Improved Quasi-Maximum Likelihood Estimation for Stochastic Volatility Models

F. Jay Breidt and Alicia L. Carriquiry
Iowa State University

Abstract. Jacquier, Polson and Rossi (1994, *J. Business and Economic Statistics*) have proposed a hierarchical model and Markov Chain Monte Carlo methodology for parameter estimation and smoothing in a stochastic volatility model, where the logarithm of the conditional variance follows an autoregressive process. In sampling experiments, their estimators perform particularly well relative to a quasi-maximum likelihood approach, in which the nonlinear stochastic volatility model is linearized via a logarithmic transformation and the resulting linear state-space model is treated as Gaussian. In this paper, we explore a simple modification to the treatment of inlier observations which reduces the excess kurtosis in the distribution of the observation disturbances and improves the performance of the quasi-maximum likelihood procedure.

Keywords. Inliers, excess kurtosis, transformations.

1 Introduction

Financial variables such as stock returns and exchange rates are often modeled as martingale differences. If a sequence of random variables observed over time is a martingale difference, then both the unconditional expectation and the conditional expectation (given the past of the series) of an observation at time t are identically equal to zero. Further, the series has no serial autocorrelation. We will call a serially uncorrelated sequence with zero mean and constant unconditional variance *white noise*. An independent and identically distributed (iid) sequence is both a martingale difference and a white noise. In the iid case, the past of the series contains no information about the present or the future, and forecasting becomes a futile exercise.

It has been shown, however, (e.g., Clark, 1973; Tauchen and Pitts, 1983; Nelson, 1988; Melino and Turnbull, 1990; Harvey, Ruiz, and Shephard, 1994) that series arising in finance and econometrics cannot always be assumed to be iid. While the martingale difference property often appears plausible, the variance in a given realization seems to change over time. In fact, it is often the case that powers of the series itself exhibit serial autocorrelation, and thus it is possible to detect and model dynamics in higher order moments of the series.

Two approaches have been proposed to model time-dependent variances. The first approach, proposed by Engle (1982) and later generalized by Bollerslev (1986) and by others, uses an autoregressive conditionally heteroscedastic (ARCH, or its generalized version, GARCH) process to model the serial autocorrelation in the variances. In this approach, the variance of the series at time t is assumed to be a deterministic function of lagged values of the squared observations and of past variances. For an excellent review of this approach, the reader should refer to Bollerslev, Chou, and Kroner (1992).

The second approach, pioneered in its earliest version in the work of Clark (1973), uses models known as stochastic volatility (SV) models. In this context, it is assumed that smooth functions of the time-dependent variances are random variables generated by an underlying stochastic process, for example an autoregressive process or a random walk. Stochastic volatility models also result from discretizing continuous-time diffusion processes such as those proposed for asset pricing (Hull and White, 1987; Harvey and Shephard, 1993). While intuitively appealing, SV models have not been popular, at least in terms of usage. The reason for the limited empirical application of these models is that, unlike the case of ARCH-type processes, the likelihood function for SV models is hard to evaluate, since it is expressed as a T-dimensional integral, where T is the number of observations.

Several methods for estimation from SV models have been proposed. A method of moments (MM) estimator, which avoids the problem of evaluating the likelihood function, was suggested by Taylor (1986), Melino and Turnbull (1990), and most recently Vetzal (1992). While easy to implement, the MM estimator was shown to be inefficient and to perform poorly over repeated sampling (Jacquier, Polson, and Rossi, 1994). Nelson (1988), Harvey and Shephard (1993), and Harvey et al. (1994), after expressing the SV model in a linear state-space form, used the usual Kalman filter recursions to estimate the parameters and the state vector in the state-space model. This approach, known as quasi-maximum likelihood (QML), produces state estimators that are best (in the mean squared error sense) among all linear estimators, and is simple to implement. However, as Harvey et al. (1994) and Jacquier et al. (1994) point out (among others), the performance of QML decreases in "noisy" series. When the variance of the underlying stochastic process is small, QML estimators can be severely biased and have high root mean squared errors (RMSE). In addition, the transformation from a SV model to the state-space version cannot be carried out for observations with a value of zero. Indeed, the results from the transformation become suspect whenever applied to inliers, where by inliers we mean any observed value that is close to zero. Different "remedies" have been proposed to accommodate inliers, where the most widely used consists of either shifting the whole series away from zero by a small value, or just adding a small amount (for example a fraction of a percent of the series mean) to all zero observations. These inlier treatments have been criticized by, for example, Nelson (1994).

Recently, Jacquier, Polson, and Rossi (1994), using a fully Bayesian framework, derived expressions for the marginal posterior distributions of the parameters and the state vector in SV models, thereby providing an optimal (under the given assumptions) solution to the estimation and smoothing problems. Their approach rests upon the formulation of the nonlinear SV model as a hierarchical model, with the prior distributions for the model parameters at the top of the hierarchy. The computational problem is resolved by means of a Markov chain sampler, specifically a cyclic independence Metropolis chain.

The method suggested by Jacquier et al. (1994) addresses several important issues. Since the nonlinear SV model need not be expressed in linear state-space form, no transformation is necessary and the inlier problem vanishes. In addition, uncertainty about true values of the model parameters is incorporated in a natural way when computing smoothed estimates of the state vector. Furthermore, Jacquier et al. have shown, via extensive sampling experiments, that the method they propose significantly outperforms QML (and MM) both in terms of bias and RMSE, particularly in those cases where the variance of the underlying process is small.

Two drawbacks to Jacquier et al.'s procedure can, however, be pointed out. In order to derive the posterior distributions of interest, it is necessary to make important assumptions regarding the model. Indeed, it would seem that the probabilistic model imposed on the observations, the parameters and the state vector is crucial, and that mis-specification of the model would lead to estimators with poor behavior. For example, the model uses a Gaussian assumption on the disturbance in the observation equation; it is not clear whether the procedure would produce reasonable results when a heavier tailed (or even skewed) distribution is, as many speculate (see, e.g., Harvey and Shephard, 1993), a more appropriate model. It is therefore questionable whether the method outlined in Jacquier et al. is robust to departures from the model's assumptions. The second drawback is purely one of convenience. Jacquier et al.'s procedure is not easy to implement, and in fact their algorithm would require serious modifications whenever the model is tailored to different applications.

In this paper, we propose a simple modification to the usual QML approach for estimation in SV models. In carrying out our work, our objective was to derive a method which, while inefficient in some cases, was robust to departures from model assumptions, and which was simple to implement. The modification we suggest consists in applying a linearizing transformation to shifted values of the observations, where the shift is determined by the slope of the function used for transformation at those points. This modified linearizing transformation addresses the inlier problem, and also improves the third and fourth moments of the distribution of transformed residuals in the observation equation. A similar transformation was proposed by Fuller (1995) in a different context. We show, through a sampling experiment, that our modified transformation significantly improves the performance of the usual QML estimator. Comparison of the behavior of our improved QML estimator to Jacquier et al.'s Bayes estimator is encouraging; in the Gaussian case, and for most parameter values, our estimator performs as well as the Bayes estimator in terms of bias. The improved QML estimator has, however, a higher RMSE than the Bayes estimator, as would be expected.

In Section 2, we present the SV model and the robust linearizing transformation. The usual QML estimator is reviewed in Section 3. Results from the sampling experiment are presented in Section 4, and conclusions and directions for future work are given in Section 5.

2 Model and transformations

Consider the simple stochastic volatility model

$$y_t = \sigma_t \xi_t = \zeta \exp(\alpha_t/2)\xi_t, \quad \alpha_t = \phi\alpha_{t-1} + \eta_t \tag{1}$$

where $\{\xi_t\}$ is iid with mean zero and variance one, $\{\eta_t\}$ is an iid $(0, \sigma_\eta^2)$ sequence of random variables independent of $\{\xi_t\}$, ζ is a positive constant, and $|\phi| < 1$. This model has been considered, for example, by Harvey et al. (1994), Kim and Shephard (1994), and Jacquier et al. (1994).

Goals are usually two-fold: to estimate the parameters ζ, ϕ, and σ_η^2, and to obtain smoothed estimates of the volatilities σ_t. Though model (1) is simple, the likelihood of $(\zeta, \phi, \sigma_\eta^2)$ given $(y_1, \ldots, y_T)$ involves a T-dimensional integral and is not easy to evaluate. Jacquier et al. (1994) have developed a Markov chain simulation methodology for likelihood-based inference in model (1). Their algorithm allows for numerical evaluation of all marginal posterior distributions of interest, thus avoiding the problem of directly evaluating the likelihood of $(\zeta, \phi, \sigma_\eta^2)$.

Nevertheless, as mentioned earlier, there remains some interest in simpler, albeit less efficient, estimation methods; at the least, simple estimation methods could be used as exploratory tools in developing a model specification, which might then suggest a more sophisticated estimation procedure.

2.1 Linearizing transformation

In theory, the series is simple to analyze after transformation, as suggested by Nelson (1988) and Harvey and Shephard (1993) among others. Consider the stationary process

$$\begin{aligned} x_t &= \log y_t^2 \\ &= \log \zeta^2 + \mathrm{E}\left[\log \xi_t^2\right] + \alpha_t + \left(\log \xi_t^2 - \mathrm{E}\left[\log \xi_t^2\right]\right) \\ &= \mu + \alpha_t + \epsilon_t, \end{aligned} \tag{2}$$

where $\{\epsilon_t\}$ is iid with mean zero and variance σ_ϵ^2. For example, if ξ_t is standard normal, then $\log \xi_t^2$ is distributed as the log of a χ_1^2 random variable. The log of its moment generating function is

$$\log \mathrm{E}\left[\exp\left\{t \log \chi_1^2\right\}\right] = \log \mathrm{E}\left[\left(\chi_1^2\right)^t\right] = t \log 2 + \log \Gamma(1/2 + t) - \log \Gamma(1/2),$$

from which we can obtain the first two moments, $\mathrm{E}\left[\log \xi_t^2\right] = -\gamma - \log 2 \simeq -1.27$ (γ is Euler's constant) and $\sigma_\epsilon^2 = \pi^2/2$, respectively. Also, the skewness is

$$\frac{\mathrm{E}\left[(\log \xi_t^2 + \gamma + \log 2)^3\right]}{\sigma_\epsilon^3} = -14\zeta(3)/\sigma_\epsilon^3 \simeq -1.5351$$

and the excess kurtosis is given by

$$\frac{\mathrm{E}\left[(\log \xi_t^2 + \gamma + \log 2)^4\right]}{\sigma_\epsilon^4} - 3 = 90\zeta(4)/\sigma_\epsilon^4 = 4,$$

where $\zeta(\cdot)$ is Riemann's zeta function (Abramowitz and Stegun 1965, §6.4).

The process $\{x_t\}$ is thus a correlated signal plus an iid non-Gaussian noise, with $\mathrm{E}[x_t] = \mu$ and

$$\gamma_x(h) = \mathrm{Cov}(x_t, x_{t+h}) = \gamma(h) + \sigma_\epsilon^2 I_{\{h=0\}},$$

where $\gamma(\cdot)$ denotes the autocovariance function of $\{\alpha_t\}$ and $I_{\{h=0\}}$ is one if $h = 0$ and zero otherwise. In spite of the non-Gaussianity of $\{\epsilon_t\}$, a reasonable estimation procedure is to maximize the Gaussian likelihood of the linear state-space model (2), a procedure known as quasi-maximum likelihood estimation (QML).

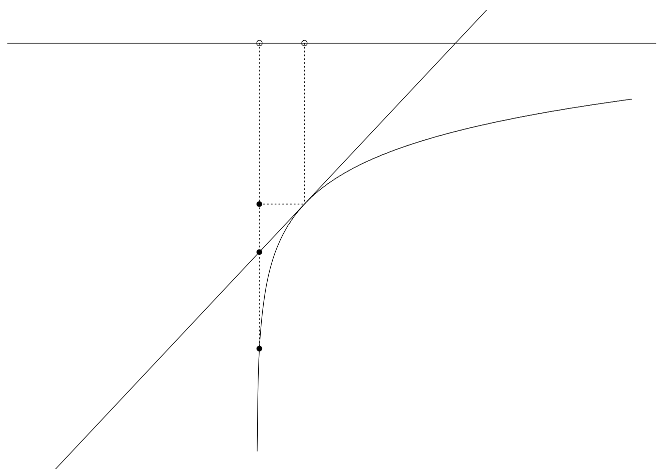


Figure 1: *Schematic diagram of the robustified transformation.*

2.2 Robust transformation

In practice, the series $\{y_t\}$ may contain zeroes (and other inliers), so the log transformation breaks down, since $\log(0) = -\infty$. Often, practitioners mean correct or work with excess returns to avoid the problem of zero observations. These procedures, which are kinds of inlier adjustments, have been criticized by Nelson (1994). An alternative inlier adjustment was presented by Fuller (1995, Example 9.3.2), who suggested evaluating $\log(\cdot)$ not at the (possibly zero) measurement z, but at $z + dz$, where dz is a small increment, and then extrapolating linearly from the point $(z + dz, \log(z + dz))$ using the slope of the tangent line, $(z + dz)^{-1}$. Evaluating the extrapolation line at z, we obtain the transformation

$$\log(z + dz) - dz(z + dz)^{-1}.$$

(See Figure 1.) In the stochastic volatility context, we obtain the robustified transformation

$$\begin{aligned} x_t^* &= \log(y_t^2 + \delta\hat{\sigma}^2) - (y_t^2 + \delta\hat{\sigma}^2)^{-1}\delta\hat{\sigma}^2 \\ &= \log\sigma_t^2 + \log(\xi_t^2 + \delta\hat{\sigma}^2\sigma_t^{-2}) - (\xi_t^2 + \delta\hat{\sigma}^2\sigma_t^{-2})^{-1}\delta\hat{\sigma}^2\sigma_t^{-2} \\ &= \mu^* + \alpha_t + \epsilon_t^*, \end{aligned} \tag{3}$$

where δ is some small constant and $\hat{\sigma}^2$ is the sample mean of the y_t^2. Note that x_t^* is bounded below by $\log\delta\hat{\sigma}^2 - 1$ and that the effect of the transformation is negligible for large y_t^2. That is, the transformation is flexible in that its effect depends on the degree of inlying of each observation. Notice too that the transformation effect is data driven, and there is no need to decide arbitrarily which observations are to be classified as inliers.

We chose $\delta = 0.005$ as the smallest value for which excess kurtosis of the $\{\epsilon_t^*\}$ was near zero across the nine sets of parameter values in Table 1. See Figure 2(c). This choice of δ reduces the skewness of $\{\epsilon_t^*\}$ substantially, as shown in Figure 2(b). Note also that the variance of $\{\epsilon_t^*\}$ is no longer $\pi^2/2$ in the Gaussian case. We treat the variance of $\{\epsilon_t^*\}$ as

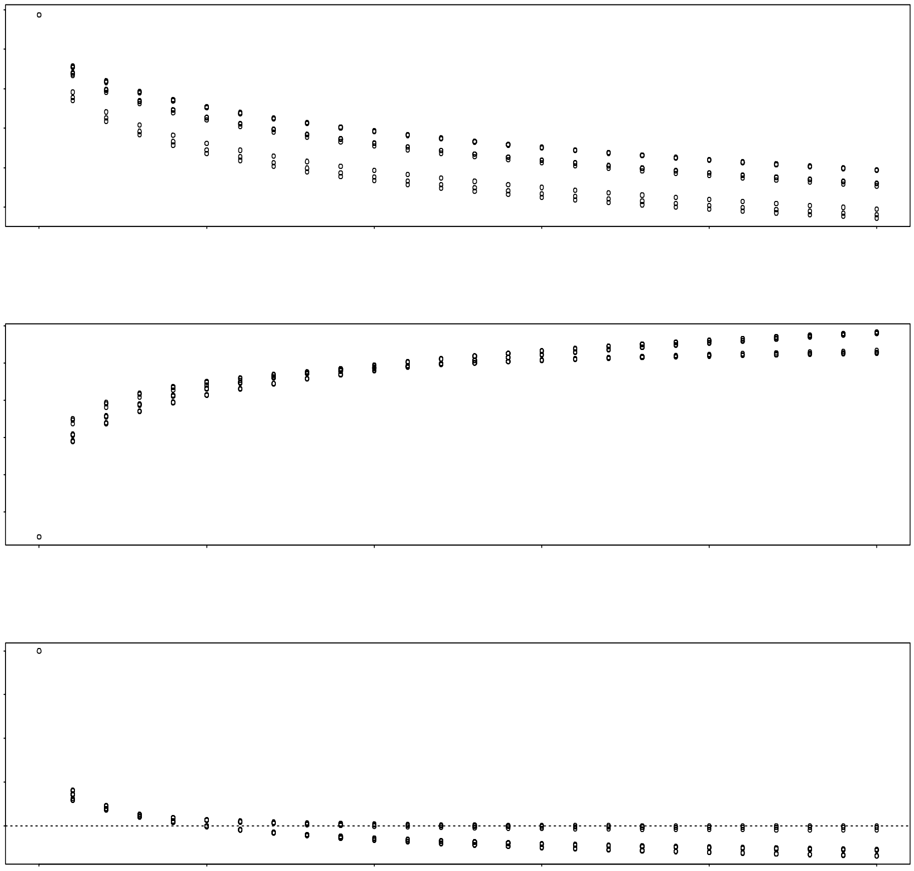


Figure 2: *Choice of δ: (a) variance of $\{\epsilon_t^*\}$ versus δ; (b) skewness of $\{\epsilon_t^*\}$ versus δ; and (c) excess kurtosis of $\{\epsilon_t^*\}$ versus δ for each of nine parameter settings in Table 1. At each parameter setting, statistics for each $\delta \neq 0$ are averages over 1000 simulated realizations of $\{\epsilon_t^*\}$ of length $T = 500$. Values at $\delta = 0$ are theoretical.*

a free parameter and estimate it from the data. We have also experimented extensively with $\delta = 0.02$, with similar results.

The new errors $\{\epsilon_t^*\}$ are no longer iid, only approximately so. Nevertheless, serial correlation in the $\{\epsilon_t^*\}$ is hard to detect; Figure 3 shows the sample autocorrelation function (ACF) for a typical realization of length $T = 500$, along with the Bartlett bounds $\pm 1.96/\sqrt{T}$. For longer realizations, the dependence disappears as the sample mean of the $\{y_t^2\}$ converges to its unconditional expectation.

Figure 4(a) compares the order statistics for 50 realizations of length $T = 500$ of $\{\epsilon_t^*\}$ (with $\phi = 0.95$ and $\sigma_\eta = 0.26$) with the order statistics of $\{\epsilon_t\}$, while Figure 4(b) compares a smoothed probability density estimate for those 50 realizations with the actual probability density function (pdf) of a $\log \chi_1^2$ random variable. Note that most of the impact of the transformation is in the lower tail of the distribution; that is, on the inlying observations in the distribution of y_t.

The transformation in (3) relies on an estimator of the unconditional variance of the process. An alternative transformation would involve using estimators of the conditional

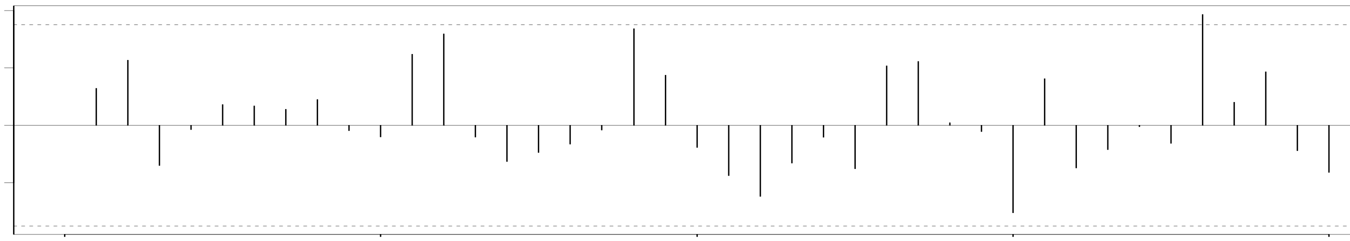
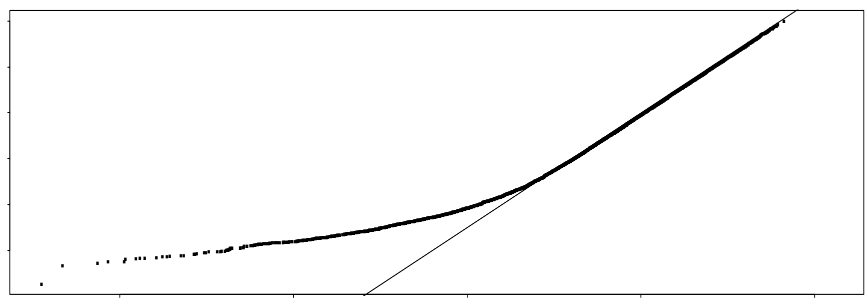


Figure 3: *Sample autocorrelation function of* $\{\epsilon_t^*\}$ *for a typical realization with* $T = 500$, $\phi = 0.95$ *and* $\sigma_\eta = 0.26$.



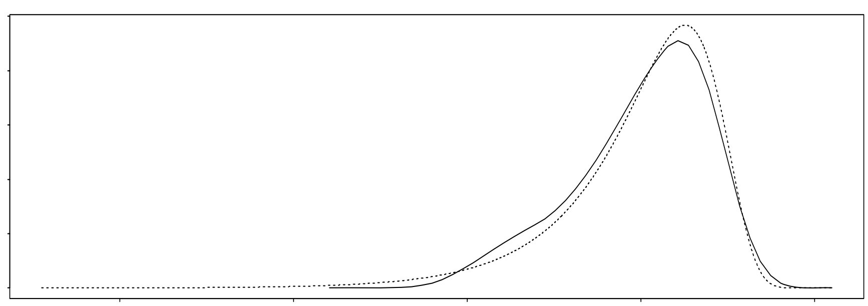


Figure 4: *Comparison of the distributions of* $\{\epsilon_t\}$ *and* $\{\epsilon_t^*\}$ *for 50 realizations with* $T = 500$, $\delta = 0.005$, $\phi = 0.95$ *and* $\sigma_\eta = 0.26$*: (a) Order statistics of* $\{\epsilon_t^*\}$ *versus order statistics of* $\{\epsilon_t\}$*, with* $45°$ *reference line; (b) Smoothed probability density estimate for* $\{\epsilon_t^*\}$ *(—) and theoretical probability density function for* $\{\epsilon_t\}$ $(\cdots)$.

variances, $\{\sigma_t^2\}$. This suggests a two-step estimation procedure: first, estimate parameters of model (1) and use these to find suitable smoothed estimates of the $\{\sigma_t^2\}$, say $\{\hat{\sigma}_t^2\}$; second, transform the original observations via

$$x_t^\dagger = \log(y_t^2 + \delta\hat{\sigma}_t^2) - (y_t^2 + \delta\hat{\sigma}_t^2)^{-1}\delta\hat{\sigma}_t^2. \tag{4}$$

Comparing the one-step robustified transformation given in (3) to the two-step estimation procedure presented in (4), it appears that the two-step estimation method would produce noticeably different parameter estimates whenever variances $\{\sigma_t^2\}$ exhibit relatively large changes over time. If the variances $\{\sigma_t^2\}$ are almost constant, then $\{\sigma_t^2\} \approx \sigma^2$ and the procedure in (4) is roughly equivalent to that given in (3).

3 Estimation and Smoothing

3.1 Kalman recursions and the quasi-likelihood method

Model (2) is in linear state-space form and (3) is approximately in linear state-space form. Predicted, filtered and smoothed values of the unobserved states α_t can thus be computed recursively via the Kalman recursions:

$$a_{t|t-1} = \phi a_{t-1}, \quad P_{t|t-1} = \phi^2 P_{t-1} + \sigma_\eta^2 \tag{5}$$

for one-step-ahead prediction and

$$a_t = a_{t|t-1} + P_{t|t-1} f_t^{-1}(x_t - a_{t|t-1}), \quad P_t = P_{t|t-1} - P_{t|t-1}^2 f_t^{-1} \tag{6}$$

for filtering, where

$$f_t = P_{t|t-1} + \sigma_\epsilon^2,$$

$a_0 = 0$ and $P_0 = \sigma_\eta^2/(1-\phi^2)$ (e.g., Harvey 1989, pp. 105–6). From these recursions, one can compute the innovations

$$\nu_t = x_t - a_{t|t-1}$$

and construct the Gaussian (quasi) log-likelihood,

$$\log \mathcal{L}(\psi) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^T \log f_t - \frac{1}{2}\sum_{t=1}^T \nu_t^2/f_t. \tag{7}$$

The smoothed estimates and their variances are given by

$$a_{t|T} = a_t + P_t^*(a_{t+1|T} - \phi a_t), \quad P_{t|T} = P_t + P_t^*(P_{t+1|T} - P_{t+1|t})P_t^*, \tag{8}$$

where

$$P_t^* = \phi P_t P_{t+1|t}^{-1}$$

(Harvey 1989, p. 154).

The state space model in equation (2) has an ARMA(1,1) reduced form,

$$(1-\phi B)(x_t - \mu) = \eta_t + \epsilon_t - \phi\epsilon_{t-1} = z_t + \theta z_{t-1}, \tag{9}$$

where $\{z_t\}$ is a white noise (WN) sequence with mean zero and variance σ_Z^2. This implies the following locally one-to-one mappings:

$$\sigma_\epsilon^2 = -\theta\sigma_Z^2\phi^{-1}, \quad \sigma_\eta^2 = (1+\theta^2)\sigma_Z^2 + (1+\phi^2)\theta\sigma_Z^2\phi^{-1}. \tag{10}$$

Thus, as an alternative to maximizing (7), we can obtain QML estimates by maximizing an ARMA(1,1) likelihood. Advantages of this approach include readily available software and insight into the nature of the likelihood surface.

The fitting procedure we used is then as follows:

- Transform $\{y_t\}$ to $\{x_t\}$ or $\{x_t^*\}$ and mean-correct the transformed series.
- Maximize the concentrated ARMA(1,1) likelihood of (ϕ, θ) given the mean-corrected transformed series, $\{X_t\}$. This likelihood (dropping irrelevant constants) is given by

  $$\ell(\phi,\theta) = -\frac{T}{2}\log\left\{\sum_{t=1}^{T}\frac{\left(X_t - \hat{X}_t\right)^2}{r_{t-1}}\right\} - \frac{1}{2}\sum_{t=1}^{T}\log r_{t-1},$$

  where

  $$\hat{X}_t = \begin{cases} 0, & t=1 \\ \theta\left(X_{t-1} - \hat{X}_{t-1}\right)/r_{t-2}, & t=2,\ldots,T \end{cases}$$

  and

  $$r_{t-1} = \begin{cases} (1+2\theta\phi+\theta^2)/(1-\phi^2), & t=1 \\ 1+\theta^2-\theta^2/r_{t-2}, & t=2,\ldots,T \end{cases}$$

  (Brockwell and Davis, 1991, §5.3).

  The MLE of σ_Z^2 is then obtained as

  $$\hat{\sigma}_Z^2 = \sum_{t=1}^{T}\frac{\left(X_t - \hat{X}_t\right)^2}{r_{t-1}}.$$

- Map $(\hat{\phi}, \hat{\theta}, \hat{\sigma}_Z^2)$ to $(\hat{\phi}, \hat{\sigma}_\epsilon^2, \hat{\sigma}_\eta^2)$.
- Using $(\hat{\phi}, \hat{\sigma}_\epsilon^2, \hat{\sigma}_\eta^2)$ in (5), (6) and (8), compute $\{a_{t|T}\}$.
- Estimate ζ via

  $$\hat{\zeta} = \left\{T^{-1}\sum_{t=1}^{T} y_t^2 \exp(-a_{t|T})\right\}^{1/2}.$$

3.2 Two-step estimation procedure

Let $\hat{\sigma}_t^2 = \hat{\zeta}^2\exp(a_{t|T})$ for $t = 1,\ldots,T$. If the $\hat{\sigma}_t^2$ are quite variable over the sample, a second round of calculations as in (4) can be carried out in which the transformation (3) is modified to

$$x_t^\dagger = \log(y_t^2 + \delta\hat{\sigma}_t^2) - (y_t^2 + \delta\hat{\sigma}_t^2)^{-1}\delta\hat{\sigma}_t^2,$$

where δ is a pre-specified constant, such as 0.005. Parameters can then be estimated as above, by mean-correcting the transformed series and fitting an ARMA(1,1).

		ϕ		
$\mathrm{Var}(\sigma_t^2)/\mathrm{E}^2[\sigma_t^2]$		0.90	0.95	0.98
10.0	ζ	0.01647	0.01647	0.01647
	σ_η	0.6750	0.4835	0.3082
1.0	ζ	0.02523	0.02523	0.02523
	σ_η	0.3629	0.2600	0.1657
0.1	ζ	0.02929	0.02929	0.02929
	σ_η	0.1346	0.0964	0.0614

Table 1: *Simulation experiment parameter values.*

4 Simulation study

4.1 Design

To assess the performance of the robust transformation in (3), and of the two-step procedure in (4), we conducted a simulation study similar to the one designed by Jacquier et al. (1994). To facilitate comparison with the results of Jacquier et al. (1994), we used for our sampling experiments the models given in Table 4 of their manuscript. The models are parameterized in terms of the ratio $\mathrm{Var}(\sigma_t^2)\,(\mathrm{E}[\sigma_t^2])^{-2}$ which, from Jacquier et al.'s empirical work, is expected to be around 1.0. Also, as found empirically by a number of authors (e.g., Harvey and Shephard, 1993; Kim and Shephard, 1994; Jacquier et al., 1994) values of ϕ between 0.9 and 0.95 are of interest. As explained by Jacquier et al. (1994), all experiments are calibrated so that $\mathrm{E}[\sigma_t^2] = 0.0009$, implying an approximate 20% annual standard deviation if the simulated series are thought of as weekly returns. We consider samples of size $T = 500$ and compute means and root mean squared errors over $N = 500$ simulated realizations for each of the six parameter settings given in Table 1. The same simulated realizations are used for testing each of the estimation procedures.

We used a different parametrization than Jacquier et al. (1994). In our formulation, the scale parameter ζ can be mapped to Jacquier et al.'s autoregressive intercept parameter β via

$$\log(\zeta^2) = \frac{\beta}{1-\phi}.$$

The cases in which the ratio of volatility variance to squared mean is 0.1 are, perhaps, unrealistically difficult without large samples or prior information. For example, a realization $\{y_t\}$ of length $T = 500$ from the model with $\phi = 0.9$ and ratio equal to 0.1 is likely to be indiscernible from a white noise series with constant conditional variance. In this case, the lag-one theoretical autocorrelations for $\{y_t^2\}$ and $\{\log y_t^2\}$ are 0.0389 and 0.0171, respectively, while the asymptotic standard error from Bartlett's formula under a white noise null hypothesis is $(500)^{-1/2} = 0.04472$. Indeed, in our 500 simulated realizations, the average lag-one sample autocorrelations for $\{y_t^2\}$ and $\{\log y_t^2\}$ are 0.033 (0.050) and 0.012 (0.046), respectively, where the simulation standard deviation appears in parentheses. Thus, practitioners carrying out exploratory analyses would be unlikely to choose a stochastic volatility model to fit to these data. Nevertheless, for the sake of completeness,

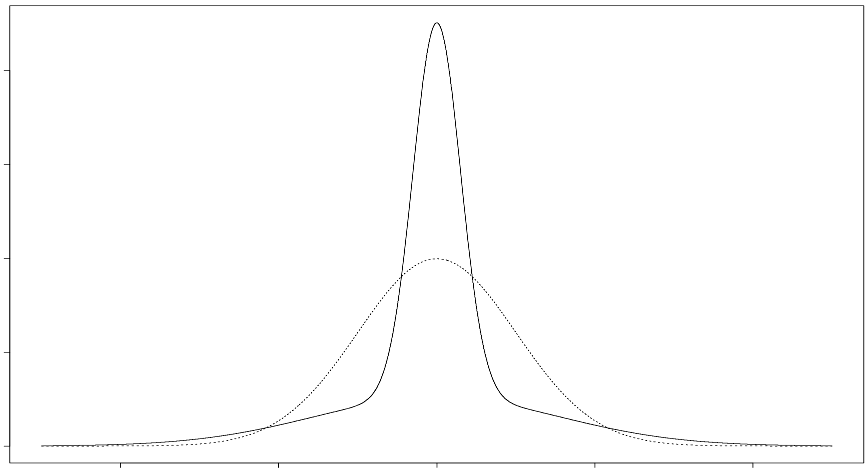


Figure 5: *Comparison of the probability density functions of contaminated normal* $(0.6N(0, 0.09) + 0.4N(0, 2.365))$ *(—) and standard normal* $(\cdots)$.

we included these cases in our simulation study.

To assess the robustness of the QML approach to distributional assumptions, we repeated our simulations with some non-Gaussian disturbance terms, $\{\xi_t\}$ iid with mean zero and variance one. In particular, we considered the contaminated normal distribution obtained by sampling from a $N(0, 0.09)$ distribution with probability 0.6 and from a $N(0, 2.365)$ distribution with probability 0.4. The pdf of this distribution is compared to the pdf of a standard normal in Figure 5. This is designed to be a very difficult case for QML since values near zero are selected with high probability.

4.2 Results

Results from the sampling experiments are presented in Tables 2, 3, and organized to be immediately comparable to those presented by Jacquier et al. (1994), and model parameters β, ϕ, and σ_η correspond to Jacquier et al.'s autoregressive intercept, autoregressive slope, and standard deviation of the underlying process, respectively. Simulation results are given for six different true values of the model parameters. Within each cell in the table, true parameter values are displayed in the first row, Jacquier et al.'s (1994) simulation results for the Bayesian (JPR Bayes) and the usual QML (JPR QML) estimators are reproduced on the second and third rows, respectively, and in the last three rows, we show our results for the usual QML (QML0) estimator, the improved, one-step QML (QML1) estimator, and the two-step QML (QML2) estimator. Entries in the table represent average parameter values over 500 replications of the experiment, and RMSE's are given below, in parenthesis.

As is usually the case, comparing results obtained from a fully Bayesian analysis to those obtained from a classical procedure is tricky. In the case of stochastic volatility, the likelihood function is often very "flat" over the parameter region, and thus, finding extreme values on the likelihood surface may heavily depend on numerical procedures and convergence criteria. In addition, it is not clear how prior information should be factored into the comparison, since whenever data contain little information about the values of parameters, whatever information is incorporated through the prior becomes crucial. Jacquier et al. (1994) point out that the prior distributions used in their sampling experiments were as diffuse as possible. However, as is clear from their discussion in Section 1.4, prior distributions were proper and informative, as would be required to guarantee integrability of the stationary distributions for their Markov chains. As it turns out, the prior information dominated the information provided by the data for those parameter settings in the bottom two rows of Table 2, resulting in significantly less biased estimates for all parameters.

A surprising result arises from the comparison of the performance of Jacquier et al.'s (1994) QML to the usual QML0 estimator we computed. Since JPR QML and QML0 are the same estimator computed from comparable samples, the difference in the performance of these two estimators was unexpected. In particular, Jacquier et al. (1994) report dismal behavior of QML for parameter values such as those in the bottom four cells of the table, whereas we found a significantly better behavior of the usual QML for those parameter settings. It would seem that the difference may be due to numerical accuracy, and that evaluation of the reduced ARMA likelihood improves the performance of the usual QML estimator.

Results obtained for the autoregressive parameter ϕ were generally good for all methods, except for those parameter values that generate sequences that are indistinguishable from white noise with constant conditional variance. At least in terms of bias, all methods appear to produce comparable results. Root mean squared errors, however, were significantly lower for those parameters computed by Jacquier et al. (1994) via the Bayesian method, as would be expected when prior information is used in the estimation procedure.

The performance of the estimates for both β and σ_η varies greatly from one cell of the table to the other, and from one method to another. The variance of the underlying process σ_η, is not easy to estimate using likelihood information alone, particularly when the true value of the parameter is very low. The biases in the estimates obtained from the series that were not distinguishable from white noise were very large, for all versions of the QML approach. The Bayesian method proposed by Jacquier et al. (1994) produced less biased estimates, a result that is likely due to the additional information incorporated through the prior distribution for σ_η. Consider, however, the two middle cells in Table 2, that represent parameter values similar to those that have been estimated from empirical studies (e.g., Melino and Turnbull, 1990; Vetzal, 1992; Jacquier et al. 1994). For these parameter values, both QML1 and QML2 behave well, exhibiting very small biases in the estimates of ϕ and of σ_η. Again, RMSE's were higher than those for the Bayes estimator.

Bias in the estimation of the intercept parameter β was high for all parameter settings and all versions of QML (see Table 2). This is in agreement with results reported by Jacquier et al. (1994), who also obtained biased intercept parameter estimates. As was mentioned in the preceding section, an alternative parametrization would include the scale parameter ζ, as given in expression (1), instead of β. Results presented in Table 3 clearly indicate that parameterizing the model in terms of ζ rather than β produces estimates for ζ that exhibit very low biases and RMSE's across all parameter settings and all versions of QML. It is worth noticing from Table 3 that the bias for ζ approaches zero, even when data are indistinguishable from white noise with constant conditional variance.

Comparison of the one-step improved QML (QML1) from expression (3) to the two-step (QML2) procedure outlined in (4) produced the expected results, as indicated in Table 2. When the true variance of the underlying autoregressive process is relatively high, time-dependent variances in the observations change noticeably over the sample period, and using the two-step approach improves the performance of the usual QML estimator. Consider, for example, those parameter settings in the four top cells in Table 2. For these cases, there seems to be an advantage when data are adjusted according to expression (4).

We present initial results for the non Gaussian case in Table 4. Data for the experiment were generated from a contaminated Gaussian distribution as was described in the preceding section, where only two of the parameter settings from the Gaussian experiment were considered, and correspond to the two middle cells in Table 2. We ran our experiment on those two sets of parameter values because they represent the more "realistic" scenarios. From Table 4 it seems apparent that QML1 performs significantly better than QML0 (the usual QML estimator), and that QML2, while better behaved than QML0, does not perform as well as the one-step QML1. Notice that QML1 is robust to departures from normality, as was hypothesized, and that even in the case we simulated, that was expected to trip the usual QML estimator, QML1 does well. Indeed, the usual QML did not perform well in the presence of inliers, as expected. However, the inlier treatment we propose was able to estimate the parameters with little bias.

5 Discussion

The usual quasi-maximum likelihood estimator has been shown to perform poorly in stochastic volatility models, particularly when variances exhibit relatively minor changes over time (Jaquier et al., 1994; Harvey et al., 1994). Jacquier et al. (1994) proposed a method for estimation for stochastic volatility models that has been shown to behave very well, at least when the data satisfy the assumptions of the model. Their procedure, however, is very intensive from the numerical point of view, and may not be robust to departures from those assumptions.

We have proposed a simple modification to the usual QML estimator which appears to improve its performance over repeated sampling. The modification really affects the treatment of inliers rather than the estimation procedure itself, which benefits mostly from the reduction in the skewness and the excess kurtosis of the distribution of the transformed disturbances in the observation equation.

We have followed Jacquier et al. (1994) regarding the design of our sampling experiment, since we were interested in comparing the performance of our improved QML estimator to their Bayes estimator. We found ourselves, however, in the usual quandary: what is a fair comparison when one method uses information exogenous to the sample and the other does not? In this particular problem, the use of prior information turns out to be crucial, since for some of the parameter values considered in the experiment, the likelihood function is very "non-informative" even relative to a diffuse prior. As we argue in Section 4.1, those parameter settings for which the prior information was dominant gave rise to samples that were not distinguishable from white noise with constant conditional variance. It would seem reasonable to speculate that no practitioner who "looks" at his or her data prior to model fitting would even attempt to estimate stochastic volatility

$\frac{\mathrm{Var}(\sigma_t^2)}{\mathrm{E}^2[\sigma_t^2]}$	Method	β	ϕ	σ_η	β	ϕ	σ_η	β	ϕ	σ_η
	true	-0.821	0.90	0.675	-0.4106	0.95	0.4835	-0.1642	0.98	0.308
	JPR Bayes	-0.679	0.916	0.562	-0.464	0.94	0.46	-0.19	0.98	0.35
		(0.22)	(0.026)	(0.12)	(0.16)	(0.02)	(0.055)	(0.08)	(0.01)	(0.06)
10	JPR QML	-0.99	0.88	0.70	-0.55	0.93	0.51	-0.11	0.99	0.33
		(0.48)	(0.06)	(0.16)	(0.32)	(0.04)	(0.12)	(0.09)	(0.01)	(0.07)
	QML0	-0.924	0.887	0.646	-0.480	0.941	0.402	-0.291	0.965	0.334
		(0.434)	(0.052)	(0.237)	(0.297)	(0.035)	(0.244)	(0.227)	(0.027)	(0.083)
	QML1	-0.941	0.886	0.617	-0.515	0.937	0.445	-0.284	0.966	00.303
		(0.368)	(0.044)	(0.130)	(0.255)	(0.030)	(0.093)	(0.211)	(0.025)	(0.067)
	QML2	-0.944	0.886	0.672	-0.515	0.938	0.483	-0.283	0.966	0.325
		(0.373)	(0.044)	(0.129)	(0.254)	(0.030)	(0.091)	(0.213)	(0.025)	(0.073)
	true	-0.736	0.90	0.363	-0.368	0.95	0.26	-0.1472	0.98	0.1657
	JPR Bayes	-0.87	0.88	0.35	-0.56	0.92	0.28	-0.22	0.97	0.23
		(0.34)	(0.046)	(0.067)	(0.34)	(0.046)	(0.065)	(0.14)	(0.02)	(0.08)
	JPR QML	-1.4	0.81	0.45	-1.0	0.86	0.35	-0.20	0.97	0.22
		(1.6)	(0.22)	(0.27)	(1.7)	(0.23)	(0.25)	(0.54)	(0.08)	(0.15)
1.0	QML0	-1.051	0.858	0.398	-0.610	0.918	0.279	-0.356	0.952	0.208
		(0.892)	(0.117)	(0.220)	(0.652)	(0.087)	(0.174)	(0.580)	(0.074)	(0.154)
	QML1	-0.986	0.867	0.341	-0.534	0.928	0.229	-0.336	0.955	0.191
		(0.762)	(0.099)	(0.173)	(0.430)	(0.057)	(0.135)	(0.540)	(0.068)	(0.125)
	QML2	-0.984	0.868	0.375	-0.540	0.927	0.252	-0.323	0.956	0.195
		(0.749)	(0.098)	(0.189)	(0.458)	(0.061)	(0.153)	(0.444)	(0.059)	(0.103)
	true	-0.706	0.90	0.135	-0.353	0.95	0.0964	-0.1412	0.98	0.0964
	JPR Bayes	-1.54	0.78	0.15	-1.12	0.84	0.12	-0.66	0.91	0.14
		(1.35)	(0.19)	(0.082)	(1.15)	(0.16)	(0.074)	(0.83)	(0.12)	(0.099)
	JPR QML	-5.5	0.23	0.33	-5.5	0.22	0.31	-3.5	0.49	0.35
		(5.6)	(0.79)	(0.39)	(6.0)	(0.85)	(0.41)	(4.6)	(0.67)	(0.46)
0.1	QML0	-2.478	0.658	0.341	-2.119	0.708	0.315	-1.199	0.836	0.229
		(3.100)	(0.424)	(0.414)	(2.950)	(0.403)	(0.401)	(2.001)	(0.266)	(0.405)
	QML1	-2.257	0.688	0.267	-1.822	0.749	0.244	-1.058	0.854	0.182
		(2.831)	(0.387)	(0.308)	(2.634)	(0.359)	(0.304)	(1.784)	(0.240)	(0.301)
	QML2	-2.229	0.692	0.281	-1.747	0.758	0.245	-1.002	0.862	0.181
		(2.805)	(0.384)	(0.334)	(2.535)	(0.348)	(0.301)	(1.726)	(0.231)	(0.301)

Table 2: *Mean and root mean squared error (in parentheses) for five estimation techniques: Bayesian (JPR Bayes) and quasi-maximum likelihood (JPR QML) results from Tables 7 and 6, respectively, of JPR (1994); QML with conventional log-squares transformation (QML0), with robust transformation (QML1), and with two-step transformation (QML2). Statistics in this table are based on 500 simulated samples, each of length* $T = 500$.

$\frac{\mathrm{Var}(\sigma_t^2)}{\mathrm{E}^2[\sigma_t^2]}$	Method	$\phi = 0.90$	$\phi = 0.95$	$\phi = 0.98$
	true ζ	0.01647	0.01647	0.01647
	QML0	0.0179	0.0197	0.0173
		(0.0052)	(0.0079)	(0.0058)
10	QML1	0.0176	0.0168	0.0172
		(0.0049)	(0.0037)	(0.0058)
	QML2	0.0161	0.0165	0.0169
		(0.0029)	(0.0036)	(0.0056)
	true ζ	0.02523	0.02523	0.02523
	QML0	0.0256	0.0257	0.0260
		(0.0029)	(0.0037)	(0.0047)
1	QML1	0.0253	0.0257	0.0259
		(0.0028)	(0.0037)	(0.0046)
	QML2	0.0251	0.0256	0.0257
		(0.0029)	(0.0038)	(0.0045)
	true ζ	0.02929	0.02929	0.02929
	QML0	0.0283	0.0283	0.0286
		(0.0026)	(0.0027)	(0.0030)
0.1	QML1	0.0284	0.0284	0.0288
		(0.0024)	(0.0026)	(0.0027)
	QML2	0.0283	0.0284	0.0287
		(0.0025)	(0.0024)	(0.0028)

Table 3: *Mean and root mean squared error (in parentheses) for estimation of the scale parameter,* ζ*, via* $\hat{\zeta}$ *for QML with conventional log-squares transformation (QML0), with robust transformation (QML1), and with two-step transformation (QML2). Statistics in this table are based on 500 simulated samples, each of length* $T = 500$.

$\frac{\mathrm{Var}(\sigma_t^2)}{\mathrm{E}^2[\sigma_t^2]}$	Method	ζ	ϕ	σ_η	ζ	ϕ	σ_η
	true	0.02523	0.90	0.363	0.02523	0.95	0.26
	QML0	0.0242	0.84	0.470	0.0247	0.91	0.331
		(0.0034)	(0.141)	(0.317)	(0.0035)	(0.099)	(0.198)
1.0	QML1	0.0242	0.85	0.392	0.0246	0.91	0.286
		(0.0033)	(0.127)	(0.225)	(0.0036)	(0.094)	(0.182)
	QML2	0.0234	0.85	0.431	0.0240	0.92	0.315
		(0.0037)	(0.155)	(0.315)	(0.0036)	(0.086)	(0.173)

Table 4: *Mean and root mean squared error (in parentheses) for contaminated Gaussian case using three estimation techniques: QML with conventional log-squares transformation (QML0), with robust transformation (QML1), and with two-step transformation (QML2). Statistics in this table are based on 500 simulated samples, each of length* $T = 500$.

parameters from data such as those. If this were a reasonable assumption, then the prior vs. no prior information argument would be less conflictive, since for all other parameter settings the likelihood function is sufficiently informative to keep from being swamped by the prior. In these cases, therefore, Jacquier et al.'s (1994) attempt to use diffuse priors for the comparison is, in fact, successful.

We hypothesize that even though QML is inefficient when observations are Gaussian, it may be robust when other models are called for. Our initial simulations using a contaminated Gaussian distribution seem to suggest that in fact this is the case. We have conducted similar simulations, not presented in this paper, using a Gumbel distribution to generate the samples. Results from those experiments are also encouraging, and will be pursued.

From an operational point of view, the usual QML estimator presents obvious advantages over the estimator that was proposed by Jacquier et al. (1994). While we wrote our own software for the simulations, any commercial software (e.g., S-Plus, STAMP, SAS) could be used to carry out the procedure we propose. Furthermore, the method is the same, *regardless* of the true distribution of the disturbances in the observation equation. Indeed, QML "knows" right from the start, that the distributional assumptions implicit in the use of the Kalman recursions are not correct, except in the unlikely case where the $\{\xi_t\}$ in model (1) are distributed as log-normal random variables. It is not clear how one would modify Jacquier et al.'s (1994) method to accommodate, for example, a non-standard distribution for the disturbances in the observation equation.

The improved QML procedures we propose (both the one-step and the two-step methods) exhibit higher root mean squared errors than the method presented by Jacquier et al. (1994). This is a serious drawback, and it would not seem possible to correct without modification of the estimation procedure itself or perhaps the model. We have obtained initial results that suggest that the performance of the improved QML estimator in terms of root mean squared error can be improved further, by considering a slightly modified version of model (1). We are currently working on this problem in collaboration with Wayne A. Fuller, and will report on it later.

Acknowledgements: The authors are grateful to Wayne A. Fuller of Iowa State University for his many helpful suggestions.

References

Abramowitz, M. and Stegun, I.A. (1965). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards Applied Mathematics Series 55.

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.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods,* 2nd ed., Springer-Verlag, New York.

Clark, P.K. (1973). A subordinated stochastic process model with finite variances for speculative prices, *Econometrica* **41**, 135–156.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.

Fuller, W.A. (forthcoming). *Introduction to Statistical Time Series,* 2nd ed., John Wiley, New York.

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models, and the Kalman Filter,* Cambridge University Press, Cambridge.

Harvey, A.C., and Shephard, N. (1993). Estimation and testing of stochastic variance models. *Unpublished manuscript, The London School of Economics.*

Harvey, A.C., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models, *Review of Economic Studies* **61**, 247–264.

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. (1994). Bayesian analysis of stochastic volatility models (with discussion). *J. Business and Economic Statistics* **12**, 371–417.

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

Melino, A. and and Turnbull, S.M. (1990). Pricing foreign currency options with stochastic volatility. *J. Econometrics* **45**, 239–265.

Nelson, D.B. (1988). Time series behavior of stock market volatility and returns. *Unpublished PhD dissertation,* Massachusetts Institute of Technology, Economics Department.

Nelson, D.B. (1994). Comment on Jacquier, Polson, and Rossi's "Bayesian analysis of stochastic volatility models". *J. Business and Economic Statistics* **12**, 403–406.

Tauchen, G., and Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica* **51**, 485–505.

Taylor, S. (1986). *Modelling Financial Time Series*, Wiley, New York.

Vetzal, K. (1992). Stochastic short rate volatility and the pricing of bonds and bond options. *Unpublished PhD dissertation*, University of Toronto, Department of Economics.