

# Inference for adaptive time series models: stochastic volatility and conditionally Gaussian state space form

CHARLES S. BOS

*Vrije Universiteit Amsterdam, De Boelelaan 1105,  
1081 HV Amsterdam, The Netherlands*  
charles.bos@nuf.ox.ac.uk

NEIL SHEPHARD

*Nuffield College, University of Oxford, Oxford OX1 1NF, UK*  
neil.shephard@nuf.ox.ac.uk

June 2003

Preliminary version – Work in progress  
Please do not cite without permission of the authors

## Abstract

In this paper we replace the Gaussian errors in the standard Gaussian, linear state space model with stochastic volatility processes. This is called a GSSF-SV model. We show that conventional MCMC algorithms for this type of model are ineffective, but that this problem can be removed by reparameterising the model. We illustrate our results on an example from financial economics and one from the nonparametric regression model. We also develop an effective particle filter for this model which is useful to assess the fit of the model.

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

## 1 Introduction

### 1.1 The model

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 a Markov chain Monte Carlo (MCMC), are extremely poor and show that a simple reparameterisation overcomes this difficulty delivering reliable methods for inference. This is the main contribution of this paper. We will illustrate the methods by two examples, one from financial econometrics and one from spline based non-parametric regression.

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

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

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

$$R_t \text{diag}(\sigma_t^2) R_t' = \begin{pmatrix} G_t \text{diag}(\sigma_t^2) G_t' & 0 \\ 0 & H_t \text{diag}(\sigma_t^2) H_t' \end{pmatrix},$$

so the errors in the transition and measurement are conditionally independent. When  $\sigma_t^2$  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

$$R_t u_t = \begin{pmatrix} y_t \\ \alpha_{t+1} \end{pmatrix} - \begin{pmatrix} Z_t \alpha_t \\ T_t \alpha_t \end{pmatrix}$$

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

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

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

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

In many models it will be convenient to assume that  $\phi$  and  $\Omega$  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  $\phi$  to the identity and  $\mu$  to a vector of zeros so that

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

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

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

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

The adaptive local level model generalises this to

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

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

**Example 2** The cubic smoothing spline (e.g. Wahba (1978) and Green and Silverman (1994)) for some data  $y_1, \dots, y_n$  finds the function  $f$  with two continuous derivatives which minimise

$$\sum_{t=1}^n \{y_t - f(x_t)\}^2 + \lambda \int_a^b \{f''(u)\}^2 du,$$

where  $\lambda$  is a fixed constant and  $a \leq x_1 \leq \dots \leq x_n \leq b$ . Here the penalty function is indexed solely by  $\lambda$ . We write the value of the function at this minimum as  $\hat{f}(x_t)$ . It is well known (e.g. Wecker and Ansley (1983)) that this function can be found as the posterior mean of the signal  $\alpha_{1t} = (1 \ 0) \alpha_t$ , where writing  $\delta_t = x_t - x_{t-1}$ , from the model

$$y_t | \alpha_t \sim N(\alpha_{1t}, \sigma_1^2), \quad \alpha_{t+1} | \alpha_t \sim N\left(\begin{pmatrix} 1 & \delta_t \\ 0 & 1 \end{pmatrix} \alpha_t, \sigma_2^2 \begin{pmatrix} \delta_t^3/3 & \delta_t^2/2 \\ \delta_t^2/2 & \delta_t \end{pmatrix}\right),$$

where

$$\lambda = \sigma_1^2 / \sigma_2^2.$$

The posterior mean (but not the posterior variance) of the signal  $s_t$  given  $y_1, \dots, y_n$  is invariant with respect to transformations of the parameters which leave  $\lambda$  unchanged. A natural generalisation of this is to an adaptive cubic spline model

$$y_t | \alpha_t, \sigma_t^2 \sim N(\alpha_{1t}, \sigma_{1t}^2), \quad \alpha_{t+1} | \alpha_t, \sigma_t^2 \sim N\left(\begin{pmatrix} 1 & \delta_t \\ 0 & 1 \end{pmatrix} \alpha_t, \sigma_{2t}^2 \begin{pmatrix} \delta_t^3/3 & \delta_t^2/2 \\ \delta_t^2/2 & \delta_t \end{pmatrix}\right).$$

In the adaptive case the optimal estimator of the signal  $\alpha_{1t}$ , the posterior mean  $\tilde{f}(x_t)$ , will have different degrees of smoothness as the variance processes change through time. For these spline models it makes sense to impose a random walk log-volatility model (2) for irregularly spaced data

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

where  $\Omega$  is diagonal.

## 1.2 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 Bos and Koopman (2002) consider the special case where  $\sigma_t^2$  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. ?). 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 (1999). These can be regarded as special cases of the above framework for in these models the  $\alpha_t$  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. 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  $\sigma_t^2$  to have a finite range of values at each time period. This type of model was additionally studied in Kim and Nelson (1999).

## 1.3 Structure of the paper

The organisation of the paper is as follows. In Section 2 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 enormous computational resources in order to deliver correct inferences. In Section 3 we introduce a reparameterisation of the model which vastly improves the algorithm. Section 4 discusses various simulated examples to compare the two algorithms, while Section 5 shows how to effectively implement a particle filter for this method. Section 5 illustrates the method on two real examples, while Section 6 concludes.

## 2 Standard parameterisation

In this paper we will write  $\theta$  as the unknown parameter vector. We often partition  $\theta$  into  $\psi$  and  $\lambda$ , where  $\psi$  indexes parameters in the  $T_t$ ,  $Z_t$  and  $G_t$  matrices, while  $\lambda$  denotes the parameters of the  $\sigma^2$  process.

## 2.1 Conventional block sampling in GSSF-SV models

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  $\sigma^2, \theta$ .
2. Update draw from  $\psi, \alpha|y, \sigma^2, \lambda$  by
  - (a) Sampling from  $\psi|y, \sigma^2, \lambda$
  - (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) and Durbin and Koopman (2002)).
3. Sampling from  $\sigma^2, \lambda|\alpha, y, \psi$  by
  - (a) Sampling from  $\sigma^2|\alpha, y, \theta$
  - (b) Sampling from  $\lambda|\sigma^2, \alpha, y, \psi \equiv \lambda|\sigma^2$
4. Goto 2.

The only non-standard parts of this sampling is the step 3. When  $\sigma_t^2$  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 3a by

- 3'. (a) Sampling from, for  $t = 1, 2, \dots, n$ ,

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

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

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

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

Finally Fruhwirth-Schnatter (1994) has argued that we should replace step 2 in the sampler by working with

2'. Update draw from  $\psi, \alpha|y, \sigma^2, \lambda$  by

(a) Sampling from  $\psi|y, \sigma^2, \alpha, \lambda$

(b) Sampling from  $\alpha|y, \sigma^2, \theta$  using the generic GSSF simulation smoother

This sampler will increase the dependence in the MCMC output, but is likely to be faster to compute as step 2'a is generally much easier to carry out.

## 2.2 Performance of standard parameterisation

### 2.2.1 Simulation design: problem 1

#### Model setup

The performance of the standard formulation can be evaluated using the local level model (3). In some practical situations, the signal-to-noise ratio is smaller than one, which we mimic in the simulation design by choosing the unconditional expectation of the volatility process of the state equation to be  $\mu_2 = -1$ , compared to  $\mu_1 = 1$  for the volatility process of the observation disturbance. The correlations in the volatility processes are put at

$$\phi = \text{diag}(.95, .9).$$

Note that these correlations are modest as compared to values which appear in many applications. For higher correlations, the effects of the SV components will be even more pronounced than what we model here.

For the disturbances of the SV processes, we fix the unconditional standard deviations  $\sigma_{if}$  of both processes at 0.25, with no cross-correlation. This implies a covariance matrix of

$$\Omega = \text{diag}(.078^2, .109^2)$$

#### Prior choice

For  $\mu$  and  $\sigma_{if}$ , conjugate priors are chosen which have the correct mean, and a standard deviation which is little informative. For  $\phi$ , a Beta prior is used to ensure that  $0 < \phi < 1$ , with most mass at large values of  $\phi$ . Table 1 summarises the parameters and their priors densities. Note that the prior of  $\sigma_{i\omega}|\phi_i$  is derived from the priors for  $\sigma_{if}$  and  $\phi$ , applying the appropriate Jacobian.

Table 1: Parameters and prior choices

$\theta$	DGP	Prior	$p_1$	$p_2$	$E(\theta)$	$\sigma(\theta)$
$\mu_1$	1	$\mathcal{N}(m_1, s_1^2)$	1	0.5	1	0.5
$\mu_2$	-1	$\mathcal{N}(m_2, s_2^2)$	-1	0.5	-1	0.5
$\sigma_{if}$	0.25	$\text{IG}(\alpha_\sigma, \beta_\sigma)$	1.3	27	.25	.25
$\phi_1$	0.95	$\text{Beta}(\alpha_\phi, \beta_\phi)$	9.5	1.5	.8	.1
$\phi_2$	0.90	$\text{Beta}(\alpha_\phi, \beta_\phi)$	9.5	1.5	.8	.1

Note that all parameters in this model refer to the volatility processes; the division of the parameter vector  $\theta = (\psi, \lambda)$  has  $\psi = \emptyset, \lambda = \{\mu_1, \mu_2, \sigma_{1f}, \sigma_{2f}, \phi_1, \phi_2\}$ .

### Sampler choices

In the conventional block sampling algorithm of section 2.1 the method of sampling  $\lambda|\sigma^2$  remains to be operationalised.

The most common choice is to use a Hastings-Metropolis-within-Gibbs step, using a Random Walk Metropolis algorithm to sample a new  $\lambda^{(i)}|\sigma^2$ . The candidate covariance matrix of the RW is constructed from the Hessian around the posterior mode of the conditional density  $P(\lambda|y, \alpha, \sigma^2, \psi)$ , with  $\alpha$  and  $\sigma^2$  the values used in the DGP.

Alternatively, a Gibbs sampler can be implemented Kim, Shephard, and Chib (1998, §2.2.1), sampling each of the elements of  $\lambda$  from the full conditionals. In case of the parameters  $\phi_i$  the full conditionals are not available precisely, as the prior is not conjugate. Therefore, sampling  $\phi_i$  from the full conditionals is done through a Hastings-Metropolis step again, using the approximate full conditional as the candidate.

A third possibility, given the fact that the posterior kernel is available in closed form, is to use the ARMS sampler of Gilks, Best, and Tan (1995) and Gilks, Neal, Best, and Tan (1997). This sampler automatically construct approximating densities for all full conditionals, and uses a Metropolis step to draw from these.

### Size of the samples

The simulated data set contains 5,000 observations, to mimic roughly the amount of data which can be expected in financial econometrics when using daily observations for 20 years.

The simulations are carried on to collect a total of 100,000 parameter vectors, after allowing the algorithms a burn-in period of 10,000 iterations. For the Hastings-Metropolis and Gibbs samplers, where the sampling of  $\lambda|\sigma^2, \alpha, y, \psi$  is relatively cheap, this step is repeated 5 times before series of  $\mu$  and  $h$  are sampled.

The Hastings-Metropolis sampler needs to draw from  $\lambda|\sigma^2, \alpha, y, \psi$  and can do this in one

step. However it can be advisable to split the sampler into two, sampling parameters for the first SV process  $\lambda_1|\sigma^2, \alpha, y, \psi$  and for the second,  $\lambda_2|\sigma^2, \alpha, y, \psi$ , separately. This alternative sampler is indicated by the label ‘H-M/Split’.

With the Gibbs sampler in general it is advisable to sample parameters with little cross-correlation. In the model at hand, it seems better to sample from  $\sigma_f$ , the unconditional standard deviation of the SV process, then from  $\sigma_\xi$ , the conditional standard deviation. All main samplers use  $\sigma_f$ , only the sampling results indicated by ‘Gibbs- $\xi$ ’ give alternative results for the Gibbs sampler using the parameterisation in terms of  $\sigma_\xi$ .

The ARMS algorithm constructs a proposal density over a grid, and then performs rejection sampling. Here we use an initial grid of 10 points, refined as necessary by the algorithm. Due to its comparative expense, the ARMS step is not repeated multiple times within one iteration of the full sampler as we do with the other samplers.

### 2.2.2 Performance of the samplers

A major obstacle to using Bayesian methods for models including stochastic volatility is the slow mixing that is generally found in the posterior sample. If the mixing is too slow, the sampler might only very slowly get to the stage of sampling from the true posterior density. As a first impression, the left panel of Figure 1 depicts the posterior density of the parameter  $\phi_1$ , based of the 100,000 drawings from the H-M, H-M/Split, Gibbs, ARMS and Gibbs- $\sigma_\xi$  samplers using the standard parameterisation.

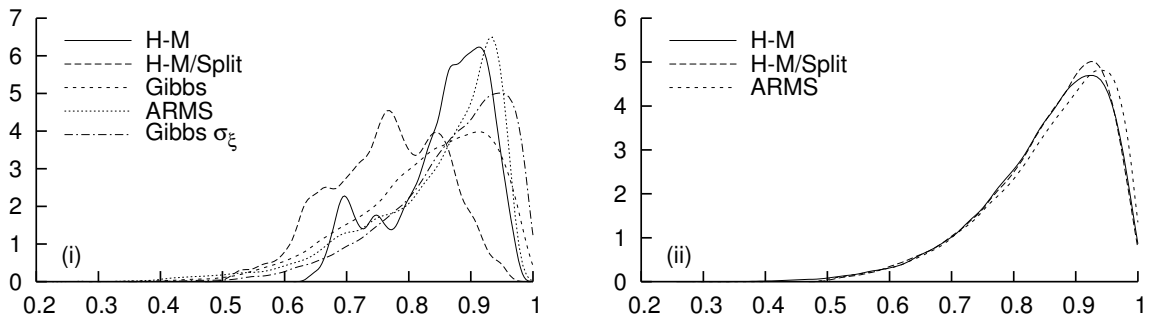


Figure 1: Marginal posterior distribution of parameter  $\phi_1$ , using (i) the standard parameterisation and (ii) using the alternative parameterisation

This graph already indicates that the samplers did not converge; continuing the samplers for 10,000,000 iterations (results not reported in the paper) does not change these results.

The problem with these samplers indeed lies in the mixing within the chains: Autocorrelation between successive drawings is high. Table 2 reports the 30th autocorrelations of the parameters using each of the samplers. The message of these correlations is consistent with the previous

density plot: Correlation remains high, even after 30 iterations. Note that only with the Gibbs samplers, correlation seems to be decreasing slightly quicker than with the other samplers.

The table also reports the estimated integrated autocorrelation times or inefficiency factors. These were highlighted in 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.

Table 2: Posterior correlation and simulation inefficiency

	H-M		H-M/Split		Gibbs		ARMS		Gibbs- $\sigma_\xi$	
	$\rho_{30}$	Ineff	$\rho_{30}$	Ineff	$\rho_{30}$	Ineff	$\rho_{30}$	Ineff	$\rho_{30}$	Ineff
$\phi_1$	0.990	2888.4	0.984	2215.8	0.804	369.9	0.976	1897.2	0.774	190.4
$\phi_2$	0.985	2792.6	0.979	2671.3	0.890	498.8	0.965	1580.7	0.881	458.5
$\mu_1$	0.973	2884.6	0.955	1902.2	0.860	1693.6	0.964	2439.6	0.774	1743.3
$\mu_2$	0.710	1411.5	0.738	1412.0	0.430	324.3	0.645	492.5	0.328	396.5
$\sigma_f/\xi_1$	0.992	3088.7	0.983	2300.6	0.724	1793.0	0.989	2668.1	0.954	2124.9
$\sigma_f/\xi_2$	0.962	2558.8	0.963	2525.6	0.513	551.8	0.929	1446.9	0.956	1474.7
Time	1:46		2:23		1:52		10:11		1:52	

The table reports the 30th order autocorrelations of the sample and the measure of simulation inefficiency, for each of the parameters, with the timing of the samplers in hours. Inefficiency measures are computed using  $B_m = 2,000$ .

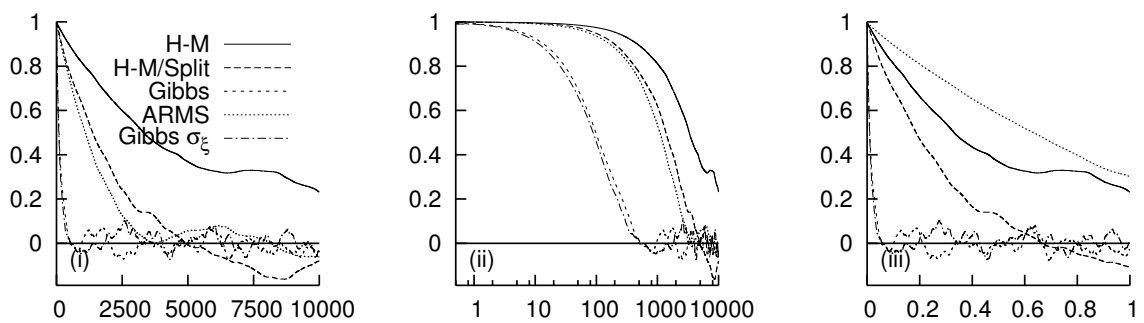


Figure 2: Autocorrelation (i) of the sample using the standard parameterisation, (ii) on a log-scale and (iii) on a time scale, taking the computational effort into account.

Figure 2 displays the autocorrelation in the sample for the parameter  $\phi_1$  in the samples, between lags 1 and 10,000. The panels display the correlations on the standard scale (i), on the log-scale (ii) and on a time scale (iii), taking the timing of the algorithms in the last row of Table

2 into account. Only the Gibbs algorithms seem to deliver (slowly) diminishing correlations. Accounting for the computational effort only changes the ranking of the ARMS algorithm as compared to the others.

### 3 Reformulation

We have seen that even if the sampling from  $\sigma^2|\alpha, y, \theta$  is carried out in a very effective way the performance of the overall sample is very poor. The reason for this is that, especially for longer data series, the information contained in the variances  $\sigma_t^2, t = 1, \dots, T$  on the parameters in the SV process is precise. Conditional on the variances  $\sigma_t^2$ , the density of  $\lambda|\sigma^2, y, \alpha, \psi$  allows for little movement between successive draws of  $\lambda$ , leading to slow mixing of the chain.

Here we reparameterise the problem in terms of the errors in the SV component of the model.

#### 3.1 Disturbance based block sampling in GSSF-SV models

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

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

Note that there is a one-to-one relation between the volatility process  $\sigma_t$  (and hence  $h_t$ ) and the  $NID(0, I)$  disturbances  $u_t$ . Therefore, the conditioning in the block sampler can also be done on  $u_t$ , which by construction contains little or no information on the value of the parameters.

The sampling algorithm now becomes:

1. Initialise  $u, \theta$ , and compute  $\sigma^2 = f(u, \theta)$  as a function of  $u$
2. Update draw from  $\theta, \alpha|y, u$  by
  - (a) Sampling from  $\theta|y, u$ .
  - (b) Sampling from  $\alpha|y, \sigma^2(u, \theta), \theta$  using the generic GSSF simulation smoother (Fruhwirth-Schnatter (1994), Carter and Kohn (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002)).
3. Recompute  $\sigma^2$  from  $u$  and  $\theta$ .
4. Sample from  $\sigma^2|\alpha, y, \theta$ .
5. Recompute  $u$  from  $\sigma^2$  and  $\theta$ .
6. Goto 2.

Notice that step 2 is subtly different from in the previous section for now sampling from  $\theta|y, \sigma^2, u$  updates all of the parameters in the model. The split into  $\theta = (\psi, \lambda)$  makes less sense here as the full conditional  $\lambda|y, u, \phi$  does not simplify any further as it did before.

Each of these steps is relatively easy to carry through. The important point here is that step 2a has considerably changed, for 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.

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 (2003). 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.

### 3.2 Performance of reformulation

The same simulation design as described in section 2.2.1 was used, where the reformulation of the model in terms of the disturbances was used.

The reparameterisation of the sampler has several effect. First of all, the conditional densities of the Gibbs sampler are no longer readily available. The densities now would have to comprise both the likelihood function of the SV model (1), the prior of the parameter, and also the transformation from  $u$  to  $\sigma^2$  and the likelihood of the GSSF model. This last likelihood is only available in closed form if we follow Frühwirth-Schnatter (1994) conditioning on the state again. Even so, the densities are of a highly nonlinear functional form, from which no simple sampling scheme is known.

The alternative is to use the ARMS sampler. This sampler uses a higher number of function evaluations of the posterior kernel in to reconstruct an approximation to the full conditional densities. As each function evaluation requires a filter to construct  $\sigma^2$  from  $u$  and  $\lambda$ , the computational effort of this sampler also increases considerably as compared to the situation in the standard formulation of the model.

The other option investigated before was using the Hastings-Metropolis sampler. As this sampler uses no more than 2 function evaluations per iteration, the computational load does not increase too much by having to filter back and forth between  $u$  and  $\sigma^2$ . Therefore, this is the most practical method to use on the model at hand. Again, also a Hastings-Metropolis algorithm is used where the sampling of  $\lambda$  is split between the parameters of the two SV processes.

The first results using these samplers are found in the second panel of Figure 1 above. The three samplers correspond closely in their estimate of the posterior density of  $\phi_1$ , which is already

a clear sign of better behaviour of the samplers.

Table 3: Posterior correlation and simulation inefficiency, using the transformed sampler

	H-M		H-M/Split		ARMS	
	$\rho_{30}$	Ineff	$\rho_{30}$	Ineff	$\rho_{30}$	Ineff
$\phi_1$	0.320	63.9	0.215	49.0	0.127	40.8
$\phi_2$	0.734	337.5	0.656	219.0	0.586	223.2
$\mu_1$	0.054	12.0	0.053	18.2	0.050	7.4
$\mu_2$	0.272	87.0	0.276	140.5	0.265	53.4
$\sigma_{f1}$	0.233	36.7	0.260	68.3	0.224	47.2
$\sigma_{f2}$	0.513	140.5	0.509	182.0	0.452	93.7
Time	3:04		5:03		19:02	

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

Table 3 displays the correlation and simulation inefficiency statistics for the H-M and ARMS samplers. These statistics indeed show the strongly increased quality of the samplers. The message springs from Figure 3 which shows the autocorrelation of the H-M, H-M/Split and ARMS samplers, as compared to the autocorrelation in the Gibbs sampler of the standard formulation (copying part of Figure 2, for lags 1-1,000). In the figure, it is clear how the lower correlation in the ARMS sampler is offset by the larger computational effort involved, and the basic and split H-M samplers perform better.

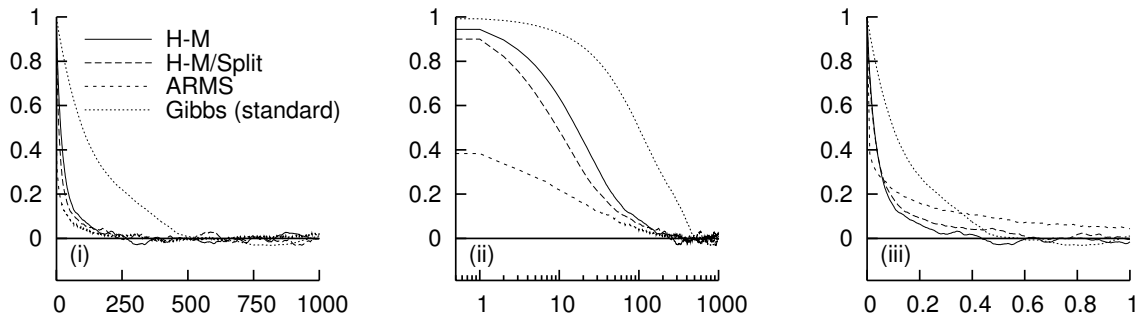


Figure 3: Autocorrelation (i) of the sample comparing the standard parameterisation with the Gibbs sampler to the reformulation and the H-M sampler, (ii) on a log-scale and (iii) on a time scale, taking the computational effort into account.

The message from the statistics and the graphs is clear: With a simple reformulation of the model, the sample correlation drops strongly, with a higher efficiency of the final sample as a result.

## 4 Particle filtering

An important feature of MCMC is that it produces samples from  $\alpha, \sigma^2, \theta | y$  and so samples from  $\alpha, \sigma^2 | y$ . 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

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

Such quantities are very important in practise for the use of sequential forecasting and model checking. A standard way of carrying this out 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 collections of particles, which are used to approximate the distribution of  $\alpha_t, \sigma_t^2 | \mathcal{F}_t$ ,

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

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

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

a mixture of normals. This implies that

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

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

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

with

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

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

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

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

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

The most important step is that we now calculate

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

Straightforward calculations show that

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

where

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

and

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

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

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

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

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

## 5 Illustrations

### 5.1 Local level with SV effects

#### 5.1.1 Modelling exchange rates

When modelling the floating exchange rates between  $n + 1$  countries, in the literature most of the attention is focused on the exchange rate vis-a-vis the dollar. The logarithm of the exchange is found to roughly follow a random walk, with possible changes in the variance of the process over time.

With multiple exchange rates, the correlation structure between the disturbance terms can be intricate, as the exchange rates together form a system, with possible correlations between all cross rates.

To simplify the model and the correlation structure, Mahieu and Schotman (1994) propose a factor structure for the log-exchange rate  $s_{ij,t}$  between countries  $i$  and  $j$  as

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

where the assumption is made that  $e_{i,t} \perp e_{j,t}$ , if  $i \neq j$ .

Using country 0 as a numeraire, a model for multiple exchange rates between  $n + 1$  countries could be

$$\begin{aligned} \begin{pmatrix} y_{10,t} \\ \vdots \\ y_{n0,t} \end{pmatrix} &= (-\mathbf{1} \quad \mathcal{I}_n) \begin{pmatrix} e_{0,t} \\ \vdots \\ e_{n,t} \end{pmatrix} \\ \begin{pmatrix} e_{0,t+1} \\ \vdots \\ e_{n,t+1} \end{pmatrix} &= \begin{pmatrix} e_{0,t} \\ \vdots \\ e_{n,t} \end{pmatrix} + H_t \epsilon_t \\ \epsilon_t &\sim N(0, \mathcal{I}_n) \end{aligned}$$

where  $H_t$  is a diagonal matrix with typical element

$$\begin{aligned} H_{ii,t} &= \exp(h_{i,t}) \\ h_{i,t+1} &= h_{i,t} + \sigma_{i,\xi} \xi_{i,t} \\ \xi_t &\sim N(0, \mathcal{I}_n). \end{aligned}$$

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 its level of variance by itself.

Alternatively, the model can be estimated without stochastic volatility by fixing all  $\sigma_{i,\xi} \equiv 0$ .

### 5.1.2 Data and estimability

The proposed model contains, for  $k$  exchange rates of length  $T$ ,  $k + 1$  unobserved factor components of length  $T$ , plus the  $k + 1$  volatilities which are second order unobserved processes. Essentially the SV processes serve to estimate the  $k + 1$  parameters  $\sigma_{i,\xi}$ , which can be expected to be considerably hard given the little degree of information available on these parameters.

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

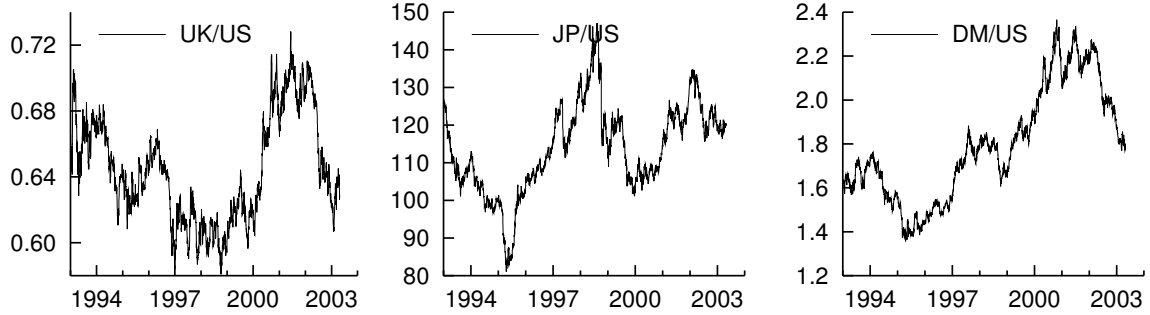


Figure 4: Exchange rates of the British Pound, Japanese Yen and German DMark against the US Dollar

To keep the estimations tractable, we concentrate on the three exchange rates of the US Dollar against the British Pound, the Japanese Yen, and the German DMark (quietly switched for the Euro at the start of 1999), with data over the period 1993/01/04–2003/04/29,<sup>1</sup>. This period contains 2610 daily observations and is depicted in Figure 4. 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  $j$ .

## 5.2 Results

The model as presented above was estimated using the simulation techniques described earlier. After a burn-in period of 10,000 iterations a million sampled parameters were collected. The priors of the standard deviations in the SV processes were Inverted Gamma-1 densities, with expectation and standard deviation equal to 0.2.

Table 4: Posterior means and variances of the factor model

	No SV		No transf.		With transf.	
	$\bar{\sigma}_\epsilon$	$\sigma(\sigma_\epsilon)$	$\bar{\sigma}_\xi$	$\sigma(\sigma_\xi)$	$\bar{\sigma}_\xi$	$\sigma(\sigma_\xi)$
$\sigma$ USD	0.123	(0.009)	0.202	(0.02)	0.210	(0.02)
$\sigma$ Pnd	0.893	(0.013)	0.299	(0.03)	0.300	(0.05)
$\sigma$ JY	9.491	(0.131)	0.284	(0.02)	0.286	(0.03)
$\sigma$ DM	1.159	(0.016)	0.091	(0.01)	0.091	(0.01)

The results are summarised in Table 4 and figures 5–7, with posterior moments, densities, autocorrelations and the underlying SV factors.

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 factors the unobserved SV processes are estimated, from which in turn the disturbances are extracted to estimated their standard deviations  $\sigma_\xi$ .

<sup>1</sup>Source: <http://pacific.commerce.ubc.ca/xr/>.

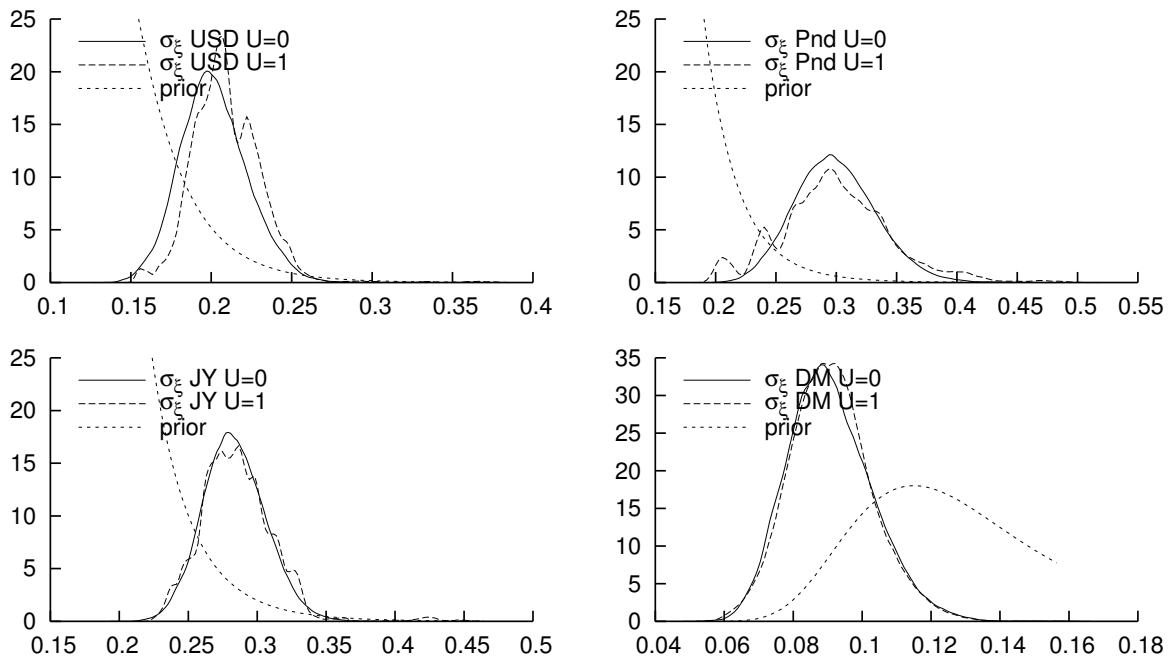


Figure 5: Posterior density of SV variance parameters for the factor model of exchange rates, without and with the transformation

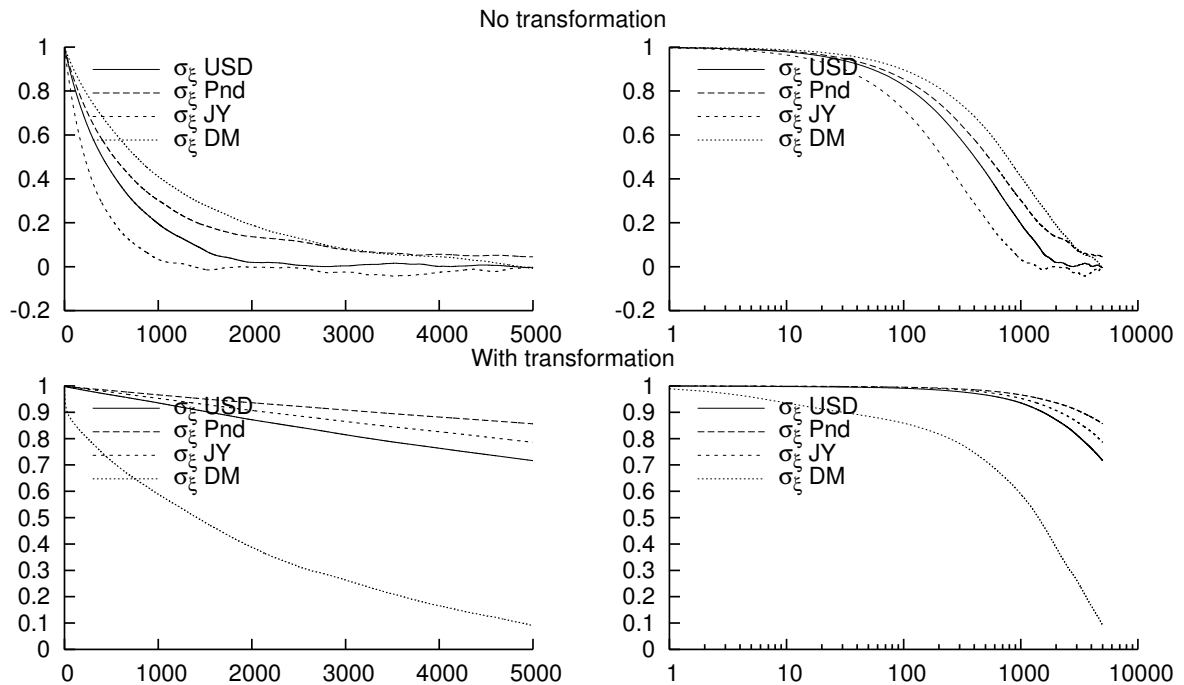


Figure 6: Autocorrelation of the sample of SV variance parameters for the factor model of exchange rates, without and with the transformation

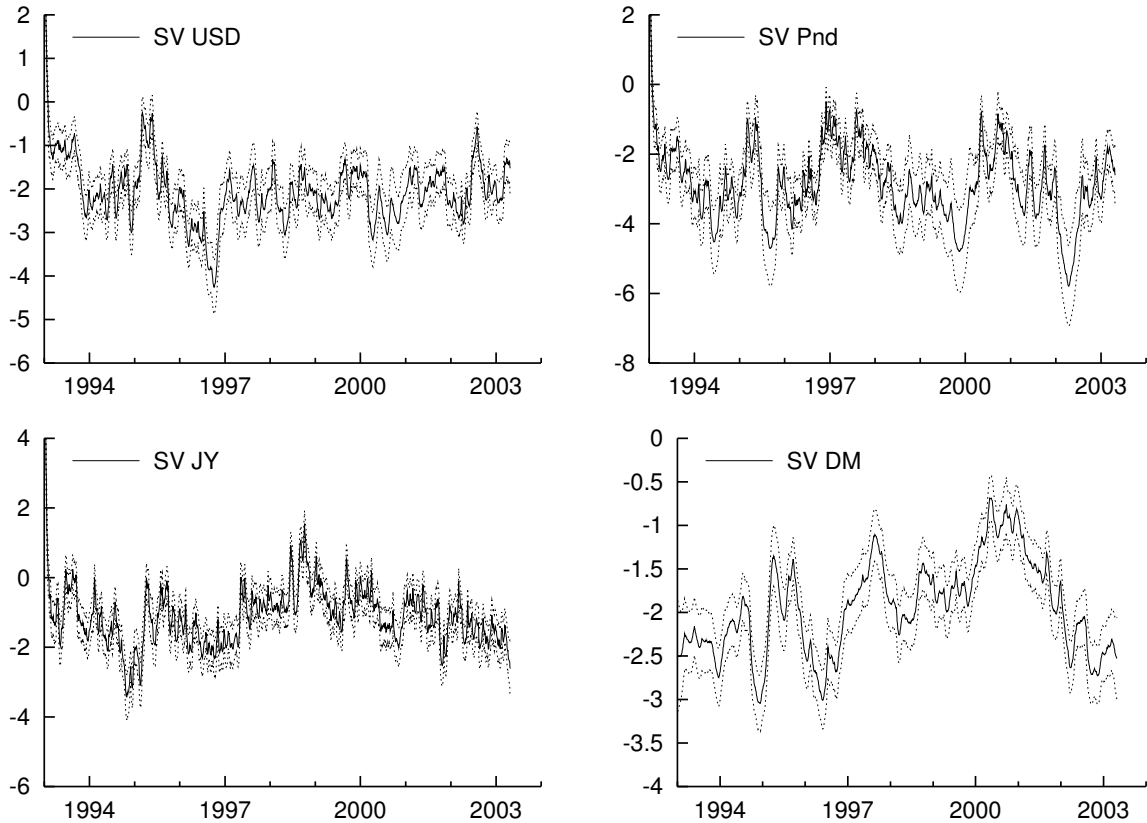


Figure 7: Stochastic volatility for the different factors

Both the method with and without the transformation result in the same posterior density, as should be the case. Note however that here the sampler with the transformation displays stronger correlation; in more extensive research not reported here it is found that this can occur for longer data series with integrated SV processes. Additional methods for breaking the correlation can be applied.

The SV factors as displayed in Figure 7 can lead to further research as to the correlation in the volatility processes between the regions. At the moment, the model assumes independent SV processes, but further a more elaborate model can be improve estimation results.

The SV factors indicate how 1996 was a year with relatively high volatility in the exchange rate markets, especially for the US Dollar. The factor pertaining to the dollar however did not seem to react to the occurrences of September 2001, at least not that the volatility jumped up. The linking of DMark into the Euro in January 1999 did not influence uncertainty much; only before the introduction of the Euro in 2001 the uncertainty peaked. This uncertainty did however not last long, and in the mean time the volatility of the Euro reached an all-time low again.

The model at hand is easily generalised to even higher more exchange rates. Multivariate

modelling of correlations in levels and variance processes can lead to a deeper understanding of the driving forces in the currency markets.

### 5.3 Regression spline

#### 5.3.1 Estimation

In this section we reanalyse the classic Silverman (1985) nonparametric regression analysis of the Schmidt, Mattern, and Schueler (1981) motorcycle dataset. The dataset consists of observations on the acceleration of the head in simulated motorcycle crash experiments, where the efficacy of a helmet in preventing serious head injury is investigated. The dataset is non-equally spaced, with multiple observations for some time periods. Also, the variability of the series is rather different between pre-, in- and post-crash periods.

Both Silverman (1985) and Harvey and Koopman (2000) fitted a cubic spline to the data. Observing the difference in variability between time periods, they adapt the weights of the observations to allow for varying variance, through a rather ad-hoc procedure. With the set-up presented in this article, the analysis can however be done in a Bayesian manner, with the stochastic volatility being a fully integrated part of the model.

In Example 2 the model was presented. We denote the standard Gaussian spline model by **0 SV**, the version with stochastic volatility on the observation variance  $\sigma_{1t}^2$  as **1 SV** and with stochastic volatility in both the observation and transition equation as **2 SV**. Prior densities are data-based, to simplify matters, in the sense that they were chosen with classical estimation results in mind. Table 5 gives the parameters of the Inverted Gamma-1 prior densities on the standard deviations in the model.

Table 5: Parameters for IG-1( $\alpha, \beta$ ) prior densities for the cubic spline model

	$\alpha$	$\beta$	$E(\theta)$	$\sigma(\theta)$	Equation
$\sigma_1$	1.5	2	2.5	1.9	Observation
$\sigma_2$	1.4	4.3	0.6	0.5	Transition
$\sigma_{\xi,1}$	2	4.5	0.4	0.2	SV Observation
$\sigma_{\xi,2}$	2	4.5	0.4	0.2	SV Transition

Using the transformation of the model and the Gibbs sampler, as explained in section 3, the posterior density of the parameters was derived. Table 6 reports the posterior means for the parameters, for the three specifications of the variance processes. For each model the sampler was run for 1,000,000 iterations. A burn-in period of 10,000 iterations was sufficient, even though the starting point for the sampler was far away from the posterior mode.

For the model with fixed variance, a high observation standard deviation  $\sigma_1$  is estimated, as at least some periods have a higher variation. This is seen in Figure 8, where the estimates for

Table 6: Posterior means of parameter estimates

	<b>0 SV</b>			<b>1 SV</b>			<b>2 SV</b>		
	$\bar{\theta}$	$\sigma_{\theta}$	$\rho_{30}$	$\bar{\theta}$	$\sigma_{\theta}$	$\rho_{30}$	$\bar{\theta}$	$\sigma_{\theta}$	$\rho_{30}$
	No transformation								
$\sigma_1$	20.636	1.64	0.02	<i>18.751</i>			<i>18.239</i>		
$\sigma_2$	0.622	0.12	0.03	0.571	0.11	0.05	<i>0.788</i>		
$\sigma_{\xi,1}$				0.338	0.07	0.56	0.337	0.07	0.53
$\sigma_{\xi,2}$							0.641	0.16	0.72
	With transformation								
$\sigma_1$				<i>18.557</i>			<i>18.216</i>		
$\sigma_2$				0.571	0.11	0.05	<i>0.719</i>		
$\sigma_{\xi,1}$				0.334	0.06	0.91	0.317	0.06	0.91
$\sigma_{\xi,2}$							0.353	0.09	0.89

Posterior mean and standard deviation of the parameters, and the 30-th order autocorrelation, in the models with 0, 1 and 2 SV components. The first panel presents results not using the transformation, the second panel applies the transformation from the SV process to the disturbances. Values in *italics* are average values for the standard deviations of the observation and transition equation as implied by the respective SV processes.

the 5, 50, and 95% quantiles of the posterior of the spline and the spline growth are depicted. In order to allow for the high variability in the middle part of the series, the interquantile range is large throughout the sample.

This figure mimics similar results in Silverman (1985) and Harvey and Koopman (2000) for the Gaussian case, with clear indication that the variance should be allowed to take lower values especially in the earlier and later parts of the data series.

Allowing for stochastic variance on the observation equation, in model **1 SV** in the second pair of columns of Table 6, the average standard deviation  $\sigma_1$  as implied by the SV model is estimated at 18.6 (in *italics*), only slightly lower than in the model without SV. However, Figure 9 plots the effective standard deviation for the observation in the second panel, and it is seen that the variability is concentrated in the middle part of the sample. The figure replicates roughly results of the aforementioned authors where they applied an adapted weighting scheme for the observations. The difference with the results presented here, is that here the construction of the results is based entirely on a probabilistic model, and hence these results are less ad-hoc.

The third model adds another stochastic variance to the growth component in panel (ii) of Figure 8. The growth component seems to be constant at zero for the first time periods, followed by swift movements until period 40, after which the movements seem to die down again. The third set of columns in Table 6 allow for such behaviour by introducing SV on the transition equation as well. The standard deviation of this SV component  $\sigma_{\xi,2}$  is estimated at values even

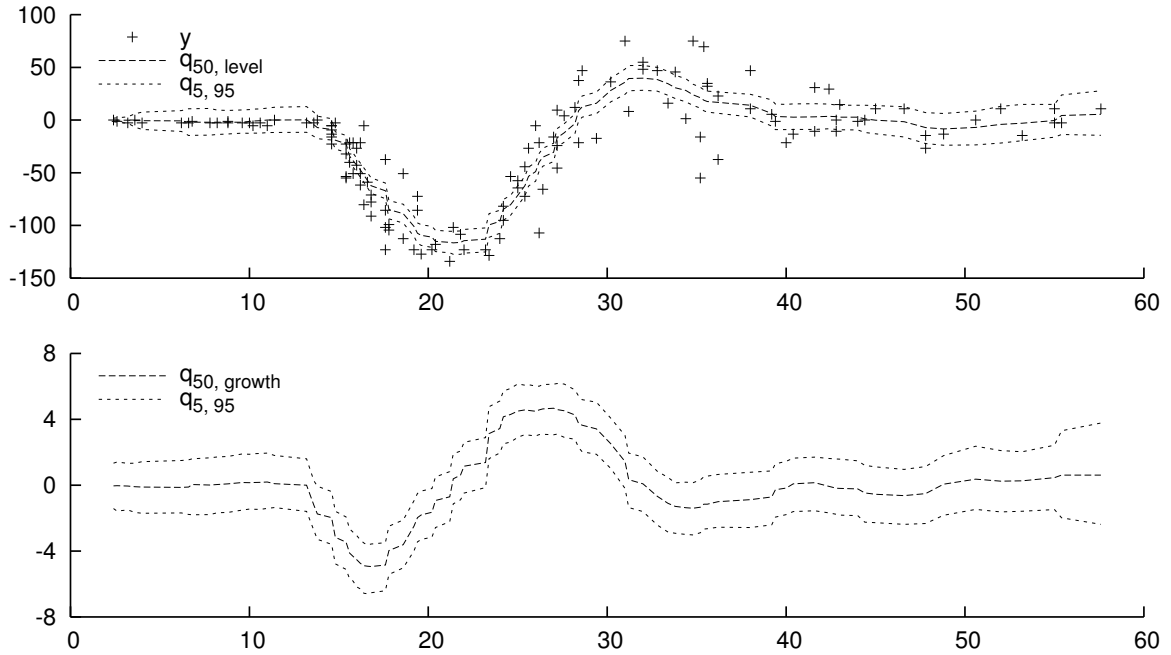


Figure 8: Cubic spline level (i), growth (ii) with quantiles, for the Gaussian model

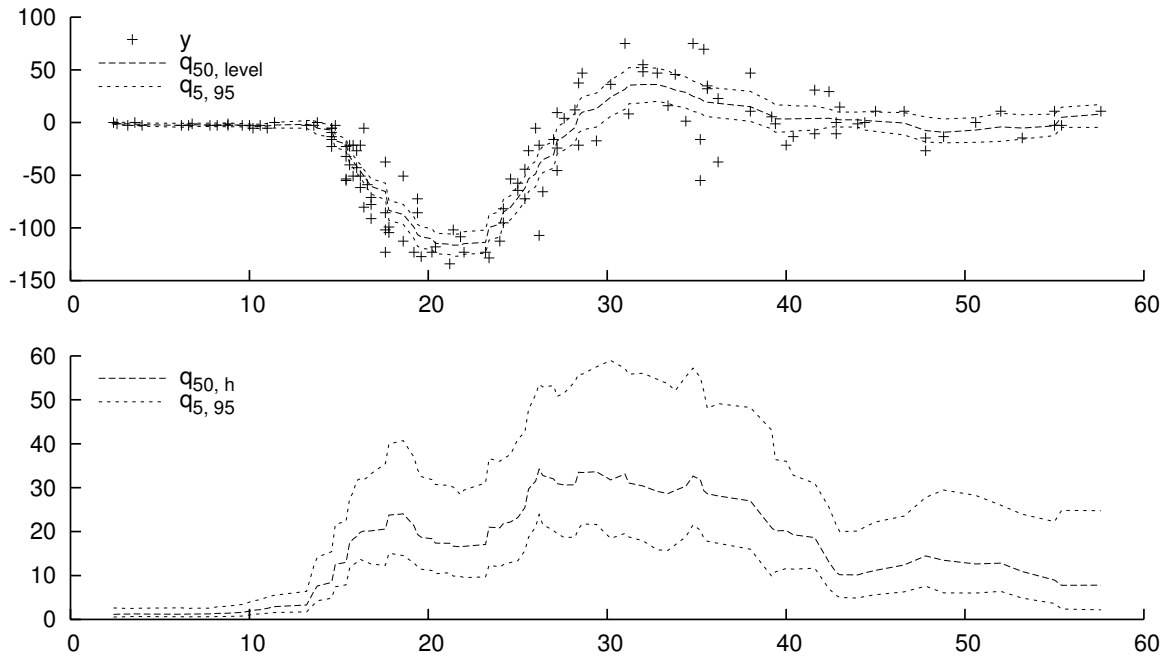


Figure 9: Cubic spline level (i) and observation standard deviation (ii) with quantiles, for the model with one stochastic variance

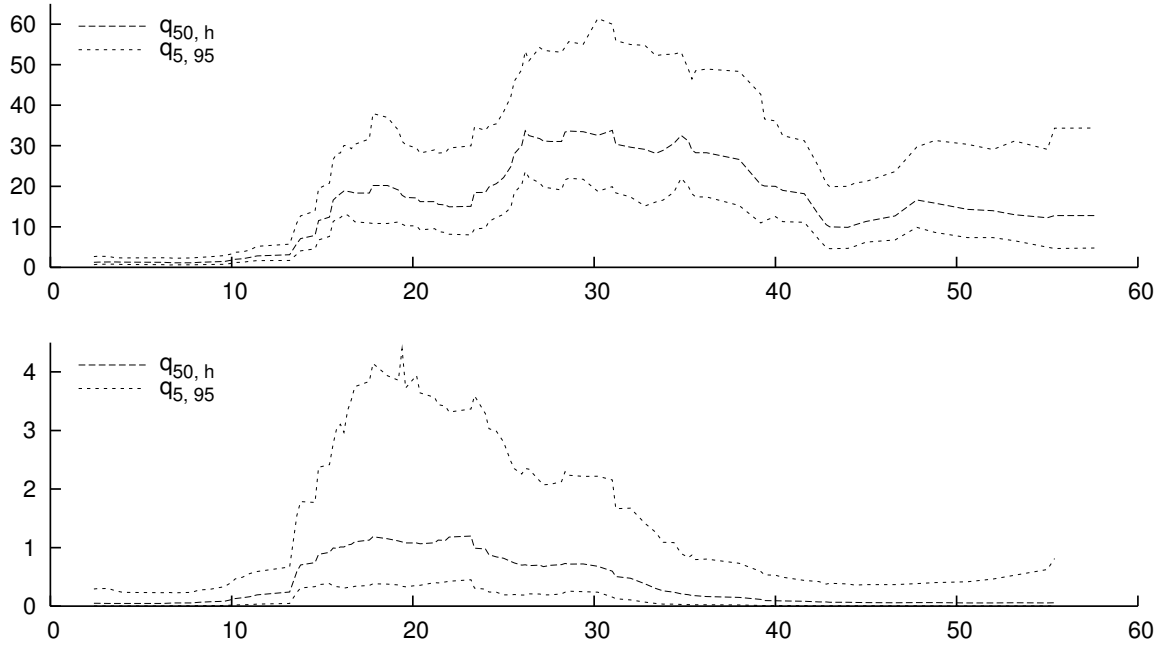


Figure 10: Observation standard deviation (i), transition standard deviation (ii) with quantiles, for the model with double stochastic variance

larger than  $\sigma_{\xi,1}$ , implying that there is more variability here. The second panel of Figure 10 displays the evolution of the variance process over time: At the start and end of the sample, the variability of the growth component is approximately zero, with in the middle positive variance, though the uncertainty concerning the variability is large.

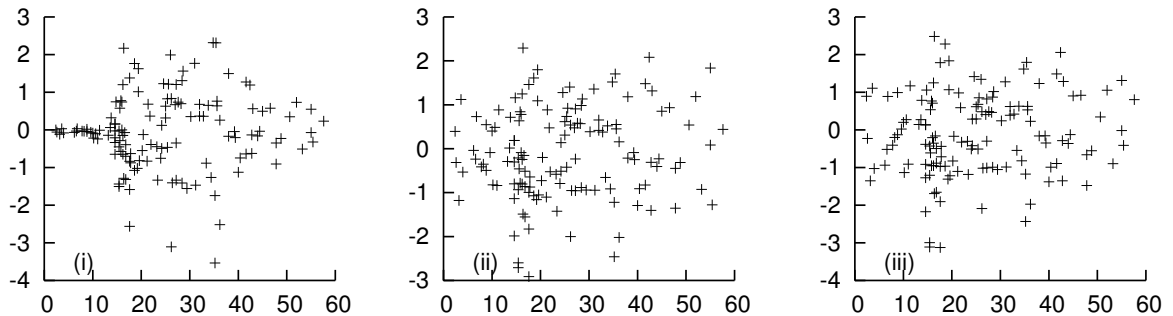


Figure 11: Average residuals for the model without SV (i), with one SV component (ii) and with two SV components (iii)

Adapting the variances throughout the sample leads to a large improvement in the distribution of the residuals of the model. Figure 11 displays the average smoothed standardised residuals. Without SV, in panel (i), the residuals clearly display heteroskedasticity, which has disappeared from panels (ii) and (iii) for models **1 SV** and **2 SV**.

Figures 8–11 are based on posterior quantities using the full data sample in the sampling

Table 7: Likelihood measures for the cubic spline models

	0 SV		1 SV		2 SV	
Loglikelihood	-433.85		-410.41		-408.07	
Log-Marginal Likelihood	-445.99		-412.18		-409.23	
Box-Ljung $Q_u, \chi^2_{10-k}$	24.89	[0.0016]	13.30	[0.1021]	8.48	[0.3879]
Box-Ljung $Q_v, \chi^2_{10-k}$	80.18	[0.0000]	6.89	[0.5487]	4.51	[0.8087]

Loglikelihood at the posterior mean, logarithm of the marginal likelihood, and the Box-Ljung statistic (with  $p$ -values) testing for autocorrelation in the  $u$  and  $v$  statistics.

algorithm. Another comparison of models is given by the marginal likelihood, given in logarithmic form in Table 7, together with the loglikelihood of the models in the posterior mean. These measures can be derived using the particle filter technique as described earlier.

According to both measures, the Gaussian model indeed fits considerably worse than the models with one or two SV components. The log-marginal likelihood of the SV 2 model is 2.95 points better than for the SV 1 model; according to Kass and Raftery (1995), values between 1 and 3 indicate ‘positive’ evidence for the alternative model, and a value of 3 would give ‘strong evidence’. Clearly there is something to say for the extra SV component in the model, though the data set is not strongly informative on this point.

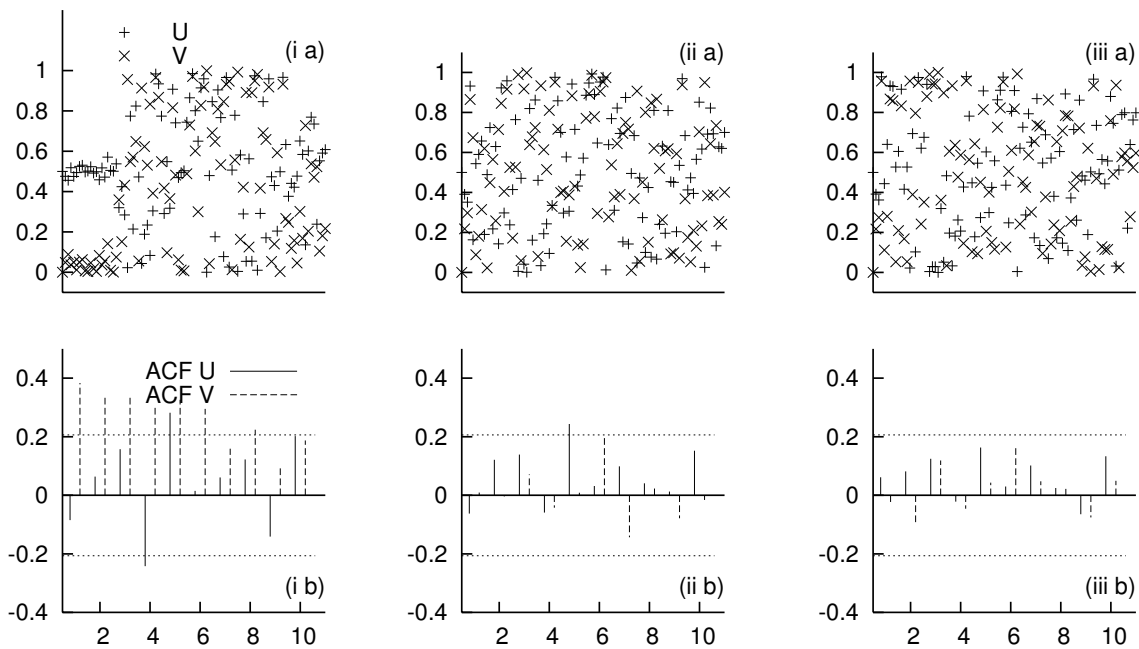


Figure 12: Probabilities  $u_t$  of observations and transformations  $v_t$  against the index, with auto-correlations

The second panel of the table is concerned with statistics  $u$  and  $v$  defined as

$$\begin{aligned} u_t &= \Pr(Y_t < y | \mathcal{F}_{t-1}), \\ v_t &= 2 \left| u_t - \frac{1}{2} \right|, \end{aligned}$$

where the probability is calculated integrating out the parameters of the model. These statistics would ideally be distributed i.i.d.  $U(0, 1)$ . Figure 12 displays the values of  $u_t$  and  $v_t$  plotted against the index, and it is obvious that there is some correlation in both series when the stochastic volatility is not modelled. With either one or two SV components, the correlation diminishes. The Box-Ljung statistics for  $u$  and  $v$  are calculated using  $\sqrt{T} \approx 10$  lags for the data set with multiple observations at the same time period replaced by their average, with a higher weight. The number of degrees-of-freedom is  $10 - k$ , where  $k = 2$  is the number of parameters in the model. For the model without SV, the hypothesis that  $v$  is uncorrelated is strongly rejected. With either one or two SV components, both tests do not reject the null of uncorrelated  $u$  or  $v$ .

## 6 Conclusion

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 and so 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.

## 7 Acknowledgements

Neil Shephard's research is supported by the UK's ESRC through the grant "Econometrics of trade-by-trade price dynamics," which is coded R00023839. 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 benefitted greatly from many discussions with Jurgen Doornik, Thomas Kittsteiner 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.

- Ameen and J. Harrison (1984). Discounted weighted estimation. *Journal of Forecasting* 3, 285–296.
- Bos, C. and S. J. Koopman (2002). Time series models with a common stochastic variance for analysing economic time series. Tinbergen Institute discussion paper, TI2002-113/4, Netherlands.
- 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 (1999). Analysis of high dimensional multivariate stochastic volatility models. Unpublished paper: Nuffield College, Oxford.
- de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika* 82, 339–50.
- 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.
- Fruhworth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis* 15, 183–202.
- Frühwirth-Schnatter, S. (2003). Efficient bayesian parameter estimation for state space models based on reparameterizations. Unpublished manuscript.
- 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.

- Glhysels, 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, 155–73.
- Gilks, W. R., R. M. Neal, N. G. Best, and K. K. C. Tan (1997). Corrigendum: adaptive rejection Metropolis sampling. *Applied Statistics* 46, 541–542.
- 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.
- Green, P. and B. W. Silverman (1994). *Nonparameteric Regression and Generalized Linear Models: A Roughness Penalty Approach*. London: Chapman & Hall.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Harvey, A. C. and S. J. Koopman (2000). Signal extraction and the formulation of unobserved components models. *Econometrics Journal* 3, 84–107.
- 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., N. Shephard, and J. A. Doornik (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Economic Journal* 2, 107–160.
- 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.
- Schmidt, G., R. Mattern, and F. Schueler (1981). Biomechanical investigation to determine physical and traumatological differentiation criteria for the maximum load capacity of head and vertebral column with and without protective helmet under the effects of impact. EEC Research Program on Biomechanics of Impacts Final report, Phase III, Project G5, Institut für Rechtsmedizin, University of Heidelberg, West Germany.
- 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.
- Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society, Series B* 47(1), 1–52.
- Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica* 65, 59–73.
- Wahba, G. (1978). Improper priors, spline smoothing, and the problems of guarding against model errors in regression. *Journal of the Royal Statistical Society, Series B* 40, 364–372.
- 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.