Inference for Adaptive Time Series Models: Stochastic Volatility and Conditionally Gaussian State Space Form

Charles S. Bos and Neil Shephard*
Tinbergen Institute and Vrije Universiteit Amsterdam
Nuffield College, University of Oxford

December 2004

Abstract

In this paper we model the Gaussian errors in the standard Gaussian, linear state space model as a stochastic volatility processes. We show that conventional MCMC algorithms for this class of models are ineffective, but that the problem can be alleviated by reparameterising the model. Instead of sampling the unobserved variance series directly, we sample in the space of the disturbances, which proves to lower correlation in the sampler and thus increases the quality of the Markov chain.

Using our reparameterised MCMC sampler, it is possible to estimate an unobserved factor model for exchange rates between a group of n countries. The underlying $n+1$ country-specific currency strength factors and the $n+1$ currency volatility factors can be extracted using the new methodology. With the factors, a more detailed image of the events around the 1992 EMS crisis is obtained.

We assess the fit of competitive models on the panels of exchange rates with an effective particle filter, and find that indeed the factor model is strongly preferred by the data.

Keywords: Markov chain Monte Carlo, particle filter, state space form, stochastic volatility.

1 Introduction

1.1 Overview

This paper shows how to statistically handle a class of conditionally Gaussian unobserved component time series models whose disturbances follow stochastic volatility (SV) processes. Unconditionally, this delivers a potentially highly non-linear model whose forecasts are adaptive through time, changing the level of optimal smoothing to locally match the properties of the data.

We will claim that standard methods for carrying out the computations required for this model class, which are based on Markov chain Monte Carlo (MCMC), can be poor in situations encountered in practise. We show that a simple reparameterisation overcomes this difficulty delivering reliable methods for inference, and investigate in what situations the reformulation improves the mixing of the sampling chain. These are the main contributions of this paper. We will illustrate

*Correspondence to: Charles Bos, Department of Econometrics and O.R., Vrije Universiteit Amsterdam, De Boelelaan 1105, NL-1081 HV Amsterdam, The Netherlands. E-mail: cbos@feweb.vu.nl. This research was performed while the first author visited Nuffield College.

the methods by an example from financial econometrics, where a panel of exchange rate series is dissected into country specific components for level and variance. Such a decomposition helps understanding events like the EMS crisis of 1992, where it is seen which currencies were affected, and how uncertainty spread.

1.2 The model

Write σ_t^2 as a vector of non-negative processes and $\sigma^2 = (\sigma_1^2, \ldots, \sigma_n^2)$, the corresponding matrix. Then we will assume that the observable process $y = (y_1, \ldots, y_n)$ follows a conditionally Gaussian state space form (GSSF) with

$$\left(\begin{array}{c} y_t \\ \alpha_{t+1} \end{array} \right) | \alpha_t, \sigma_t^2 \sim N \left\{ \left(\begin{array}{c} Z_t \alpha_t \\ T_t \alpha_t \end{array} \right), R_t diag\left(\sigma_t^2\right) R_t' \right\},$$

where Z_t, T_t and R_t are non-stochastic matrices. Throughout, to simplify the exposition, we will assume that

$$R_t diag\left(\sigma_t^2\right) R_t' = \left(\begin{array}{cc} G_t diag\left(\sigma_t^2\right) G_t' & 0 \\ 0 & H_t diag\left(\sigma_t^2\right) H_t' \end{array} \right),$$

so the errors in the transition and measurement are conditionally independent. When σ_t^2 is an unobserved exogenous Markov chain then this is a special case of the conditionally Gaussian state space form introduced independently and concurrently by Carter and Kohn (1994) and Shephard (1994b). We will denote this class a GSSF-SV to show that $y|\sigma^2$ can be written as a Gaussian state space model and that unconditionally

$$R_t u_t = \left(\begin{array}{c} y_t \\ \alpha_{t+1} \end{array} \right) - \left(\begin{array}{c} Z_t \alpha_t \\ T_t \alpha_t \end{array} \right)$$

follows a Harvey, Ruiz, and Shephard (1994) type multivariate SV model. In particular we will assume that

$$u_t = \varepsilon_t \odot \sigma_t, \quad \varepsilon_t \overset{i.i.d.}{\sim} N(0, I),$$

where $\odot$ is a Hadamard product. Reviews of the literature on state space models are given in Harvey (1989), Kitagawa and Gersch (1996), West and Harrison (1997), Durbin and Koopman (2001), while the corresponding literature on SV processes is discussed in Ghysels, Harvey, and Renault (1996) and Shephard (1996).

The main model we will work with is where

$$h_{it} = \log \sigma_{it}^2$$

follows a short memory Gaussian process. The most important example of this, which we will focus on, is where h_t follows a vector autoregression

$$h_{t+1} = \gamma + \phi \left(h_t - \gamma \right) + \omega_t, \quad \omega_t \sim NID(0, \Omega). \tag{1}$$

In many models it will be convenient to assume that ϕ and Ω are diagonal matrices. When the aim is solely to smooth the data, rather than predict future values, it often makes sense to simplify the model by setting ϕ to the identity matrix and γ to a vector of zeros so that

$$h_{t+1} = h_t + \omega_t, \quad \omega_t \sim NID(0, \Omega). \tag{2}$$

Throughout we will write $\alpha = (\alpha_1, \ldots, \alpha_n)$, $h = (h_1, \ldots, h_n)$ and $\omega = (\omega_1, \ldots, \omega_n)$.

Example 1 *A traditional Gaussian local level model (e.g. Muth 1961, Harvey 1989 and West and Harrison 1997) has*

$$y_t|\alpha_t \sim N\left(\alpha_t, \sigma_1^2\right), \quad \alpha_{t+1}|\alpha_t \sim N\left(\alpha_t, \sigma_2^2\right).$$

The adaptive local level model generalises this to

$$y_t|\alpha_t, \sigma_t^2 \sim \mathcal{N}\left(\alpha_t, \sigma_{1t}^2\right), \quad \alpha_{t+1}|\alpha_t, \sigma_t^2 \sim \mathcal{N}\left(\alpha_t, \sigma_{2t}^2\right). \tag{3}$$

In a static model, where σ_t^2 is constant through time, then $\mathrm{E}\left(\alpha_{n+s}|y_1, \ldots, y_n\right)$ for, $s > 0$, only depends upon the signal-to-noise ratio $q = \sigma_2^2/\sigma_1^2$. The amount of discounting of past data we use to produce forecasts is constant through time. When σ_t^2 changes through time, the degree of discounting changes through time, adapting to the data.

This example is one of the most basic GSSF-SV models, and it is used in the next sections in a simulation exercise. The application in Section 4 can be seen as a more elaborate version of this adaptive local level model.

The model above is just one special case of the GSSF-SV models. Enlarging a general model with time varying volatility allows the amount of smoothing to change over time. Another example, not elaborated here, would be to write a cubic smoothing spline in GSSF (Wecker and Ansley 1983) with a time-varying level of smoothness.

1.3 The literature

The idea of allowing the variance of components in state space models to change through time is not new. Ameen and Harrison (1984), Shephard (1994a), West and Harrison (1997) and Koopman and Bos (2004) consider the special case where σ_t^2 is a scalar. This allows all the variances of the components to inflate and deflate through time. This added flexibility is potentially very useful, but it does not allow the signal-to-noise ratios to change much through time and so will have a limited impact on mean forecasts. Shephard (1994b, p. 122) mentioned the possibility of allowing the variance of the transition model to change through time and use a non-stationary volatility model to deal with it. However, he did not implement his strategy for this class of models. Highly related work includes Uhlig (1997) and West and Harrison (1997, Ch. 10). There is quite some

work on large dimensional factor SV models. Leading references include Aguilar and West (2000), Pitt and Shephard (1999c), Chib, Nardari, and Shephard (2002). These can be regarded as special cases of the above framework for in these models the α_t process does not have any memory. Harvey, Ruiz, and Sentana (1992) wrote about state space models with ARCH errors terms, however they were not able to prove any properties about their proposed filter and estimation strategies. Bos, Mahieu, and Van Dijk (2000) combine the state space model with a single SV process, and compare its effectiveness with other specifications for the disturbance densities. Carter and Kohn (1994) and Shephard (1994b) independently and concurrently introduced conditionally Gaussian state space models where one could condition on Markov indicator variables, which allowed the σ_t^2 to have a finite range of values at each time period. This type of model was additionally studied in Kim and Nelson (1999).

1.4 Structure of the paper

The organisation of the paper is as follows. In Section 2.1 we discuss a standard approach to designing MCMC algorithms for this type of problem. We will show this method is rather ineffective, delivering algorithms which need large computational resources in order to deliver correct inferences. In Section 2.2 we introduce a reparameterisation of the model which vastly improves the algorithm. Section 2.3 discusses the performance of the algorithms in a simulation exercise using the adaptive local level model 3 above, while Section 3 shows how to effectively implement a particle filter for this method. Section 4 applies the methods on a real world example to dissect a system of exchange rates into country specific factors, allowing a more detailed analysis of the occurrences around the EMS currency crisis in September 1992. Section 5 concludes.

2 Block sampling in GSSF-SV models

In this paper we will write θ as the unknown parameter vector. We often partition θ into ψ and λ, where ψ indexes parameters in the T_t, Z_t and G_t matrices, while λ denotes the parameters of the σ^2 process.

2.1 Conventional block sampling

The GSSF-SV model is a special case of the conditionally Gaussian state space form introduced by Carter and Kohn (1994) and Shephard (1994b). This class has a convenient blocking structure which considerably aids the implementation of MCMC techniques. In particular their methods suggest the following standard algorithm.

1. Initialise σ^2, θ.

2. Update draw from $\theta, \alpha | y, \sigma^2$ by

   (a) Sampling from $\theta | y, \sigma^2$;

   (b) Sampling from the multivariate normal distribution $\alpha | y, \sigma^2, \theta$ using the generic GSSF simulation smoother (Fruhwirth-Schnatter 1994; Carter and Kohn 1994; de Jong and Shephard 1995; Durbin and Koopman 2002).

3. Resample a draw from $\sigma^2 | \alpha, y, \theta$.

4. Goto 2.

The only non-standard part of this sampling is step 3. When σ_t^2 is Markovian and discrete then we can sample from $\sigma^2 | \alpha, y, \theta$ in a single block, as creatively emphasised by Carter and Kohn (1994). Outside that case we have to resort to more brute force MCMC (e.g. in this type of context Carlin, Polson, and Stoffer 1992) by replacing 3 by

3'. Sampling from, for $t = 1, 2, \ldots, n$,

$$\sigma_t^2 | \sigma_{t-1}^2, \sigma_{t+1}^2, y_t, \alpha_t, \alpha_{t+1}, \theta.$$

Sampling from this density can be carried out in a number of ways. We use a method based on the sampler discussed in detail by Kim, Shephard, and Chib (1998), although other methods such as those highlighted by Jacquier, Polson, and Rossi (1994) and Geweke (1994) could be used. This works with the h_t parameterisation and notes that

$$f(h_t | h_{t-1}, h_{t+1}, y_t, \alpha_t, \alpha_{t+1}) \propto f(h_t | h_{t-1}, h_{t+1}) f(y_t | \alpha_t, h_t) f(\alpha_{t+1} | \alpha_t, h_t),$$

which is relatively simple for

$$h_t | h_{t-1}, h_{t+1} \sim N(\gamma + \Sigma \phi' \Omega^{-1} \left\{ (h_{t+1} - \gamma) + (h_{t-1} - \gamma) \right\}, \Sigma), \quad \Sigma = (\Omega^{-1} + \phi' \Omega^{-1} \phi)^{-1}$$

Proposals can be made from this density, either using many univariate draws (which seems always a good idea if Ω and ϕ are diagonal) or all at once. Then they can be accepted using a Hastings-Metropolis step in the usual way.

Step 2a allows for many different implementations. In the following we consider

2a(i). Sample all of $\theta | y, \sigma^2$ at once, using a Hastings-Metropolis step [indicated in the following by $\theta | \sigma^2$ HM] or a step from the so-called Adaptive Rejection Metropolis Sampler, see Gilks, Best, and Tan (1995) [$\theta | \sigma^2$ ARMS];

2a(ii). Sample each element of θ conditional on the others, using a Gibbs sampling scheme [$\theta | \sigma^2$ Gibbs];

2a(iii). Split the density $(\theta|y,\sigma^2) = (\psi|y,\sigma^2) \times (\lambda|\sigma^2)$. Note that the series σ^2 contains all possible information on λ, hence $(\psi|y,\sigma^2) \equiv (\psi|y,\sigma^2,\lambda)$ and $(\lambda|\sigma^2) \equiv (\lambda|y,\psi,\sigma^2)$ $[\psi,\lambda|\sigma^2]$;

2a(iv). If the SV model contains h stochastic volatility series, the density $(\lambda|\sigma^2)$ can usually be split into $(\lambda_i|\sigma_i^2)$, sampling the parameters pertaining to each SV sequence separately. This tends to lower correlation in the final chain $[\theta_i|\sigma_i^2]$.

2a(v). Additionally condition on the states, sampling from $\theta|y,\sigma^2,\alpha$. Fruhwirth-Schnatter (1994) argued that this increases the dependence in the MCMC output, but allows faster and easier simulation of especially $\psi|y,\alpha,\sigma^2$ $[\theta|\sigma^2,\alpha]$.

2.2 Disturbance based block sampling

Even if the sampling from $\sigma^2|\alpha,y,\theta$ is carried out in an effective way the performance of the overall sampler can be poor. The slow mixing of the samplers is caused in part by the choice of conditioning variables in the Gibbs chain. The variance series $\sigma_t^2, t = 1,\ldots,T$ is very informative on the parameters in the SV process. Conditional on the variances σ_t^2, the density $\lambda|\sigma^2,y,\alpha,\psi$ allows for little movement between successive draws of λ, leading to slow exploration of the parameter space.

This section presents a reformulation of the sampling scheme in terms of the errors of the SV process. The idea behind the transformation is to condition on elements which contain as little information as possible on the parameters of the process.

In the original SV specification (1) the volatility process was defined in terms of

$$\omega_t = \Omega^{\frac{1}{2}} u_t \sim NID(0,\Omega), \qquad u_t \sim NID(0,I)$$

Note that there is a one-to-one relation between the volatility process σ_t (and hence h_t) and the white noise disturbances u_t. Therefore, the conditioning in the block sampler can also be done on u_t, which by construction contains less information on the value of the parameters.

The sampling algorithm now becomes:

1. Initialise u,θ, and compute $\sigma^2 = f(u,\lambda)$ as a function of u.
2. Update draw from $\theta,\alpha|y,u$ by
   (a) Sampling from $\theta|y,u$;
   (b) Sampling from $\alpha|y,\sigma^2(u,\lambda),\theta$ as before.
3. Recompute σ^2 from u and θ, sample from $\sigma^2|\alpha,y,\theta$, and reconstruct $u = f^{-1}(\sigma^2,\lambda)$.
4. Repeat from 2.

Notice that step 2 is subtly different from the previous section for now we are no longer conditioning on the time-changing variances. Instead we are conditioning on the standardised disturbances for the log-variances and so as the parameters change so do the conditional variances. The split into $\theta = (\psi, \lambda)$ makes less sense here as the full conditional $\lambda|y, u, \psi$ does not simplify any further as it did before in option (2a(iii)). Furthermore, the full conditional densities needed for the Gibbs sampler in (2a(ii)) are not known in closed form when we condition on the disturbances u. The other possibilities, of applying HM, ARMS, splitting between the h SV sequences and conditioning on the state are still valid. Methods are indicated similarly, e.g. $[\theta|u$ HM] for the Hasting-Metropolis sampler conditioning on the disturbances, sampling all parameters at once.

There has been very little research into the effect of reparameterisation on the convergence of MCMC algorithms. The only two papers we know of are Pitt and Shephard (1999a) and the excellent Frühwirth-Schnatter (2004). The latter paper is relevant here as the author has a section on designing samplers based on the errors of the process rather than the states. This work was carried out in the case of the GSSF.

2.3 Performance of the formulations in a simulation exercise

Settings

The performance of the standard formulation will be evaluated using the adaptive local level model (3), with separate stochastic volatility series for the observation and transition equations. The model only contains parameters in the volatility part, so here $\theta \equiv \lambda$. We choose the settings of the model to mimic situations found in financial time series, i.e.

- Persistent time varying volatility on the observations;
- Relatively slow moving underlying trend component;
- Variability in the trend volatility;
- Long data series.

Table 1: Parameters and prior choices

This table about here

Table 1 reports the parameter values chosen for the data generating process, together with the prior specifications. The priors are mildly informative, to ensure existence of all sampling densities, while not influencing the posterior density much.

The simulated data set contains 5,000 observations, which for a series of daily observations would correspond to a time span of roughly 20 years.

The algorithms were run in order on to collect a total of 100,000 posterior draws, after allowing the algorithms a burn-in period of 20,000 iterations for initialising the SV sequences and getting rid of the effect of the starting values of the samplers. Simulations are performed with the methods specified in Sections 2.1–2.2.

For the Hastings-Metropolis and Gibbs samplers, where the sampling of the parameters $\theta|\sigma^2, y$ is relatively cheap, this step is repeated 5 times before series of α and h are updated. The HM sampler draws candidate parameter vectors from a random walk candidate density, with a covariance matrix proportional to a numerical approximation of the target density. The approximation is updated every 100 iterations. The ARMS algorithm constructs a proposal density over a grid. We choose an initial grid of 10 points.

Performance of the samplers

As explained before, a major obstacle for Bayesian methods is the mixing of the sampling chain. Especially with models including unobserved stochastic volatility components, the correlation in the chain can be high. In this section, we look at the performance of the algorithms from three different angles.

This figure about here

Figure 1: Posterior density of ϕ_2 according to different samplers

First, Figure 1 displays the posterior distribution of the autoregressive parameter ϕ_2 of the SV process for the state equation, as estimated using the different sampling methods. The lefthand panel displays the posterior constructed using methods applying the conventional block sampling algorithms of Section 2.1 whereas on the right the reformulation in disturbances of Section 2.2 was used. For the model at hand, the methods $[\theta|\sigma^2$ HM] and $[\theta|\sigma^2, \alpha$ HM] correspond exactly, as all parameters in the model are within the SV process; conditioning on the state does not alter the likelihood. When conditioning on u instead of σ^2, there is a difference in performance.

This figure about here

Figure 2: Autocorrelation of draws of ϕ_2 according to different samplers

From the graph it is seen how the methods conditioning as usual on the volatility process itself, differ strongly in their estimate of the posterior density, even though the underlying true posterior density is the same for all simulations. This indicates that even after such a long sample, convergence of the sampling scheme is not complete yet.

The samplers on the righthand side correspond well, and the estimated posterior densities are smooth. This is a better sign of full convergence of the sampler after the 100,000 drawings collected, and hence of better properties of the sampling schemes based on the reformulation.

Second, as low mixing implies strong correlation in the chain of sampled parameter vectors, it is instructive to look at the correlation of the individual parameters. Figure 2 displays the autocorrelation of the samples of ϕ_2, applying the different sampling schemes. Again, the lefthand panel displays results for the standard formulation in terms of σ^2, with different choices for the sampler of $\theta, \alpha | y, \sigma^2$, while the righthand panel follows the alternative formulation. Note that on the left the autocorrelation is plotted until lag 10,000, on the right up to lag 1,000. Indeed, for this model, conditioning on the disturbances in the volatility equations leads to strongly faster mixing of the chain.

Table 2: Inefficiency of the sampling procedures

This table about here

Finally, Table 2 displays the estimated integrated autocorrelation times or inefficiency factors. These were used in Shephard and Pitt (1997) and Kim, Shephard, and Chib (1998). Note that Geweke (1989) prefers to report the inverse of this number. The measure compares the variance of the sample mean, adapted for correlation in the chain, to the variance of the mean when the correlation is not accounted for, as

$$\hat{R}_{B_m} = 1 + \frac{2B_m}{B_m - 1} \sum_{i=1}^{B_m} K\left(\frac{i}{B_m}\right) \hat{\rho}(i)$$

with $K(j)$ the Parzen kernel and B_m the bandwidth. A low value of R is preferable, while a value of one indicates that the sampler delivers an uncorrelated set of draws. In the table the inefficiency factors are calculated using a bandwidth of $B_m = 2,000$. The conclusions are the same as before: Conditioning on the volatility disturbances increases the efficiency of the sampling strongly.

2.4 Expected performance in different situations

Even though the results in the previous section are encouraging, the question is if such a strong improvement can be expected in every situation. To understand this question, it is important to realise where the gains are coming from.

Following Tanner (1996, Section 4.4.2) we can use the 'missing information principle' to decompose the log-posterior density $p(\theta|y)$ into parts pertaining to the augmented posterior and to the posterior of the augmentation itself, which is for the original formulation

$$\log p(\theta|y) = \log p(\theta|y, \sigma^2) - \log p(\sigma^2|y, \theta) + C.$$

The degree of mixing of the Gibbs chain depends on the information in the steps, i.e. on the expectation with respect to the density $p(\sigma^2|y, \theta)$ of the augmentation parameters of

$$-\frac{\partial^2 \log p(\theta|y)}{\partial\theta\partial\theta'} = -\frac{\partial^2 \log p(\theta|y, \sigma^2)}{\partial\theta\partial\theta'} + \frac{\partial^2 \log p(\sigma^2|y, \theta)}{\partial\theta\partial\theta'}$$

and similarly for the reformulation based on conditioning on u instead of on h. If the expected information

$$\int_h -\frac{\partial^2 \log p(\theta|y,\sigma^2)}{\partial\theta\partial\theta'} p(\sigma^2|y,\theta)\partial\sigma^2 > \int_u -\frac{\partial^2 \log p(\theta|y,u)}{\partial\theta\partial\theta'} p(u|y,\theta)\partial u,$$

then the reformulation indeed reduces the informativeness of the augmentation variable, and a gain in the sampling performance can be expected.

To check this conjecture, the adaptive local level model (3) was simulated using a stochastic volatility component only on the observation equation. The autocorrelation ϕ is fixed at 0.95, the mean of the SV equation at $\gamma = 1$, and the standard deviation of the transition equation $\sigma_2 = 0.5$. The unconditional standard deviation of the SV equation,

$$\sigma_{\text{SV}} = \frac{\sigma_\omega}{\sqrt{1-\phi^2}}$$

is allowed to range from 0.05 to 5. Figure 3 displays the determinant of the information matrix for the different choices of σ_{SV}, for the original parameterisation and the reformulation.

This figure about here

Figure 3: Expected determinant of the information matrix of density of the parameters in the Adaptive Local Level model, with SV on the observation equation, for varying values of the standard deviation of the SV process.

The figure indicates how, for SV processes with low unconditional variance, the density of $p(\theta|y,\sigma^2)$ can be expected to contain much more information on θ than the reformulated density $p(\theta|y,u)$, and hence the reformulation can be expected to deliver gains in the mixing of the Gibbs chain. On the other hand, for higher variances σ^2_{SV}, including integrated SV processes with $\phi = 1$, the original formulation can be expected to work just as well, or better.

Notice that this simulation exercise is only concentrating on the effect of the formulation on the parameters of the SV process itself. For a general model, with parameters relating to the mean and others relating to the variance equation, one cannot immediately conclude that one method will always be preferable to the other. Practical experience should be used to guide the sampler choice.

3 Particle filtering

An important feature of MCMC is that it produces samples from $\alpha, \sigma^2, \theta|\mathcal{F}_T$, conditioning on all information available at the end of the sample, $\mathcal{F}_T = \{y_1, \ldots, y_T\}$. Of course this is very useful in terms of summarising important features of the model and the data. MCMC methods do not, on the other hand, produce effective methods for sequentially sampling from

$$\alpha_t, \sigma^2_t|\mathcal{F}_t, \theta, \quad t = 1, 2, \ldots, n.$$

Such quantities, conditioning only upon information available at time t, are very important in practise for the use of sequential forecasting and model checking. Sequential forecasting underlies Bayesian decision-making under uncertainty, where future actions may depend on the present, i.e. filtered, unobserved components α_t, σ_t^2. For model checking, extensive use is made of the marginal likelihood contrasting competing models using Bayes factors (Aitkin 1991; Kass and Raftery 1995). An estimate of the marginal likelihood can be computed from the conventional likelihood, $\mathcal{L}(\theta; y)$, which in turn needs a prediction error decomposition. This decomposition itself needs filtered state estimates of α_t and σ_t^2.

A standard way of computing these filtered state estimates is via a particle filter (e.g. Gordon, Salmond, and Smith 1993, Pitt and Shephard 1999b and Doucet, de Freitas, and Gordon 2001). In this case the model has a lot of structure which allows us to carry out particle filtering in a very fast way. This work follows the ideas discussed in, for example, Pitt and Shephard (1999b) and Chen and Liu (2000).

We will argue by induction. Consider a collection of particles, or sample, which is used to approximate the distribution of $\alpha_t, \sigma_t^2 | \mathcal{F}_t$,

$$\sigma_t^{2(i)}, f_N\left(\alpha_t | \mathcal{F}_t; a_t^{(i)}, P_t^{(i)}\right), \quad i = 1, 2, \ldots, M.$$

This implies, in particular, that the particle approximation to $\alpha_t, \sigma_t^2 | \mathcal{F}_t$ is

$$\widehat{f}\left(\alpha_t, \sigma_t^2 | \mathcal{F}_t\right) = \sum_{i=1}^{M} f_N\left(\alpha_t | \mathcal{F}_t; a_{t|t}^{(i)}, P_{t|t}^{(i)}\right) I\left(\sigma_t^2 = \sigma_t^{2(i)}\right),$$

a mixture of normals. Consequently

$$\widehat{f}\left(\alpha_t | \sigma_t^2 = \sigma_t^{2(i)}, \mathcal{F}_t\right) = f_N\left(\alpha_t | \mathcal{F}_t; a_{t|t}^{(i)}, P_{t|t}^{(i)}\right),$$

We treat this approximation as if it is true, which implies straightforwardly that

$$\widehat{f}\left(\alpha_{t+1} | \sigma_t^2 = \sigma_t^{2(i)}, \mathcal{F}_t\right) = f_N\left(\alpha_{t+1} | \mathcal{F}_t; a_{t+1|t}^{(i)}, P_{t+1|t}^{(i)}\right)$$

with

$$a_{t+1|t}^{(i)} = T_t a_{t|t}^{(i)} \quad \text{and} \quad P_{t+1|t}^{(i)} = T_t P_{t|t}^{(i)} T_t' + H_t diag\left(\sigma_t^{2(i)}\right) H_t'.$$

We propagate the volatility process forward using simulation. For each $\sigma_t^{2(i)}$ we generate R daughters by simulating forward

$$\sigma_{t+1}^{2(i,j)} \sim \sigma_{t+1}^2 | \sigma_t^{2(i)}, \quad j = 1, 2, \ldots, R.$$

This produces the approximation to the density of $f\left(\alpha_{t+1}, \sigma_{t+1}^2 | \mathcal{F}_t\right)$ of

$$\widehat{f}\left(\alpha_{t+1}, \sigma_{t+1}^2 | \mathcal{F}_t\right) = \left\{ \sum_{i=1}^{M} f_N\left(\alpha_{t+1} | \mathcal{F}_t; a_{t+1|t}^{(i)}, P_{t+1|t}^{(i)}\right) \left\{ \frac{1}{R} \sum_{j=1}^{R} I\left(\sigma_{t+1}^2 = \sigma_{t+1}^{2(i,j)}\right) \right\} \right\}.$$

The most important step is that we now calculate

$$\begin{aligned}\widehat{f}\left(\alpha_{t+1},\sigma_{t+1}^2,i,j|\mathcal{F}_{t+1}\right) &\propto f_N\left(\alpha_{t+1}|\mathcal{F}_t;a_{t+1|t}^{(i)},P_{t+1|t}^{(i)}\right)I\left(\sigma_{t+1}^2=\sigma_{t+1}^{2(i,j)}\right)\\ &\quad\times f_N(y_{t+1}|Z_{t+1}\alpha_{t+1},G_{t+1}diag\left(\sigma_{t+1}^2\right)G_{t+1}').\end{aligned}$$

It follows that

$$\widehat{f}\left(\alpha_{t+1},\sigma_{t+1}^2,i,j|\mathcal{F}_{t+1}\right)=\left(\frac{w_{i,j}}{\sum_{k=1}^{M}\sum_{l=1}^{R}w_{k,l}}\right)f_N\left(\alpha_{t+1}|\mathcal{F}_{t+1};a_{t+1|t+1}^{(i,j)},P_{t+1|t+1}^{(i,j)}\right)$$

where

$$\begin{aligned}w_{i,j}&=f_N\left(v_{t+1}^{(i)}|0,F_{t+1}^{(i,j)}\right),\\ v_{t+1}^{(i)}&=y_{t+1}-Z_{t+1}a_{t+1|t}^{(i)},\quad F_{t+1}^{(i,j)}=Z_{t+1}P_{t+1|t}^{(i)}Z_{t+1}'+G_{t+1}diag\left(\sigma_{t+1}^{2(i,j)}\right)G_{t+1}',\end{aligned}$$

and

$$\begin{aligned}a_{t+1|t+1}^{(i,j)}&=a_{t+1|t}^{(i)}+P_{t+1|t}^{(i)}Z_{t+1}'\left\{F_{t+1}^{(i,j)}\right\}^{-1}v_{t+1}^{(i)},\\ P_{t+1|t+1}^{(i,j)}&=P_{t+1|t}^{(i)}-P_{t+1|t}^{(i)}Z_{t+1}'\left\{F_{t+1}^{(i,j)}\right\}^{-1}Z_{t+1}P_{t+1|t}^{(i)}.\end{aligned}$$

We need to sample from this density to produce the new set of particles, in order to complete the algorithm. This is straightforward, we sample with replacement from the discrete distribution

$$\sigma_{t+1}^{2(i,j)},\quad i=1,2,\ldots,M;\quad j=1,2,\ldots,R,$$

with probabilities proportional to $w_{i,j}$. Associated with each of these discrete particles are the distributions $\alpha_{t+1}|\mathcal{F}_{t+1};a_{t+1|t+1}^{(i,j)},P_{t+1|t+1}^{(i,j)}$. Relabelling all the particles produces the sample

$$\sigma_{t+1}^{2(i)},f_N\left(\alpha_{t+1}|\mathcal{F}_{t+1};a_{t+1}^{(i)},P_{t+1}^{(i)}\right),\quad i=1,2,\ldots,M.$$

The likelihood of the model is delivered as a by-product of the particle filter. It is calculated as

$$\mathcal{L}(y_{t+1}|\theta,\mathcal{F}_t)\approx\frac{1}{RM}\sum_i\sum_j w_{i,j}.$$

4 Factoring exchange rates

This section applies the sampling methods and the particle filter of Sections 2.1-2.2 and 3 on two data sets. First, Section 4 uses a GSSF-SV model to dissect exchange rates into country-specific components. Applying the techniques to data series of the British Pound, the German DMark and the Japanese Yen against the U.S. Dollar, around the time of the 1992 crisis in the European Monetary System, allows us to analyse in detail the uncertainty in the market, and the extend to which it affected these currencies specifically.

Modelling correlations

The models based on the General State Space Form with Stochastic Volatility promise great flexibility in modelling time series with intricate dependencies in both means and variances. One area where the application of such models can be of great interest is the field of exchange rate research.

For instance in Lyons (2001, Ch. 6), in his introduction to exchange rate models, the purchasing power parity (PPP) is presented as a first way to relate exchange rates to the structure of the economy. The PPP can be written as

$$S_{\mathrm{UK/US}} = P_{\mathrm{UK}}/P_{\mathrm{US}}$$

relating the spot exchange rate between Dollar and Pound ($S_{\mathrm{UK/US}}$) to the (consumer or other) price indices in the respective countries (P_{UK} and P_{US}, respectively). As is well known, it is hard to prove empirically that the PPP actually holds, as it is unclear for what price indices or on what time span the relation should be valid. Hence, the discourse in Lyons (2001) continues with models of exchange rates themselves, without looking into the separate components in the exchange rates.

When multiple exchange rates are taken into account, there always is the problem of correctly modelling the correlation between series. As both $S_{\mathrm{UK/US}}$ and $S_{\mathrm{YY/US}}$ relate to the dollar, both levels and volatilities of the exchange rates are surely strongly related, see e.g. Beine (2004).

A clean way to model this relation was presented in Mahieu and Schotman (1994). They propose a factor structure for the logarithm of the exchange rates $s_{ij,t} = \log S_{ij,t}$ between countries i and j as

$$s_{ij,t} = e_{i,t} - e_{j,t},$$

where the country specific components $e_{i,t}$ are assumed to follow random walks (which implies the assumption of unpredictable returns on the exchange rates) with time varying variance. Using country 0 as a numeraire, a full model for multiple exchange rates between $n+1$ countries could be

$$\begin{pmatrix} s_{10,t} \\ \vdots \\ s_{n0,t} \end{pmatrix} = \begin{pmatrix} -\mathbf{1} & \mathcal{I}_n \end{pmatrix} \begin{pmatrix} e_{0,t} \\ \vdots \\ e_{n,t} \end{pmatrix} \tag{4}$$

$$\begin{pmatrix} e_{0,t+1} \\ \vdots \\ e_{n,t+1} \end{pmatrix} = \begin{pmatrix} e_{0,t} \\ \vdots \\ e_{n,t} \end{pmatrix} + H_t^{\frac{1}{2}} \epsilon_t \qquad \epsilon_t \sim N(0, \mathcal{I}_{n+1}) \tag{5}$$

where H_t is a diagonal matrix with typical element

$$H_{ii,t} = \exp(h_{i,t})$$
$$h_{i,t+1} - \gamma_i = \phi_i(h_{i,t} - \gamma_i) + \sigma_{i,\omega}\omega_{i,t}$$
$$\omega_t \sim N(0, \mathcal{I}_{n+1}).$$

The values of the SV processes $h_{i,t=0}$ at the start of the process should be initialised diffusely, such that the process can choose the initial level of variance by itself.

This model, indicated below as Factor-SV model, allows us to extract the country specific factors for both levels and volatilities out of a set of exchange rates. With the factors, one can analyse the effect central bank interventions have on the exchange rates (Dominguez 1998; Beine, Laurent, and Palm 2004), but now for the first time on each country specifically. Likewise, one can e.g. measure which countries are hit hardest during currency crises (Evans and Lyons 2003).

The next sections apply the above model to exchange rate data on the US Dollar, the German DMark, the Japanese Yen and the British Pound, over the years 1991-1993, around the period of the crisis in the European Monetary System (EMS). Over this period the model is contrasted to a the standard stochastic volatility model for exchange rate returns, or, using the previous notation, modelling

$$\begin{pmatrix} s_{10,t+1} \\ \vdots \\ s_{n0,t+1} \end{pmatrix} = \begin{pmatrix} s_{10,t} \\ \vdots \\ s_{n0,t} \end{pmatrix} + H_t\epsilon_t \qquad \epsilon_t \sim N(0, \mathcal{I}_n) \tag{6}$$

where H_t is a diagonal matrix as before, though now of size $n \times n$. Note how for this model the numeraire of the exchange rates matters, in contrast to the factor model. Also, for simplicity, we do not take the correlations between the exchange rates explicitly into account. Effectively we estimate a panel of univariate random walk-SV (RW-SV) models.

Data and estimability

The proposed factor model contains, for n exchange rates of length T, $n+1$ unobserved factor components of length T, plus the $n+1$ volatilities which are second order unobserved processes. Essentially the SV processes serve to estimate the $3 \times (n+1)$ parameters $\sigma_{i,\omega}, \phi_i$ and γ_i, which can be expected to be considerably hard given the low degree of information available on these parameters.

With $n = 1$ exchange rate, the model is not identifiable as it is not possible to distinguish between the two country level factors. With $n > 1$, theoretically the numeraire factor can be identified as a driving force within all exchange rates; for larger values of n more information on $e_{0,t}$ and hence on the other factors is available.

In Mahieu and Schotman (1994) the model on exchange rates is estimated in a classical framework. The estimation procedure applied in their article however does not allow to estimate jointly all unobserved processes, and can serve only as an approximation.

This figure about here

Figure 4: Exchange rates of the German DMark, British Pound and Japanese Yen against the US Dollar, with $1/1/1991 \equiv 100$

To keep estimation tractable, we concentrate on the three exchange rates of the German DMark, the Japanese Yen and the British Pound against the US Dollar, over the period 2/1/1991–31/12/1993.[1] This period contains 755 daily observations. The series are depicted in Figure 4, with the exchange rates scaled to 100 at the start of the sample. In the model we use the transformation $s_{ij,t} = 100 \ln S_{ij,t}$, with $S_{ij,t}$ the exchange rate between countries i and j at day t.

Posterior estimates

The model as presented above is estimated using a subsample of the simulation techniques described earlier. Conditioning either on the stochastic volatilities σ^2, or on the disturbances u in the SV equations, the parameters θ are sampled using a Hastings-Metropolis step. The parameters are either sampled all at once, or splitting the parameter vector into elements concerning each of the volatility sequences separately, and splitting between the AR-parameters ($\psi_i, \sigma_{\xi,i}$ and the level parameters γ_i (for simplicity, in the tables and figures this splitting is denoted $\theta_i|\sigma^2$ or $\theta_i|u$). The final sample contains 100,000 parameter vectors, collected after a burn-in period of 10,000 iterations.

The priors of the standard deviations in the SV processes are Inverted Gamma-1 densities, with expectation and standard deviation equal to 0.2. The ϕ_i parameters used a Beta-prior with expectation 0.86 and standard deviation 0.1, while γ_i was supposed to be normally distributed around 0 with $\sigma_\gamma = 2$.

Table 3: Posterior statistics of the Factor-SV model

This table about here

Table 4: Posterior statistics of the standard RW-SV model

This table about here

The results for the multi-SV model are summarised in Table 3 and Figure 5, with the posterior mean, highest posterior density region, efficiency measures for the various samplers, and a figure of the posterior density of selected parameters. Statistics for the alternative model with standard

[1] Source: Reuters FX, interbank exchange rate at 16h GMT+1.

This figure about here

Figure 5: Posterior densities of parameters relating to the US volatility factor resulting from the competing sampling methods

RW-SV are given in Table 4, while discussion of the estimated underlying factors around the time of the EMS crisis is postponed until later in this section.

Concerning the posteriors, we can conclude that the data is indeed informative on the parameters in the SV process, as the posterior shifts away from the prior. This is quite an accomplishment, as the estimation is very indirect: From the exchange rates, through the unobserved level factors the unobserved SV processes are estimated, from which in turn the disturbances are extracted to estimate their standard deviations σ_ω, autocorrelations ϕ and location parameters γ.

The quality of the sample from the posterior density differs between the samplers. For instance in Figure 5 the estimated posterior densities of $\sigma_{\omega,\mathrm{US}}, \phi_{\mathrm{US}}$ and γ_{US} indicate troublesome convergence for the standard formulation conditioning on the volatility sequences σ^2. The two alternatives using the reformulated sampler correspond with each other, and with the standard samplers in case those converged.

A similar picture is displayed from the inefficiency measures. Overall, efficiency increases when switching to the reformulation, though some parameters like ϕ_{UK}, and also the γ's, seem to be estimated more easily in the original formulation. This is related to the findings in Section 2.4, where for higher unconditional variation in the SV series the original formulation has lower expected information in the SV series σ^2 than in the disturbances u.

Note that this model and data series is still relatively simple. In Figure 5 it already is seen that convergence for this model, using the standard sampling procedure may not have been complete. At the amount of correlation in the chain, full convergence may take a very long time, too long to be practical. On larger data sets, covering a larger time period and/or at a higher frequency, with more exchange rates included, and possibly with further explanatory variables in the model, the sampling problems are exacerbated strongly. Without the alternative formulation of the sampling procedure, a valid sample from the posterior distribution of the parameters may not be attainable.

For the model with the standard SV specification, without any underlying factors, posterior statistics are reported in Table 4. Even though this model is simpler, as no underlying level components have to be estimated, and there are only 3 SV components with their parameters, efficiency of the estimated parameters is not clearly better than in the case of the factor model. Conclusions on the performance of the samplers correspond with earlier results on the factor models: Overall, the reformulation does increase the efficiency of the chain.

Table 5: Likelihood estimates

This table about here

The particle filter of Section 3 can be used to estimate the marginal likelihood of the models. Table 5 reports the log-likelihood at the posterior mode, and the corresponding log-marginal likelihood computed using the kernel method (Kass and Raftery 1995; Bos 2002). For the particle filter a set of $m = 1,000$ particles with $r = 100$ daughters was used. A difference in log-marginal likelihood of 613 points indicates that the factor model is strongly favoured above the standard SV model. Of course, allowing for further correlation in the standard SV model could partly close the gap between the models, but the message from the likelihoods is that the factor model indeed is an interesting alternative for standard modelling procedures, especially now that it comes within reach computationally.

Resulting factors and the crisis

The advantage of the multivariate factor model for exchange rates is that it is able to extract estimates of the strength and uncertainty pertaining to each of the currencies separately. Figures 6 and 7 display these level e_i and volatility $\sigma_{i,t} = \exp(h_{i,t}/2)$ factors, over the 1991-1993 period. Note that the factors displayed here are derived from the MCMC sampling results, and hence condition on the full data set. Similar graphs, conditional only on past data, can be made using the particle filter of Section 3.

This figure about here

Figure 6: Level factors extracted for the currencies over the 1991-1993 period, with 1-standard deviation error bound.

This figure about here

Figure 7: Volatility factors extracted for the currencies over the 1991-1993 period, with 1-standard deviation error bound.

For the means, it is utterly clear that it was the British Pound which devaluated in September 1992. Though there was some increased fluctuation in the strength of the German DMark, this effect was not strong, and may in part be caused by estimation uncertainty.

The graph on the uncertainty in the exchange rate markets is more telling. First of all, the only three 'important' currencies are considered to be the U.S. Dollar, the Japanese Yen and the German DMark (or Euro, nowadays). The British Pound is seen to have hardly any variation of itself before September 1992, as it was completely pegged to the DMark; any movement of the Pound exchange rate was due to movements in other currencies.

In September, the EMS crisis hits. With the devaluation of the Pound, its volatility explodes to levels about fifteen times higher than before. As the whole EMS is involved, also the DMark reacts to the uncertainty in the system, but to a lesser extent. Neither the Dollar or the Yen show any reaction, as the the crisis only involves the European currencies.

The factors shown here are estimated using the Hastings-Metropolis algorithm with the reformulated sampler, sampling the parameters separately over all regions (as indicated by $\theta_i|u_i$ earlier). The results for the other methods are virtually equal; even after a relatively short simulation, the estimate of the currency factors is already formed, and does not increase much in precision by lengthening the sample size. As the data is clearly informative on these factors, this opens up the scope for estimating the influence of e.g. central bank interventions on these factors, or whether it is possible to use factor estimates to predict the (probability of) a crisis.

5 Concluding remarks

In this paper we have focused on the GSSF-SV class of adaptive time series models. We have shown that standard MCMC methods can be ineffective in this context, showing slow convergence and high correlation in resulting sampling chains. Therefore, we have designed a reparameterisation of the sampler. This delivers a method which allows us to routinely carry out likelihood based inference using a palette of parameterisations, in order to choose the one with best characteristics for the problem at hand. We back this up with an effective particle filter which allows us to carry out on-line forecasting and diagnostic checking for this model. We illustrated the methods on simulated and real data.

Using simulated data, the effect of the reformulation was clear, in that the simulated chain displayed overwhelmingly better mixing properties. In the real data example, the advantage of the reformulation is found to be strongest when the unobserved SV process is relatively persistent with low unconditional variance, with the effect of the change in parameterisation being different for the various parameters in the model. It is left for future research to check whether it is advisable to allow different parameters within a model to be sampled using differing conditioning variables.

The GSSF-SV class of models was found to be valuable in modelling a panel of exchange rates, allowing the exchange rates to be dissected into country-specific level and variance factors. This decomposition promises to be a fruitful starting point for future analysis of mayor economic events in the exchange markets, as it allows to indicate where changes in level or variance stem from, from one specific country, from a group of countries, or from all countries jointly. This way, events like currency crisis, or also the effectiveness of central bank interventions, can be investigated at a level of detail which was previously unattainable.

6 Acknowledgements

Neil Shephard's research is supported by the UK's ESRC through the grant "High frequency financial econometrics based upon power variation". All the calculations made in this paper are

based on software written by the authors using the Ox language of Doornik (2001) in combination with SsfPack of Koopman, Shephard, and Doornik (1999). This paper benefited greatly from many discussions with Jurgen Doornik, Thomas Kittsteiner, Siem Jan Koopman and Bent Nielsen.

References

Aguilar, O. and M. West (2000). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of Business and Economic Statistics 18*, 338–357.

Aitkin, M. (1991). Posterior Bayes factors. *Journal of the Royal Statistical Society, Series B 53*(1), 111–142.

Ameen and J. Harrison (1984). Discounted weighted estimation. *Journal of Forecasting 3*, 285–296.

Beine, M. (2004). Conditional covariances and direct central bank interventions in the foreign exchange markets. *Journal of Banking and Finance 28*(6), 1385–1411.

Beine, M., S. Laurent, and F. C. Palm (2004). Central Bank forex interventions assessed using realized moments. Core Discussion Paper 2004/1, UCL, Louvain-la-Neuve.

Bos, C. S. (2002). A comparison of marginal likelihood computation methods. In W. Härdle and B. Ronz (Eds.), *COMPSTAT 2002 – Proceedings of Computational Statistics*, pp. 111–117.

Bos, C. S., R. J. Mahieu, and H. K. Van Dijk (2000). Daily exchange rate behaviour and hedging of currency risk. *Journal of Applied Econometrics 15*(6), 671–696.

Carlin, B. P., N. G. Polson, and D. Stoffer (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modelling. *Journal of the American Statistical Association 87*, 493–500.

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

Chen, R. and J. Liu (2000). Mixture kalman filters. *Journal of the Royal Statistical Society, Series B 62*, 493–508.

Chib, S., F. Nardari, and N. Shephard (2002). Analysis of high dimensional multivariate stochastic volatility models. Unpublished manuscript.

de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*, 339–50.

Dominguez, K. M. (1998). Central bank intervention and exchange rate volatility. *Journal of International Money and Finance 17*, 161–190.

Doornik, J. A. (2001). *Ox: Object Oriented Matrix Programming, 3.0*. London: Timberlake Consultants Press.

Doucet, A., N. de Freitas, and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*, 603–616.

Evans, M. D. D. and R. K. Lyons (2003). How is macro news transmitted to exchange rates? Working paper 9433, NBER.

Fruhwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis 15*, 183–202.

Frühwirth-Schnatter, S. (2004). Efficient Bayesian parameter estimation for state space models based on reparameterizations. In A. C. Harvey, S. J. Koopman, and N. Shephard (Eds.), *State Space and Unobserved Component Models: Theory and Applications*. Cambridge University Press.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica 57*, 1317–39.

Geweke, J. (1994). Comment on Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics 12*, 397–399.

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

Gilks, W. R., N. G. Best, and K. K. C. Tan (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics 44*, 455–472.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEEE-Proceedings F 140*, 107–113.

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

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

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

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

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Kim, C.-J. and C. R. Nelson (1999). *State-Space Models with Regime Switching. Classical and Gibbs-Sampling Approaches with Applications*. Cambridge: MIT.

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

Kitagawa, G. and W. Gersch (1996). *Smoothness Priors Analysis of Time Series*. New York: Springer Verlag.

Koopman, S. J. and C. S. Bos (2004). State space models with a common stochastic variance. *Journal of Business and Economic Statistics 22*(3), 346–357.

Koopman, S. J., N. Shephard, and J. A. Doornik (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal 2*, 107–160.

Lyons, R. K. (2001). *The Microstructure Approach to Exchange Rates*. MIT Press.

Mahieu, R. J. and P. C. Schotman (1994). Neglected common factors in exchange rate volatility. *Journal of Empirical Finance 1*, 279–311.

Muth, J. (1961). Rational expectations and the theory of price movements. *Econometrica 29*, 315–35.

Pitt, M. K. and N. Shephard (1999a). Analytic convergence rates and parameterisation issues for the Gibbs sampler applied to state space models. *Journal of Time Series Analysis 21*, 63–85.

Pitt, M. K. and N. Shephard (1999b). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association 94*, 590–599.

Pitt, M. K. and N. Shephard (1999c). Time varying covariances: a factor stochastic volatility approach (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 6*, pp. 547–570. Oxford: Oxford University Press.

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

Shephard, N. (1994b). Partial non-Gaussian state space. *Biometrika 81*, 115–31.

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

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–67.

Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions* (3 ed.). New York: Springer-Verlag.

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

Wecker, W. E. and C. F. Ansley (1983). The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association 78*, 81–89.

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

Figures and tables

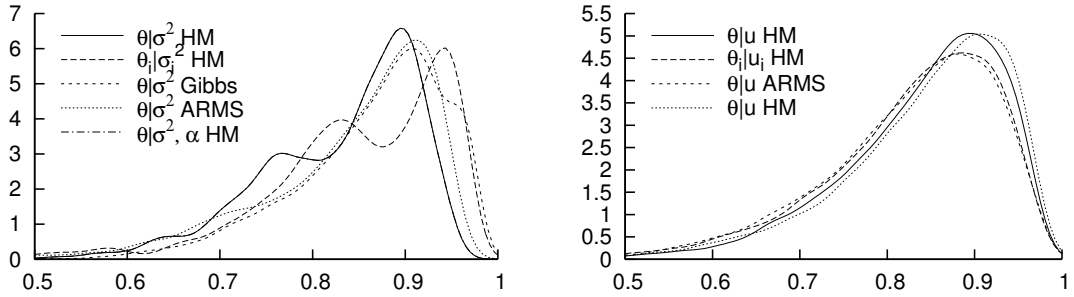

Figure 1: Posterior density of ϕ_2 according to different samplers

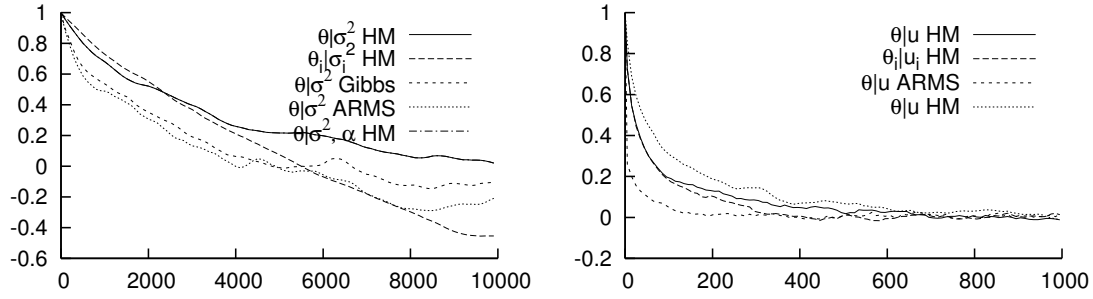

Figure 2: Autocorrelation of draws of ϕ_2 according to different samplers

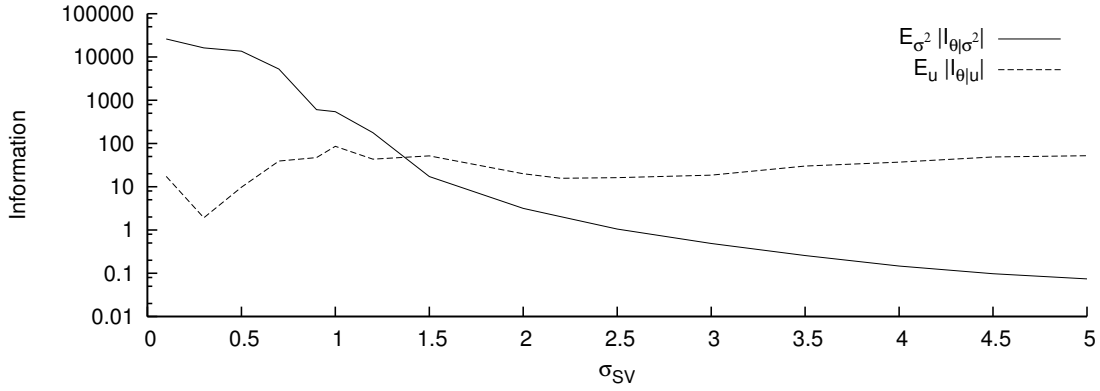

Figure 3: Expected determinant of the information matrix of density of the parameters in the Adaptive Local Level model, with SV on the observation equation, for varying values of the standard deviation of the SV process.

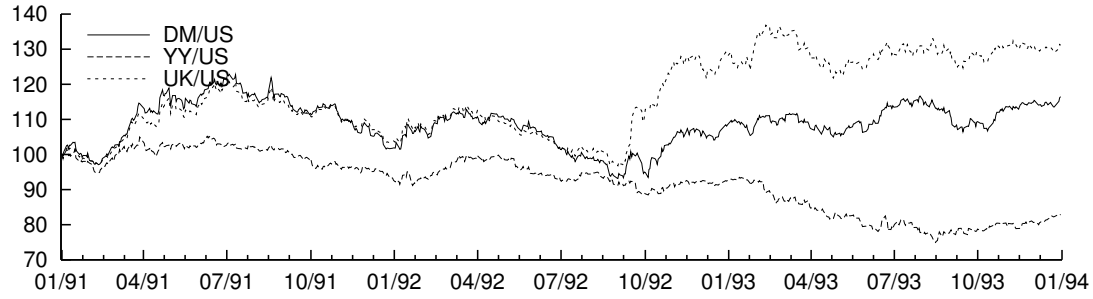

Figure 4: Exchange rates of the German DMark, British Pound and Japanese Yen against the US Dollar, with $1/1/1991 \equiv 100$

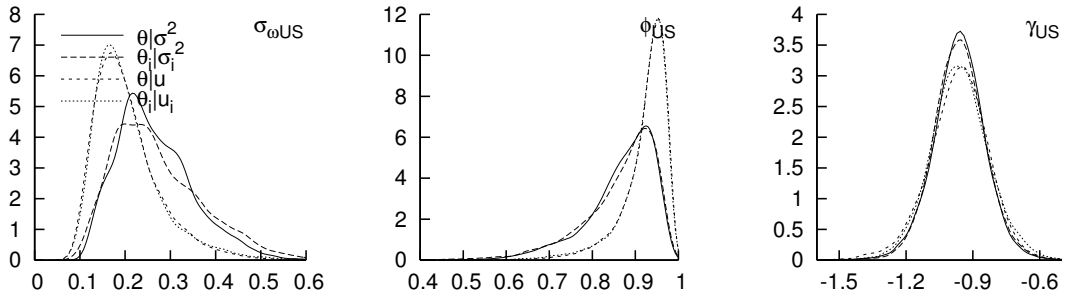

Figure 5: Posterior densities of parameters relating to the US volatility factor resulting from the competing sampling methods

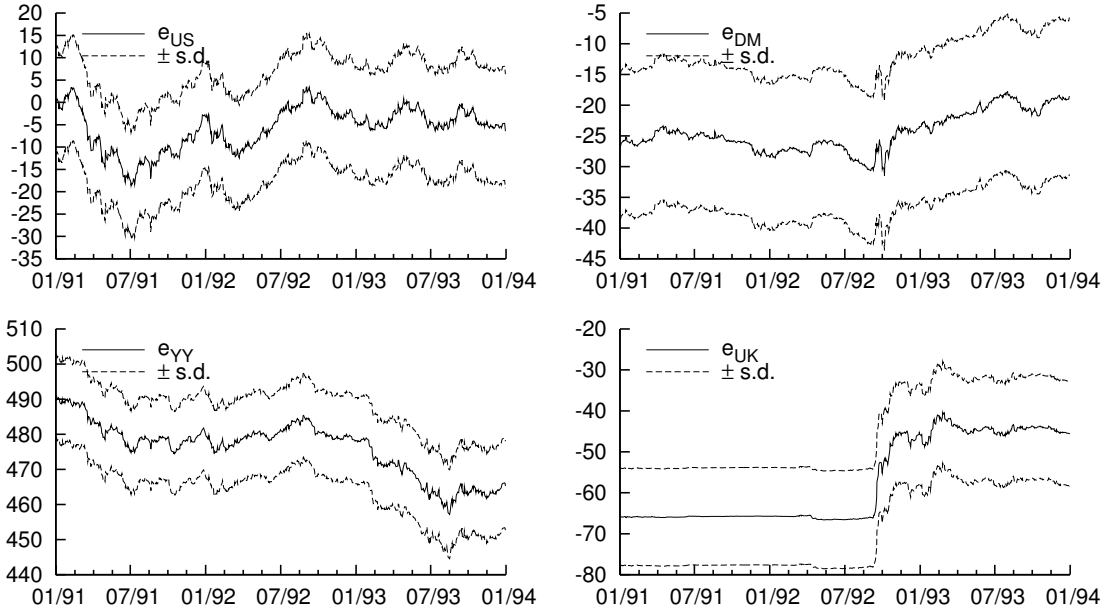

Figure 6: Level factors extracted for the currencies over the 1991-1993 period, with 1-standard deviation error bound.

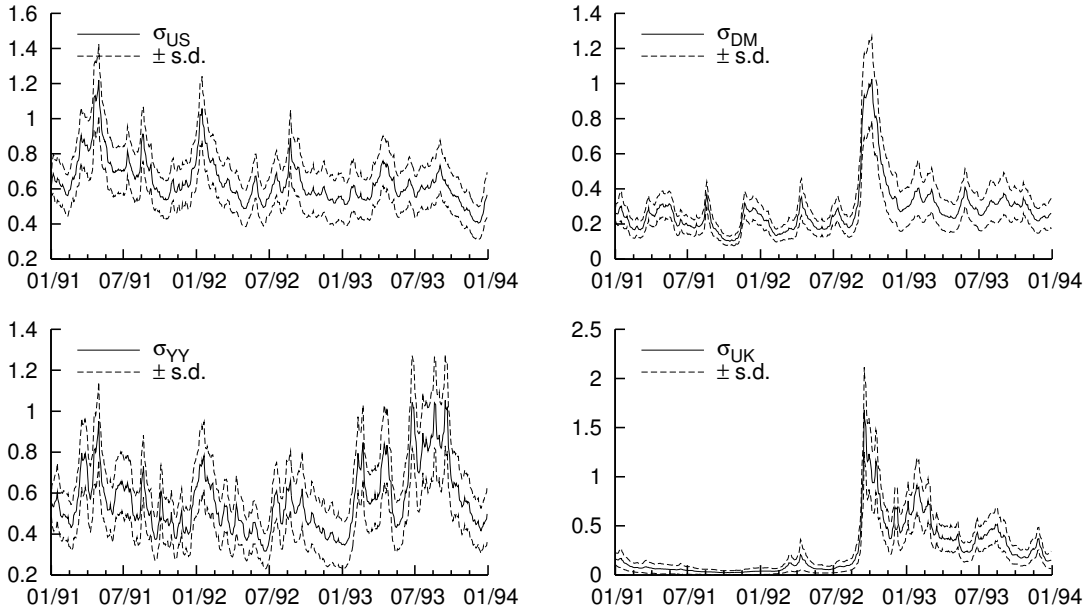

Figure 7: Volatility factors extracted for the currencies over the 1991-1993 period, with 1-standard deviation error bound.

Table 1: Parameters and prior choices

θ	DGP	Prior	p_1	p_2	E(θ)	$\sigma(\theta)$
γ_1	1	$\mathcal{N}(g_1, s_1^2)$	$g_1 = 0$	$s_1 = 2$	0	2
γ_2	-1	$\mathcal{N}(g_2, s_2^2)$	$g_2 = 0$	$s_2 = 2$	0	2
$\sigma_{\omega 1}$	0.05	IG($\alpha_{\sigma 1}, \beta_{\sigma 1}$)	$\alpha_{\sigma 1} = 2$	$\beta_{\sigma 1} = 75$	.1	.05
$\sigma_{\omega 2}$	0.1	IG($\alpha_{\sigma 2}, \beta_{\sigma 2}$)	$\alpha_{\sigma 2} = 2$	$\beta_{\sigma 2} = 75$	.1	.05
ϕ_1	0.95	Beta($\alpha_{\phi 1}, \beta_{\phi 1}$)	$\alpha_{\phi 1} = 9.5$	$\beta_{\phi 1} = 1.5$	.86	.1
ϕ_2	0.90	Beta($\alpha_{\phi 2}, \beta_{\phi 2}$)	$\alpha_{\phi 2} = 9.5$	$\beta_{\phi 2} = 1.5$	.86	.1

The table reports the parameter values used in the DGP, together with the prior density, the choices for the hyperparameters of the prior density in columns p_1 and p_2, and the resulting prior expectation and standard deviation.

Table 2: Inefficiency of the sampling procedures

	$\theta\|\sigma^2$ HM	$\theta_i\|\sigma_i^2$ HM	$\theta\|\sigma^2$ Gibbs	$\theta\|\sigma^2$ ARMS	$\theta\|\sigma^2, \alpha$ HM	$\theta\|u$ HM	$\theta_i\|u_i$ HM	$\theta\|u$ ARMS	$\theta\|u, \alpha$ HM
ϕ_1	358.86	480.55	310.53	113.53	358.86	44.26	54.02	16.38	44.77
ϕ_2	491.82	525.93	401.82	374.11	491.82	26.91	23.06	7.82	29.69
$\sigma_{\omega 1}$	588.65	563.36	419.32	271.36	588.65	55.09	43.67	10.82	59.70
$\sigma_{\omega 2}$	636.61	645.72	415.56	418.44	636.61	29.89	19.08	2.96	17.27
γ_1	172.77	271.11	202.84	93.58	172.77	14.94	16.41	3.49	20.80
γ_2	546.53	538.50	457.69	522.27	546.53	7.69	7.14	2.08	11.11

Table reports the inefficiency measures (Shephard and Pitt 1997) for the different sampling procedures, at a window width of $B_m = 2,000$.

Table 3: Posterior statistics of the Factor-SV model

				Inefficiency			
	$\overline{\theta}$	95% conf		$\theta\|\sigma^2$	$\theta_i\|\sigma_i^2$	$\theta\|u$	$\theta_i\|u_i$
σ_ω(US)	0.206	[0.11,	0.41]	338.93	218.65	46.75	62.99
σ_ω(EU)	0.290	[0.19,	0.42]	108.93	103.65	80.81	53.86
σ_ω(YY)	0.288	[0.16,	0.48]	313.90	231.19	44.11	43.77
σ_ω(UK)	0.410	[0.25,	0.62]	232.11	205.38	120.53	173.26
ϕ(US)	0.920	[0.75,	0.98]	280.95	169.84	55.55	68.68
ϕ(EU)	0.960	[0.92,	0.99]	56.54	61.08	132.37	107.44
ϕ(YY)	0.916	[0.80,	0.98]	256.12	170.31	73.68	79.25
ϕ(UK)	0.986	[0.97,	0.99]	89.83	82.23	367.33	331.51
γ(US)	−0.959	[−1.22,	−0.69]	16.55	28.76	110.43	93.41
γ(EU)	−2.747	[−3.31,	−2.15]	18.66	15.37	125.06	195.98
γ(YY)	−1.283	[−1.62,	−0.96]	34.78	15.38	73.88	87.62
γ(UK)	−3.722	[−5.22,	−2.43]	8.60	5.61	277.75	306.04

Table reports the posterior mean and 95% highest posterior density region of the parameters of the factor model for exchange rates. The righthand panel reports the inefficiency measures at window size 2,000, using HM sampling with the classical ($\theta|\sigma^2$) or reformulated ($\theta|u$) conditioning, either sampling parameters jointly (θ) or separated by region (θ_i).

Table 4: Posterior statistics of the standard RW-SV model

				Inefficiency			
	$\overline{\theta}$	95% conf		$\theta\|\sigma^2$	$\theta_i\|\sigma_i^2$	$\theta\|u$	$\theta_i\|u_i$
σ_ω(EU/US)	0.207	[0.11,	0.39]	303.73	343.03	28.12	54.85
σ_ω(YY/US)	0.238	[0.13,	0.41]	157.92	112.20	84.52	113.23
σ_ω(UK/US)	0.204	[0.11,	0.38]	296.23	219.83	53.00	68.18
ϕ(EU/US)	0.915	[0.76,	0.98]	257.58	339.85	38.64	60.77
ϕ(YY/US)	0.935	[0.84,	0.99]	117.71	79.83	106.04	151.64
ϕ(UK/US)	0.933	[0.81,	0.98]	237.00	245.50	71.26	85.82
γ(EU/US)	−0.610	[−0.85,	−0.34]	32.19	323.00	78.10	119.44
γ(YY/US)	−1.012	[−1.37,	−0.14]	10.74	61.12	133.45	412.11
γ(UK/US)	−0.651	[−0.96,	−0.32]	25.31	139.72	86.86	142.92
σ_ω(EU/US)	0.201	[0.11,	0.38]	303.73	154.35	28.12	43.40
σ_ω(YY/US)	0.241	[0.14,	0.40]	157.92	132.15	84.52	90.06
σ_ω(UK/US)	0.204	[0.11,	0.39]	296.23	238.74	53.00	71.20
ϕ(EU/US)	0.921	[0.78,	0.98]	257.58	136.99	38.64	45.52
ϕ(YY/US)	0.934	[0.84,	0.99]	117.71	99.30	106.04	124.79
ϕ(UK/US)	0.934	[0.81,	0.98]	237.00	215.89	71.26	82.86
γ(EU/US)	−0.610	[−0.86,	−0.35]	32.19	17.66	78.10	95.62
γ(YY/US)	−1.039	[−1.36,	−0.52]	10.74	10.34	133.45	205.59
γ(UK/US)	−0.647	[−0.93,	−0.35]	25.31	25.47	86.86	92.44

See Table 3 for a description of the entries in the table.

Table 5: Likelihood estimates

	Factor-SV	RW-SV
Log likelihood	−1796.81	−2413.14
Log marginal likelihood	−1817.27	−2430.21

Table reports the log likelihood at the posterior mode of the parameters θ, as evaluated with the particle filter, and the logarithm of the marginal likelihood, for the Factor-SV and standard RW-SV models.