Discussion Paper 96

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis
Research Department
P.O. Box 291
Minneapolis, Minnesota 55480-0291

December 1994

Stochastic Volatility and the Distribution of Exchange Rate News

Ronald Mahieu*

University of Limburg and
Queen Mary and Westfield College

Peter Schotman*

University of Limburg

ABSTRACT

This paper studies the empirical performance of stochastic volatility models for twenty years of weekly exchange rate data. We concentrate on the effects of the distribution of the exchange rate innovations for parameter estimates and for estimates of the latent volatility series. We approximate the density of the log of exchange rate innovations by a mixture of normals. The major findings of the paper are that: (i) explicitly incorporating fat-tailed innovations increases the estimates of the persistence of volatility dynamics; (ii) estimates of the latent volatility series depend strongly on the estimation technique; (iii) the estimation error of the volatility time series is so large that finance applications to option pricing should be interpreted with care. We reach these conclusions using three different estimation techniques: quasi maximum likelihood, simulated EM, and a Bayesian procedure based on the Gibbs sampler.

*Ronald Mahieu is the corresponding author at the Limburg Institute of Financial Economics (LIFE), University of Limburg, P.O. Box 616, 6200 MD Maastrich, Netherlands, FAX: (+31) 43 258530, E-Mail: R.Mahieu@BERFIN.Rulimburg.nl. This research was started while Peter Schotman was a visitor at the Institute of Empirical Macroeconomics in Minneapolis. He thanks the Institute for its hospitality and John Geweke for helpful suggestions. All errors are, of course, our own.

1 Introduction

Many high frequency financial time series show time varying volatility. The most popular way to describe the time variation is the AutoRegressive Conditional Heteroskedasticity (ARCH) model introduced by Engle (1982), or one of its variants (GARCH, Bollerslev (1986), EGARCH, Nelson (1991)). An alternative approach, which has become more popular recently, is the stochastic volatility model, where the variance is modeled as an unobserved component that follows some stochastic process. It has been proposed because it directly connected to the type of diffusion processes used in asset pricing theory in finance (see, *e.g.*, Melino and Turnbull (1990)). Initial research in these directions was performed by Clark (1973) and Tauchen and Pitts (1983). It has been advocated in econometrics by Taylor (1986), Harvey, Ruiz and Shephard (1994), Andersen (1994a), and many others. The relations between the ARCH and stochastic volatility approaches have recently been analyzed by Andersen (1992, 1994b) and others.

From an econometric viewpoint the practical drawback of the stochastic volatility model is the intractability of the likelihood function. Because the variance is an unobserved component, the likelihood function is only available in the form of a multiple integral. Also, Quasi Maximum Likelihood (QML) and Method of Moments estimators are not very reliable, see Jacquier, Polson and Rossi (1994) and Andersen (1994). Exact likelihood oriented methods require simulations and are thus computer intensive, see Danielsson (1994) and Jacquier, Polson and Rossi (1994).

The first purpose of the paper is to compare different estimation techniques applied to an empirical dataset consisting of weekly exchange rate changes for six currencies over a 20 year period. The estimation techniques differ in their distributional assumptions about exchange rate innovations. Most of the computational problems stem from the assumption that the innovation of the underlying variable has a normal distribution, which translates to an awkward log chi-square distribution when the model is written in a linear state space form. This implication is ignored in the QML method, but fully implemented by Jacquier, Polson and Rossi (1994) and Kim and Shephard (1994).

The normality assumption is not only computationally impractable, it is probably also empirically unjustified. The distribution of exchange rate news is fat-tailed as is widely established in the literature. Like ARCH models, stochastic volatility can explain part of the fat-tailedness. But given the evidence for ARCH models one would expect that time varying volatility does not fully account for the tail behavior, in line with Baillie and Bollerslev (1989) and Bollerslev and Engle (1986). The second purpose of this paper is to replace the normally distributed innovation with a specification that is computationally more tractable and empirically viable.

Estimation of stochastic volatility models consists of two stages: parameter estimation, and estimation of the latent volatility time series. Methods that work well for estimating the parameter vector, are not necessarily also good in estimating the latent time series. For finance applications the main interest is in the volatility time series itself. The series is estimated by some smoothing algorithm, which also produces standard errors of the volatility estimate. This enables us to compare the different models and estimation techniques with respect to the estimated volatility series. Focussing directly on the output of the model – the volatility series – sheds light on issues like the efficiency gain from a simulation smoother over the Gaussian linear Kalman smoother, and the effect of some forms of misspecication. Because the interest is in estimating the latent volatility series at every time period, asymptotic arguments are of limited value. Consequently distributional assumptions become important for this purpose.

Like the GARCH(1,1) model in the conditional heteroskedasticity literature, most work on stochastic volatility deals exclusively with the basic univariate first order autoregressive volatility model. For stock prices Gallant, Hsieh and Tauchen (1994) conclude that this basic model is severely misspecified, not only because of distributional assumptions on innovations, but also because of its dynamic specification. Similar results might hold for exchange rates. The third aim of the paper is to perform a battery of diagnostic LM-tests to search for deviations of the basic stochastic volatility model.

The remainder of the paper is organized as follows. The model specification and the estimators are discussed in section 2. We consider four different estimators: Quasi Maximum Likelihood, two different simulated EM techniques, and a Bayesian method based on the Gibbs sampler. Section 3 describes the exchange rate data. Section 4 contains the parameter estimation results. Section 5 reports the results for the estimation of the latent volatility series. Section 6 concludes.

2 Models and Estimators

Let S_t be a bilateral exchange rate, and define $s_t = \Delta \ln S_t$. Assuming that the change in the log of the exchange rate is unpredictable, the standard stochastic volatility model is written

$$s_t = \exp(h_t/2)\epsilon_t \tag{1}$$

$$h_t = \beta + \rho(h_{t-1} - \beta) + \eta_t, \tag{2}$$

where $\exp(h_t)$ is the variance of s_t at time t, and where the innovations ϵ_t and η_t have mean zero, with variances equal to one and σ^2 respectively. The usual assumption is that ϵ_t and η_t are normally distributed. The exchange rate then obtains its fat tailed distribution by the mixing of ϵ_t and $\exp(h_t/2)$.

The estimation of stochastic volatility (SV) models has been the main obstacle for application of this type of model. Because of the latent volatility, likelihood analysis amounts to evaluating an integral with dimension equal to the number of observations:

$$L(Y_T;\theta) \propto \int f(Y_T|X_T,\theta)f(X_T|\theta)\mathrm{d}X_T, \tag{3}$$

where Y_T contains all the data, X_T is the vector with all the latent volatilities and θ contains the parameters of the SV model. In this equation the second density in the integral can be considered as a prior over X_T, specified by the transition equation (2). From a computational viewpoint specification (1) with normality for ϵ_t is inconvenient, since the likelihood function can only be written in integral form.

The most straightforward way to estimate the SV model is Quasi Maximum Likelihood (QML). This is the approach followed by Harvey and Shephard (1993a,b), Ruiz (1993), Harvey, Ruiz and Shephard (1994), Taylor (1994) and Mahieu and Schotman (1994). The QML method starts by transforming the measurement equation (1). Let $y_t = \ln s_t^2$, then (1) can be written in the linear form

$$y_t = h_t + \xi_t, \tag{4}$$

where $\xi_t = \ln \epsilon_t^2$. In the QML approach the density of ξ_t is approximated by a normal density with mean -1.27 and variance $\pi^2/2$. QML estimates can be then obtained by standard numerical optimization techniques, since the full likelihood becomes Gaussian. The QML estimator is not efficient, since the transformed error term ξ_t will be extremely skewed to the left, if the underlying ϵ_t is normal. Not only can this result in inefficient parameter estimates, but the standard Kalman smoother might also produce poor estimates for the state variable h_t conditional on the parameters of the process. Some of the drawbacks of QML have been documented by Andersen (1994) and Jacquier, Polson and Rossi (1994). The latter also document the performance of the (Generalised) Method of Moments estimator. The problem with the GMM estimator for the stochastic volatility model is that it is sensitive to the number and choice of sample moments, and that some moments depend on the distributional assumptions regarding ϵ_t.[1]

Recently, several methods have been developed to deal with the multiple integration problem in (3). These methods heavily rely on simulation techniques. Danielsson (1994) and Danielsson and Richard (1993) develop an importance sampling technique to estimate the integral. This method is still very computer intensive. Furthermore, extensions to multivariate models are not straightforward.

Jacquier, Polson and Rossi (1994) combine a Gibbs sampler with the Metropolis algorithm to obtain the marginal posterior densities of the parameters in (1) and (2), and also the exact posterior distribution of the variance series given the observed data s_t. They show that simulation from the exact densities can be done. Still the approach remains computationally demanding, and tailor made designed for the specific likelihood.

Shephard (1993b) and Kim and Shephard (1994) retain the convenient linear form of the state space form, and approximate the log-chisquared distribution of ξ_t by a prespecified mixture of seven normals. Shephard (1994) describes efficient algorithms for statistical inference in this class of what he calls partial non-Gaussian state space models, and provides examples of how to take advantage of the linear and Gaussian parts. The mixture is specified as

$$\xi_t = \xi(z_t), \qquad z_t = 1, 2, \ldots, M \tag{5}$$

$$\xi(i) \sim \mathrm{N}(\mu_i, \omega_i^2) \tag{6}$$

Using the mixture model the value of the likelihood is estimated using the multimove Gibbs sampler (see Tanner (1991)). Conditional on a time series of indicators z_t, $(t = 1, \ldots, T)$, the standard Kalman recursions can be used to simulate the states h_t from the conditional density $f(X_T | Y_T, Z_T, \theta)$. Conversely, given a time series for the states, a posterior odds calculation gives the multinomial distributions $f(z_t | Y_T, X_T, \theta)$, from which new regime indicators can be drawn. This method iteratively evaluates the conditional densities of the random variables in the model. Details of the iterations are given in appendix A. Tanner (1991) and Zellner and Min (1992) discuss some convergence criteria for the Gibbs sampler.

From a statistical viewpoint there is however no reason to insist on the log chi-squared distribution of ξ_t, and neither on normality of ϵ_t.[2] In this paper we make a slight generalization of the mixture framework by allowing ξ_t to be generated by a

[1]Optimal choices of moments are discussed in Gallant, Hsieh and Tauchen (1994), and require simulation estimators of the likelihood and score of an auxialiary model.

[2]There are economic reasons though to be more concerned about the normality, since without normality some of the motivations for the stochastic volatility model like in Clark (1973) and Tachen and Pitts (1983), lose much of their appeal.

Table 1: Mixture parameters
for log chi-squared distribution

i	weight	μ_i	σ_i
1	0.70	-0.2172	1.1052
2	0.25	-3.0461	1.5705
3	0.05	-6.4818	3.0002

flexible mixture of normals. If the distribution of ϵ_t is symmetric, there is in principle no loss of information, since it is always possible to calculate the implied distribution of $|\epsilon_t| = \sqrt{\exp(\xi_t)}$. The mixture can generate densities with a wide range of third and fourth moments.

The main feature of the data are the "inliers", *i.e.* observations with almost zero change for which the log transformation generates large negative outliers. Even as few as two normals can model the empirically observed (and also expected) negative skewness of ξ_t. A mixture with the first normal centered around zero, covering the bulk of the data, and the other centered around a large negative mean to accommodate the outliers, seems a good candidate for the distribution of ξ_t. Its main drawback is that such a distribution is likely to be bimodal. After some experimentation we therefore settled for a mixture of three normals with weights $p = (0.70, 0.25, 0.05)$. The middle element in the mixture provides a smooth blending to a unimodal density.

With three elements in the mixture the error distribution contains six free parameters. More than three elements in the mixture causes problems for the convergence of the estimator due to overparameterization. The weights p_i of the different elements of the mixture are specified a priori, as inference on these parameters is intractable.

A mixture of three normals is also not a serious limitation in case the error distribution happens to be log chi-squared. We performed a small simulation study to estimate the optimal parameters for data generated by a log chi-squared. The resulting mixture distribution parameters are given in Table 1. Figure 1 plots the mixture distribution together with the actual log chi-squared distribution. The main difference is the slightly higher mode of the mixture distribution. It looks like a three element mixture strikes a balance between flexibility and the number of parameters.

Although the transformation from ξ_t back to ϵ_t is well defined, it is very sensitive to exact specification of especially the right tail of the density of ξ_t. For example, under the QML assumption $\xi_t \sim \mathrm{N}(-1.27, \pi^2/2)$, the implied kurtosis of ϵ_t is equal

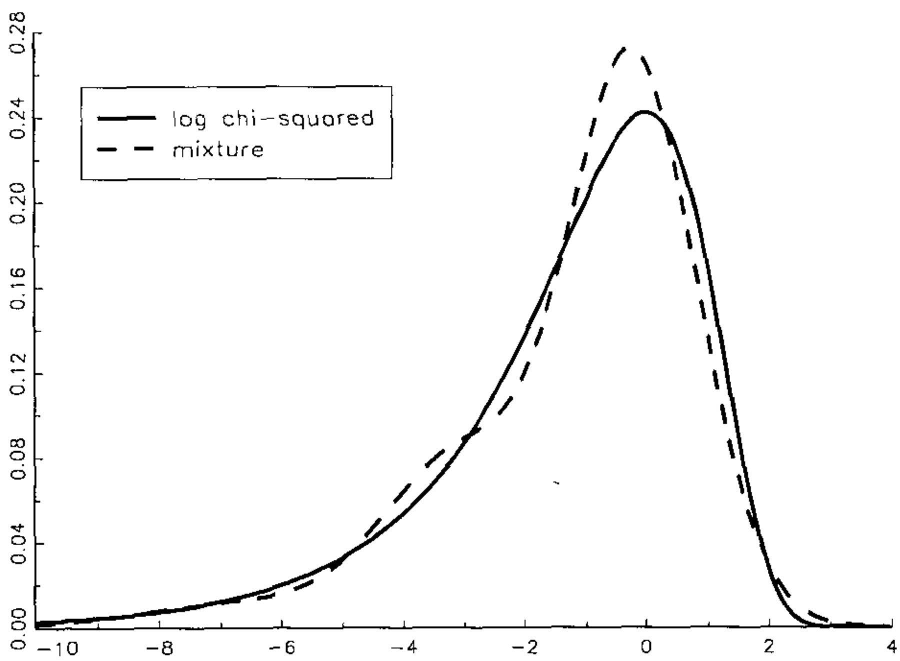


Figure 1: **Densities of measurement error**

to $\exp(\pi^2/2) = 139$ instead of 3.[3] For the mixture in Table 1 the implied kurtosis is 15.[4] In general the tails of ϵ_t become fatter the larger the variance of ξ_t, ceteris paribus, but given the sensitivity it is hard to draw firm conclusions.

Since the unconditional mean of the measurement error ξ_t is a free parameter in the mixture distribution, the constant term in the transition equation (2) becomes unidentified. We therefore respecify the variance dynamics without a constant term:

$$x_t = \rho x_{t-1} + \eta_t, \tag{7}$$

where $x_t = h_t - \beta$. Similarly the measurement equation (4) is also redefined as $y_t = x_t + \xi_t$.

The models will be estimated by simulation. We compare three different estimators. The first is the simulated EM algorithm described in Kim and Shephard (1994), which converges to the maximum likelihood estimator. With this estimator (SIEM1) the parameters of the mixture are fully specified a priori to mimic the log-chisquared distribution. An alternative is the simulated EM method (SIEM2) with free parameters in the three elements of the mixture. The SIEM2 model allows flexibility in the error distribution, and can accommodate fat-tailed distributions of the underlying exchange rate innovation ϵ_t. Since both SIEM1 and QML are nested within the general SIEM2, these two specification can be tested by a likelihood ratio test.[5]

We estimate the likelihood value by decomposing the likelihood function as in Kim and Shephard (1994)

$$\log L(\theta|Y_T) = \log \Pr(Z_T) - \log \Pr(Z_T|Y_T;\theta) + \log f(Y_T|Z_T;\theta) \tag{8}$$

Only the second term on the right hand requires extra simulation from the multimove Gibbs sampler. The relatively straightforward way to compute the likelihood value allows us to get standard errors for the parameters. We can numerically differentiate the likelihood under the null hypothesis and consequently find both the gradient and hessian. Furthermore LM tests can be performed.

Collecting all the parameters in the vector θ the expectation step of the simulated EM algorithm gives a new estimate θ^*,

$$\theta^* = \arg\max_{\theta} \frac{1}{N} \sum_{i=1}^{N} \log f(Y_T|X_T^{(i)}, Z_T^{(i)}; \theta), \tag{9}$$

[3]See appendix B for a derivation.

[4]Increasing the number of elements in the mixture does not seem to make much difference. The seven element mixture reported in Kim and Shephard (1994) also implies a kurtosis of 14 for the normal random variable ϵ_t, although their mixture excellently describes the first four moments of ξ_t.

[5]QML is nested within SIEM2 by having the same parameters in each piece of the mixture.

where the superscript denotes the ith drawing from the Gibbs sampler. Performing the maximization in (9), the parameters ρ and σ^2 (and β in the case of fixed mixture parameters) can be found by doing an OLS regression of x_t on x_{t-1} (see appendix A. New mixture parameters are found as the conditional means of the measurement error $\xi(z_t = i)$. After estimating the parameters the multimove Gibbs sampler is used again to get new state vectors X_T. This procedure is repeated until convergence. Parameter estimates for the model with a fixed mixture distribution for ξ are obtained analogously.

The final estimation method we consider is the Bayesian Gibbs sampling algorithm, in which we cycle through simulating the states, the mixture indicators, and the parameter vector θ. In the Bayesian procedure we assume very diffuse proper conditionally conjugate priors. The prior means of μ_i and ω_i are set at the values of the approximating mixture in Table 1. The prior mean of ρ is equal to one. The prior variances of the μ_i's and ρ are equal to 1000, while the prior degrees of freedom in the inverted gamma priors for ω_i is equal to 5. We need proper priors in order to ensure a proper posterior. With an improper prior on ω_i the posterior does not exist, and it is this pathology that we must avoid. Simulation from the conditional density of θ given the states and the data works well, since the model is linear. Again all details are in appendix A.

3 Data

The data consist of weekly observations of the bilateral exchange rates among the major currencies (US dollar, UK pound sterling, Japanese yen, and German deutsche mark). The sample period is January 3, 1973 until February 9, 1994 (1102 observations). The original data are in pound sterling; cross rates have been constructed. All data are taken from DATASTREAM and are sampled on Wednesdays. If Wednesday is a holiday Thursday is taken. The data are transformed to $y_t = \ln[(\Delta \ln S_t - \bar{s})^2]$, where $\bar{s}$ is the sample mean of $\Delta \ln S_t$.

Table 2 provides summary statistics. The main features of the transformed data are the negative skewness, and the persistent autocorrelations. Negative skewness is implied by the log transformation, and related to the "inlier" problem. The variance of the transformed data is much higher than that of a log-chisquared (which is equal to $\pi^2/2$)), so that either the volatility series h_t is itself highly volatile or normality of

Table 2: Summary statistics of $y_t = \ln[(\Delta \ln S_t - \bar{s})^2]$

	JP/US	GE/US	UK/US	GE/JP	UK/JP	GE/UK
mean	-1.203	-0.770	-1.003	-0.926	-0.979	-1.494
variance	6.555	5.954	6.383	5.318	6.461	5.817
skewness	-0.971	-1.202	-0.968	-1.257	-1.278	-0.904
kurtosis	1.225	1.022	0.895	2.850	2.731	1.181
minimum	-13.03	-13.15	-11.80	-15.04	-14.93	-11.83
maximum	4.871	4.129	4.310	3.613	4.525	3.853
normality	246.7*	452.7*	208.7*	662.6*	641.8*	214.0*
LB(10)	219.2*	65.65*	277.7*	34.33*	34.90*	67.95*
LB(20)	383.6*	127.1*	413.9*	54.87*	69.24*	94.10*
			Autocorrelations y_t			
1	0.184	0.080	0.172	0.078	0.074	0.132
2	0.142	0.098	0.170	0.071	0.081	0.096
3	0.163	0.115	0.152	0.105	0.064	0.090
4	0.167	0.055	0.202	0.047	0.077	0.114
5	0.127	0.096	0.123	0.063	0.016	0.035
10	0.104	0.027	0.129	-0.012	0.019	0.053
20	0.087	0.062	0.073	0.097	0.040	0.002

NOTES: The kurtosis is measured as excess kurtosis. *Normality* is the Jarque-Bera test for normality. *LB(m)* is the Ljung-Box test taking m autocorrelations. '*' denotes significance at a 1% level.

the measurement error in (4) is violated.

We include all six possible bilateral rates, because in this way we can model all variances and all possible covariances. For example, the covariance between the yen/pound and mark/pound can be written as the identity

$$\begin{aligned} \mathrm{Cov}(\Delta \ln S^{JP/UK}, \Delta \ln S^{GE/UK}) &= \frac{1}{2}\Big(\mathrm{Var}(\Delta \ln S^{JP/UK}) \\ &\quad + \mathrm{Var}(\Delta \ln S^{GE/UK}) - \mathrm{Var}(\Delta \ln S^{JP/GE})\Big), \end{aligned}$$

which shows that the covariance between two pound denominated exchange rates can be obtained through the variance of the cross rate.

4 Parameter Estimates

Tables 3 and 4 summarise the parameter estimation results of the four estimators/ models: QML, simulated EM with fixed mixture (SIEM1), simulated EM with flexible mixture (SIEM2), and the Bayesian Gibbs sampler (BAYES).

In general the parameter estimates are similar, both across currencies as well as estimators. This is as expected due to the large sample of more than twenty years of weekly data and the tight parameterization. The high value of ρ implies persistent logarithmic volatility series. However, when the measurement error is kept fixed (QML and SIEM1) the persistence is much lower for some series, notably the pound/yen and pound/mark series. This might be due to outliers, which are harder to accomodate by QML and the fixed mixture. Another explanation is the importance of the identifying restriction $\omega^2 = \pi^2/2$ in QML and SIEM1, as we will discuss later in this section. In general the persistence of the volatility rises when the measurement error distribution is flexible. Also, when ρ is small, the estimate of the volatility innovation variance σ^2 increases. The estimates of ρ are never significantly different from unity using Dickey-Fuller critical values. The Bayesian posterior means of ρ are somewhat lower than the ML estimates in SIEM2. This is reflected in the marginal posterior densities of ρ in Figure 2, which are all skewed to the left.[6]

Since the SIEM2 model nests both the QML and the SIEM1 model, likelihood ratio tests can be computed. As reported in Table 5 both the normal (QML) and the log-chisquared distributions can be firmly rejected in every case against the free six parameter mixture. The QML and SIEM1 estimators restrict the second and higher moments, while the first moment is unrestricted (because of β). Since the tests solely involve the shape of the measurement error distribution, it is of interest to compare the first few moments of the mixture distributions. The formulas for the moments of a mixture of normals are listed in appendix B. Table 6 reports the implied moments of the measurement errors. Overall the implied variance estimates of the flexible mixture are larger than those of the log-chisquared distribution. The high variance of the measurement error ξ_t suggests that the distribution of the ϵ_t in (4) has fatter tails than the normal.[7] The BAYES moments are computed by averaging over the conditional moments in a run of the Gibbs sampler. The BAYES moments are mostly close to the maximum likelihood results, although the implied variances are somewhat smaller. There is no discernible pattern of differences for the higher order moments for the six series. Figure 3 shows a plot of the estimated measurement densities.[8] The

[6] One way to increase the posterior mean is by adopting a different prior, for instance one that is proportional to $(1-\rho^2)^{-1/2}$. Such a prior would implicitly arise, if it is assumed that the initial condition in the state vector has variance proportional to $(1-\rho^2)^{-1}$ (see Schotman (1994) and Kim and Shephard (1994)).

[7] Gallant, Hsieh and Tauchen (1994) reach a similar conclusion in an application with stock price data.

[8] The Bayesian densities are obtained as the average of the conditional densities over a run of the Gibbs sampler.

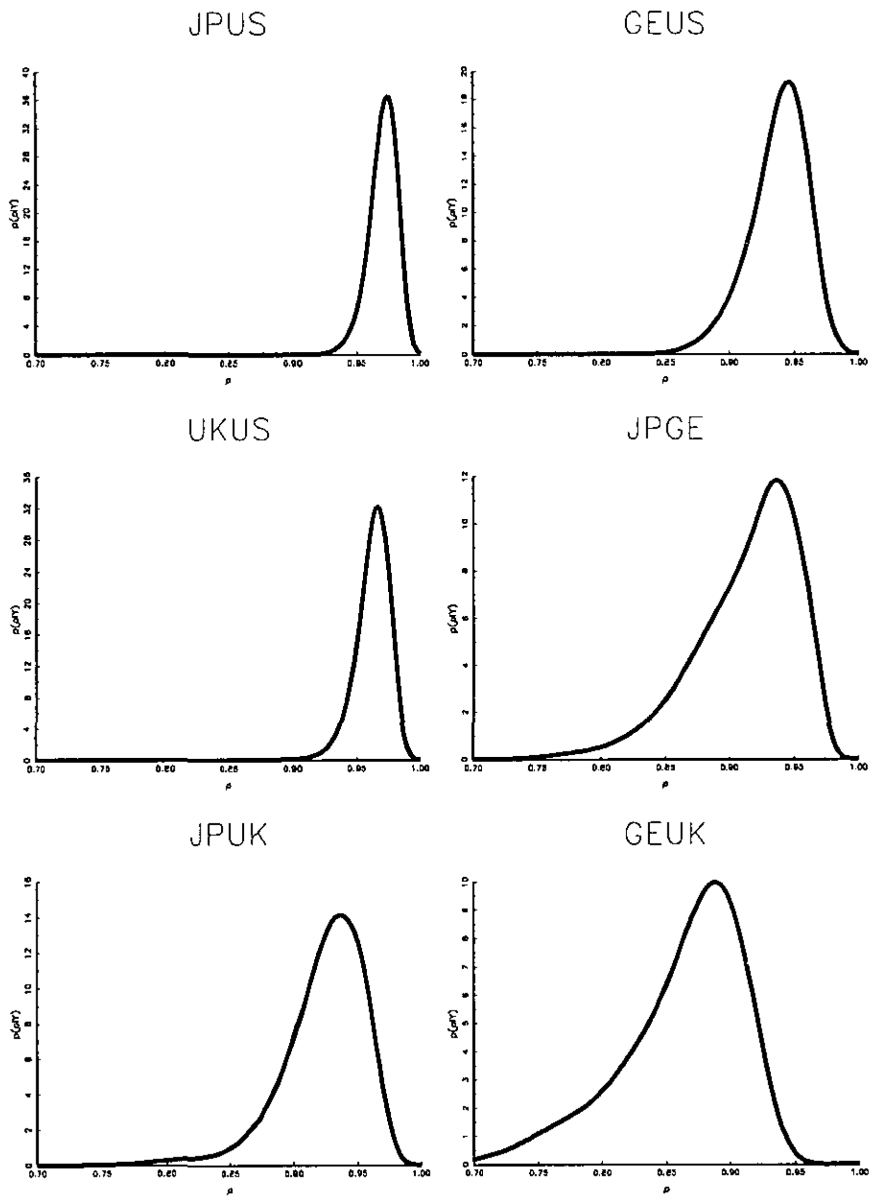


Figure 2: **Posterior densities of ρ**

These posterior densities are based on a single run of 5000 iterations from the Gibbs sampler.

Table 3: Parameter estimates of stochastic volatility model

$$
\begin{aligned}
y_t &= h_t + \xi_t & \xi_t \sim \sum_i p_i \mathrm{N}(\mu_i, \omega_i^2) \\
h_t &= \beta(1-\rho) + \rho h_{t-1} + \eta_t & \eta_t \sim \mathrm{N}(0, \sigma_2)
\end{aligned}
$$

	JP/US	GE/US	UK/US	GE/JP	UK/JP	GE/UK
Quasi Maximum Likelihood (QML)						
ρ	0.976	0.967	0.960	0.985	0.653	0.884
	(0.015)	(0.027)	(0.015)	(0.009)	(0.391)	(0.061)
β	0.050	0.482	0.238	0.369	0.292	-0.226
	(0.211)	(0.199)	(0.258)	(0.190)	(0.101)	(0.130)
σ	0.225	0.198	0.311	0.092	0.842	0.419
	(0.081)	(0.097)	(0.055)	(0.027)	(0.612)	(0.142)
Simulated EM, fixed mixture (SIEM1)						
ρ	0.878	0.928	0.952	0.921	0.584	0.768
	(0.019)	(0.018)	(0.010)	(0.021)	(0.048)	(0.034)
β	0.150	0.591	0.395	0.362	0.349	-0.194
	(0.138)	(0.123)	(0.200)	(0.111)	(0.068)	(0.090)
σ	0.558	0.293	0.330	0.285	0.822	0.650
	(0.040)	(0.036)	(0.029)	(0.039)	(0.055)	(0.050)
Simulated EM, flexible mixture (SIEM2)						
ρ	0.979	0.975	0.967	0.954	0.957	0.930
	(0.008)	(0.010)	(0.009)	(0.016)	(0.016)	(0.020)
σ	0.190	0.148	0.251	0.200	0.190	0.285
	(0.029)	(0.027)	(0.028)	(0.035)	(0.040)	(0.042)
μ_1	-0.042	0.410	0.205	0.102	0.133	-0.353
	(0.266)	(0.176)	(0.227)	(0.137)	(0.140)	(0.129)
μ_2	-3.122	-2.724	-2.190	-2.547	-2.736	-3.441
	(0.282)	(0.199)	(0.230)	(0.156)	(0.174)	(0.144)
μ_3	-6.651	-6.877	-6.561	-6.819	-7.355	-6.048
	(0.429)	(0.416)	(0.282)	(0.344)	(0.447)	(0.244)
ω_1	1.755	1.281	1.417	1.277	1.775	1.512
	(0.097)	(0.075)	(0.081)	(0.080)	(0.105)	(0.090)
ω_2	2.578	2.759	1.537	2.040	3.439	1.436
	(0.230)	(0.234)	(0.133)	(0.182)	(0.297)	(0.144)
ω_3	6.134	7.981	2.005	5.865	10.085	2.385
	(1.214)	(1.520)	(0.377)	(0.865)	(1.840)	(0.403)

NOTES: Numbers in parenthesis are robust standard errors. Parameters in fixed mixture are as given in Table 1. The mixture weights were set to (0.70, 0.25, 0.05). In the QML model ξ_t has mean -1.27 and variance $\pi^2/2$. In the SIEM2 model β is restricted to zero for identification.

Table 4: Gibbs Sampler: Bayesian Posterior Moments

$$\begin{aligned} y_t &= h_t + \xi_t \qquad \xi_t \sim \sum_i p_i \mathrm{N}(\mu_i, \omega_i^2) \\ h_t &= \rho h_{t-1} + \eta_t \quad \eta_t \sim \mathrm{N}(0, \sigma_2) \end{aligned}$$

	JP/US	GE/US	UK/US	GE/JP	UK/JP	GE/UK
ρ	0.971	0.938	0.963	0.912	0.924	0.862
	(0.013)	(0.022)	(0.014)	(0.041)	(0.033)	(0.050)
σ	0.274	0.267	0.285	0.300	0.270	0.454
	(0.046)	(0.047)	(0.044)	(0.082)	(0.059)	(0.101)
μ_1	-0.091	0.359	0.188	0.092	0.113	-0.390
μ_2	-3.137	-2.778	-2.800	-2.587	-2.796	-3.415
μ_3	-6.389	-5.906	-6.727	-6.574	-6.807	-6.883
ω_1	1.315	1.113	1.208	1.108	1.305	1.174
ω_2	1.691	1.754	1.366	1.474	1.847	1.374
ω_3	2.706	2.979	3.182	2.534	3.220	2.146

NOTES: Posterior means and standard deviations (in parentheses) of the parameters are based on a run of 5000 simulations from the Gibbs sampler. The weights of the mixture are (0.70, 0.25, 0.05).

Table 5: Likelihood Ratio Tests

	JP/US	GE/US	UK/US	GE/JP	UK/JP	GE/UK
QML (normal)	153.0	263.6	134.4	227.0	244.2	109.0
SIEM1 ($\ln(\chi^2)$)	153.4	76.0	42.4	50.4	234.4	108.2

NOTE: The likelihood ratio statistic has a $\chi^2(5)$ distribution for both hypotheses. The 5% and 1% critical values are 11.1 and 15.1 respectively.

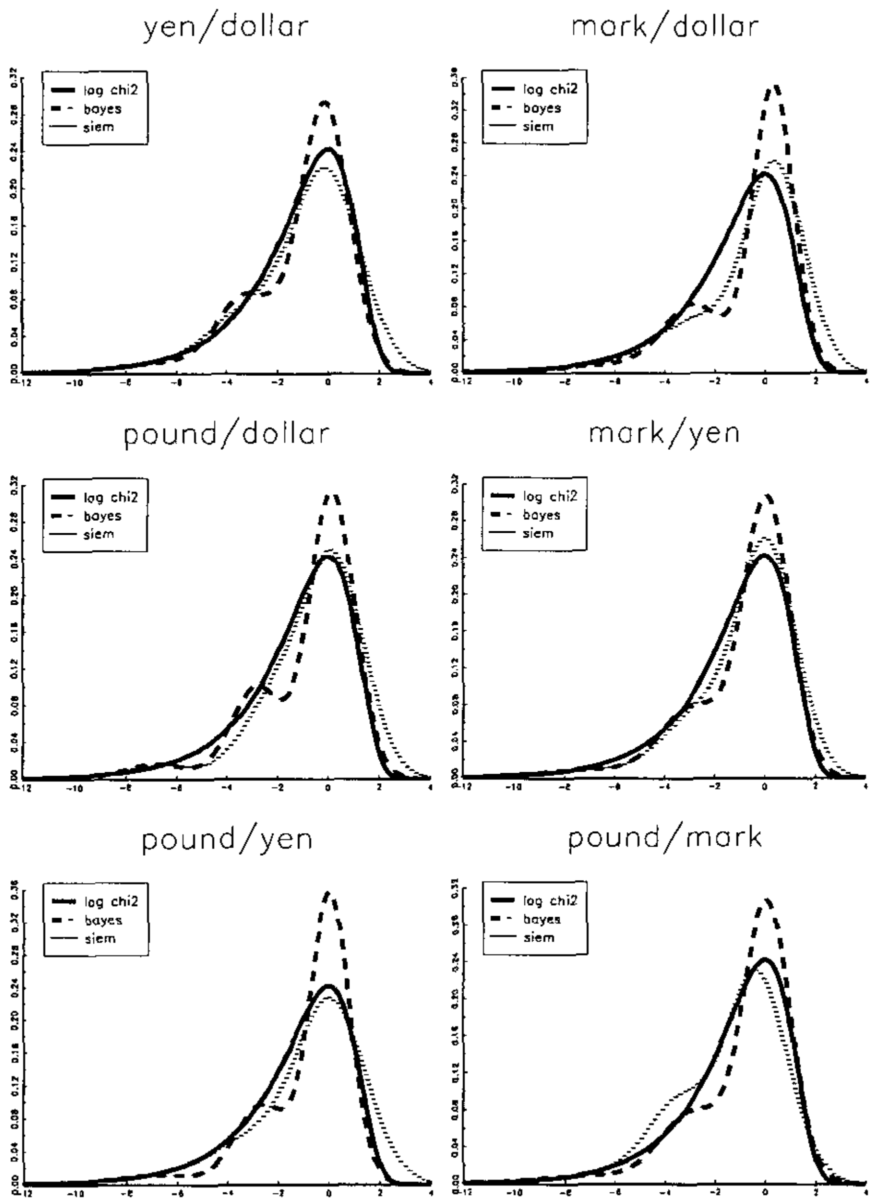


Figure 3: **Implied measurement error densities**

The figures show the mixture densities from the SIEM2 and Bayesian algorithms compared with the log chi-squared.

Table 6: Moments of Measurement Error Distribution

	JP/US	GE/US	UK/US	GE/JP	UK/JP	GE/UK	SIEM1	QML
Variance								
SIEM2	5.524	5.778	5.042	4.830	6.276	5.289	4.843	4.935
BAYES	5.511	5.349	5.087	4.767	5.987	4.937		
Skewness								
SIEM2	-1.094	-1.500	-1.082	-1.482	-1.512	-0.681	-1.478	0
BAYES	-1.133	-1.386	-1.106	-1.475	-1.436	-1.142		
Excess Kurtosis								
SIEM2	1.847	3.244	2.129	3.454	3.776	0.842	3.514	0
BAYES	2.045	2.741	1.608	3.422	3.457	1.860		

NOTES: The SIEM2 moments are computed using the parameter estimates in Table 3. The SIEM1 moments are the same for all exchange rates and were fixed in advance using the parameters in Table 1. The BAYES moments are computed as the average of the conditional moments over the iterations of the Gibbs sampler. Skewness and kurtosis are normalized, and the reported kurtosis is in excess of the normal kurtosis.

BAYES densities have a somewhat sharper mode, and are sometimes still bimodal.

In Table 7 we present some selected moment statistics of the actual data in order to further compare the four models. The main conclusion from the table is that all four estimators/models fit well for the mean and variance of the log transformed data. Major differences between the estimators show up for the excess kurtosis, where the BAYES estimator implies much less kurtosis than is actually in the data.

The implied autocorrelations are curious. For example, QML and SIEM1 provide the lowest estimates of ρ for the pound/yen volatility in Table 3. But even so they imply the highest first order autocorrelations for the data. The explanation is that the measurement error variance has been fixed for QML and SIEM1, and is much lower than the unrestricted estimate of SIEM2. Since y_t follows an ARMA(1,1) process, parameterized by ρ, σ^2 and ω^2, restricting $\omega^2 = \pi^2/2$, as in QML and SIEM1, will affect the estimates of the remaining two parameters ρ and σ^2, even asymptotically. The first order autocorrelation of y_t is

$$\mathrm{AR}(1) = \mathrm{corr}(y_t, y_{t-1}) = \frac{\rho\sigma_x^2}{\sigma_x^2 + \omega^2}$$

Keeping the variance of y_t fixed, $\mathrm{Var}(y) = \sigma_x^2 + \omega^2$, we obtain

$$\mathrm{AR}(1) = \rho(1 - \omega^2/\mathrm{Var}(y)) \tag{10}$$

From (10) we see that increasing ω^2 must lead to a larger value of ρ to keep the sample first order autocorrelation AR(1) constant. This is exactly what we find in

Table 7: Data and implied model moments

	JP/US	GE/US	UK/US	GE/JP	UK/JP	GE/UK
			Mean			
$\ln(y^2)$	-1.203	-0.770	-1.003	-0.926	-0.979	-1.494
QML	-1.220	-0.788	-1.032	-0.901	-0.978	-1.496
SIEM1	-1.088	-0.647	-0.843	-0.876	-0.889	-1.432
SIEM2	-1.142	-0.738	-0.732	-0.906	-0.959	-1.410
BAYES	-1.167	-0.738	-0.905	-0.911	-0.960	-1.471
			Variance			
$\ln(y^2)$	6.555	5.954	6.383	5.318	6.461	5.817
QML	6.002	5.539	6.168	5.219	6.171	5.738
SIEM1	6.201	5.463	6.007	5.377	5.868	5.875
SIEM2	6.435	6.117	6.013	5.275	6.704	5.889
BAYES	6.629	5.995	6.322	5.349	6.530	5.779
			Skewness			
$\ln(y^2)$	-0.971	-1.202	-0.968	-1.257	-1.278	-0.904
QML	0	0	0	0	0	0
SIEM1	-1.020	-1.234	-1.070	-1.264	-1.108	-1.106
SIEM2	-0.886	-1.402	-0.831	-1.230	-1.369	-0.580
BAYES	-0.865	-1.177	-0.800	-1.252	-1.273	-0.907
			Kurtosis			
$\ln(y^2)$	1.225	2.022	0.895	2.850	2.731	1.181
QML	0	0	0	0	0	0
SIEM1	1.127	2.244	1.441	2.350	1.598	1.363
SIEM2	0.649	2.583	1.786	2.457	2.959	0.039
BAYES	0.337	1.544	-0.123	2.065	2.385	0.356
			First order autocorrelation			
$\ln(y^2)$	0.184	0.080	0.172	0.078	0.074	0.132
QML	0.174	0.105	0.192	0.054	0.131	0.124
SIEM1	0.192	0.105	0.184	0.091	0.102	0.135
SIEM2	0.132	0.071	0.156	0.080	0.061	0.095
BAYES	0.164	0.101	0.189	0.099	0.077	0.125
			Second order autocorrelation			
$\ln(y^2)$	0.142	0.098	0.170	0.071	0.081	0.096
QML	0.170	0.102	0.184	0.053	0.086	0.104
SIEM1	0.169	0.098	0.176	0.084	0.060	0.104
SIEM2	0.129	0.069	0.151	0.077	0.058	0.088
BAYES	0.160	0.095	0.182	0.091	0.071	0.108

NOTES: QML, SIEM1, SIEM2, BAYES represent the Quasi Maximum Likelihood, the Simulated EM with fixed and free mixture parameters and the Bayesian methods for estimating the univariate SV model, respectively. See text for details on computing moments.

Table 8: Diagnostic LM tests

	JP/US	GE/US	UK/US	GE/JP	UK/JP	UK/GE
Cross volatility tests						
JP/US	—	6.34	4.49	4.49	7.32	0.70
GE/US	0.88	—	9.57	4.89	1.29	0.68
UK/US	3.39	5.78	—	3.07	9.27	1.35
GE/JP	0.61	8.85	1.92	—	7.79	3.84
UK/JP	4.56	1.96	4.41	16.52	—	8.24
UK/GE	0.17	0.17	1.04	1.74	2.92	—
Higher order dynamics						
x_{t-2}	63.11	68.78	98.39	82.35	73.45	85.56

NOTES: The LM statistics test for inclusion of a cross-volatility x^*_{t-1}, or a further lag of own volatility in the volatility equation (x_{t-2}). The scores are all computed by simulation under the null hypothesis of the free mixture model.

the estimates in Table 3, and most clearly for the pound/yen rate. In general ω^2 will be higher than $\pi^2/2$ if the distribution of the underlying exchange innovation ϵ_t has fat tails. This way the distribution of ϵ_t and the estimates of the volatility persistence are connected. QML and SIEM1 can be seen as a restriction on one of the parameters in an ARMA(1,1) process.[9]

An ARMA(1,1) process is fully identified by the variance and the first two autocorrelations. Ignoring all distributional identifying restrictions one could estimate the parameters of the stochastic volatility model from these three moments. However, the second order autocorrelation is larger than the first order autocorrelation, so that the implied estimate $\rho = \text{AR}(2)/\text{AR}(1)$ is larger than one and thus infeasible. This means that a GMM estimator with more moments is required. Given the number of observations the GMM criterion function will reject the overidentifying moment conditions, implying misspecification of the simple first order stochastic volatility model. The same conclusion also follows from comparing the sample autocorrelations and the implied autocorrelations in Table 7. None of the estimators provides a uniformly good fit for all six exchange rates.[10]

[9] One could consider a modified QML estimator where the measurement error variance ω^2 is a free parameter. This estimator was used in Mahieu and Schotman (1994). The modified QML estimator does yield higher estimates of the volatility persistence ρ, but generally not as high as SIEM2. The implicit outlier correction of SIEM2, due to the high estimated kurtosis, must explain the remaining (small) differences.

[10] Since the model does not seem to fit the lower order moments care must be taken in a Generalized Method of Moments estimation. Both the choice and the number of moments seem to be important. Gallant, Hsieh and Tauchen (1994) develop a statistical method to choose the moments. The score function of an auxiliary probability model is used to employ a standard GMM analysis. Melino and

Instead of a GMM based misspecification analysis, we computed a set of Lagrange Multiplier test, taking the free mixture SIEM2 model as the null hypothesis. The tests take the form of adding an explanatory variable in the volatility equation (2),

$$x_t = \rho x_{t-1} + \gamma x^*_{t-1} + \eta_t, \tag{11}$$

where x^*_{t-1} is the additional variable. Candidate additional variables are a second lag of the own volatility (x_{t-2}) of a currency in order to test for dynamic misspecification, and the lagged volatility of another exchange in order to test for spillover effects. We computed LM statistics using the simulated states from the SIEM2 model and numerical differentiation of the likelihood. The scores were then computed from a simulation with the multimove Gibbs sampler under the null hypothesis. The statistics are asymptotically distributed as a chi-square with one degree of freedom.

The table shows that spillovers seem to be present in a number of cases, especially for the problematical pound/yen rate. Note also that among exchange rates that do not have a certain currency in common (the anti-diagonal in the table) no evidence for spillovers exist. These results direct the interest towards a multivariate model of exchange rates as in Mahieu and Schotman (1994). Dynamic misspecification of the volatility process is rejected overwhelmingly for all series.

5 Volatility Estimates

Even though the first order stochastic volatility model appears misspecified, it has a strong theoretical appeal in finance applications, and remains an improvement on the assumption of constant volatility. It is therefore of interest to investigate the estimated volatilities of the different estimators.

The volatilities can be estimated in three different ways. Conditional on the parameters the Kalman smoother produces estimates under the assumption that the measurement error ξ_t is normally distributed. An alternative is to run the Gibbs sampler conditional on the parameters. The Gibbs sampler alternates between simulating from the states using $f(X_T|Z_T, \theta)$ and the mixture indicators $f(Z_T|X_t, \theta)$ like in the moment steps of the Simulated EM algorithm. The third volatility estimate is the Bayesian posterior mean of X_T, which is obtained as part of the full Gibbs sampler in the BAYES estimator. The first two volatility estimates are conditional

Turnbull (1990) use 47 moments in estimating a simple SV model.

Table 9: Summary statistics of volatility estimates

	JP/US	GE/US	UK/US	GE/JP	UK/JP	GE/UK
SIEM2 simulation estimates: $\hat{x}_t$						
SD($\hat{x}_t$)	0.837	0.576	0.938	0.540	0.508	0.586
$\hat{\sigma}_x$	0.932	0.666	0.985	0.667	0.655	0.775
Standard errors of $\hat{x}_t$: $\bar{P}_t$						
SIEM2	0.421	0.343	0.459	0.394	0.408	0.487
BAYES	0.549	0.485	0.578	0.497	0.497	0.626
SIEM2 (simulated) minus SIEM2 (Kalman)						
std. dev.	0.244	0.266	0.244	0.241	0.505	0.240
AR(1)	0.082	0.121	0.041	0.067	0.647	0.063
SIEM2 (simulated) minus QML (Kalman)						
std. dev.	0.248	0.267	0.270	0.385	0.330	0.264
AR(1)	0.156	0.213	0.195	0.512	0.321	0.254
SIEM2 (Kalman) minus QML (Kalman)						
std. dev.	0.082	0.111	0.121	0.250	0.450	0.159
AR(1)	0.018	0.004	0.055	0.038	0.759	0.061
SIEM2 (Kalman) minus BAYES						
std. dev.	0.091	0.145	0.103	0.106	0.082	0.166
AR(1)	0.110	0.379	0.075	0.230	0.136	0.445

NOTES: SD($\hat{x}_t$) is the sample standard deviations of the time series $\hat{x}_t$ of estimated log-volatilities, using the multimove Gibbs sampler as the smoothing algorithm; σ_x is the estimated standard deviation of x_t using the parameter estimates in table 3; $\bar{P}_t$ is the square root of the sample mean of the estimated error variances of $x_t - \hat{x}_t$. The lower part of the table provides information on the time series of differences between the SIEM2 simulated states and estimates obtained by the standard Kalman smoother. AR(1) is the first order autocorrelation of the series of deviations.

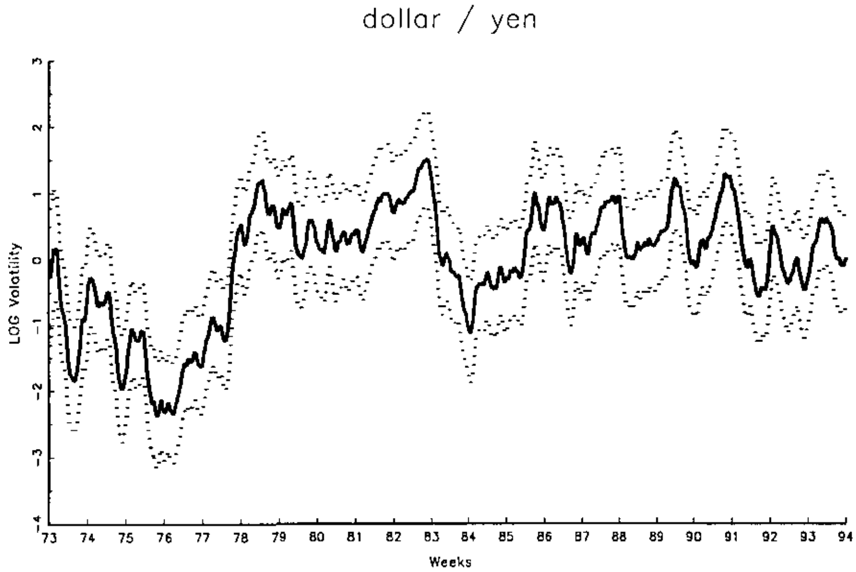
on the parameters, while the Bayesian estimate incorporates parameter uncertainty. The simulation estimators also provide estimates of the actual variances $\exp(x_t)$.[11]

In table 9 we present the some summary statistics of the estimated log-volatilities. The total variance σ_x^2 can approximately be split in two components,

$$\sigma_x^2 = \mathrm{Var}(\hat{x}_t) + \bar{P}^2,$$

where $\bar{P}^2$ is the sample average of the variances of the estimation errors $(x_t - \hat{x}_t)$, which is obtained for each period t as the variance of the simulated $x_t^{(i)}$. All three

[11]Because the unconditional mean β is not identified in the BAYES and SIEM2 models, we subtract the sample mean of the smoothed x_t. Similarly we also correct the sample mean of the estimated $\exp(x_t)$ by multiplying with a constant such that it equals the unconditional sample variance of the exchange rates $\Delta \ln S_t$.



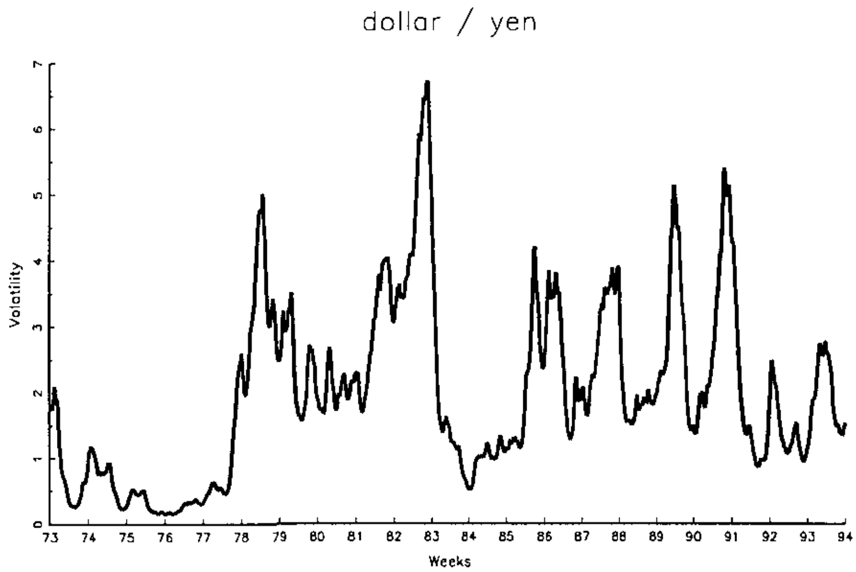


Figure 4: **dollar / yen**

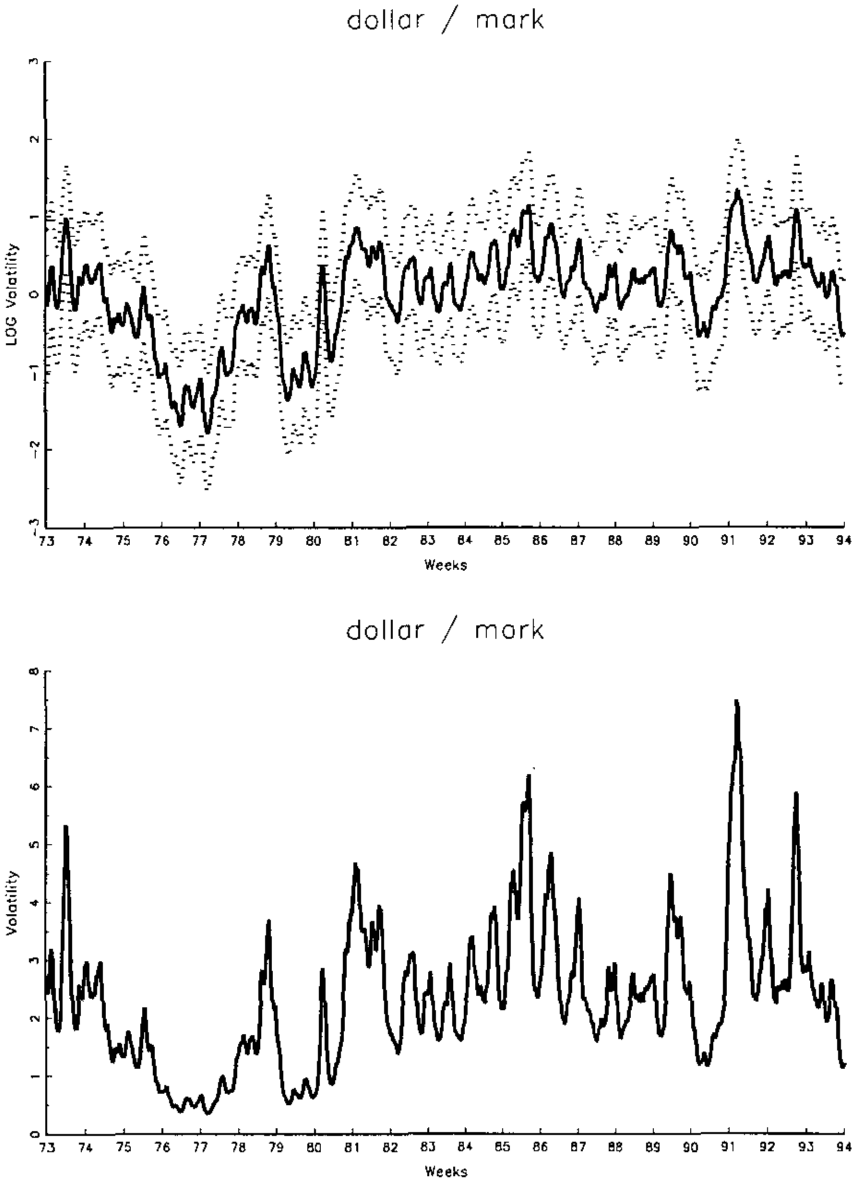
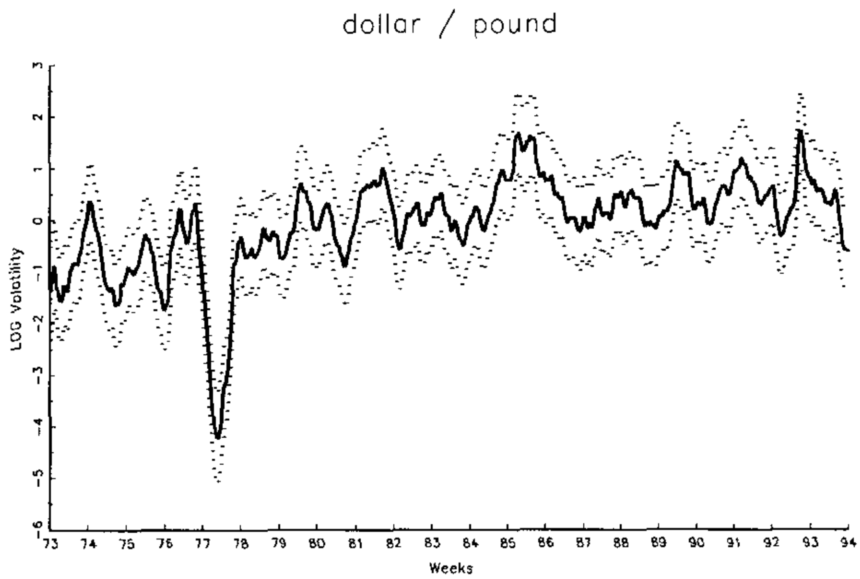


Figure 5: **dollar / mark**



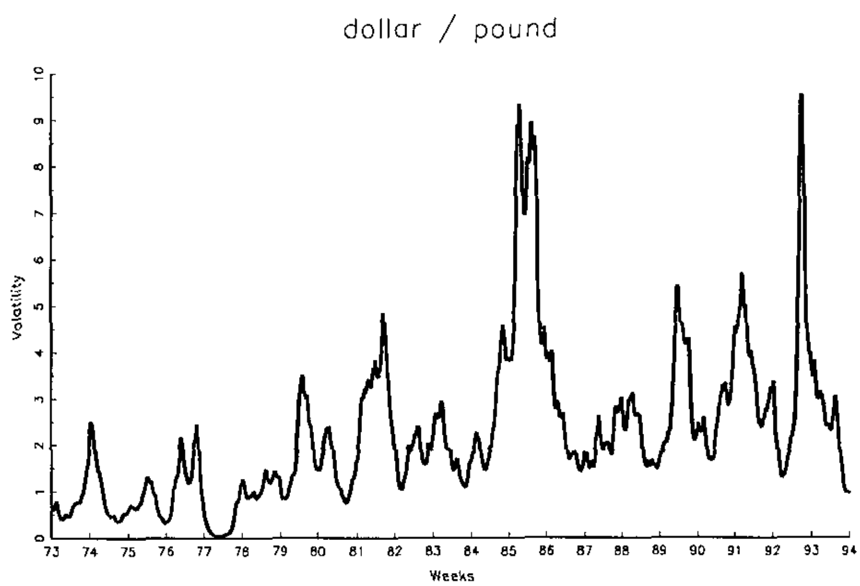
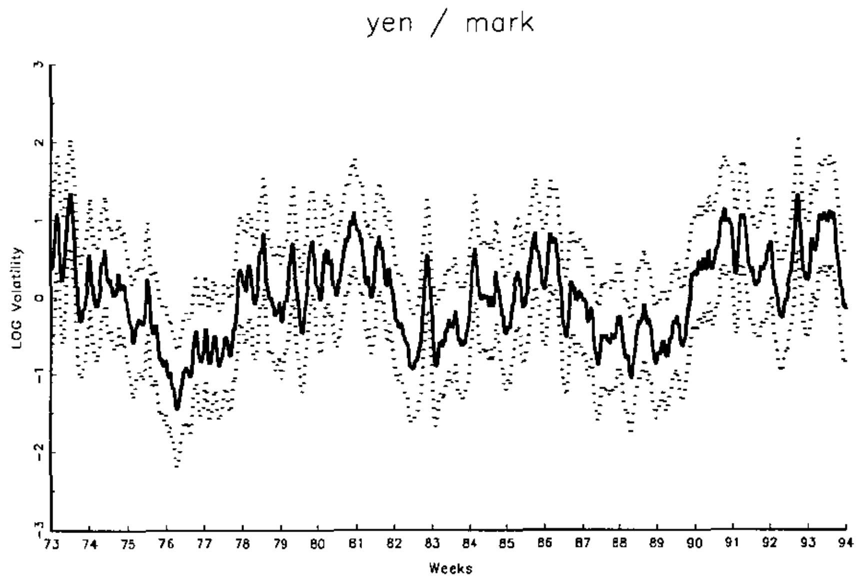


Figure 6: **dollar / pound**



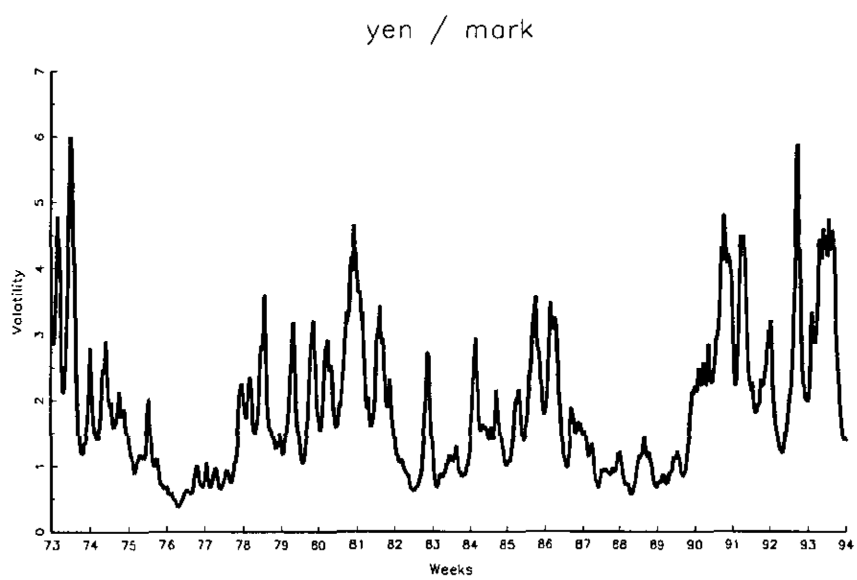
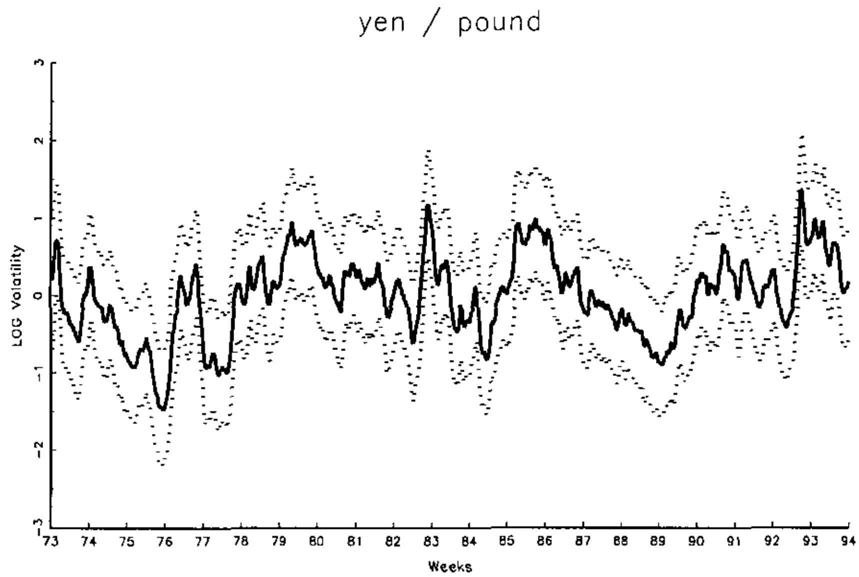


Figure 7: **mark / yen**



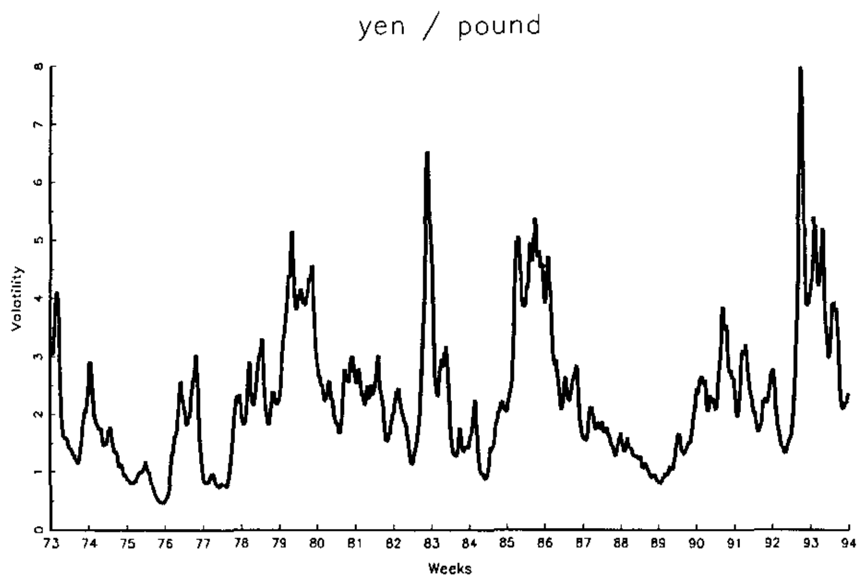


Figure 8: **pound / yen**

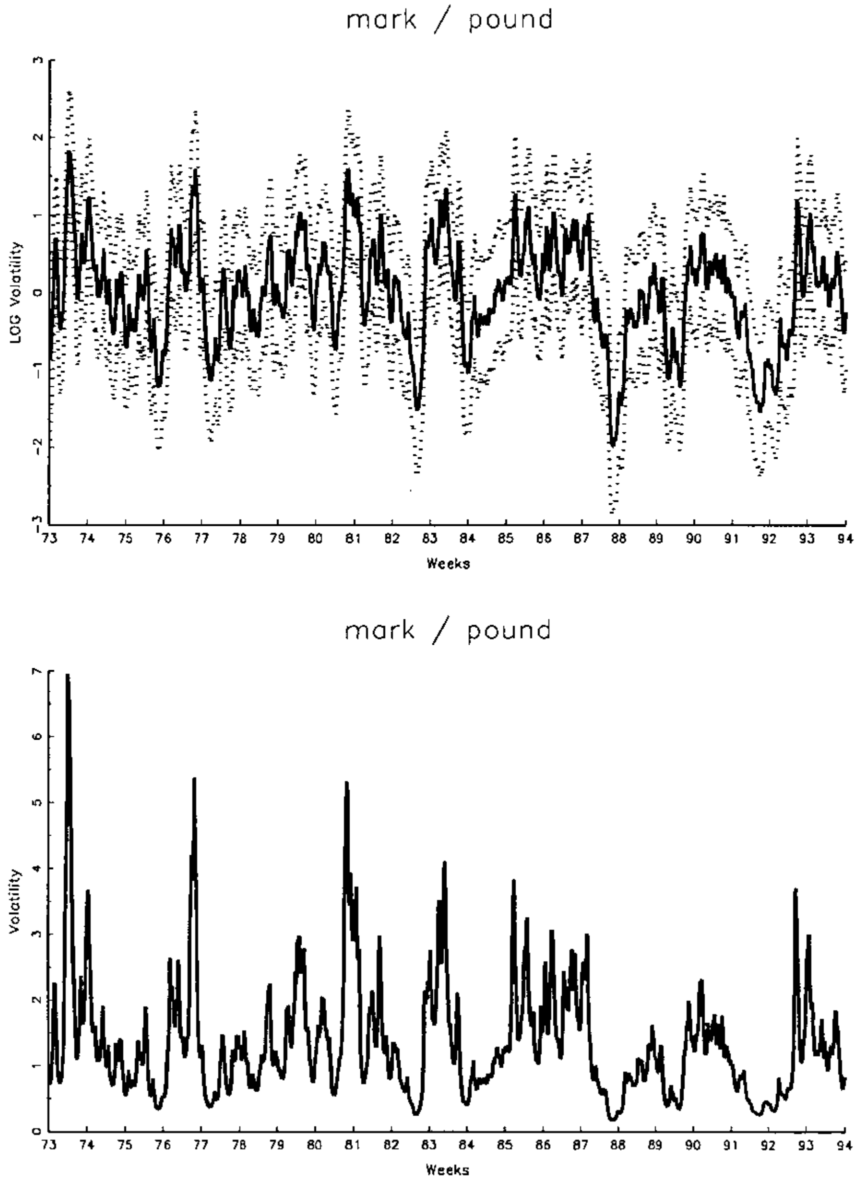


Figure 9: **mark / pound**

components are reported in table 9. The lowest relative estimation error variance $\bar{P}^2/\sigma_x^2$ is obtained for the dollar/pound rate, where the sample variance of $\hat{x}_t$ accounts for 90% of the total variation of the volatility. The highest error ratios are for the cross rates, where the sample variance of $\hat{x}_t$ is only between 60% and 70% of the total variance. The estimation error, conditional on the parameters, is also big in absolute value. Noting that the numbers refer to log-volatility we find that an average 95% confidence interval for a weekly volatility ranges roughly from a half to two times the estimated volatility. The BAYES estimator has even larger standard errors as it also incorporates the parameter uncertainty. For any financial application this is a very wide range, and would also lead to enormous price ranges for implied option prices. These numbers are further evidence for the scope for much improvement in volatility models.

The lower part of the table considers the differences between the SIEM2 simulation smoother, and alternative smoothing algorithms. Given that all series have mean zero by construction, we use the standard deviation of the time series of the difference between two series as a distance measure. These standard errors are large. Even if the same parameters are used, the Kalman smoother produces a series that deviates strongly from the simulation smoother. The differences between the BAYES estimator and the SIEM2 simulation smoother are much smaller. There are also much smaller differences between the two series produced by the Kalman smoother. From this we conclude that the issue of the right smoother is much more important than the parameter estimator in this model with such long time series.

Table 9 also shows that the dollar/yen and dollar/pound exchange rates exhibit much wider fluctuations in the volatility than the other four series. The fluctuations are shown in figures 4 to 9, which present the smoothed log volatilities x_t of the BAYES estimator in the top panel, and the actual volatilities $\exp(x_t)$ (scaled to match the unconditional sample variance) in the lower panel.[12] The dollar/pound volatility has a large dip in 1977, when the exchange rate was very stable. The peaks of the dollar/pound volatility in 1985 and 1993 appear very sharply in the lower panel. The high pound volatility starting in the fall of 1992 is also visible in the pound/mark and pound/yen rates, identifying this episode as pound volatility.

[12]The smoothed volatilities in the lower panel have been obained as the average of the simulated $\exp(x_t^{(i)})$ over a run of the Gibbs sampler.

6 Summary and Conclusions

In the paper we empirically studied the performance of the first order stochastic volatility model using a dataset of weekly exchange rates over the last twenty years. The model has been estimated by estimation methods that differ in the specification of the distribution of the standardized exchange rate innovations. Our first finding is that for the number of observations we employ the estimation is of secondary importance for the problem of estimating the key parameters. The only empirical regularity is that the point estimates of the persistence of the volatility increase if the error distribution is explicitly modelled through a mixture of normals.

The other results of the paper pertain to the estimation of the time series of volatilities. First we find that different smoothing algorithms produce very different estimates, even if the parameters of the underlying process are the same. Again the differences arise from explicit consideration of the measurement error density in the state space model for the log volatility. The most disturbing finding is that even the most efficient simulation smoothers produce very large standard errors for the volatility estimates. From this we conjecture that implied option prices in various finance applications will also be subject to large estimation error. We leave the finance implications for future work.

Finally, diagnostic tests reveal several forms of misspecification of the standard stochastic volatility model. Multivariate specifications and higher order dynamics seem particularly attractive empirical extensions.

APPENDIX

A Simulation Algorithms

The algorithms listed below draw heavily on Shephard (1993b, 1994). The Bayesian Gibbs sampler for the stochastic volatility model has the following steps:

1. Let $\mu_t = \mu_{z_t}$ and $\omega_t^2 = \omega_{z_t}^2$, $t = 1, \ldots, T$ be the mean and variance associated with the current draw for the index of the element of the mixture distribution. Conditional on z_t and the parameters σ, μ, and ω the Kalman filter produces an estimate of the filtered states $\hat{x}_t = \mathrm{E}[x_t|Y_t]$. The recursion is

$$\hat{x}_t = \rho\hat{x}_{t-1} + \frac{p_t^2}{\omega_t^2}(y_t - \rho\hat{x}_{t-1} - \mu_t) \qquad \text{(A1)}$$

$$p_t^2 \quad = \quad \frac{\omega_t^2(\rho^2 p_{t-1}^2 + \sigma^2)}{\rho^2 p_{t-1}^2 + \sigma^2 + \omega_t^2}, \tag{A2}$$

where p_t^2 is the conditional variance of $\hat{x}_t$. The initial conditions of the filter are:

$$\hat{x}_1 = y_1 - \mu_1$$
$$p_1^2 = \omega_1^2$$

2. Given the filtered estimates $\hat{x}_T$ and p_T, the final state x_T is generated as a draw from $\mathrm{N}(\hat{x}_T, p_T)$. A backward recursion produces the new time series of simulated states. The recursion is taken from Shephard (1994). For $t = T-1, \ldots, 1$, generate

$$x_t \sim \mathrm{N}\left[\hat{x}_t + \frac{\rho p_t^2}{\rho^2 p_t^2 + \sigma^2}(x_{t+1} - \rho \hat{x}_t)\ ,\ \frac{\sigma^2}{\rho^2 p_t^2 + \sigma^2}\right] \tag{A3}$$

3. Given a simulation for the time series $X_T = (x_1, \ldots x_T)'$ a new draw for the autocorrelation parameter ρ is generated from the conditional posterior of ρ, which is truncated normal with mean and variance parameters $\hat{\rho}$ and $\hat{S}_{\hat{\rho}}$ that are obtained by the OLS regression of x_t on x_{t-1}. Since the prior of ρ is truncated to the interval (-1,1) a draw from the conditional normal is only accepted if it falls in this region.

4. Given ρ and X_T a new value for σ is drawn from the inverted Gamma distribution with degrees of freedom $(T - 1 + d_0)$ and sum of squares $(s_0 + \mathrm{SSR})$, where SSR is the sum of squared residuals of the regression of x_t on x_{t-1} in 3, and where d_0 and s_0 are the prior degrees of freedom and prior sum of squares in the prior on σ.

5. Conditional on the data Y_T and the states X_T, the innovations ξ_t are observable as $y_t - x_t$. Given the regime indicators z_t, the posterior of the means μ_i $(i = 1, 2, 3)$ of the mixtures is implicitly given by the regression models

$$\xi(i) = \mu_i \iota + v(i), \tag{A4}$$

where $\xi(i)$ denotes a vector of length T_i with elements corresponding to $z_t = i$, and where ι is a T_i vector of ones. The error term in this model has mean zero and variance ω_i^2. Since the prior on μ_i is $\mathrm{N}(m_i, V_{0i}^2)$, the conditional posterior of μ_i is also normal with mean and variance given by

$$\hat{\mu}_i = \frac{T_i V_{0i}^2 \bar{\xi}_i + \omega_i^2 m_i}{T_i V_{0i}^2 + \omega_i^2}, \quad \text{and} \quad \left(\frac{1}{V_{0i}^2} + \frac{T_i}{\omega_i^2}\right)^{-1},$$

respectively, with $\bar{\xi}_i$ the sample mean of $\xi(i)$. These densities are used to draw new values for the μ_i. The unconditional mean β of ξ follows as $\beta = \sum_{i=1}^{K} p_i \mu_i$.

6. Analogolously to drawing σ we draw ω_i from an inverted Gamma distribution, in this case with degrees of freedom parameter $(T_i - 1 + d_0)$ and sum of squares $(\omega_{0i}^2 + \mathrm{SSR}_i)$.

7. The joint distribution of ξ_t and z_t is given by

$$f(\xi, z) \propto P_i \frac{1}{\omega_i} \exp\left(-\frac{(\xi - \mu_i)^2}{2\omega_i^2}\right), \quad \text{if } z = i, \quad i = 1, \ldots, N$$

with the same constant of proportionality for all $i = 1, \ldots, K$. The conditional distribution of z given ξ is multinomial and follows from the posterior odds ratios

$$K_{ij} = \frac{P_i}{P_j} \frac{\omega_i^{-1}}{\omega_j^{-1}} \exp\left(-\frac{(\xi - \mu_i)^2}{2\omega_i^2} + \frac{(\xi - \mu_j)^2}{2\omega_j^2}\right) \tag{A5}$$

Conditional on ξ_t we draw a new value for z_t by drawing a uniform random number u, and inverting the cumulative distribution for each t. Defining $K_{ii} = 1$ the posterior probabilities of $z = i$ follow as

$$\hat{P}_j = \left(\sum_{i=1}^{K} K_{ij}\right)^{-1} \tag{A6}$$

This completes a single iteration of the Gibbs sampler. The results in the empirical application are based on a single run of 5200 iterations, of which the first 200 were discarded.

The Simulated EM algorithm (SIEM) with free and fixed mixture parameters is set out below. Like in the Bayesian case this algorithm is an application of the multimove Gibbs sampler.

1. Conditional on the parameter vector θ the simulation smoother, consisting of repeating steps 1, 2 and 7 above, is executed in N parallel runs, each run taking 5 steps starting from the Kalman smoother solution. This way we obtain simulated states $x_t^{(j)}$.

2. New estimates of the parameters in the transition equation are obtained as (see Shephard (1993a)):

$$\hat{\rho} = \frac{\sum_{j=1}^{N} \sum_{t=2}^{T} x_t^{(j)} x_{t-1}^{(j)}}{\sum_{j=1}^{N} \sum_{t=2}^{T} (x_{t-1}^{(j)})^2} \tag{A7}$$

$$\hat{\sigma}^2 = \frac{1}{(T-1)N} \sum_{j=1}^{N} \sum_{t=2}^{T} (x_t^{(j)} - \hat{\rho} x_{t-1}^{(j)})^2 \tag{A8}$$

3. New estimates of the mixture parameters are obtained from the conditional moments of the simulated measurement errors given the mixture indicators:

$$\hat{\mu}_i = \frac{\sum_{j=1}^{N} \sum_{t \in I_i^{(j)}} \xi_t^{(j)}}{\sum_{j=1}^{N} T_i^{(j)}} \tag{A9}$$

$$\hat{\omega}_i^2 = \frac{\sum_{j=1}^{N} \sum_{t \in I_i^{(j)}} (\xi_t^{(j)} - \hat{\mu}_i)^2}{\sum_{j=1}^{N} T_i^{(j)}}, \tag{A10}$$

where $I_i^{(j)} = \{t : z_t^{(j)} = i\}$.

The number of iterations N of the Gibbs sampler at each iteration of the Gibbs sampler was set at 128 initially, increased to 1024 when the parameter estimates had reached a stable region. The starting values for the SIEM1 algorithm (which does not contain step 3) are at the final QML estimates. The starting values for the SIEM2 estimator are the final estimates of the SIEM1 estimators.

B Moments

The first four non-central moments of a random variable z distributed as a mixture of normals with means and variances μ_i and $\omega_i^2, (i=1,\ldots,K)$ are given by

$$\mathrm{E}[z^j] = \sum_{i=1}^{K} p_i (j^{th} \text{ moment of } \mathrm{N}(\mu_i, \omega_i^2)),$$

where p_i is the weight of the i^{th} normal density in the mixture. This results in the following four moments

$$\begin{aligned} m_1 = \mathrm{E}[z_t] &= \sum_{i=1}^{K} p_i \mu_i \\ m_2 = \mathrm{E}[z_t^2] &= \sum_{i=1}^{K} p_i \left(\omega_i^2 + \mu_i^2\right) \\ m_3 = \mathrm{E}[z_t^3] &= \sum_{i=1}^{K} p_i \left(3\mu_i\omega_i^2 + \mu_i^3\right) \\ m_4 = \mathrm{E}[z_t^4] &= \sum_{i=1}^{K} p_i \left(3\omega_i^4 + 6\mu_i^2\omega_i^2 + \mu_i^4\right) \end{aligned} \tag{B1}$$

The non-central moments of the independent variable of the simple univariate stochastic volatility models can be found using the moments above. Consider the model

$$\begin{aligned} y_t &= x_t + \xi_t \\ x_t &= \rho x_{t-1} + \eta_t, \end{aligned} \tag{B2}$$

where ξ_t is independent of η_t at all leads and lags. The first four unconditional moments of y_t and the first two time series moments are given by

$$\begin{aligned} \mathrm{E}[y_t] &= m_1 \\ \mathrm{E}[y_t^2] &= \sigma_x^2 + m_2 \\ \mathrm{E}[y_t^3] &= 3m_1\sigma_x^2 + m_3 \\ \mathrm{E}[y_t^4] &= 3\sigma_x^4 + 6m_2\sigma_x^2 + m_4 \\ \mathrm{E}[y_t y_{t-1}] &= m_1^2 + \rho\sigma_x^2 \\ \mathrm{E}[y_t y_{t-2}] &= m_1^2 + \rho^2\sigma_x^2, \end{aligned} \tag{B3}$$

where $\sigma_x^2 = \sigma^2/(1-\rho^2)$. In case the measurement equation contains the constant term β one should replace all μ_i in (B1) by $\mu_i + \beta$. The central moments can be computed from these moments in the usual way.

To obtain the moments of $\epsilon_t = \pm e^{\xi_t/2}$ we assume that the distribution of ϵ_t is symmetric so that all odd moments are zero. The even moments are found analogous to (B1) by using the properties of the individual elements in the mixture:

$$\begin{aligned} \mathrm{E}[\epsilon_t^2] &= \mathrm{E}[e^{\xi_t}] = \sum_{i=1}^{K} p_i \exp(\mu_i + \omega_i^2/2) \\ \mathrm{E}[\epsilon_t^4] &= \mathrm{E}[e^{2\xi_t}] = \sum_{i=1}^{K} p_i \exp(2\mu_i + 2\omega_i^2) \end{aligned} \tag{B4}$$

References

ANDERSEN, T.G. (1992), Volatility, working paper 144, Northwestern University.

ANDERSEN, T.G. (1994a), Return Volatility and Trading Volume: an Information Flow Interpretation of Stochastic Volatility, working paper, Northwestern University.

ANDERSEN, T.G. (1994b), Stochastic Autoregressive Volatility: a Framework for Volatility Modeling, *Mathematical Finance*, **4**, 75–102.

BAILLIE, R.T. AND T. BOLLERSLEV (1989), The Message in Daily Exchange Rates: A Conditional Variance Tale, *Journal of Economic and Business Statistics*, **7**, 297–305.

BOLLERSLEV, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, **31**, 307–328.

CLARK, P.K. (1973), The Subordinated Stochastic Process Model with Finite Variance for Speculative Prices, *Econometrica*, **41**, 135–155.

DANIELSSON, J. (1994), Stochastic Volatity in Asset Prices: Estimation with Simulated Maximum Likelihood, *Journal of Econometrics*, **61**, forthcoming.

DANIELSSON, J. AND J.-F. RICHARD (1993), Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models, *Journal of Applied Econometrics*, **8**, 153–173.

ENGLE, R.F. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, **50**, 987–1007.

ENGLE, R.F. AND T. BOLLERSLEV (1986), Modeling the Persistence of Conditional Variances, *Econometric Reviews*, **5**, 1–50.

GALLANT, A.R., D. HSIEH, AND G.E. TAUCHEN (1994), Estimation of Stochastic Volatility Models with Suggestive Diagnostics, working paper, Duke University.

HARVEY, A.C., E. RUIZ, AND N. SHEPHARD (1994), Multivariate Stochastic Variance Models, *Review of Economic Studies*, **61**, 247–264.

HARVEY, A.C. AND N. SHEPHARD (1993a), The Econometrics of Stochastic Volatility, working paper, Nuffield College, Oxford.

HARVEY, A.C. AND N. SHEPHARD (1993b), Estimation and Testing of Stochastic Variance Models, working paper, Nuffield College, Oxford.

JACQUIER, E., N.G. POLSON, AND P.E. ROSSI (1994), Bayesian Analysis of Stochastic Volatility Models, *Journal of Economics and Business Statistics*, forthcoming.

KIM, S. (1993), Persistence in Conditional Variance: Estimation using a General Stochastic Volatility Framework, working paper, Princeton University.

KIM, S. AND N. SHEPHARD (1994), Stochastic Volatility: Optimal Likelihood Inference and Comparison with ARCH Models, working paper, Nuffield College, Oxford.

MAHIEU, R.J. AND P.C. SCHOTMAN (1994), Neglected Common Factors in Exchange Rate Volatility, *Journal of Empirical Finance*, **1**, 279–311.

MELINO, A. AND S.M. TURNBULL (1990), Pricing Foreign Currency Options with Stochastic Volatility, *Journal of Econometrics*, **45**, 239–265.

NELSON, D.B. (1991), Conditional Heteroskedasticity in Asset Pricing: a New Approach, *Econometrica*, **59**, 347–370.

RUIZ, E. (1993), Stochastic Volatility versus Autoregressive Conditional Heteroskedasticity, working paper 93-44, Universidad Carlos III de Madrid.

SCHOTMAN, P.C. (1994), Priors for the AR(1) Model: Parameterization Issues and Time Series Considerations, *Econometric Theory*, **10**, 579–595.

SHEPHARD, N. (1993a), Fitting Nonlinear Time-Series Models with Applications to Stochastic Variance Models, *Journal of Applied Econometrics*, **8**, 135–152, special issue.

SHEPHARD, N. (1993b), Maximum Likelihood Estimation of Regression Models with Stochastic Trend Components, *Journal of the American Statistical Association*, **88**, 590–595.

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

TANNER, M.A. (1991), *Tools for Statistical Inference: Observed Data and Data Augmentation Methods*, Lecture Notes in Statistics 67, Springer-Verlag.

TAUCHEN, G.E. AND M. PITTS (1983), The Price Variability-Volume Relationship on Speculative Markets, *Econometrica*, **51**, 485–505.

TAYLOR, S.J. (1986), *Modelling Financial Time Series*, Wiley.

TAYLOR, S.J. (1994), Modeling Stochastic Volatility: a Review and Comparative Study, *Mathematical Finance*, **4**, 183–204.

ZELLNER, A. AND C. MIN (1992), Gibbs Sampler Convergence Criteria (gsc^2), working paper, University of Chicago.