AN EMPIRICAL APPLICATION OF STOCHASTIC VOLATILITY MODELS

RONALD J. MAHIEU[a] AND PETER C. SCHOTMAN[b]*

[a]*Erasmus University Rotterdam*
[b]*Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands. Email: p.schotman@berfin.unimaas.nl*

SUMMARY

This paper studies the empirical performance of stochastic volatility models for twenty years of weekly exchange rate data for four major currencies. We concentrate on the effects of the distribution of the exchange rate innovations for both parameter estimates and for estimates of the latent volatility series. The density of the log of squared exchange rate innovations is modelled as a flexible mixture of normals. We use three different estimation techniques: quasi-maximum likelihood, simulated EM, and a Bayesian procedure. The estimated models are applied for pricing currency options. The major findings of the paper are that: (1) explicitly incorporating fat-tailed innovations increases the estimates of the persistence of volatility dynamics; (2) the estimation error of the volatility time series is very large; (3) this in turn causes standard errors on calculated option prices to be so large that these prices are rarely significantly different from a model with constant volatility. © 1998 John Wiley & Sons, Ltd.

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 Heteroscedasticity (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 modelled as an unobserved component. It has been proposed because it is 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), Tauchen and Pitts (1983), and Taylor (1986). It has been advocated in econometrics by Harvey, Ruiz, and Shephard (1994), and Jacquier, Polson, and Rossi (1994) among others.[1]

From an econometric viewpoint a practical drawback of stochastic volatility models is the intractability of the likelihood function. Because (1) the variance is an unobserved component and (2) the model is non-Gaussian, 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; Andersen and Sørensen, 1996). Exact likelihood-oriented methods require simulations and are thus computer intensive (see Danielsson, 1994; Jacquier, Polson, and Rossi, 1994).

---

* Correspondence to: Peter Schotman, LIFE, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands

This paper initially circulated under the title 'Stochastic volatility and the distribution of exchange rate news'.

[1] The relations between the ARCH and stochastic volatility approaches have been analysed by Andersen (1992, 1994) and Kim, Shephard, and Chib (1998). Ghysels, Harvey, and Renault (1996) provide a useful review of the literature.

The first purpose of the paper is to compare different estimation techniques applied to an empirical dataset 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 into an awkward logarithmic 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, Shephard, and Chib (1998).

The computational aspects are not our primary focus. We are more concerned with the implications of the normality assumption for the innovations. 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 behaviour (see Baillie and Bollerslev, 1989; Engle and Bollerslev, 1986). Part of the fat-tailedness is attributable to time aggregation. Our observations are weekly, whereas the stochastic volatility model is often formulated in continuous time. But even in a continuous time model Gallant, Hsieh, and Tauchen (1994) still find evidence for non-normal exchange rate diffusions. Kim, Shephard, and Chib (1998) find that in daily sterling/dollar series the non-normality is not severe and primarily due to a few outliers. Deviations from normality have not been as widely explored as in the ARCH literature, though. The second purpose of this paper is to employ a flexible mixture specification for the innovation that is computationally 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 suited for 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. Focusing 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 misspecification. 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.

Our main motivation stems from potential applications to option pricing. Hull and White (1987) derive an option valuation formula if the underlying asset follows a stochastic volatility process. The crucial input to their option formula is the average volatility over the lifetime of the option. For short-term options the expected volatility will be approximately the current volatility, while for longer term (at-the-money) options the average expected future volatility depends crucially on the persistence and variance of the volatility process. The higher the persistence of the volatility process, the more the long-term volatility depends on current volatility. For our exchange rate data we find that estimates of the persistence vary systematically with the estimation method. It turns out that the differences are due to the auxiliary assumption about normality of exchange innovations conditional on the volatility. We explain how ignoring the occurrence of outliers, and thus *not* modelling the deviations from normality, leads to biased low estimates of the persistence in the volatility process.

In the empirical work we will compare different estimators of volatility. The comparison allows a decomposition of the estimation error in a part that is due to the smoothing algorithm and a

part related to the parameter values obtained from a particular estimation method. We will also compare the volatilities conditional on the parameters of the process with a full Bayesian estimator of volatility, thus incorporating parameter uncertainty. A similar analysis for implied volatilities, obtained from pricing at-the-money options, shows how well we can estimate average volatility given only a time series of observations on the returns of the underlying asset.

The remainder of the paper is organized as follows. The model specification and the estimators are discussed in Sections 2 and 3. Section 4 describes the exchange rate data. Section 5 contains the parameter estimation results. Section 6 reports the results for the estimation of the latent volatility series. In Section 7 we concentrate on the option application. Section 8 presents conclusions.

2. STOCHASTIC VOLATILITY MODEL

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

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

$$h_{t+1} = \rho h_t + \eta_t \tag{2}$$

where $\psi^2 \exp(h_t) \equiv \zeta_t^2$ is the variance of s_t for period 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, and mutually uncorrelated.[2] The exchange rate obtains its unconditional fat-tailed distribution by the mixing of ε_t and $\exp(h_t/2)$. The scale parameter ψ determines the unconditional variance of s_t.

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:

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

where Y_T contains all the data for a sample of T observations, $H_T \equiv (h_1, \ldots, h_T)'$ is the vector with all the latent volatilities, and $\theta = (\rho, \sigma, \psi)'$ contains the parameters of the SV model. In this equation the second density in the integral can be considered as a prior over H_T, specified by the transition equation (2). Since the integral in equation (3) cannot be solved analytically, maximum likelihood methods will be computationally intensive. That is one reason why alternative estimators have been developed.

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

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

[2] Correlation between ε_t and η_t introduces the leverage effect (see Nelson, 1991), which is important in applications to stock returns, but less so for applications to exchange rates.

where $\beta \equiv \ln \psi^2$ and $\xi_t \equiv \ln \varepsilon_t^2$. In the QML approach the density of ξ_t is approximated by a normal density with mean $\mu = -1 \cdot 2704$ and variance $\omega^2 = \pi^2/2$. QML estimates can be calculated by standard numerical optimization techniques, using the Kalman filter.

Studies on the performance of the QML estimator have been concerned with estimating θ. 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. The GMM methodology is an alternative to QML, and can provide a more efficient estimator of the parameter vector θ. If the main interest is in estimating the parameter vector θ, then QML or GMM will be satisfactory.[3] GMM, however, does not solve the problem of estimating the latent time series h_t or the volatility $\exp(h_t)$. Additionally, the standard Kalman smoother might produce poor estimates for the state variable h_t conditional on the parameters of the process. Unlike the problem of estimating the parameters θ, estimating each individual element h_t remains a small sample problem. An estimator $\hat{h}_t$ does not converge to the true unobserved state variable as $T \to \infty$. The Kalman smoother produces the best linear estimator of h_t, but if the ξ_t innovations are non-normal, accurate estimation of the time series of volatilities requires the entire conditional distribution of h_t given the data and the parameters.

Several methods have been developed to deal with the multiple integration problem in equation (3). These methods rely heavily on simulation techniques. Danielsson (1994) and Danielsson and Richard (1993) develop an importance sampling technique to estimate the integral. In a Bayesian setting Jacquier, Polson, and Rossi (1994) combine a Gibbs sampler with the Metropolis–Hastings algorithm to obtain the marginal posterior densities of the parameters in equations (1) and (2), and also the exact posterior distribution of the variance series given the observed data s_t.

In this paper we build on a suggestion by Kim, Shephard, and Chib (1998). These authors retain the convenient linear form of the state space model, and approximate the log-chi-squared 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 approach of Kim, Shephard, and Chib (1998) is very flexible and lends itself easily for extensions to accommodate alternative distributions of ε_t. The mixture is specified as

$$\xi_t = \xi(z_t) \qquad z_t = 1, 2, \ldots, K$$
$$\xi(i) \sim \mathrm{n}(\mu_i, \omega_i^2) \tag{5}$$
$$\Pr(z_t = i) = p_i$$

Using the mixture model the latent volatility series can be estimated through simulation by the multi-move Gibbs sampler. This sampler cycles through steps that simulate the mixture indicators z_t and the states h_t. Conditional on a time series of mixture indicators $Z_T = (z_1, \ldots, z_T)'$ the standard Kalman recursions can be used to simulate the states h_t from the conditional density $f(H_T | Y_T, Z_T; \theta)$. Conversely, given a time series for the states, a posterior odds calculation gives the multinomial distributions $f(z_t | Y_T, H_T; \theta)$, from which new regime indicators can be drawn. Given the parameters the log-volatility h_t is estimated by averaging over the draws $h_t^{(j)}$ ($j = 1, \ldots, M$). The volatility itself is estimated by averaging over $\exp(h_t^{(j)})$. All other functions of the volatility can be estimated similarly.

---

[3] The merits and drawbacks of QML and GMM have been documented by Jacquier, Polson, and Rossi (1994), and Andersen and Sørensen (1996). See also Ghysels, Harvey, and Renault (1996) and the references cited therein.

The mixture model also allows fast optimization of the likelihood function with respect to θ using the simulated EM algorithm (see Tanner, 1996; Kim, Shephard, and Chib, 1998). Conditional on Z_T the likelihood is completely Gaussian, and can be easily computed using the Kalman filter. Maximization of the likelihood proceeds by cycling through an **E**stimation and a **M**aximization step. In the estimation step the relevant part of the log-likelihood is estimated by simulation conditional on a set of draws $Z_T^{(j)}$, $(j = 1, \ldots, M)$,

$$\ln L(Y_T \mid \theta) = \frac{1}{M} \sum_{j=1}^{M} \ln f(Y_T \mid Z_T^{(j)}; \theta) \qquad (6)$$

Keeping the draws $Z_T^{(j)}$ fixed the likelihood function is maximized with respect to θ. When a new parameter vector θ^* has been found, the Gibbs sampler is run to obtain a new set of $Z_T^{(j)}$ using the new parameter vector θ^*. The EM steps are repeated until convergence.

The classical estimators of the volatility are conditional on the parameters. To investigate the effect of parameter uncertainty on the volatility and option value estimates, we compare the classical methods with a fully Bayesian procedure. The Bayesian Gibbs sampling algorithm cycles through simulators for the states, the mixture indicators, and the parameter vector θ. In the Bayesian procedure we assume very diffuse proper conditionally conjugate priors. Simulation from the conditional density of θ given the states and the data works well, since the model is linear. More details on the estimation procedures can be found in the Appendix.

3. FLEXIBLE MIXTURE MODEL

The skewed log-chi-squared distribution for ξ_t comes from the normality assumption for ε_t. From a statistical viewpoint there is, however, no reason to insist on the normality of ε_t. In fact, Gallant, Hsieh, and Tauchen (1994) find that the innovations for the Dmark/dollar exchange rate are still fat-tailed in a stochastic volatility model.[4] Therefore we extend the mixture model to accommodate a wide range of shapes of the distribution of ξ_t and hence the distribution of ε_t.

The generalization allows ξ_t to be generated by a 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 $|\varepsilon_t| = \sqrt{\exp(\xi_t)}$.[5] In general the tails of ε_t become fatter the larger the variance of ξ_t, *ceteris paribus*. Also, the fatter the tails of ε_t, the more symmetric the implied density of ξ_t. For example, if ε_t has a Cauchy distribution, the density of ξ_t is symmetric (Mahieu and Schotman, 1994). Ghysels, Harvey, and Renault (1996) give a detailed characterization of the relation between the moments of ε_t and ξ_t.

Since each element in the mixture adds three more nuisance parameters, we have to keep the number of elements small to avoid overparameterization and convergence problems. After some experimentation we settled for a mixture of three normals with fixed weights $p = (0 \cdot 70, \ 0 \cdot 25, \ 0 \cdot 05)$. This choice is motivated by some of the main features of the data. The negative skewness of

---

[4] Evidence on fat tails in stochastic volatility models is, however, less clear than for GARCH models. Kim, Shephard, and Chib (1998) find that normality of ε_t is a good approximation for some daily exchange rate return series. Partly this is due to their 'offset' parameter c in taking the logtransform $y_t = \ln(s_t^2 + c)$. The parameter $c > 0$ reduces the impact of very small values of s_t, which would cause huge negative outliers otherwise. The offset parameter effectively shifts mass away from small values $|s_t|$ to larger values.

[5] In practice, due to the exponential form, the transformation is very sensitive to exact specification of especially the right tail of the density of ξ_t.

Table I. Mixture parameters for log-chi-squared distribution

i	p_i	μ_i	ω_i
1	0·70	−0·2172	1·1052
2	0·25	−3·0461	1·5705
3	0·05	−6·4818	3·0002

Notes: The mixture density is $\Sigma_{i=1}^3 p_i \mathrm{n}(\mu_i, \omega_i^2)$. Parameters have been estimated by simulating 100,000 random numbers from a log-chi-squared distribution, and maximizing the likelihood function of the three-element mixture for these data.

ξ_t is caused by the 'inliers', i.e. observations with almost zero change ($s_t \approx 0$) for which the transformation $y_t = \ln s_t^2$ generates large negative outliers. Even as few as two normals could model the skewness of ξ_t: the first normal centred around zero, covering the bulk of the data, and the other centred around a large negative mean to accommodate the outliers. Its main drawback is that such a distribution is likely to be bimodal. We therefore add a third element to the mixture to provide a smooth blending to a unimodal density.

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 I. In the upper panel of Figure 1 we plot the mixture distribution together with the actual log-chi-squared distribution. The main difference is the slightly higher mode of the mixture distribution. The middle panel of Figure 1 shows the $\chi^2(1)$ distribution together with the exponential mixture distribution. The lower panel shows the logarithmic ratio of chi-square and exponential mixture distributions. It can be seen that our mixture gives a reasonable fit. It looks like a three-element mixture strikes a balance between flexibility and the number of parameters.[6]

Since the unconditional mean of the measurement error ξ_t is a free parameter in the mixture distribution ($\mu = \Sigma_{i=1}^K p_i \mu_i$), the constant term β in the measurement equation (4) becomes unidentified. We therefore identify β and $\psi^2 = \exp(\beta)$ from the unconditional variance of s_t.

There is a second reason for the extension of standard volatility models. Fixing the distribution of the measurement error means that in particular the variance of ξ_t is restricted. In QML, GMM and the mixture of Kim, Shephard, and Chib (1998) the variance of ξ_t is equal to $\omega^2 = \pi^2/2$. With fat-tailed exchange rate innovations this restriction will be rejected by the data. But this will also affect the estimates of the other parameters (ρ and σ^2), even asymptotically, since a larger value for ω^2 changes the implied autocorrelation function of y_t. The state space model is an AR(1) process with measurement error, or equivalently an ARMA(1,1) process, parameterized by ρ, σ^2 and ω^2. The variance and first-order autocorrelation of y_t are

$$\mathrm{Var}(y_t) = \frac{\sigma^2}{1 - \rho^2} + \omega^2 \tag{7}$$

$$\mathrm{Corr}(y_t, y_{t-1}) = \frac{\rho\sigma^2}{\sigma^2 + (1 - \rho^2)\omega^2} \tag{8}$$

---

[6] Our method differs from the approach followed in Escobar and West (1995). These authors also provide estimation procedures for determining the mixture weight parameters p_i. We set the weights *a priori*.
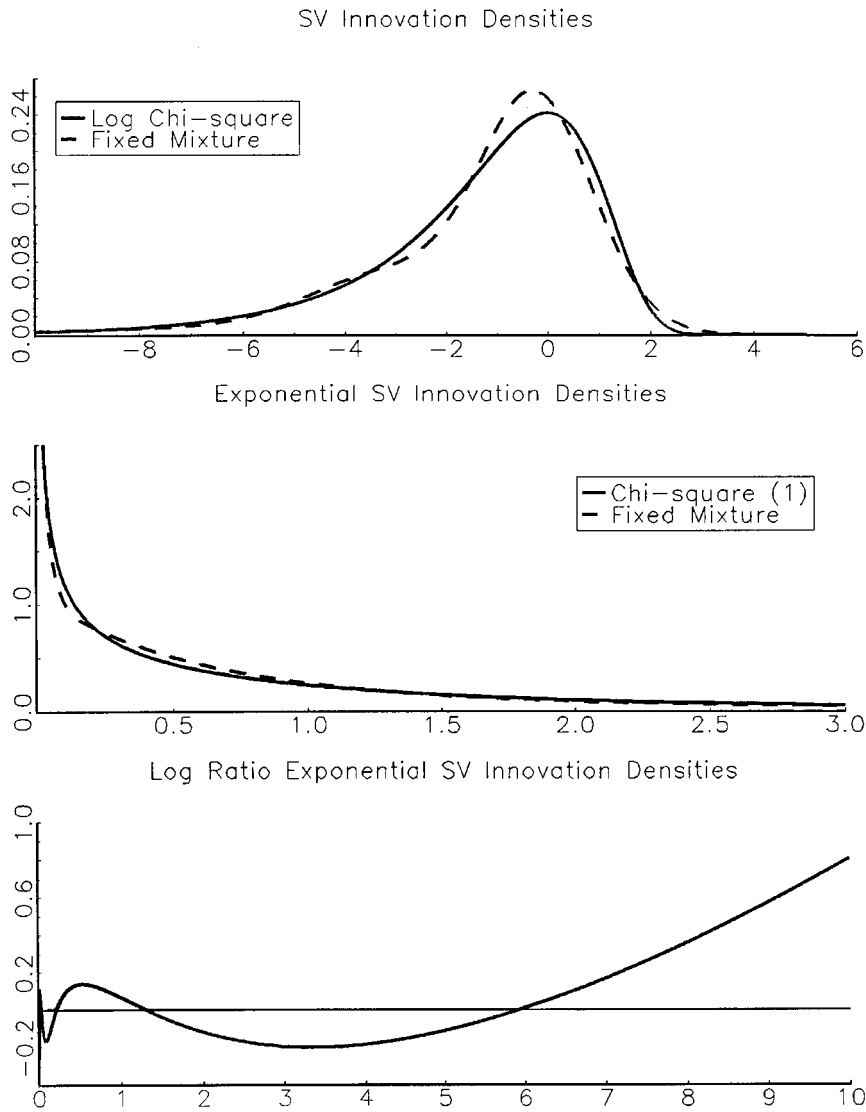
Figure 1. Densities of measurement error. The mixture approximation is based on the values in Table I

When the parameters match the sample variance $\text{Var}(y_t)$ and the first order autocorrelation AR(1), we have

$$\frac{\text{Corr}(y_t, y_{t-1})}{\rho} = 1 - \frac{\omega^2}{\text{Var}(y_t)} \qquad (9)$$

which shows that ρ is an increasing function of ω^2. An increase in ω^2 must lead to a larger value of ρ to keep the sample first-order autocorrelation AR(1) constant. So one effect of the extended

model with an unrestricted ω^2 will most likely be an increase in the estimated persistence ρ. In that way fat-tailedness (of ε_t!) and persistence are connected in the stochastic volatility model. The increase in estimates of ρ are a second moment property, and will already show up in the QML estimates with unrestricted measurement error variance.

Estimation of the flexible mixture model proceeds analogous to the estimation algorithms for the fixed mixture. The parameter vector θ is augmented to the eight-dimensional vector $(\rho, \sigma, \mu', \omega')$, where μ is the (3×1) vector of means of the mixture, and ω the (3×1) vector of standard deviations. Using the EM algorithm the optimal value of θ is found by iteratively maximizing equation (6).

In the Bayesian algorithm we assume independent normal/inverse gamma conjugate priors for each element of μ and ω. The prior means of μ_i and ω_i are set at the values of the approximating mixture in Table I. The prior variances of the μ_i's 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, since with an improper prior on ω_i the posterior does not exist.

4. DATA

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

The weekly sampling frequency is chosen as a compromise between using as many data as possible and avoiding additional modelling issues. Daily data would in general be preferable for studies that focus on second moments. However, except for the two most recent years, the data from DATASTREAM are only reported in discrete tick sizes. As a result, daily data contain many days with 'no change', which is a discrete phenomenon that would have to be modelled separately. Volatility over the interval of a week is much larger, so that 'no change' observations are less frequent in weekly data. For example, for the yen/pound exchange rate the tick size is 0·25 in DATASTREAM. With daily data 14% of changes are zero, and for another 10% and 12% of the data we see changes of 0·25 and 0·50 respectively. For weekly data the percentages are much smaller: 3%, 3% and 7% respectively. It also appears that the distribution of daily returns is much more fat-tailed than the distribution of weekly returns. A second problem with the use of daily data (or even higher frequencies) is the treatment of weekends and other seasonal day-of-the-week effects.

Table II 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-chi-squared (which is equal to $\pi^2/2$), so that either the logarithmic volatility series h_t is itself highly volatile or normality of the measurement error in equation (4) is violated. The first-order autocorrelation is generally low, but higher-order autocorrelations are always of similar magnitude as the first-order autocorrelation. That would roughly identify all series as ARMA(1,1) or equivalently AR(1) with measurement error, consistent with the specification of the first-order stochastic volatility model.

Table II. Summary statistics

	JP/US	GE/US	UK/US	GE/JP	UK/JP	GE/UK
Statistics of $s_t = \Delta \ln S_t$						
Std dev.	10·12	10·96	10·78	9·78	10·84	8·34
Statistics of $y_t = \ln[(\Delta \ln S_t - \bar{s})^2]$						
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
Autocorrelations of 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 standard deviation of s_t is reported in units of per cent per annum. JP is yen, US is dollar, GE is mark and UK is sterling. Skewness and kurtosis of y_t are scaled by the standard deviation of y_t. The kurtosis is measured as excess kurtosis.

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

$$\text{Cov}\left(s_t^{\text{JP/UK}}, s_t^{\text{GE/UK}}\right) = \frac{1}{2}\left\{\text{Var}\left(s_t^{\text{JP/UK}}\right) + \text{Var}\left(s_t^{\text{GE/UK}}\right) - \text{Var}\left(s_t^{\text{JP/GE}}\right)\right\} \qquad (10)$$

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

5. PARAMETER ESTIMATES

In the empirical analysis we compare four different sets of classical parameter estimates. The first is the QML estimator with the measurement error fixed at $\omega^2 = \pi^2/2$ as in Harvey, Ruiz, and Shephard (1994) (QML1). The second is the QML estimator with unrestricted ω^2 as in Harvey and Shephard (1993) (QML2). The third is the simulated maximum likelihood estimator for the fixed three-element mixture with the parameters from Table I (SIEM1). The last one is the ML estimator for the flexible mixture with free parameters in the three elements of the mixture (SIEM2).

Table III summarises the parameter estimates for the four classical methods and the posterior means of the Bayesian algorithm. We concentrate the discussion on the parameters ρ and σ of the volatility process. In general, the parameter estimates are similar, both across currencies as well as estimators. The high value of ρ implies persistent logarithmic volatility series. The estimates of ρ are never significantly different from unity using Dickey–Fuller critical values. The estimates of ρ

Table III. Parameter estimates $h_{t+1} = \rho h_t + \eta_t \quad \eta_t \sim \mathrm{N}(0, \sigma^2) \quad y_t = \ln s_t^2 = h_t + \xi_t \quad f(\xi_t) = \Sigma_{i=1}^3 p_i \mathrm{N}(\mu_i, \omega_i^2)$

	JP/US	GE/US	UK/US	GE/JP	UK/JP	UK/GE
(a) Persistence: ρ						
QML1	0·976	0·967	0·960	0·985	0·653	0·884
(restricted)	(0·015)	(0·027)	(0·015)	(0·009)	(0·391)	(0·061)
QML2	0·983	0·978	0·963	0·985	0·981	0·895
(unrestricted)	(0·010)	(0·016)	(0·014)	(0·010)	(0·015)	(0·050)
SIEM1	0·878	0·928	0·952	0·921	0·584	0·768
(fixed mixture)	(0·016)	(0·013)	(0·009)	(0·014)	(0·042)	(0·027)
SIEM2	0·979	0·975	0·967	0·954	0·957	0·930
(flexible mixture)	(0·007)	(0·007)	(0·008)	(0·011)	(0·011)	(0·014)
Bayesian	0·971	0·938	0·963	0·912	0·924	0·862
(flexible mixture)	(0·013)	(0·022)	(0·014)	(0·041)	(0·033)	(0·050)
(b) Volatility standard deviation: σ						
QML1	0·225	0·198	0·311	0·092	0·842	0·419
(restricted)	(0·081)	(0·097)	(0·055)	(0·027)	(0·612)	(0·142)
QML2	0·182	0·150	0·296	0·090	0·118	0·389
(unrestricted)	(0·051)	(0·062)	(0·055)	(0·031)	(0·057)	(0·125)
SIEM1	0·558	0·293	0·330	0·285	0·822	0·650
(fixed mixture)	(0·024)	(0·018)	(0·017)	(0·018)	(0·036)	(0·028)
SIEM2	0·190	0·148	0·251	0·200	0·190	0·285
(flexible mixture)	(0·013)	(0·012)	(0·013)	(0·015)	(0·016)	(0·018)
Bayesian	0·274	0·267	0·285	0·300	0·270	0·454
(flexible mixture)	(0·046)	(0·047)	(0·044)	(0·082)	(0·059)	(0·101)
(c) Measurement error variance: ω^2						
QML2	5·540	5·426	5·163	5·027	6·074	5·046
SIEM2	5·524	5·778	5·042	4·830	6·276	5·289
Bayesian	5·511	5·349	5·087	4·767	5·987	4·937

Notes: For QML1 and SIEM1 the measurement error variance is equal to $\pi^2/2 = 4\cdot935$ as opposed to QML2, where this variance is estimated as the free parameter ω^2. The parameters for SIEM1 refer to a fixed three-dimensional mixture for the measurement error with variance equal to $\pi^2/2$. The results for SIEM2 refer to a flexible three-dimensional mixture with variance $\omega^2 = \Sigma_{i=1}^3 p_i(\omega_i^2 + (\mu_i - \bar{\mu})^2)$, where $\bar{\mu} = \Sigma_{i=1}^3 p_i\mu_i$. Standard errors are in parentheses.

are uniformly higher when the measurement error variance is estimated as a free parameter (compare QML1 and QML2), or when the mixture density is left unrestricted (compare SIEM1 and SIEM2). These results are consistent with the analysis of the measurement error restriction $\omega^2 = \pi^2/2$ discussed in Section 3 (see equation (9)). The clearest example is the pound/yen series, where ρ increases from 0·58 under SIEM1 to 0·96 under SIEM2. The pound/yen series is also the one for which the measurement error variance deviates most from $\pi^2/2$. Together with the increase in ρ we see a drop in the volatility parameter σ.

The QML estimator can be interpreted as the maximum likelihood estimator for a model with ξ_t generated by a mixture with the same parameters in each element of the mixture. It is therefore nested within the more general SIEM2 model. Since both SIEM1 and the two QML models are nested within SIEM2, these specifications can be tested by a likelihood ratio test. Table IV reports likelihood ratio tests. The results clearly show that the flexible mixture SIEM2 fits the data much better than any of the other models. The improvement is mainly due to the increased flexibility in the shape of the measurement error distribution. Of course, the mixture fits better than a normal (with or without a restriction on the variance), but the flexible SIEM2 mixture also obtains a

Table IV. Likelihood ratio tests

	df	JP/US	GE/US	UK/US	GE/JP	UK/JP	GE/UK
QML1 (restricted)	5	150·6	259·3	113·9	226·9	242·0	80·0
QML2 (unrestricted)	4	144·1	255·0	112·9	226·8	225·4	79·8
SIEM1 (fixed mixture)	5	179·9	113·2	63·3	156·7	240·7	189·8

Notes: The entries report the likelihood ratio statistic comparing each model with the flexible mixture SIEM2 model. The number of restrictions is given by df. The 5% and 1% critical values of a $\chi^2(5)$ distribution are 11·1 and 15·1 respectively.

much higher likelihood value than the fixed SIEM1 mixture. Surprisingly, the variance restriction that distinguishes QML2 from QML1 can only be rejected for three exchange rates (JP/US, GE/US and UK/JP). Details on the computation of the likelihood values are presented in the Appendix.

Figure 2 shows the estimated measurement error densities. The main difference between a log-chi-square and the flexible mixture is in the shift of the right tail of the density of ξ. Because positive values of ξ get more weight, the implied density of $\varepsilon = \exp(\xi)^{1/2}$ will have extremely fat tails.[7]

The Bayesian posterior means of ρ are somewhat lower than the ML estimates in the flexible SIEM2.[8] Otherwise the Bayesian posterior means are not very different from the SIEM2 maximum likelihood estimates. We will use the Bayesian version of the model to investigate the effects of parameter uncertainty on volatility estimates. This is the topic to which we turn in the next sections.

6. VOLATILITY ESTIMATES

The main reason for estimating stochastic volatility models is that we are interested in the time series of volatilities. The basic issue in this section is the accuracy of the volatility estimates themselves.

Volatilities can be estimated in three different ways: (1) by a Kalman smoother conditional on the parameters assuming that the measurement error ξ_t is Gaussian; (2) by a simulation smoother conditional on the parameters using the mixture distribution of the measurement error density $f(\xi_t)$; (3) by a fully Bayesian simulation smoother using the posterior density of the parameters. The Bayesian smoother averages both over the mixture indicators $Z_T^{(j)}$ as well as parameter draws $\theta^{(j)}$ from the conditional posterior $f(\theta \mid Z_T^{(j)}, H_T^{(j)})$. In this section we compare the various methods for our exchange rate data.

In Section 5 we considered various sets of parameter estimates and model specifications. Conditional on each set of parameter estimates we could compute the volatilities either by method (1) or (2), creating a two-way classification. In order to limit the number of tables and graphs, we will concentrate primarily on the SIEM2 parameter estimates.

---

[7] The implied densities of $\varepsilon = \pm \exp(\xi)^{1/2}$ do not look very sensible. Because the Jacobian of the transformation is zero at $\varepsilon = 0$, the implied densities of ε go to zero for every mixture by construction. The kurtosis of ε implied by the estimated mixtures is of the order 10^{20} for some exchange rates. For this reason we focus entirely on the volatility dynamics and not on the distribution of exchange rate changes.

[8] 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, Shephard, and Chib (1998).
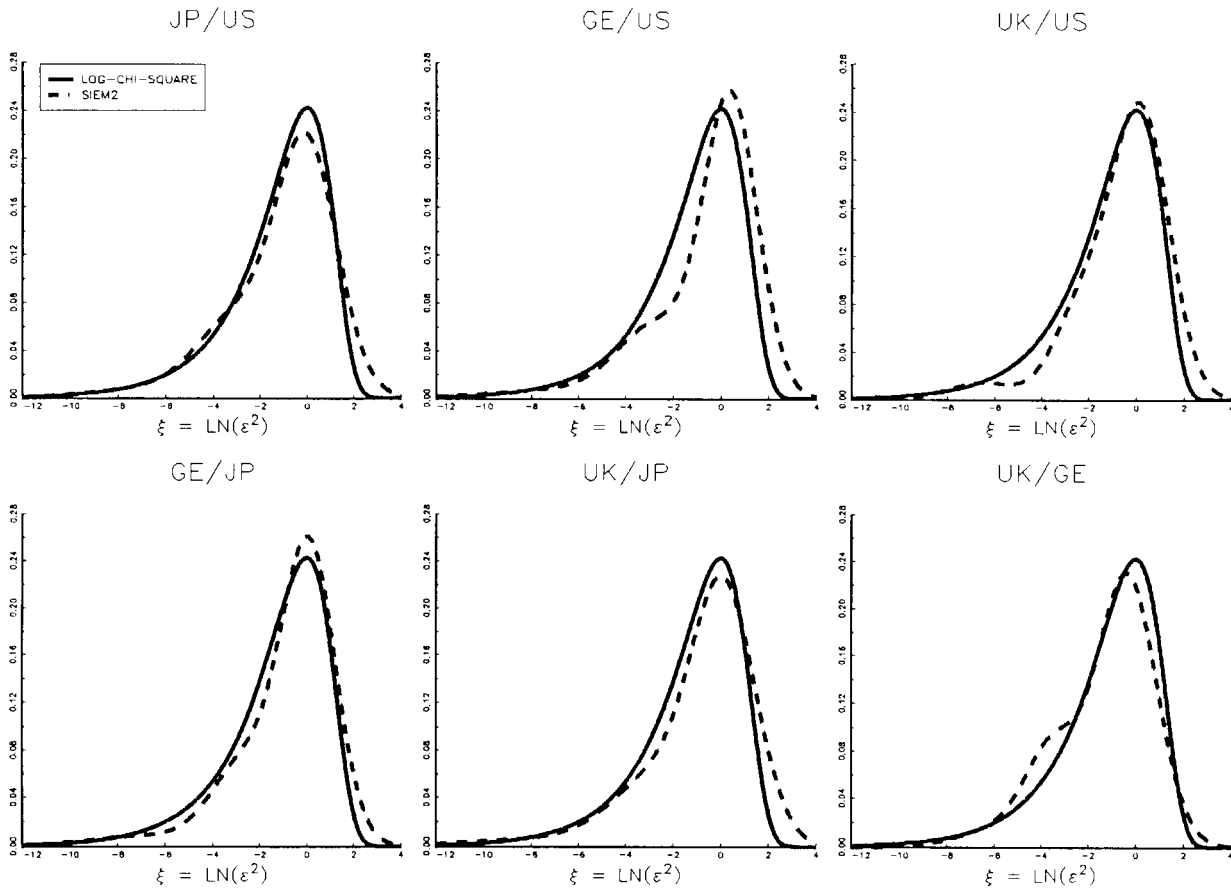
Figure 2. Implied measurement error densities

6.1. Log-volatility

Examining the precision of the estimates of the log-volatilities h_t is based on the decomposition

$$h_t = \hat{h}_{t|T} + (h_t - \hat{h}_{t|T}) \tag{11}$$

where $\hat{h}_{t|T}$ is a shorthand for the smoothed estimate $\hat{E}[h_t | Y_T]$. The smoothing algorithm that we used is described in De Jong and Shephard (1995). The variance of the left-hand side of equation (11) can be estimated from the parameters of the process,

$$\text{Var}(h_t) = \frac{\sigma^2}{1 - \rho^2} \equiv V_h^2 \tag{12}$$

This is the overall variance of the log-volatility. The variance of the first term on the right-hand side of equation (11) is estimated from the sample variance of the time series of smoothed log-volatilities,

$$\text{Var}(\hat{h}_{t|T}) = \frac{1}{T} \sum_{t=1}^{T} \hat{h}_{t|T}^2 \tag{13}$$

Finally, the variance of the estimation errors $(h_t - \hat{h}_{t|T})$ is obtained from the simulation smoother. First, note that for the Kalman smoother the standard error of the volatility estimator is directly available in the recursion from $\text{Var}(h_t - \hat{h}_{t|T}) = P_{t|T}$. For the simulation smoothers the standard errors are obtained by averaging over the results of the Kalman smoother conditional on a set of mixture indicators $Z_T^{(j)}$ at the jth iteration of the Gibbs sampler. Conditional on $Z_T^{(j)}$ the model is Gaussian, and the Kalman smoother produces the conditional means $h_{t|T}^{(j)}$ and variances $P_{t|T}^{(j)}$. The variance of the volatility estimator at time t is computed as

$$P_t = \frac{1}{M} \sum_{j=1}^{M} P_{t|T}^{(j)} + \frac{1}{M} \sum_{j=1}^{M} (\hat{h}_{t|T} - h_{t|T}^{(j)})^2 \tag{14}$$

which uses the fact that the total overall variance equals the average conditional variance plus the variance of the conditional means. For reporting purposes we calculate the average variance over the sample as

$$\bar{P}^2 = \frac{1}{T} \sum_{t=1}^{T} P_t.$$

The variances of all three components in equation (11) have been computed independently, and are reported in Table V. The variances of the two components on the right-hand side in equation (11) therefore do not necessarily add up to the total variance implied by the parameter estimates on the left-hand side. Since the numbers refer to log-volatility the standard errors $\bar{P}$ can be interpreted as the percentage error in estimating the variance at a particular time t. Thus the value of 0·421 for the yen/dollar rate means that there is a 42% error on the variance. These standard errors are large for every exchange rate. For the Bayesian smoother they are even larger due to parameter uncertainty. With such high standard errors we can understand why some

Table V. Summary statistics of log-volatility estimates

	JP/US	GE/US	UK/US	GE/JP	UK/JP	GE/UK
SIEM2 parameters, simulation smoother:						
V_h	0·932	0·666	0·985	0·667	0·655	0·775
$\text{sd}(\hat{h}_{t\mid T})$	0·837	0·576	0·938	0·540	0·508	0·586
$\bar{P}$	0·421	0·343	0·459	0·394	0·408	0·487
Bayes:						
V_h	1·060	0·786	0·864	0·764	0·723	0·899
$\text{sd}(\hat{h}_{t\mid T})$	0·865	0·613	0·910	0·553	0·518	0·660
$\bar{P}$	0·549	0·485	0·578	0·497	0·497	0·626
SIEM2 parameters, Kalman smoother:						
V_h	0·932	0·666	0·985	0·667	0·655	0·775
$\text{sd}(\hat{h}_{t\mid T})$	0·879	0·554	1·028	0·505	0·404	0·714
$\bar{P}$	0·467	0·408	0·524	0·448	0·460	0·543

Notes: $V_h = \sigma/\sqrt{\{1-\rho^2\}}$ is the unconditional standard deviation of h_t implied by the SIEM2 parameter estimates in Table III. The Bayesian estimate of V_h^2 is obtained as the average of $\sigma^{2(j)}/(1-\rho^{2(j)})$. $\text{sd}(\hat{h}_{t\mid T})$ is the sample standard deviation of the time series $\hat{h}_{t\mid T}$ of estimated log-volatilities. For the simulation smoothers $\bar{P}^2$ is the sample mean of the estimated error variances of $\hat{h}_{t\mid T}$. For the Kalman smoother $\bar{P}^2$ is the steady-state variance of the estimation error $(h_t - \hat{h}_{t\mid T})$.

studies find that econometric volatility models do such a poor job in predicting variance (see, for example, West and Cho, 1995; Figlewski, 1994).

The lowest relative estimation error variance $\bar{P}^2/V_h^2$ is obtained for the dollar/pound rate, where the sample variance of $\hat{h}_{t\mid T}$ accounts for 90% of the total variation of the volatility. This is also the series for which we find the highest time-varying volatility. The highest error ratios are for the non-dollar cross-rates, where the sample variance of $\hat{h}_{t\mid T}$ is between 60% and 70% of the total variance.

The standard errors are primarily a function of the parameter estimates. The larger the measurement error variance, the harder it is to estimate the underlying state variable. On the other hand, the larger the persistence ρ, the more information about h_t can be extracted from the data around time t. For comparison Table V also reports the variance estimates from the Kalman smoother. Note that the standard error on the log-volatilities differs substantially from the values that were found for the simulation smoothers. Because the numbers for V_h and $\bar{P}$ are very close to the simulation estimates, we conclude that the large standard errors do not arise from the distributional properties of the measurement error ξ_t, but are indeed due to the low signal/noise ratio. In other words, no matter whether one applies the simulation smoother or the straight-forward Kalman smoother, one would always find that the log-volatilities cannot be estimated with great accuracy (given the empirically relevant parameters for the volatility process). Section 7 investigates what consequences this has for financial applications like option pricing.

6.2. Volatility

In this section we discuss how the results for the log-volatilities translate to actual volatilities $\zeta_t = \exp(h_t/2)$. The results are reported graphically in Figures 3 to 5. Each figure is a scatter plot of the volatility estimates from two methods. The horizontal axis represents the time series of estimates $\hat{\zeta}_t$ ($t = 1, \ldots, T$) from one method, and the vertical axis the corresponding estimates
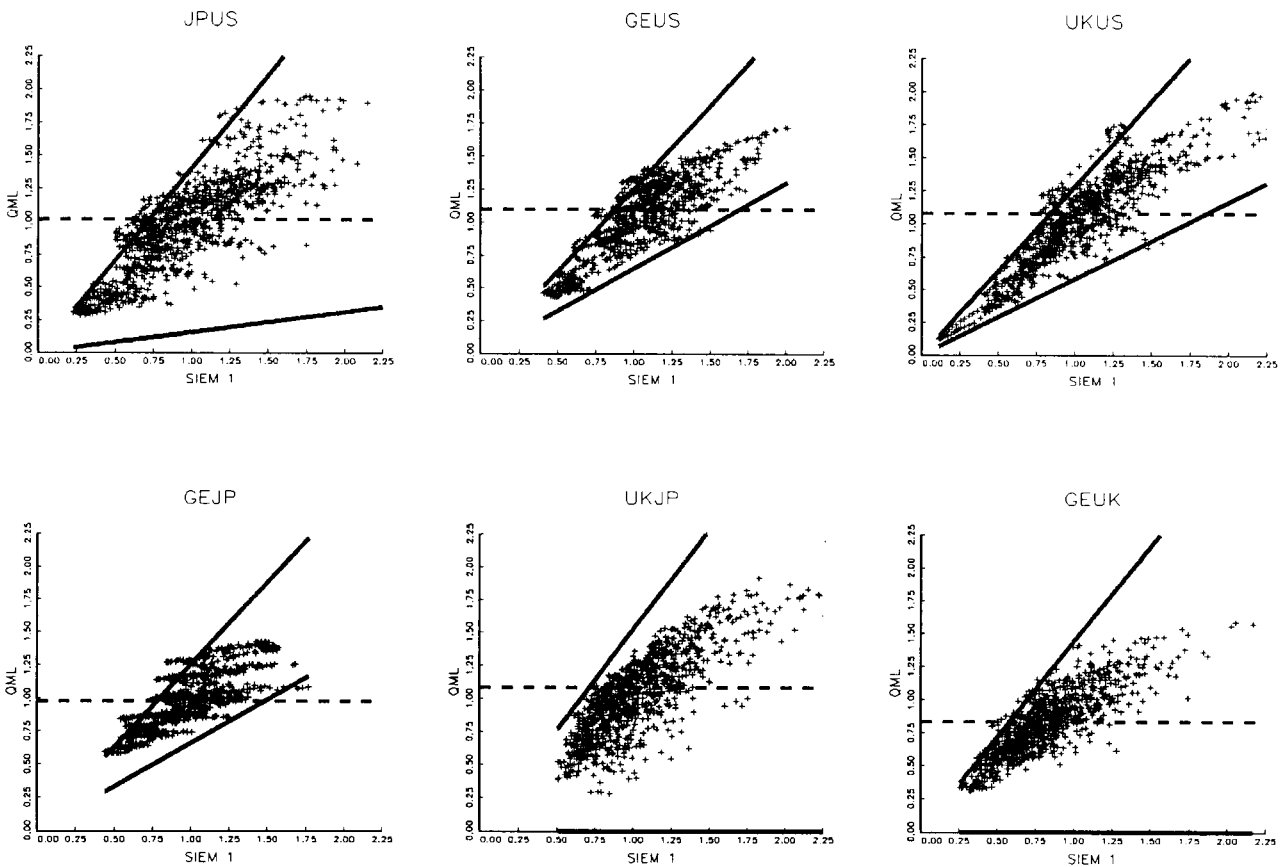
Figure 3. Volatility estimates: Kalman versus simulation smoother. The pluses (+) are the smoothed estimates of the volatility ζ_t from the Kalman smoother (vertical axis) against the simulation smoother (horizontal axis). Both smoothers use the parameter values of the SIEM1 model. The solid lines represent a one standard error margin above and below the estimates. The dashed line is the unconditional variance of $\ln s_t^2$. Both axes have a square root scale
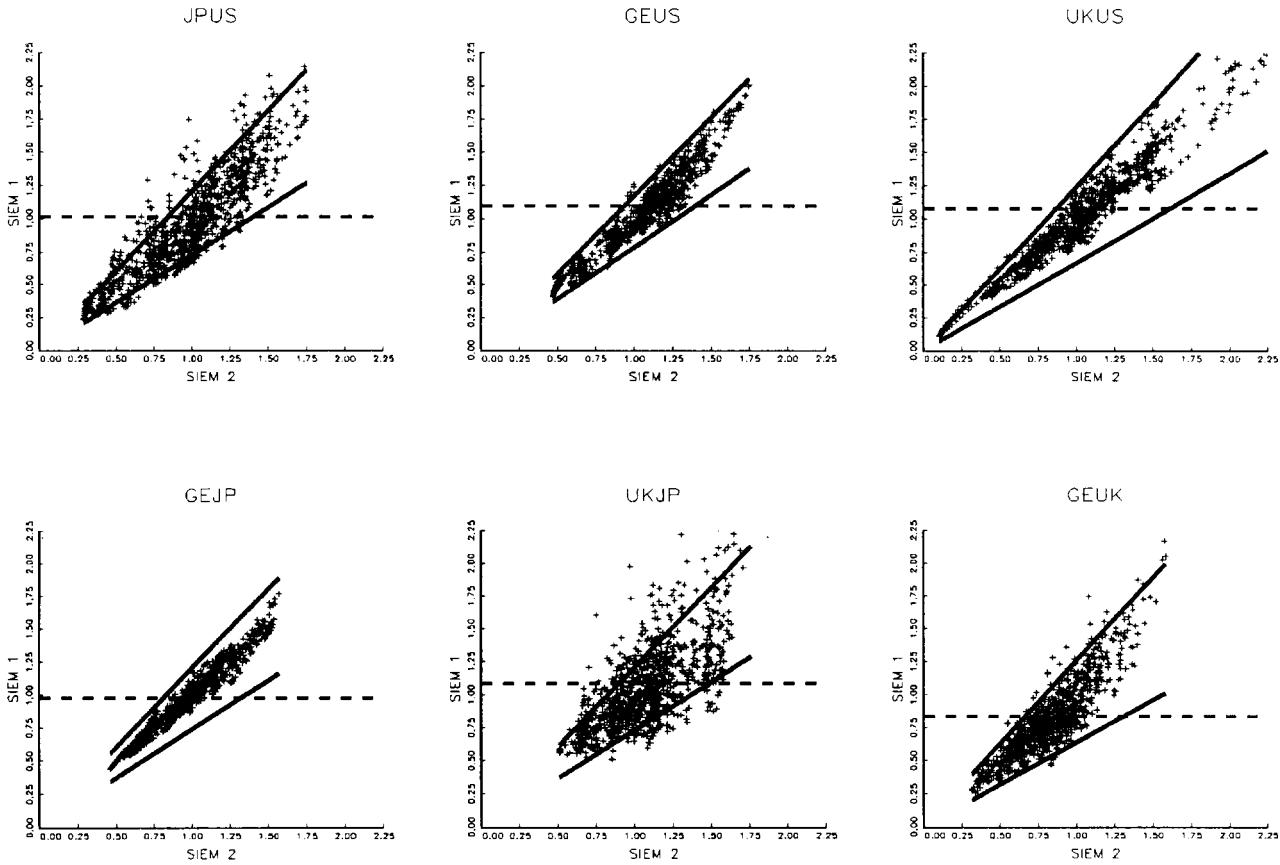
Figure 4. Volatility estimates: SIEM1 versus SIEM2. The pluses (+) are the smoothed estimates of the volatility ζ_t from the SIEM1 (vertical axis) and SIEM2 (horizontal axis) simulation smoothers. The solid lines represent a one standard error margin above and below the estimates using the SIEM2 parameters. The dashed line is the unconditional variance of s_t. Both axes have a square root scale
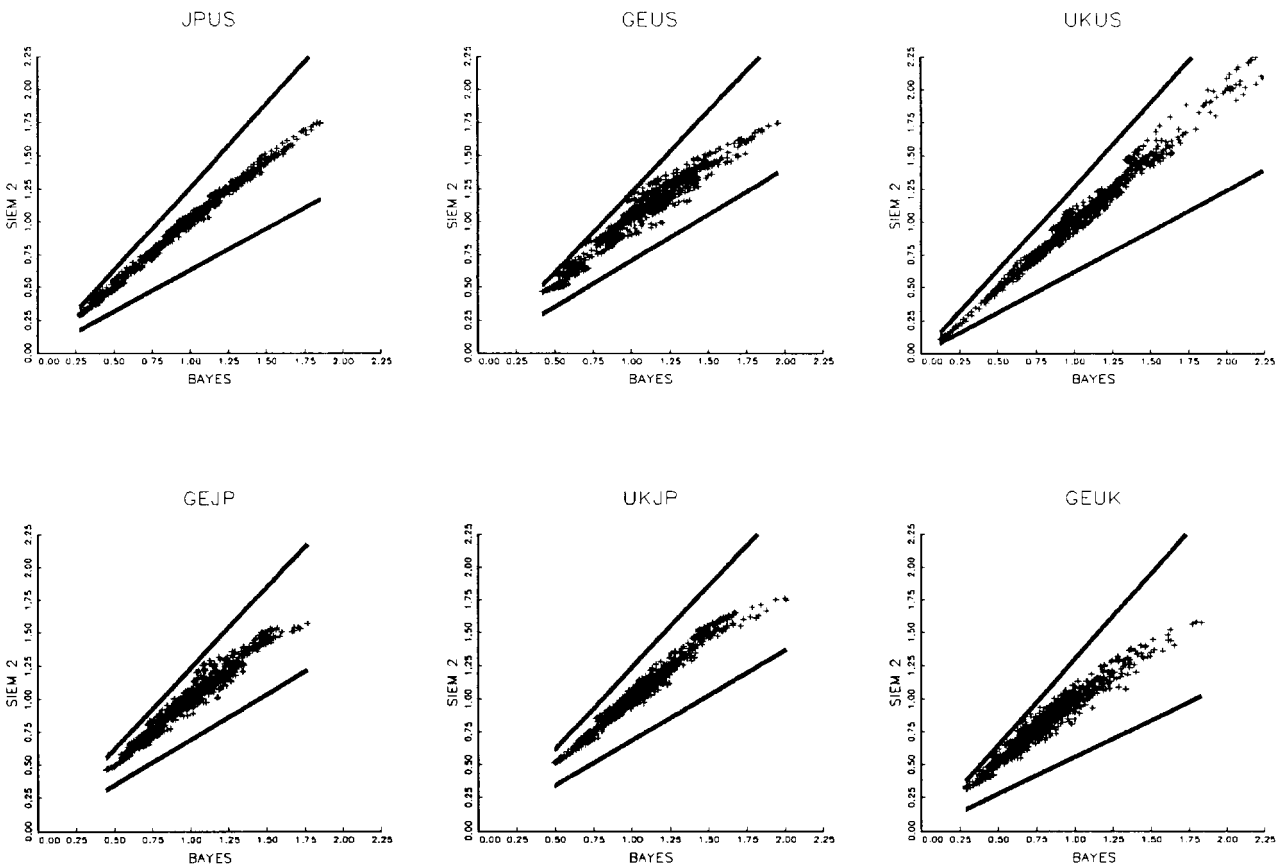
Figure 5. Volatility estimates: SIEM2 versus Bayes. The pluses (+) are the smoothed estimates of the volatility ζ_t from the SIEM2 (vertical axis) and Bayesian (horizontal axis) simulation smoothers. The solid lines represent a one standard error margin above and below the estimates using the Bayesian parameters. The dashed line is the unconditional variance of $\ln s_t$. Both axes have a square root scale

from another. If both methods would produce the same volatility estimates, all points on the scatter would lie on the 45° line through the origin. If volatility is constant all dots in the graph would be cluttered around a single point.

For the Kalman smoother the volatility estimates are calculated using the properties of the lognormal distribution as

$$\hat{\zeta}^2_{t|T} = \tilde{\psi} \exp(\hat{h}_{t|T} + \frac{1}{2} P_{t|T}) \tag{15}$$

The scaling factor $\tilde{\psi}$ is chosen such that $\tilde{\psi}^2 E[e^{h_t}]$ equals the unconditional sample variance of the relevant exchange rate. Such a scaling factor is applied to all smoothing methods, which are thus standardized to produce the same unconditional variance. For the simulation smoothers the volatility is estimated by

$$\hat{\zeta}^2_{t|T} = \frac{1}{M} \sum_{j=1}^{M} \zeta^{2(j)}_{t|T} \tag{16}$$

where $\zeta^{2(j)}_{t|T}$ is the Kalman smoother estimate conditional on the mixture indicators $Z_t^{(j)}$. The volatility estimates on the horizontal axis are plotted against a one-standard error band $(\sqrt{\hat{\zeta}^2_{t|T} - v_t}, \sqrt{\hat{\zeta}^2_{t|T} + v_t})$.[9] The standard errors v_t are computed analogously to equation (14) as

$$v_t^2 = \frac{1}{M} \sum_{j=1}^{M} \zeta^{2(j)}_{t|T} \exp(P_{t|T}^{(j)}) + \frac{1}{M} \sum_{j=1}^{M} (\hat{\zeta}^2_{t|T} - \zeta^{2(j)}_{t|T})^2 \tag{17}$$

The standard error bands are represented as the solid lines in the figures.[10,11]

Figure 3 compares the Kalman smoother with the simulation smoother using the SIEM1 parameter estimates in both methods. The first conclusion from Figure 3 is that there are substantial differences between the Kalman smoother and the simulation smoother. Although the parameters are the same in both methods, the points are widely dispersed around the ideal 45° line, where both smoothers would give the same volatility estimates. So in applications the choice of smoothing technique matters for the point estimates. The second conclusion from the figure is in line with the previous results for the log-volatilities: volatility is not estimated with any precision at all. The one standard error band on both sides of the fixed SIEM volatility estimates is extremely large. It is therefore no surprise that almost all Kalman smoother volatility estimates fit within the bands. Even the unconditional volatility is within the band for about 75% of the observations, i.e. the sample variance of s_t is within the interval $(\hat{\zeta}_{t|T} - v_t, \hat{\zeta}_{t|T} + v_t)$ 75% of the time.

Figure 3 considered different smoothing methods using the same parameters. In Figure 4 we compare the same smoothing technique (simulation) but for different sets of parameters. The first

---

[9] The square root appears because the graphs use a volatility instead of a variance scale.

[10] Convergence of the Gibbs sampler for the standard errors v_t appears to be much slower than for the point estimates $\hat{\zeta}_{t|T}$. For the flexible SIEM estimates we needed about 1000 iterations before the v_t estimates stabilize.

[11] For graphical purposes the lower and upper bounds are shown as solid lines through use of a non-parametric kernel smoother. If $l_t = \sqrt{\hat{\zeta}^2_{t|T} - v_t}$ is the lower part of the standard error band corresponding to $\hat{\zeta}_{t|T}$, then the plotted lower bound line is the function $f(\zeta)$ obtained from $f(\zeta) = \Sigma K(\zeta - \hat{\zeta}_{t|T})l_t/\Sigma K(\zeta - \hat{\zeta}_{t|T})$ with $K(x)$ a kernel. If for the lower bounds the function value is negative, it is plotted as zero (variance cannot be negative).

set of parameters are the SIEM1 estimates, where we assume that the measurement error density is a log-chi-squared, approximated by a fixed mixture of normals. The alternative set of parameters are the SIEM2 estimates, which are based on the flexible mixture. Two things stand out from Figure 4. First, the volatility estimates from the two models are highly correlated, despite the difference in parameter estimates (see Table III above). The points are generally much closer to the 45° line than in the previous figure. Comparing Figures 3 and 4 it appears that the smoothing method is more important than the parameter values.

The second point to notice is that the standard error bands of SIEM2, shown in Figure 4 are much smaller than those of SIEM1, shown in Figure 3. This decrease in standard errors is related to the increased persistence estimates in Table III. The higher the value of ρ, the smoother the volatility series. With a large value for ρ neighbouring observations $t + s$ are more informative about the volatility at time t. Higher persistence also comes together with a lower estimate of σ, and this smaller innovation variance of h_t further reduces the standard errors of the volatility estimates.

Finally, Figure 5 compares SIEM2 with the Bayesian volatility estimates. These estimates are very closely related (often almost on the 45° line), again indicating that parameter uncertainty is less of an issue than the smoothing method.

Summarizing, we find that the point estimates of ζ_t^2 are not very sensitive to the precise parameter values, but more sensitive to the smoothing method. However, the estimated standard errors v_t very strongly depend on the parameters and less on the smoothing method. The dependence of standard errors on the parameters also shows up in the Kalman smoother variance.

7. APPLICATION TO OPTION PRICING

An important application of the stochastic volatility model is the pricing of options. The large standard errors of the volatility estimates do not necessarily carry over to option prices. Option prices depend on the average expected volatility over the length of the option contract, and this averaging should reduce standard errors. In the limit, the average volatility over a long horizon converges to the unconditional variance, which is known without error when conditioning on the parameters of the process. On the other hand, if the persistence parameter ρ of the volatility process is close to unity, convergence to the unconditional volatility will be slow, and the errors of the smoothing estimator will be highly autocorrelated. In that case the option prices and deltas will also have large standard errors.

Let C be the value of a European call option on a currency. The value of the call option with maturity τ, exercise price X, and domestic interest rate r, is given by

$$C(S_t, h_t) = e^{-r\tau} \mathrm{E}_t[\max(S_{t+\tau} - X, 0)] \tag{18}$$

The expected payoff is a function of both the current log-volatility h_t and the current spot exchange rate S_t. The expectation is taken with respect to the risk-neutral density of the exchange rate. Hull and White (1987) assume a continuous time stochastic volatility process and show that the value of the option depends on the expected average variance of the exchange rate over the

remaining life of the option, conditional on the current volatility. Assuming that volatility risk is not priced they obtain the pricing formula:[12]

$$C(S_t, h_t) = e^{-r\tau} \int_0^\infty \mathrm{BS}(w_{t+\tau}) p(w_{t+\tau} \,|\, h_t) \, \mathrm{d}w_{t+\tau}$$
$$= e^{-r\tau} \mathrm{E}_t[\mathrm{BS}(w_{t+\tau})] \tag{19}$$

where $\mathrm{BS}(w_{t+\tau})$ is the Black–Scholes value

$$\mathrm{BS}(w_{t+\tau}) = F_t \mathrm{N}(d_1) - X \mathrm{N}(d_2) \tag{20}$$

in which F_t is the forward exchange rate applying to time $t + \tau$, and d_1 and d_2 are defined as

$$d_1 = \frac{\ln(F_t/X) + \frac{1}{2}w_{t+\tau}^2}{w_{t+\tau}} \tag{21}$$

$$d_2 = d_1 - w_{t+\tau} \tag{22}$$

$$w_{t+\tau}^2 = \int_t^{t+\tau} \exp(h_s^2) \, \mathrm{d}s \tag{23}$$

The expectation is taken with respect to the conditional density $p(w_{t+\tau} \,|\, h_t)$ of the total lifetime volatility $w_{t+\tau}$ given the current log-volatility h_t. Since it is assumed that volatility risk is not priced the density $p(w_{t+\tau} \,|\, h_t)$ coincides with the actual density. Note that the expectation is taken conditional on knowing the current log-volatility h_t, i.e. assuming that the market knows the volatility. The density $p(w_{t+\tau} \,|\, h_t)$ is a function of the parameters of the stochastic volatility process, but does not involve any data information. The econometric problem is to estimate the option value given time-series data of exchange rates, but without directly observing the volatility h_t. Volatility is only available through the stochastic volatility model.

   The pricing formula simplifies if the option is at-the-money, i.e. the exercise price X is equal to the forward rate F_t:

$$C(S_t, h_t) = e^{-r\tau} F_t \mathrm{E}_t[2\mathrm{N}(\tfrac{1}{2}w_{t+\tau}) - 1] \tag{24}$$

The option value is homogeneous of degree one in the discounted forward rate $e^{-r\tau}F_t$, so this can be set equal to one without loss of generality.

   In discrete time we replace the integral in equation (23) by the summation (see Amin and Ng, 1993):

$$w_{t+n}^2 = \sum_{i=1}^n \exp(h_{t+i}) \tag{25}$$

---

[12] The valuation of European-style currency options under stochastic volatility is analogous to the derivation for stock options in Ghysels, Harvey, and Renault (1996) and Renault (1996). The only difference is that currency options pay a 'dividend' rate equal to the foreign interest rate (see Hull, 1996, ch. 12).

where n is the number of discrete time periods until maturity of the option. The value of the option can be computed by direct simulation using the following steps:

(1) Draw $h_t^{(j)}$ from the simulation smoother.
(2) Conditional on the initial condition $h_t^{(j)}$, draw the ith path ($i = 1, \ldots, K$) of future volatilities $h_{t+s}^{(ij)}$ from the recursion

$$h_{t+s}^{(ij)} = \rho h_{t+s-1}^{(ij)} + \eta_{t+s-1}^{(ij)} \tag{26}$$

for $s = 1, \ldots, n$. From these compute $w_{t+n}^{(ij)}$ according to equation (25), and the Black–Scholes value $\mathrm{BS}(w_{t+n}^{(ij)})$.
(3) If $i < K$, select next i and go to step (2).
(4) Compute the Hull–White option value conditional on $h_t^{(j)}$ as

$$\mathrm{HW}(h_t^{(j)}) = \frac{1}{K} \sum_{i=1}^{K} \mathrm{BS}(w_{t+n}^{(ij)}) \tag{27}$$

Also compute the implied volatility $c_t^{(j)}$ by solving $\mathrm{HW}(h_t^{(j)}) = \mathrm{BS}(c_t^{(j)})$.
(5) If $j < M$, select next j and go to step (1).
(6) Compute the option value conditional on observing the data Y_T as the average $(1/M)\Sigma_{j=1}^{M}\mathrm{HW}(h_t^{(j)})$. For reporting purposes also compute the mean and variance of the implied volatilities, denoted by c_t and v_t respectively.

The algorithm is not necessarily efficient for practical use, since there exist several simplifying approximations for at-the-money options (see, for example, Ghysels, Harvey, and Renault, 1996). The inner loop (steps 2 and 3) with K independent sample paths for the volatility is introduced to facilitate the computations of standard errors. By fully integrating over possible volatility paths in step 2 we obtain for each draw $h_t^{(j)}$ the exact corresponding option value and the implied volatility, apart from numerical integration error that is negligible for large enough K. The simulation variance of the option price and the implied volatility is thus only due to the error in estimating h_t from the time-series data Y_T.

The algorithm has been applied to price a 26-weeks at-the-money call option for each of the six exchange rates and for each week in the sample. At each sample point we thus obtain the smoothed value of the option (or equivalently the implied volatility) and a standard error reflecting the precision of the smoothed estimate.[13]

The results are summarized in Figure 6 for the SIEM2 estimates, and in Figure 7 for the Bayesian analysis. Each graph shows the cumulative empirical distribution function of the implied volatility (solid line). The two thin lines in the figures show the cumulative sample distributions of the lower and upper end of a plus and minus two standard error interval around the implied volatility, i.e. $F(c_t - 2v_t)$ and $F(c_t + 2v_t)$, with $F(\cdot)$ the empirical distribution function. The vertical dashed line in the figures denotes the unconditional volatility.

The main conclusion we draw from the figures is that the unconditional volatility is very often included in the confidence intervals. Apart from 5% of the observations with extreme low volatility, the implied volatility of the three cross-rates (GEJP, UKJP and GEUK) is never significantly different from the unconditional volatility. For the three dollar rates the number of

---

[13] The parameter K was set to 500. The random numbers for $\eta_{t+s}^{(ij)}$ were common for each time period observation t, but refreshed for each iteration j of the smoother.
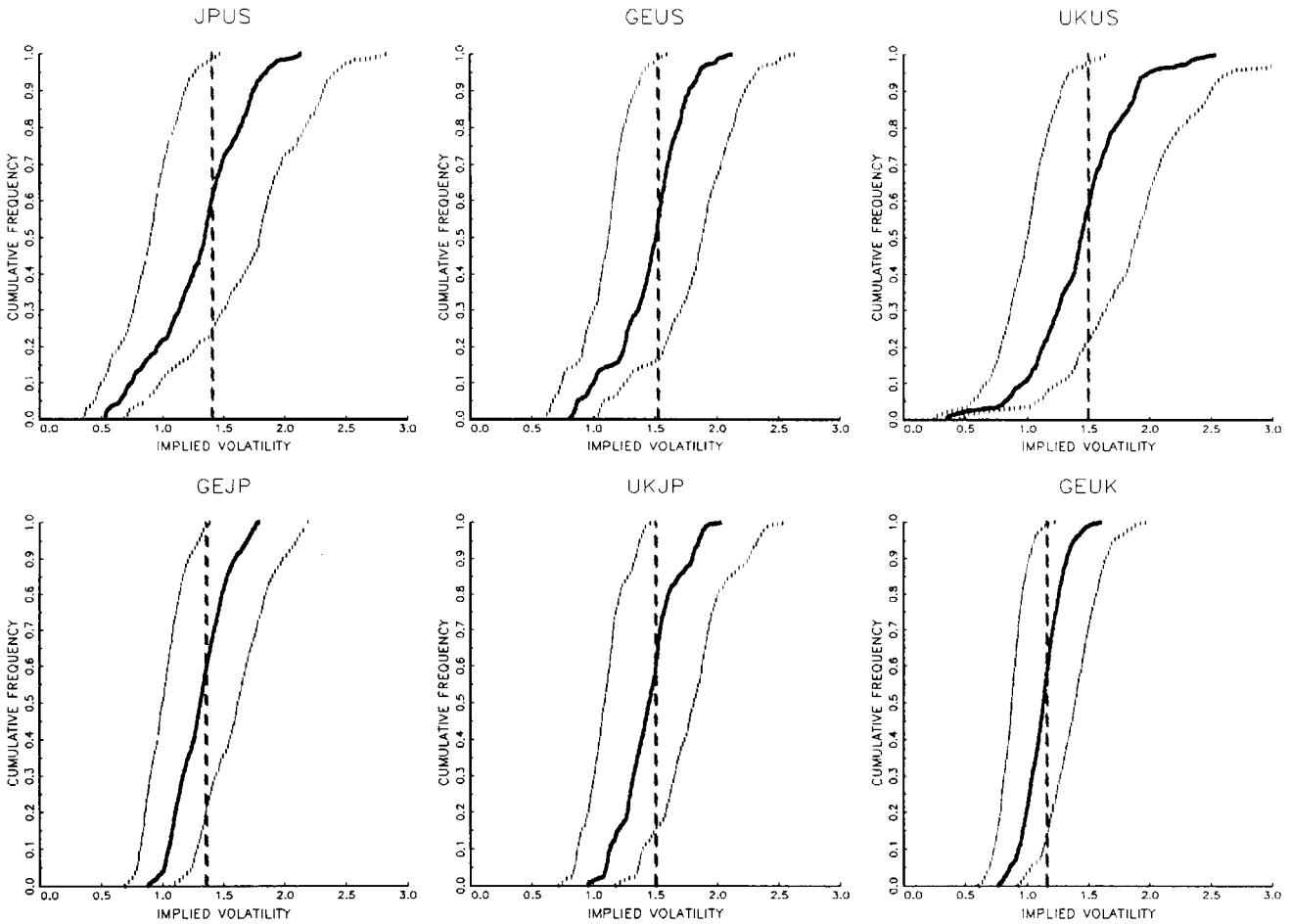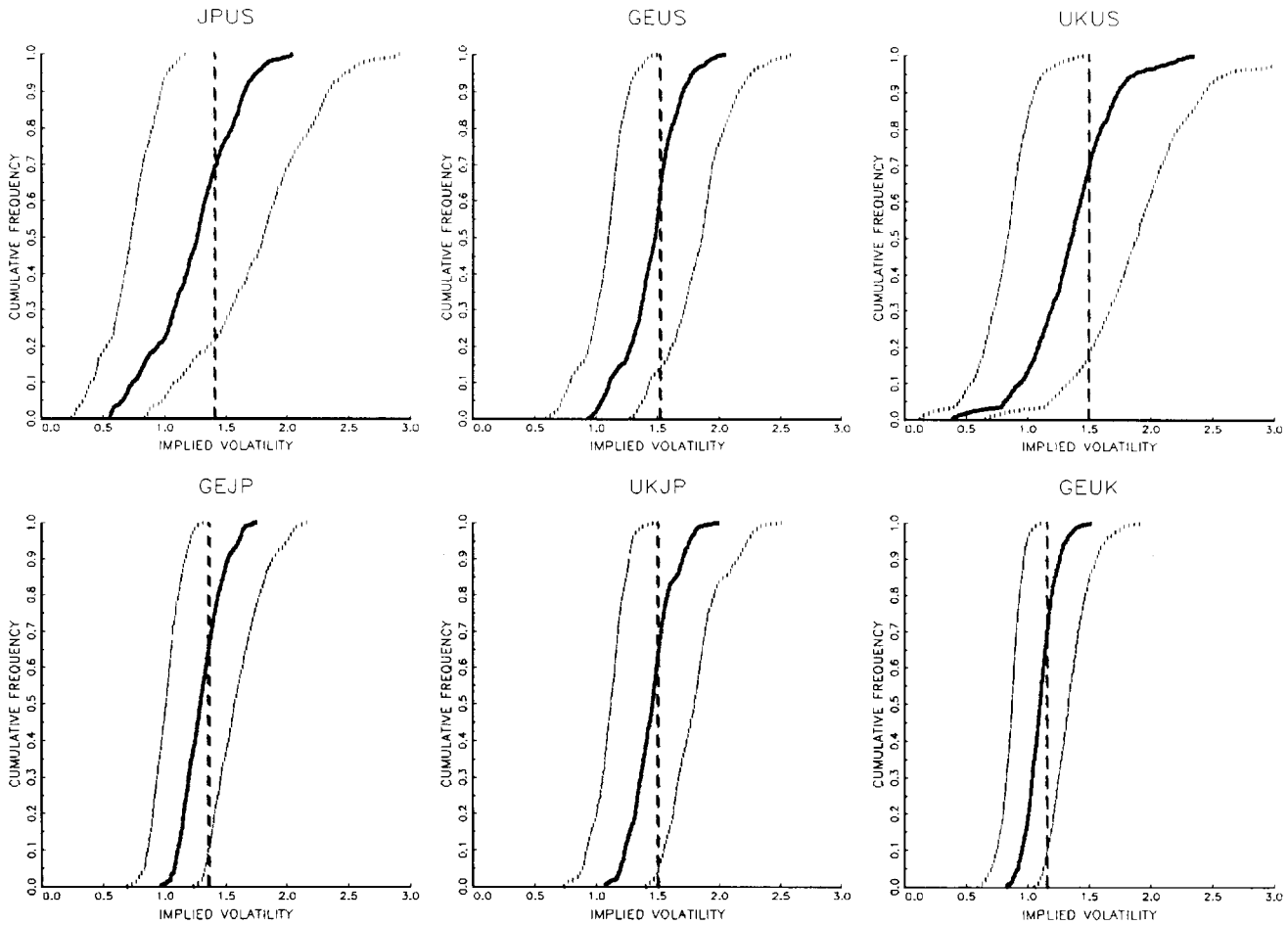
Figure 6. Option values: SIEM2. The empirical distribution function of the implied volatility for a 26-week at-the-money call option for each of the six exchange rates shown. Option values, implied volatility and standard errors are computed according to the algorithm described in Section 7. The solid line represents the implied volatility, the two thin lines are lower and upper ends of a plus and minus two standard error confidence band. The dashed vertical marks the unconditional volatility

Figure 7. Option values: Bayes. The empirical distribution function of the implied volatility for a 26-week at-the-money call option for each of the six exchange rates shown. Option values, implied volatility and standard errors are computed according to the algorithm described in Section 7. The solid line represents the implied volatility, the two thin lines are lower and upper ends of a plus and minus two standard error confidence band. The dashed vertical marks the unconditional volatility

significant deviations from unconditional volatility is between 15% and 20%. The SIEM2 and Bayesian results are very similar, indicating that parameter uncertainty does not play a major role in the results. This is the same conclusion as we drew in Section 6. The problem is that the univariate time-series analysis does not lead to accurate estimates of the volatility at a particular point in time. Averaging over 26 weeks, as in the option calculations, reduces the standard errors, but at the same time also pulls the average volatility towards the unconditional volatility.

The differences between the dollar rates and the cross-rates are due to the different degree of persistence. The larger ρ, the larger the difference between the lifetime volatility $w_{t+\tau}$ and the unconditional volatility. The significant differences appear at the low volatilities because of the exponential specification of the process. Standard errors of volatility are proportional to the volatility itself.

The general conclusion is that volatility estimates from a univariate time-series model have such large standard errors that they are hardly informative for the purpose of option pricing or currency hedging.

8. SUMMARY AND CONCLUSIONS

In this paper we have empirically studied the performance of the first-order stochastic volatility model using a dataset of weekly exchange rates. The model has been estimated for different specifications of the distribution of the standardized exchange rate innovations. Our first finding is that estimates of the persistence of the volatility process depend crucially on the stochastic specification of the model.

The other results 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.

The implications of the large standard errors are investigated in an application to option pricing. Using the Hull and White (1987) framework and our simulation procedures we computed values for at-the-money currency options with a maturity of 26 weeks. The standard error bounds around the option values are mostly wide enough that the Black–Scholes price based on the unconditional volatility is not significantly different. We conclude that volatility estimates based on a univariate time-series model have such large standard errors that they are hardly informative for the purpose of option pricing or currency hedging. These results explain some of the problems with volatility forecasting.

APPENDIX: COMPUTATIONAL DETAILS

The SIEM algorithms in the text are based on the following decomposition of the likelihood function (Kim *et al.*, 1998):

$$\ln L(\theta \mid Y_T) = \ln \Pr(Z_T) - \ln \Pr(Z_T \mid Y_T; \theta) + \ln f(Y_T \mid Z_T; \theta) \qquad \text{(A1)}$$

with Y_T the observed data and Z_T the latent data. In the stochastic volatility models that we consider in the text the latent data consist of the mixture indicators Z_T. As in Dempster *et al.*

(1977), the likelihood function is maximized by considering the term $\ln f(Y_T | Z_T; \theta)$ only. Note that values for this part of the likelihood function can be found by running the Kalman filter, since the state space of the SV model becomes Gaussian when conditioning on the mixture indicators. A multimove Monte Carlo Markov Chain (MCMC) can be used to find vectors $Z_T^{(i)}$, $i = 1, \ldots, M$. The objective function in the moment step of the simulated EM algorithm is

$$\frac{1}{M} \sum_{i=1}^{M} \ln f(Y_T | Z_T^{(i)}; \theta) \tag{A2}$$

When a new parameter vector θ^* has been found, the multimove MCMC must be run again replacing θ with θ^*. The EM algorithm is stopped when the absolute difference in subsequent parameter vectors $|\theta - \theta^*|$ is smaller than a previously set constant.

Standard errors of the parameters can be found by inverting

$$\left[-\frac{\partial^2 \ln L(\theta | Y_T)}{\partial\theta\partial\theta'} \bigg|_{\theta=\theta^*} \right] \tag{A3}$$

Using the decomposition of the likelihood function we find that

$$\frac{\partial^2 \ln L(\theta | Y_T)}{\partial\theta\partial\theta'} = \frac{\partial^2 \ln f(Y_T | Z_T; \theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \ln \Pr(Z_T | Y_T; \theta)}{\partial\theta\partial\theta'} \tag{A4}$$

Tanner (1996) shows that

$$\frac{\partial^2 \ln \Pr(Z_T | Y_T; \theta)}{\partial\theta\partial\theta'} = \mathrm{Var}\left(\frac{\partial \ln f(Y_T | Z_T; \theta)}{\partial\theta} \right) \tag{A5}$$

which in our case can be approximated by

$$\frac{1}{M} \sum_{i=1}^{M} \left(\frac{\partial \ln f(Y_T | Z_T^{(i)}; \theta)}{\partial\theta} \right)' \left(\frac{\partial \ln f(Y_T | Z_T^{(i)}; \theta)}{\partial\theta} \right) - \left\{ \frac{1}{M} \sum_{i=1}^{M} \frac{\partial \ln f(Y_T | Z_T^{(i)}; \theta)}{\partial\theta} \right\}^2 \tag{A6}$$

The derivatives can be computed numerically using common random variables.

To update the parameters in the Simulated EM cases, we have run the multimove Gibbs sampler for 500 iterations. To diminish the correlations between subsequent draws from the chain we used only every fifth draw. Thus M in equation (A2) is equal to 100. We took the estimated parameters from the unrestricted QML estimation procedure as start values for the SIEM algorithms. The optimization we used in equation (A2) is a quasi-Newton method. Convergence of the algorithms is determined by checking that subsequent values of parameters and functions (A2) are close.[14] For the simulated EM model with flexible mixture parameters the standard errors are conditional on the covariance structure of the $\hat{\mu}(z_t)$ and $\hat{\omega}^2(z_t)$. The SIEM algorithms were programmed in Fortran 90 and executed on a DEC Alpha workstation. Computation time for 100 sweeps of the algorithms was about 22 seconds.

---

[14] In fact we applied the conditions $|\theta^* - \theta| < 0.01$ and $|\ln L(Y_T|\theta^*) - \ln L(Y_T|\theta)| < 0.01$. The values of $\ln L(Y_T|\theta^*)$ and $\ln L(Y_T|\theta)$ are computed with common random variables.

In order to compute likelihood values for $\ln s_t^2$ we again consider the decomposition of the likelihood function (A1). In the stochastic volatility models that we consider in the text the latent data consist of the mixture indicators Z_T. The Simulated EM algorithm of Kim, Shephard, and Chib (1998) is used to obtain maximum likelihood estimates $(\hat{\theta})$ for the parameter vector θ. In order to obtain an estimate of the likelihood function itself we apply the ideas of Dempster *et al.* (1977). They integrate both sides of the likelihood equation (A1) with respect to the density $\Pr[Z_T | Y_T; \theta]$. Evaluating the likelihood in $\hat{\theta}$, we can rewrite equation (A1) as

$$
\begin{aligned}
\ln L(\hat{\theta} | Y_T) = & \int_{Z_T} \ln \Pr(Z_T) \Pr[Z_T | Y_T; \hat{\theta}] \, \mathrm{d}Z_T \\
& - \int_{Z_T} \ln \Pr(Z_T | Y_T; \hat{\theta}) \Pr[Z_T | Y_T; \hat{\theta}] \, \mathrm{d}Z_T \qquad \text{(A7)} \\
& + \int_{Z_T} \ln f(Y_T | Z_T; \hat{\theta}) \Pr[Z_T | Y_T; \hat{\theta}] \, \mathrm{d}Z_T
\end{aligned}
$$

Note that the Gibbs sampling scheme includes the density $\Pr[Z_T | Y_T; \hat{\theta}]$ for making mixture indicator draws $Z_T^{(j)}$, $(j = 1, \ldots, M)$. Consequently, the draws from the MCMC can be used directly to estimate the three parts of the likelihood in equation (A7). The first part can be approximated by using the mixture frequencies p_i, $(i = 1, \ldots, K)$:

$$
\int_{Z_T} \ln \Pr(Z_T) \Pr[Z_T | Y_T; \hat{\theta}] \, \mathrm{d}Z_T \approx \frac{1}{M} \sum_{j=1}^{M} \sum_{t=1}^{T} \ln p_{z_t^{(j)}}
$$

The second part uses the multinomial distribution $\Pr(Z_T | Y_T; \hat{\theta})$, which in our case is proportional to

$$
\Pr(z_t = i | y_t; \hat{\theta}) \propto \frac{p_i}{\hat{\omega}_i} \exp\left(-\frac{1}{2\hat{\omega}_i^2} (y_t - h_t - \hat{\beta} - \hat{\mu}_i)^2 \right) \quad i = 1, \ldots, K.
$$

In order to compute conditional frequencies f_i we normalize in the following way:

$$
f_i \equiv \frac{\Pr(z_t = i | y_t; \hat{\theta})}{\displaystyle\sum_{k=1}^{K} \Pr(z_t = k | y_t; \hat{\theta})}, \quad i = 1, \ldots, K.
$$

The likelihood part is then approximated as

$$
\int_{Z_T} \ln \Pr(Z_T | Y_T; \hat{\theta}) \Pr[Z_T | Y_T; \hat{\theta}] \, \mathrm{d}Z_T \approx \frac{1}{M} \sum_{j=1}^{M} \sum_{t=1}^{T} \ln f_{z_t^{(j)}}
$$

The last part of the likelihood $f(Y_T | Z_T, \hat{\theta})$ is conditionally Gaussian and can be easily calculated by applying the Kalman filter.

The parameters for the Bayesian algorithm were found by running the Gibbs sampler for 50,000 iterations. The information from every tenth iteration is used in the computation of posterior densities and volatilities. The prior means of μ_i and ω_i are set at the values of the

approximating mixture in Table I. The prior variances of the μ_i's are equal to 1000, while the prior degrees of freedom in the inverted gamma priors for ω_i is equal to 5. The prior for ρ is a truncated normal with mean one and variance 1000. It is truncated at its mode $\rho = 1$, so that we restrict the volatility process to be stationary *a priori*.

Using GAUSS 3.2 on a Pentium 200 PC the Bayesian smoother runs about 300 iterations per minute for our model with 1100 observations.

REFERENCES

Amin, K. I. and V. K. Ng (1993), 'Option valuation with systematic stochastic volatility', *Journal of Finance*, **48**, 881–909.

Andersen, T. G. (1992), 'Volatility', Working paper 144, Kellogg Graduate School of Management.

Andersen, T. G. (1994), 'Stochastic autoregressive volatility: a framework for volatility modeling', *Mathematical Finance*, **4**, 75–102.

Andersen, T. G. and B. E. Sørensen (1996), 'GMM estimation of a stochastic volatility model: a Monte Carlo study', *Journal of Business and Economics Statistics*, **14**, 328–352.

Baillie, R. T. and T. Bollerslev (1989), 'The message in daily exchange rates: a conditional variance tale', *Journal of Business and Economics Statistics*, **7**, 297–305.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics*, **31**, 307–328.

Clark, P. K. (1973), 'A subordinated stochastic process model with finite variance for speculative prices', *Econometrica*, **41**, 135–155.

Danielsson, J. (1994), 'Stochastic volatility in asset prices: estimation with simulated maximum likelihood', *Journal of Econometrics*, **64**, 375–400.

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.

Dempster, A. P., N. M. Laird and D. B. Rubin (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

Engle, R. F. (1982), 'Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation', *Econometrica*, **50**, 987–1008.

Engle, R. F. and T. Bollerslev (1986), 'Modeling the persistence of conditional variances', *Econometric Reviews*, **5**, 1–50.

Escobar, M. D. and M. West (1995), 'Bayesian density estimation and inference using mixtures', *Journal of the American Statistical Association*, **90**, 577–588.

Figlewski, S. (1994), 'Forecasting volatility using historical data', working paper, NYU.

Gallant, A. R., D. Hsieh and G. E. Tauchen (1994), 'Estimation of stochastic volatility models with diagnostics', unpublished manuscript, Duke University.

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

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 (1993), 'Estimation and testing of stochastic variance models', unpublished manuscript, Nuffield College, Oxford.

Hull, J. C. (1996), *Options, Futures, and other Derivatives*, Prentice Hall, Englewood Cliffs, NJ.

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

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

Jong, P. De and N. Shephard (1995), 'Efficient sampling from the smoothing density in time series models', *Biometrika*, **82**, 339–350.

Kim, S. and N. Shephard (1994), 'Stochastic volatility: optimal likelihood inference and comparison with ARCH models', working paper, Nuffield College, Oxford.

Kim, S., N. Shephard and S. Chib (1998), Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies*, 65, forthcoming.

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.

Renault, E. (1996), 'Econometric models of option pricing errors', in D. M. Kreps and K. F. Wallis (eds), *Advances in Economics and Econometrics: Theory and Applications*, Cambridge University Press, Cambridge.

Schotman, P. C. (1994), 'Priors for the AR(1) model: parameterization issues and time series considerations', *Econometric Theory*, **10**, 579–595.

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

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

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, New York.

West, K. D. and D. Cho (1995), 'The predictive ability of several models of exchange rate volatility', *Journal of Econometrics*, **69**, 367–391.