ISSN 1440-771X

AUSTRALIA

DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS

Bayesian Analysis of the Stochastic Conditional Duration Model

Chris M. Strickland, Catherine S. Forbes and Gael M. Martin

Working Paper 14/2003

Bayesian Analysis of the Stochastic Conditional Duration Model

Chris M. Strickland, Catherine S. Forbes and Gael M. Martin*

Department of Econometrics and Business Statistics, PO Box 11E

Monash University, Vic., 3800, Australia;

August 12, 2003

Abstract

A Bayesian Markov Chain Monte Carlo methodology is developed for estimating the stochastic conditional duration model. The conditional mean of durations between trades is modelled as a latent stochastic process, with the conditional distribution of durations having positive support. The sampling scheme employed is a hybrid of the Gibbs and Metropolis Hastings algorithms, with the latent vector sampled in blocks. The suggested approach is shown to be preferable to the quasi-maximum likelihood approach, and its mixing speed faster than that of an alternative single-move algorithm. The methodology is illustrated with an application to Australian intraday stock market data.

Key Words: Transaction data, Latent factor model, Non-Gaussian state space model, Kalman filter and simulation smoother.

JEL Codes: C11, C15, C41.

*This research has been supported by a Monash University Research Grant and an Australian Research Council Discovery Grant. The authors would also like to thank Ralph Snyder and participants in seminars at Monash University and the 2003 Australasian Meetings of the Econometric Society, for some very helpful comments on an earlier draft of the paper.

1 Introduction

The increasing availability of data at the transaction level for financial commodities has allowed researchers to model the microstructure of financial markets. New models and inferential methods have been developed to enable the analysis of intraday patterns and the testing of certain microstructure hypotheses to occur.

The present paper contributes to this growing literature by presenting a methodology for estimating a particular dynamic model for durations between stock market trades: the stochastic conditional duration (SCD) model. In contrast to the autoregressive conditional duration (ACD) model of Engle and Russell (1998), in which the conditional mean of the durations is modelled as a conditionally deterministic function of past information, the SCD model treats the conditional mean of durations as a stochastic latent process, with the conditional distribution of durations defined on a positive support. As such, the contrast between the two specifications mimics the contrast between the generalized autoregressive conditional heteroscedasticity (GARCH) and stochastic volatility (SV) frameworks for capturing the conditional volatility of financial returns. In particular, as is the case with the SV model, the SCD model presents a potentially more complex estimation problem than its alternative, by augmenting the set of unknowns with a set of unobservable latent factors.

Whilst several modifications of the original ACD specification have been put forward (see Bauwens et al, 2000, for a recent summary), the literature that focusses on the SCD model is much less advanced, with the first introduction of the model into the literature occurring in Bauwens and Veradas (2002).[1] The latter authors present a quasi-maximum likelihood (QML) technique for estimating the SCD model. They also compare the empirical performance of the SCD model and a particular specification of the ACD model, concluding that the SCD model is preferable according to a number of different criteria.

In this paper, a Bayesian methodology for estimating the SCD model is presented. The unobservable latent factors are integrated out of the joint posterior distribution

[1] Durbin and Koopman (2001) suggest the use of a latent variable model for durations, but do not develop the idea further. An alternative latent variable approach to modelling durations is developed in Ghysels, Gourieroux and Jasiak (1998). They present a two factor duration model, referred to as the Stochastic Volatility Duration Model, which accommodates distinct dynamic processes for the first two conditional moments. We focus on the simpler SCD model in this paper, as it provides a more manageable basis for the development of the proposed Bayesian inferential method. Extension of our approach to more general multi-factor models is the topic of on-going research.

via a hybrid Gibbs/Metropolis-Hastings (MH) Markov Chain Monte Carlo (MCMC) sampling scheme. Along the lines suggested in Durbin and Koopman (2000, 2001), the non-Gaussian state-space representation of the model is approximated by a linear Gaussian model in the neighbourhood of the posterior mode of the latent process. This approximating model defines the candidate distribution from which blocks of the latent process are drawn, via the application of the Kalman filter and simulation smoother of de Jong and Shephard (1995). The latent factor draws are then accepted with a particular probability, according to the MH algorithm. The MH subchains associated with the latent factor blocks are embedded into an outer Gibbs chain in the usual way, with estimates of all posterior quantities of interest produced from the draws after convergence of the hybrid algorithm.

The structure of the paper is as follows. Section 2 describes the SCD model. Section 3 then outlines the MCMC scheme, including the details of the approximation used in the production of a candidate distribution for the vector of latent factors. For the purpose of comparison with the multi-move, Kalman filter-based sampler, an alternative single-move sampler is also outlined, whereby the latent factors are sampled one element at a time. In Section 4 a controlled experiment using simulated data, designed to compare the mixing performance of the multi-move and single move samplers, is conducted. In line with expectations, it is found that the multi-move approach has superior mixing performance relative to the single move sampler. This result also tallies with the comparable results for alternative MCMC algorithms for SV models reported in Kim, Shephard and Chib (1998). In Section 5, the repeated sampling behavior of the Bayesian estimation method is compared with the QML approach adopted by Bauwens and Veradas (2002), via a small-scale Monte Carlo experiment. The experiments are based on a sample size of $N = 10000$, to be representative of the typically large sample sizes that are associated with transaction data. The findings indicate that the Bayesian method is superior overall, in terms of both bias and efficiency, as compared with the QML approach. An empirical illustration of the Bayesian method is then described in Section 6. The multi-move sampler is used to produce draws from the posterior distribution of the SCD model of durations between trades in the shares of the Australian firm Broken Hill Proprietary (BHP) Limited, for the month of August 2001. Some conclusions and proposed extensions are given in Section 7.

2 A Stochastic Conditional Duration Model

Denoting by τ_i the time of the ith transaction, the ith duration between successive trades at times τ_{i-1} and τ_i is defined as $x_i = \tau_i - \tau_{i-1}$. The SCD model for x_i is then given by

$$x_i = \exp(\psi_i)\varepsilon_i, \tag{1}$$

$$\psi_{i+1} = \mu + \phi(\psi_i - \mu) + \sigma_\eta \eta_i, \tag{2}$$

for $i = 1, 2, ..., N$, where ε_i is assumed to be an identically and independently distributed ($i.i.d.$) random variable with positive support. For the purposes of this paper it is assumed that ε_i has an exponential distribution with mean (and variance) equal to one. However other suitable distributions are possible; see, for example Bauwens and Veradas (2002). Using (1), the assumption of an exponential distribution for ε_i implies that x_i is also (conditionally) exponential, with conditional mean $\exp(\psi_i)$ and conditional variance $\exp(2\psi_i)$.[2] It is also assumed that

$$\eta_i \sim i.i.d. N(0, \sigma_\eta^2), \tag{3}$$

with ε_i and η_i independent for all i.[3]

The latent factor, ψ_i, in (2) is assumed to be generated by a stationary process, with $|\phi| < 1$. As such, the unconditional mean of the latent factor process is equal to μ. The parameter ϕ is a measure of persistence in the latent process, whilst σ_η^2 captures its variance. Stationarity implies that the initial state has a marginal distribution given by

$$\psi_1 \sim N\left(\mu, \frac{\sigma_\eta^2}{1-\phi^2}\right). \tag{4}$$

[2] Note that one consequence of the use of an exponential distribution for ε_i is that the specification of the dynamic process in (2) for the logarithm of the conditional mean of x_i, also implies the same dynamic structure for the logarithm of the standard deviation of x_i. That is, the so-called dispersion ratio is equal to one. Alternative distributions for ε_i could be adopted in order to allow for either under or over-dispersion; see Bauwens and Veradas (2002) for more discussion. The two factor model of Ghysels et al (1998) avoids the imposition of a constant dispersion ratio.

[3] An alternative parameterization of (1) and (2) involves setting μ to zero and estimating a constant scaling factor $\beta = \exp(\mu)$ in (1). In the SV context, Kim, Shephard and Chib (1998) have shown that this type of parameterization leads to slower convergence for an MCMC algorithm. Pitt and Shephard (1999) provide a theoretical explanation for the slow convergence.

3 Markov Chain Monte Carlo Estimation

An MCMC algorithm is utilised to estimate the SCD model. The advantage of a Gibbs-based MCMC approach over other exact simulation methods, such as importance sampling for example, is that the former method allows complex multivariate densities to be broken up into lower dimensional densities. These lower dimensional densities may then be amenable to direct simulation or, at least, allow for the easier selection of good proposal densities. This is particularly important in this context given the sample sizes that are associated with transaction data, as sampling the entire state vector in one block is not efficient. An MCMC algorithm is presented for estimating the SCD model which is a hybrid of the Gibbs and MH algorithms. As noted above, the algorithm involves sampling the latent state vector in blocks of size greater than 1 (i.e. implementing a multi-move), via the application of the Kalman Filter and simulation smoother to the linear Gaussian approximation to the model. A second algorithm, based on single-move sampling of the states, is presented for comparison purposes. All algorithms are implemented using the GAUSS software.

3.1 A Multi-Move MCMC Algorithm

Defining the $N-$dimensional vector of durations as $x = (x_1, x_2, \ldots, x_N)'$ and the $N-$dimensional latent state vector as $\psi = (\psi_1, \psi_2, \ldots \psi_N)'$, the joint posterior for the full set of unknowns in the SCD model is given by

$$p(\psi, \theta|x) \propto p(x|\psi) \times p(\psi|\theta) \times p(\theta), \tag{5}$$

where $p(x|\psi)$ denotes the joint density function of the observations given the latent volatilities, $p(\psi|\theta)$ denotes the joint density of the latent state vector, conditional on the vector of unknown parameters, $\theta = (\phi, \mu, \sigma_\eta)'$, and $p(\theta)$ is the prior density for the set of unknown parameters. Given the distributional specifications in (1) and (2), it follows that

$$p(x|\psi) = \prod_{i=1}^{N} p(x_i|\psi_i) \tag{6}$$

and

$$p(\psi|\theta) = \left\{ \prod_{i=1}^{N-1} p(\psi_{i+1}|\psi_i, \theta) \right\} \times p(\psi_1|\theta), \tag{7}$$

where

$$p(x_i|\psi_i) = \exp(-\psi_i) \exp\left\{-x_i \exp(-\psi_i)\right\}, \tag{8}$$

for $i = 1, 2, \ldots, N$,

$$p(\psi_{i+1}|\psi_i, \theta) \propto \exp\{\frac{-1}{2\sigma_\eta^2}\left[\psi_{i+1} - \mu - \phi(\psi_i - \mu)\right]^2\}, \tag{9}$$

for $i = 1, 2, \ldots, N-1$, and $p(\psi_1|\theta)$ is as defined in (4). Standard priors for σ_η, ϕ, and μ are used, with details given below[4].

In employing a Gibbs-based MCMC sampler simulated draws from the full conditional distribution relating to each block of unknowns must be obtained. In the multi-move sampler, all of the latent states are sampled in blocks of size greater than one.[5] Since the SCD model is a partially non-Gaussian state space model, the difficulty is in finding a good candidate density for producing a draw of the block of states. One approach outlined by Shephard (1994) and Carter and Kohn (1994), and implemented by Kim, Chib and Shephard (1998) for the SV model, is to approximate the non-Gaussian density in the measurement equation, (1), with a mixture of normal densities. This approach is, however, model specific and given the many different possible distributional assumptions that could be adopted for durations, we have chosen to develop a sampling scheme based a more general approximation method. The methodology employed is outlined in Durbin and Koopman (2000, 2001), whereby the non-Gaussian density for each observation is approximated by a Gaussian density with the same mode. The curvature of the approximating Gaussian density is equated to that of the non-Gaussian density at the mode. The approximation is performed via an iterative Kalman filter which is operationally quite simple.

The steps of the multi-move sampler are summarized as follows:

1. Initialize $\psi, \phi, \mu, \sigma_\eta$.
2. Sample $\sigma_\eta|x, \psi, \phi, \mu$.
3. Sample $\phi|x, \psi, \mu, \sigma_\eta$.
4. Sample $\mu|x, \psi, \phi, \sigma_\eta$.
5. Sample $\psi|\theta$, where ψ is broken up into blocks of size greater than one.

[4] See also Kim, Shephard and Chib (1998).

[5] Sampling latent factors jointly, rather than one at a time, was first proposed by Carter and Kohn (1994) and Frühwirth-Schnatter (1994).

6. Repeat steps 2 to 5 until convergence has been achieved.

The four conditional posteriors of σ_η, ϕ, μ and ψ respectively, including the sampling algorithm required to draw from each conditional, are detailed below.

3.1.1 Sampling σ_η

A conjugate inverted-gamma prior for σ_η is adopted, such that

$$p(\sigma_\eta) \sim IG\left(\frac{\sigma_r}{2}, \frac{S_\sigma}{2}\right), \tag{10}$$

with σ_r and S_σ representing hyperparameters. Given that

$$p(\sigma_\eta|x, \psi, \phi, \mu) \propto p(\psi|\theta)p(\sigma_\eta), \tag{11}$$

where $p(\psi|\theta)$ is as specified in (7), it follows that the full conditional posterior distribution for σ_η is given by

$$\sigma_\eta|x, \psi, \phi, \mu \sim IG\left\{\frac{N+\sigma_r}{2}, \frac{S_\sigma + (\psi_1-\mu)^2(1-\phi^2) + \sum_{i=1}^{N-1}(\psi_{i+1}-\mu-\phi(\psi_i-\mu))^2}{2}\right\}. \tag{12}$$

Draws from (12) can be obtained directly using a standard simulation algorithm.

3.1.2 Sampling ϕ

The prior for ϕ is derived from the beta density function by extending the density over the (-1,1) interval. The prior density for ϕ is thus given by

$$p(\phi) \propto \left\{\frac{1+\phi}{2}\right\}^{\alpha-1}\left\{\frac{1-\phi}{2}\right\}^{\beta-1}, \tag{13}$$

where $\alpha, \beta > 0.5$ are shape parameters for the resultant *stretched* beta distribution. The prior in (13) imposes the stationarity restriction on ϕ, whilst α and β can be selected to assign reasonably high prior weight to values of ϕ that imply a fair degree of persistence in the latent process, as tallies with previous empirical results in the durations literature. The conditional posterior for ϕ is then defined as

$$p(\phi|x, \psi, \mu, \sigma_\eta) \propto p(\psi|\theta)p(\phi). \tag{14}$$

A beta prior is not conjugate for $p(\psi|\theta)$. Hence, the conditional posterior for ϕ in (14) does not mimic the normal structure of $p(\psi|\theta)$, as a function of ϕ, which results from (7). However, it is straightforward to define a candidate density based on that normal structure, with the draws reweighted via an MH algorithm. Defining the candidate density q as a normal density with mean

$$\widehat{\phi} = \frac{\sum_{i=1}^{N-1}(\psi_{i+1}-\mu)(\psi_i-\mu)}{\sum_{i=1}^{N-1}(\psi_i-\mu)^2} \tag{15}$$

and variance

$$\sigma_\phi^2 \;=\; \sigma_\eta^2 \left\{ \sum_{i=1}^{N-1}(\psi_i-\mu)^2 \right\}^{-1}, \tag{16}$$

the steps of the MH Algorithm, inserted at iteration j of the Gibbs chain, are:

1. Specify $\phi^{(j-1)}$ as an initial value for the algorithm.
2. Draw a candidate ϕ^* from $N(\widehat{\phi}, \sigma_\phi^2)$.
3. Accept $\phi^{(j)} = \phi^*$, with probability equal to $\min\left(1, \frac{w(\phi^*|.)}{w(\phi^{(j-1)}|.)}\right)$, where $w(\phi|.) = \frac{p(\phi|.)}{q(\phi|.)}$, $p(\phi|.)$ denotes the conditional posterior in (14), evaluated at the relevant argument, and $q(\phi|.)$ is the corresponding candidate ordinate.
4. Otherwise accept $\phi^{(j)} = \phi^{(j-1)}$.

3.1.3 Sampling μ

A uniform prior over $\mathbb{R}^1$ is defined for μ. Hence, the conditional posterior for μ is normal, with mean

$$\widehat{\mu} = \frac{\sigma_\mu^2}{\sigma_\eta^2}\left\{(1-\phi^2)\psi_1 + (1-\phi)\sum_{i=1}^{N-1}(\psi_{i+1}-\phi\psi_i)\right\} \tag{17}$$

and variance

$$\sigma_\mu^2 = \sigma_\eta^2\left\{(T-1)(1-\phi)^2 + (1-\phi^2)\right\}^{-1}. \tag{18}$$

Draws from this conditional can be obtained directly using a standard simulation algorithm.

3.1.4 Sampling ψ

A blocking scheme for ψ is defined such that

$$\psi = (\psi_{_1} \dots \psi_{k_1}, \psi_{k_{1+1}} \dots \psi_{k_2}, \psi_{k_{2+1}} \dots, \dots \psi_{k_K}, \psi_{k_{K+1}} \dots \psi_N),$$

where $k_1, k_2, \dots, k_K$ are the knot points separating the $(K+1)$ blocks. The knots at each iteration are selected stochastically via the following formula,

$$k_l = round(N/(K+1) \times (l + U_l)) \quad l = 1, \dots, K, \tag{19}$$

where U_l is uniformly distributed $(0, 0.5)$, and K is chosen to satisfy the condition $N/(K+1) > 2$.[6] The selection of K is based on a compromise between the simulation efficiency gains of using a larger average block size against the higher associated rejection rate in the algorithm.

Defining $\psi_{B_l} = (\psi_{k_{(l-1)^+1}} \dots \psi_{k_l})$, $l = 1, \dots, K+1$, with $k_0 = 0$ and $\psi_{k_{0+1}} = \psi_{_1}$, the steps of the sampling scheme for ψ, inserted at iteration j of the Gibbs chain, are:

1. Sample $\psi^j_{B_1}|x, \psi^{j-1}_{B_2}, \theta$.
2. Sample $\psi^j_l|x, \psi^j_{l-1}, \psi^{j-1}_{l+1}, \theta$, for $l = 2, 3, ..., K$.
3. Sample $\psi^j_{B_{K+1}}|x, \psi^j_{B_K}, \theta$.

For each block ψ_{B_l} a linear Gaussian approximation to the non-Gaussian state space model represented by (1) and (2) is produced. The Gaussian approximation then serves as a candidate model from which a candidate draw for ψ_{B_l} is produced. The candidate draw is accepted with a probability determined by the appropriate MH ratio. The methodology employed to produce the linear Gaussian approximating model was first suggested by Durbin and Koopman (2000, 2001) in the context of general non-Gaussian state space models. The approximating model is derived in such a way that the mode of the candidate density associated with this model, $q(\psi_{B_l}|x, \psi_{B_{l-1}}, \psi_{B_{l+1}}, \theta)$, is equivalent to the mode of the actual conditional posterior for ψ_{B_l} as based on the non-Gaussian model, $p(\psi_{B_l}|x, \psi_{B_{l-1}}, \psi_{B_{l+1}}, \theta)$. In order to minimize the notational complexity associated with the description of this component of

[6] We set K such that $N/(K+1) = 20$ in all applications of the algorithm in the paper. See also Shephard and Pitt (1997) for a related selection rule.

the algorithm, it is assumed for the moment that $K = 0$, i.e. that ψ is simulated as a single block of size N.

The process of matching the modes begins with an initial specification of an approximating measurement equation as

$$\widetilde{x}_i = \psi_i + \widetilde{\varepsilon}_i, \tag{20}$$

for $i = 1, 2, \ldots, N$, where $\widetilde{\varepsilon}_i \sim N(0, \widetilde{H}_i)$ and both $\widetilde{x}_i$ and $\widetilde{H}_i$ are defined as particular functions of x_i and an initial trial value of ψ_i, $\widetilde{\psi}_i$. As is demonstrated below, these functions are updated via an iterative procedure in such a way that the modes of $q(\psi|x,\theta)$ and $p(\psi|x,\theta)$ are ultimately equated. The mode of the candidate density $q(\psi|x,\theta)$ is the solution to the vector equation $\frac{\partial \ln q(\psi|x,\theta)}{\partial \psi} = 0$. Equivalently, it is the solution to the vector equation $\frac{\partial \ln q(\psi,x|\theta)}{\partial \psi} = 0$. Given the linear Gaussian model in (20), the assumption of the density in (4) for ψ_1, and the form of the linear Gaussian state equation in (2), it follows that

$$\begin{aligned}
\ln q(\psi, x|\theta) &= \text{constant} - \frac{1}{2}\left(\frac{1-\phi^2}{\sigma_\eta^2}\right)(\psi_1 - \mu)^2 \\
&\quad -\frac{1}{2\sigma_\eta^2}\sum_{i=1}^{N-1}\left(\psi_{i+1} - [\mu + \phi(\psi_i - \mu)]\right)^2 \\
&\quad -\frac{1}{2}\sum_{i=1}^{N}\frac{(\widetilde{x}_i - \psi_i)^2}{\widetilde{H}_i}.
\end{aligned} \tag{21}$$

Differentiating with respect to ψ_i and setting the result equal to zero yields the equations

$$\begin{aligned}
\frac{\partial \ln q(\psi, x|\theta)}{\partial \psi_i} &= (d_i - 1)\left(\frac{1-\phi^2}{\sigma_\eta^2}\right)(\psi_1 - \mu) \\
&\quad -d_i\left(\frac{1}{\sigma_\eta^2}\right)\left(\psi_i - [\mu + \phi(\psi_{i-1} - \mu)]\right) \\
&\quad +\left(\frac{\phi}{\sigma_\eta^2}\right)\left(\psi_{i+1} - [\mu + \phi(\psi_i - \mu)]\right) \\
&\quad +\frac{(\widetilde{x}_i - \psi_i)}{\widetilde{H}_i} \\
&= 0,
\end{aligned} \tag{22}$$

$i = 1, 2, \ldots, N$, where $d_1 = 0$ and $d_i = 1$ for $i = 2, \ldots, N$. Since $q(\psi|x,\theta)$ is Gaussian, the solution to (22) is equivalent to the mean of $q(\psi|x,\theta)$ which can, in turn, be

produced via the application of the Kalman filter and smoother to the model defined by (20) and (2).

Similarly, for the non-Gaussian model, the mode of $p(\psi|x,\theta)$ is the solution to the vector equation $\frac{\partial \ln p(\psi|x,\theta)}{\partial \psi} = 0$, and therefore equivalently, to the vector equation $\frac{\partial \ln p(\psi,x|\theta)}{\partial \psi} = 0$. Given the model in (1) and (2) and the distributional assumption in (4) for ψ_1,

$$\begin{aligned}
\ln p(\psi, x|\theta) &= \text{constant} - \frac{1}{2}\left(\frac{1-\phi^2}{\sigma_\eta^2}\right)(\psi_1 - \mu)^2 \\
&\quad -\frac{1}{2\sigma_\eta^2}\sum_{i=1}^{N-1}\left(\psi_{i+1} - [\mu + \phi(\psi_i - \mu)]\right)^2 \\
&\quad -\sum_{i=1}^{N} h(x_i|\psi_i), \qquad (23)
\end{aligned}$$

where $h(x_i|\psi_i) = -\ln p(x_i|\psi_i)$. Again, differentiating with respect to ψ_i and setting the result to zero produces the first order conditions,

$$\begin{aligned}
\frac{\partial \ln p(\psi, x|\theta)}{\partial \psi_i} &= (d_i - 1)\left(\frac{1-\phi^2}{\sigma_\eta^2}\right)(\psi_1 - \mu) \\
&\quad -d_i\left(\frac{1}{\sigma_\eta^2}\right)\left(\psi_i - [\mu + \phi(\psi_{i-1} - \mu)]\right) \\
&\quad +\left(\frac{\phi}{\sigma_\eta^2}\right)\left(\psi_{i+1} - [\mu + \phi(\psi_i - \mu)]\right) \\
&\quad -\frac{\partial h(x_i|\psi_i)}{\partial \psi_i} \\
&= 0, \qquad (24)
\end{aligned}$$

for $i = 1, 2, \ldots, N$, with d_i as defined above. The approximate model in (20) is to be chosen in such a way that the solution to (22) is equivalent to the solution to (24). To achieve this the term $\frac{\partial h(x_i|\psi_i)}{\partial \psi_i}$ in (24) is linearized around the trial value $\widetilde{\psi}_i$ as follows

$$\frac{\partial h(x_i|\psi_i)}{\partial \psi_i} \approx \left.\frac{\partial h(x_i|\psi_i)}{\partial \psi_i}\right|_{\psi_i=\widetilde{\psi}_i} + \left.\frac{\partial^2 h(x_i|\psi_i)}{\partial \psi_i^2}\right|_{\psi_i=\widetilde{\psi}_i}(\psi_i - \widetilde{\psi}_i). \qquad (25)$$

Substituting (25) into (24), and rearranging, an explicit expression for $\frac{\partial \ln q(\psi,x|\theta)}{\partial \psi_i}$ is

obtained,

$$\begin{aligned}\frac{\partial \ln q(\psi, x|\theta)}{\partial \psi_i} &= (d_i - 1)\left(\frac{1-\phi^2}{\sigma_\eta^2}\right)(\psi_1 - \mu) \\ &\quad -d_i\left(\frac{1}{\sigma_\eta^2}\right)\left(\psi_i - [\mu + \phi(\psi_{i-1} - \mu)]\right) \\ &\quad +\left(\frac{\phi}{\sigma_\eta^2}\right)\left(\psi_{i+1} - [\mu + \phi(\psi_i - \mu)]\right) \\ &\quad +\frac{(\widetilde{x}_i - \psi_i)}{\widetilde{H}_i} \\ &= 0, \end{aligned} \tag{26}$$

where $\dot{h}_i = \frac{\partial h(x_i|\psi_i)}{\partial \psi_i}$, $\ddot{h}_i = \frac{\partial^2 h(x_i|\psi_i)}{\partial \psi_i^2}$, $\widetilde{H}_i = \ddot{h}_i^{-1}$ and $\widetilde{x}_i = \psi_i - \widetilde{H}_i \dot{h}$. Given the form of the density in (8),

$$h(x_i|\psi_i) = \psi_i + x_i \exp(-\psi_i), \tag{27}$$

$$\dot{h}_i = 1 - x_i \exp(-\psi_i), \tag{28}$$

$$\ddot{h}_i = x_i \exp(-\psi_i), \tag{29}$$

$$\widetilde{H}_i = x_i^{-1} \exp(\psi_i) \tag{30}$$

and

$$\widetilde{x}_i = \psi_i - x_i^{-1} \exp(\psi_i) + 1. \tag{31}$$

The iterative procedure is thus based on the following steps:

1. Initialize $\widetilde{H}_i$ and $\widetilde{x}_i$, via the initial trial value of ψ_i, $\widetilde{\psi}_i$.
2. Run the Kalman filter and smoother based on (20) and (2) to produce the mode of $q(\psi|x, \theta)$.
3. Substitute the mode of $q(\psi|x, \theta)$ into (24) and check whether the N first order conditions are satisfied.

4. If the first order conditions are not satisfied recalculate $\widetilde{H}_i$ and $\widetilde{x}_i$ using (27) to (31), as based on the output of the Kalman filter and smoother in 2. (i.e. the current mode of $q(\psi|x,\theta)$)

5. Repeat from step 2 until the first order conditions in (24) are satisfied.

Once the linear Gaussian approximating model has been obtained, that is the measurement equation as defined by (20), with $\widetilde{H}_i$ and $\widetilde{x}_i$ derived via the above iterative procedure, a candidate draw of ψ, ψ^*, is produced from $q(\psi|x,\theta)$. Drawing from $q(\psi|x,\theta)$ is implemented through the use of the Kalman filter and simulation smoother of de Jong and Shephard (1995).

In summary then, and reverting to the consideration of producing the jth draw of block B_l, $l = 1, 2, \ldots, K+1$, of the latent vector ψ at iteration j of the Gibbs sampler, the steps are:

1. Initialize ψ_{B_l}.

2. Run the iterative procedure described above to produce $\widetilde{H}$ and $\widetilde{x}$.[7]

3. Define the approximating measurement equation as (20), for the $(k_l - k_{(l-1)})$ elements in the block ψ_{B_l}.

4. Generate a candidate $\psi^*_{B_l}$ from $q(\psi_{B_l}|x, \psi_{B_{l-1}}, \psi_{B_{l+1}}, \theta)$ using the Kalman filter and simulation smoother.

5. Accept $\psi^{(j)}_{B_l} = \psi^*_{B_l}$, with probability equal to $\min\left(1, \frac{w(x_{B_l}|\psi^*_{B_l})}{w(x_{B_l}|\psi^{(j-1)}_{B_l})}\right)$, where $w(x_{B_l}|\psi_{B_l}) = \frac{p(x_{B_l}|\psi_{B_l})}{q(x_{B_l}|\psi_{B_l})}$.[8]

6. Otherwise accept $\psi^{(j)}_{B_l} = \psi^{(j-1)}_{B_l}$.

---

[7] $\widetilde{H}$ and $\widetilde{x}$ are re-dimensioned appropriately to match the blocking of ψ.

[8] Note that $\frac{p(\psi_{B_l}|x,\psi_{B_{l-1}},\psi_{B_{l+1}},\theta)}{q(\psi_{B_l}|x,\psi_{B_{l-1}},\psi_{B_{l+1}},\theta)} \propto \frac{p(x_{B_l}|\psi_{B_l})p(\psi_{B_l}|\psi_{B_{l-1}},\psi_{B_{l+1}},\theta)}{q(x_{B_l}|\psi_{B_l})q(\psi_{B_l}|\psi_{B_{l-1}},\psi_{B_{l+1}},\theta)} \propto \frac{p(x_{B_l}|\psi_{B_l})}{q(x_{B_l}|\psi_{B_l})}$ since $p(\psi_{B_l}|\psi_{B_{l-1}},\psi_{B_{l+1}},\theta) = q(\psi_{B_l}|\psi_{B_{l-1}},\psi_{B_{l+1}},\theta)$.

3.2 A Single-Move MCMC Algorithm

For the purpose of comparison, an alternative Gibbs-MH scheme is provided which is defined by the blocking scheme $(\psi_1, \psi_2, \ldots, \psi_N, \phi, \mu, \sigma_\eta)$. This form of algorithm is referred to as a single-move sampler, as the latent factor at each time point t is sampled individually. The problem with this approach is that the high correlation between the components of $\psi|x, \theta$ impacts on the Markov chain, restricting its movement across the joint parameter space. For example, in the context of partially non-Gaussian state space models, Shephard (1994) has shown the single-move sampler to have inferior simulation efficiency when compared with a multi-move sampler based on a blocking of the latent states. It is of interest to quantify the relative efficiency performance of the two samplers in the durations context.

The following sampling scheme is used in the single-move sampler:

1. Initialize ψ, ϕ, μ.
2. Sample $\sigma_\eta|x, \psi, \phi, \mu$.
3. Sample $\phi|x, \psi, \mu, \sigma_\eta$.
4. Sample $\mu|x, \psi, \sigma_\eta, \phi$.
5. Sample $\psi_1|x, \psi_2, \theta$.
6. Sample $\psi_i|x, \psi_{i-1}, \psi_{i+1}, \theta$, for $i = 2, \ldots, N-1$.
7. Sample $\psi_N|x, \psi_{N-1}, \theta$.
8. Repeat steps 2 to 7 until convergence has been achieved.

With the parameters sampled as described in Section 3.1, the single-move sampler involves replacing the sampling algorithm for ψ outlined in Section 3.1.4 by the following scheme.

3.2.1 Sampling ψ_i

In the single move sampler the ith state ψ_i is sampled individually, for $i = 1, 2, \ldots, N$. As a consequence of the Markov nature of the state process in (2), ψ_i is dependent upon the values of ψ_{i-1} and ψ_{i+1}. The full conditional distribution for ψ_i,

$i = 2, 3, \ldots N-1$, is

$$p(\psi_i|x, \psi_{i+1,}\psi_{i-1}, \theta) \propto p(x_i|\psi_i, \theta)p(\psi_i|\psi_{i+1}, \psi_{i-1}, \theta), \tag{32}$$

where

$$p(\psi_i|\psi_{i+1}, \psi_{i-1}, \theta) \quad \propto \quad p(\psi_{i+1}|\psi_i, \theta)p(\psi_i|\psi_{i-1}, \theta). \tag{33}$$

From (2) it follows that

$$\psi_i|\psi_{i+1}, \psi_{i-1}, \theta \sim N(\omega_i, \nu_i^2), \tag{34}$$

with

$$\omega_i = \frac{\mu + (\psi_{i+1} - \mu) + \phi(\psi_{i-1} - \mu)}{1 + \phi^2}, \tag{35}$$

$$\nu_i^2 = \frac{\sigma_\eta^2}{1 + \phi^2}. \tag{36}$$

The conditional posterior of the initial value ψ_1, is

$$p(\psi_1|x, \psi_2, \theta) \propto p(x_1|\psi_1)p(\psi_1|\psi_2, \theta), \tag{37}$$

where

$$p(\psi_1|\psi_2, \theta) \propto p(\psi_2|\psi_1, \theta)p(\psi_1|\theta) \tag{38}$$

is normal with mean and variance given respectively by

$$\omega_1 = \frac{\phi(\psi_2 - \mu + 2\phi\mu) + \mu}{1 + 2\phi^2}, \tag{39}$$

and

$$\nu_1^2 = \frac{\sigma_\eta^2}{1 + 2\phi^2}. \tag{40}$$

The conditional posterior of ψ_N is

$$p(\psi_N|x, \psi_{N-1}, \theta) \propto p(x_N|\psi_N, \theta)p(\psi_N|\psi_{N-1}, \theta), \tag{41}$$

where $p(\psi_N|\psi_{N-1}, \theta)$ is normal with mean

$$\omega_N = \mu + \phi(\psi_{N-1} - \mu) \tag{42}$$

and variance

$$\nu_N = \sigma_\eta^2. \tag{43}$$

To construct a candidate density a function $m(\psi_i|x_i) = \ln\ p(x_i|\psi)$ is defined, where $p(x_i|\psi_i)$ is given by (8). An approximation to $m(\psi_i|x_i)$ is obtained using a first-order Taylor series expansion around ω_i as defined in (35). Combining this approximation with the normal density associated with (34) yields a normal candidate density $q(\psi_i|x, \psi_{i+1}, \psi_{i-1}, \theta)$ with mean

$$\omega_i^* = \omega_i + \nu_i^2(x_i \exp\{-\mu\} - 1) \tag{44}$$

and variance ν_i^2.

The steps of the MH Algorithm, inserted at iteration j of the Gibbs chain for $i = 1, \ldots, N$, are:

1. Draw a candidate ψ_i^* from $N(\omega_i^*, \nu_i^2)$
2. Accept $\psi_i^{(j)} = \psi_i^*$, with probability $\min\left(1, \frac{w(x|\psi_i^*)}{w(x|\psi_i^{(j-1)})}\right)$, where $w(x|\psi_i) = \frac{p(x|\psi_i)}{q(x|\psi_i)}$.[9]
3. Otherwise accept $\psi_i^{(j)} = \psi_i^{(j-1)}$.

4 Simulation Efficiency

In this section we report the results of a simulation experiment designed to ascertain the relative simulation efficiency of the two MCMC samplers outlined in the paper, based respectively on the multi-move and single-move algorithms. To evaluate simulation efficiency the inefficiency factor is calculated; see also Kim, Shephard and Chib (1998). The inefficiency factor may be interpreted as the magnitude of the variance of the sample mean of the MCMC chain, relative to the variance of the mean of a hypothetical independently distributed sample of draws. To calculate the inefficiency factor the following formula is utilised,

$$\widehat{IF} = 1 + 2\frac{B}{B-1}\sum_{i=1}^{B} K_{QS}\left(\frac{i}{B}\right)\widehat{\rho}_i, \tag{45}$$

[9]Note that as with the multi-move sampler the calculation of the MH transition probability is simplified because $p(\psi_i|\psi_{i+1}, \psi_{i-1}, \theta) = q(\psi_i|\psi_{i+1}, \psi_{i-1}, \theta)$.

where $\widehat{\rho}_i$ is the estimate of the correlation at lag i of the MCMC iterates, K_{QS} is the Quadratic Spectral (QS) kernel and B is the bandwidth.[10] The QS kernel is defined as

$$K_{QS}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right). \tag{46}$$

To select the bandwidth B the automatic bandwidth selector of Andrews (1991) is used, which estimates the bandwidth as a function of the data. For the QS kernel the automatic bandwidth selector is defined as

$$\widehat{B} = 1.3221(\widehat{\alpha}(2)M)^{1/5}, \tag{47}$$

where M is the number of iterations in the Markov Chain and

$$\widehat{\alpha}(2) = \frac{4\widehat{\rho}_a^2\widehat{\sigma}_a^4}{(1-\widehat{\rho}_a)^8} \Big/ \frac{\widehat{\sigma}_a^4}{(1-\widehat{\rho}_a)^4}\,. \tag{48}$$

The terms $\widehat{\rho}_a$ and $\widehat{\sigma}_a$ in (48) are estimated by running a first-order autoregressive linear regression on the draws of the Markov Chain, where $\widehat{\rho}_a$ is the autoregressive coefficient and $\widehat{\sigma}_a$ is the estimated standard error.

In the simulation experiment, both the multi-move and single-move samplers, based on a sample size of 5000, are run for 25000 iterations, with a burn in period of 10000 iterations. The parameter settings used in generating the artificial data are $\phi = 0.95$, $\sigma_\eta = 0.1$ and $\mu = 0.0334$. These parameter values are similar to those reported in Section 6 for the empirical application of the SCD model. The hyperparameters σ_r and S_σ in (10) are set to 3 and 0.03 respectively, implying a prior mean of 0.12 and variance 0.0017 for σ_η. The hyperparameters α and β in (13) are set to 15 and 1.2 respectively, implying a prior mean of 0.85 and a variance of 0.016 for ϕ. In both cases these prior settings seem reasonable, given published empirical findings on durations. Experimentation with different prior settings leads to the conclusion that for the large sample sizes used in the analysis in this and subsequent sections, the results are robust to the precise prior specifications.

Table 1 contains the results of the experiment. The factors for the multi-move sampler as relate to σ_η and μ are both markedly smaller than the corresponding factors for the single-move sampler. The factors for ϕ are approximately the same for both samplers. Based on the estimated inefficiency factors for σ_η, approximately

[10] We select the QS kernel as Andrews (1991) finds it to be superior in terms of an asymptotic truncated mean squared error criterion, relative to other kernels.

Table 1: Inefficiency Factors for the Multi-move and Single-move Algorithms

Parameter	Multi-move	Single-move
ϕ	56.93	55.38
σ_η	346.54	643.28
μ	3.11	4.78

35000 iterations in the multi-move sampler and 65000 iterations in the single-move sampler are required to limit the Monte Carlo error to be less than 1% of the percent of the variation of the error which is related to the data.

5 Sampling Experiments

A small-scale Monte Carlo experiment is conducted to assess the sampling properties of the Bayesian simulation method and to compare these properties with those of the QML approach adopted by Bauwens and Veradas (2002) in their analysis of the SCD model. Earlier research by Jaquier, Polson and Rossi (1994) in an SV setting shows that the QML approach works poorly with relatively small sample sizes (i.e. $N = 500$), showing bias and inefficiency relative to the Bayesian MCMC method. With a larger sample size of $N = 2000$, however, Jaquier et al find little bias in both the QML and Bayesian estimators but find that the Bayesian estimator produces efficiency gains over the QML estimator. In the Monte Carlo experiment conducted here a sample size of $N = 10000$ is employed to be representative of the typically large sample sizes that are associated with transaction data. Artificial data is generated for the model specified in (1) and (2), with parameter settings $\phi = \{0.95,\ 0.90\}$, $\sigma_\eta = \{0.1,\ 0.3\}$ and $\mu = 0.0334$. Again, these parameter settings correspond to a range of values that are representative of the estimated parameter values reported for the empirical study undertaken in Section 6. The hyperparameters σ_r, S_σ, α and β are set to the values described in the previous section. Bayesian point estimates of the parameters are produced using the marginal posterior means estimated from the draws of the MCMC algorithm.

The QML approach is based on a logarithmic transformation of the measurement equation (1) which produces a linear relationship between the transformed durations and the state. The transformed measurement equation has the following form,

$$\ln(x_i) = c_i + \psi_i + \zeta_i, \tag{49}$$

where $c_i = E[\ln \varepsilon_i]$ and ζ_i has a zero mean and variance equal to $Var[\ln \varepsilon_i]$. The QML estimation method involves constructing the likelihood function via the Kalman filter, by treating ζ_i as though it were $i.i.d.N(0, Var[\ln \varepsilon_i])$. When ε_i is assumed to be exponentially distributed with a mean of 1, $E[\ln \varepsilon_i] = -\gamma^*$, with $\gamma^* \approx 0.5772 =$ Euler's constant, and $Var[\ln \varepsilon_i] = \frac{\pi^2}{6}$; see Johnson, Kotz and Balakrishnan (1994). Standard asymptotic theory implies that the QML estimator will be consistent yet inefficient. This corresponds with the simulation findings of Jaquier et al (1994) cited earlier for the SV context, who find little evidence of bias with larger sample sizes, yet find the QML estimator to be inefficient relative to their exact Bayesian estimator.

The number of replications for each parameter setting is 100. To reduce the computational burden, the MCMC algorithm is implemented with a burn-in period of only 2000 iterations after which the next 5000 iterations are stored[11]. The results are reported in Table 2. The true parameter values are shown in the second column and the Monte Carlo (MC) mean and root mean squared error (RMSE) for the MCMC and QML methods respectively, reported in the subsequent columns. The MC mean shows the MCMC sampler to have negligible bias for all parameter settings. In contrast, the QML estimator still shows clear bias for one particular setting, namely for $\phi = 0.9$ and $\sigma_\eta = 0.1$, even with a sample size of 10000. As indicated by the ratios of RMSE's reported in the last column in the table, the MCMC method is more accurate than the QML method in nine of the twelve cases. For both ϕ and σ_η^2 the MCMC approach is clearly dominant, whilst for μ there are mixed results, but with little difference between the two estimators. Overall, the gains in accuracy for both ϕ and σ_η^2 in particular, provide support in favor of the exact Bayesian approach.

6 An Illustrative Empirical Application

The Bayesian methodology for estimating the SCD model is illustrated using transaction data for BHP Limited, an Australian listed company. Trade durations are

[11] The burn in period of 2000 is chosen by a preliminary visual inspection of the iterates, which have clearly converged after 2000 iterations.

Table 2: Repeated Sampling Performance of the Bayesian (MCMC) and QML Methods. Results Based on 100 Replications of Samples of Size N=10000.

Parameter	True Value	MC Mean		RMSE		Relative RMSE
		MCMC	QML	MCMC	QML	QML/MCMC
ϕ	0.95	0.9489	0.9461	0.0081	0.0135	1.6700
σ_η	0.1	0.0999	0.1036	0.0089	0.0161	1.8106
μ	0.0334	0.0351	0.0355	0.0247	0.0235	0.9515
ϕ	0.95	0.9497	0.9492	0.0041	0.0046	1.1295
σ_η	0.3	0.3006	0.2989	0.0109	0.0125	1.1429
μ	0.0334	0.0272	0.0308	0.0652	0.0579	0.8877
ϕ	0.9	0.8934	0.8594	0.0257	0.1450	5.6498
σ_η	0.1	0.1021	0.1124	0.0158	0.0411	2.6025
μ	0.0334	0.0341	0.0313	0.0134	0.0139	1.0355
ϕ	0.9	0.9000	0.8993	0.0087	0.0090	1.0326
σ_η	0.3	0.2981	0.3003	0.0141	0.0165	1.1676
μ	0.0334	0.0301	0.0302	0.0309	0.0306	0.9907

initially calculated for the month of August 2001, amounting to $N = 48190$ observations. Only trades between 10:20 a.m. and 4:00 p.m. are recorded. Zero trade durations are not included; see also Hautsch and Pohlmeier (2002). This filtering reduces the length of the time series to $N = 27746$ observations. The intraday pattern in the duration data is modelled using a cubic smoothing spline, $g(x_i)$, where the roughness penalty is selected using generalized cross-validation[12]; see also Engle and Russell (1998). The adjusted durations are are then constructed as

$$\widehat{x}_i = \frac{x_i}{g(x_i)}. \tag{50}$$

The full series of adjusted durations, as well as the first 5000 observations, are plotted in Figure 1.

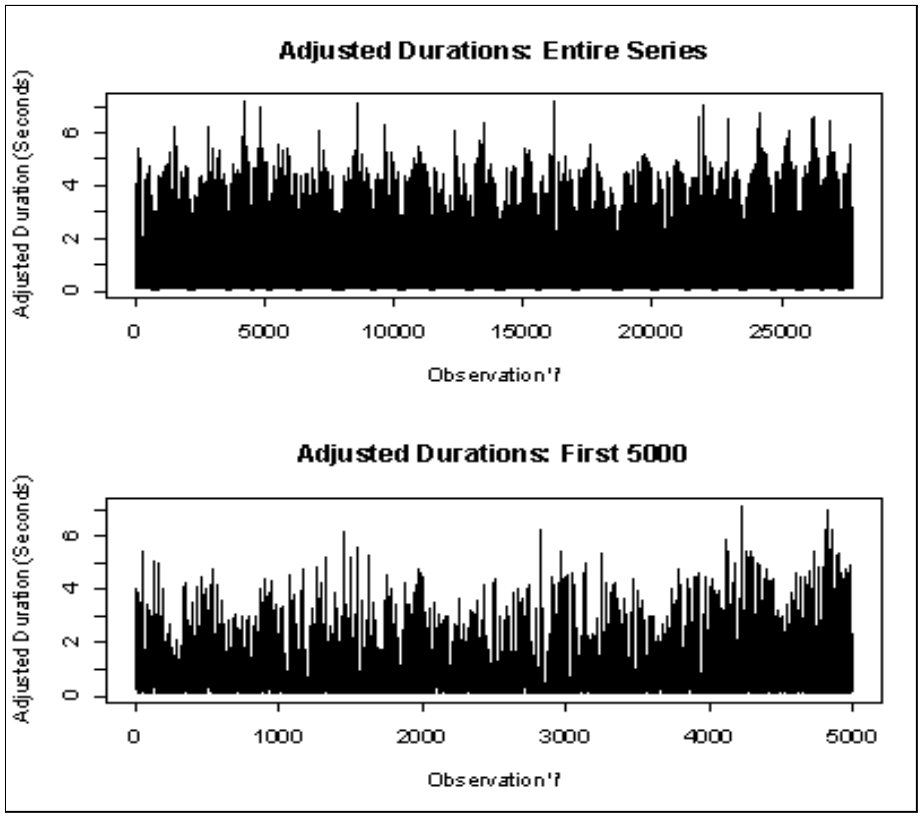

Figure 1. Adjusted Durations for August, 2001.

The changes over time in both the magnitude and variability of the durations are obvious, with clustering in both properties also evident. Given the assumption of

[12]The smoothing spline is estimated using the 'fields' package in the 'R' software.

an exponential distribution, the Markov process for the latent factor ψ_i models time variation in both the conditional mean and the conditional standard deviation of the durations. Hence, this process should capture the observed features in the series.

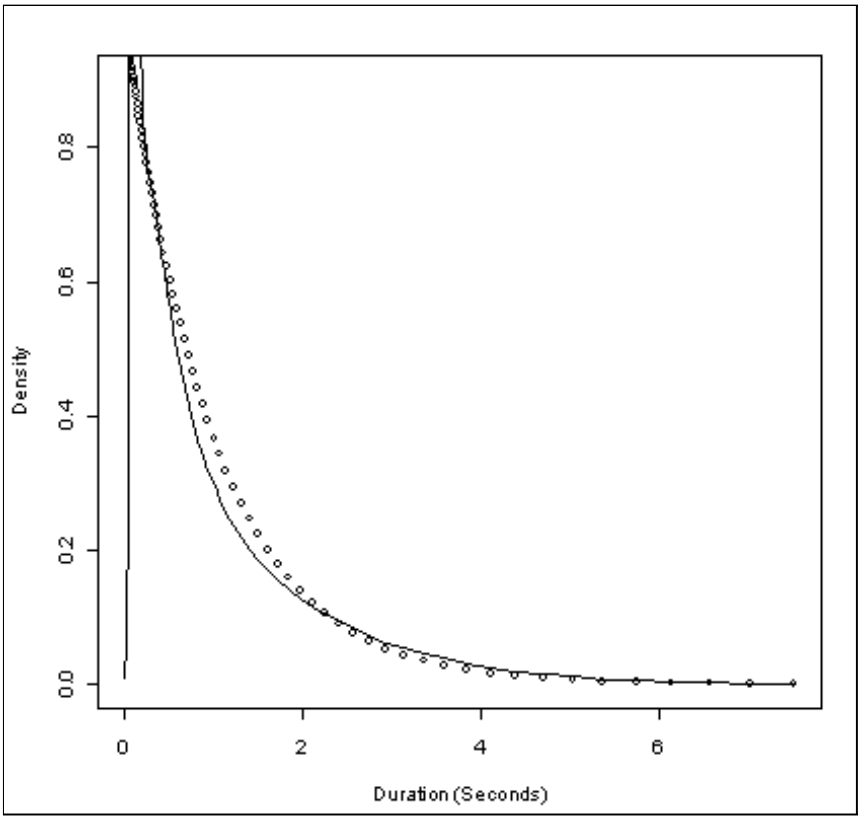

Figure 2: Kernel Density Estimate of the Adjusted Durations (——); Exponential Density (· · · · · ·).

Figure 2 depicts the disparity between a kernel density estimate of the data and an exponential density with a mean equal to the unconditional mean of the adjusted durations. It is clear that the shoulder of the exponential density is narrower than that of the kernel density estimate of the adjusted series. In addition, there is a larger frequency of observations near zero than is associated with the fitted exponential density. The goodness of fit is evaluated using the Kolmogorov-Smirnov goodness of fit test, with the test rejecting the null hypothesis that the observed data comes from the specified exponential distribution.[13]

[13]See Sheskin (2000) for details. The calculated value of the test statistic of 0.065 exceeds the crititical value of 0.008, leading to rejection of the null hypothesis at the 5% significance level.

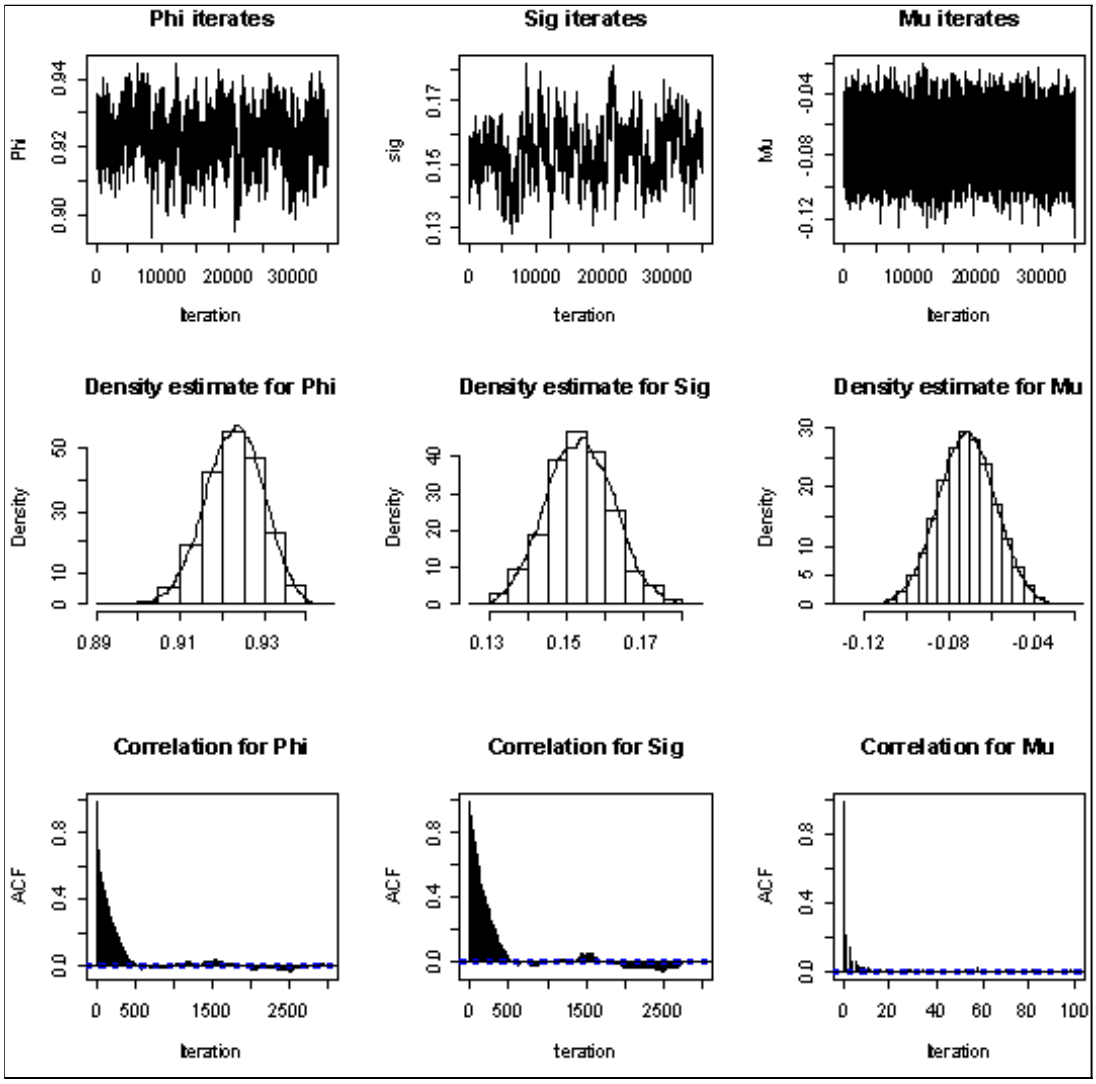

Figure 3. Graphical Output from the MCMC Multi-move Sampler: August 2001 Durations.

Figure 3 contains graphical summaries of the output of the multi-move MCMC sampler. All results are based on 35000 draws after a burn-in of 5000 draws. The hyperparameters σ_r, S_σ, α and β are set to the values previously described in Section 4. The top row contains plots of the iterates from the marginal distributions for ϕ, σ_η and μ respectively. The second row contains estimates of the marginal posterior densities, whilst the autocorrelation function (ACF) of the iterates is presented in the third row for ϕ, σ_η and μ respectively. The plots of the iterates and the ACF functions show a reasonably high amount of correlation for both ϕ and σ_η, whilst there is a very low amount of autocorrelation present in the draws of μ.

Table 3 reports the marginal posterior means and standard deviations (SD), the inefficiency factors and the correlation matrix of the iterates. The high degree of simulation inefficiency indicated by the inefficiency factors for ϕ and σ_η is consistent with the slow decrease in the ACF's for these two parameters. The marginal posterior mean of ϕ indicates that the (logarithm) of the conditional mean (and standard deviation) of durations is quite persistent.

Table 3: Output from the MCMC Multi-move Sampler: August, 2001 Durations

Parameter	Marginal Posterior Mean	Marginal Posterior SD	Inefficiency Factor	Correlation Matrix		
				ϕ	σ_η	μ
ϕ	0.9232	0.0072	80.6119	1.000	-0.8764	0.1225
σ_η	0.1529	0.0092	388.2285	-0.8764	1.000	-0.1589
μ	-0.0716	0.0138	2.4432	0.1225	-0.1589	1.000

Finally, in Figure 4 a plot of the kernel density estimate of the standardised residuals,

$$\widehat{\varepsilon}_i = \frac{\widehat{x}_i}{\exp(\widehat{\psi}_i)}, \tag{51}$$

is compared with an exponential distribution with the same mean as the unconditional mean of the adjusted durations. In (51), the conditional mean at each observational point, $E(x_i|\psi_i) = \exp(\psi_i)$, $i = 1, 2, \ldots, N$, is evaluated using the marginal posterior mean estimate of ψ_i, denoted by $\widehat{\psi}_i$. The graph clearly shows that the conditional distribution provides a very good match for the exponential distribution. The goodness of fit test statistic still rejects the null. However, the test statistic here is smaller than the test statistic given earlier, indicating stronger support for the null hypothesis of an exponential distribution than when the dynamic behaviour in durations is not modelled using (2).[14]

[14]The calculated value of the test statistic is 0.048, thereby still leading to rejection of the null hypothesis at the 5% significance level. However, given the size of the sample, any deviation between the actual and theoretical distributions is likely to be associated with rejection of the theoretical distribution. See Conover (1980) for more on this point.

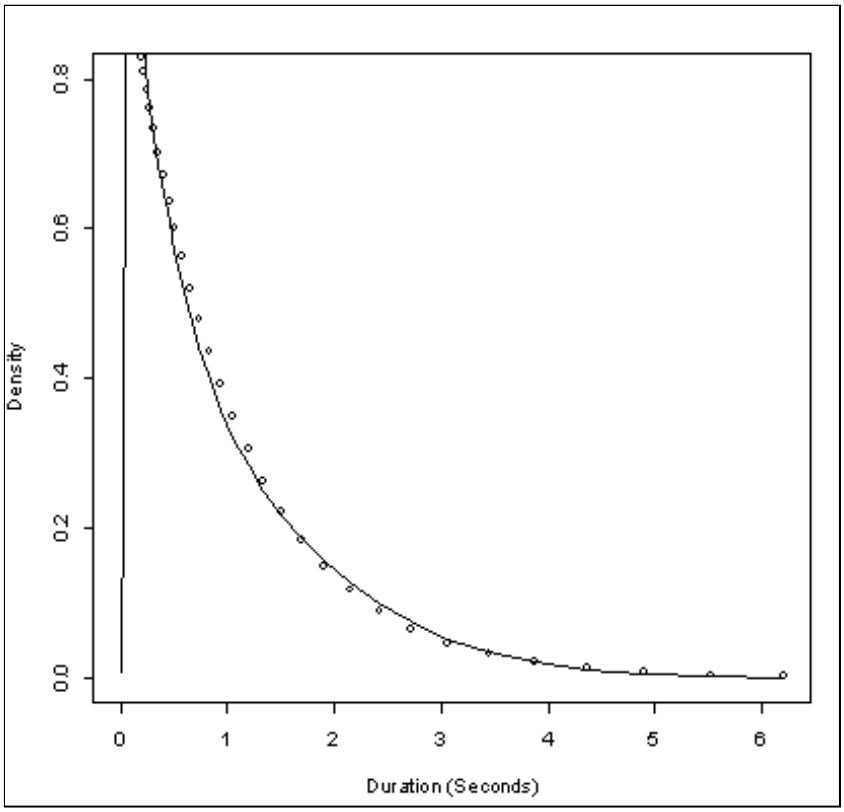

Figure 4. Kernel Density Estimate of the standardised (Adjusted) Durations (——); Exponential Density (· · · · · ·).

7 Conclusions

In this paper an MCMC estimation methodology for the SCD model has been introduced. The methodology exploits the state space representation of the latent factor model for durations. The multi-move sampler proposed has been shown to possess substantially better mixing properties than an alternative single-move sampler. This result corresponds with theoretical and empirical findings in other applications of partially non-Gaussian state space models. The exact MCMC approach has also been compared with the approximate QML procedure using a small-scale Monte Carlo experiment. The results indicate that the MCMC approach tends to outperform the QML approach in terms of both bias and efficiency.

Application of the Bayesian methodology to empirical duration data on BHP trades indicates a high degree of persistence in the conditional mean (and standard

deviation) of durations. Once the data is adjusted for the dynamic behaviour captured by the latent factor process, the distribution of the durations fits the assumed exponential distribution more closely than when such dynamic behaviour is not modelled.

Possible extensions to the methodology include the use of a wider range of distributional assumptions for conditional durations, in particular those that cater for varying degrees of dispersion. The allowance for more complex dynamics in the latent factor process could also be investigated, including the accommodation of long memory dynamic behaviour in durations, as well as the use of more flexible multi-factor models such as that proposed by Ghysels et al (1998). Along the lines suggested by Durbin and Koopman (2001), the estimation of the intraday seasonal pattern could be directly incorporated into the MCMC scheme, rather than the data being filtered in a preliminary step. Of particular interest would be the inclusion of additional regressors in the duration model in order to test various market microstructure hypotheses.

References

[1] Andrews, D. W. K. (1991), "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

[2] Bauwens, L. and Veradas, D. (2002), "The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations," forthcoming in the *Journal of Econometrics.*

[3] Carter, C, K. and Kohn, R. (1994), "On Gibbs Sampling for State Space Models," *Biometrica*, 81, 541-353.

[4] Conover, W. J. (1980), *Practical nonparametric statistics*, second edition, John Wiley and Sons, New York.

[5] De Jong, P. and Shephard, N. (1995), "The Simulation Smoother for Time Series Models," *Biometrica,* 82, 339-350.

[6] Durbin, J. and Koopman, S. J. (2000), "Time Series Analysis of Non-Gaussian Observations Based on State Space Models from both Classical and Bayesian Perspectives," *Journal of the Royal Statistical Society*, 62, 3-56.

[7] Durbin, J. and Koopman, S. J. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press.

[8] Engle, R. F., and Russell, J. R. (1998) "Autoregressive Conditional Duration: A new Approach for Irregularly Spaced Transaction Data," *Econometrica*, 66, 987-1007.

[9] Frühwirth-Schnatter, S. (1994)., "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, 15, 183-202.

[10] Hautsch, N and Pohlmeier, W. (2002), "Econometric Analysis of Financial Transaction Data: Pitfalls and Opportunities," Allgemeines Statistisches Archiv, 86, 5-30.

[11] Ghysels, E., Gourieroux, C. and Jasiak, J. (1998), "Stochastic Volatility Duration Models," *Working Paper*, CREST, Paris. Forthcoming, *Journal of Econometrics.*

[12] Jacquier, E., Polson, N. G. and Rossi, P. E. (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics,* 12, 69-87.

[13] Johnson, N.L., Kotz, S., and Balakrishnan, N. (1994). *Distributions in Statistics: Continuous Univariate Distributions* (Vol. 1, 2nd edition), John Wiley and Sons, New York.

[14] Kim, S,. Shephard, N,. and Chib, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361-393.

[15] Pitt, M. K. and Shephard, N. (1999), "Analytic Convergence Rates and Parameterisation Issues for the Gibbs sampler applied to state space models," *Journal of Time Series Analysis*, 20, 63-85.

[16] Shephard, N. (1994) "Partial Non-Gaussian State Space," *Biometrica*, 81, 115-131.

[17] Shephard, N. and Pitt, M. (1997), "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrica*, 84, 653-667.

[18] Sheskin, D. (2000), *Handbook of Parametric and Nonparametric Statistical Procedures* (2nd edition), Chapman and Hall, Florida.

[19] Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *Annals of Statistics*, 22, 1701-1762.