Inference for Adaptive Time Series Models

Charles S. Bos[1]
Neil Shephard[2]

[1] *Faculty of Economics and Business Adminstration, Vrije Universiteit Amsterdam, and Tinbergen Institute,*
[2] *Nuffield College, University of Oxford, Oxford, UK.*

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

CHARLES S. BOS

Tinbergen Institute and Vrije Universiteit Amsterdam, De Boelelaan 1105,
1081 HV Amsterdam, The Netherlands
cbos@feweb.vu.nl

NEIL SHEPHARD

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

January 2004

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 spline, 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), can be extremely 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. The paper proposes a method which can be applied such that an improvement in the sampler can be expected in most cases. These are 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 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

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

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 μ 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

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

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 $q = \sigma_2^2 / \sigma_1^2$. Hence 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.

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

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

where λ is a fixed constant and $a \le x_1 \le \ldots \le x_n \le b$. Here the penalty function is indexed solely by λ. We write the value of the function at this minimum as $\widehat{f}(x_t)$. 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

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

where

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

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

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

In the adaptive case the optimal estimator of the signal α_{1t}, the posterior mean $\widetilde{f}(x_t)$, 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 Ω 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 (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 (1999). 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).

1.3 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. Sections 2.2 and 3.2 discuss various simulated examples to compare the two algorithms, while Section 4 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 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 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 σ^2, θ.

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 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 3a by

3'. (a) 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

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

Proposals can be made from this density, either using many univariate draws (which seems always a good idea if Ω and ϕ 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

Simulation design

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

$$\phi = diag(.95, .9).$$

Note that these correlations are modest as compared to values which appear in many applications. We would expect the strengths of the point we make using this simulation to become even more important with stronger correlations in the volatility processes.

For the disturbances of the SV processes, we fix the unconditional standard deviations $\sigma_{if} = \sigma_{i\omega}/\sqrt{1 - \phi_i^2} = 0.25$ for both processes, with no cross-correlation.

Prior choice

For μ and σ_{if}, conjugate priors are chosen which have the correct mean, and a standard deviation which is little informative. For ϕ, a Beta prior is used to ensure that $0 < \phi < 1$, with most mass at large values of ϕ. 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 σ_{if} and ϕ, applying the appropriate Jacobian.

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

Table 1: Parameters and prior choices

θ	DGP	Prior	p_1	p_2	E(θ)	$\sigma(\theta)$
μ_1	1	$\mathcal{N}(m_1, s_1^2)$	1	0.5	1	0.5
μ_2	-1	$\mathcal{N}(m_2, s_2^2)$	-1	0.5	-1	0.5
σ_{if}	0.25	IG($\alpha_\sigma, \beta_\sigma$)	1.3	27	.25	.25
ϕ_1	0.95	Beta(α_ϕ, β_ϕ)	9.5	1.5	.8	.1
ϕ_2	0.90	Beta(α_ϕ, β_ϕ)	9.5	1.5	.8	.1

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 α and σ^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 λ from the full conditionals. In case of the parameters ϕ_i the full conditionals are not available precisely, as the prior is not conjugate. Therefore, sampling ϕ_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 (conditional) 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 constructs 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 μ and h are updated.

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,1 $\lambda_2|\sigma^2, \alpha, y, \psi$, 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 σ_f, the unconditional standard deviation of the SV process, then from σ_ω, the conditional standard deviation. All main samplers use σ_f, only the sampling results indicated by 'Gibbs-ω' give alternative results for the Gibbs sampler using the parameterisation in terms of σ_ω.

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.

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 ϕ_1, based of the 100,000 drawings from the H-M, H-M/Split, Gibbs, ARMS and Gibbs-σ_ω samplers using the standard parameterisation.
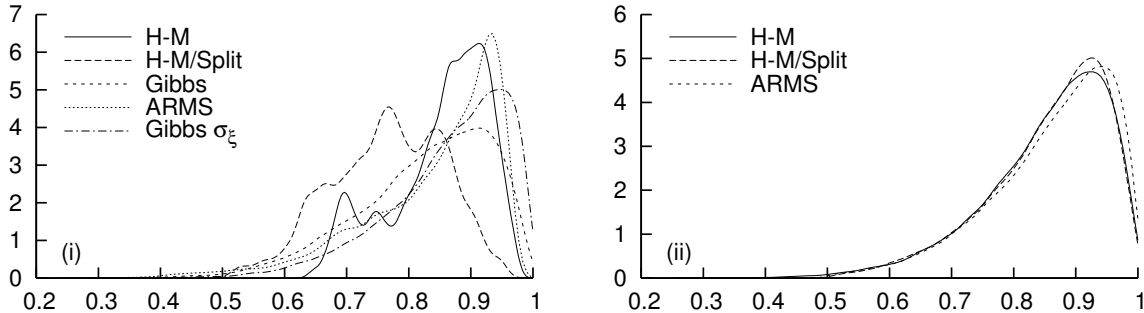

Figure 1: Marginal posterior distribution of parameter ϕ_1, using (i) the standard parameterisation and (ii) using the alternative parameterisation

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 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.

Figure 2 displays the autocorrelation in the sample for the parameter ϕ_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 an effective way the performance of the overall sample can be poor. The slow mixing of the samplers is caused in part

Table 2: Posterior correlation and simulation inefficiency

	H-M		H-M/Split		Gibbs		ARMS		Gibbs-σ_ω	
	ρ_{30}	Ineff	ρ_{30}	Ineff	ρ_{30}	Ineff	ρ_{30}	Ineff	ρ_{30}	Ineff
ϕ_1	0.990	2888.4	0.984	2215.8	0.804	369.9	0.976	1897.2	0.774	190.4
ϕ_2	0.985	2792.6	0.979	2671.3	0.890	498.8	0.965	1580.7	0.881	458.5
μ_1	0.973	2884.6	0.955	1902.2	0.860	1693.6	0.964	2439.6	0.774	1743.3
μ_2	0.710	1411.5	0.738	1412.0	0.430	324.3	0.645	492.5	0.328	396.5
$\sigma_{f/\omega 1}$	0.992	3088.7	0.983	2300.6	0.724	1793.0	0.989	2668.1	0.954	2124.9
$\sigma_{f/\omega 2}$	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 $B_m = 2,000$.
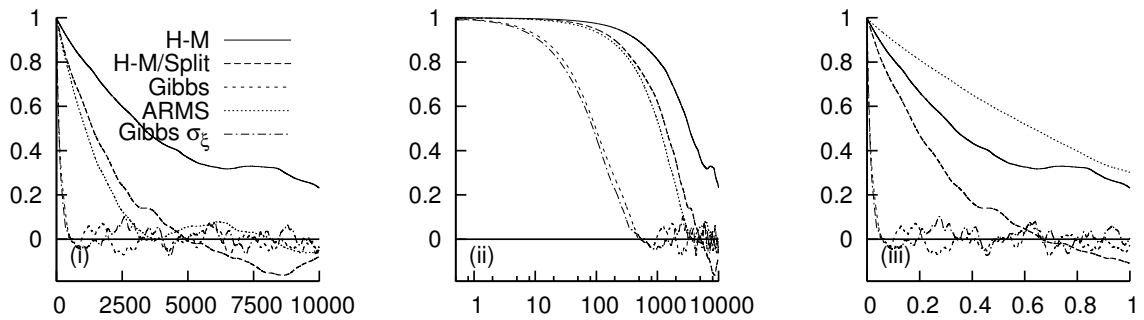

Figure 2: Autocorrelation (i) of the sample using the standard parameterisation, (ii) on a log-scale and (iii) on a time scale, taking the computational effort into account.

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.

The next 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. The performance of the reformulation on the simulation design is checked in Section 3.2, which is followed by a note on the situations in which an improvement of the sampler can be expected.

3.1   Disturbance based block sampling in GSSF-SV models

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 little or no information on the value of the parameters.

The sampling algorithm now becomes:

1. Initialise u, θ, 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 σ^2 from u and θ.

4. Sample from $\sigma^2 | \alpha, y, \theta$.

5. Recompute u from σ^2 and θ.

6. Goto 2.

Notice that step 2 is subtly different from 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 changed considerably, 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 (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.

3.2 Performance of reformulation

The same simulation design as described in Section 2.2 was used, with the reformulation of the model in terms of the disturbances.

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 σ^2 and the likelihood of the GSSF model. This last likelihood is only available in closed form if we follow Fruhwirth-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 σ^2 from u and λ, 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 σ^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 λ 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 ϕ_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 ρ_{30}	H-M Ineff	H-M/Split ρ_{30}	H-M/Split Ineff	ARMS ρ_{30}	ARMS Ineff
ϕ_1	0.320	63.9	0.215	49.0	0.127	40.8
ϕ_2	0.734	337.5	0.656	219.0	0.586	223.2
μ_1	0.054	12.0	0.053	18.2	0.050	7.4
μ_2	0.272	87.0	0.276	140.5	0.265	53.4
σ_{f1}	0.233	36.7	0.260	68.3	0.224	47.2
σ_{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.

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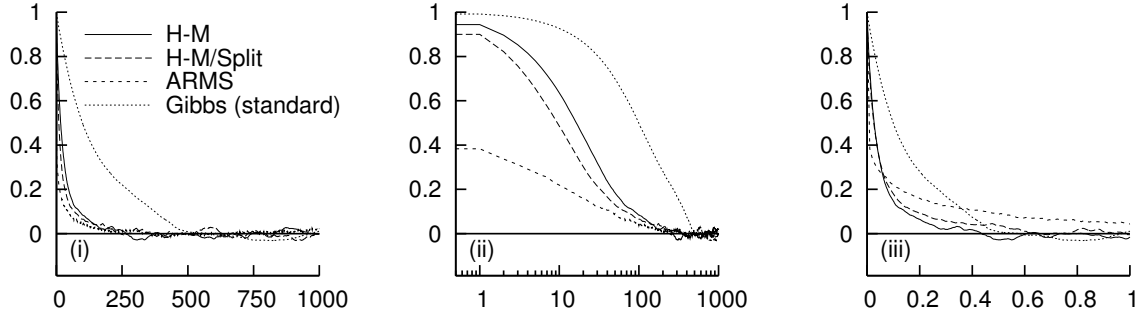
Figure 3: Autocorrelation (i) of the sample comparing the standard parameterisation with the Gibbs sampler to the reformulation and the H-M sampler, (ii) on a log-scale and (iii) on a time scale, taking the computational effort into account.

3.3 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 anytime. To understand this question, it is important to realise where the gains are coming from.

Following Tanner (1996) 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

$$\log p(\theta|y) = \log p(\theta|y, h) - \log p(h|y, \theta) + C.$$

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 $p(h|y, \theta)$ of the augmentation parameters of

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

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

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

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 specification of the SV equation was

$$h_{t+1} - \mu = \phi(h_t - \mu) + (1 - \phi^2)\sigma_{\text{SV}}\omega_t, \qquad \omega_t \sim \mathcal{N}(0, 1)$$

with $\phi = 0.95, \mu = 1$ and unconditional standard deviation σ_{SV} ranging from 0.05 to 5. Figure 4 displays the determinant of the information matrix for the different choices of σ_{SV}, for the original parameterisation and the reformulation.

The figure indicates how, for SV processes with low unconditional variance, the density of $p(\theta|y, h)$ 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_{SV}^2, including integrated SV processes with $\phi = 1$, the original formulation can be expected to work just as well, or better.
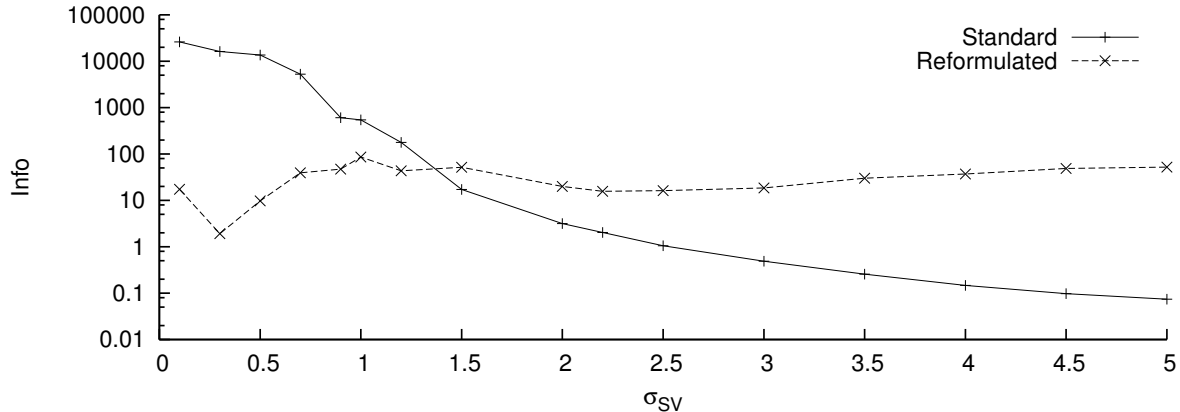
11

Figure 4: 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 variability of the SV process.

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.

In Section 5, on the illustrations using real world data sets, it is found that a combination of the methods works remarkably well. The simple combination, switching evenly between the reformulation and the original representation already delivers robust mixing close to the optimal choice for cases with extreme (lack of) persistence in the SV component. A second solution evaluates the conditional information matrices $\mathcal{I}_h(\theta) = -\frac{\partial^2 \log p(\theta|y,h)}{\partial\theta\partial\theta'}$ and $\mathcal{I}_u(\theta) = -\frac{\partial^2 \log p(\theta|y,u)}{\partial\theta\partial\theta'}$. The choice for the formulation conditioning on either u or h can be made random, with a probability depending on their standard deviations as implied by the covariance matrix $\hat{\Sigma}_u = \mathcal{I}_u(\theta)^{-1}$, according to

$$p(\text{Use reformulation}) = \frac{\sum \log(\text{diag}(\Sigma_u(\theta))^{1/2})}{\sum \log(\text{diag}(\Sigma_u(\theta))^{1/2}) + \log(\text{diag}(\Sigma_h(\theta))^{1/2})}.$$

The probability can be adapted in a straightforward manner if several SV processes are to be sampled, for each of which the choice to use the reformulation can be made separately. Furthermore, it is advisable to adapt the probabilities only once every l iterations, with l sufficiently large, as the evaluation of the information matrices is computationally costly.

4   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

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

Such quantities, conditioning only upon information available at time t, are very important in practise for the use of sequential forecasting and model checking. 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 $\alpha_t, \sigma_t^2 | \mathcal{F}_t$,

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

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. This implies that

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

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

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

Straightforward calculations show 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

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

and

$$
\begin{aligned}
a^{(i,j)}_{t+1|t+1} &= a^{(i)}_{t+1|t} + P^{(i)}_{t+1|t} Z'_{t+1} \left\{ F^{(i,j)}_{t+1} \right\}^{-1} v^{(i)}_{t+1}, \\
P^{(i,j)}_{t+1|t+1} &= P^{(i)}_{t+1|t} - P^{(i)}_{t+1|t} Z'_{t+1} \left\{ F^{(i,j)}_{t+1} \right\}^{-1} Z_{t+1} P^{(i)}_{t+1|t}.
\end{aligned}
$$

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

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

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

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

5 Illustrations

This section applies the sampling methods and the particle filter of Sections 2-3 and 4 on two data sets. First, Section 5.1 uses a GSSF-SV model to dissect exchange rates into country-specific factors, and investigates whether it is useful to model also a country-specific stochastic variance component.

It is followed, in Section 5.2, by an application using the cubic spline model (2) on a data set with measurements on the acceleration of the head in simulated motorcycle accidents. The observations in this data set are highly irregularly spaced in time, and display strong and sudden changes in variability. The question here is whether one SV component on the observation equation is enough, or if it is advisable to allow the state component to display time varying variance as well.

5.1 Local level with SV effects

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 $e_{i,t} \perp\!\!\!\perp e_{j,t}$, if $i \neq j$.

Using country 0 as a numeraire, taking random walks for the country factors, a model for

multiple exchange rates between $n + 1$ countries could be

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

$$\begin{pmatrix} e_{0,t+1} \\ \vdots \\ e_{n,t+1} \end{pmatrix} = \begin{pmatrix} e_{0,t} \\ \vdots \\ e_{n,t} \end{pmatrix} + H_t \epsilon_t \qquad (4)$$

$$\epsilon_t \sim N(0, \mathcal{I}_n)$$

where H_t is a diagonal matrix with typical element

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

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.

Alternatively, the model can be estimated with e.g. a common stochastic volatility, when H_t is a diagonal matrix with elements

$$H_{ii,t} = \sigma_{\epsilon_i} \exp(h_t)$$
$$h_{t+1} = \phi h_t + \sigma_\omega \omega_t$$
$$\omega_t \sim N(0, 1).$$

The possibility of $\sigma_{i,\omega} \equiv 0$, leading to fixed variances, is not considered here, as for exchange rates this assumption is unrealistic.

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 $3 \times (k + 1)$ 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 $k = 1$ 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.

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.[1] This period contains 2610 daily observations and is depicted in Figure 5, 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.

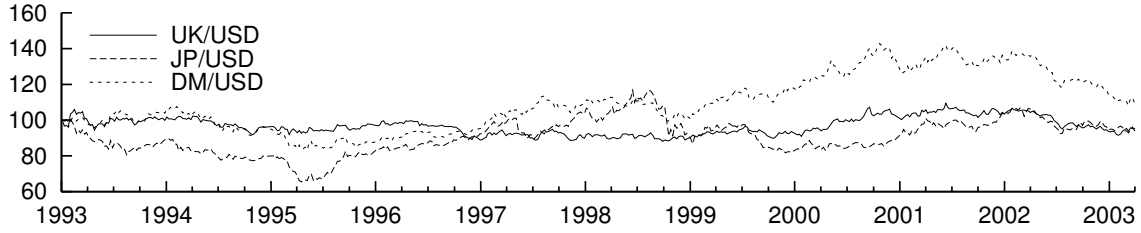[1]Source: `http://pacific.commerce.ubc.ca/xr/`.

Figure 5: Exchange rates of the British Pound, Japanese Yen and German DMark against the US Dollar, with $1/1/1993 \equiv 100$

Posterior estimates

The model as presented above is estimated using the simulation techniques described earlier. After a burn-in period of 10,000 iterations 100,000 sampled parameters are collected. 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 distributed around 0 with $\sigma_\gamma = 2$.

Table 4: Posterior statistics of the multi SV factor model

	$\overline{\theta}$	95% conf		ρ_{500} Original	Reform	50/50	$p(\Sigma)$
σ_ω(US)	0.191	[0.10,	0.33]	0.84	0.40	0.28	0.34
σ_ω(UK)	0.249	[0.14,	0.40]	0.80	0.48	0.44	0.39
σ_ω(JP)	0.276	[0.20,	0.37]	0.52	0.30	0.32	0.24
σ_ω(DM)	0.124	[0.08,	0.19]	0.76	0.28	0.32	0.25
ϕ(US)	0.955	[0.89,	0.99]	0.73	0.47	0.26	0.32
ϕ(UK)	0.961	[0.92,	0.99]	0.63	0.54	0.45	0.35
ϕ(JP)	0.937	[0.90,	0.97]	0.41	0.39	0.32	0.25
ϕ(DM)	0.980	[0.96,	0.99]	0.52	0.55	0.30	0.21
γ(US)	-2.040	[-2.25,	-1.81]	0.08	0.41	0.00	0.03
γ(UK)	-2.742	[-3.07,	-2.43]	0.08	0.38	0.02	0.01
γ(JP)	-1.140	[-1.33,	-0.94]	0.03	0.25	0.01	0.01
γ(DM)	-1.957	[-2.26,	-1.65]	0.04	0.80	0.00	0.05

Reported are the mean and 95% confidence bounds of the posterior of the parameters in the model, together with the 500th lag autocorrelation in the sample. Autocorrelations are reported for the samplers using the original, the reformulated and two combined samplers, either switching deterministically between the methods or adapting the probabilities relating to the standard deviations of the vector of parameters $\theta|h, y$ resp. $\theta|u, y$.

The results for the multi-SV model are summarised in Table 4 and Figures 6–7, with posterior moments, densities, autocorrelations and the underlying (SV) factors. Statistics for the alternative model with common SV are given in Table 5.

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 estimate their standard deviations σ_ω, autocorrelations ϕ and location parameters γ.

All sampling methods eventually result in the same posterior density, as should be the case;

16

Table 5: Posterior statistics of the common SV factor model

	$\overline{\theta}$	95% conf		ρ_{100} Original	Reform	50/50	$p(\Sigma)$
σ_ϵ(US)	0.367	[0.35,	0.39]	0.05	0.02	0.03	0.07
σ_ϵ(UK)	0.267	[0.25,	0.28]	0.02	0.01	0.02	0.04
σ_ϵ(JP)	0.578	[0.55,	0.61]	0.05	0.03	0.03	0.08
σ_ϵ(DM)	0.371	[0.35,	0.39]	0.03	0.02	0.02	0.06
σ_ω	0.474	[0.40,	0.55]	0.53	0.16	0.18	0.22
ϕ	0.737	[0.66,	0.81]	0.44	0.23	0.16	0.22

See Table 4 for an explanation of the entries in the table.
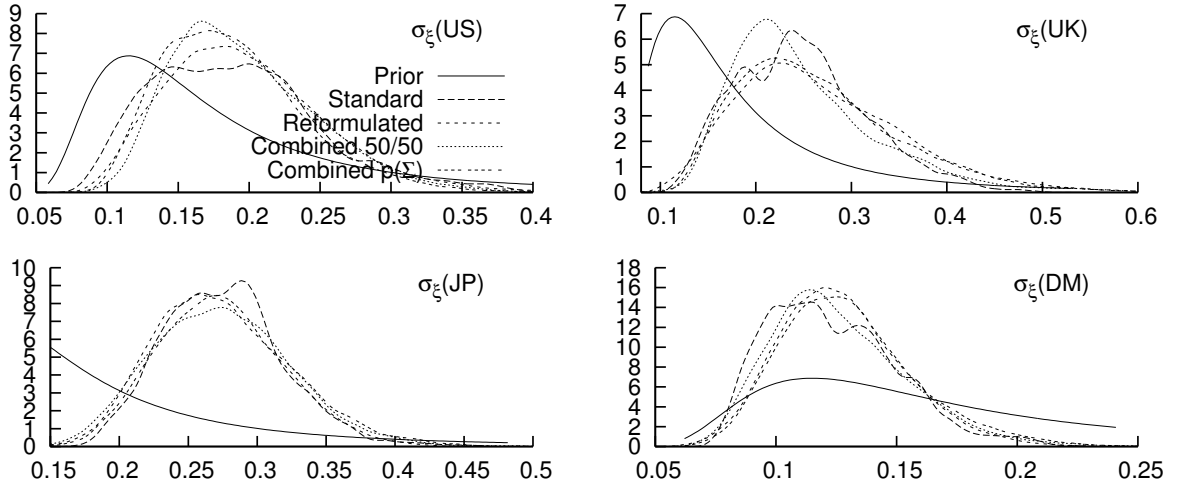
Figure 6: Posterior density of SV standard deviations for the factor model of exchange rates, without and with the transformation

see Figure 6 for the full posterior densities of the parameters $\sigma_{\omega,i}$. Even for such a long sample, the estimates of the posterior density of σ_ω(US) and σ_ω(UK) are multimodal, an indication that the original simulation scheme did not yet converge fully. The reformulated sampler displays clearly lower autocorrelations ρ_{500} in Table 4 for the parameters σ_ω and ϕ, whereas for the γ's the original formulation works better. The columns marked 'Combined 50/50' and 'Combined $p(\Sigma)$' switch between the two formulations, with the first alternating evenly between the original and the reformulated samplers, whereas the second chooses randomly from the formulations with probabilities proportional to the conditional standard deviations of the parameters, conditional on either h or u (see also Section 3.3).

In Figure 7 the autocorrelation in the posterior sample is displayed. The original formulation is seen to work well for the group of γ parameters, while the other parameters display
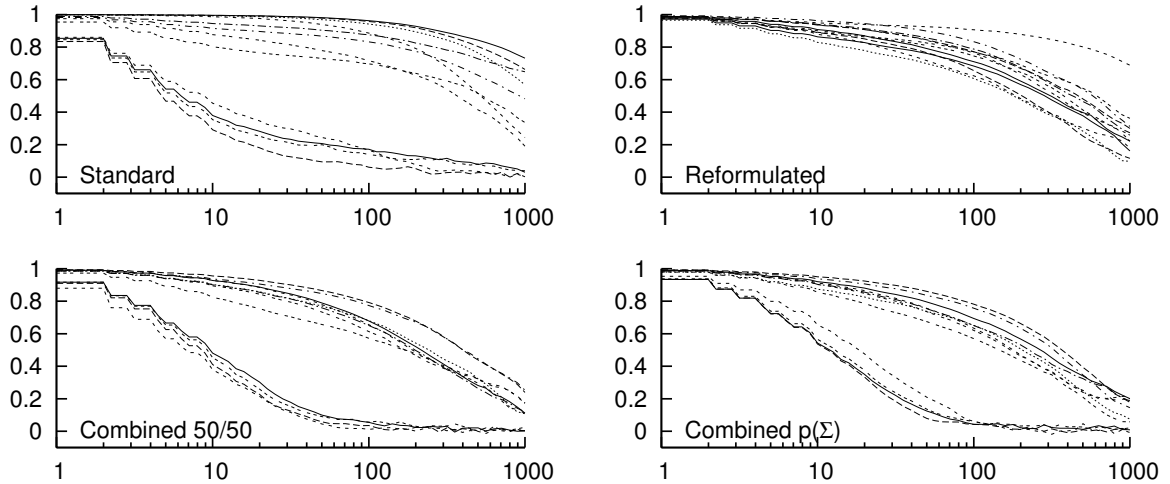
Figure 7: Autocorrelation of the sample of SV variance parameters for the factor model of exchange rates, with the four sampling schemes

strong autocorrelation also at higher lags. With the reformulation, the autocorrelation of most parameters drops off towards zero at a steady rate, with a clear advantage for most parameters. It is seen that a combination of the two sampling schemes has the advantage of letting the ρ_γ drop off quickly, roughly along the lines of the original sampler, whereas the other parameters show less autocorrelation as well. For this application, there is no clear difference between the deterministic and the probabilistic switching rule, though this may depend on the application at hand.

Table 5 displays the posterior results for the model with only one, common, SV component. The model contains only 6 parameters, and is estimated more easily. The autocorrelations drop off to zero considerably quicker, though at lag 100 the reformulated sampler displays lower autocorrelation than the original or combined methods. Note however the estimates for the parameters of the SV component: When the countries are forced to undergo the same stochastic variance process, this process is less well defined, with $\phi \approx 0.74 \ll .95$ and σ_ω higher than previously found. The persistence in the SV process is minimal, implying that there is no clear joint SV process common to the exchange rates between the four countries.

Table 6: Likelihood measures for the FX models

	Multi SV	Common SV
Log Likelihood	-1244.5	-1276.0
Log Marginal Likelihood	-1285.5	-1297.6
Log Jackknife Likelihood	-5817.5	-6132.0

18

The particle filter of Section 4 can be used to compute the likelihood in a high density point, e.g. the posterior mean. In Table 6 the filter was used with $m = 1,000$ particles and $r = 100$ proposals for each particle. From the likelihood estimate (the logarithm of) the marginal likelihood, which is the likelihood of the data after integrating out the parameters, can be derived (see Kass and Raftery (1995) and Bos (2002)). A third statistic in the table is the jackknifed loglikelihood, defined as

$$\log \mathcal{L}_{\text{JK}} = \sum_t \log \mathcal{L}(y_t | \mathcal{F}_{T \setminus t}, \theta) = \sum_t \int \log \mathcal{L}(y_t | \mathcal{F}_{T \setminus t}, \theta, h) P(h | \mathcal{F}_{T \setminus t}, \theta) \partial h,$$

which again is evaluated at the posterior mean θ, with the SV processes integrated out through simulation. This jackknifed likelihood measures the extent to which each observation fits in with the other observations, and has the advantage that it can be calculated without the use of the particle filter.

The estimates in Table 6 indicate that the likelihood of the multi-SV model is higher, and also on grounds of the marginal likelihoods the multi-SV model is strongly preferable to the model with only one single SV component, with a difference of 12 points on a log-scale. The same message is given by the jackknifed likelihood, which indicates that each of the observations fits better into the model with multiple SV than with common SV.

Economic results and discussion

Model (4) takes the exchange rates apart in country specific factors with their respective SV components. Figure 8 displays the underlying country components e_i for the US, UK, Japan and Germany/EU for the period of the data sample. The separation of the exchange rates in country factors allows the researcher to compare the strength of currencies: E.g. it is seen that the weak performance of the Euro, especially after 2001, is partly caused by the weakness of the European currency itself, but that in the mean time the Dollar itself grew stronger, especially in 2002 and the beginning of 2003. Likewise, the British Pound is seen to have been stable over the last 10 years apart from a few sudden shifts, e.g. in April 1995.

Similarly, the model delivers an estimate for the standard deviations pertaining to the specific currencies. These estimates are displayed in Figure 9. The factors indicate heightened uncertainty about the exchange rates in the second quarter of 1995, for all countries. For Germany, or the Euro area, uncertainty increased slightly just before the introduction of the Euro banknotes; quickly after January 2001 the markets noticed that the introduction went smoothly, and uncertainty again diminished. For the US, one could remark that the events of 9/11/01 do not seem to have led to increased exchange rate uncertainty, or at least it was not noticeable. For Japan, the banking problems give rise to sharply rising uncertainty in 1998.

For the model with one common SV, the posterior standard deviation factor is plotted in Figure 10. From this figure it is clear that no persistent common SV is derived for these countries, and from an economic viewpoint the multi-SV model should be preferred.

The model at hand is easily generalised to even 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.

5.2   Regression spline

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
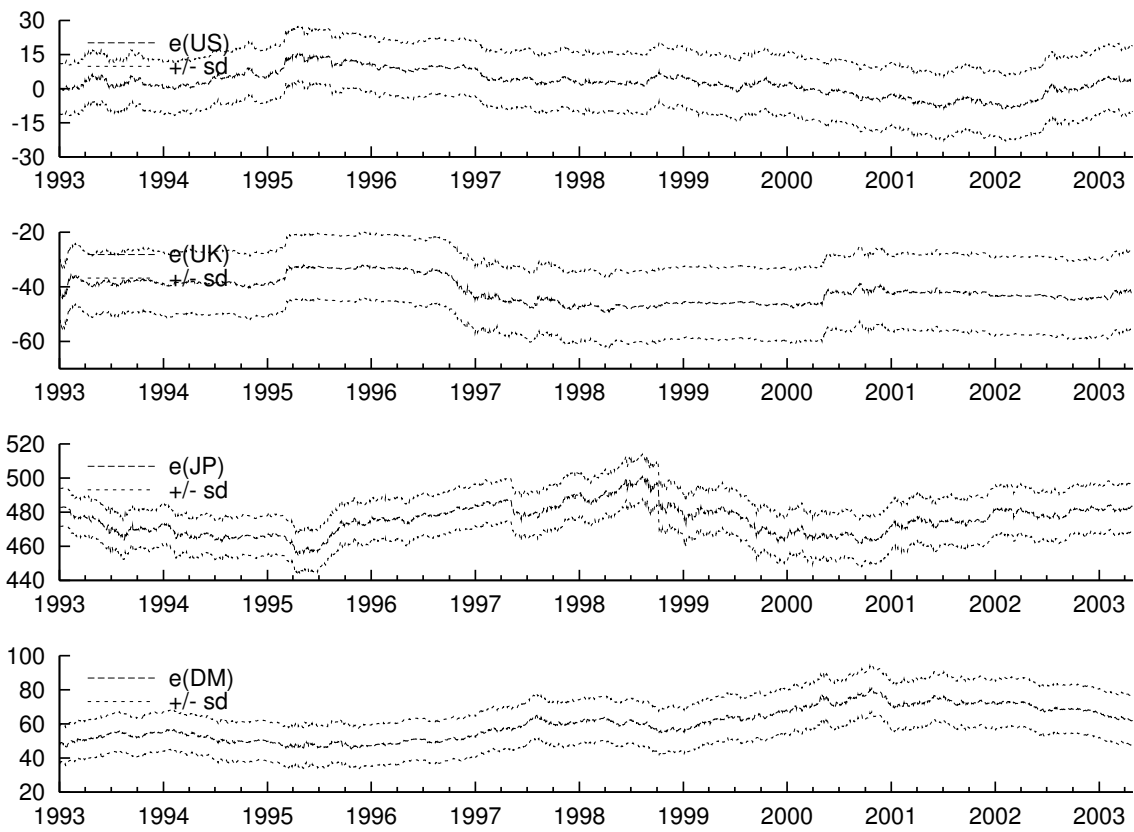
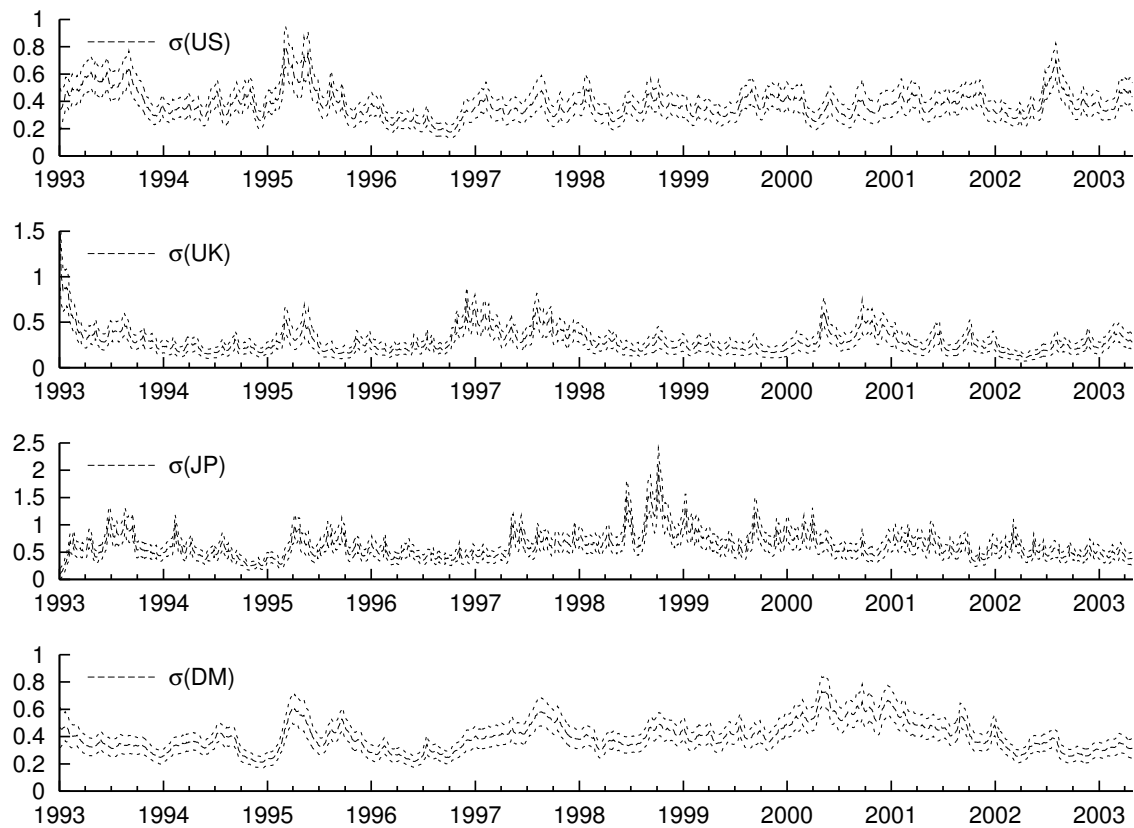Figure 8: Posterior mean of underlying country factors, ± one standard deviation

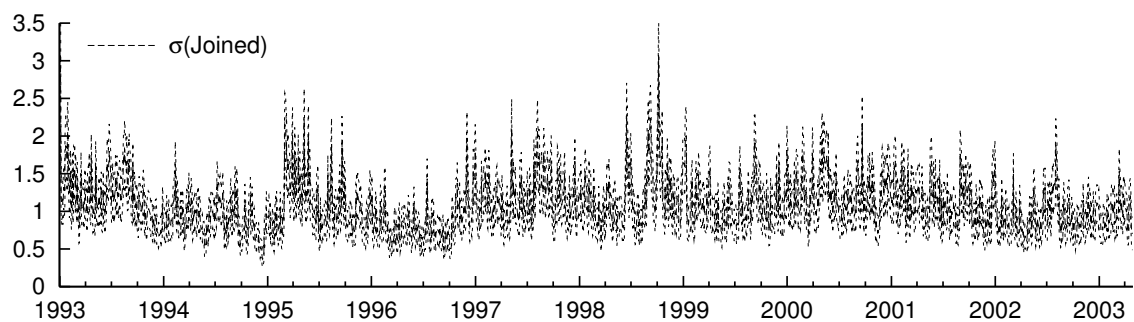Figure 9: Stochastic volatility for the different factors, ± one standard deviation, for the multi SV model

Figure 10: Stochastic volatility for the different factors, ± one standard deviation, for the common SV model

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 σ_{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 7 gives the parameters of the Inverted Gamma-1 prior densities on the standard deviations in the model.

Table 7: Parameters for IG-1(α, β) prior densities for the cubic spline model

	α	β	$E(\theta)$	$\sigma(\theta)$	Equation
σ_1	1.5	2	2.5	1.9	Observation
σ_2	1.4	4.3	0.6	0.5	Transition
$\sigma_{\omega,1}$	2	4.5	0.4	0.2	SV Observation
$\sigma_{\omega,2}$	2	4.5	0.4	0.2	SV Transition

Using the both the conventional sampling scheme as the transformation of the model, as explained in Sections 2.1 and 3, the posterior density of the parameters was derived. Also correlations in the posterior sample were computed using two combinations of the conventional and reformulated samplers. Table 8 reports the posterior means for the parameters, for the three specifications of the variance processes. For each model the sampler was run for 100,000 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.

Results

For the model with fixed variance, a high observation standard deviation σ_1 is estimated, as at least some periods have a higher variation. This is seen in Figure 11, where the estimates for 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 panel of Table 8, the average standard deviation σ_1 as implied by the SV model is estimated at 19.6 (in italics), only slightly lower that in the model without SV. However, Figure 12 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 we are able to derive the results based entirely on a probabilistic model, whereas their weighting scheme was more or less ad-hoc.

For the model with **1 SV**, the 30-th order autocorrelations are reported for the samplers

Table 8: Posterior means of parameter estimates

		$\overline{\theta}$	95% conf		ρ_{30} Original	Reform	50/50	$p(\Sigma)$
0 SV	σ_1	21.989	[19.40,	25.12]	-0.02			
	σ_2	0.631	[0.43,	0.93]	0.02			
1 SV	σ_1	*19.59*						
	σ_2	0.587	[0.41,	0.84]	-0.02	-0.01	-0.01	-0.00
	$\sigma_{\omega,1}$	0.321	[0.23,	0.45]	0.25	0.92	0.41	0.43
2 SV	σ_1	*19.39*						
	σ_2	*0.68*						
	$\sigma_{\omega,1}$	0.330	[0.23,	0.46]	0.26	0.92	0.47	0.60
	$\sigma_{\omega,2}$	0.489	[0.28,	0.83]	0.63	0.91	0.63	0.63

Reported are the mean and 95% confidence bounds of the posterior of the parameters in the model, together with the 30th lag autocorrelation, in the models with 0, 1 and 2 SV components. Values in italics are average values for the standard deviations of the observation and transition equation as implied by the respective SV processes. Autocorrelations are reported for the samplers using the original, the reformulated and two combined samplers, either switching deterministically between the methods or adapting the probabilities relating to the standard deviations of the vector of parameters. $\theta|h, y$ resp. $\theta|u, y$.
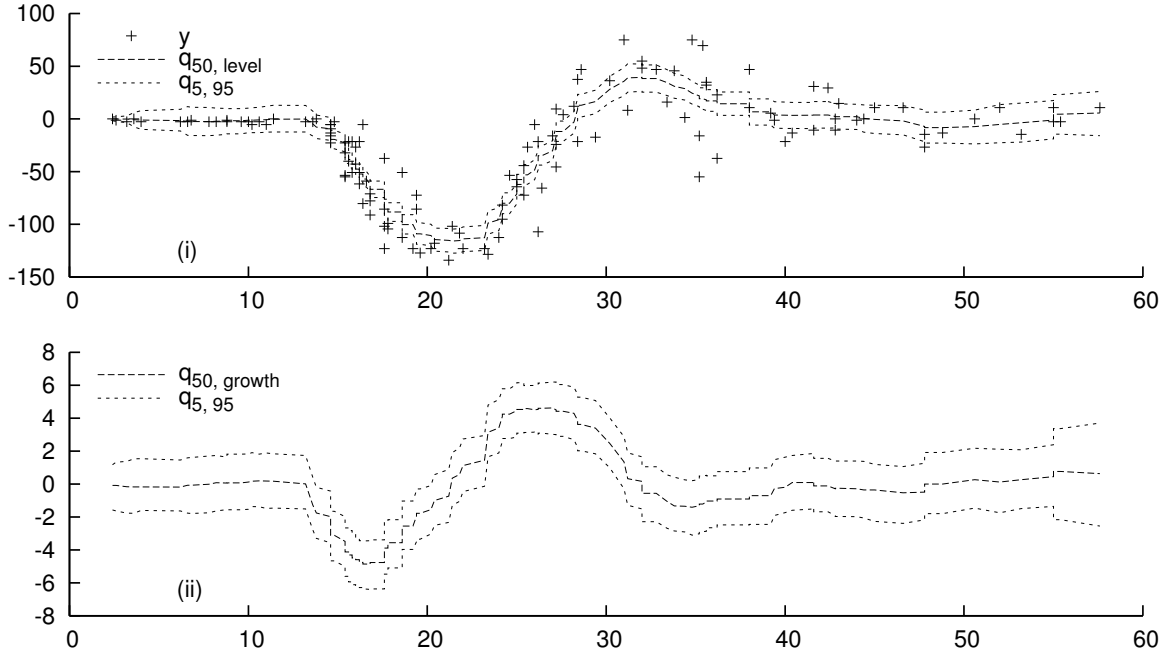

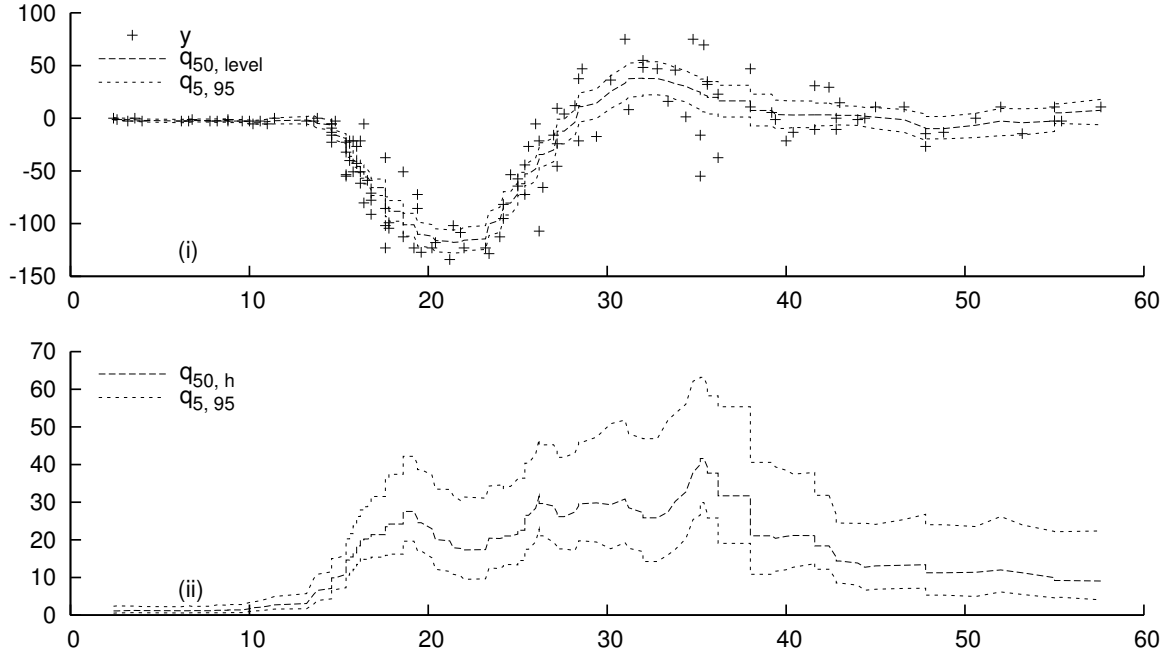Figure 11: Cubic spline level (i), growth (ii) with quantiles, for the Gaussian model

23

Figure 12: Cubic spline level (i) and observation standard deviation (ii) with quantiles, for the model with one stochastic variance

without the transformation, with the transformation, and with two combinations, in the last four columns of Table 8. In accordance with the theoretical results in Section 3.3, the method without the transformation is performing best, followed closely by the two combinations of methods. The reason for this finding lies in the fact that the SV component estimated is here of the I(1) type, with $\phi \equiv 1$, and hence the unconditional variance of the SV process is infinite. In such a situation, theoretical results indicated that the original formulation can be expected to perform best. Note that for the parameter σ_2, the standard deviation of the transition equation, the autocorrelations drops sufficiently quickly; especially for the SV parameter σ_ω the method of simulation matters.

The third model adds another stochastic variance to the growth component in panel (ii) of Figure 11. 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 panel in Table 8 allows for such behaviour by introducing SV on the transition equation as well. The standard deviation of this SV component $\sigma_{\omega,2}$ is estimated at values even larger than $\sigma_{\omega,1}$, implying that there is more variability here. The second panel of Figure 13 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.

Adapting the variances throughout the sample leads to a large improvement in the distribution of the residuals of the model. Figure 14 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**.

Diagnostic checking on the fitted model

Figures 11–14 are based on posterior quantities using the full data sample in the sampling algorithm. Another comparison of models can be made using likelihood measures, see also the corresponding results in Section 5.1.
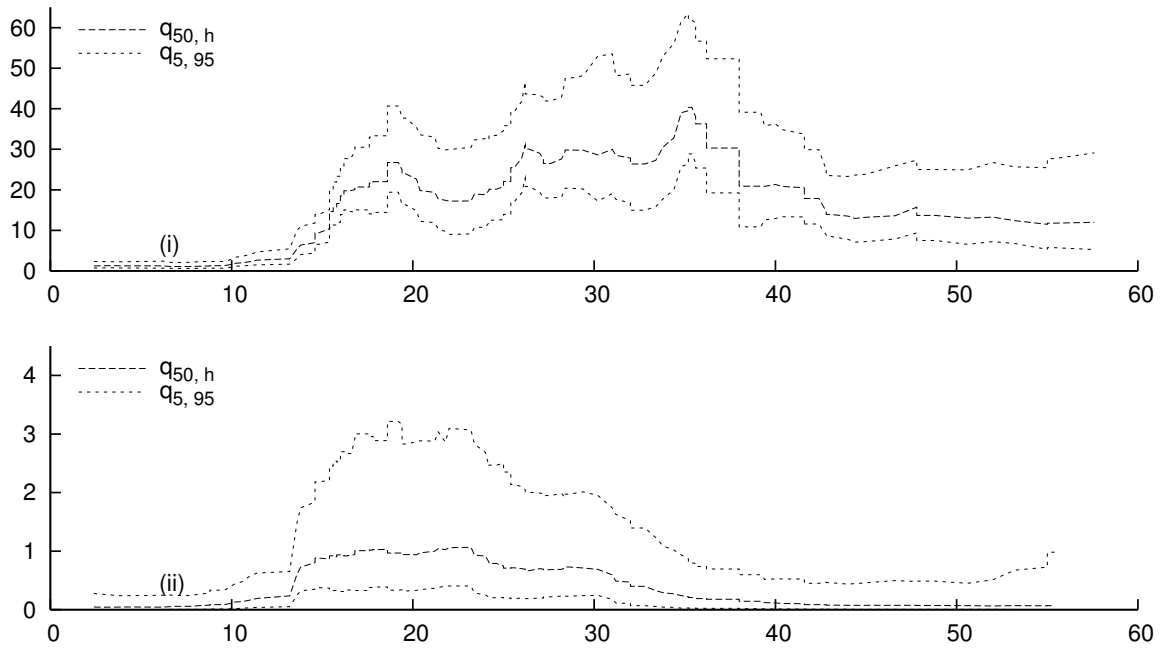
24

Figure 13: Observation standard deviation (i), transition standard deviation (ii) with quantiles, for the model with double stochastic variance
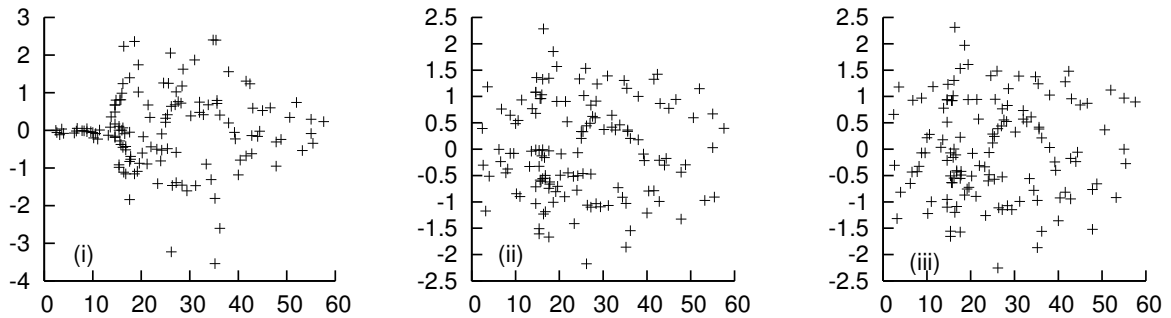

Figure 14: Average standardised smoothed residuals for the model without SV (i), with one SV component (ii) and with two SV components (iii)

Table 9: Likelihood measures for the cubic spline models

	0 SV	**1 SV**	**2 SV**
Log Likelihood	-625.93	-595.30	-597.87
Log Marginal Likelihood	-638.48	-597.18	-599.16
Log Jackknife Likelihood	-1152.70	-690.43	-691.79
Box-Ljung Q_u	8.27	10.15	7.18
Box-Ljung Q_v	142.99	12.52	5.33

Loglikelihood at the posterior mean, logarithm of the marginal likelihood, and the Box-Ljung statistic testing for 12th order autocorrelation in the u and v statistics.

25

According to the standard loglikelihood and log-marginal likelihood measures, in the first panel of Table 9, the Gaussian model indeed fits considerably worse than the models with one or two SV components. The log-marginal likelihood of the **2 SV** model is comparable to the the **1 SV** model, but there is no strong indication of the need for the extra SV component. The jackknife likelihood gives a further indication that the model with one SV component is preferable.
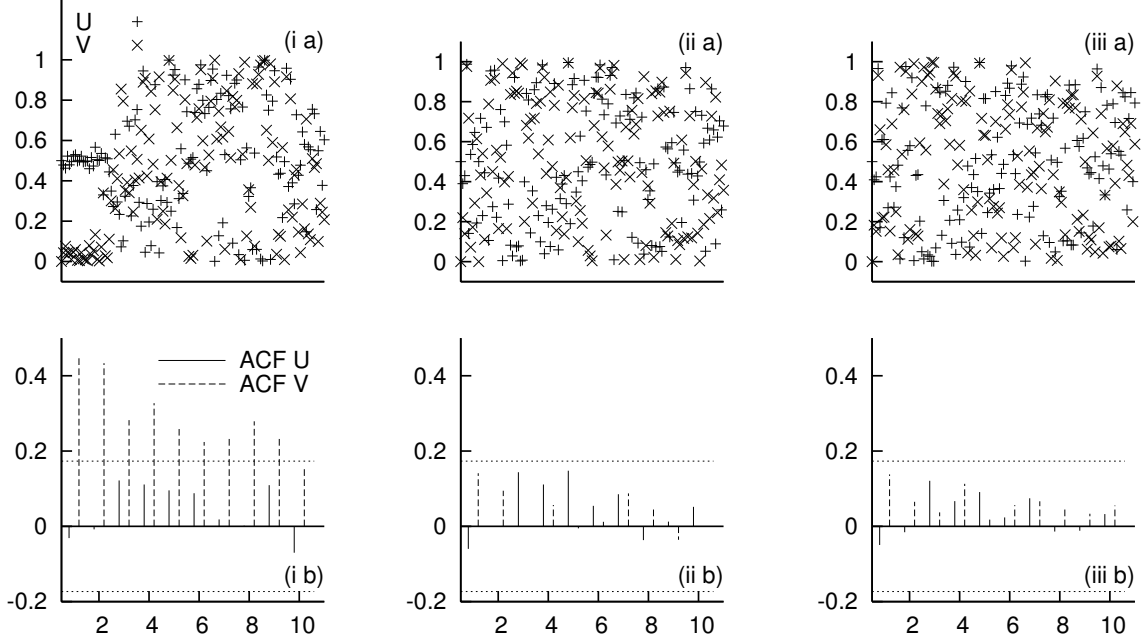

Figure 15: Probabilities u_t of observations and transformations v_t against the index, with auto-correlations

The second panel of Table 9 is concerned with statistics u and v defined as

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

where the probability is calculated integrating out the SV process(-es) of the model and conditioning on the posterior mean. They can easily be computed as a by-product of the particle filter.

These u and v statistics would ideally be distributed i.i.d. $U(0,1)$, and can be used for checking SV models as is done in Kim, Shephard, and Chib (1998). Figure 15 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 disappears. The Box-Ljung statistics for u and v are calculated using $\sqrt{T} \approx 12$ lags. In standard models, the statistic would be χ^2 distributed with $12 - k$ degrees-of-freedom, where $k = 2$ is the number of parameters in the model. For the model without SV, the hypothesis that v is uncorrelated seems to be strongly rejected, even if the test is not exact in this setting. With either one or two SV components, both tests take on acceptable values for testing the null of uncorrelated u and v.

Discussion

The analysis displays again the viability of the Bayesian approach to modelling state space models with multiple stochastic variance processes, even for data sets where the data is unevenly spaced through time, with possibly multiple observations at one time period. In this application, it was found that both the conventional as the reformulated algorithm work. As the SV processes are highly variable, the reformulation by itself did not lower correlations in the chain of sampled parameter values. A combination of the two methods however worked fine, and is a robust solution in general.

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.

Using simulated data, the effect of the reformulation was clear, in that the simulated chain displayed strongly better mixing properties. In the real data examples, 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. In practise, a combination of both the original and alternative formulations can be expected to give better overall results, in terms of delivering good mixing properties irrespective of the model at hand.

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 further analysis for the economic analysis of mayor 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.

When applied to a motorcycle data set with highly irregularly spaced observations, the addition of the SV component allowed for a robust estimation of the state, without having to resort to an ad-hoc procedure to provide a variance estimate ex-ante. The combined sampling methods in this case deliver a posterior sample with good mixing properties, and the marginal likelihoods computed using the particle filter indicate that a model with only one stochastic variance component is sufficient to cover the most salient features of the data.

7   Acknowledgements

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. 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. and S. J. Koopman (2004). State space models with a common stochastic variance. *Journal of Business and Economic Statistics*. Forthcoming.

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 (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.

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*, 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. *Econometrics 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 traumatiological 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.

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.

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.