Bayesian Analysis of Stochastic Volatility Models

Article *in* Journal of Business and Economic Statistics · February 1994

DOI: 10.2307/1392199 · Source: RePEc

4 authors, including:

Peter E. Rossi
University of California, Los Angeles
134 PUBLICATIONS **15,125** CITATIONS

Journal of Business & Economic Statistics, October 1994, Vol. 12, No. 4

Editor's Note: The following article was the *JBES* Invited Address presented at the Joint Statistical Meetings, Toronto, Ontario, August 14–18, 1994.

Bayesian Analysis of Stochastic Volatility Models

Eric JACQUIER
Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853

Nicholas G. POLSON and Peter E. ROSSI
Graduate School of Business, University of Chicago, Chicago, IL 60637

New techniques for the analysis of stochastic volatility models in which the logarithm of conditional variance follows an autoregressive model are developed. A cyclic Metropolis algorithm is used to construct a Markov-chain simulation tool. Simulations from this Markov chain converge in distribution to draws from the posterior distribution enabling exact finite-sample inference. The exact solution to the filtering/smoothing problem of inferring about the unobserved variance states is a by-product of our Markov-chain method. In addition, multistep-ahead predictive densities can be constructed that reflect both inherent model variability and parameter uncertainty. We illustrate our method by analyzing both daily and weekly data on stock returns and exchange rates. Sampling experiments are conducted to compare the performance of Bayes estimators to method of moments and quasi-maximum likelihood estimators proposed in the literature. In both parameter estimation and filtering, the Bayes estimators outperform these other approaches.

KEY WORDS: Bayesian inference; Markov-chain Monte Carlo; Method of moments; Nonlinear filtering; Quasi-maximum likelihood; Stochastic volatility.

Interest in models with stochastic volatility dates at least to the work of Clark (1973), who proposed an iid mixture model for the distribution of stock-price changes. Unobservable information flow produces a random volume of trade in the Clark approach. Tauchen and Pitts (1983) and Gallant, Hsieh, and Tauchen (1991) noted that if the information flows are autocorrelated, then a stochastic volatility model with time-varying and autocorrelated conditional variance might be appropriate for price-change series. Stochastic volatility models also arise as discrete approximations to various diffusion processes of interest in the continuous-time asset-pricing literature (Hull and White 1987; Melino and Turnbull 1990; Wiggins 1987).

The purpose of this article is to develop new methods for inference and prediction in a simple class of stochastic volatility models in which logarithm of conditional volatility follows an autoregressive (AR) times series model. Unlike the autoregressive conditional heteroscedasticity (ARCH) and generalized ARCH (GARCH) models [see Bollerslev, Chou, and Kroner (1992) for a survey of ARCH modeling], both the mean and log-volatility equations have separate error terms. The ease of evaluating the ARCH likelihood function and the ability of the ARCH specification to accommodate the time-varying volatility found in many economic time series has fostered an explosion in the use of ARCH models. On the other hand, the likelihood function for stochastic volatility models is difficult to evaluate, and hence these models have had limited empirical application.

The current literature on inference and prediction for stochastic volatility models is brief. Rather than pursuing a likelihood-based approach, Taylor (1986), Melino and Turnbull (1990), and Vetzal (1992) relied on the method of moments (MM) to avoid the integration problems associated with evaluating the likelihood directly. As is well known, the MM may be inefficient relative to a likelihood-based method of inference. This problem is particularly severe in the case of stochastic volatility models because the score function cannot be computed to suggest which moments should be used for MM estimation. Nelson (1988), Harvey, Ruiz, and Shephard (1994), and Ruiz (1994) employed approximate linear filtering methods to produce a quasi-maximum likelihood (QML) estimator. As Harvey et al. (1994) and Ruiz (in press) pointed out, the accuracy of the normality approximation used in the filtering approach will depend on where the true parameters lie in the parameter space. Specifically, the approximation will worsen as the volatility equation variance decreases. Finally, Danielsson (in press) and Danielsson and Richard (1992) developed new methods for approximating the integral used in evaluating the likelihood function. Without a direct and "error-free" method for evaluating the likelihood, it is difficult to gauge the accuracy of the integral approximations proposed in these articles.

In applications of stochastic volatility models to financial data, prediction and filtering as well as parameter estimation are major goals of the analysis. The literature only offers approximate filtering solutions to the problem of inferring about the unobservable volatilities and predicting future volatility. Furthermore parameter estimates are routinely

"plugged in" to multistep-ahead prediction formulas, and often the contribution of parameter uncertainty to forecast variability is not accounted for. Finally, all of the techniques discussed in the stochastic volatility literature rely on asymptotic approximations to conduct inference.

We propose a new Bayesian approach in which the latent volatility structure is directly exploited to conduct finite-sample inference and calculate predictive distributions. We augment the stochastic volatility parameters with the time series of volatilities and construct a Markov chain that can be used to draw directly from the joint posterior distribution of the model parameters and unobservable volatilities. Our algorithm combines the idea of data augmentation as advanced by Tanner and Wong (1987) with a hybrid Metropolis independence chain (see Tierney 1991). In less than one half-hour of workstation central processing unit (CPU) time, we compute simulation-based estimates of posterior quantities of interest to a very high degree of accuracy.

The rest of the article is outlined as follows. Section 1 describes our algorithm and discusses its theoretical properties. Section 2 discusses the problem of smoothing and developing predictive distributions of future variances. Section 3 applies the method to analysis of stock-return data. Section 4 reports sampling experiments that compare the performance of our likelihood-based Bayes estimator with both MM and QML procedures. Section 5 provides results on filtering performance.

1. THE MODEL AND MARKOV-CHAIN MONTE CARLO APPROACH

The general stochastic volatility model views the time series of the data, y, as a vector generated from a probability model, $p(y \mid h)$, where h is a vector of volatilities. Each data point y_t has variance h_t, which is time dependent. The volatilities h are unobserved and are assumed to be generated by the probability mechanism, $p(h \mid \omega)$. The density of the data is a mixture over the h distribution, $p(y \mid \omega) = \int p(y \mid h)p(h \mid \omega)dh$.

Carlin, Polson, and Stoffer (1992) used a Gibbs sampling procedure for analysis of nonlinear state-space models, allowing for nonnormal error densities but with a time invariant scale h. Our focus here is on modeling variation in h over time rather than elaborations of the process governing the mean. McCulloch and Tsay (1993) considered a class of priors for variance changes in which the ratios of volatilities have a random-walk component based on an inverse gamma innovation. Their focus was more on the filtering problem with the process parameters specified a priori. In our approach, we wish to consider prediction and filtering with a hierarchical structure in which we infer about the volatility equation parameters rather than just fixing them to implement a filtering procedure.

Uhlig (1991) introduced a different volatility model in which the ratio of volatilities has a Beta distribution. Uhlig also provided a multivariate model in which the covariance structures have a generalized multivariate Beta distribution. In Uhlig's models, the range of possible variance ratios is restricted to a finite interval and has a complicated interaction with the variance of the volatility. The advantage of Uhlig's approach is that it provides exact expressions for the marginal likelihood. One must rely on high-dimensional numerical integration, however, to provide a solution to the problem of inferring about the Beta distribution range parameters as well as solving the prediction and filtering problem.

1.1 Simple Stochastic Volatility Models and Comparisons to ARCH Approaches

To focus discussion on the key aspects of modeling and estimating stochastic volatility models, we will start with a simple model in which the conditional variance of a series $\{y_t\}$ follows a log-AR(1) process. Jacquier, Polson, and Rossi (1993) considered priors and methods for the general multivariate case: $y_t = \sqrt{h_t}u_t$, $\ln h_t = \alpha + \delta \ln h_{t-1} + \sigma_v v_t$, and $(u_t, v_t) \sim$ independent $N(0, 1)$. Here $\omega' = (\alpha, \delta, \sigma_v)$. In our Bayesian simulation framework, it will be a simple matter to introduce exogenous regressors into the mean equation as well as to accommodate an AR(p) process for the log variance.

Although the preceding model is quite parsimonious, it is capable of exhibiting a wide range of behavior. Like ARCH/GARCH models, the model can give rise to a high persistence in volatility (sometimes referred to as "volatility clustering"). Even if $\delta = 0$, the model is a variance mixture that will give rise to excess kurtosis in the marginal distribution of the data. In ARCH/GARCH models with normal errors, the degree of kurtosis is tied to the roots of the variance equation; as the variances become more autocorrelated, the degree of mixing also increases. In the ARCH/GARCH literature, it has become common (e.g., see Nelson 1991) to use nonnormal innovation densities to accommodate the high kurtosis of various financial time series. In the stochastic volatility model, the σ_v parameter governs the degree of mixing independently of the degree of smoothness in the variance evolution.

Jacquier et al. (1993) demonstrated how to accommodate correlation between the mean and variance equations errors, which will introduce an asymmetry into the conditional variance function of the sort documented in the EGARCH literature. Although adding correlation is an interesting extension, not all economic time series display a "leverage" effect. Evidence from the EGARCH literature suggests that the leverage effect is small for interest-rate and exchange-rate series. Vetzal (1992) found a small and insignificant correlation in his analysis of the treasury bill (T-bill) series. Gallant, Rossi, and Tauchen (1992) found that the leverage effect in a stock index series is sensitive to conditioning arguments and outliers.

There are important differences between likelihood functions for the standard GARCH and stochastic volatility models. To illustrate this point, we fit GARCH and stochastic volatility models to weekly returns on a portfolio formed of the smallest decile of listed New York Stock Exchange

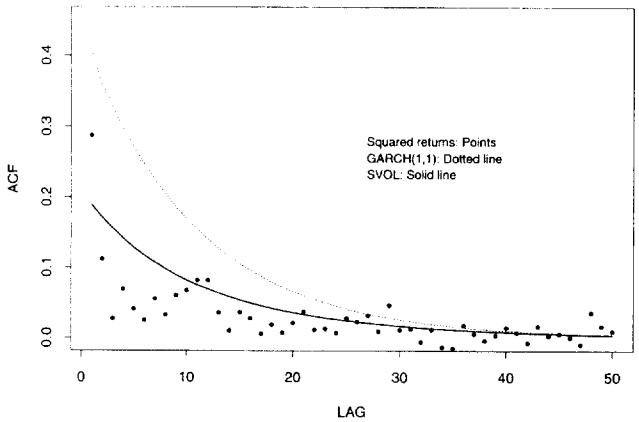

Figure 1. ACF of Squared Returns, Fitted GARCH, and SVOL Models, Weekly Decile 1 Portfolio: Squared Returns, Points; GARCH (1,1), Dotted Line; SVOL, Solid Line.

(NYSE) stocks (see Sec. 3 for a complete analysis of other portfolios and individual stocks). Based on the Schwarz criterion, we select a GARCH(1, 1) model that was fitted by maximum likelihood. The stochastic volatility model is fitted using the Bayes estimators developed in this article (see Table 1, Sec. 3, for the estimates). Figure 1 compares the autocorrelation of the squared returns with the implied theoretical autocorrelations of the fitted stochastic volatility model (see Appendix A for the expression) and the GARCH model [see Bollerslev (1986) for derivation of GARCH autocorrelation function (acf)]. The stochastic volatility and GARCH acf's are computed by inserting the Bayes or maximum likelihood estimators (MLE's) in the theoretical expressions. The fitted GARCH and stochastic volatility models have very different implied acf's with the stochastic volatility model in closer conformance with the data. This serves to illustrate that the likelihood functions for the GARCH and stochastic volatility models put different weights on various moments functions. This point of view is further corroborated by Hsieh's (1991) findings that ARCH/GARCH/EGARCH filters do not remove all nonlinear dependencies in the series as measured by a battery of BDS-inspired statistics. On the other hand, Hsieh found that when an autoregressive stochastic volatility filter is applied to the data, no remaining nonlinear dependencies are found. Danielsson (in press) also provided some goodness-of-fit evidence that the stochastic volatility model may perform better than some variants of ARCH models including the EGARCH model.

Our goal in this article is to provide a practical method for analysis of stochastic volatility models and not to resolve the controversy over whether ARCH or stochastic volatility models provide a better approximation to the data-generating process for economic and financial data. Practical difficulties have kept investigators from using the stochastic volatility model even though it offers a natural alternative to ARCH and has some advantages in the smoothing and prediction of unobserved variances. It is our view that the limited evidence available at this time suggests that stochastic volatility models are a promising alternative to various ARCH variants.

1.2 The Marginal Likelihood Dilemma

As noted by many authors, the fundamental problem with even the simplest stochastic volatility specification is that the marginal likelihood over the parameters of the stochastic volatility process is defined by a T-dimensional integral $\ell(\omega) = \int p(y \mid h)p(h \mid \omega)dh$, where $\omega' = (\alpha, \delta, \sigma_v, \rho)$, $y' = (y_1, \ldots, y_T)$ and $h' = (h_1, \ldots, h_T)$. The stochastic volatility model can be thought of as specifying a prior over the sequence of $\{h_t\}$. Our prior consists of the view that the volatilities evolve smoothly (for large positive δ) according to the AR process. A naive strategy for approximating ℓ would be to simulate from the "prior" $p(h \mid \omega)$ and average the conditional likelihood, $p(y \mid h)$, over these draws, $\hat{\ell}(\omega) = \Sigma_p(y \mid h_i)/H$, where h_i $(i = 1, \ldots, H)$ are simulation draws from the $p(h \mid \omega)$ model. In more than a few dimensions, it is well known that this sort of integration strategy will not work. Because we do not know where the likelihood has mass, we are drawing from the prior to compute the expectation of the conditional likelihood, $p(y \mid h)$. The prior may be centered far away from the conditional likelihood. As we see in Section 1.3, one important advantage of our Markov-chain simulation approach is that draws are made from the posterior rather than prior distribution of $h \mid \omega$.

Inspired by the ideas of importance sampling, Danielsson and Richard (1992) proposed a more accurate integration strategy to calculate $\ell(\omega)$. They approximated ℓ and then used these approximations along with a derivative-free optimizer to produce estimates. Although they were able to assess the accuracy of their method in evaluating the likelihood for $\delta = 0$ (this is the iid mixing model), the unavailability of the likelihood for any other value of the parameters makes it difficult to evaluate the accuracy of their methods for any other value of δ. Even if it is possible to evaluate the marginal likelihood with low error, asymptotic theory must be used to approximate the sampling distribution of the approximate MLE. Another alternative would be to use the Danielsson and Richard strategy to evaluate the likelihood in a more standard Bayesian analysis. Monte Carlo methods of numerical integration could be used to compute posterior expectations of functions of interest such as posterior means. In effect, this strategy would nest the Danielsson and Richard integration strategy inside of a standard importance sampling approach to conducting posterior analysis. Because thousands of likelihood evaluations would be required to set up the importance function and perform the Monte Carlo integration, however, a Bayesian analysis based on the Danielsson–Richard simulated likelihood method is not yet computationally feasible, based on the timing estimates for likelihood evaluation given by Danielsson (in press; 1993 personal communication).

1.3 Bayesian Analysis and Markov-Chain Monte Carlo

We view the specification of the stochastic volatility model as a hierarchical structure of conditional distributions. The hierarchy is specified by a sequence of three distributions, the conditional distribution of $y \mid h$, the conditional distribution of $h \mid \omega$, and the marginal or prior distribution of ω. In this

view, the stochastic volatility equation, governed by $p(h \mid \omega)$, is a prior with hyperparameters ω. The joint posterior distribution of h and ω is proportional to the product of these three distributions. In this sense, we have augmented the parameters, ω, with h to form a large parameter vector, $(\omega, h) \in \Omega \times H$. This idea is termed data augmentation and was pioneered in Markov-chain Monte Carlo (MCMC) by Tanner and Wong (1987). The joint posterior is given by Bayes theorem $\pi(h, \omega \mid y) \propto p(y \mid h)p(h \mid \omega)p(\omega)$. From this joint posterior, the marginal $\pi(\omega \mid y)$ can be used to make inferences about the stochastic volatility parameters, and the marginal $\pi(h \mid y)$ provides the solution to the "smoothing" problem of inferring about the unobserved volatilities. To compute these marginal distributions, we construct a Markov chain with invariant distribution π. Extremely long samples are computable in short order on any workstation, enabling us to construct very precise simulation-based estimates of any posterior quantity. Mueller (1991) developed Markov-chain algorithms for analysis of factor ARCH and other dynamic models for which the marginal likelihood is easily computed. Mueller's methods do not extend to the stochastic volatility case.

Breaking the joint posterior into various conditional distributions is the key to constructing the appropriate Markov-chain sampler. For example, conditional on h, the posterior distribution of $\omega, p(\omega \mid y, h)$, is simple to compute from standard Bayesian treatment of linear models. If it were also possible to sample directly from $p(h \mid \omega, y)$, then we could easily construct a Markov chain by alternating back and forth between drawing from $p(h \mid \omega, y)$ and $p(\omega \mid h, y)$. This process of alternating between conditional distributions produces a cyclic chain, a special case of which is the Gibbs sampler (see Gelfand and Smith 1990).

In the case of the stochastic volatility model, it is not possible to sample directly from $p(h \mid \omega, y)$ at low cost. Instead, we decompose the joint distribution, $p(h \mid \omega, y)$ further into the set of conditionals $p(h_t \mid h_{-t}, \omega, y)$, where h_{-t} denotes the rest of the h vector other than h_t. In our case, it is difficult to sample from these univariate conditional distributions. Instead, our approach is to construct a hybrid method that uses a series of Metropolis accept/reject independence chains which do not directly sample from the conditionals $p(h_t \mid h_{-t}, \omega, y)$ but nonetheless provide the posterior as the stationary distribution of the chain. Our algorithm is known as a cyclic independence Metropolis chain. We now give the details on how to implement such a Markov chain.

1.4 The Algorithm

To construct the chain, we first sample from $p(\omega \mid h, y)$ and then sample indirectly from $p(h \mid \omega, y)$. We specify a standard natural conjugate prior for (β, σ_v), $\beta' = (\alpha, \delta)$, $p(\beta, \sigma) = p(\beta \mid \sigma)$, $p(\sigma)$, $\beta \mid \sigma \sim N(\overline{\beta}, \sigma^2 A^{-1})$, $\sigma \sim IG(\nu_0, s_0^2)$. The posterior for $\omega \mid h$ is available from standard linear-models theory. We can easily draw the ω vector at one time by drawing from the appropriate multivariate normal and gamma random variables. Clearly, we can extend the AR(1) model for the volatilities to include higher-order returns, keeping the standard linear-models framework. One obvious advantage of the Bayesian approach for the general AR (p) model is that we can put on "smoothness" or "damping" priors on the AR coefficients to deal with the problem of excessive sampling variability caused by the addition of more model parameters.

To draw from $h \mid \omega, y$ requires more effort. We can easily exploit the Markovian structure of the stochastic volatility model to break down the joint posterior of the entire h vector by considering the series of univariate conditional densities, $p(h_t \mid h_{t-1}, h_{t+1}, \omega, y_t)$, $t = 1, \ldots, T$. If it were possible to draw directly from these univariate densities, our algorithm would reduce to a Gibbs sampler in which we would draw successively from $p(\omega \mid h, y)$ and then each of the T univariate conditionals in turn to form one step in the Markov chain. Because T draws from these univariate conditional densities are required for one step in the chain and then the chain has to be iterated for several thousand steps, however, it is imperative that the univariate sampling methods be highly efficient.

The univariate conditional densities have an unusual form, which is produced by the normal form for the conditional sampling density and the lognormal forms for the volatility equations,

$$
\begin{aligned}
(*) \quad p(h_t \mid h_{t-1}, h_{t+1}, \omega, y_t) &\propto p(y_t \mid h_t)p(h_t \mid h_{t-1})p(h_{t+1} \mid h_t) \\
&\propto h_t^{-.5} \exp\{-.5y_t^2/h_t\} 1/h_t \\
&\qquad \exp\{-(\ln h_t - \mu_t)^2/(2\sigma^2)\},
\end{aligned}
$$

where $\mu_t = (\alpha(1 - \delta) + \delta(\ln h_{t+1} + \ln h_{t-1}))/(1 + \delta^2)$ and $\sigma^2 = \sigma_v^2/(1 + \delta^2)$. Here we have combined both of the lognormal terms and completed the square on $\ln h_t$. In these cases, when the density is not of a standard form, it is natural to consider an accept/reject sampling method. An ideal accept/reject density, $q(x)$, is a density for which there exists a constant such that $p(x) \leq cq(x)$ for all x and for which the ratio $p(x)/q(x)$ is relatively constant over the range of x in which p has most of its mass. In this application, the fact that there is no analytic expression for the normalizing constant of the conditional density further complicates matters. The normalizing constant is a function of the conditioning arguments, which will vary as the sampler progresses. Thus, even with a valid accept/reject density that truly dominates p, we would have to find the constant c required to raise q over p for each time period, which could involve an optimization problem at each draw. The solution commonly taken in the random-number-generation literature (e.g., see Ripley 1987, chap. 3) is to develop customized bounding procedures for each special case. These ideas do not adapt well to simulating from conditional densities with unknown and changing normalizing constants. Furthermore, even if the blanketing constant could be found at low computational cost, it may be difficult to find a valid accept/reject density that would be efficient in the sense of accepting a high percentage of the draws.

Our solution is to modify the Markov chain by allowing repeats of points in the sample sequence producing what is called a Metropolis chain. It should be emphasized that although the Metropolis chain repeats points in the draw

sequence the invariant distribution of the chain is absolutely continuous. First, we choose a candidate simulation density, q, which is cheap to sample from and closely mimics the shape of the univariate conditional density, p. We do not actually have to guarantee dominance, however. Second, we introduce a Metropolis independence chain to handle situations in which the sample points from q are such that dominance fails [see Tierney (1991) for a discussion of independence Metropolis chains]. Shephard (1993) used a Metropolis chain to implement the E step of a simulated EM algorithm for the random-walk special case of the stochastic volatility models considered here. To implement the simulated EM approach, a decision rule must be established that determines the number of steps that the Metropolis algorithm is iterated at each E step. Because there is no theoretical guidance in selection of this rule, it must be established that a proposed rule provides a large enough number of Metropolis steps to endow the simulated EM with adequate sampling properties for all relevant regions of the sampling space.

We choose a candidate simulation density to be the kernel of the independence chains by exploiting the special form of the conditional density function. The first term in the density expression $(*)$ is the density of the inverse of a gamma-distributed random variable (this is to be distinguished from the inverted gamma density): $X \sim 1/Z$, $Z \sim \text{gamma}(\phi, 1/\lambda)$; $P_x(x) = \lambda^{\phi}/\Gamma(\phi)\, x^{-(\phi+1)} \exp(-\lambda/x)$. We approximate the lognormal term by matching the first and second moments of the lognormal to the moments of the inverse gamma. The two inverse gamma density forms can then be combined to form one inverse gamma with parameters, $\phi = (1 - 2\exp(\sigma^2))/(1 - \exp(\sigma^2)) + .5$; $\lambda = (\phi - 1)(\exp(\mu_t + .5\sigma^2)) + .5y_t^2$. The inverse gamma is a good choice for a simulation "blanketing" density because its right tail is algebraic, allowing it to dominate the lognormal density on the right. Our experiments with a lognormal blanketing density have shown the danger of choosing a blanketing density that does not dominate in the right tail.

The only remaining problem is to devise a method for choosing c. Because we no longer require that c be chosen so that cq dominates p, we choose c so as to trade off too frequent rejection from high c values with too frequent "staying" from the Metropolis mechanism. We compute the ratio of the unnormalized p density and normalized q density at the three points centered on the mode of the inverse gamma distribution. The median of these ratios is used to calculate c. We find that we accept between 70% and 80% of all draws when they are constructed in this manner and "stay" less than 5% of the time. This indicates that the inverse gamma density is a good approximation to the conditional density.

To construct the Markov chain on the full state space, $(h, \omega) \in H \times \Omega$, we piece together T Metropolis independence accept/reject chains to handle each coordinate in the h vector and the generation from $p(\omega \mid h, y)$. Each chain samples from the accept/reject density at each step and modifies one h_t coordinate. We construct a chain that moves from the point (h, ω) to (h', ω') by cycling through T Metropolis chains to update h to h' and then draw $w \mid h,' y$ to update ω to ω'. We now establish that this hybrid chain has invariant distribution π, the joint posterior of (h, ω).

1.5 Convergence of the Cyclic Metropolis Chain

The basic idea behind MCMC is to specify a Markov chain, with transition kernel P that is ergodic (irreducible and aperiodic) with π as its stationary distribution. Two widely used chains are the Gibbs sampler (Geman and Geman 1984) and the Metropolis algorithm (see Tierney 1991 for a general discussion). The desired distribution π can be simulated by picking an arbitrary initial state X_0 and then applying the transition kernel of the chain giving rise to the sequence $\{X_t\}$. The general Metropolis algorithm works as follows. Let $Q(x, y)$ be the Metropolis transition kernel. Suppose that the chain is currently at $X_t = x$. Then generate a candidate point as the new point, X_{t-1}. We accept this new point, y, with probability $\alpha(x, y)$, where $\alpha(x, y) = \min\{(p(y)Q(x, y))/(p(x)Q(y, x)), 1\}$. Here we only need to know π up to a proportionality constant, $\pi(x) \propto p(x)$. A Metropolis independence chain is fashioned by sampling candidate steps, Y, from a fixed density that does not depend on x (hence the term "independence" chain). For the independence chain, with transition kernel $Q(x, y) = f(y)$, the acceptance probability is given by $\alpha(x, y) = \min\{w(y)/w(x), 1\}$, where $w(z) = p(z)/f(z)$. If rejection sampling is used, then $f(x) \propto \min\{p(x), cq(x)\}$, where q is the accept/reject density.

If $cq(x)$ were to dominate for all x, then this chain would simply "pass through" iid accept/reject draws and provide an iid sampling method. If, however, cq does not dominate p, then there is the chance that the chain will simulate an ordinate at which $w(y)/w(x)$ will be less than 1. In these circumstances, the chain may actually choose to stay at a point for more than one draw. The intuition behind the working of the Metropolis chain is that, although accept/reject sampling works by rejecting draws so as to reduce the sampling density at points where cq is much larger than p, the Metropolis builds up mass at points where p dominates cq by repeating values.

To implement a Metropolis independence chain on the whole state space $(H \times \Omega)$ would be inadvisable because, although we can calculate the joint posterior of h and ω up to a constant of proportionality, accept/reject sampling in high dimensions can be very inefficient. Our solution is to define the transition kernel as the product of T kernels for each of the T elements of the h vector along with the conditional distribution of $\omega \mid h$, P: $P_1 P_2 \cdots P_T P_{\omega \mid h}$, where P_t updates $h_t \mid \omega$ and $P_{\omega \mid h}$ updates ω. Although each of the P_t chains is reducible because they only update one coordinate at a time), the product has a transition kernel that is positive everywhere and is therefore irreducible and aperiodic (see Tierney 1991, p. 5). It only remains to check that the posterior, π, is the invariant distribution. Standard arguments involving the time reversibility of the general Metropolis chain (see Tierney 1991, p. 3) can be used to verify that the posterior, π, is the unique invariant distribution of the chain.

Although it is easily verified that our chain is ergodic and converges to the correct stationary distribution, in practice we

must be satisfied that the draws are coming from a distribution close to the stationary distribution. This amounts to determining how long the chain must be run to dissipate the effects of the initial state, X_0. Analysis of the second eigenvalues of the chain (Applegate, Kannan, and Polson 1991) can provide exact information on the run length required to ensure that the chain is within a given tolerance of the invariant distribution. For most models, however, these eigenvalues cannot be calculated. A body of current research seeks to obtain tight bounds on these eigenvalues. For our problem, these bounds are not sufficiently tight to be useful. Therefore, we must rely on simulation experience to assess the speed of convergence.

One common practice, which we follow in our empirical applications, is to discard the first T^* draws under the assumption that the chain will have converged to within a very close tolerance to the invariant distribution. In practice, T^* must be chosen by experience with simulated data. We simulate data from a known stochastic volatility model and verify that the sampler converges in distribution after a specified number of draws. After extensive experimentation, we find that the sampler converges very rapidly to the stationary distribution. The excellent sampling performance of the Bayes estimator reported in Section 4 provides an indirect confirmation that the sampler has converged quickly.

Another issue that has received some attention (see Geweke 1992; McCulloch and Rossi in press) is the information content of a given draw sequence. We use the empirical distribution of the draws to estimate posterior quantities of interest [it is also possible to average the conditional distributions as suggested by Gelfand and Smith (1990)]. It is important to ensure that these posterior estimates are sufficiently precise for the purpose at hand. In the case of stochastic volatility models, however, a vast number of draws (exceeding 50,000) can be drawn in a matter of hours on even a slow workstation, rendering these considerations less important.

In summary, it is important to remember that these sorts of Markov-chain samplers converge to the required posterior distribution under only mild and verifiable conditions. Furthermore, allocation of more computer time to the simulation will unambiguously increase the amount of information regarding the posterior and provide better inferences. No amount of additional computing resources, however, can improve the quality of inference obtained from asymptotic procedures.

2. SMOOTHING AND FORECASTING

One important use of stochastic volatility models is to infer about values of unobservable conditional volatility both within the sample (smoothing) and out of sample (prediction). For example, option-pricing applications often require some sort of estimates of conditional volatility, and "event"-style studies may wish to relate specific events to changes in volatility. To provide a solution to the smoothing problem, we must compute the conditional distribution of $h_t \mid y$, where $y' = (y_1, \ldots, y_T)$ is the whole data vector. If this distribution were available, we could simply use $E[h_t \mid y]$ as the smoothed estimate of h_t. This density also summarizes our uncertainty about the unobservable h_t. An analytical expression for $p(h_t \mid y)$ is unavailable, even conditional on the volatility parameters ω. We will see, however, that a Monte Carlo estimate of this density is available using our draws from the joint posterior, $\pi(h, \omega)$.

When an investigator uses the MM or simulated likelihood approach to produce parameter estimates, a nonlinear filtering problem has to be solved after the estimation to produce smoothed estimates of h_t. For example, Melino and Turnbull (1990) used approximate Gaussian filtering methods to extract smoothed estimates conditional on MM parameter estimates. These approximate Gaussian filtering methods do not necessarily provide the optimal nonlinear filter. In a related approach, Foster and Nelson (1992) constructed an approximate nonlinear filter based on the EGARCH model.

Our method avoids any asymptotic approximations and provides draws that can be used to solve the smoothing problem as a natural by-product of the simulation process [see Carlin et al. (1992) and McCulloch and Tsay (1993) for related approaches]. Our Monte Carlo (MC) sampler has invariant distribution, which is the joint posterior of (h, ω). The solution to the smoothing problem is the marginal posterior of $h_t, p(h_t \mid y)$. We simply use the draws of the h vector to form estimates of these marginals for all T observations. These draws must be made to construct the chain so that they are available at no additional computational cost. Furthermore this marginal distribution directly accounts for parameter uncertainty because $p(h \mid y) = \int p(h \mid \omega, y)\, p(\omega \mid y) d\omega$. The solution to the nonlinear filtering problem is simply $E[h_t \mid y]$, which we can approximate to any desired degree of accuracy by using the draws from the marginal distribution, $p(h_t \mid y)$.

Our MC sampling framework can be extended to include forecasting up to an arbitrary step ahead. The goal of any forecasting exercise is to compute the predictive distribution of a future vector of observations given the past observations. In the case of volatility forecasting, our goal is to compute the predictive density of a vector of future volatilities given the sample data, $p(h_f \mid y)$, where $h_f' = (h_{T+1}, \ldots, h_{T+k})$. Our approach is to compute the joint posterior distribution of the h vector including both sample and future h's and simply marginalize on the future h's, which will be the predictive distribution. Define $y_f' = (y_{T+1}, \ldots, y_{T+k+1}), h^{*\prime} = (h', h_f')$, and $y^* = (y', y_f')$. The key insight is that, given y^*, we have already shown how to draw from $p(h^* \mid y^*, \omega)$ and that $p(y_f \mid y, h^*, \omega)$ is simple to draw from. We now piece together three collections of conditional distributions to construct the following chain:

1. $y_f \mid y, h^*, \omega$ (iid normal draws)
2. $h^* \mid y^*, \omega$
3. $\omega \mid h^*$

The second and third conditionals are sampled using the cyclic Metropolis approach outlined in Section 1.

To make forecasts using a standard filtering framework, estimates of h_T are formed using filtering methods, and then

estimated parameters are used in conjunction with the model to forecast future variances. There are now two sources of error that must be considered, error in the filtering estimates of h_T and parameter estimation error. Even if the sampling variability in the parameter estimates were negligible, uncertainty in the level of h_T must be reckoned with. There is no reason to believe that asymptotic normal approxmations would result in accurate prediction intervals, especially for multistep-ahead forecasts that will involve nonlinearities in the estimate of the persistence parameter, $\hat{\delta}$.

3. ANALYSIS OF SELECTED STOCKS AND PORTFOLIOS

To illustrate our Markov-chain simulation method and to compare findings from the stochastic volatility models with findings in the ARCH/GARCH literature, we analyze 15 data sets on a variety of stock portfolios, selected individual stocks, and exchange rates. For the stock data, we rely on weekly stock returns constructed by compounding daily returns from the Center for Research in Security Prices (CRSP) daily NYSE stock files. Weekly returns were used to minimize the asynchronous trading and bid–ask bounce problems in daily data. In addition, there are variance shifts due to weekends, holidays, and day-of-the-week effects that are tangential to the main point of the analysis. By using weekly returns, we hope to minimize these problems while still retaining much of the information on short-term volatility shifts that is present in the daily data.

We first prefilter the returns series to take out an AR(1) term and monthly systematic shifts in the mean returns, $R_t^* = R_t - a - bR_{t-1} - \Sigma d_i Mnth(i)_t$. These residuals are then analyzed via the Markov-chain algorithm outlined in Section 1. If desired, one could accommodate these systematic shifts in the mean of the series in the Markov-chain method by allowing $y_t = x_t'\gamma + \sqrt{h_t}\varepsilon_t$ and conducting an analysis of the joint posterior of $p(\gamma, h, \omega \mid y)$. This is easily achieved by introducing the conditional posterior distribution of $\gamma \mid h, \omega$, and y, which can be computed from the standard theory of linear models. although it is straightforward to allow a linear mean function, we expect that information about the mean is approximately independent of information regarding the variance parameters.

Table 1 presents estimates of the stochastic volatility parameters for the standard CRSP equal-weighted (EW) and value-weighted (VW) indices of NYSE stocks as well as three decile portfolios corresponding to the 1, 5, and 10 deciles of stock as measured by market capitalization. The posterior mean of δ lies between .91 and .95, exhibiting a high persistence in conditional variances typical of estimates in the ARCH literature. Below the posterior mean and standard deviation is the 95% Bayesian credibility interval that is constructed from the .025 and .975 percentiles of the simulated posterior distribution. It is hard to directly interpret the posterior distribution of σ_v. A parameter that is more interpretable is the coefficient of variation (we use the square) of the volatility process, var $(h)/E[h]^2 = \exp(\sigma_v/(1-\delta^2)) - 1$. Since the MC sampler provides draws of the parameters, it is a straightforward computational task to compute the posterior distribution of any function of the parameters. We do not have to rely on delta-method asymptotics. Examination of the posterior distribution of the coefficient of variation suggests that the small stocks (which drive the EW index) have greater variability in the stochastic volatility equation than for the high-capitalization stocks that greatly influence the VW index.

Table 1. Posterior Analysis for Selected Portfolios

	Posterior means (standard deviation)				
Parameter	*EW*	*VW*	*D1*	*D5*	*D10*[a]
α	−.69	−.39	−.56	−.71	−.56
	(.12)	(.11)	(.12)	(.36)	(.18)
δ	−.91	−.95	−.93	−.91	−.93
	(.015)	(.013)	(.016)	(.046)	(.022)
	[.88, .94]	[.92, .97]	[.89, .96]	[.81, .96]	[.89, .97]
σ_v	.39	.23	.32	.32	.29
	(.025)	(.026)	(.032)	(.095)	(.056)
V_h/E_h^2	1.61	.80	1.1	.92	.93
	(.38)	(.24)	(.28)	(.27)	(.25)
Shock	7.8	14.	9.6	9.4	11.
half-life	(1.4)	(4.3)	(2.3)	(4.5)	(4.5)

NOTE: Weekly returns, 7/62–12/91. T = 1,540. All returns are prefiltered to remove AR (1) + monthly seasonals from mean equation.
[a]EW ≡ equal-weighted NYSE; VW ≡ value-weighted; *Dx* is decile *x* portfolio. D1 is the decile of smallest stocks.
[b]Brackets denote 95% Bayes probability interval.

Analysis of the posterior distribution of model parameters for stocks selected from the 1st and 5th decile is given in Table 2. As might be expected, the individual stocks have greater variation in the level of persistence as measured by widely varying posterior distributions of δ and σ_v. For this small sample of stocks, there seems to be a relationship between the level of market capitalization and the level of predictability in the stochastic variance equation. All three middle-decile stocks have posterior distributions massed on higher levels of the coefficient of variation than the high-cap stocks. One interesting measure is the half-life of a shock to volatility, $\log(.5)/\log(\delta)$, the posterior moments of which are presented in the last row of the table. These measures vary widely from about 3 weeks to over 26 weeks. It is also interesting to observe that, although the posterior distribution of δ is tightly massed, the posterior distribution of the half-life is very diffuse.

The adequacy of asymptotic normal approximations to the posterior is addressed in Figure 2, which shows the marginal posteriors of δ, var$(h)/E[h]^2$, and the half-life for the EW index and Texaco. All of these distributions show pronounced skewness. This casts some doubt on the usefulness of asymptotic approximations for conducting either Bayesian or frequentist inference.

We computed MM estimates of the model parameters using the set of moments outlined in Section 4. We do not report the MM estimates; instead we summarize the qualitative differences. Some of the MM estimates of δ are very different

Table 2. Posterior Analysis for Selected Stocks

	Posterior means (standard deviation)					
Parameter	*IBM*	*GM*	*TEX*	*Broad*	*Commetco*	*Bearings*[a]
α	−1.2	−.22	−.38	−.39	−.78	−.12
	(.20)	(.079)	(.10)	(.093)	(.13)	(.16)
δ	.83	.97	.95	.93	.87	.81
	(.028)	(.011)	(.014)	(.017)	(.021)	(.024)
	[.77, .88][b]	[.94, .99]	[.92, .97]	[.89, .96]	[.83, .91]	[.76,.86]
σ_v	.40	.14	.23	.25	.43	.53
	(.025)	(.023)	(.028)	(.029)	(.027)	(.028)
V_h/E_h^2	.67	.43	.73	.56	1.3	1.35
	(.12)	(.52)	(.21)	(.15)	(.26)	(.22)
Shock	3.8	26.0	14.	9.7	5.3	3.4
half-life	(.69)	(15.)	(4.1)	(2.5)	(1.0)	(.49)

NOTE: Weekly returns, 7/62–12/91. T = 1540. All returns are prefiltered to remove AR(1) and monthly seasonals from the mean equation.

[a]Tex ≡ Texaco; Broad ≡ Broad Inc.; Commetco ≡ Commercial Metals Company; Bearings ≡ Bearing Inc. Stocks are listed from largest to smallest capitalization in terms of 1991 market value. The 1991 market capitalizations are $50.8 billion, $17.8 billion, $15.8 billion, $621 million, $230 million, and $137.8 million, respectively.

[b]Brackets denote 95% Bayes probability interval.

from the Bayes estimates; for example, the MM estimate of δ for IBM is .996 (compare with .83 from the MC sampler). In most cases, the asymptotic standard errors computed for the MM estimates are larger than the posterior standard deviations reported in Tables 1 and 2. Although the posterior standard deviations and MM asymptotic errors are not strictly

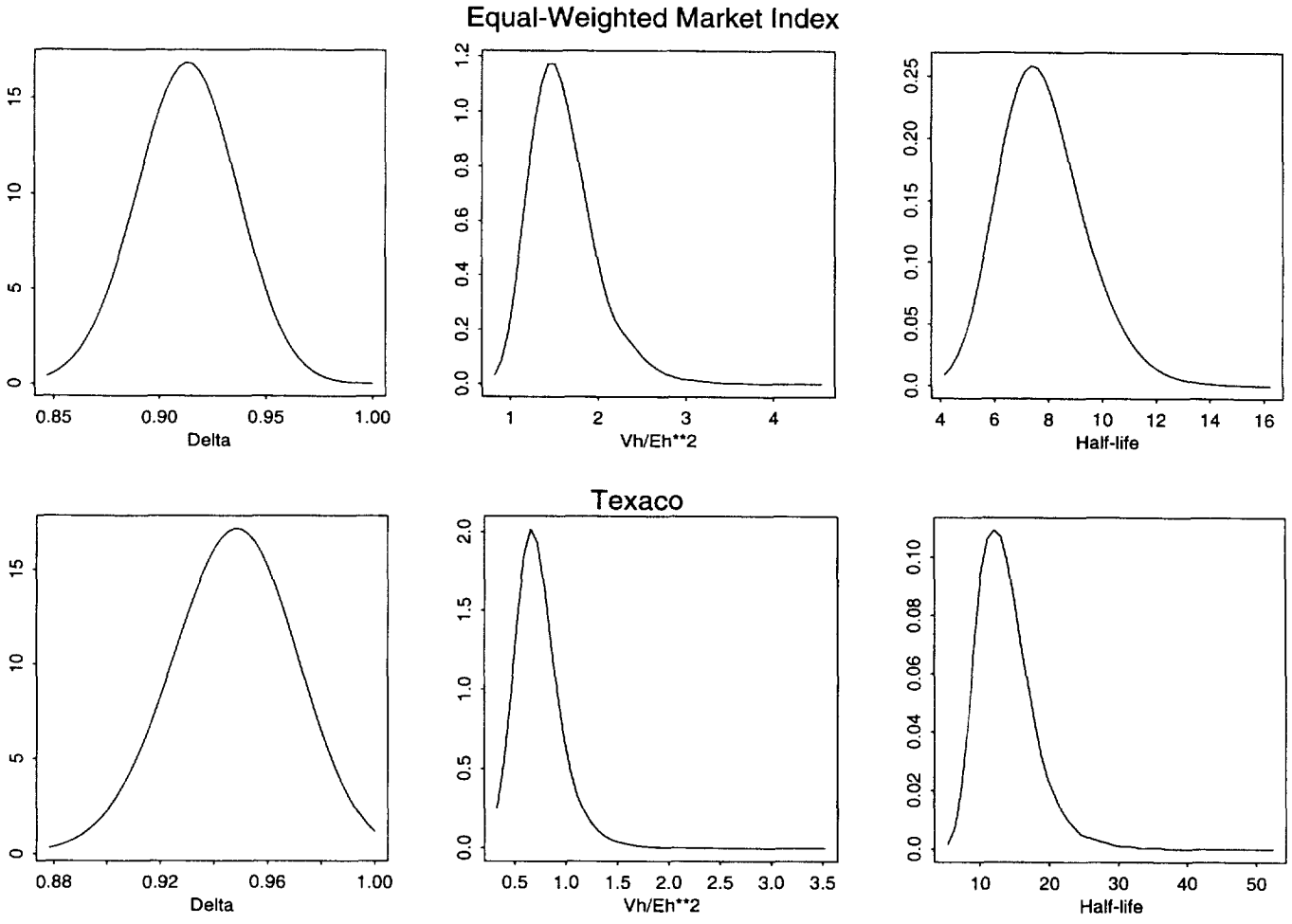

Figure 2. Posterior Distributions of Selected Parameters.

comparable, this does suggest that there is important information missing in the set of moments used in computing the MM estimates. We also found evidence of multiple interior local optima in the MM criterion surface. For example, in the analysis of the Commetco data, we found two interior optima, one at ($\widehat{\alpha} = -5.5$, $\widehat{\delta} = .12$, $\widehat{\sigma}_v = .83$) with an MM criterion value of 20.4 and another at the point (−.129, .98, .119) with value 19.3.

In Table 3, we present a Bayesian analysis of daily time series that have received attention in the stochastic volatility literature. We examine the daily changes in the log of the S&P 500 stock-price index studied by Gallant, Rossi, and Tauchen (1992) and daily spot exchange rates for the £/$, DM/$, and CD/$ series. The S&P 500 series is prefiltered to remove systematic components from the mean and variance as discussed by Gallant et al. (1992). The £/$ and DM/$ series are noon spot prices from the Board of Governors of the Federal Reserve System, supplied by David Hsieh. The CD/$ series was obtained from Melino and Turnbull (1990). The Bayes estimates of δ for the CD/$ exchange rate agree closely with MM estimates computed by Melino and Turnbull. Our estimates of the coefficient of variation, however, are 10 times larger (1.12 vs. .14). For the DM/$ and £/$ rates, Harvey et al. (1994) found much higher δ values for the £/$ series. It should be pointed out that the Harvey et al. study was conducted on a subperiod of our sample and with data from a different source. The exchange-rate data exhibit a high degree of persistence in volatility although the posterior is massed well away from the unit-root case. Some care should be exercised in interpreting these persistence findings because level shifts in these series brought about by calendar effects and central bank interventions [especially for the U.K. series; see Gallant et al. (1991)] can easily lead to spurious unit-root findings. Our analysis of the S&P series finds a higher level of persistence for the daily series as compared to the weekly stock portfolios, a fact frequently noted in the ARCH literature. The posterior distribution of δ, however, is massed well away from 1 (the posterior probability that $\delta > .99 = .005$).

Taken as a whole, these findings suggest that there may be important differences in the sampling properties of Bayes and other estimation procedures. For this reason, we designed and carried out a set of simulation experiments comparing the sampling performance of the posterior means as Bayes estimators as compared to MM estimators and estimators based on the QML filtering approach of Nelson (1988) and Harvey et al. (1994). These sampling experiments are reported in Section 4.

Table 3. Posterior Analysis for Selected Daily Financial Series

	Posterior means (standard deviation)			
Parameter	*S&P 500*	*£/$*	*DM/$*	*CD$/$*
α	−.002	−.36	−.56	−.61
	(.004)	(.12)	(.012)	(.12)
δ	.97	.96	.95	.95
	(.008)	(.012)	(.013)	(.009)
	[.96, .99]*	[.94, .98]	[.92, .97]	[.93, .97]
σ_v	.15	.17	.13	.26
	(.017)	(.03)	(.03)	(.025)
V_h/E_h^2	.56	.52	.69	1.1
	(.19)	(.12)	(.14)	(.23)
Shock	29.0	21.6	13.	14.4
half-life	(10.0)	(8.8)	(3.5)	(3.0)

NOTE: S&P 500-daily change in log (S&P 500) index, filtered to remove calendar effects as documented by Gallant, Rossi, and Tauchen (1992); 1/2/80–12/30/87; T = 2,023. £/$, DM/$-daily noon spot rates from the Board of Governors of the Federal Reserve System, supplied by David Hsieh; 1/2/80–5/31/90; T = 2,614. CD$/$-daily noon interbank market spot rates from Bank of Canada, supplied by Melino and Turnbull (1990); 1/2/75–12/10 86, T = 3,011.

* [,] denotes 95% Bayes probability interval.

4. SAMPLING EXPERIMENTS

To compare our Bayes estimator to other procedures, we devised a series of simulation experiments designed to gauge performance over a fairly wide range of relevant parameter values. Table 4 summarizes the six sets of parameter settings used in the experiments. As discussed previously we have found it most convenient to parameterize the model in terms of the coefficient of variation of h. Our empirical analysis of stock data and exchange-rate data suggests that δ values between .9 and .98 are relevant and that the coefficient of variation fluctuates between .5 and 1.0. To examine the sensitivity of the methods to the size of the stochastic component in the volatility equation, we consider cells symmetrically positioned around the central $\text{var}(h)/E[h]^2 = 1.0$ cell. All experiments are calibrated so that $E[h] = .0009$. If we think of the simulated data as weekly returns, this implies an approximately 20% annual standard deviation, which is typical of many stocks. We consider sample sizes of $T = 500$ for all six cells, and we run one set of simulations at $T = 2{,}000$ for the central cells to confirm that our findings are consistent across sample sizes. The same set of simulated data sets are used for all three estimators considered.

A review of the findings in the stochastic volatility literature supports our choice of parameter values. All studies find a coefficient of variation of less than 1.0 with δ ranging from around .8 to .995. Melino and Turnbull (1990) analyzed the CD/$ exchange rate (T = 3,011, daily from 1975–1986) and reported $\widehat{\delta} = .91$ and $V_h/E_h^2 = .14$. Danielsson (in press) studied the daily S&P 500 (T = 2,202, 1980–1987) and re-

Table 4. Sampling Experiment Parameter Values

		δ		
Var(h)/E[h]²		*.9*	*.95*	*.98*
10	α	−.821	−.4106	−.1642
	σ_v	.675	.4835	.3082
1	α	−.736	−.368	−.1472
	σ_v	.3629	.26	.1657
.1	α	−.7061	−.353	−.1412
	σ_v	.135	.0964	.0614

NOTE: All of the models are calibrated so that E_h = .0009—that is, about 22% annual standard deviation for weekly data.

ported $\widehat{\delta} = .96$ and $V_h/E_h^2 = .34$. Vetzal (1992) used weekly T-bill data and got $\widehat{\delta} = .94$ and $V_h/E_h^2 = .53$. Harvey et al. (1994) studied daily exchange rates from 1981 to 1985 ($T = 946$) and reported δ estimates from .958 to .995 with V_h/E_h^2 ranging from .47 to .74.

Two methods that have received prominent attention in the literature on estimating stochastic volatility models are the MM and the quasi-likelihood Kalman filtering (QML) approach. Both the MM and QML methods are nonlikelihood-based methods that are less efficient than our likelihood-based Bayes method. The purpose of the simulation experiments is to measure the extent of this inefficiency. In the recent generalized MM (GMM) literature, there has been considerable discussion of an efficiency/robustness trade-off in which one might be willing to use inefficient methods in exchange for robustness with respect to departures from the distributional assumptions used in formulating the likelihood. The MM approach considered here, however, uses higher-order moment conditions that depend critically on normality assumptions.

We do not consider the simulated likelihood method (QAM) of Danielsson and Richard (1992) in our simulation experiments for practical computational reasons. Using the number of draws reported by Danielsson (in press), we conservatively estimate that it would take at least three Sparc 2 workstation CPU years to run our full set of simulation experiments. Further refinements of the Danielsson and Richard method may make a full-blown simulation study feasible, and we leave this for future study.

The simulation studies allow for comparison of various methods not only in parameter estimation but also in producing "smoothed" estimates of the unobserved variances. Although our MCMC Bayes procedure yields solutions to the smoothing problem as a natural by-product of the method, the MM, QML, and QAM approaches offer no direct solution to the filtering problem. In the non-Bayesian approach, a method of parameter estimation must be coupled with a nonlinear filtering method to solve the smoothing problem. In the simulation experiments, standard approximate Kalman-filtering methods are compared to the Bayes solution to the filtering problem.

4.1 Method of Moments

For simple stochastic volatility models such as the one considered here, analytic expressions are available for a large collection of different moments. For more complicated models in which analytic moment expressions are not available, a simulated MM methodology could be employed. The real problem in implementing an MM approach is the choice of moments. Although many collections of these moments are sufficient to identify the stochastic volatility parameters, there is always the question of efficiency loss due to excluding information in the likelihood. The score function of the log-likelihood should suggest which moments should be used. Since the score is not available for these problems, we must guess which set of moments is "sensible." For extremely large samples, we might specify moments from some set of basis functions, which would, one hopes, approximate any score function. A very large number of basis functions and lags might be required, however.

We select the same basic set of moments considered by Melino and Turnbull (1990). We use $\{E[|y_t|^k], k = 1, \ldots, K\}$, $\{E[|y_t y_{t-k}|], k = 1, \ldots, L_a\}$, and $\{E[y_t^2 y_{t-k}^2], k = 1, \ldots, L\}$. The moments of marginal distribution of y_t primarily serve to identify the mean of h, and the autocovariances of the squares and absolute values help to identify δ. As will be discussed, we experimented with the number of lags in the moments of the autocovariance type. Appendix A presents the well-known analytic expressions for these moments as well as the derivatives of the moment expressions with respect to the stochastic volatility parameters. These derivatives are used in optimization problems as well as in the computation of the asymptotic covariance matrix of the method of moments estimator.

The MM estimator is based on the classical MM approach, except that the non-iid character of the moment discrepancies is taken into account in forming the weighting matrix, as in the work of Hansen (1982). Using the notation in Appendix A, define $m_{ij} = \Sigma f_{ijt}(y)/T$, where m_{ij} indicates moment of type i and power/order j. For example, $m_{11} = \Sigma|y_t|/T$ with $f_{ijt}(y) = |y_t|$. Let $m_{ij}(\omega)$ be the population moment expression. The MM estimator is defined as $\widehat{\omega}_{\mathrm{MM}} = \arg\min g'(\omega) W g(\omega)$, where g is the vector of moment deviations and $g = \Sigma g_t/T$, $g_t = [m_{ijt} - m_{ij}(\omega)]$, an $m \times 1$ vector, where m is the number of moments used. W is the weighting matrix standard in the GMM literature, $\widehat{W}^{-1} = 1/T\Sigma_k \Sigma_t w(k, L) g_t g_{t+k}'$, and $w(k, L)$ are the Bartlett weights for up to order K. The MM estimates are based on the first four moments of $|y_t|$, 10 lags of $E[y_t^2 y_{t-k}^2]$, and 10 lags of $E[|y_t||y_{t-k}|]$. Although there is some correlation among the 24 moment discrepancies, there is no evidence of singularity in the weighting matrix.

It is common practice to iterate the computation of the weighting matrix by starting with the identity matrix and then computing new weighting matrices based on the current parameter estimates. Some iterate this process until it has converged in the sense of producing small changes in the criterion function and parameter estimates from iteration to iteration. We found that estimates converge after only a few iterations of the weighting matrix so that weighting matrix issues are less important for this class of problems. In forming the MM estimator, we used five iterations of the weight matrix for $T = 500$ and four for $T = 2{,}000$.

As is standard in this literature, the asymptotic variance matrix of $\widehat{\omega}_{\mathrm{MM}}$ is computed as $(D'W^{-1}D)^{-1}$. D is the $3 \times m$ Jacobian matrix of the moment expressions. For the simple stochastic volatility model, it should be possible to calculate W exactly rather than using the standard consistent estimator, although this would be a tedious calculation. More important, we would have to evaluate the exact expression for W at the MM estimates. It is not clear which of these consistent variance estimators should be superior, the standard spectral estimation approach or evaluating the exact W at $\widehat{\omega}_{\mathrm{MM}}$.

To compute the MM estimates, we employ a state-of-the-

Table 5. Mean and Root Mean Squared Error of the MM Estimator

$Var(h)/E[h]^2$	α	δ	σ_v	α	δ	σ_v	α	δ	σ_v
10	*−.821*	*.9*	*.675*	*−.4106*	*.95*	*.4835*	*−.1642*	*.98*	*.308*
	−1.47	.83	.59	−.72	.916	.42	−.09	.99	.19
	(1.46)	(.17)	(.23)	(.61)	(.07)	(.18)	(.18)	(.02)	(.14)
1	*−.736*	*.9*	*.363*	*−.368*	*.95*	*.26*	*−.1472*	*.98*	*.1657*
	−1.0	.87	.24	−.51	.93	.16	−.128	.982	.104
	(1.25)	(.17)	(.19)	(.83)	(.11)	(.16)	(.26)	(.036)	(.11)
.1	*−.706*	*.9*	*.135*	*−.353*	*.95*	*.0964*	*−.1412*	*.98*	*.0614*
	−5.0	.30	.06	−4.7	.34	.05	−2.25	.65	.063
	(6.3)	(.88)	(.12)	(6.5)	(.9)	(.09)	(4.7)	(.67)	(.090)

NOTE: The statistics in this table are based on 500 simulated samples, each consisting of a time series of length 500. For each cell, the top row (italics) shows the true value of the parameters. The following two rows show the mean and root mean squared error (in parentheses).

art optimizer specifically designed for a sum-of-squares objective function. We use the NPSOL routine from Stanford Optimization Laboratory as implemented in NAG library routine E04UPF. Analytic derivatives are used throughout. Inspection of the moment expressions (derived under the assumption of stationarity) immediately suggests that numerical problems can occur for values of delta close to 1 as well as for large values of σ_v (the moments contain the expression $\exp\{\sigma_v^2/(1-\delta^2)\}$). Furthermore it is well known that σ_v is an absorbing state in the sense that once $\sigma_v = 0$ the MM estimation procedure will just try to match the mean volatility. Our approach to this problem is to place bounds on the parameter space to keep δ from approaching too close to 1 and to keep sigma from getting near 0. Since our optimizer is based on a sequence of quadratic programs that approximate the augmented Lagrangian problem, it is a simple matter to add bounds constraints. Typically, we require $|\widehat{\delta}_{\mathrm{MM}}| < .995$ and $\widehat{\sigma}_v > .001$. These bounds are very far from the parameter values used in the simulations. Another possible solution to this problem would be to work with moments of the log series such as $\log|y_t|$ and $\log(|y_t y_{t-k}|)$. Although this strategy will reduce the overflow problems, it will not eliminate them.

For many simulated data sets, the configuration of sample moments implies a minimum to the MM criterion at a point outside the stationary region. In our procedure, the optimizer will hit the bounds imposed on the problem. There is a serious problem with corner solutions for any sample size in cells with high δ and low coefficient of variation. At $T = 500$, over half of the simulated samples result in corner solutions for $\delta = .95$ or .98 and $\mathrm{var}(h)/E[h]^2 = .1$. Both δ and σ_v bounds may be binding, although never at the same time. Even for the middle row of the design, $\mathrm{var}(h)/E[h]^2 = 1.0$, there are larger numbers of δ estimates at the bounds. Obviously, this problem will only worsen for data generated with models showing higher persistence. These problems with corner solutions are less severe but still persist in samples of size 2,000, but they eventually disappear if we examine extremely large samples of 32,000.

Given the large number of corner solutions, it is all the more important to examine the sensitivity of the MM estimates to the starting point for the optimizer. For all simulated samples in the experiment, we started from the true parameter vector. One view is that this gives the MM the "best chance" to find the optimum that we expected to be fairly close to the true parameter values. This argument only holds if there is very little sampling variability. It is possible that the MM criterion surface has multiple local minima. To check this, we took a few samples of 2,000 observations and defined a fine grid over the bounded Ω parameter space. The optimizer was then started from every point on this grid. We could find no evidence of more than one interior minimum. From some starting points, however, the optimizer would stop at the bounds with valid Kuhn–Tucker conditions. From still other starting points and for the same data set, the optimizer would converge at an interior optimum with lower criterion value. Finally, for some data sets, we find the minimum of the criterion at the bounds. Thus our experience to date does not rule out the possibility of either a nonconvex criterion surface or the presence of global optima outside the stationary region.

Table 5 shows the absolute sampling performance of the MM estimator for $T = 500$. The table shows the mean and root mean squared error (RMSE) for each of the six cells. The MM estimator exhibits substantial bias, especially for the estimates of σ_v. In addition, there is very substantial sampling variability that results in RMSE's as large as the true parameter value for both δ and σ_v. It should be emphasized that the actual MM estimator investigated in Table 4 uses the strong prior information in the assumption of stationarity to attenuate the sampling variability by imposing stationarity bounds.

To assess the adequacy of the standard asymptotic approximations to the sampling distribution of the MM estimator, we computed standardized values of the MM estimates—for example, $(\widehat{\delta}_{\mathrm{MM}} - \delta)/\widehat{\sigma}$, where $\widehat{\sigma}$ is the asymptotic standard error. Figure 3 shows the sampling distribution of the standardized MM estimates of δ and σ_v along with the $N(0,1)$ asymptotic distribution. For $T = 500$, there is a substantial difference between the asymptotic and exact sampling distributions. For σ_v, the substantial bias of the MM estimate is evident even for $T = 2{,}000$.

It is also useful to look at the correlations between the parameter estimates. For all cells, $\widehat{\alpha}$ and $\widehat{\delta}$ are extremely highly

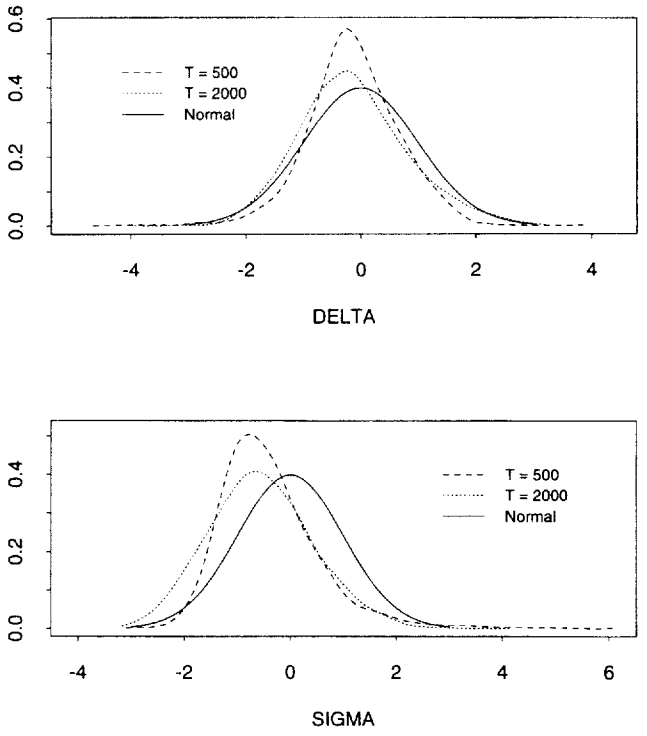

*Figure 3. Standardized MM Parameter Estimates Var(h)/E[h]**2 = 1, Delta = .9: ---, T = 500; -----, T = 2,000; ———, Normal.*

correlated (over .99). for high δ and low variability, there can also be high ($\approx -.95$) correlations between $\widehat{\sigma}_v$ and $\widehat{\delta}$. Although α and δ are separately identified, the high correlation indicates that, in practice, these parameters are underidentified. It is possible that this high correlation induces a ridge (actually a valley) in the MM criterion surface that could cause problems for the optimizer. In an effort to regularize the surface, we experimented with reparameterizing to $(\alpha/(1-\delta), \delta, \sigma_v)$. Although this new parameterization eliminates much of the correlation between the parameters, the sampling distribution of $\widehat{\delta}$ computed from this parameterization is indistinguishable from the distribution resulting from the standard $(\alpha, \delta, \sigma_v)$ parameterization.

Some might suggest that the poor performance of the MM estimator might be due to inclusion of insufficient lags in the autocovariance-type moments. We used 10 lags of both the autocorrelations in the squares and absolute values. Experimentation with a larger number of lags did not affect the sampling performance. It appears that the problem with the MM is the selection of the form of the moments and not the number of lags.

4.2 The Quasi-Maximum Likelihood Estimator

Nelson (1988) and Harvey et al. (1994) suggested a QML method that is based on a Kalman-filtering approach. The basic stochastic volatility model can be written as a system of two linear equations:

$$\log(y_t^2) = \log(h_t) + \log(\varepsilon_t^2)$$
$$\log(h_t) = \alpha + \delta \log(h_{t-1}) + \sigma_v v_t,$$
$$(\varepsilon_t, v_t) \sim \text{independent } N(0, 1).$$

This is a linear dynamic model in the standard-measurement-equation/state-evolution-equation form. If $\eta_t = \log(\varepsilon_t^2)$ were normally distributed, standard linear filtering theory could be used to evaluate the likelihood and solve the prediction/smoothing problem (see Appendix B for a complete discussion and review of the filtering approach). Since η_t is not normally distributed, linear filtering methods can only approximate the results of the true nonlinear optimal filter. As Harvey et al. (1994) and Ruiz (1994) pointed out, the adequacy of the approximation depends critically on the value of σ_v. For large values of σ_v, the systematic component $[\log(h_t)]$ in the measurement equation will dominate the η_t error term, the normality approximation may be adequate, and the linear filtering approach will be close to optimal. For small values of σ_v, however, the normality approximate will break down, and a linear filtering approach may produce estimates with poor sampling properties relative to a likelihood-based procedure. The real advantage of the QML approach is its speed and adaptibility to many situations. The very fact that we are able to conduct large-scale sampling experiments with our Bayesian methods suggests that computational speed is not an important advantage for the QML procedure in this context.

Table 6. Mean and Root Mean Squared Error of the QML Estimator

$Var(h)/E[h]^2$	α	δ	σ_v	α	δ	σ_v	α	δ	σ_v
10	*−.821*	*.9*	*.675*	*−.4106*	*.95*	*.4835*	*−.1642*	*.98*	*.308*
	−.99	.88	.70	−.55	.93	.51	−.11	.99	.33
	(.48)	(.06)	(.16)	(.32)	(.04)	(.12)	(.09)	(.01)	(.07)
1	*−.736*	*.9*	*.363*	*−.368*	*.95*	*.26*	*−.1472*	*.98*	*.1657*
	−1.4	.81	.45	−1.0	.86	.35	−.20	.97	.22
	(1.6)	(.22)	(.27)	(1.7)	(.23)	(.25)	(.54)	(.08)	(.15)
.1	*−.706*	*.9*	*.135*	*−.353*	*.95*	*.0964*	*−.1412*	*.98*	*.0614*
	−5.5	.23	.33	−5.5	.22	.31	−3.5	.49	.35
	(5.6)	(.79)	(.39)	(6.0)	(.85)	(.41)	(4.6)	(.67)	(.46)

NOTE: The statistics in this table are based on 500 simulated samples, each consisting of a time series of length 500. For each cell, the top row (italics) shows the true value of the parameters. The following two rows show the mean and root mean squared error (in parentheses).

Table 7. Mean and Root Mean Squared Error of the Bayes Estimator

$Var(h)/E[h]^2$	α	δ	σ_v	α	δ	σ_v	α	δ	σ_v
10	*−.821*	*.9*	*.675*	*−.4106*	*.95*	*.4835*	*−.1642*	*.98*	*.308*
	−.679	.916	.562	−.464	.94	.46	−.19	.98	.35
	(.22)	(.026)	(.12)	(.16)	(.02)	(.055)	(.08)	(.01)	(.06)
1	*−.736*	*.9*	*.363*	*−.368*	*.95*	*.26*	*−.1472*	*.98*	*.166*
	−.87	.88	.35	−.56	.92	.28	−.22	.97	.23
	(.34)	(.046)	(.067)	(.34)	(.046)	(.065)	(.14)	(.02)	(.08)
.1	*−.706*	*.9*	*.135*	*−.353*	*.95*	*.0964*	*−.141*	*.98*	*.0614*
	−1.54	.78	.15	−1.12	.84	.12	−.66	.91	.14
	(1.35)	(.19)	(.082)	(1.15)	(.16)	(.074)	(.83)	(.12)	(.099)

NOTE: The statistics in this table are based on 500 simulated samples, each consisting of a time series of length 500. For each cell, the top row (italics) shows the true value of the parameters. The following two rows show the mean and root mean squared error (in parentheses).

Table 6 presents the sampling performance of the QML estimator for our experimental design. The sampling experiment confirms our intuition that the performance of the QML estimator is best for large values of σ_v. For a coefficient of variation of 10, the QML estimator exhibits little bias and small variability. For smaller values of the coefficient of variation, however, the performance of both of the QML and MM estimators deteriorates rapidly.

It is interesting to compare the relative performance of the MM and QML estimators. The overall impression is that the QML and MM estimators have similar performance with the QML dominating the MM estimator only for the high volatility cells ($\text{var}(h)/E[h]^2 = 10.0$). Ruiz (in press) compared the MM and QML approach on the basis of relative *asymptotic* efficiency and found that the QML approach has very high relative efficiency as compared to the MM estimator for large σ_v values. For smaller σ_v values, the QML is dominated by the MM as measured by asymptotic relative efficiency. Our finite-sample results suggest that the performance of QML for high σ_v cells is only marginally superior to the MM. It should be emphasized that we do not consider exactly the same set of parameter values. Ruiz considered values of $\sigma_v = 1$ that imply coefficients of variation still higher than ours. We know of no estimates from a stochastic volatility model with a coefficient of variation even as high as 10, as considered in our high-volatility cells. Most studies, including our own, find coefficients of variation between .1 and 2 or so. Finally, Ruiz pointed out that the asymptotic variance estimates dramatically understate the true sampling variability. This suggests that there may be large differences between actual sampling performance and asymptotic relative efficiency. In addition, the asymptotic relative efficiency measure does not consider the substantial biases that we find for both the MM and QML estimators.

4.3 Bayes Estimators

Table 7 presents the sampling performance of the Bayes estimator. The Bayes estimator is constructed using a draw sequence of length 4,000 from the Markov chain constructed according to the algorithm outlined in Section 1. To assure convergence to the stationary distribution, the first 1,500 draws are discarded, leaving a sequence of length 2,500 to compute an estimate of the posterior mean. To simulate a sequence of this length takes approximately 14 minutes of Sparc 10 CPU time.

In the literature on MCMC (e.g., see McCulloch and Rossi in press), there is extensive discussion of the dependence of results on initial conditions. We experimented with different starting points [the pair (ω, h)] and found no discernible differences between starting points. Examination of draw sequences from the Markov chain shows that the effect of initial conditions is rapidly dissipated in the first hundred or fewer draws. In all of the simulation experiments, we start from the point ($\omega = (-.5, .5, .1)$, and $h_t = y_t^2$). We employ extremely diffuse but proper priors centered over the starting point for ω. Our prior standard deviations exceed the posterior standard deviations by a factor of at least 100.

The sampling performance of the Bayes estimator is dramatically superior to that of either the MM or QML estimators. In each of the six cells, both the bias and sampling variance of the Bayes estimator is smaller than either the MM or QML estimator. To summarize the relative performance of the Bayes estimator across experimental cells, we describe the distribution of the ratio of RMSE in Table 8. These summary statistics show that, for the parameter settings considered in the sampling experiments, the Bayes estimator dominates both the MM and QML procedures with RMSE's of less than one-half of these other procedures. This superior sampling performance also demonstrates that the convergence of our Markov chain is rapid and reliable.

Table 8. The Distribution of the Ratio of RMSE

Parameter	Relative RMSE	Median	Range
α	MM: Bayes	3.8	[1.9, 6.6]
	QML: Bayes	4.1	[1.1, 5.5]
δ	MM: Bayes	3.7	[1.8, 6.5]
	QML: Bayes	4.1	[1.1, 5.6]
σ_v	MM: Bayes	1.9	[.95, 3.3]
	QML: Bayes	3.8	[1.2, 5.5]

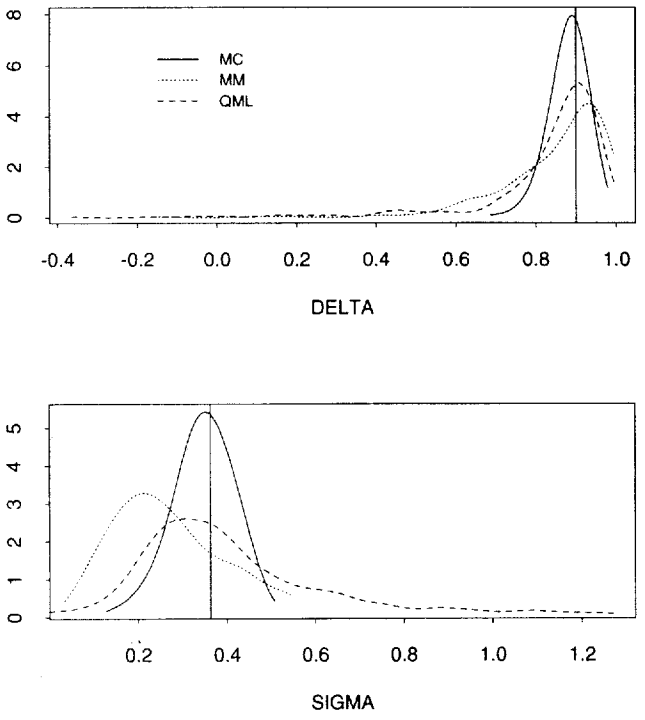

*Figure 4. Density of MC Posterior Means, QML, and MM Estimates, T = 500, Var(h)/E[h]** = 1, Delta = .9: ———, Bayes; -----, MM; – – –, QML.*

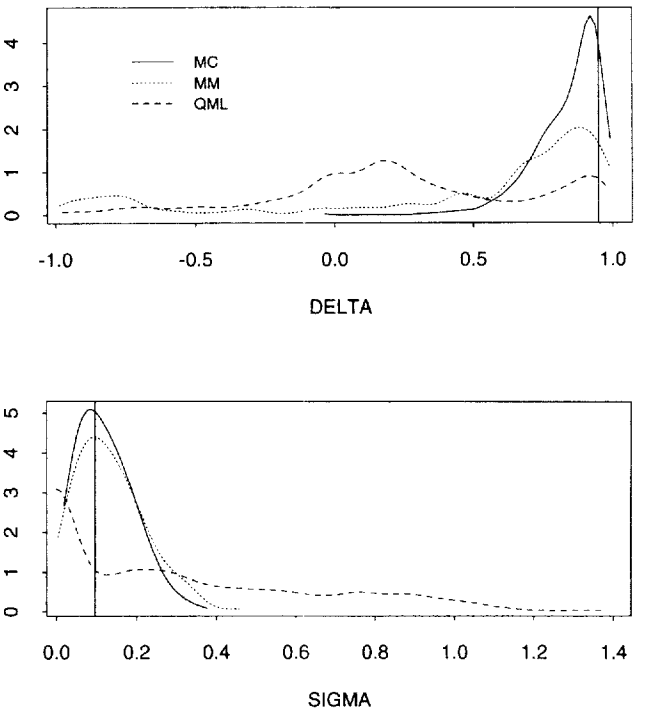

*Figure 5. Density of MC Posterior Means, QML, and MM Estimates, T = 500, Var(h)/E[h]**^{2} = 1,Delta = .95: ———, Bayes; -----, MM; – – –, QML.*

The only cell in which the QML has a sampling performance near to the Bayes estimator is the $\delta = .98$ and $V_h/E_h^2 = 10$ cell and, in this cell, only for estimation of α and δ. We know of no data set that appears to have come from this region of the parameter space. The MM estimator never comes near to the performance of the Bayes estimator for α and δ, but it does perform slightly better in the estimation of σ_v for the high persistence, very low-volatility cell ($\delta = .98, V_h/E_h^2 = .1$). This performance may be somewhat deceptive, however, because over two-thirds of the samples in this cell have an MM surface that gives an optimum at the lower bound for $\sigma_v(.001)$. In effect, this winsorizes the MM estimator to a very favorable, in terms of RMSE, parameter estimate. If all MM estimates were truncated at the lower bound, the RMSE would be .06, even lower than the actual performance.

Figures 4 and 5 present the sampling distribution of MM, QML, and Bayes estimators for two cells in the experimental design. The substantial biases in the MM estimates of σ_v are evident in both figures. In addition, the Bayes estimator shows a much tighter (but nonnormal with substantial skewness) sampling distribution than either the QML or MM estimators. Figure 5 illustrates how dramatically the performance of both the MM and QML estimators degrades in situations with high persistence and low volatility in the variance equation. This situation is typical in financial data that display highly predictable short-term volatility.

Table 9 examines one of the central cells in the design for $T = 2{,}000$. As should be expected, the relative performance of the Bayes estimator is similar to that in the small sample with $T = 500$. For $T = 2{,}000$, there is little bias for any of the estimators; the superiority of the Bayes estimator ($\frac{1}{3}$ of RMSE of others) is achieved by dramatically lower sampling variability.

4.4 Refinements of the MM and QML Approaches

Some recent developments in the MM literature hold some promise of improving the performance of the MM estimator considered here. Andersen and Sorensen (1993) made two refinements of the basic MM procedure: (1) They employed a penalty function to avoid nonstationary parameter estimates, and (2) they used a slightly modified weighting matrix using the weights proposed by Andrews (1991). Andersen and Sorensen replicated a portion of our sampling experiment

Table 9. Comparison of Method of Moments, QML, and Bayes Estimators

	α	δ	σ_v
Method	*−.736*	*.9*	*.363*
MM	−.86	.88	.31
	(.42)	(.06)	(.10)
QML	−.853	.88	.383
	(.46)	(.06)	(.11)
Bayes	−.762	.896	.359
	(.15)	(.02)	(.034)

NOTE: The statistics in this table are based on 500 simulated samples. The table shows the mean and the root mean squared error (in parentheses). The true parameter values are in italics, $T = 2{,}000$, $\text{var}(h)/E[h]^2 = 1.0$.

design, facilitating direct comparison with our results. They achieved about a 25% reduction in RMSE over the numbers reported here, which still puts an MM procedure at a disadvantage relative to Bayes procedure for all except the $\delta = .98, V_h/E_h^2 = .1$ cell. In this cell, the Andersen and Sorensen procedure could result in a modest improvement over our current Bayes procedure.

Gallant and Tauchen (1992) suggested ways of using both parametric and nonparametric procedures to suggest which moments are important in evaluating the score function for the stochastic volatility model. Their procedure involves simulating from the stochastic volatility model at a given parameter setting and using semiparametric methods to approximate the score. Although the Gallant and Tauchen approach has some appealing asymptotic properties, its value in the estimation of stochastic volatility models remains to be determined from a thorough simulation analysis. Finally, this sort of approach may help in parameter estimation but does not help solve the nonlinear filtering and prediction problem.

Our implementation of the QML procedure assumes that it is known that the mean equation error is normal. We view this as giving the maximum possible advantage to QML procedure. It is possible, however, to make σ_η^2 an unknown parameter. Because the linear filter is an approximate method, it might be the case that by freeing this variance parameter we could achieve a better approximation. For this reason, we reran the QML sampling experiments with this parameter free in the algorithm. Overall, the results did not change much at all, with modest improvements in parameter estimation in some cells and some degradation of performance in others. Over all nine cells in the design, including σ_η^2 as a free parameter changed the RMSE for parameter estimation by no more than 15%.

Table 10. RMSE Smoothing Performance of Approximate Kalman Filtering Versus Bayes Solution

$Var(h)/E[h]^2$	Method	δ = .9	.95	.98
10	Approx. Kalman			
	True parameters	24.2	21.1	13.5
	MM estimate	25.4	22.7	16.4
	QML estimate	23.6	22.0	13.9
	Bayes solution	21.1	17.0	12.2
1	Approx. Kalman			
	True parameters	6.74	6.04	5.60
	MM estimate	7.74	7.08	6.52
	QML estimate	7.14	6.64	6.19
	Bayes solution	5.9	5.26	5.04
.1	Approx. Kalman			
	True parameters	2.6	2.5	2.28
	MM estimate	3.0	2.9	2.55
	QML estimate	3.5	3.7	4.83
	Bayes solution	2.58	2.46	2.27

NOTE: The statistics in this table are based on 500 simulated samples, each consisting of a time series with $T = 500$. RMSE $\times$ 10,000 is displayed.

Table 11. Filtering Performance

Relative RMSE	Median	Range
Approx. K filter given true parameters: Bayes	1.11	[1.004, 1.24]
Approx. K filter given MM estimates: Bayes	1.29	[1.12, 1.35]
Approx. K filter given QML estimates: Bayes	1.26	[1.12, 2.13]

5. SMOOTHING PERFORMANCE

As noted previously, one of the unique advantages of our method is that it provides an exact solution to the smoothing problem of estimating the unobserved variances. The smoothing problem is the problem of computing the posterior distribution of h_t given y, $p(h_t \mid y)$. A natural smoothed estimate would be $\widehat{h}_t = E[h_t \mid y]$, which can be computed directly (up to simulation error) using the sequence of Markov-chain draws. In the stochastic volatility literature, approximate Kalman-filtering methods are used to approximate this solution to the smoothing problem. To evaluate the smoothing performance of various procedures, we compare our Bayesian MCMC solution to the smoothing problem to approximate Kalman-filtering methods. We consider approximate Kalman-filtering (see Appendix B for details) conditional on the true parameters, MM estimates, and the QML estimates.

Table 10 summarizes the smoothing performance. To evaluate the smoothing performance, we compute the grand RMSE $= [(1/NT)\sum_i \sum_t (h_t - \widehat{h}_t)^2]^{.5}$ over all $N = 500$ samples for $t = 100, \ldots, 400$. Again we can use relative RMSE as a performance criterion for smoothing; see Table 11. It is remarkable that the Bayes smoothing solution dominates not only approximate Kalman filtering based on the MM and QML estimates but also the approximate Kalman-filtering solution using the true parameters.

As outlined in Section 4.4, we also considered the variant of QML, in which σ_η^2 is left as a free parameter. Freeing σ_η^2 unambiguously worsened the filtering performance. For high persistence and low volatility cells, use of the free σ_η^2 parameter can result in a very substantial degradation in filtering performance of up to a 400% increase in RMSE (for the $\delta = .98$, $V_h/E_h^2 = .1$ cell).

Finally, it should be emphasized that we can obtain estimates of the entire posterior distribution of $h_t \mid y$ rather than just the mean. This distribution can be used to characterize the full level of uncertainty in the smoothing estimates rather than resorting to asymptotic "plug-in" computations in which the estimation error in the parameter estimates is ignored and an asymptotic normal approximation is used.

ACKNOWLEDGMENTS

Support from the Johnson Graduate School of Management, Cornell University, the Q Group Research Fund, and Graduate School of Business, University of Chicago, is grate-

fully acknowledged. We received useful comments from Torben Andersen, Orazio Attanasio, Phil Braun, Peter Carr, Rob Engle, Dan Nelson, Rob McCulloch, George Tauchen, and Mike West. We thank David Hsieh and Angelo Melino for supplying us with exchange-rate data. We thank Ken Vetzal for supplying his moment calculations, which we used to check our derivations.

APPENDIX A: MOMENTS OF THE STOCHASTIC LOG-VOLATILITY MODEL

The model is

$$y_t = \sqrt{h_t}e_t = \sigma_t e_t, \qquad e_t \sim N(0,1)$$
$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t, \qquad v_t \sim N(0,1).$$

Then,

$$\log h_t \sim N\left(\frac{\alpha}{1-\delta}, \frac{\sigma_v^2}{1-\delta^2}\right) \sim N(\mu_h, \sigma_h^2).$$

The model in this appendix does not have mean parameters. Moreover, the correlation between e_t and v_t is assumed to be 0. For each type of moment, we write its expectation (m) and the derivative of a discrepancy function (g) with respect to each parameter. The discrepancy function is the difference between the sample moment ($\hat{m}$) and the expected value of the moment.

$E(y_t^m)$

We first compute the expectations of the moments for several values of m:

$$m_{1i} = E(y_t^i) = 0\text{: no parameter information for } i \text{ odd}$$
$$m_{12} = E(y_t^2) = E(h_t e_t^2) = E(h_t)E(e_t^2) = e^{\mu_h + 1/2\sigma_h^2}$$
$$m_{14} = E(y_t^4) = E(h_t^2 e_t^4) = E(h_t^2)E(e_t^4) = 3e^{2\mu_h + 2\sigma_h^2}$$
$$m_{16} = E(y_t^6) = E(h_t^3 e_t^6) = E(h_t^3)E(e_t^6) = 15e^{3\mu_h + 9/2\sigma_h^2}.$$

Now compute the derivatives of the discrepancy functions:

$$g_{12} = \tfrac{1}{T}\sum y_t^2 - e^{\mu_h + 1/2\sigma_h^2} \equiv \hat{m}_{12} - m_{12}$$
$$\frac{\partial g_{12}}{\partial \alpha} = -m_{12}\frac{1}{1-\delta}$$
$$\frac{\partial g_{12}}{\partial \delta} = -m_{12}\left(\frac{\alpha}{(1-\delta)^2} + \frac{\delta\sigma_v^2}{(1-\delta^2)^2}\right)$$
$$= -m_{12}\left(\frac{\mu_h}{1-\delta} + \frac{\delta\sigma_h^2}{1-\delta^2}\right)$$
$$\frac{\partial g_{12}}{\partial \sigma_v} = -m_{12}\frac{\sigma_v}{1-\delta^2};$$
$$g_{14} = \frac{1}{T}\sum y_t^4 - 3e^{2\mu_h + 2\sigma_h^2} \equiv \hat{m}_{14} - m_{14}$$
$$\frac{\partial g_{14}}{\partial \alpha} = -m_{14}\frac{2}{1-\delta}$$
$$\frac{\partial g_{14}}{\partial \delta} = -m_{14}\left(\frac{2\alpha}{(1-\delta)^2} + \frac{4\delta\sigma_v^2}{(1-\delta^2)^2}\right)$$
$$= -m_{14}\left(2\mu_h 1 - \delta + \frac{4\delta\sigma_h^2}{1-\delta^2}\right)$$
$$\frac{\partial g_{14}}{\partial \sigma_v} = -m_{14}\frac{4\sigma_v}{1-\delta^2};$$
$$g_{16} = \frac{1}{T}\sum y_t^2 - 15e^{3\mu_h + 9/2\sigma_h^2} \equiv \hat{m}_{16} - m_{1}$$
$$\frac{\partial g_{16}}{\partial \alpha} = -m_{16}\frac{3}{1-\delta}$$
$$\frac{\partial g_{16}}{\partial \delta} = -m_{16}\left(\frac{3\alpha}{(1-\delta)^2} + \frac{9\delta\sigma_v^2}{(1-\delta^2)^2}\right)$$
$$\frac{\partial g_{16}}{\partial \sigma_v} = -m_{16}\frac{9\sigma_v}{1-\delta^2}.$$

$E(|y_t^m|)$

$$m_{21} = E(|y_t|) = E(\sigma_t|e_t|) = E(\sigma_t)E(|e_t|)$$
$$= \sqrt{\frac{2}{\pi}}e^{1/2\mu_h + 1/8\sigma_h^2}$$
$$m_{23} = E(|y_t|^3) = E(\sigma_t^3|e_t^3|) = E(\sigma_t^3)E(|e_t|^3)$$
$$= 2\sqrt{\frac{2}{\pi}}e^{3/2\mu_h + 9/8\sigma_h^2};$$
$$g_{21} = \frac{1}{T}\sum |y_t| - m_{21}$$
$$\frac{\partial g_{21}}{\partial \alpha} = -m_{21}\frac{1}{2(1-\delta)}$$
$$\frac{\partial g_{21}}{\partial \delta} = -m_{21}\left(\frac{\alpha}{2(1-\delta)^2} + \frac{\delta\sigma_v^2}{4(1-\delta^2)^2}\right)$$
$$= -m_{21}\left(\frac{\mu_h}{2(1-\delta)} + \frac{\delta\sigma_h^2}{4(1-\delta^2)}\right)$$
$$\frac{\partial g_{21}}{\partial \sigma_v} = -m_{21}\frac{\sigma_v}{4(1-\delta^2)};$$
$$g_{23} = \frac{1}{T}\sum |y_t|^3 - m_{23}$$
$$\frac{\partial g_{23}}{\partial \alpha} = -m_{23}\frac{3}{2(1-\delta)}$$
$$\frac{\partial g_{23}}{\partial \delta} = -m_{23}\left(3\alpha 2(1-\delta)^2 + \frac{9\delta\sigma_v^2}{4(1-\delta^2)^2}\right)$$
$$= -m_{21}\left(\frac{3\mu_h}{2(1-\delta)} + \frac{9\delta\sigma_h^2}{4(1-\delta^2)}\right)$$
$$\frac{\partial g_{23}}{\partial \sigma_v} = -m_{23}\frac{9\sigma_v}{4(1-\delta^2)}.$$

$E(y_t^2 y_{t-i}^2)$

$$E(y_t^2 y_{t-i}^2) = E(h_t e_t^2 h_{t-i} e_{t-i}^2)$$
$$= E(h_t h_{t-i}) E(e_t^2) E(e_{t-i}^2) = E(h_t h_{t-i}).$$

Note that

$$\log h_t = \delta^i \log h_{t-i} + \sum_{\tau=0}^{i-1} (\alpha + \sigma_v v_{t-\tau})\delta^\tau$$

and

$$E(y_t^2 y_{t-i}^2) = E\left[h_{t-i}^{\delta^i+1} \exp\left\{\sum_{\tau=0}^{i-1}(\alpha + \sigma_v v_{t-\tau})\delta^\tau\right\}\right]$$
$$= E(h_{t-i}^{\delta^{i+1}}) E\left[\exp\left\{\sum_{\tau=0}^{i-1}(\alpha + \sigma_v v_{t-\tau}\delta^\tau)\right\}\right]$$
$$= \exp\left\{(1-\delta^i)\mu_h + \frac{1}{2}(1+\delta^i)^2\sigma_h^2 + \sum_{\tau=0}^{i-1}\left(\alpha\delta^\tau + \frac{1}{2}\sigma_v^2\delta^{2\tau}\right)\right\}$$
$$= \exp\left\{(1+\delta^i)\mu_h + \frac{1}{2}(1+\delta^i)^2\sigma_h^2 + \alpha\frac{1-\delta^i}{1-\delta} + \frac{1}{2}\sigma_v^2\frac{1-\delta^{2i}}{1-\delta^2}\right\}$$
$$= \exp\left\{2\mu_h + \sigma_h^2(1+\delta^i)\right\} = m_{4,i}.$$

The derivatives are

$$\frac{\partial g_{4i}}{\partial \alpha} = -m_{4i}\frac{2}{1-\delta}$$
$$\frac{\partial g_{4i}}{\partial \delta} = -m_{4i}\left(\frac{2\mu_h}{1-\delta} + \frac{\sigma_h^2}{1-\delta^2}(i\delta^{i-1} - i\delta^{i+1} + 2\delta + 2\delta^{i+1})\right)$$
$$\frac{\partial g_{4i}}{\partial \sigma_v} = -m_{4i}2\sigma_v\frac{1+\delta^i}{1-\delta^2}.$$

$E(|y_t y_{t-i}|)$

$$E(|y_t y_{t-i}|) = E(\sigma_t|e_t|\sigma_{t-i}|e_{t-i}|)$$
$$= E(\sigma_t\sigma_{t-i})E|e_t|E|e_{t-i}|$$
$$= \frac{2}{\pi}E(\sigma_t\sigma_{t-i}).$$

Note that

$$\log\sigma_t = \delta^i \log\sigma_{t-i} + \sum_{\tau=0}^{i-1}\left(\frac{\alpha}{2} + \frac{\sigma_v}{2}v_{t-\tau}\right)\delta^\tau$$

and

$$E|y_t y_{t-i}| = \frac{2}{\pi}E\left[\sigma_{t-i}^{\delta^i+1} e^{\sum_\tau^{i-1}(\alpha+\sigma_v v_{t-\tau})\frac{\delta^\tau}{2}}\right]$$
$$= \frac{2}{\pi}E(\sigma_{t-i}^{\delta^i+1})E\left[\exp\left\{\sum_\tau^{i-1}(\alpha+\sigma_v v_{t-\tau})\frac{\delta^\tau}{2}\right\}\right]$$
$$= \frac{2}{\pi}\exp\left\{(1+\delta^i)\frac{\mu_h}{2} + (1+\delta^i)^2\frac{\sigma_h^2}{8} + \sum_{\tau=0}^{i-1}\left(\frac{\alpha\delta^\tau}{2} + \frac{\sigma_v^2\delta^2\tau}{8}\right)\right\}$$
$$= \frac{2}{\pi}\exp\left\{(1+\delta^i)\frac{\mu_h}{2} + (1+\delta^i)^2\frac{\sigma_h^2}{8} + \frac{\alpha}{2}\frac{1-\delta^i}{1-\delta} + \frac{\sigma_v^2}{8}\frac{1-\delta^{2i}}{1-\delta^2}\right\}$$
$$= \frac{2}{\pi}\exp\left\{\mu_h + \frac{\sigma_h^2}{4}(1+\delta^i)\right\} \equiv m_{5i}.$$

The derivatives are

$$\frac{\partial g_{5i}}{\partial \alpha} = -m_{5i}\frac{1}{1-\delta}$$
$$\frac{\partial g_{5i}}{\partial \delta} = -m_{5i}\left(\frac{\mu_h}{1-\delta} + \frac{\sigma_h^2}{4(1-\delta^2)}(i\delta^{i-1} - i\delta^{i+1} + 2\delta + 2\delta^{i+1})\right)$$
$$\frac{\partial g_{5i}}{\partial \sigma_v} = -m_{5i}\sigma_v\frac{1+\delta^i}{2(1-\delta^2)}.$$

Corr (y_t^2, y_{t-i}^2)

$$\rho_i = \text{corr}(y_t^2, y_{t-i}^2) = \left(\exp\{\sigma_h^2\delta^i\} - 1\right)/\left(3\exp\{\sigma_h^2\} - 1\right).$$

APPENDIX B: APPROXIMATE KALMAN FILTERING AND QML ESTIMATION

B.1 The Kalman Filter

The stochastic volatility model,

$$y_t = (h_t)^{.5}\varepsilon_t, \qquad \varepsilon_t \sim N(0,1),$$

and

$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t, \qquad v_t \sim N(0,1),$$

can be rewritten as

$$\log\left(y_t^2\right) = -1.27 + \log h_t + \eta_t,$$
$$E[\eta_t] = 0, \quad \text{var}(\eta_t) = \pi^2/2,$$

and

$$\log h_t = \alpha + \delta \log h_{t-1} + \sigma_v v_t, \quad v_t \sim N(0,1).$$

If the distribution of η_t is approximated by a normal distribution the preceding system becomes a standard dynamic linear model, to which the Kalman filter can be applied. The Kalman filter requires three sets of equations—a prediction and updating set which are run forward through the data, and smoothing equations, which are run backward through the data. We follow the standard notation of Anderson and

Moore (1979). Let $\overline{\log h_{t|t-1}}$ be the prediction of $\log h_t$ based on the information available at time $t-1$. $\Omega_{t|t-1}$ is the variance of the prediction. Let $\overline{\log h_{t|t}}$ be the update that uses the information at time t and Ω_t the variance of the update. The equations that recursively compute the predictions and updatings are given by

$$\overline{\log h_{t|t-1}} = \alpha + \delta\overline{\log h_{t-1|t-1}},$$
$$\Omega_{t|t-1} = \delta^2\Omega_{t-1|t-1} + \sigma_v^2,$$

and

$$\overline{\log h_{t|t}} = \overline{\log h_{t|t-1}} + \frac{\Omega_{t|t-1}}{f_t}\left[\log\left(y_t^2\right) + 1.27 - \overline{\log h_{t|t-1}}\right]$$
$$\Omega_{t|t} = \Omega_{t|t-1}\left(1 - \Omega_{t|t-1}/f_t\right),$$

where $f_t = \Omega_{t|t-1} + \pi^2/2$. Once the predictions and updates are computed for $t = 1, \ldots, T$, we can obtain the smoothed estimates, $\overline{\log h_{t|T}}$, which is the estimate of $\log h_t$ given all information in the sample. $\Omega_{t|T}$ denotes the variance of $\overline{\log h_{t|T}}$. The smoothing equations are

$$\overline{\log h_{t|T}} = \overline{\log h_{t|t}} + P_t\left[\overline{\log h_{t+1|T}} - \overline{\log h_{t+1|t}}\right]$$
$$\Omega_{t|T} = \Omega_{t|t} + P_t^2(\Omega_{t+1|T} - \Omega_{t+1|t})$$

where $P_t = \delta\Omega_{t|t}/\Omega_{t+1|t}$. The system is initialized at the unconditional values, $\Omega_0 = \sigma_v^2/(1-\delta^2)$ and $\overline{\log h_0} = \alpha/(1-\delta)$

The prediction, updating, and smoothing estimates of h_t are computed using standard properties of the lognormal distribution.

As discussed in Section 4.4, it is possible to refine the QML to make σ_η^2 a free parameter and avoid an assumption that η is normally distributed.

B.2 QML Estimation With the Kalman Filter

The quasi-likelihood is defined and computed using the predictive error decomposition (see Harvey 1981) $\ell(\alpha, \delta, \sigma_v) \propto -1/2\Sigma\log f_t - 1/2\Sigma e_t^2/f_t$, where f_t is the prediction error variance just defined and e_t is the one-step-ahead prediction error, $e_t = \log(y_t^2) + 1.27 - \overline{\log h_{t|t-1}}$.

[*Received August 1993. Revised October 1993.*]

REFERENCES

Andersen, T., and Sorensen, B. (1993), "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," working paper, Northwestern University, J. L. Kellogg Graduate School of Management.

Anderson, B. O., and Moore, J. B. (1979), *Optimal Filtering*, Englewood Cliffs, NJ: Prentice-Hall.

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Applegate, D., Kannan, R., and Polson, N. (1991), "Random Polynomial Time Algorithms for Sampling From Joint Distributions," working paper, Carnegie Mellon University, Statistics Dept.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992), "ARCH Modeling in Finance," *Journal of Econometrics*, 52, 5–59.

Carlin, B., Polson, N. G., and Stoffer, D. (1992), "A Monte Carlo Approach to Non-normal and Nonlinear State-Space Modelling," *Journal of the American Statistical Association*, 87, 493–500.

Clark, P. K. (1973), "A Subordinated Stochastic Process Model With Finite Variances for Speculative Prices," *Econometrica*, 41, 135–156.

Danielsson, J. (in press), "Stochastic Volatility in Asset Prices: Estimation With Simulated Maximum Likelihood," *Journal of Econometrics*, 54.

Danielsson, J., and Richard, J. F. (1992), "Quadratic Acceleration for Monte Carlo Likelihood Evaluation," unpublished manuscript, University of Iceland, Dept. of Economics.

Foster, D., and Nelson, D. (1992), "Asymptotic Filtering Theory for Univariate ARCH Models," Technical Reports, University of Chicago, Statistics Research Center, Graduate School of Business.

Gallant, A. R., Hsieh, D., and Tauchen, G. (1991), "On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rates, 1974–83," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics, Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, eds. W. A. Barnett, J. Powell, and G. Tauchen, Cambridge, U.K.: Cambridge University Press.

Gallant, A. R., Rossi, P. E., and Tauchen, G. (1992), "Stock Prices and Volume," *The Review of Financial Studies*, 5, 199–242.

Gallant, A. R., and Tauchen G. (1992), "Which Moments to Match," unpublished manuscript, University of North Carolina, Dept. of Economics.

Gelfand, A., and Smith, A. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in *Bayesian Statistics 4*, eds. J. M. Bernardo et al., Oxford, U.K.: Oxford University Press, pp. 169–193.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Harvey, A. C. (1981), *Time Series Models*, New York: John Wiley.

Harvey, A. C., Ruiz, E., and Shephard, N. (1994), "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

Hsieh, D. (1991), "Chaos and Nonlinear Dynamics: Applications to Financial Markets," *Journal of Finance*, 46, 1839–1877.

Hull, J., and White, A. (1987), "The Pricing of Options on Assets With Stochastic Volatilities," *Journal of Finance*, 42, 281–300.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1993), "Priors and Models for Multivariate Stochastic Volatility," unpublished manuscript, University of Chicago, Graduate School of Business.

McCulloch, R. E., and Rossi, P. E. (in press), "An Exact Likelihood Analysis of the Multinomial Probit Model," *Journal of Econometrics*, 54.

McCulloch, R. E., and Tsay, R. (1993), "Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series," *Journal of the American Statistical Association*, 88, 968–978.

Melino, A., and Turnbull, S. (1990), "Pricing Foreign Currency Options With Stochastic Volatility," *Journal of Econometrics*, 45, 7–39.

Mueller, P. (1991), "A Dynamic Vector ARCH Model for Exchange Rate Data," unpublished manuscript, Duke University, Institute for Statistics and Decision Sciences.

Nelson, D. B. (1988), "Time Series Behavior of Stock Market Volatility and Returns," unpublished Ph.D. dissertation, Massachusetts Institute of Technology, Economics Dept.

——— (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Ripley, B. (1987), *Stochastic Simulation*, New York: John Wiley.

Ruiz, E. (1994), "Quasi-Maximum Likelihood Estimation of Stochastic Variance Models," *Journal of Econometrics*, 63, 284–306.

Shephard, N. (1993), "Fitting Nonlinear Time Series Models With Applications to Stochastic Variance Models," *Journal of Applied Econometrics*, 8, S135–S152.

Tanner, T., and Wong, W. (1987), "The Calculation of Posterior Distributions By Data Augmentation," *Journal of the American Statistical Association*, 82, 528–549.

Tauchen, G., and Pitts, M. (1983), "The Price Variability-Volume Relation-

ship on Speculative Markets," *Econometrica*, 51, 485–505.
Taylor, S. (1986), *Modelling Financial Time Series*, New York: John Wiley.
Tierney, L. (1991), "Markov Chains for Exploring Posterior Distributions," Technical Report 560, University of Minnesota, School of Statistics.
Vetzal, K. (1992), "Stochastic Short Rate Volatility and the Pricing of Bonds and Bond Options," unpublished Ph.D. thesis, University of Toronto, Dept. of Economics.
Wiggins, J. (1987), "Options Values Under Stochastic Volatility: Theory and Empirical Evidence," *Journal of Financial Economics* 19, 351–372.
Uhlig, H. (1991), "BVARTEC—Bayesian Vector Autoregressions With Time Varying Error Covariances," working paper, Princeton University, Dept. of Economics.