Bayesian Analysis of the Stochastic Conditional Duration Model

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

March 14, 2005

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. Regressors are included in the latent process model in order to allow additional variables to impact on durations. The sampling scheme employed is a hybrid of the Gibbs and Metropolis Hastings algorithms, with the latent vector sampled in blocks. Candidate draws for the latent process are generated by applying a Kalman filtering and smoothing algorithm to a linear Gaussian approximation of the non-Gaussian state space representation of the model. Monte Carlo sampling experiments demonstrate that the Bayesian method performs better overall than an alternative quasi-maximum likelihood approach. The methodology is illustrated using Australian intraday stock market data, with Bayes factors used to discriminate between different distributional assumptions for durations.

Key Words: Transaction data; Latent variable model; Non-Gaussian state space model; Markov chain Monte Carlo; Kalman filter and simulation smoother.

---

*Corresponding Author: Department of Econometrics and Business Statistics, PO Box 11E, Monash University, Vic., 3800, Australia. Email: gael.martin@buseco.monash.edu.au. Phone: 61 3 9905 1189. Fax: 61 3 9905 5474.

1 Introduction

The increased 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 function of a stochastic latent process, with the conditional distribution of durations defined on a positive support. As such, the contrast between the two specifications is similar to the contrast between the generalized autoregressive conditional heteroscedasticity (GARCH) and stochastic volatility (SV) frameworks for capturing the conditional volatility of financial returns.

Whilst several modifications of the original ACD specification have been put forward (see Bauwens *et al.*, 2004, for a recent summary), the literature that focusses on the SCD model is less advanced. Although Durbin and Koopman (2001) suggest the use of a latent variable model for durations, they do not develop the idea further, with the first published analysis of the model occurring in Bauwens and Veredas (2004). 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. Variants of the SCD model include a two-factor version of the model developed in Ghysels, Gourieroux and Jasiak (2004) (first appearing in draft form in 1998) and the

SCD model with 'leverage effect' outlined in Feng, Jiang and Song (2004), with these models estimated using simulated method of moments and Monte Carlo maximum likelihood respectively. A comparison of the forecasting performance of a range of duration models, including the SCD and ACD specifications, is presented in Bauwens *et al.*

As is the case with the SV model, the 'parameter-driven' SCD model presents a potentially more complex estimation problem than its 'observation-driven' alternative, by augmenting the set of unknowns with a set of unobservable latent variables. However, as is also argued in the case of the SV/GARCH dichotomy, with the advent of more sophisticated computing resources, the extra computational burden associated with that complexity is no longer such an issue. Moreover, as is highlighted in Ghysels *et al.* (2004), the ability of the latent variable framework, in which the SCD model is nested, to be extended to the multi-factor case is a crucial advantage over the ACD framework, in which the dynamic behavior in all higher order moments is necessarily tied to the dynamic behavior in the conditional mean. This is a particularly important point as it relates to the first two moments of durations, given their association with the separate features of liquidity and liquidity risk respectively. Consequently, the SCD model should be viewed both as a serious competitor to the ACD model and as the starting point for more sophisticated modelling of durations data.

In this paper, a Bayesian methodology for estimating the SCD model is presented. The unobservable latent variables are integrated out of the joint posterior distribution 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 neighborhood 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; see, for example, Carter and Kohn (1994), Frühwirth-Schnatter (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002). The latent variable draws are then accepted with a particular probability, according to the MH algorithm. The MH subchains associated with the latent variable 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 methodology presented here is very general. In particular, with relatively minor alterations, the MCMC algorithm can accommodate a range of conditional distributions. This means that we can readily allow for alternative specifications for the durations data and use Bayes factors to choose between the different models. In addition, this flexibility means that the methodology is applicable beyond the durations context. For example, in the case of financial returns, the use of an appropriate conditional distribution in combination with a latent variable model for the variance, constitutes a valid SV specification. Hence, the methodology that we present is an alternative to existing methods for estimating SV models; see, for example, Jacquier, Polson and Rossi (1994, 2004), Kim, Shephard and Chib (1998), Meyer and Yu (2000) and Liesenfeld and Richard (2003). Furthermore, the algorithm is equally applicable to different data types, including discrete and binary data. As such, in the context of transactions data, it could be used to estimate, for example, the dynamic behavior in tick count or binary price change data; see Rydberg and Shephard (2003). Whilst this level of generality is also a feature, in principle, of the Durbin and Koopman (2000, 2001) importance sampling-based algorithm, as is noted below, that algorithm is likely to be computationally inefficient compared to MCMC when applied to the large data sets associated with transaction data.

The structure of the paper is as follows. Section 2 describes the SCD model, including the way in which microstructure effects can be readily incorporated by augmenting the state equation with observed regressors. Sec-

4

tion 3 then outlines the MCMC scheme, with details of the approximation used in the production of a candidate distribution for the vector of latent variables provided in Appendix I. In Section 4, the sampling behavior of the Bayesian estimation method is compared with that of the QML approach adopted by Bauwens and Veredas (2004), via a 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, in general, superior to the QML approach, in terms of both bias and efficiency. Two empirical illustrations of the Bayesian method are then described in Section 5, based on durations data for two Australian firms, Broken Hill Proprietary (BHP) Limited and News Corporation (NCP), for the month of August 2001. In both cases three alternative conditional distributions are specified for durations, with Bayes factors, calculated using the Savage-Dickey density ratio, finding in favour of the gamma distribution for the BHP data and the Weibull distribution in the case of the NCP data. Trade volume is found to have no impact on durations for BHP yet does have an impact in the case of NCP. Some conclusions and proposed extensions are given in Section 6.

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 $x_i = \tau_i - \tau_{i-1}$. Defining $\mathbf{x} = (x_1, x_2, \ldots, x_N)'$ as the $N-$ dimensional vector of durations, the SCD model for $\mathbf{x}$ is defined as

$$p(\mathbf{x}|\boldsymbol{\psi},\lambda) \;=\; \prod_{i=1}^{N} p(x_i|\psi_i,\lambda) \tag{1}$$

$$\psi_i \;=\; \mathbf{W}'_{i-1}\boldsymbol{\delta} + \phi\psi_{i-1} + \sigma_\eta\eta_i, \tag{2}$$

$$\eta_i \;\sim\; i.i.d.N(0,1), \tag{3}$$

for $i = 2, \dots, N$, where x_i, conditional on ψ_i, is assumed to be independent of η_i for all i. The latent variable process in (2) is assumed to have finite variance, with $|\phi| < 1$. The $(1 \times k)$ vector $\mathbf{W}'_i$ contains the ith row of the observed $(N \times k)$ regressor matrix $\mathbf{W}$, with $\boldsymbol{\delta}$ the associated $(k \times 1)$ vector of coefficients. The assumption of $|\phi| < 1$ implies that the initial state has a marginal distribution given by

$$\psi_1 \sim N\left(\frac{\mathbf{W}'_1\boldsymbol{\delta}}{1-\phi}, \frac{\sigma_\eta^2}{1-\phi^2}\right), \tag{4}$$

where $\mathbf{W}'_1$ is the $(1 \times k)$ row vector of initial values for the regressors. The decomposition in (1) implies that, conditional on the N-dimensional state vector $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_N)'$, the durations are identically and independently distributed ($i.i.d.$), such that dependence in the durations is generated solely through the dynamic latent process in (2). The probability density function (pdf) in (1) is assumed to have positive support, with the distributions entertained in the paper being exponential, Weibull and gamma respectively. The scalar λ assumes the role of the additional distributional parameter in the case of the latter two distributions.

3 Markov Chain Monte Carlo Estimation

3.1 Discussion

The similarity between the SCD model and the SV model provides some motivation for the use of an MCMC approach to estimation. Jacquier *et al.* (1994)

show that an MCMC algorithm applied to a particular SV specification is superior, in terms of both bias and efficiency, to both maximum likelihood and generalized method of moment approaches, both of which have asymptotic justification only. An advantage of a Gibbs-based MCMC algorithm over the other finite sample simulation methods, such as the importance sampling algorithms of Daniellson (1994), Sandmann and Koopman (1998), Durbin and Koopman (2000, 2001) and Liesenfeld and Richard (2003), derives from the ability of the former method to allow the state vector to be treated in sub-blocks, with the draw of any one sub-block conditioned on the previous draw of the remaining sub-blocks. As is demonstrated in Pitt (2000) in the context of the SV model, this approach, compared with sampling the entire state vector in one block, can have a large impact on the ability of the simulation scheme to traverse the joint space of the latent states. This point has particular weight in the SCD context, given the sample sizes and, hence, the dimension of the state vector associated with transaction data. Extension to a multi-factor version of the SCD would highlight this issue even further. Similarly, the alternative grid-based maximum likelihood method presented by Fridman and Harris (1998) for an SV model, although a potential contender in the single-factor SCD case, would suffer the usual curse of dimensionality common to algorithms that use deterministic integration methods, if extended to a multi-factor version of the SCD model.

Results from the SV literature also help to motivate the particular form of MCMC algorithm specified: namely, a multi-move sampler. Specifically, drawing the latent variables in blocks of size greater than one, rather than individually, exploits the correlation between the latent states, thereby enabling more efficient coverage of the joint space of the latent variables; see Shephard (1994) for further discussion on this point. For numerical evidence of the superiority of the multi-move over the single-move sampler, in the case of an SV model, see Kim, Shephard and Chib (1998). Comparable results in the case of the SCD model are also available from the authors upon request.

7

3.2 A Multi-Move MCMC Algorithm

Defining the vector of unknown parameters in the latent process in (2) as $\boldsymbol{\theta} = (\boldsymbol{\delta}', \phi, \sigma_\eta)'$, the joint posterior for the full set of unknowns in the SCD model is given by

$$p(\boldsymbol{\psi}, \boldsymbol{\theta}, \lambda | \mathbf{x}, \mathbf{W}) \propto p(\mathbf{x}|\boldsymbol{\psi}, \lambda) \times p(\boldsymbol{\psi}|\mathbf{W}, \boldsymbol{\theta}) \times p(\boldsymbol{\theta}) \times p(\lambda), \qquad (5)$$

where $p(\boldsymbol{\psi}|\mathbf{W}, \boldsymbol{\theta})$ denotes the joint pdf of $\boldsymbol{\psi}$ conditional on $\mathbf{W}$ and $\boldsymbol{\theta}$, and $p(\boldsymbol{\theta})$ and $p(\lambda)$ are the prior pdfs for $\boldsymbol{\theta}$ and λ, with $\boldsymbol{\theta}$ and λ assumed to be *a priori* independent. The joint pdf $p(\mathbf{x}|\boldsymbol{\psi}, \lambda)$ is as defined in (1), whilst from the state recursion in (2) and the distributional specification in (3), it follows that

$$p(\boldsymbol{\psi}|\mathbf{W}, \boldsymbol{\theta}) = \left\{ \prod_{i=2}^{N} p(\psi_i|\mathbf{W}_{i-1}, \psi_{i-1}, \boldsymbol{\theta}) \right\} \times p(\psi_1|\mathbf{W}_1, \boldsymbol{\theta}), \qquad (6)$$

where

$$p(\psi_i|\mathbf{W}_{i-1}, \psi_{i-1}, \boldsymbol{\theta}) \propto \exp\{\frac{-1}{2\sigma_\eta^2} \left[\psi_i - \mathbf{W}_{i-1}'\boldsymbol{\delta} - \phi\psi_{i-1} \right]^2\}, \qquad (7)$$

for $i = 2, \ldots, N$, and $p(\psi_1|\mathbf{W}_1, \boldsymbol{\theta})$ follows from (4).

The specific form of $p(x_i|\psi_i, \lambda)$ in (1) depends on the distribution assumed to generate the durations, conditional on the latent states. The form of $p(x_i|\psi_i, \lambda)$ corresponding to each of the conditional distributions to be considered in the paper is given in second column of Table 1. For each distribution the conditional mean is equal to $\exp(\psi_i)$. All higher order conditional moments are also functionally dependent on ψ_i; see Johnson, Kotz and Balakrishnan (1994) for all relevant formulae. Note that for the exponential distribution λ is a redundant parameter, whilst $\lambda = \gamma > 0$ for the Weibull distribution and $\lambda = \zeta > 0$ for the gamma distribution. These specifications for λ are given in the final column of the table. In the cases were λ is applicable, a lognormal prior is specified for λ in order to ensure positivity. Standard priors for σ_η, ϕ, and $\boldsymbol{\delta}$ are used, with details given below; see also

Table 1: Densities of Alternative Conditional Distributions for Durations

Distribution	$p(x_i|\psi_i,\lambda)$	λ
Exponential	$\exp(-\psi_i)\exp\left\{-x_i\exp\left(-\psi_i\right)\right\}$	n.a.[a]
Weibull	$\gamma\left(\frac{\Gamma(1+\gamma^{-1})}{\exp(\psi_i)}\right)^{\gamma}x_i^{(\gamma-1)}\exp\left\{-\left[\frac{x_i}{\exp(\psi_i)}\Gamma\left(1+\gamma^{-1}\right)\right]^{\gamma}\right\}$	γ
Gamma	$\left(\frac{\zeta}{\exp(\psi_i)}\right)^{\zeta}x_i^{(\zeta-1)}\exp\left\{-\frac{x_i\zeta}{\exp(\psi_i)}\right\}/\Gamma\left(\zeta\right)$	ζ

(a) n.a. = not applicable

Kim, Shephard and Chib (1998) for justification of the prior assumptions adopted here.

In employing a Gibbs-based MCMC sampler, simulated draws from the full conditional distribution relating to each block of unknowns must be obtained. As already noted, in the multi-move sampler adopted here all of the latent states are sampled in blocks of size greater than one. Since the SCD model is a non-Gaussian state space model, the difficulty is in finding a good candidate density for the block of states. One approach outlined by Shephard (1994) and Carter and Kohn (1994), and implemented by Kim, Shephard and Chib (1998) for the SV model, is to approximate the non-Gaussian density in the measurement equation, (1), by 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 on a more general approximation method. The approximation employed is that used by Durbin and Koopman (2000, 2001) in an importance sampling context. Specifically, the non-Gaussian density for each observation is approximated by a Gaussian density with the same mode, with the curvature of the ap-

proximating Gaussian density equated to that of the non-Gaussian density at the mode. The approximation is performed via an iterative Kalman filter, details of which are provided in Appendix I.

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

1. Initialize $\boldsymbol{\psi}$ and λ.

2. Sample $\boldsymbol{\theta}|\mathbf{x}, \mathbf{W}, \boldsymbol{\psi},\lambda$.

3. Sample $\lambda|\mathbf{x}, \mathbf{W}, \boldsymbol{\psi}, \boldsymbol{\theta}$.

4. Sample $\boldsymbol{\psi}|\mathbf{x}, \mathbf{W}, \boldsymbol{\theta}, \lambda$, where $\boldsymbol{\psi}$ is broken up into blocks of size greater than one.

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

The three conditional posteriors of $\boldsymbol{\theta}$, λ and $\boldsymbol{\psi}$ respectively, including the sampling algorithm required to draw from each conditional, are detailed below.

3.2.1   Sampling $\boldsymbol{\theta}$

Given (5), the full conditional for $\boldsymbol{\theta} = (\boldsymbol{\delta}', \phi, \sigma_\eta)'$ is given by

$$p(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\psi}, \mathbf{W}) \propto p(\boldsymbol{\psi}|\mathbf{W}, \boldsymbol{\theta}) \times p(\boldsymbol{\theta}), \tag{8}$$

where $p(\boldsymbol{\psi}|\mathbf{W}, \boldsymbol{\theta})$ is defined as in (6) and $p(\boldsymbol{\theta}) = p(\boldsymbol{\delta}) \times p(\phi) \times p(\sigma_\eta)$, assuming *a priori* independence between the elements of $\boldsymbol{\theta}$. In the empirical analysis, an inverted-gamma prior for σ_η is adopted, such that

$$\sigma_\eta \sim IG\left(\frac{\sigma_r}{2}, \frac{S_\sigma}{2}\right), \tag{9}$$

with σ_r and S_σ representing hyperparameters; see Zellner (1996). The prior for ϕ is derived from the beta density function by extending the density over the (-1,1) interval and is thus given by

$$p(\phi) \propto \left\{\frac{1+\phi}{2}\right\}^{\phi_1-1} \left\{\frac{1-\phi}{2}\right\}^{\phi_2-1}, \tag{10}$$

where $\phi_1, \phi_2 > 0.5$ are shape parameters for the resultant *stretched* beta distribution. The prior in (10) imposes the stationarity restriction on ϕ, whilst ϕ_1 and ϕ_2 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. A uniform prior over $\mathbb{R}^k$ is defined for $\boldsymbol{\delta}$.

Given the use of the beta prior for ϕ and the density in (4) for the initial state, $p(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\psi}, \mathbf{W})$ does not have a standard form from which draws of $\boldsymbol{\theta}$ can be made directly. As such, we use an MH algorithm to draw $\boldsymbol{\theta}$ indirectly, via a candidate density specified as follows. Define

$$\mathbf{Z} = \begin{bmatrix} w_{1,1} & \ldots & w_{1,k} & \psi_1 \\ \vdots & & \vdots & \vdots \\ w_{N-1,1} & \ldots & w_{N-1,k} & \psi_{N-1} \end{bmatrix},$$

where $w_{i,j}$ is the i, j^{th} element in $\mathbf{W}$. Further defining $\boldsymbol{\beta} = (\boldsymbol{\delta}', \phi)'$, we express (2) as

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{u}, \tag{11}$$

where $\mathbf{y} = (\psi_2, \ldots, \psi_N)'$ and $\mathbf{u} \sim N(0, \sigma_\eta^2 \mathbf{I}_{N-1})$. Adopting a standard non-informative prior for $(\boldsymbol{\beta}', \sigma_\eta)'$ of the form $p(\boldsymbol{\beta}', \sigma_\eta) \propto \frac{1}{\sigma_\eta}$, the candidate for $\boldsymbol{\theta}$ is given by the normal-gamma distribution for $(\boldsymbol{\beta}', \sigma_\eta)'$ that results from the normal regression structure of (11); see, for example, Zellner (1996).

The steps of the MH algorithm for the jth iteration of the Gibbs sampler are as follows:

11

1. Specify $\boldsymbol{\theta}^{(j-1)}$ as an initial value for the algorithm.

2. Draw a candidate $\boldsymbol{\theta}^*$ from the normal-gamma candidate distribution.

3. Accept $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^*$, with probability equal to $\min\left(1, \frac{w(\boldsymbol{\theta}^*|.)}{w(\boldsymbol{\theta}^{(j-1)}|.)}\right)$, where $w(\boldsymbol{\theta}|.) = \frac{p(\boldsymbol{\theta}|.)}{q(\boldsymbol{\theta}|.)}$, $p(\boldsymbol{\theta}|.)$ denotes the conditional posterior in (8), evaluated at the relevant argument, and $q(\boldsymbol{\theta}|.)$ is the corresponding candidate ordinate.

4. Otherwise accept $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^{(j-1)}$.

3.2.2   Sampling λ

In the two cases in which λ features in the conditional distribution (i.e. for the Weibull and gamma distributions), the restrictions $\gamma > 0$ and $\zeta > 0$ are imposed via the adoption of a lognormal prior distribution. Since the same form of algorithm is used for drawing λ whether $\lambda = \gamma$ or $\lambda = \zeta$, the generic sampling scheme for λ is described here, with the lognormal prior density denoted by $p(\lambda)$. The full conditional posterior distribution for λ, with pdf

$$p(\lambda|\mathbf{x}, \mathbf{W}, \boldsymbol{\psi}, \boldsymbol{\theta}) \propto p(\mathbf{x}|\boldsymbol{\psi}, \lambda) \times p(\lambda), \tag{12}$$

has no closed-form representation. The orientational bias Monte Carlo (OBMC) algorithm is used to draw λ; see Liu (2001) for details. The version of the OBMC algorithm applied here can be viewed as a generalization of the random walk MH algorithm in which multiple candidate draws are taken. The steps of the algorithm inserted at iteration j of the Gibbs chain are:

1. Specify $\lambda^{(j-1)}$ as an initial value for the algorithm.

2. Draw L candidates $\lambda_1^*, \ldots, \lambda_L^*$ as $\lambda^{(j-1)} + \varepsilon_l$, where ε_l, $l = 1, 2, \ldots, L$, is a draw from a normal distribution with a mean of 0 and a variance of σ_ε^2.

3. Construct a probability mass function (pmf) by assigning to each $\lambda_1^*, \ldots, \lambda_L^*$ a probability proportional to $p(\lambda_l|.)$, $l = 1, 2, \ldots, L$, where $p(\lambda_l|.)$ denotes the pdf in (12) evaluated at the relevant argument. Select λ^{**} randomly from this distribution.

4. Draw $L-1$ reference points $r_1, \ldots, r_{L-1}$ from $\lambda^{**} + \varepsilon_l$ and set $r_L = \lambda^{(j-1)}$.

5. Accept $\lambda^{(j)} = \lambda^{**}$ with probability equal to $\min\left(1, \frac{p(\lambda_1^*|.) + \cdots + p(\lambda_L^*|.)}{p(r_1|.) + \cdots + p(r_L|.)}\right)$

6. Otherwise accept $\lambda^{(j)} = \lambda^{(j-1)}$.

3.2.3 Sampling ψ

A blocking scheme for ψ is defined such that

$$\psi = (\psi_1 \ldots \psi_{k_1}, \psi_{k_1+1} \ldots \psi_{k_2}, \psi_{k_2+1} \ldots, \ldots \psi_{k_K}, \psi_{k_K+1} \ldots \psi_N)',$$

where $k_1, k_2, \ldots, k_K$ are the knot points separating the $(K+1)$ blocks. The knots at each iteration are selected stochastically following Shephard and Pitt (1997) and Elerian, Chib, and Shephard (2001), with the expected block size, ϖ, used as a tuning parameter. The selection of ϖ is a compromise between the simulation efficiency gains of using a larger average block size and the higher associated rejection rate in the algorithm.

Defining $\psi_{B_l} = (\psi_{k_{(l-1)}+1} \ldots \psi_{k_l})'$, $l = 1, \ldots, K+1$, 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 $\psi_{B_1}^{(j)}|\mathbf{x}, \mathbf{W}, \psi_{B_2}^{(j-1)}, \boldsymbol{\theta}^{(j)}, \lambda^{(j)}$.

2. Sample $\psi_{B_l}^{(j)}|\mathbf{x}, \mathbf{W}, \psi_{B_{l-1}}^{(j)}, \psi_{B_{l+1}}^{(j-1)}, \boldsymbol{\theta}^{(j)}, \lambda^{(j)}$, for $l = 2, 3, ..., K$.

3. Sample $\psi_{B_{K+1}}^{(j)}|\mathbf{x}, \mathbf{W}, \psi_{B_K}^{(j)}, \boldsymbol{\theta}^{(j)}, \lambda^{(j)}$.

For each block ψ_{B_l} a linear Gaussian approximation to the non-Gaussian state space model represented by (1) and (2) is produced. This approximating

13

model comprises (2), in combination with a linear Gaussian approximating measurement equation, defined as

$$\widetilde{x}_i = \psi_i + \widetilde{\varepsilon}_i, \tag{13}$$

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 ψ_i. All details of the iterative procedure used to derive $\widetilde{x}_i$ and $\widetilde{H}_i$ are provided in Appendix I; see also Durbin and Koopman (2000, 2001). The linear Gaussian approximation then serves as a candidate model from which a candidate draw for $\boldsymbol{\psi}_{B_l}$ is produced. In brief, the approximating model is derived in such a way that the mode of the candidate density associated with this model, $q(\boldsymbol{\psi}_{B_l} | \mathbf{x}, \mathbf{W}, \boldsymbol{\psi}_{B_{l-1}}, \boldsymbol{\psi}_{B_{l+1}}, \boldsymbol{\theta}, \lambda)$, is equal to the mode of the actual conditional posterior for $\boldsymbol{\psi}_{B_l}$ as based on the non-Gaussian model, $p(\boldsymbol{\psi}_{B_l} | \mathbf{x}, \mathbf{W}, \boldsymbol{\psi}_{B_{l-1}}, \boldsymbol{\psi}_{B_{l+1}}, \boldsymbol{\theta}, \lambda)$. Drawing from $q(\boldsymbol{\psi}_{B_l} | \mathbf{x}, \mathbf{W}, \boldsymbol{\psi}_{B_{l-1}}, \boldsymbol{\psi}_{B_{l+1}}, \boldsymbol{\theta}, \lambda)$ is implemented through the use of the Kalman filter and simulation smoother.

The steps taken to draw $\boldsymbol{\psi}_{B_l}$, $l = 1, 2, \ldots, K + 1$, at iteration j of the Gibbs sampler, are:

1. Initialize $\boldsymbol{\psi}_{B_l}$.

2. Run the iterative procedure described in Appendix I to produce $\widetilde{H}$ and $\widetilde{x}$.

3. Define the approximating measurement equation as (13), for the $(k_l - k_{(l-1)})$ elements in the block $\boldsymbol{\psi}_{B_l}$.

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

5. Accept $\boldsymbol{\psi}_{B_l}^{(j)} = \boldsymbol{\psi}_{B_l}^*$, with probability equal to $\min\left(1, \frac{w(\mathbf{x}_{B_l} | \boldsymbol{\psi}_{B_l}^*)}{w(\mathbf{x}_{B_l} | \boldsymbol{\psi}_{B_l}^{(j-1)})}\right)$, where

$w(\mathbf{x}_{B_l} | \boldsymbol{\psi}_{B_l}) = \frac{p(\mathbf{x}_{B_l} | \boldsymbol{\psi}_{B_l})}{q(\mathbf{x}_{B_l} | \boldsymbol{\psi}_{B_l})}$ and $\mathbf{x}_{B_l}$ denotes the block of $\mathbf{x}$ corresponding to $\boldsymbol{\psi}_{B_l}$. Note that

$$\frac{p(\boldsymbol{\psi}_{B_l} | \mathbf{x}, \mathbf{W}, \boldsymbol{\psi}_{B_{l-1}}, \boldsymbol{\psi}_{B_{l+1}}, \boldsymbol{\theta}, \lambda)}{q(\boldsymbol{\psi}_{B_l} | \mathbf{x}, \mathbf{W}, \boldsymbol{\psi}_{B_{l-1}}, \boldsymbol{\psi}_{B_{l+1}}, \boldsymbol{\theta}, \lambda)} \propto \frac{p(\mathbf{x}_{B_l} | \boldsymbol{\psi}_{B_l}) p(\boldsymbol{\psi}_{B_l} | \mathbf{W}, \boldsymbol{\psi}_{B_{l-1}}, \boldsymbol{\psi}_{B_{l+1}}, \boldsymbol{\theta}, \lambda)}{q(\mathbf{x}_{B_l} | \boldsymbol{\psi}_{B_l}) q(\boldsymbol{\psi}_{B_l} | \mathbf{W}, \boldsymbol{\psi}_{B_{l-1}}, \boldsymbol{\psi}_{B_{l+1}}, \boldsymbol{\theta}, \lambda)}$$

$$\propto \frac{p(\mathbf{x}_{B_l} | \boldsymbol{\psi}_{B_l})}{q(\mathbf{x}_{B_l} | \boldsymbol{\psi}_{B_l})},$$

as $p(\boldsymbol{\psi}_{B_l} | \mathbf{W}, \boldsymbol{\psi}_{B_{l-1}}, \boldsymbol{\psi}_{B_{l+1}}, \boldsymbol{\theta}, \lambda) = q(\boldsymbol{\psi}_{B_l} | \mathbf{W}, \boldsymbol{\psi}_{B_{l-1}}, \boldsymbol{\psi}_{B_{l+1}}, \boldsymbol{\theta}, \lambda)$, due to the fact that the latent process is the same under both the actual and candidate models.

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

4 Sampling Experiments

Monte Carlo experiments are 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 Veredas (2004) in their analysis of the SCD model. In these experiments we only consider the case where the conditional distribution is exponential. We also omit regressors from the state equation so that the scalar parameter δ represents the intercept in (2). Earlier research by Jacquier *et al.* (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 a Bayesian MCMC method. With a larger sample size of $N = 2000$, Jacquier *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 experiments 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 $\phi = \{0.95, 0.90\}$,

15

$\sigma_\eta = \{0.1,\ 0.3\}$ and $\delta = 0.033$. These parameter settings correspond to a range of values that are reasonably representative of some of the estimated parameter values reported in Section 5, and in the empirical study undertaken by Bauwens and Veredas (2004). In order to render the comparison between the Bayesian and classical methods as fair as possible, we report results based on diffuse priors for the parameters. Specifically, uniform priors are specified for δ and ϕ, with the prior on ϕ truncated to the $(-1,1)$ interval. The hyperparameters σ_r and S_σ in (9) are set to 1.0001 and 0.01 respectively, implying a prior mean of 0.18 and variance of 99.97 for σ_η. For the sampling experiments, the stochastic knot algorithm is based on Pitt and Shephard (1997), with an average block size of $\varpi = 20$. Bayesian point estimates of the parameters are produced using the marginal posterior means estimated from the draws of the MCMC algorithm.

Re-expressing the measurement equation in the exponential case as

$$x_i = \exp(\psi_i)\varepsilon_i, \tag{14}$$

for $i = 1, 2, \ldots, N$, where ε_i is an $i.i.d$ exponential variable with mean (and variance) equal to one, the QML approach is based on a logarithmic transformation of (14), which produces a linear relationship between the transformed durations and the state. The transformed measurement equation has the following form,

$$\ln(x_i) = c_i + \psi_i + \zeta_i, \tag{15}$$

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 $i.i.d.N(0, Var[\ln \varepsilon_i])$. With ε_i assumed here to be exponentially distributed with a mean of one, $E[\ln \varepsilon_i] = -\gamma^*$, with $\gamma^* =$ Euler's constant ≈ 0.5772, and $Var[\ln \varepsilon_i] = \frac{\pi^2}{6}$; see Johnson *et al.* (1994). Standard asymptotic theory implies that the QML estimator will be consistent yet inefficient. This corresponds with the simulation findings of Jacquier et al (1994) cited earlier for the SV context,

Table 2: Repeated Sampling Performance of the Bayesian MCMC and QML Methods. Results are Based on 10000 Replications of Samples of Size N=10000. Diffuse Priors are used in the Bayesian Algorithm

Parameter	True Value	MC Mean		RMSE		Relative RMSE
		MCMC	QML	MCMC	QML	QML/MCMC
ϕ	0.950	0.948	0.947	0.009	0.012	1.333
σ_η	0.100	0.101	0.102	0.010	0.015	1.500
δ	0.033	0.033	0.033	0.023	0.024	1.044
ϕ	0.950	0.950	0.949	0.004	0.005	1.250
σ_η	0.30	0.300	0.301	0.010	0.013	1.300
δ	0.033	0.034	0.033	0.061	0.061	1.000
ϕ	0.900	0.891	0.851	0.028	0.161	5.750
σ_η	0.100	0.103	0.112	0.017	0.044	2.588
δ	0.033	0.033	0.033	0.015	0.016	1.067
ϕ	0.900	0.900	0.889	0.008	0.012	1.500
σ_η	0.300	0.300	0.301	0.013	0.020	1.539
δ	0.033	0.033	0.033	0.032	0.033	1.031

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 used for each parameter setting is 10000. 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. The burn-in of 2000 iterations is sufficient to establish convergence, as gauged by a visual assessment of the stationarity in the time series of iterates for all parameters. The results are reported in Table 2. The true parameter values are shown in the second column of the table 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 method still shows clear bias in the estimation of ϕ and σ_η, 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 as, or more accurate than the QML method in all twelve cases. To assess the robustness of the results to prior specification, the Bayesian algorithm was also run using the more informative priors on ϕ and σ_η that are used in the empirical exercise in Section 5. The results (not reported) are very similar to those based on the diffuse priors. Hence, the sampling experiments provide quite strong support in favor of the exact Bayesian approach.

5 Illustrative Empirical Applications

5.1 Data Description

The Bayesian methodology for estimating the SCD model is illustrated using transaction data for two Australian listed companies: BHP and NCP. Trade durations are initially calculated for the month of August 2001, amounting to $N = 48190$ and $N = 17691$ observations for BHP and NCP respectively. Only trades between 10:20 a.m. and 4:00 p.m. are recorded. Zero trade durations are not included, following Engle and Russell (1998). This filtering reduces the length of the time series to $N = 27746$ observations for BHP and $N = 13832$ observations for NCP. 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. The smoothing spline is estimated using the 'fields' package in the 'R' software. The adjusted

durations are then constructed as

$$\widehat{x}_i = \frac{x_i}{g\left(x_i\right)}, \tag{16}$$

again following Engle and Russell (1998).

For both data sets trading volume is incorporated as a regressor in accordance with market microstructure theory. For example, from Easley and O'Hara (1987, 1992) it follows that large trades indicate the presence of informed traders in the market. The presence of such informed traders, viewed as evidence of the existence of new 'information', in turn stimulates general trading, thereby leading to a negative relationship between durations and trading volume. Note that as with the duration series, the intraday pattern in trading volume is estimated using a cubic smoothing spline and removed assuming the same multiplicative type of relationship.

5.2 Empirical Results

Tables 3 and 4 contain the empirical results for the BHP and NCP data respectively. The results are based on 50000 iterations that were stored after discarding the initial 5000 iterations. The latter represents a conservative choice of burn-in period, with the time series of iterates for all parameters well and truly indicating convergence by this stage in the chain. The hyperparameters σ_r and S_σ in (9) are set to 3 and 0.03 respectively, implying a prior mean of 0.12 and variance 0.0017 for σ_η. The hyperparameters ϕ_1 and ϕ_2 in (10) are set to 15 and 1.5 respectively, implying a prior mean of 0.82 and a variance of 0.02 for ϕ. Uniform priors are adopted for δ_1 and δ_2. The prior mean for both γ and ζ is 1, whilst the prior variance is approximately 6.4. In the OBMC algorithm used to estimate γ and ζ, the tuning parameters L and σ_ε are set to 5 and 0.1 respectively. The value for the tuning parameter ϖ is distribution-specific and is reported at the bottom of Tables 3 and 4. Selection of values for all tuning parameters is based on a preliminary analysis of a subsample of the data, using just a few thousand

iterations, with the inefficiency factor (see Appendix II and Chib and Greenberg, 1996) used as the selection criterion. In the tables, for each parameter we report consecutively: the marginal posterior mean, the 95% highest posterior density (HPD) interval (in parentheses), the MC standard error and the inefficiency factor. The MC standard error is calculated using standard time series techniques discussed in Andrews (1991) and Chib and Greenberg (1996). The parameters δ_1 and δ_2 are respectively the intercept coefficient and trading volume coefficient in the state equation (2).

In the lower panel of each table we also report the Bayes factor for each model relative to the model based on the exponential conditional distribution. As the exponential distribution is nested in both the Weibull and the gamma distributions, associated with the values of $\gamma = 1$ and $\zeta = 1$ respectively, the Savage-Dickey density ratio can be used to calculate Bayes factors; see Verdinelli and Wasserman (1995) for details. Given the *a priori* assumption that the models are equally probable, the Bayes factor is equivalent to the posterior odds ratio for the respective models. Beneath each Bayes factor the posterior probability associated with each of the three different models is reported. Using the sensitivity tests outlined in Kass (1993), the Bayes factors are found to be robust with respect to the prior specification for γ and ζ. The tables also state the time taken for the MCMC sampler to produce 1000 iterations given each specific distributional assumption.

For both series the estimates of ϕ indicate a fairly high level of persistence in the latent variable process, with there being a reasonable degree of variation in the persistence estimates across the different conditional distributions, in particular for the NCP data. This highlights the potential influence of the distributional specification on inferences about dynamics in non-Gaussian data of this type; see also McCabe, Martin and Freeland (2004) for more on this point. For both data sets, the point estimates of the distributional parameters γ and ζ differ from the value of one associated with the exponential distribution, with the values of $\gamma = 1$ as $\zeta = 1$ excluded from the 95% HPD

20

interval in three of the four cases The value of $\zeta = 1$ is only just included in the 95% interval for ζ in the NCP case. This lack of support for the exponential model is substantiated by the posterior model probabilities, which assign small posterior probability to the exponential model for both sets of data. In the case of the BHP series, the gamma distribution is assigned the highest posterior probability, followed by the Weibull and exponential distributions in that order, whilst for NCP the Weibull distribution has by far the largest posterior probability, with much less posterior weight given to the exponential and Weibull distributions. Clearly, these results suggest that a conditional distribution that allows for overdispersion is required in addition to the correlated latent variable, in order to fit the unconditional overdispersion in the data.

In Table 5 we report the unconditional mean, standard deviation, and level of dispersion of the (adjusted) durations, as implied by each estimated model (denoted respectively by $\widehat{\mu}_M$, $\widehat{\sigma}_M$ and $\widehat{\sigma}_M/\widehat{\mu}_M$), along with the corresponding dispersion ratio for the data ($\widehat{\sigma}_D/\widehat{\mu}_D$). For details on the calculation of the model-based quantities, see Bauwens and Veredas (2004). In the NCP case, as tallies with the posterior model probabilities, the best fitting model, in the sense of that which produces the best match of the dispersion in the data, is that based on a Weibull distribution. For the BHP data, all three of the models tend to overestimate the level of dispersion, with there being little to choose between alternatives according to this criterion.

With reference to Table 3, the 95% HPD interval for δ_2 covers the value of zero. Consequently we conclude that in the case of the BHP data, trading volume in the previous period does not impact on the current latent state and, hence, on the current duration. In contrast, the HPD interval reported in Table 4 for the corresponding parameter in the NCP case excludes zero, leading to the conclusion that trading volume does impact negatively on the state variable for this data set. Given the monotonically increasing relationship between ψ_i and the conditional mean of durations that obtains for each

21

Table 3: Estimation Results for BHP Based on 100000 Iterations of the MCMC Sampler (Burn-in of 10000 Iterations)

Parameter		Exponential	Weibull	Gamma
ϕ	Mean[a]	0.921	0.888	0.862
	95% HPDI [b]	(0.908,0.935)	(0.868,0.907)	(0.839,0.885)
	MC Error[c]	0.000	0.001	0.001
	IF Factor[d]	219.051	272.163	265.551
σ_η	Mean	0.155	0.204	0.240
	95% HPDI	(0.137,0.173)	(0.178,0.229)	(0.210, 0.266)
	MC Error	0.001	0.001	0.001
	IF Factor	305.968	435.835	399.085
δ_1	Mean	-0.006	-0.010	-0.014
	95% HPDI	(-0.007,-0.004)	(-0.013,-0.007)	(-0.017,-0.010)
	MC Error	0.000	0.000	0.000
	IF Factor	57.378	92.018	92.937
δ_2	Mean	0.000	0.000	0.000
	95% HPDI	(-0.001,0.000)	(-0.001,0.000)	(-0.001, 0.000)
	MC Error	0.000	0.000	0.000
	IF Factor	25.056	19.995	15.935
γ	Mean	-	1.041	-
	95% HPDI	-	(1.027,1.054)	-
	MC Error	-	0.000	-
	IF Factor	-	34.308	-
ζ	Mean	-	-	1.119
	95% HPDI	-	-	(1.095,1.142)
	MC Error	-	-	0.000
	IF Factor	-	-	28.025
	ϖ	8	6	6
	Bayes Factor	1.0000	4.632	12.505
	$P(M_j|\mathbf{x})$[e]	0.055	0.255	0.689
	Time[f]	105	315	178

(a) Marginal posterior mean; (b) 95% Highest posterior density interval; (c) Monte Carlo standard error; (d) Inefficiency factor; (e) Posterior probability of model M_j, $j = 1, 2, 3$; (f) Number of seconds on average for 1000 iterations of the MCMC sampler on a pentium 4 with a 3.0 GHz processor and 1 Gb of RAM

conditional distribution, we conclude that an increase in NCP trading volume is associated with a decrease in the duration in the following period, as is consistent with Easley and O'Hara (1987, 1992).

As is common, the inefficiency factors vary across parameters, data sets and models. In general, the added complexity associated with both the Weibull and gamma distributions leads to a higher degree of correlation in the iterates and, consequently, to higher inefficiency factors.

6 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 variable model for durations and has been shown to be readily adaptable to different choices of distributional assumption for the measurement equation. The MCMC approach has been compared with the QML procedure using Monte Carlo experiments. 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 and NCP trades indicates a reasonably high degree of persistence in the conditional mean of durations. Bayes factors provide rankings for the relative goodness of fit for each model, favouring the gamma distribution for BHP data and the Weibull distribution in the NCP case. The prefered model in the latter case is also the model that best matches the level of overdispersion in the data. This preference given to the more flexible distributions over the exponential distribution indicates that heterogeneity in the durations data is a product of more than just the level of persistence in the series. There are mixed results with respect to the relationship that trading volume has with the duration between trades. For BHP no relationship is found to hold, whilst the theoretical negative relationship is supported by the NCP data.

Table 4: Estimation Results for NCP Based on 100000 Iterations of the MCMC sampler (Burn-in of 10000 Iterations)

Parameter		Exponential	Weibull	Gamma
ϕ	Mean[a]	0.673	0.905	0.683
	95% HPDI[b]	(0.634,0.710)	(0.865,0.937)	(0.606,0.765)
	MC Error[c]	0.001	0.002	0.004
	IF Factor[d]	77.2044	543.714	386.75
σ_η	Mean	0.568	0.240	0.554
	95% HPDI	(0.530,0.606)	(0.188,0.310)	(0.459,0.642)
	MC Error	0.001	0.004	0.005
	IF Factor	101.692	756.851	516.422
δ_1	Mean	-0.078	-0.007	-0.074
	95% HPDI	(-0.095,-0.062)	(-0.016,0.000)	(-0.106,-0.042)
	MC Error	0.000	0.000	0.001
	IF Factor	32.513	126.245	174.601
δ_2	Mean	-0.013	-0.009	-0.013
	95% HPDI	(-0.019,-0.006)	(-0.013,-0.005)	(-0.019,-0.007)
	MC Error	0.000	0.000	0.000
	IF Factor	8.419	33.650	112.772
γ	Mean	-	0.878	-
	95% HPDI	-	(0.861,0.901)	-
	MC Error	-	0.000	-
	IF Factor	-	70.794	-
ζ	Mean	-	-	0.991
	95% HPDI	-	-	(0.933,1.043)
	MC Error	-	-	0.001
	IF Factor	-	-	112.772
	ϖ	8	8	11
	Bayes Factor	1.0000	4.283	0.031
	$p(M_j|\mathbf{x})$[e]	0.188	0.806	0.006
	Time[f]	39	168	91

(a) Marginal posterior mean; (b) 95% Highest posterior density interval; (c) Monte Carlo standard error; (d) Inefficiency Factor; (e) Posterior probability of model M_j, $j = 1, 2, 3$; (f) Number of seconds on average for 1000 iterations of the MCMC sampler on a pentium 4 with a 3.0 GHz processor and 1 Gb of RAM

Table 5: Model-based and Data Dispersion Ratios

Statistic	Conditional Distribution		
	Exponential	Weibull	Gamma
BHP			
Model-based Statistics			
$\widehat{\mu}_M$	1.006	0.997	1.137
$\widehat{\sigma}_M$	1.166	1.155	1.330
$\widehat{\sigma}_M/\widehat{\mu}_M$	1.159	1.158	1.170
Data Dispersion Ratio			
$\widehat{\sigma}_D/\widehat{\mu}_D$	1.060	1.060	1.060
NCP			
Model-based Statistics			
$\widehat{\mu}_M$	1.081	1.155	1.068
$\widehat{\sigma}_M$	1.745	1.701	1.712
$\widehat{\sigma}_M/\widehat{\mu}_M$	1.615	1.472	1.604
Data Dispersion Ratio			
$\widehat{\sigma}_D/\widehat{\mu}_D$	1.431	1.431	1.431

(a) $\widehat{\mu}_M$ = model-based estimate of the mean of durations

(b) $\widehat{\sigma}_M$ = model-based estimate of the standard deviation of durations

(c) $\widehat{\mu}_D$ = sample mean of durations

(d) $\widehat{\sigma}_D$ = sample standard deviation of durations

Possible extensions to the methodology include the allowance for more complex dynamics in the latent variable process, in particular the long memory dynamics that characterize certain durations series (see Bauwens *et al.*, 2004), as well as the use of more flexible multi-factor models such as that proposed by Ghysels *et al.* (2004). 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. Further research could also be undertaken into the suitability of the normality assumption for the errors of the latent variable, especially given the empirical evidence in Bauwens and Veredas (2004) that normality may not always be an appropriate assumption in the case of durations data.

Acknowledgements

References

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

[2] Bauwens, L. and Veredas, D. (2004), "The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations," *Journal of Econometrics,* 199, 381-412.

[3] Bauwens, L., Giot, P., Grammig, J. and Veredas, D. (2004), "A Comparison of Financial Duration Models Via Density Forecasts", *International Journal of Forecasting,* 20, 589-609.

[4] Carter, C.K. and Kohn, R. (1994), "On Gibbs Sampling for State Space Models," *Biometrica*, 81, 541-553.

[5] Chib, S. and Greenberg, E. (1996), "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory*, 12, 407-431.

[6] Daniellson, J. (1994), "Stochastic Volatility in Asset Prices Estimation with Simulated Maximum Likelihood", *Journal of Econometrics*, 64, 375-400.

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

[8] 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.

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

[10] Durbin, J. and Koopman, S.J. (2002), "A Simple and Efficient Simulation Smoother for Time Series Models," *Biometrika*, 81, 341-353.

[11] Easley, D. and O'Hara, M. (1987), "Price, Trade Size, and Information in Securities Markets", *Journal of Financial Economics*, 19, 69-90.

[12] Easley, D. and O'Hara, M. (1992), "Time and the Process of Security Price Adjustment", *Journal of Finance*, 47, 577-605.

[13] Elerian, O., Chib, S. and Shephard, N., (2001), "Likelihood inference for discretely observed nonlinear diffusions", *Econometrica*, 69, 959-993.

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

[15] Feng, D., Jiang, J. J. and Song, P. (2004), "Stochastic Conditional Duration Models with "Leverage Effect" for Financial Transaction Data", *Journal of Financial Econometrics*, 2, 390-421.

[16] Fridman, M. and Harris, L. (1998), "A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models', *Journal of Business and Economics Statistics,* 114, 429-434.

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

[18] Ghysels, E., Gourieroux, C. and Jasiak, J. (2004), "Stochastic Volatility Duration Models," *Journal of Econometrics,* 119, 413-433.

[19] 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.

[20] Jacquier, E., Polson, N.G. and Rossi, P.E. (2004), "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors", *Journal of Econometrics*, 122, 185-212.

[21] 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.

[22] Liesenfeld, R. and Richard, J-F. (2003), "Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics", *Journal of Empirical Finance,* 10, 505-531.

[23] Liu, J. (2001), "Monte Carlo Strategies in Scientific Computing", *Springer-Verlag*, New York.

[24] Kass, R. E. (1993), "Bayes Factors in Practice", *The Statistician*, 2, 551-560.

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

[26] McCabe, B.P.M., Martin, G.M. and Freeland, K.R. (2004), "Testing for Dependence in Non-Gaussian Time Series Data". *Department of Econometrics and Business Statistics Working Paper No. 13/04, Monash University.*

[27] Meyer, R. and Yu, J. (2000), "BUGS for a Bayesian Analysis of Stochastic Volatility Models', *Econometrics Journal* 3, 198-215.

[28] Pitt, M. (2000), "Discussion of Durbin and Koopman (2000)," *Journal of the Royal Statistical Society*, Series B 62, 30-32.

[29] Rydberg, T.H. and Shephard, N. (2003), "Dynamics of Trade-by-trade Price Movements: Decomposition and Models", *Journal of Financial Econometrics,* 1, 2-25.

[30] Sandmann, G. and Koopman, S. J. (1998), "Estimation of Stochastic Volatility Models via Monte Carlo maximum likelihood", *Journal of Econometrics*, 87, 271-301.

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

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

[33] Verdinelli, I. and Wasserman, L. (1995), "Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio," *Journal of the American Statistical Association*, 90, 614-618.

[34] Zellner, A. (1996), "An Introduction to Bayesian Inference in Econometrics," *John Wiley and Sons Inc.*, New York, 371.

Appendix I: Production of the Approximating Measurement Equation

In order to minimize the notational complexity associated with the description of this component of the algorithm, it is assumed for the moment that $K = 0$, i.e. that $\boldsymbol{\psi}$ is simulated as a single block of size N. The process of defining the linear Gaussian approximation on which the candidate model for $\boldsymbol{\psi}$ is based begins with an initial specification of an approximating measurement equation in (13), where 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 $q(\boldsymbol{\psi}|\mathbf{x}, \mathbf{W}, \boldsymbol{\theta}, \lambda)$ and $p(\boldsymbol{\psi}|\mathbf{x}, \mathbf{W}, \boldsymbol{\theta}, \lambda)$ are ultimately equated.

The mode of the candidate density $q(\boldsymbol{\psi}|\mathbf{x}, \mathbf{W}, \boldsymbol{\theta}, \lambda)$ is the solution to the vector equation $\frac{\partial \ln q(\boldsymbol{\psi}|\mathbf{x}, \mathbf{W}, \boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\psi}} = 0$. Equivalently, it is the solution to the vector equation $\frac{\partial \ln q(\boldsymbol{\psi}, \mathbf{x}|\mathbf{W}, \boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\psi}} = 0$. Given the linear Gaussian model in (13), the assumption of the density in (4) for ψ_1, and the form of the linear Gaussian state equation in (13), it follows that

$$
\begin{aligned}
\ln q(\boldsymbol{\psi}, \mathbf{x}|\mathbf{W}, \boldsymbol{\theta}, \lambda) \;=\; & \text{constant} \; - \frac{1}{2}\left(\frac{1-\phi^2}{\sigma_\eta^2}\right)(\psi_1 - W_1\delta)^2 \\
& - \frac{1}{2\sigma_\eta^2}\sum_{i=2}^{N}\left(\psi_i - W_{i-1}\delta - \phi\psi_{i-1}\right)^2 \\
& - \frac{1}{2}\sum_{i=1}^{N}\frac{(\widetilde{x}_i - \psi_i)^2}{\widetilde{H}_i}.
\end{aligned} \tag{17}
$$

Table 6: Specifications for the Approximating Measurement Equation in (14)

Conditional Distribution	$\widetilde{H}_i = \left[\frac{\partial^2 h(x_i \mid \psi_i, \lambda)}{\partial \psi_i^2} \right]^{-1}$
Exponential	$\widetilde{H}_i = x_i^{-1} \exp(\psi_i)$
Weibull	$\widetilde{H}_i = \left[\frac{x_i}{\exp(\psi_i^{k-1})} \Gamma\left(1 + \gamma^{-1}\right) \right]^{-\gamma}$
Gamma	$\widetilde{H}_i = \frac{\exp(\psi_i^{k-1})}{\gamma x_i}$
	$\widetilde{x}_i = \psi_i - \widetilde{H}_i \left| \frac{\partial h(x_i \mid \psi_i, \lambda)}{\partial \psi_i} \right|$
Exponential	$\widetilde{x}_i = \psi_i - x_i^{-1} \exp(\psi_i) + 1$
Weibull	$\widetilde{x}_i = \psi_i - \gamma \left[\frac{x_i}{\exp(\psi_i^{k-1})} \Gamma\left(1 + \gamma^{-1}\right) \right]^{-\gamma} + 1$
Gamma	$\widetilde{x}_i = \psi_i - \frac{\exp(\psi_i)}{x_i} + \frac{1}{\gamma}$

Differentiating with respect to ψ_i, ignoring the initial and final states, and setting the result equal to zero, yields the equations

$$
\begin{aligned}
\frac{\partial \ln q(\boldsymbol{\psi}, \mathbf{x} | \mathbf{W}, \boldsymbol{\theta}, \lambda)}{\partial \psi_i} &= -d_i \left(\frac{1}{\sigma_\eta^2} \right) \left(\psi_i - W_{i-1}\delta - \phi\psi_{i-1} \right) \\
&\quad + \left(\frac{\phi}{\sigma_\eta^2} \right) \left(\psi_{i+1} - W_i\delta - \phi\psi_i \right) \\
&\quad + \frac{(\widetilde{x}_i - \psi_i)}{\widetilde{H}_i} \\
&= 0,
\end{aligned}
\tag{18}
$$

$i = 2, \ldots, N-1$, where $d_i = 1$ for $i = 2, \ldots, N-1$. As $q(\boldsymbol{\psi}|\mathbf{x}, \mathbf{W}, \boldsymbol{\theta}, \lambda)$ is Gaussian, the solution to (18) occurs at the mean of $q(\boldsymbol{\psi}|\mathbf{x}, \mathbf{W}, \boldsymbol{\theta}, \lambda)$ which can, in turn, be produced via the application of the Kalman filter and smoother to the model defined by (13) and (2).

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

$$
\begin{aligned}
\ln p(\boldsymbol{\psi}, \mathbf{x} | \mathbf{W}, \boldsymbol{\theta}, \lambda) &= \text{constant} - \frac{1}{2} \left(\frac{1 - \phi^2}{\sigma_\eta^2} \right) (\psi_1 - W_1\delta)^2 \\
&\quad - \frac{1}{2\sigma_\eta^2} \sum_{i=2}^{N} \left(\psi_i - W_{i-1}\delta - \phi\psi_{i-1} \right)^2 \\
&\quad - \sum_{i=1}^{N} h(x_i|\psi_i, \lambda),
\end{aligned}
\tag{19}
$$

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

$$
\begin{aligned}
\frac{\partial \ln p(\boldsymbol{\psi}, \mathbf{x}|\mathbf{W}, \boldsymbol{\theta}, \lambda)}{\partial \psi_i} &= -d_i \left(\frac{1}{\sigma_\eta^2}\right)\left(\psi_i - W_{i-1}\delta - \phi\psi_{i-1}\right) \\
&\quad + \left(\frac{\phi}{\sigma_\eta^2}\right)\left(\psi_{i+1} - W_i\delta - \phi\psi_i\right) \\
&\quad - \frac{\partial h(x_i|\psi_i, \lambda)}{\partial \psi_i} \\
&= 0, \qquad\qquad\qquad\qquad\qquad\qquad (20)
\end{aligned}
$$

for $i = 2, \ldots, N-1$, with d_i, $i = 2, \ldots, N-1$, as defined above. The approximate model in (13) is to be chosen in such a way that the solution to (18) is also the solution to (20). To achieve this, the term $\frac{\partial h(x_i|\psi_i, \lambda)}{\partial \psi_i}$ in (20) is linearized around the trial value $\widetilde{\psi}_i$ as follows,

$$
\frac{\partial h(x_i|\psi_i, \lambda)}{\partial \psi_i} \approx \left.\frac{\partial h(x_i|\psi_i, \lambda)}{\partial \psi_i}\right|_{\psi_i = \widetilde{\psi}_i} + \left.\frac{\partial^2 h(x_i|\psi_i, \lambda)}{\partial \psi_i^2}\right|_{\psi_i = \widetilde{\psi}_i} (\psi_i - \widetilde{\psi}_i). \qquad (21)
$$

Substituting (21) into (20), and rearranging, an explicit expression for $\frac{\partial \ln q(\boldsymbol{\psi}, \mathbf{x}|\mathbf{W}, \boldsymbol{\theta}, \lambda)}{\partial \psi_i}$ in (18) is obtained, with $\widetilde{x}_i = \psi_i - \widetilde{H}_i \dot{h}$ and $\widetilde{H}_i = \ddot{h}_i^{-1}$, where $\dot{h}_i = \frac{\partial h(x_i|\psi_i, \lambda)}{\partial \psi_i}$ and $\ddot{h}_i = \frac{\partial^2 h(x_i|\psi_i, \lambda)}{\partial \psi_i^2}$. The form of both $\widetilde{x}_i$ and $\widetilde{H}_i$, for the different conditional distributions specified in Table 1, is given in Table 6.

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 (13) and (2) to produce the mode of $q(\boldsymbol{\psi}|\mathbf{x}, \mathbf{W}, \boldsymbol{\theta}, \lambda)$.

3. Substitute the mode of $q(\boldsymbol{\psi}|\mathbf{x}, \mathbf{W}, \boldsymbol{\theta}, \lambda)$ into (20) and check whether the first order conditions are satisfied.

4. If the first order conditions are not satisfied recalculate $\widetilde{H}_i$ and $\widetilde{x}_i$ using Table 6, as based on the output of the Kalman filter and smoother in 2. (i.e. the current mode of $q(\boldsymbol{\psi}|\mathbf{x}, \mathbf{W}, \boldsymbol{\theta}, \lambda)$)

5. Repeat from step 2 until the first order conditions in (20) are satisfied.

6. Use $\widetilde{H}_i$ and $\widetilde{x}_i$ so derived to define the candidate model for $\boldsymbol{\psi}$.

Appendix II: Inefficiency Factors

The inefficiency factor may be interpreted as the magnitude of the variance of the mean of the correlated MCMC iterates, relative to the variance of the mean of a hypothetical sample of independent draws. To calculate the inefficiency factor the following formula is used,

$$\widehat{IF} = 1 + 2\frac{B}{B-1}\sum_{i=1}^{B} K_{QS}\left(\frac{i}{B}\right)\widehat{\rho}_i, \tag{22}$$

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. We use 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. The QS kernel is defined as

$$K_{QS}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right). \tag{23}$$

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

$$\widehat{B} = 1.3221(\widehat{\alpha}(2)M)^{1/5}, \tag{24}$$

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

$$\widehat{\alpha}(2) = \frac{4\widehat{\rho}_a^2\widehat{\sigma}_a^4}{(1-\widehat{\rho}_a)^8}\bigg/\frac{\widehat{\sigma}_a^4}{(1-\widehat{\rho}_a)^4}. \tag{25}$$

The terms $\widehat{\rho}_a$ and $\widehat{\sigma}_a$ in (25) 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.