given by

$$f(y_1, ..., y_n|M) = \prod_{t=1}^{n} f(y_t|D_{t-1}, M),$$

can be used to assess model fit. The densities however must be estimated.

Consider state parameters $\alpha = (\alpha_1, ..., \alpha_n)$ and hyperparameter θ. Then,

$$\begin{aligned} f(y_t|D_{t-1}, M) &= \int\int f(y_t|\alpha_t, \theta, D_{t-1}, M) f(\alpha_t, \theta|D_{t-1}, M) d\alpha_t d\theta \\ &\doteq \frac{1}{M}\sum_{j=1}^{M} f(y_t|\alpha_t^{(j)}, \theta^{(j)}, D_{t-1}, M) \end{aligned}$$

where $(\alpha_t^{(j)}, \theta^{(j)}) \sim f(\alpha_t, \theta|D_{t-1}, M)$. Once sampling from this distribution is achieved, it would provide means for on-line model updating which can be very useful for real-time applications. Instead, the authors used $(\alpha_t^{(j)}, \hat{\theta})$ where $\alpha_t^{(j)} \sim f(\alpha_t|\hat{\theta}, D_{t-1}, M)$ disregarding uncertainty about θ. It would be nice to extend their SIR scheme for updating α_t's by including θ as well.

Data analysis. It would be nice to be able to compare between models with $K = 1$ factor and $K = 2$ factors. In particular, $f(y_t|D_{t-1}, K = j)$ could be evaluated and displayed.

A nice feature of the similar paper by Aguilar and West (1998) is the consideration of different portfolio allocations. In that paper, comparison in terms of cumulative returns between different allocations and different models are provided. This kind of comparison based on the variable users are most interested on is one of the main concerns in this area. This exercise could be performed here with the filtered distributions, instead of the smoothed ones by Aguilar and West (1998).

Another practical issue concerns modeling major currency devaluations. One possibility is provide by the fat-tailed (eg t-Student) distribution. Another one is the use switching regimes (Hamilton and Susmel, 1994). In the SV context, it would mean taking $\log\sigma_t^2 = \phi_0 + \phi_1 \log\sigma_{t-1}^2 + k_t\delta_t + u_t$ where δ_t could be the indicator of a major devaluation of a given currency, subject to a Markov chain form, and k_t is the size of the volatility jump. This can be tied in with the model suggested by Merton (1990, ch. 9) to cope with jumps in asset prices. His model however does not lead to a multiplicative effect on the volatilities as indicated above.

THOMAS LEONARD *(University of Edinburgh, UK)*

I would like to add my congratulations to the authors for an excellent advance. During the discussion it was also brought to my attention that my multivariate normal model for log-variances (Leonard, 1975) which specifically includes dynamic random walk models simultaneously on the means and log variances, has since and much later been termed "stochastic volatility" by the economists. I am delighted. I am pleased that this paper has now been made even more famous by Steve Fienberg's after dinner speech, where it was cited three times!

When I extended my more general ideas to the matrix logarithms of covariance matrices (Leonard and Hsu, 1992, Chiu 1994, Chiu, Leonard and Tsui, 1996) I found that the dynamic models were similarly flexible. The authors have an alternative generalisation of stochastic volatility, which would certainly be worth comparing with my generalization and which directly addresses the log-variances. However a model of the form

$$\underset{\sim}{A_t} = \underset{\sim}{A_{t-1}} + \underset{\sim}{\varepsilon_t}$$

on the log-covariance matrices, where the upper triangular elements of the symmetric matrices $\underset{\sim}{\varepsilon_t}$ possess independent multivariate normal distributions, provides an obvious multivariate definition of stochastic volatility. Why get more complicated?

REPLY TO THE DISCUSSION

We would like to thank Professors Chib, Gamerman, Moreira and Leonard for their comments on our paper. Here we will respond to their discussion in that order.

Siddhartha Chib argues principally for two points: that a more efficient MCMC sampling scheme can be constructed for this problem and that one could use Bayes factors for determining the number of factors in the multivariate SV model. We completely agree with the latter point and no doubt this will be a prominent feature of our later work on this model. We are currently developing the so called "fully adapted" particle filter for the multivariate SV model which should deliver reliable estimates of $f(y|\theta, M)$, which are smooth in θ. This can then be used as part of the Chib(1995) method as

$$f(y|M) = \frac{f(y|\theta^*, M)f(\theta^*|M)}{f(\theta^*|y, M)},$$

where the denominator is evaluated using a kernel density estimation procedure based around posterior draws of θ and θ^* is usually taken as the posterior mean. However, if a fairly large amount of initial data is used to construct the prior, as Professor Chib suggests at the end of his comment, then it may be that a simple Monte Carlo estimate of

$$f(y|M) = \int f(y|\theta, M)f(\theta|M)d\theta$$

may be efficient, depending upon the relative information content of the prior and posterior for θ.

On the former point, Professor Chib argues that the Kim, Shephard and Chib(1998) (KSC) MCMC method could be used in this context, and properly implemented this method may deliver a sampler which is much faster at converging than the one we develop. We agree that this is an interesting avenue to develop (as KSC themselves point out) and it would be very interesting to perform an in-depth comparison of the KSC sampling method with our own in this context. We do mention here, however, that the KSC method does have the drawback that it does not seem able to easily deal with the generalisation of the SV model to the case where there is leverage (correlation between ε_t and η_t). This is more or less straightforward in our approach. Further, we have been working on refining the approach we advocate in the paper and it does seem that by appropriate modification it is possible to improve the performance of our method quite significantly. We will report on these improvements in the literature shortly.

Professor Gamerman's and Professor Moreira's comments are quite wide ranging and stimulating. Their points cover: GARCH models, stationarity, priors, doubly dynamic models, updating and data analysis. We will take them in order.

Of course there is an extremely large literature on ARCH models which we did not mention due to space limitations — although we pointed the readers towards reviews of the relevant papers. The so called factor ARCH model they mention is important and it is helpful to read Sentana(1998) and the references contained within that paper, for he compares the properties of our type of model structure to the one proposed by Ng, Engle and Rothschild (1992). In addition, we did discuss the relationship between our work and that of Aguilar and West(1998) in our paper.

Professor Gamerman and Professor Moreira argue that we should provide more information about the possibility of unit roots in the posterior. In this paper, as elsewhere, we have imposed stationarity in our prior for ϕ, in each of the state equations. This is the only parameter that we imposed a strong prior upon. We have argued elsewhere that if there is evidence for unit roots in these types of models then this indicates some type of structural break in the process rather than evidence to believe that the log-variance is really a random walk or an explosive autoregression. If a unit root really drives log-variance then, in the long term, we would observe either infinitely large variances or variances of 0. This is logically and empirically inconsistent with financial theory and data respectively. Hence we disagree with the implication of the comment on this point. They also express a wish that the univariate graphical summaries of the persistence parameters be included. Unfortunately, restrictions on space meant that we were unable to do that in this paper.

The point on hierarchical priors is an interesting one and we hope that some other researchers will explore the usefulness of this in practice for these types of models.

We agree that the doubly dynamic models are interesting. We would refer the reader to the papers by Carter and Kohn(1994) and Shephard(1994b) who discuss them in some detail. Particular the second of these references points out the possibility of having SV models and dynamic linear models linked together. Of course from a finance theory viewpoint it would also be useful to allow the time varying mean to depend on the changing covariance matrix — as we point out in the paper.

The discussion about on-line learning for both the parameters and states is important. We simply do not know how to do this at the moment — hence we condition on some estimate of the parameters before carrying out the calculations via a particle filter. This is clearly unsatisfactory. The only work we know which is close to being able to carry out the required calculation is Gerlach, Carter and Kohn(1996). However, we have not tried this method out in practice on our problem.

Finally, we agree that carrying out an asset allocation exercise using filtering would be informative, but we have not done that yet. Also the generalisation to fat tails is important as we mention in the paper. Discrete Markov chains are always interesting, but we have not attempted to fit them in our context.

We would like to thank Professor Leonard for the references to previous work. We have now had the chance of reading his early Technometrics paper which is indeed very impressive given it was written two decades ago. Although one can see connections to SV models with the types of models he was advocating in those days, we think that the discussion was far from explicit. Further, other earlier informal discussions of these types of models exist. An example of this is the highly influential paper on subordination by Clark (1973) which many refer to as the first general paper on SV models. Our reference to the Taylor(1982) paper is the first reference we know of a discrete time SV model written explicitly down for speculative prices in the modern form.

ADDITIONAL REFERENCES IN THE DISCUSSION

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* **81**, 541–53.

Chib, S. (1995), Marginal likelihood from the Gibbs output. *J. Amer. Statist. Assoc.* **90**, 1313–1321.

Chiu, Y. M., Leonard, T. and Tsui, K. W. (1996). The Matrix-Logarithmic Covariance Model. *J. Amer. Statist. Assoc.* **91**, 198–210.

Chiu, Y. M. (1994). Log-Covariance Matrix Model. Ph.D. Thesis, University of Wisconsin - Madison.

Clark, P. K. (1973). A subordinated stochastic process model with fixed variance for speculative prices. *Econometrica* **41**, 135–156.

Gerlach, R., Carter, C. K. and Kohn, R. (1996). Diagnostics for time series analysis. *Tech. Rep.*, Australian Graduate School of Management, University of New South Wales.

Hamilton, J. D. and Susmel, R. (1994). Autoregressive conditional hetoroscedasticity and changes of regime. *Journal of Econometrics* **64**, 307–333.

Harvey, A. C., Ruiz, E. and Sentana, E. (1992). Unobserved component time series models with ARCH disturbances. *Journal of Econometrics* **52**, 129–158.

Leonard, T. (1975). A Bayesian approach to the linear model with unequal variances. *Technometrics* **17**, 95–102.

Leonard, T. and Hsu, J. S. J. (1992). Bayesian Inference for a Covariance Matrix. *The Annals of Statistics* **20**, 1669–1696.

Merton, R. C. (1990). *Continuous Time Finance*. Basil Blackwell: Oxford.

Ng, V., Engle, R. F. and Rothshild, M. (1992). A multi-dynamic-factor model for stock returns. *Journal of Econometrics* **52**, 245–267.

Sentona, E. (1998). The relation between conditionally heteroskedastic factor models and factor GARCH models. *Econometrics Journal* **1**, 1–9.

Valle, C. A. and Migon, H. S. (1998). Bayesian analysis of dynamic GARCH models. *Tech. Rep.*, Statistical Laboratory, UFRJ. (in preparation).