SIMULATION BASED LIKELIHOOD INFERENCE FOR LIMITED DEPENDENT PROCESSES

AURORA MANRIQUE

Department of Economics, University of Salamanca, Spain

amg@gugu.usal.es

AND

NEIL SHEPHARD

Nuffield College, Oxford OX1 1NF, UK

neil.shephard@nuf.ox.ac.uk

April 8, 1998

Abstract

This paper looks at the problem of performing likelihood inference for limited dependent processes. Throughout we use simulation to carry out either classical inference through a simulated score method (simulated EM algorithm) or Bayesian analysis.

A common theme is to develop computationally robust methods which are likely to perform well for any time series problem. The central tools we use to deal with the time series dimension of the models are the scan sampler and the simulation signal smoother.

Some key words: Disequilibrium models, Markov chain Monte Carlo, Scan sampler, Tobit model.

1 INTRODUCTION

1.1 General problem

Suppose $s \in R^{n \times p}$ is jointly Gaussian with a density $f_S(s; \theta)$ indexed by a finite dimensional parameter θ. The task is to use likelihood methods to infer about θ in problems where there is a form of time dependence in s. The difficulty will be that we observe only a coarse version of s, $h(s)$, which we call y, where the support of y is a proper subset of the support of s. Recently there has been extensive work in the econometric literature on the use of simulation to estimate these types of coarse models in contexts where the time dimension n is typically small. This very interesting literature is reviewed by Hajivassiliou and Ruud (1994).

In this paper we will focus on time series problems. We take n as the length of the p dimensional time series and we suppose that n is typically large compared to p. Our desire is to develop methods which deliver likelihood based estimators in $O(n)$ computations. We will be able to carry this out by exploiting the common structure of time series models.

To focus ideas we will write down a general Gaussian model for s, which will be assumed to follow a Gaussian state space form (see, for example, Harvey (1989) and West and Harrison (1997)). This allows us to efficiently handle all Gaussian time series models except those which are fractional. The structure of the model will be that $s = (s_1', ..., s_n')'$, where $s_t = (s_{t1}, ..., s_{tp})'$ is a p-dimensional vector. Then

$$\begin{aligned} s_t &= c_t + Z_t \alpha_t + G_t u_t, \quad t = 1, ..., n, \\ \alpha_{t+1} &= d_t + T_t \alpha_t + H_t u_t, \quad u_t \sim NID(0, \sigma^2 I), \quad t = 0, 1, ..., n, \\ \alpha_0 &= 0. \end{aligned} \tag{1}$$

Typically c_t, Z_t, G_t, d_t, T_t and H_t will be assumed to be fixed and indexed by θ, while we will write the signals $c_t + Z_t \alpha_t$ as μ_t. Often we use the notation $\mu = (\mu_1', ..., \mu_n')'$ and $\alpha = (\alpha_1', ..., \alpha_n')'$. We call the s_t the latent time series. Examples of models which can be put in this framework are vector autoregressions, unobserved component models, continuous time models observed irregularly and moving averages. Notice that s_t can be, and will in practice often be, non-stationary.

The advantage of insisting that the time series part of the model is placed into a Gaussian state space form (SSF) is that it automatically means that a number of computational problems have been solved, allowing us to generically compute in $O(n)$ the following: (1) the joint density $f_S(s; \theta)$ by the Kalman filter, (2) simulate from $\alpha | s; \theta$, or $\mu | s; \theta$ using the simulation state and signal smoothers (de Jong and Shephard (1995)), (3) given θ, evaluate the density of a single s_t (and hence be able to simulate from it), conditional on the most recent values of all the other latent variables by the scan sampler (de Jong (1997)). Each of these algorithms will prove to

be helpful in dealing with non-Gaussian problems. The first two are detailed in the Appendix, while the other will be introduced in section 2 of the paper.

1.2 Limited dependent processes

We call y_t a Gaussian limited dependent process if $y_t = h(s_t)$, and $h(\cdot)$ is not a one-to-one function. If y_t is multivariate then the process could include some variables which are identical to elements of the s_t vector. The following subsections give examples of this setup.

1.2.1 Tobit and Probit processes

Maddala (1983) reviews the very extensive literature on the univariate Tobit model, which writes $y_t = \max(0, s_t)$, and the Probit model which has $y_t = I(s_t > 0)$. A Tobit process is simply the time series extension of a standard univariate Tobit regression model and so the univariate latent process is only observed when it is positive. A univariate Probit process is quite a rich alternative to simple Markov chains for even if s_t is a first order autoregression, y_t is not Markovian.

1.2.2 Disequilibrium processes

Models which allow for markets not to instantaneously clear have a long tradition in economics. One econometric formulation of this phenomena is the disequilibrium model where the univariate output, y_t, is the minimum of demand and supply. Suppose $s_t = (s_{1t}, s_{2t})'$ contains supply and demand. Then $y_t = \min(s_{1t}, s_{2t})$. Recent work in this area includes Laroque and Salanié (1993), Lee (1997a) and Lee (1997b).

1.2.3 Bid/ask price dynamics

In most markets assets are traded only at a fixed number of prices. An example of this is the New York Stock Exchange where some stocks have had tick sizes of as much as 1/8th of a dollar. For high frequency data such discretisations could considerably distort the econometric analysis and we have to explicitly model this institutional feature of the data.

In a recent paper Hasbrouck (1996) suggested a Gaussian limited dependent process for the analysis of the dynamic behaviour of bid and ask quotes. Let μ_t be the unobserved implicit efficient price of a security, $\mu_t - \beta_t$ the bid price in the absence of discreteness restrictions and $\mu_t + \alpha_t$ the corresponding ask price. Here β_t, α_t reflects the non-negative cost of quote exposure for small trades. Hasbrouck (1996) assumed that $\{\log \mu_t, \log \beta_t, \log \alpha_t\}$ followed a Gaussian process. As the market prices occur on a discrete mesh the observed bid and ask prices are

$$b_t = Floor(\mu_t - \beta_t) \qquad a_t = Ceiling(\mu_t + \alpha_t),$$

where the *Floor* function rounds down to a whole integer and *Ceiling* rounds up.

1.3 Outline of paper

In the next section we detail basic inference issues. Then we discuss parameter estimation in the context of a Tobit process in section 3, and for a simple dynamic disequilibrium model in section 4. Results of some Monte Carlo experiments are reported. Section 5 concludes and finally, the Appendix details various algorithms featured in the paper.

2   LIKELIHOOD INFERENCE

2.1   Basic results

2.1.1   Simulated EM algorithm

In general it is easy to state what we must carry out to perform likelihood inference. The score function is given by the usual EM algorithm result due to Louis (1982). We write the complete data likelihood as $f(s; \theta)$, while we will work with respect to the posterior distribution of the complete data $F(s|y; \theta)$. Notice that this distribution only has positive probability at the points s for which $y = h(s)$.

The EM algorithm (see Dempster, Laird, and Rubin (1977)), involves two steps: the E-step (expectation step) and the M-step (maximisation step). In the E-step, we calculate

$$Q\left\{\theta, \theta^{(i)}; y\right\} = \int_{y=h(s)} \log f(s; \theta) dF(s|y; \theta),$$

that is the expectation of the log-likelihood function for the complete data with respect to the distribution of $s|y; \theta^{(i)}$, where $\theta^{(i)}$ denotes the parameter estimate after i iterations of the algorithm. In the M-step, we maximise $Q\left\{\theta, \theta^{(i)}; y\right\}$ with respect to θ, inducing the mapping

$$\theta^{(i+1)} = \arg\max_{\theta} Q\left\{\theta, \theta^{(i)}; y\right\}.$$

The algorithm is iterated until $\left\|\theta^{(i+1)} - \theta^{(i)}\right\| < \varepsilon$, for some small ε, $(\varepsilon = 10^{-10}$ in all our experiments) or $\left|Q\left\{\theta^{(i+1)}, \theta^{(i)}; y\right\} - Q\left\{\theta^{(i)}, \theta^{(i)}; y\right\}\right|$ is sufficiently small. Under weak regularity conditions (see, for example, Wu (1983)) this EM algorithm converges to a (local) maximum in the log-likelihood function. Notice

$$\frac{\partial \log f(y; \theta)}{\partial \theta} = \int_{y=h(s)} \frac{\partial \log f(s; \theta)}{\partial \theta} dF(s|y; \theta) = \frac{\partial Q(\theta, \theta; y)}{\partial \theta}.$$

Generally $F(s|y; \theta)$ is intractable, however it will be seen that we can simulate from it, giving the possibility of unbiasedly estimating the score $\partial Q(\theta, \theta; y)/\partial \theta$. We will write in general

$$\frac{\partial \widetilde{Q}\left(\theta, \theta^{(i)}; y\right)}{\partial \theta} = \frac{\partial \log \widehat{f(y; \theta)}}{\partial \theta}\Bigg|_{\theta=\theta^{(i)}} = \frac{1}{R} \sum_{j=1}^{R} \frac{\partial \log f(s^j; \theta)}{\partial \theta}, \qquad \text{where} \qquad s^j \sim F\left\{s|y; \theta^{(i)}\right\}.$$

4

This estimator will typically converge to the true score as $R \to \infty$. Iterating this simulation based estimator

$$\frac{\partial \widetilde{Q} \left\{ \theta^{(k+1)}, \theta^{(k)}; y \right\}}{\partial \theta} = \frac{\partial \widetilde{Q} \left\{ \theta, \theta^{(k)}; y \right\}}{\partial \theta} \Bigg|_{\theta = \theta^{(k+1)}} = 0,$$

performs a series of simulated EM updates. This is studied in Qian and Titterington (1991) and Chan and Ledolter (1995), while earlier work on this subject includes Bresnahan (1981), Wei and Tanner (1990) and Ruud (1991). A textbook exposition of this material is given in Tanner (1996, Ch. 4). Typically this literature considers the use of simulation as a convenient way to approximate expectations in the EM algorithm and it is presumed that in practice the number of simulations R will be large in order to reduce the approximation error.

Another literature has also grown up using this technique which suggests this procedure is useful when R is small. This works off the result that averaging over possible simulations from $F(s|y; \theta_0)$ and data sets, the expected value of $Q(\theta_0, \theta_0; y)$ is zero. This extends the usual result that the score at the true value has zero expectation and can be used to construct an estimating equation. If we assume that the simulations will be continuous as a function of θ then the estimating equation

$$0 = \frac{1}{R} \sum_{j=1}^{R} \frac{\partial \log f\left(s^j; \theta\right)}{\partial \theta} \Bigg|_{\theta = \widehat{\theta}}, \qquad \text{where} \qquad s^j \sim F\left(s|y; \widehat{\theta}\right),$$

gives a consistent estimator $\widehat{\theta}$ which is typically asymptotically normal with a covariance matrix that depends on the information matrix and the amount of simulation. The suggestion of using this approach with a small R and allowing $n \to \infty$ to produce approximate distribution theory seems to have first occurred in the econometrics literature due to the work of Hajivassiliou and McFadden (1996). In econometrics, this estimator is called the simulated scores estimator. As $R \to \infty$ it becomes fully efficient and is an EM algorithm.

The requirement that the simulations from $F(s|y; \theta)$ be continuous in θ is quite restrictive as it means that we cannot use rejection type methods or binary indicators to generate the simulations.

Furthermore, it is important to note that convergence of the iterates requires that the same underlying random variates are used to simulate the latent data throughout all iterations of the algorithm. If these random numbers are not fixed, then the sequence of parameter estimates will be random.

2.1.2 Bayesian estimator

Bayesian estimators, which focuses on the posterior density of $\theta|y$, give an alternative likelihood approach. The posterior is basically intractable when tackled directly, however again the use

of simulation delivers easy methods. Suppose we can simulate from the distribution of $s|y; \theta$ and in addition that we can simulate from the distribution of $\theta|s$. This second task is usually possible (although it is often difficult) as $s|\theta$ is Gaussian and can be evaluated using the Kalman filter. Then we can set up a simple Markov chain Monte Carlo (MCMC) sampler to analyse these problems. A booklength review of this literature is given in Gilks, Richardson, and Spiegelhalter (1996). Using this setup the sampler will proceed as follows: (1) Initialise θ; (2) Sample s from $s|y; \theta$; (3) Sample θ from $\theta|s$; (4) Repeat from step 2.

This type of algorithm will, under some rather weak conditions, converge to a draw from $(s', \theta')'|y$ using Markov chain Monte Carlo results. Averaging subsets of these simulations will lead to Bayesian estimators of the parameters. The estimators, based on the mean, median or mode of the posterior density of $\theta|y$, are typically efficient (if viewed as estimators from a sampling viewpoint) as shown by, for example, Barndorff-Nielsen and Cox (1994, Ch. 4) for a wide class of prior distributions. Further, if the prior distributions are believed then this approach gives a completely self contained method of performing exact finite sample inference.

Sometimes sampling from $\theta|s$ is generically difficult, even though we can evaluate the likelihood $f(s; \theta)$. An alternative would be to add another line of simulation and so switch to the following sampler: (1) Initialise θ; (2) Sample s from $s|y; \theta$; (3) Sample α from $\alpha|s; \theta$; (4) Sample θ from $\theta|\alpha, s$; (5) Repeat from step 2. Despite performing the additional sampler at step 3, in order that the updating of the parameters is from $\theta|\alpha, s$ rather than the difficult density $\theta|s$, the rate of convergence of the sampler does not appear to be much slower in experiments we have conducted. Hence this is the approach we advocate in this paper for the generic problem.

2.1.3 Maximum simulated likelihood estimators

Recently a number of authors have developed methods which can be used to handle non-Gaussian SSF models, estimating the likelihood function using importance sampling methods. The resulting log of the estimated likelihood is then numerically maximised. Generally, these methods require the amount of simulation used to estimate the log-likelihood to increase as the sample size rises in order to get good asymptotic properties.

An early paper on this topic in econometrics is Danielsson and Richard (1993). A simpler approach, initiated by Shephard and Pitt (1997), has recently been further developed by Durbin and Koopman (1997). This so-called SPDK importance sampler has been applied to stochastic volatility models by Sandmann and Koopman (1998). These methods are not widely applicable to limited dependent processes as they rely on the existence of significant measurement error in the models in order to construct good suggestion densities, which may not be present here.

A set of alternative sequential importance samplers have been investigated in the context of

disequilibrium models in a recent series of papers by Lee (1997a) and Lee (1997b). His preferred method is the use of a smooth importance sampler which is similar in spirit to the methods developed for panel data models. It has the disadvantage that it is computationally expensive for long time series unless the dynamics of the limited dependent process is quite simple (Lee (1997b, pp. 291-2)).

An alternative importance sampling approach is the use of particle filters, which has recently attracted a great deal of attention in the statistics literature. See the papers by Gordon, Salmond, and Smith (1993), Pitt and Shephard (1997), Berzuini, Best, Gilks, and Larizza (1997) and Kim, Shephard, and Chib (1998). It has the disadvantage that the log estimated likelihood function has a large number of points in the parameter space which are not continuous, although such discontinuities are typically very small with reasonably large simulation sizes.

2.2 Simulating from $s|y; \theta$

2.2.1 Conditional structure

The task will be to simulate from the distribution of $s|y; \theta$. Our suggestion is to design Markov chain Monte Carlo methods to handle these types of problems. These methods will attempt to be as general as we can make them and be reasonably efficient across a wide variety of problems.

The basic structure of our approach is that we intend to sample a single s_t, given the most recent updated values of all the other latent points $s_{\backslash t} = (s_1, ..., s_{t-1}, s_{t+1}, ..., s_n)'$, the observations y, and θ, repeating this operation for $t = n, ..., 1$. However, this simplifies due to the structure of the model since

$$s_t|s_{\backslash t}, y; \theta \overset{L}{=} s_t|s_{\backslash t}, y_t; \theta,$$

where $\overset{L}{=}$ denotes an equality in law. Now $s_t|s_{\backslash t}, y; \theta$ has a simple distribution which is that of $s_t|s_{\backslash t}; \theta$ subject to the constraint that $y_t = h(s_t)$. Hence the problem reduces to one of simply efficiently computing the mean and variance of the Gaussian distribution of $s_t|s_{\backslash t}; \theta$ and then sampling from the constrained space induced by knowing y_t.

2.2.2 Scan sampler

At first sight the problem of evaluating the density of $s_t|s_{\backslash t}; \theta$ is the jackknife problem for SSF models and was solved by de Jong (1989). However, the problem here is different as we want to evaluate this density and repeatedly sample from it (for $t = n, ..., 1$) in only $O(n)$ operations. Given θ, at every iteration of the sampler, s_t will be updated conditional on the most recent values of all the other latent points. Hence, for example, given initial values for $s_1, ..., s_n$, say $s_1^{(0)}, ..., s_n^{(0)}$, if we evaluate $s_n|s_{\backslash n}; \theta$ and then sample $s_n^{(1)}$ from $s_n|s_{\backslash n}, y_t; \theta$ the next draw we have

to make requires us to evaluate the distribution of

$$s_{n-1} | s_1^{(0)}, ..., s_{n-2}^{(0)}, s_n^{(1)}; \theta.$$

Thus for each density evaluation the conditioning variable will change. This is a much harder problem to solve than jackknifing which, if directly applied, would deliver an $O(n^2)$ operation.

Fortunately this type of problem has recently been solved by de Jong (1997) in a result he called the scan sampler. Let $s^{(k)}$ denote the latent series at the k-th step of the sampler, then the scan sampler evaluates the density of $s_t | s_{\backslash t}; \theta$, where $s_{\backslash t} = \left\{ s_1^{(k)}, ..., s_{t-1}^{(k)}, s_{t+1}^{(k+1)}, ..., s_n^{(k+1)} \right\}$, and simulates $s_t^{(k+1)}$ from it (for $t = n, ..., 1$). It can be stated as follows:

Scan sampler (due to de Jong (1997))

1. Initialise s at say $s^{(0)}$. Set $k = 0$.

2. Sample $s^{(k+1)}$ with $s_t^{(k+1)} \sim s_t | s_{\backslash t}; \theta$ where $s_{\backslash t} = \left\{ s_1^{(k)}, ..., s_{t-1}^{(k)}, s_{t+1}^{(k+1)}, ..., s_n^{(k+1)} \right\}'$, $t = n, ..., 1$, as follows:

   (a) Apply the Kalman filter to $s_1^{(k)}, ..., s_n^{(k)}$ to calculate F_t^{-1}, v_t and K_t. Here v_t is the one-step ahead prediction errors, F_t is the corresponding mean square error and K_t is the Kalman gain. Details of how each of these are computed is given in the Appendix.

   (b) Set $t = n$, $r_n = 0$ and $N_n = 0$.

   (c) Compute with $L_t = T_t - K_t Z_t$

$$
\begin{aligned}
e_t &= F_t^{-1} v_t - K_t' r_t, & s_t^{(k+1)} &\sim N\left\{ s_t^{(k)} - D_t^{-1} e_t, D_t^{-1} \right\} \\
r_{t-1} &= Z_t' F_t^{-1} v_t + L_t' r_t - V_t' \left\{ s_t^{(k)} - s_t^{(k+1)} \right\}, & N_{t-1} &= Z_t' F_t^{-1} Z_t + L_t' N_t L_t,
\end{aligned}
\tag{2}
$$

   where

$$D_t = F_t^{-1} + K_t' N_t K_t, \qquad V_t = F_t^{-1} Z_t - K_t' N_t L_t.$$

   (d) Let $t = t - 1$. If $t \geq 1$, then repeat from step 2(c).

3. Write $k = k + 1$ and repeat from step 2.

The result of step 2 is a new latent time series $s^{(k+1)}$ which comes about from a complete sweep of the scan sampler. The scan sampler has close similarities to the simulation signal smoother of de Jong and Shephard (1995) which draws from $\mu | s; \theta$.

It is important to note that this scan sampler derives random variables $s_t | s_{\backslash t}; \theta$ from a density which is strictly positive on R^p under the simple and very weak assumption that there are no

linear combinations of s_t which are perfectly predictable from $s_{\backslash t}$ for any t. If such a model were to exist, then it is likely that we could reformulate it to satisfy the condition.

Of course the scan sampler we have just stated is an unconditional sampler, but we can easily amend it by simply replacing the sampling step

$$s_t^{(k+1)} \sim N\left\{s_t^{(k)} - D_t^{-1}e_t, D_t^{-1}\right\} = f\left\{s_t|s_1^{(k)}, ..., s_{t-1}^{(k)}, s_{t+1}^{(k+1)}, ..., s_n^{(k+1)}; \theta\right\},$$

by the drawing from the above normal density subject to the constraint that $h\left\{s_t^{(k+1)}\right\} = y_t$. Thus the scan sampler has completely dealt with the time series aspect in handling these models. We will return to this constrained sampling problem after the next subsection.

2.2.3   Alternatives to the scan sampler

The use of MCMC methods means that it is possible to use a huge variety of different methods to provide simulations from $s|y; \theta$. Although each alternative will usually have a geometric rate of convergence to $s|y; \theta$ it does not imply that each is as general or as good as the other. Here we highlight two alternatives which are in general inferior, in terms of speed of convergence of the sampler, to the use of the scan sampler but could be employed in this context. A thorough empirical comparison of their relative performance for the examples given below is given in detail in Manrique (1997), while a theoretical explanation can be found in the results of Liu, Wong, and Kong (1994) on the Gibbs sampler using reduced conditionals. Here we just mention the results.

Signal simulation smoothing   In some recent work, de Jong and Shephard (1995) showed how to efficiently sample in $O(n)$ from the np dimensional normal distribution of $\mu|s$. Clearly if $G_t G_t'$ is non-singular for all t then we might use the following scheme to sample from $s|y$: (1) Initialise s; (2) Sample $\mu \sim \mu|s$; (3) Sample $s \sim s|\mu, y$; (4) Goto 2.

The advantage of this scheme is that the $\{s_t|\mu_t, y_t; \theta\}$ are conditionally independent, which should be relatively easy to sample from. The difficulty with this approach arises when $G_t G_t'$ is close to being singular, for in this case knowing μ is nearly equivalent to knowing s and so we might expect the sampler to converge quite slowly. In the limit, as $G_t G_t'$ becomes singular it would not converge at all. This is an important scenario, for example if s_t was an autoregression moving average of any order, then this is exactly the situation we have just described. Indeed an algorithm based on the simulation signal smoother will only work on models with explicit measurement error.

Single move state sampler   The same problem occurs when we work with a univariate MCMC sampler for the states, as suggested by Carlin and Polson (1992) in this context. This

is often called a single move sampler, as it makes suggestions which if accepted would only change a single state at a time. Their idea, following the more general methods discussed in Carlin, Polson, and Stoffer (1992), was to use the conditional independence structure of the states and sample in the following way: (1) Initialise α; (2) Sample s from $s|\alpha, y; \theta$; (3) Sample $\alpha_t|\alpha_{\backslash t}, s; \theta \stackrel{L}{=} \alpha_t|\alpha_{t-1}, \alpha_{t+1}, s_t; \theta$, for $t = 1, ..., n$; (4) Repeat from step 2.

This sampler is never better than that of the simulation signal smoothing algorithm as the state sampler performs an unnecessary Gibbs sampling step in line 3 and consequently the result of Liu, Wong, and Kong (1994) applies. This states that directly integrating out variables in the Gibbs sampler never slows its convergence rate and almost always increases its speed.

2.2.4 Sampling $s_t|y_t, s_{\backslash t}; \theta$

The use of scan sampling has reduced the problem of sampling from $s|y; \theta$ down to the task of drawing from $s_t|s_{\backslash t}; \theta \sim N(\gamma_t, \Xi_t)$ subject to constraints that $y_t = h(s_t)$. For some simple problems this can be carried out by directly. In general this is a difficult problem and has generated its own literature (see Hajivassiliou, McFadden, and Ruud (1996)). In our context we have to rely on a Markov chain Monte Carlo technique to make suggestions for possible new samples for $s_t|s_{\backslash t}; \theta$ or for elements of that vector.

There has been a great deal of recent work on the special case of this problem where $y_t = h(s_t)$ can be represented by the requirement that

$$a_0(y_t) \leq As_t \leq a_1(y_t),$$

for some boundary vectors a_0, a_1 and some positive definite matrix A. This setup covers important cases such as the multinomial Probit, discretised asset prices and Tobit models, although we cannot put disequilibrium models into this framework. The simplest of general procedures, which can be carried out without rejection, is to sequentially use truncated Gaussian draws inside a Gibbs sampler. This is discussed in Hajivassiliou, McFadden, and Ruud (1996, pp. 110-1). Notice there is no need to iterate until convergence if the scan sampler is being used, as it is itself a Gibbs sampler.

For more complicated functions $y_t = h(s_t)$ it is more difficult to be very prescriptive, and hence we will deal with it in an ad hoc way. However, it may be inevitable that we will not be able to smoothly simulate from $s_t|y_t, s_{\backslash t}; \theta$, which will mean classical methods will become difficult and we will typically resort to Bayesian methods. Indeed this will be what we do when we work with the disequilibrium models in section 4.

3 EXAMPLE: TOBIT

3.1 Basics

In this section we extend the Tobit structure $y_t = \max(0, s_t)$, to allow for a long time series dimension in s_t. An example of this is where

$$s_t = \mu + \phi s_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim NID\left(0, \sigma_\varepsilon^2\right), \tag{3}$$

with $|\phi| < 1$. The $\{s_t\}$ are Markovian, while the $\{y_t\}$ are not. This contrasts with a simple alternative process written directly in terms of the y_t where

$$y_t = \max(0, \mu + \phi y_{t-1} + \varepsilon_t), \tag{4}$$

which is trivial to estimate using a conditional (on y_1) likelihood as the conditional ML estimator is easily found by numerical optimisation.

Our model is more complicated due to its non-Markovian nature. This type of structure appears in many aspects of econometrics and so it is interesting to study ways of estimating it. (3) is easier to work with from a modelling building viewpoint — for example it is obvious how to build a continuous time model of this type, while it would be difficult using the specification in (4).

Chib (1992) proposes a MCMC method for the Tobit regression model, where

$$s_t = x_t'\beta + \varepsilon_t, \quad \varepsilon_t \sim NID\left(0, \sigma^2\right)$$

and the regressors are strongly exogenous. Here $\theta = (\beta', \sigma^2)'$. Chib studied setting $\mu_t = x_t'\beta$ and then running: (1) Initialise s; (2) Sample θ from $\theta|s$, and then construct μ; (3) Sample s from $s|\mu, y; \theta$; (4) Repeat from step 2.

This sampler is exactly of the form of the simulation signal smoother based algorithm discussed in the previous section for the time series extension of this regression model. It will generally work if $\sigma^2 > 0$ which is not an unreasonable assumption in the context of this model. However, it becomes less convincing in the time series context.

3.2 Sampling $s_t|s_{\backslash t}, y_t; \theta$

For the Tobit model, sampling from $s_t|s_{\backslash t}, y_t; \theta$ is straightforward, for $s_t|s_{\backslash t}, y_t = y_t$, if $y_t > 0$, and

$$s_t|s_{\backslash t}, y_t; \theta \sim TN_{s_t < 0}\left(\gamma_t, \Xi_t\right), \quad \text{if} \quad y_t = 0,$$

where γ_t and Ξ_t are, respectively, the mean and variance matrix of the Gaussian distribution of $s_t|s_{\backslash t}; \theta$, and the notation $TN_{s_t < 0}$ means that s_t has a truncated normal distribution and

$s_t < 0$. As we can sample continuously from truncated normal distributions by applying the probability integral transform theorem (see for example Devroye (1986, p. 39))[1], the resulting Gibbs sampler will be smooth in the parameters and will converge to $s|y;\theta$.

To illustrate the relative performance of the possible MCMC samplers for this problem, we consider a simple experiment on $y_t = \max(0, s_t)$, where s_t is defined by

$$
\begin{aligned}
s_t &= \beta + \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim NID\left(0, \sigma_\varepsilon^2\right), \quad \eta_t \sim NID\left(0, \sigma_\eta^2\right), \\
\alpha_{t+1} &= \phi\alpha_t + \eta_t, \quad \alpha_1 \sim N\left\{0, \sigma_\eta^2/\left(1 - \phi^2\right)\right\}.
\end{aligned}
$$

We assume $|\phi| < 1$, $\{\varepsilon_t\}$ and $\{\eta_s\}$ are mutually uncorrelated for all t and s.

The experiment uses a data set of size 50 generated by this model using the parameter values $\beta = -0.5$, $\phi = 0.95$, $\sigma_\varepsilon^2 = 0.1$ and $\sigma_\eta^2 = 0.1$. The expected number of censored observations is 34 while the actual number in the sample is 30. We run the scan, signal simulation smoothing and single move state samplers for $50,000$ iterations, discarding the first $1,000$ results, for the above set of parameter values. In the state sampler we initialise $\alpha_t = s_t$. In the signal simulation smoothing algorithm and the scan sampler we initialise $s_t = y_t$ if $y_t > 0$ and -0.5 otherwise. The resulting drawings from $s_4|y$ $(y_4 = 0)$ were inputted into a correlogram and are reported in Figure 1. The idea is to represent the correlation in the sampler once it has reached equilibrium. As we expected, the results show that the state sampler has a higher degree of correlation than the signal simulation smoothing algorithm, which in turn, is more correlated than the scan sampler. In all cases the correlation is modest, attributable to the fact that a very simple model is being considered.

3.3 Estimation

The state space framework allows either a fully Bayesian or a simulated EM algorithm, as the scan sampler yields simulations which are continuous functions of the model parameters. Once a prior is stated, the Bayesian analysis of this model is algorithmically straightforward. The simulated EM algorithm also takes on a simple form. The log-likelihood for any latent s can be computed as $\log f(\theta; s)$ can be computed by the Kalman filter. Therefore, the function to be maximised, possibly numerically, in the M-step, is

$$
\widetilde{Q}\left(\theta, \theta^{(i)}; y\right) = \frac{1}{R} \sum_{j=1}^{R} \log f\left(s^j; \theta\right), \quad \text{where} \quad s^j \sim F\left\{s|y; \theta^{(i)}\right\}.
$$

---

[1]Precisely, if ξ is a random variable with distribution function F, and η is a truncated version of ξ to the region $[a, b]$ with distribution function G, then we can sample from η by

$$
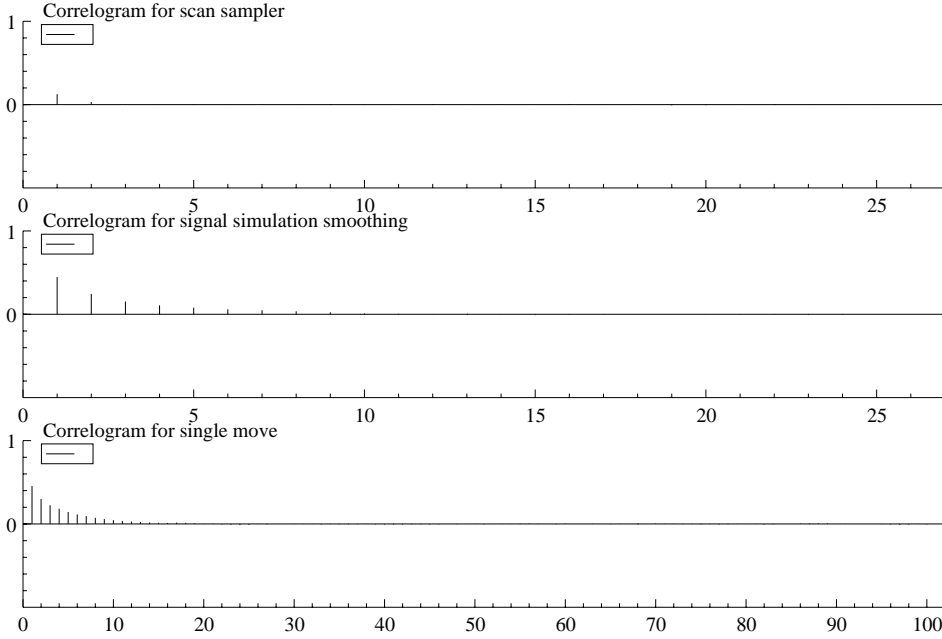F^{-1}\left[F(a) + u\left\{F(b) - F(a)\right\}\right], \quad \text{where} \quad u \sim U[0, 1].
$$

Figure 1: *Correlograms of $y_4|Y_{50}$. Indicates rate of convergence. From top to bottom: scan sampler, signal simulation smoother (multi-move) sampler and single-move sampler.*

The solution to this maximisation problem will give $\theta^{(i+1)}$. The process will be iterated until convergence.

3.4   A Monte Carlo study

In this subsection we will illustrate likelihood inference on the Tobit structure where s_t is defined by

$$
\begin{aligned}
s_t &= x_t'\beta + \alpha_t + \varepsilon_t, & \varepsilon_t &\sim NID\left(0, \sigma_\varepsilon^2\right), \\
\alpha_{t+1} &= \phi\alpha_t + \eta_t, & \eta_t &\sim NID\left(0, \sigma_\eta^2\right),
\end{aligned}
\tag{5}
$$

with $|\phi| < 1$, $\{\varepsilon_t\}$ and $\{\eta_s\}$ independent for all t and s and $\alpha_1 \sim N\left\{0, \sigma_\eta^2/\left(1 - \phi^2\right)\right\}$.

Example 1. We take $x_t = 1$ for all t in the measurement equation. The artificial set of data is generated according to (5) with $\alpha_0 = 0$ for different values of the parameter values.

We perform both Bayesian and classical inference on the parameters of the model. We will take as true values $\beta = 0.25$, $\sigma_\varepsilon^2 = 0.5$, $\sigma_\eta^2 = 0.1$, and $\phi = 0.95$ and use $n = 200$. Thus $\theta = \left(\beta, \sigma_\varepsilon^2, \sigma_\eta^2, \phi\right)'$.

3.4.1   Bayesian inference

When we update the parameters we use the following conditional structure (1) $\beta|s, \alpha, \sigma_\varepsilon^2$, (2) $\sigma_\varepsilon^2|s, \alpha, \beta$, (3) $\sigma_\eta^2|\alpha, \phi$, and (4) $\phi|\alpha, \sigma_\eta^2$. The first three have straightforward conjugate distribu-

13

tions which we use. In particular, we use a non-informative prior distribution for β, and take $L_0\chi_p^{-2}$ for σ_ε^2, and $L_1\chi_q^{-2}$ for σ_η^2. Here the notation $L_0\chi_p^{-2}$ denotes a random variable which is L_0 divided by a chi-squared variable with p degrees of freedom. Throughout we take $q = p = 5$ and $L_0 = p \times 1$ and $L_1 = q \times 1$. We use $2Beta(\zeta_1, \zeta_2) - 1$ as a prior family for ϕ, following Kim, Shephard, and Chib (1998). This implies $E(\phi) = \{2\zeta_1 / (\zeta_1 + \zeta_2)\} - 1$. We take $\zeta_1 = 10$, $\zeta_2 = 2$ so that ϕ has a prior mean of 0.66 and a standard deviation of 0.207. To sample ϕ we adopt the procedure used in Kim, Shephard, and Chib (1998), although a more general procedure such as Chib and Greenberg (1994) could have been used.

The rate of the convergence of the sampler will vary depending on the true parameter values. When the noise in the measurement equation is small, the sampler will converge slowly, especially for small sample sizes, because the data hardly helps to discriminate between β and the states. We also expect slower convergence of the sampler as the autoregression coefficient and/or the noise in the transition equation increase, and for negative values of β.
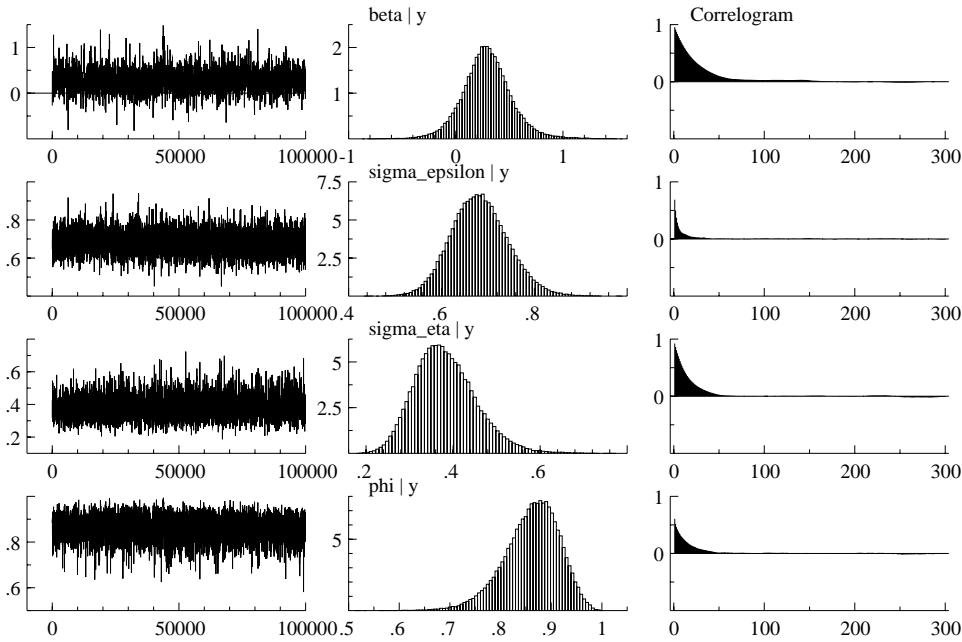

Figure 2: *Multi-move Gibbs sampler for a censored model with $\beta = 0.25$, $\sigma_\varepsilon = 0.707$, $\sigma_\eta = 0.316$, $\phi = 0.95$. Sample size $n = 200$. Left graphs: the simulation against iteration number. Middle graphs: histograms of the resulting marginal distributions and estimated densities. Right graphs: the corresponding correlograms for the iterations.*

Figure 2 and Table 1 present the results of the signal simulation smoothing algorithm for the Tobit process given in (5). The number of censored observations in the sample is 76. The

	TRUTH	Mean	MCse	Ineff	Covariance and *Correlation*			
$\beta\|y$	.25	.2920	.003578	23.9	.0536	*-.0496*	*-.046*	*.092*
$\sigma_\varepsilon\|y$	.707	.6823	.000519	7.2	-.000701	.00372	*-.360*	*.275*
$\sigma_\eta\|y$	.316	.3827	.000979	19.4	-.000753	-.00155	.00495	*-.571*
$\phi\|y$	.95	.8610	.000579	11.3	.00116	.000913	-0.00218	.00296

Table 1: Summaries of Figure 2. Sample size $n = 200$. MCse denotes the Monte Carlo standard error of the simulation estimator of mean of the posterior density. Throughout these standard errors are computed using 250 lags and $100,000$ iterations. Numbers in italics are correlations rather than covariances. Ineff denotes the estimated inefficiency factor.

initial parameter values are set at $\beta = -0.2$, $\sigma_\varepsilon^2 = 0.4$, $\sigma_\eta^2 = 0.1$, and $\phi = 0.9$. We iterated the Gibbs sampler on the states for 500 iterations and then the parameters and states for $1,000$ more iterations before recording any answers. The next $100,000$ iterations are recorded.

The summary statistics of Table 1 report inefficiency factors, or integrated autocorrelations, of the sampler. These are estimated as the variance of the sample mean from the MCMC sampling scheme, using a Parzen window, (see, for example, Priestley (1981, Ch.6)) to account for the serial correlation in the draws[2] relative to the variance of a hypothetical sampler which draws independent random variables from the posterior, (the posterior variance divided by the number of iterations). Note that if the sample was *i.i.d.*, then the inefficiency factor would be one.

The inefficiency factor can be a useful diagnostic, although not the only one, in measuring how well the chain mixes. If we require the Monte Carlo error in estimating the mean of the posterior to be no more than one percentage of the variation of the error due to the data, i.e., $\{MCse(\theta_i)\}^2 / Var(\theta_i|y) \leq 0.01$, for $i = 1, ..., 4$, where θ_i denotes the i-th component of θ, $MCse(\theta_i)$ the Monte Carlo standard error, and $Var(\theta_i|y)$ the posterior variance given y, then inefficiency factors in the region of 30 suggest that the model can be quite precisely analysed in about 3000 iterations of the MCMC algorithm.

An interesting feature of the result is that the regression coefficient β is not particularly well

---

[2]The suggested estimate of the Monte Carlo standard error using a chain of length N, $\widehat{\tau}_N$, is computed as

$$\widehat{\tau}_N = 1 + \frac{2N}{N-1} \sum_{i=1}^{B_M} K\left(\frac{i}{B_M}\right) \widehat{\rho}(i),$$

where $\widehat{\rho}(i)$ is an estimate of the autocorrelation at lag i of the MCMC sampler, B_M represents the bandwith, and K the Parzen kernel given by

$$K(z) = \begin{cases} 1 - 6z^2 + 6z^3, & z \in \left[0, \frac{1}{2}\right] \\ 2(1-z)^3, & z \in \left[\frac{1}{2}, 1\right] \\ 0, & \text{elsewhere.} \end{cases}$$

estimated, while ϕ is really poorly estimated.

Example 2. Our second set of experiments deals with a censored model where the latent variable s_t is defined by equations (5) with a 2-dimensional $\beta = (\beta_1, \beta_2)'$. For each t, $x_t = (x_{1t}, x_{2t})$ is a 2×1 matrix of exogenous variables. In our experiment, $x_{1t} = 1$ for all t, and x_{2t} are i.i.d.. $U(0,1)$ random variables. As usual, $\alpha_0 = 0$ is imposed in the transition equation to generate the states.

The true parameter values are $\beta = (0.25, 0.5)'$, $\sigma_\varepsilon^2 = 0.5$, $\sigma_\eta^2 = 0.1$, $\phi = 0.95$. In this case $\theta = \left(\beta', \sigma_\varepsilon^2, \sigma_\eta^2, \phi\right)'$. The number of censored observations in the sample is 62 out of a sample of size 200. The initial parameter values in the sampler are set to $\beta = (0.1, 0.4)'$, $\sigma_\varepsilon^2 = 0.4$, $\sigma_\eta^2 = 0.1$, and $\phi = 0.9$.

We iterated the Gibbs sampler on the states for 500 iterations and then the parameters and states for $1,000$ more iterations before recording any answers. The next $100,000$ iterations are recorded. Results for a sample of size 200 are given in Figure 3. The posterior means, covariance and correlation matrices, as well as the Monte Carlo standard errors, are presented in Table 2.
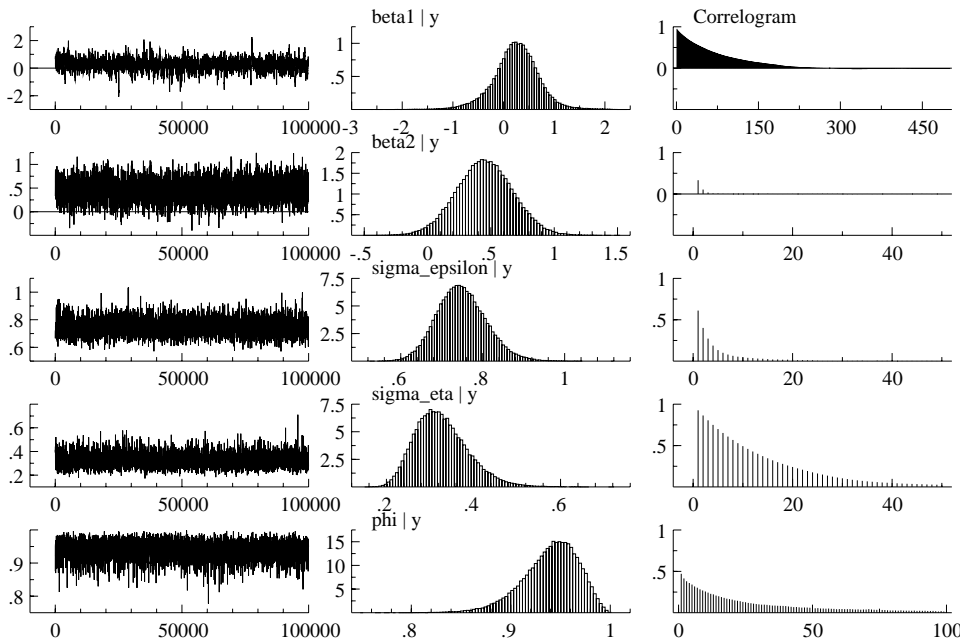

Figure 3: *Multi-move Gibbs sampler for a censored model with $\beta = (0.25, 0.5)'$, $\sigma_\varepsilon^2 = 0.5$, $\sigma_\eta^2 = 0.1$, $\phi = 0.95$. Sample size $n = 200$. Left graphs: the simulation against iteration number. Middle graphs: histograms of the resulting marginal distributions and estimated densities. Right graphs: the corresponding correlograms for the iterations.*

The results suggest that the sampler does not change its behaviour very much with small

16

	TRUTH	Mean	MCse	Ineff	Covariance and *Correlation*				
$\beta_1\|y$	.25	.2432	.00790	30	.206	*-.243*	*-.0440*	*-.0243*	*-.0518*
$\beta_2\|y$	.5	.4519	.000994	2	-.02449	.0490	*.0150*	*.0384*	*-.00219*
$\sigma_\varepsilon\|y$	.707	.7487	.000428	5	-.00120	.000199	.00357	*-.240*	*.186*
$\sigma_\eta^2\|y$	.316	.3270	.000838	19	-.000671	.000517	-.000870	.00369	*-.537*
$\phi\|y$	.95	.9400	.000286	10	-.000659	-1.36e-005	.000311	-.000912	.000783

Table 2: Summaries of Figure 3. Sample size $n = 200$. MCse denotes the Monte Carlo standard error of the simulation estimator of the mean of the posterior density. Throughout these standard errors are computed using 250 lags and $100,000$ iterations. Numbers in italics are correlations rather than covariances. Ineff denotes estimated inefficiency factor.

changes in the model. Again this model could be reasonably analysed in $3,000$ iterations, while the regression coefficients are again poorly estimated.

3.4.2 Classical inference

To illustrate the simulated EM algorithm we will work with the following example.

Example 3 (example 1 reconsidered). The true parameter values are $\beta = 0.25$, $\sigma_\varepsilon^2 = 0.5$, $\sigma_\eta^2 = 0.1$, and $\phi = 0.95$ in the high persistence case and $\phi = 0.6$ in the low persistence case.

Non-negativity constraint for the variances are imposed by the reparameterisations $\sigma_\varepsilon^2 = \exp(\theta_1)$ and $\sigma_\eta^2 = \exp(\theta_2)$, and the stationarity constraint for ϕ is imposed by the reparameterisation $\phi = \theta_3 / (1 + \theta_3^2)$. The parameter β is unconstrained.

For each data set y, the initial value of θ is set to be $\theta^{(0)} = (0, 1, 1, 0)'$. We have checked that the results of the experiment are independent of these initial values. Given the parameter estimate at step i of the algorithm, $\theta^{(i)}$, the scan sampler is applied, with initial values $s_t^{(0)} \sim TN_{s_t<0}\left[\beta^{(i)}, \sigma_\varepsilon^{(i)2} + \sigma_\eta^{(i)2} / \left\{1 - \phi^{(i)2}\right\}\right]$ if $y_t = 0$, to get R scan samples $s_t^1, ..., s_t^R$, $t = n, ..., 1$, with $s_t^j \sim s_t | s_{\backslash t}, y; \theta^{(i)}$. These values are used to compute

$$\widetilde{Q}\left\{\theta, \theta^{(i)}; y\right\} = \frac{1}{R} \sum_{j=1}^{R} \log f\left(s^j; \theta\right).$$

In the M-step of the algorithm, $\widetilde{Q}\left\{\theta, \theta^{(i)}; y\right\}$ is maximised using the quasi-Newton method developed by Broyden, Fletcher, Goldfarb and Shanno (BFGS) and implemented in Ox (see Doornik (1996, pp. 114-8)). Analytical first derivatives, derived by Koopman and Shephard (1992), were used combined with a linear line search if necessary. The maximum number of iterations used in this maximisation routine within the M-step of the EM algorithm is 250. The solution to this maximisation problem will give $\theta^{(i+1)}$. The process is iterated until convergence.

The experiment just described is replicated for 250 data sets. Results for $R = 1$, $R = 5$ and $R = 10$ are presented in Table 3. We find that in this simple case there is little gain in taking
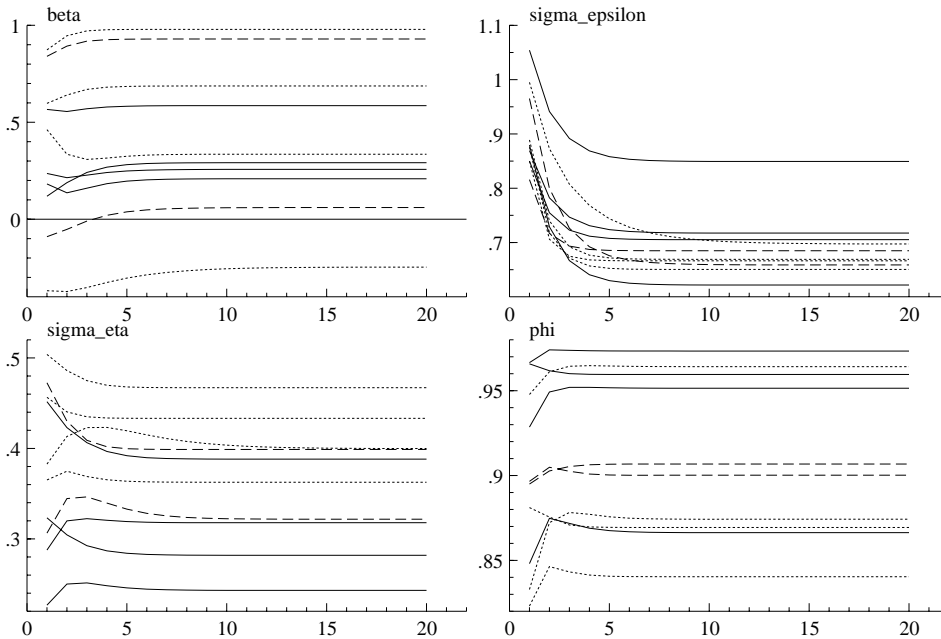
Figure 4: *EM estimates against iteration number for* 10 *random samples of size* 200 *taking* $R = 5$ *scan draws. Illustrates the rapid convergence of the simulated EM algorithm.*

R much bigger than 10. For each data set, we consider 20 iterations of the algorithm.

Figure 4 plots the simulated EM parameter estimates against iteration number for the first 10 data sets taking $R = 5$. These estimates reflect that in most cases, a fairly small number is sufficient for convergence. This suggests that by using the scan sampler we have set the EM algorithm up to have only a small amount of missing data.

The results of an experiment using 200 data sets and $R = 5$ scan draws for different sample sizes are reported in Table 4. For each data set, we consider a maximum of 50 iterations of the EM algorithm, although most replications in fact converge within 25 iterations. The replications for which the algorithm did not converge are not included in the computations of the mean and estimated standard deviations.

The simulated EM algorithm often estimates σ_ε^2 as zero, especially for the low persistence case and small sample sizes. The variance of the estimates of β are much higher for $\phi = 0.95$ than for $\phi = 0.6$. In the low persistence case, the algorithm yields very high variance estimates of the autoregression coefficient.

Example 4 (example 2 reconsidered). In our second experiment, the true parameters values are taken to be $\beta = (0.25, 0.5)'$, $\sigma_\varepsilon^2 = 0.5$, $\sigma_\eta^2 = 0.1$, $\phi = 0.95$. The initial values of the parameters are set to $\theta^{(0)} = (0, 0, 1, 1, 0)'$.

R	Converge	β	σ_ε	σ_η	ϕ
1	100	.2449	.6963	.3471	.9056
		(.400)	(.114)	(.108)	(.065)
5	100	.2539	.6966	.3370	.9091
		(.382)	(.078)	(.080)	(.055)
10	99.6	.2532	.6933	.3362	.9056
		(.382)	(.0829)	(.082)	(.080)
TRUTH		.25	.707	.316	.95

Table 3: 250 Monte Carlo replications as an experiment to study the sampling behaviour of the simulated EM algorithm. Sample size $n = 200$. Uses a variety of values of R. Converge denotes the proportion of replications which converged. Non-bracketed numbers are the average values of the estimates over the Monte Carlo. Figures in brackets denote the empirical standard error of the estimator over the 250 replications.

The sampler is performed as in Example 3 but in this case the BFGS method maximises the log-likelihood for the 'augmented' data using numerical derivatives and at each step, we consider $R = 10$ scan samples. For each data set, the algorithm is iterated until convergence for a maximum of 50 times. In Table 5 we only report results for $n = 200$. The sampler gets very high variance estimates of the regression coefficient.

4   EXAMPLE: GAUSSIAN DYNAMIC DISEQUILIBRIUM

4.1   Modelling framework

We consider in this section a dynamic single-market disequilibrium model specified by a demand function, a supply function and a minimum condition $y_t = \min(s_{1t}, s_{2t})$, where y_t denotes the quantity transacted at period t, s_{1t} the quantity demanded at period t, s_{2t} denotes the quantity supplied at period t. We assume that s_t, which includes s_{1t}, s_{2t} as elements, follows a Gaussian SSF. This process is generally called a linear Gaussian dynamic disequilibrium model. This framework could be extended to allow y_t to be multivariate with $y_{jt} = s_{j+1t}$, $j = 1, ..., P$, where now s_t is a Gaussian SSF. This would be a time series system which allowed a component to be the result of a disequilibrium. For simplicity of exposition we remove this possibility here as the generalisation to cover that problem is technically trivial.

The scan sampler can be used to find the mean and covariance of the Gaussian $s_t|s_{\setminus t}; \theta$. Thus all we have to do is to simply sample which of s_{1t} and s_{2t} is exactly equal to y_t and then sample the other according to a truncated normal distribution. In our problem sampling a single s_t,

n	Converge	β	σ_ε	σ_η	ϕ	Converge	β	σ_ε	σ_η	ϕ
100	99.6	.2852	.6450	.3667	.8479	97.5	0.2426	0.4146	0.4855	0.3582
		(.499)	(.167)	(.143)	(.138)		(.117)	(.325)	(.263)	(.346)
200	100	.2539	.6966	.3369	.9091	98	0.2354	0.4979	0.4537	0.4057
		(.382)	(.078)	(.080)	(.055)		(.089)	(.304)	(.245)	(.340)
500	100	.2262	.7043	.3246	.9351	100	0.2443	0.6152	0.3737	0.5573
		(.274)	(.045)	(.052)	(.024)		(.052)	(.225)	(.196)	(.223)
TRUTH		.25	.707	.316	.95		.25	.707	.316	.6

Table 4: 250 Monte Carlo replications as an experiment to study the sampling behaviour of the simulated EM algorithm. On the left $\phi = 0.95$, and on the right $\phi = 0.6$. Uses $R = 5$ scan draws. Converge denotes the proportion of the 250 replications which converge. Non-bracketed numbers are the average values of the estimates over the Monte Carlo. Figures in brackets denote the empirical standard error of the estimator over the replications.

	β_1	β_2	σ_ε	σ_η	ϕ
mean	.2656	.4965	.6910	.3314	.9130
standard deviation	(.400)	(.225)	(.065)	(.071)	(.049)
TRUTH	.25	.5	.707	.316	.95

Table 5: 250 Monte Carlo replications as an experiment to study the sampling behaviour of the EM algorithm with $R = 10$. Sample size $n = 200$. Uses $R = 10$ scan draws. Non-bracketed numbers are the average values of the estimates over the Monte Carlo. Figures in brackets denote the empirical standard error of the estimator over the 250 replications.

conditional on all the other latent points, the observations and some parameter values, involves computing the probability of y_t being equal to say s_{1t}, given all the other latent points and y_t; choosing which element of s_t is equal to y_t; and then, sampling from a truncated normal distribution. Precisely, we have (suppressing dependence on θ)

$$\lambda_t = \frac{f_1\left(y_t|s_{\backslash t}\right) \times \Pr\left(s_{2t} > s_{1t}|s_{\backslash t}, s_{1t} = y_{1t}\right)}{f_1\left(y_t|s_{\backslash t}\right) \cdot \Pr\left(s_{2t} > s_{1t}|s_{\backslash t}, s_{1t} = y_t\right) + f_2\left(y_t|s_{\backslash t}\right) \cdot \Pr\left(s_{1t} > s_{2t}|s_{\backslash t}, s_{2t} = y_t\right)}, \quad (6)$$

where $\lambda_t = \Pr\left\{s_{1t} < s_{2t}|s_{\backslash t}, \min\left(s_{1t}, s_{2t}\right) = y_t\right\}$ denotes the probability of having excess supply at time t, conditional on all the other components of s_t and the observed y_t. Here $f_1(y_t|s_{\backslash t})$ is the density of $s_{1t}|s_{\backslash t}$ evaluated at y_t, while $f_2(y_t|s_{\backslash t})$ is the corresponding density for $s_{2t}|s_{\backslash t}$. Note that λ_t can be easily computed as $s_t|s_{\backslash t} \sim N\left(\gamma_t, \Xi_t\right)$. This probability enables us to sample which of the s_{1t} and s_{2t} is exactly equal to y_t, (e.g. by using a draw ν from a uniform distribution on $[0, 1]$ and comparing it to λ_t; that is, we set $s_{1t} = y_t$ if $\nu < \lambda_t$, and $s_{2t} = y_t$ otherwise). If

$s_{1t} = y_t$, then conditional on $s_{1t} < s_{2t}$,

$$s_{2t}|s_{\backslash t}, s_{1t} = y_t \sim TN_{s_{2t}>s_{1t}} \left(\bar{\gamma}_t, \bar{\Xi}_t \right),$$

for some mean and variance which are functions of the elements of γ_t and Ξ_t. Similar results hold if $s_{2t} = y_t$.

An important point is that this simulator will not be continuous with respect to the parameters, as λ_t depends on the parameters being estimated, and choosing which element of s_t is equal to y_t is a binary choice.

4.2 Parameter estimation

4.2.1 The literature

Due to computational complexity and to the lack of computationally tractable statistical methods, disequilibrium models have mostly been specified and estimated without dynamic structures. Classic papers in this context include Fair and Jaffee (1972), Goldfeld and Quandt (1975), Goldfeld and Quandt (1978), Maddala and Nelson (1974) and Hartley and Mallela (1977). A common difficulty revealed by this literature is that there can be a very significant problem of the ML estimator falling on a boundary of the parameter space, and often there are multiple maxima in the likelihood function.

With autocorrelation in disturbances or lagged latent dependent variables, likelihood functions of such models involve high dimensional integrals, typically of order n. Numerical approaches run into problems when n is larger than 4. Simulation techniques open up the ability to look at dynamic models. The simulated scores method developed by Hajivassiliou and McFadden (1996), directly simulates the derivatives of the log-likelihood function.

The first paper to use simulation techniques for the estimation of dynamic disequilibrium models is Laroque and Salanié (1993), who propose the use of dynamic simulated pseudo-maximum likelihood methods. This technique relies on using dynamic simulations of the endogenous variables conditionally on the exogenous variables to compute the first and second order moments of the endogenous variables, and maximising the resulting pseudo likelihood function to estimate the parameters. It requires the exogenous variables to be strongly exogenously for the parameters of interest.

We start the rest of this section by revising Laroque and Salanie's simulated pseudo-maximum likelihood method. Throughout this section, for any variable z, z^t will denote $(z_1, ..., z_t)'$. Then, we suggest how to perform Bayesian inference using the sampler introduced above. The next subsection will compare Laroque and Salanie's estimator to the Bayesian estimator in a Monte Carlo experiment. More general dynamics, including the model in Lee (1997a, Section 6), are

studied in Manrique (1997). In particular, in a Monte Carlo experiment Manrique (1997) compares Lee's approach to the Bayesian estimator we advocate in this paper. Both approaches yield biased estimators, and biases are approximately the same magnitude. Although the Bayesian estimator is not particularly well behaved for very small sample sizes, (e.g. $n = 30$), for $n = 50$ or $n = 100$ the Bayesian estimators appear to be more precise than Lee's estimates. Of course for large n they should be roughly equivalent.

4.2.2 Laroque and Salanie's dynamic simulated pseudo-maximum likelihood method

Laroque and Salanié (1993) propose a dynamic simulated pseudo-maximum likelihood method to deal with a very general class of dynamic non-linear models, which encompasses limited dependent variable models with lagged endogenous variables, both observed and latent, with or without serial correlation. This paper extends Laroque and Salanié (1989) who studied static disequilibrium models.

In short, dynamic pseudo-maximum likelihood (PML) methods consists of using dynamic simulations of the model, computing a pseudo likelihood function that depends only on the first two moments of the endogenous variables. Specifically, we start from the following reduced form model,

$$
\begin{aligned}
y_t &= g(x_t, y_{t-1}, s_{t-1}, \varepsilon_t, b_0) \\
s_t &= g^*(x_t, y_{t-1}, s_{t-1}, \varepsilon_t, b_0) \\
\varepsilon_t &= R_0 \varepsilon_{t-1} + u_t
\end{aligned}
$$

where y_t is the vector of observed endogenous variables, s_t is a vector of latent variables, x_t is a vector of strongly exogenous variables and ε_t are the structural disturbances. It is assumed that the innovations u_t are $i.i.d.$ with known distribution. The parameter $\theta_0 = (b_0, R_0)$ fully describes the data generation process.

Laroque and Salanié (1993) focus on the so-called dynamic PML2 estimator which uses $\{g_t\}$ is a sequence of transformations of the output of dynamic simulations of the model, starting at period 1 with given initial values for period 0 and replacing lagged dependent variables with their simulated values,

$$
\begin{aligned}
Y_t &= vec(y_t, ..., y_{t-k}), \\
G_t(x^t, u^t, \theta) &= vec\left\{ g_t(x^t, u^t, \theta), ..., g_{t-k}(x^{t-k}, u^{t-k}, \theta) \right\},
\end{aligned}
$$

for some integer $k \geq 0$, and F_t, V_t are the first two moments of G_t conditional on x^t, that is,

$$
F_t(x^t, \theta) = E\left\{ G_t(x^t, u^t, \theta) | x^t \right\}, \quad \text{and} \quad V_t(x^t, \theta) = Var\left\{ G_t(x^t, u^t, \theta) | x^t \right\}.
$$

Then the PML2 estimator corresponds to the value of θ which minimises

$$
l^n(\theta) = \frac{1}{2n} \sum_{t=1}^{n} \left[\left\{ Y_t - F_t(x^t, \theta) \right\}' V_t(x^t, \theta)^{-1} \left\{ Y_t - F_t(x^t, \theta) \right\} + \log \left| V_t(x^t, \theta) \right| \right]
$$

over θ.

Note that x_t needs to be strongly exogenous so that in the conditional expectations that we compute, the distribution of u^t can be taken independent of the conditioning variables. Under standard regularity conditions, this estimator is consistent and has satisfactory asymptotic properties. However, these conditions may not hold when some variables are not stationary.

As it will generally be impossible to write down an analytical form for the dynamic simulations G_t and their moments F_t and V_t, the function $l^n(\theta)$ can be approximated by simulations. This leads to the dynamic simulated pseudo-maximum likelihood (SPML) method.

Thus, at each period t, we draw H values of the innovations u_t, u_t^h. These draws, that must be independent over t and over h, will be held fixed during the minimisation of the criterion function. For each $h = 1, ..., H$, we then recursively compute a dynamic simulation path for the endogenous variables y_t^h. The dynamic SPML2 estimator minimises

$$l^{Hn}(\theta) = \frac{1}{2n} \sum_{t=1}^{n} \left[\left\{ Y_t - F_t^H(x^t, \theta) \right\}' V_t^H(x^t, \theta)^{-1} \left\{ Y_t - F_t^H(x^t, \theta) \right\} + \log \left| V_t^H(x^t, \theta) \right| \right]$$

over θ, where $F_t^H(x^t, \theta)$ and $V_t^H(x^t, \theta)$ stand for the empirical moments of $(Y_t^h)_{h=1,...,H}$.

Despite the generality and simplicity of the dynamic SPML method (the only requirements for applying it are that one can draw the innovations from their distribution and that it is possible to solve the model for the values of the endogenous variables), this technique yields simulators that are only as differentiable as the functions g and g^* with respect to the parameters, and so, they are discontinuous even in Probit models.

Laroque and Salanié (1989) study consistency and the asymptotic distribution of the static SPML2 estimator. Further study is required to extend those theoretical results to the dynamic SPML2 estimator.

Laroque and Salanié (1993) test the dynamic simulated pseudo maximum likelihood method on Monte Carlo generated data for the disequilibrium model defined as

$$\begin{aligned}
s_{1t} &= a_1 x_{1t} + \phi s_{1,t-1} + \sigma_1 \varepsilon_{1t}, \\
s_{2t} &= a_2 x_{2t} + \sigma_2 \varepsilon_{2t}, \\
y_t &= \min(s_{1t}, s_{2t}),
\end{aligned} \tag{7}$$

where y_t is an observable variable, for $t = 1, ..., n$, $s_t = (s_{1t}, s_{2t})'$ is a latent endogenous variable, ε_{1t} and ε_{2t} are independent, serially uncorrelated disturbances with $N(0,1)$ distributions. They assume $|\phi| < 1$ to ensure stationarity, and draw $x_t = (x_{1t}, x_{2t})'$ from

$$x_{1t} = 2.5(1 + \nu_t) \quad \text{and} \quad x_{2t} = 5, \tag{8}$$

with $\nu_t \sim N(0,1)$ and uncorrelated with $(\varepsilon_{1t}, \varepsilon_{2t})$. The implication is that the regressors are strongly exogenous. The true parameter values are $a_1 = 1$, $\sigma_1^2 = 1$, $\phi = 0.5$, $a_2 = 1$, and $\sigma_2^2 = 1$.

Finally,

$$s_{10} = (1 - \phi)^{-1} \{2.5 (1 + \nu) a_1 + \sigma_1 \varepsilon\}, \tag{9}$$

with $\nu, \varepsilon \sim NID(0,1)$. This set-up ensures that $E(s_{1t}) = E(s_{2t})$, so that there is a roughly equal mix of regimes.

They report estimation results obtained using a simple SPML approach based on $Y_t = y_t$, (i.e., $k = 1$), for 200 randomly generated 50-observations samples. The first two moments are approximated using $H = 10$, 20, or 50 dynamic simulations. The method appears to be quite tractable and it converges in most samples. However, it is rather difficult to obtain accurate estimates of the standard errors. To avoid the numerical computation of second derivatives, Laroque and Salanie use an approximation that relies on first derivatives but it is only valid when both the sample size and the number of simulations are infinite. Unfortunately, this yields mean estimates of the standard errors that are appreciably lower than the empirical standard deviation of the estimated parameters. In an earlier unpublished version of the paper, the authors apply the dynamic PML method to dynamic variants of the disequilibrium model of the US labour market with wage and price adjustment of Quandt and Rosen (1986).

4.2.3 Bayesian estimation

We resort to Bayesian estimation of dynamic disequilibrium models as the simulations are not continuous with respect to the parameters.

Assuming a known prior for θ, a generic Markov chain Monte Carlo sampler would take on the following form: (1) Initialise θ; (2) Sample s from $s|y; \theta$; (3) Sample θ from $\theta|s$; (4) Repeat from step 2.

4.3 A Monte Carlo study

For sake of simplicity, we only report results for the dynamic disequilibrium model considered in Laroque and Salanié (1993). More general dynamics, including the model in Lee (1997a), are studied in Manrique (1997).

4.3.1 Monte Carlo design

In our Monte Carlo experiments, we work with the simple disequilibrium model given in (7)-(8) with

$$s_{10} \sim N\left(\frac{2.5 a_1}{1 - \phi}, \frac{\sigma_1^2}{1 - \phi^2}\right), \tag{10}$$

where $x_t = (x_{1t}, x_{2t})'$ is strongly exogenous.

This is precisely the model considered by Laroque and Salanié (1993). We will interpret s_{1t} as latent demand. Demand depends on its lagged value. The true parameter values are $a_1 = 1$, $a_2 = 1$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$ and $\phi = 0.5$. The number of periods n is equal to 50.

Our aim will be to replicate the results of Laroque and Salanié (1993) in the next subsection, and then compare them, in a simple Monte Carlo experiment, to the Bayesian estimator we advocate in this situation.

In order to start the algorithms we need to set initial values for the parameters. These values were established in the same way as in Laroque and Salanie's paper. Thus, they were drawn from the uniform distributions on $[0.5, 1.5]$ for a_1, σ_1, a_2, and σ_2, and from the uniform distribution on $[0.25, 0.75]$ for ϕ.

4.3.2 Laroque and Salanie's method

For small samples $l^{Hn}(\theta)$ is not particularly well behaved. Laroque and Salanié (1993) identify three possible sources of spurious minima when applying the dynamic simulated pseudo-maximum likelihood to estimate this disequilibrium model, the 'zero-variance' minima, when the algorithm strays in a region in which σ_1^2 and/or σ_2^2 is close to zero, the 'one-sided' minima, when all observations are classified to belong to the same regime, and the cases where the autoregressive coefficient being equal or greater than 1 in absolute value. All three occur frequently in the Monte Carlo experiment.

Laroque and Salanie report results of a Monte Carlo experiment using 200 samples of size 50 and taking $H = 10$, $H = 20$, and $H = 50$ obtained with a simple SPML approach based only on $Y_t = y_t$, (i.e., $k = 1$). We perform 500 replications using the same sample size and the same amount of simulation. In order to balance the experiment favourably for the Laroque and Salanie procedure, we bounded the variances to be greater than or equal to 0.1, we imposed $|\phi| < 1$ and when we had numerical convergence problems or boundary estimates we typically double H and then applied the method again (until $H = 50$ when we stopped changing the parameters). Using this method we managed to get a lower percentage of non-converging simulation experiments than that reported in Laroque and Salanié (1993).

Using 500 replications we get the following results reported in Figure 5 and Table 6. The results indicate that typically the estimators become more precise as H increases, but that in the cases of a_1, σ_1 and σ_2 the parameters are not very precisely defined. The numerical optimisation procedure continues to fail a worryingly large percentage of the time, although this also falls as H increases.
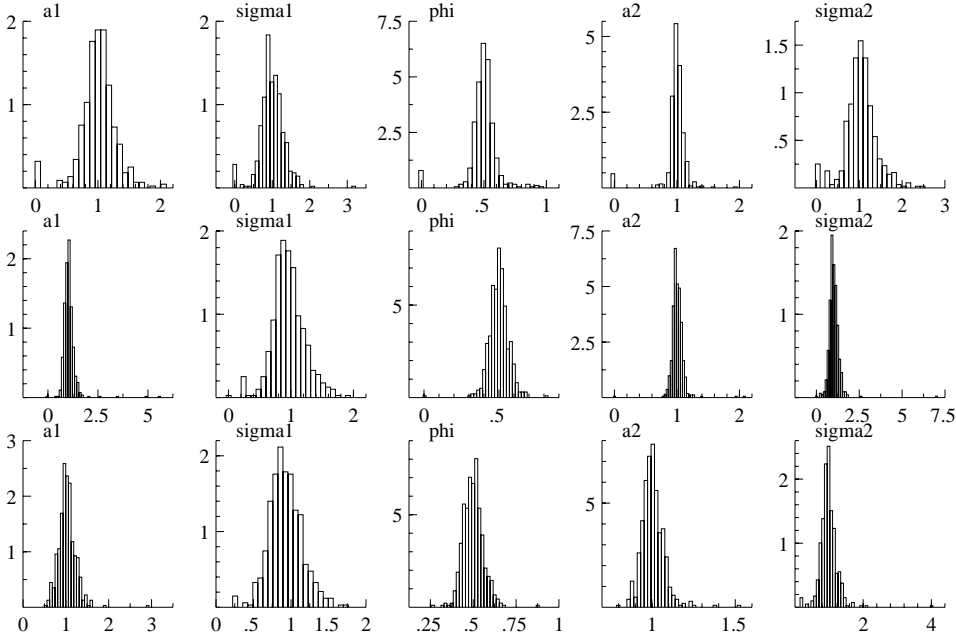
25

Figure 5: *Histograms of the 500 replications of the Laroque and Salanie's estimator of the disequilibrium model using $n = 50$. Top graphs have $H = 10$, middle $H = 20$ and bottom $H = 50$. The true parameters are 1.0, 1.0, 0.5, 1.0 and 1.0 as we go from the left to the right.*

4.3.3 Bayesian method

To start the scan sampler we need to set the initial values of demand and supply, $s_t = (s_{1t}, s_{2t})'$ for the whole sample period. We do this by setting both demand and supply equal to their observed minimum value. Initial conditions to run the Kalman filter on $s_1, ..., s_n$ are set to be the unconditional mean and unconditional variance, respectively. Alternatively, a diffuse initial condition could have been used.

We outline here how to sample from the posterior distribution $\theta | s$. In this framework $\theta = (a_1, \sigma_1^2, \phi, a_2, \sigma_2^2)'$. Given the structure of the model, we can estimate the parameters in the demand and supply equations separately. Thus, the parameter θ is partitioned into $\theta = (\theta_1, \theta_2)'$ where $\theta_1 = (a_1, \sigma_1^2, \phi)'$ and $\theta_2 = (a_2, \sigma_2^2)'$, and therefore step 3 of the MCMC sampler for Bayesian estimation can be divided into two parts: (3a) sample θ_1 from $\theta_1 | s_1$. where s_1. $= (s_{11}, ..., s_{1n})'$ is the $n \times 1$ vector of the demand quantities; (3b) sample θ_2 from $\theta_2 | s_2$. where s_2. $= (s_{21}, ..., s_{2n})'$ is the $n \times 1$ vector of the supply quantities.

When we update the parameters, we use the following conditional structure: (1) $a_1 | s_1, \sigma_1^2, \phi$, (2) $\sigma_1^2 | s_1, a_1, \phi$, (3) $\phi | s_1, a_1, \sigma_1^2$, (4) $a_2 | s_2, \sigma_2^2$, and (5) $\sigma_2^2 | s_2, a_2$. All the parameters but ϕ have straightforward conjugate distributions which we use. In particular, we use a non-informative

H	Converge	a_1	σ_1	ϕ	a_2	σ_2
10	97.2	1.029	1.019	0.5098	1.012	1.064
		(.234)	(.281)	(.082)	(.100)	(.323)
20	99.8	1.058	0.9597	0.5012	1.008	1.024
		(.357)	(.256)	(.061)	(.096)	(.366)
50	100.0	1.030	0.9276	0.4996	1.005	0.9847
		(.205)	(.219)	(.059)	(.068)	(.260)
TRUTH		1.0	1.0	0.5	1.0	1.0

Table 6: Summaries of Figure 5. Laroque and Salanie's method for various values of H. Throughout we use $n = 50$ and perform 500 replications. Figures in brackets are the estimated standard errors of the method, computed using the simulation. Converge denotes the proportion of the replications which converge.

prior distribution for a_1 and a_2, take $L_0 \cdot \chi_p^{-2}$ for σ_1^2, and $L_1 \cdot \chi_q^{-2}$ for σ_2^2. Throughout we take $q = p = 5$ and $L_0 = p \times 1$ and $L_1 = q \times 1$. As before, we use $2Beta(\zeta_1, \zeta_2) - 1$ with $\zeta_1 = 10$, $\zeta_2 = 2$ as a prior family for ϕ.

Illustration We first look at a Monte Carlo experiment on a single data set. Later we will look at the sampling experiment for 500 replications.

Figure 6 and Table 7 give the results. We iterate the sampler on the demand and supply series for 500 iterations, then on the parameters and demand and supply series for a further 1,000 iterations, before recording any answers. The next 10,000 iterations are recorded. The correlograms die out very quickly. The parameter a_1 is positively correlated with σ_1, and negatively correlated with ϕ and a_2. An interesting feature is the small inefficiency factor for ϕ. The inefficiency factors of less than 10 for all the parameters, suggest that these models can be estimated reasonably precisely with only about 1,000 iterations of the scan sampler.

Sampling behaviour of Bayesian estimators Now we repeat the experiment 500 times to analyse the sampling behaviour of the parameter estimates. We use the same data as was used in the previous subsection on Laroque and Salanie's procedure. Hence the only difference in the estimates is due to the use of different procedures and the fact that both are simulation based estimators.

We iterate the sampler on the demand and supply series for 100 iterations, then on the parameters and demand and supply series for a further 50 iterations, before recording any answers. The next 1,000 iterations are recorded. The results of the experiment are presented in Table 8 and Figure 7. The histograms of the estimates are quite symmetric. The sampler yields biased estimates. Biases are rather small for such a small sample size. Results are exactly
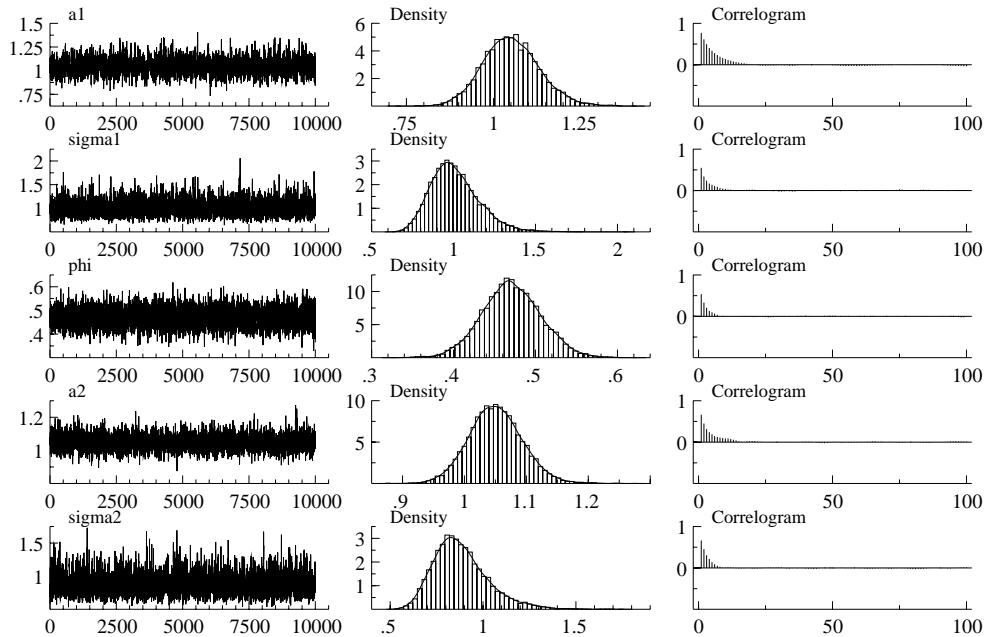
Figure 6: *Gibbs sampler for a dynamic disequilibrium model with $n = 50$, $a_1 = 1$, $\sigma_1 = 1$, $\phi = 0.5$, $a_2 = 1$ and $\sigma_2 = 1$. Left graphs: the simulation against iteration number. Middle graphs: histograms of the resulting marginal distributions and estimated densities. Right graphs: the corresponding correlograms for the iterations.*

comparable with the earlier Table 6 reported for Laroque and Salanie's estimator. The Bayesian estimator seems to be about 2 or 3 times as efficient. Further, the Bayesian estimator was computed without convergence problems.

5   CONCLUSION

In this paper we have developed methods to perform likelihood inference for some Gaussian limited dependent processes. We have designed quite efficient MCMC methods which deliver simulations from the distribution of the latent data given the observed data and a fixed parameter value. The basic approach has been to simulate from $s_t|s_{\backslash t}, y; \theta$, that is, to sample a single s_t conditional on the most recent updated values of all the other latent points, the observations and some parameter estimates, repeating this operation backwards. The problem can then be reduced to one of simply computing the Gaussian distribution of s_t given all the other latent points, and then sampling from the constrained space induced by knowing y_t.

The state space framework enables either Bayesian inference or classical inference through a simulated score method (simulated EM algorithm) when the simulations are continuous functions

28

	TRUTH	Mean	MCse	Inefficiency	
$a_1|y$	1.0	1.051	.00208	6.7	
$\sigma_1|y$	1.0	1.009	.00342	5.4	
$\phi_1|y$	.5	.4708	.000596	2.8	
$a_2|y$	1.0	1.050	.00109	6.4	
$\sigma_2|y$	1.0	.8808	.00306	4.4	
		Covariance and *Correlation*			
$a_1|y$	.00649	*.171*	*-.398*	*-.220*	*.0188*
$\sigma_1|y$	.00203	.0217	*.0573*	*-.0742*	*-.0628*
$\phi_1|y$	-.00115	.000302	.00128	*-.0735*	*-.0316*
$a_2|y$	-.000762	-.000471	-.000113	.00186	*.0339*
$\sigma_2|y$	.000222	-.00136	-.000166	.000214	.0215

Table 7: Summaries of Figure 6. MCse denotes the Monte Carlo standard error of the simulation estimator of mean of the posterior density. Throughout these standard errors are computed using 500 lags and 10,000 iterations. Numbers in italics are correlations rather than covariances.

	a_1	σ_1	ϕ	a_2	σ_2
	0.9833	1.027	0.5139	1.014	1.054
	(.131)	(.132)	(.048)	(.074)	(.161)
TRUTH	1.0	1.0	0.5	1.0	1.0

Table 8: Summaries of Figure 7. Bayesian posterior mean estimator using 1,000 iterations. Throughout we use $n = 50$ and perform 500 replications. Results correspond exactly to the Monte Carlo design for the simulation reported in Figure 5 for the Laroque and Salanie's estimator. Figures in brackets are the estimated standard errors of the method, computed using the replications.

of the model parameters. Results of some Monte Carlo experiments have been reported.

The approach we have advocated can be generalised in a number of directions. Firstly, this type of method can be used within the context of long panels, where the time series dimension is computationally important. Second, some forms of non-Gaussian dynamics can be introduced without disrupting the structure of the models we have discussed here. In particular the conditional Gaussian or partially non-Gaussian time series models of Carter and Kohn (1994) and Shephard (1994) can be used. This allows, for example, for time series dynamics which experience outliers or breaks to be modelled. Further, stochastic volatility is easily accommodated in this framework.
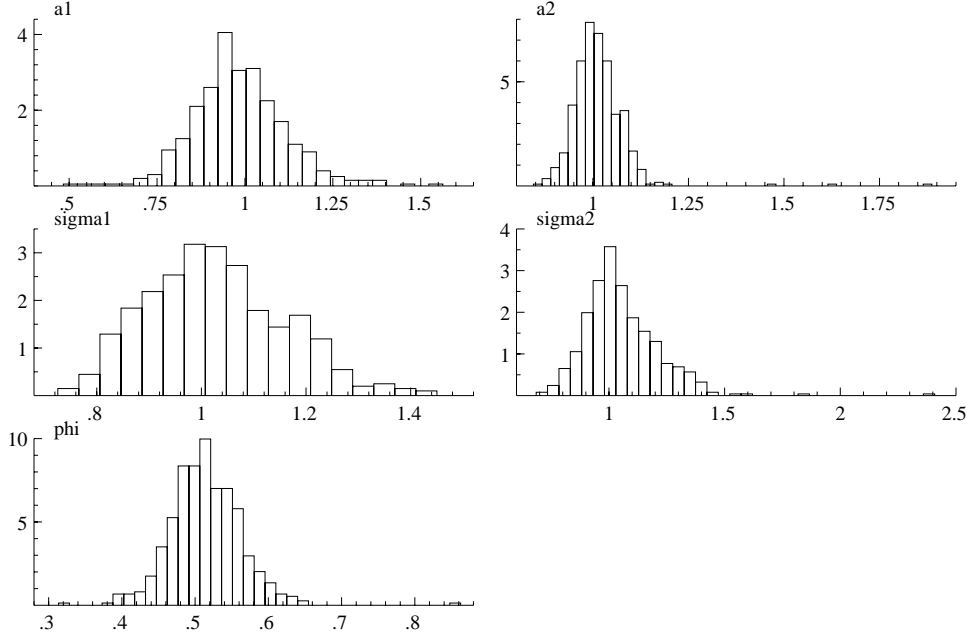
Figure 7: *Histogram of Bayesian estimator using 1,000 iterations. Throughout we use $n = 50$ and perform 500 replications. True parameter values are $a_1 = 1$, $a_2 = 1$, $\sigma_1 = 1$, $\sigma_2 = 1$ and $\phi = 0.5$.*

6 COMPUTATIONAL APPENDIX

6.1 The model

This section details the algorithms given in this paper for general Gaussian state space form for the time series s_t given in (1). Typically c_t, Z_t, G_t, d_t, T_t and H_t will be assumed to be fixed and indexed by a small dimensional parameter θ, while we will write the signals $c_t + Z_t \alpha_t$ as μ_t. For simplicity, we set $\sigma^2 = 1$.

6.2 Kalman filter

The Kalman filter (see, for example, de Jong (1989)) has the following form:

1. Initialise $t = 0$, $a_{1|0} = 0$, and $P_{1|0} = H_0 H_0'$

2. Compute

$$
\begin{array}{llll}
v_t & = & s_t - c_t - Z_t a_{t|t-1}, & F_t & = & Z_t P_{t|t-1} Z_t' + G_t G_t', \\
K_t & = & \left(T_t P_{t|t-1} Z_t' + H_t G_t' \right) F_t^{-1}, & L_t & = & T_t - K_t Z_t, \\
a_{t+1|t} & = & d_t + T_t a_{t|t-1} + K_t v_t, & P_{t+1|t} & = & T_t P_{t|t-1} L_t' - H_t G_t' K_t' + H_t H_t'.
\end{array}
\tag{11}
$$

3. Write $t = t + 1$. Repeat from 2 if $t \leq n$.

30

Here K_t is called the Kalman gain, while v_t and F_t are the one-step ahead prediction error (or innovation) and its mean square error, respectively.

The KF outputs allows the computation of the log-likelihood function via the prediction decomposition, for, ignoring constants

$$\log f\left(s_1,...,s_n;\theta\right) = \sum_{t=1}^{n} \log f\left(s_t|s_1,...,s_{t-1};\theta\right) = -\frac{1}{2}\sum_{t=1}^{n} \log |F_t| - \frac{1}{2}\sum_{t=1}^{n} v_t' F_t^{-1} v_t. \tag{12}$$

6.3   Simulation state smoother

The simulation state smoother draws from the state smoothing distribution $\alpha|s;\theta$.

Set R_t to equal to the non-zero rows of the H_t matrix. An example of this is an AR(2) with measurement error. Using an obvious notation:

$$z_t = \left(\begin{array}{cc} 1 & 0 \end{array}\right), G_t = \left(\begin{array}{cc} \sigma_\varepsilon & 0 \end{array}\right), T_t = \left(\begin{array}{cc} \phi_1 & 1 \\ \phi_2 & 0 \end{array}\right), H_t = \left(\begin{array}{cc} 0 & \sigma_\eta \\ 0 & 0 \end{array}\right), R_t = \left(\begin{array}{cc} 0 & \sigma_\eta \end{array}\right).$$

The simulation state smoother (due to de Jong and Shephard (1995)) has the following form:

1. Requires that v_t, F_t and K_t be stored from the KF run on $s_1,...,s_n$.

2. Set $t = n$, $r_n = 0$ and $N_n = 0$.

3. Compute with $L_t = T_t - K_t Z_t$

$$\begin{array}{llll} C_t & = & R_t R_t'\left(I - N_t\right)R_t R_t', & \kappa_t & \sim & N(0,C_t), \\ r_{t-1} & = & Z_t' F_t^{-1} v_t + L_t' r_t - V_t' C_t^{-1} \kappa_t, & N_{t-1} & = & Z_t' F_t^{-1} Z_t + L_t' N_t L_t + V_t' C_t^{-1} V_t, \end{array}$$
$$\tag{13}$$

   where $V_t = R_t R_t' N_t L_t$.

4. Record $\eta_t = R_t R_t' r_t + \kappa_t$. Repeat for $t = n,...,1$.

The end condition is calculated by $C_0 = P_{1|0} - P_{1|0} N_0 P_{1|0}$, $\kappa_0 \sim N(0,C_0)$, yielding $\eta_0 = P_{1|0} r_0 + \kappa_0$. The resulting η_t corresponds to a draw from $H_t u_t | s; \theta$.

Once the disturbances of the states are drawn it is possible to reconstruct the state via the running of the forward recursion starting with $a_{0|n} = 0$,

$$a_{t+1|n} = d_t + T_t a_{t|n} + \eta_t, \qquad t = 0,...,n-1.$$

6.4   Simulation signal smoother

The simulation signal smoother (due to de Jong and Shephard (1995)) draws from $\mu|s;\theta$.

Setting $r_n = 0$ and $N_n = 0$, for $t = n,...,1$, and writing $D_t = F_t^{-1} + K_t' N_t K_t$, $n_t = F_t^{-1} v_t - K_t' r_t$,

$$\begin{array}{llll} C_t & = & \Sigma_t - \Sigma_t D_t \Sigma_t, & \kappa_t & \sim & N(0,C_t), \\ r_{t-1} & = & Z_t' F_t^{-1} v_t + L_t' r_t - V_t' C_t^{-1} \kappa_t, & N_{t-1} & = & Z_t' F_t^{-1} Z_t + L_t' N_t L_t + V_t' C_t^{-1} V_t. \end{array} \tag{14}$$

31

where $V_t = \Sigma_t \left(D_t Z_t - K_t' N_t T_t \right)$ and we store $\Sigma_t n_t + \kappa_t$. Then $s_t - \Sigma_t n_t - \kappa_t$ is a draw from the signal $\mu_t | s; \theta$.

All the programs have been written using the object orientated matrix programming language Ox of Doornik (1996).

References

Barndorff-Nielsen, O. E. and D. R. Cox (1994). *Inference and Asymptotics.* London: Chapman & Hall.

Berzuini, C., N. G. Best, W. R. Gilks, and C. Larizza (1997). Dynamic conditional independence models and Markov chain Monte Carlo methods. *J. American Statistical Association 92*, 140312.

Bresnahan, T. F. (1981). Departures from marginal-cost pricing in the American automobile industry, estimates from 1977-1978. *J. Econometrics 17*, 201–227.

Carlin, B. P. and N. G. Polson (1992). Monte Carlo Bayesian methods for discrete regression models and categorical time series. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4.* Oxford: Oxford University Press.

Carlin, B. P., N. G. Polson, and D. Stoffer (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modelling. *J. American Statistical Association 87*, 493–500.

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*, 541–53.

Chan, K. S. and J. Ledolter (1995). Monte Carlo EM estimation for time series models involving counts. *J. American Statistical Association 89*, 242–52.

Chib, S. (1992). Bayes inference in the Tobit censored regression model. *J. Econometrics 51*, 79–99.

Chib, S. and E. Greenberg (1994). Bayes inference for regression models with ARMA(p,q) errors. *J. Econometrics 64*, 183–206.

Danielsson, J. and J. F. Richard (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *J. Applied Econometrics 8*, S153–S174.

de Jong, P. (1989). Smoothing and interpolation with the state space model. *J. American Statistical Association 84*, 1085–8.

de Jong, P. (1997). The scan sampler. *Biometrika 84*, 929–37.

de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*, 339–50.

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

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.

Doornik, J. A. (1996). *Ox: Object Oriented Matrix Programming, 1.10*. London: Chapman & Hall.

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space model. *Biometrika 84*, 669–84.

Fair, R. C. and D. M. Jaffee (1972). Methods of estimation for markets in disequilibrium. *Econometrica 40*, 497–514.

Gilks, W. K., S. Richardson, and D. J. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

Goldfeld, S. M. and R. E. Quandt (1975). Estimation in a disequilibrium model and the value of information. *J. Econometrics 3*, 325–348.

Goldfeld, S. M. and R. E. Quandt (1978). Some properties of the simple disequilibrium model with covariance. *Economic Letters 1*, 343–346.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993). A novel approach to non-linear and non-Gaussian Bayesian state estimation. *IEE-Proceedings F 140*, 107–13.

Hajivassiliou, V. and D. McFadden (1996). The method of simulated scores, with application to models of external debt crises. *Econometrica 64*. Forthcoming.

Hajivassiliou, V. A., D. McFadden, and P. A. Ruud (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *J. Econometrics 72*, 85–134.

Hajivassiliou, V. A. and P. A. Ruud (1994). Classical estimation methods for LDV models using simulation. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Volume 4*, pp. 2383–441. Amsterdam: North-Holland.

Hartley, H. O. and P. Mallela (1977). The asymptotic properties of a maximum likelihood estimator for a model of markets in disequilibrium. *Econometrica 45*, 1205–1220.

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

Hasbrouck, J. (1996). The dynamics of discrete bid and ask quotes. Unpublished paper: Stern Business School, New York University.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*. Forthcoming.

Koopman, S. J. and N. Shephard (1992). Exact score for time series models in state space form. *Biometrika 79*, 823–6.

Laroque, G. and B. Salanié (1989). Estimation of multi-market disequilibirum fix-price models: An application of pseudo maximum likelihood methods. *Econometrica 57*, 831–860.

Laroque, G. and B. Salanié (1993). Simulation-based estimation of models with lagged latent variables. *J. Applied Econometrics 8*, S119–S133.

Lee, L.-f. (1997a). Simulation estimation of dynamic switching regression and dynamic disequilibrium models — some Monte Carlo results. *J. Econometrics 78*, 179–204.

Lee, L.-f. (1997b). A smooth likelihood simulator for dynamic disequilibrium models. *J. Econometrics 78*, 257–294.

Liu, J., W. H. Wong, and A. Kong (1994). Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes. *Biometrika 81*, 27–40.

Louis, T. A. (1982). Finding observed information using the EM algorithm. *J. Royal Statistical Society B 44*, 98–103.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press.

Maddala, G. S. and F. D. Nelson (1974). Maximum likelihood methods for models of markets in disequilibrium. *Econometrica 42*, 1013–1030.

Manrique, A. (1997). Econometric analysis of limited dependent time series. Unpublished, D.Phil. Thesis, Oxford University.

Pitt, M. K. and N. Shephard (1997). Filtering via simulation based on auxiliary particle filters. Unpublished paper, Nuffield College, Oxford.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. London: Academic Press.

Qian, W. and D. M. Titterington (1991). Estimation of parameters in hidden Markov Chain models. *Philosophical Transactions of Royal Society of London A. 337*, 407–28.

Quandt, R. E. and H. Rosen (1986). Unemployment, disequilibrium and the short-run Phillips curve: an econometric approach. *J. Applied Econometrics 1*, 235–253.

Ruud, P. (1991). Extensions of estimation methods using the EM algorithm. *J. Econometrics 49*, 305–341.

Sandmann, G. and S. J. Koopman (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *J. Econometrics*. Forthcoming.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika 81*, 115–31.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–67.

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

Wei, G. C. G. and M. A. Tanner (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. American Statistical Association 85*, 699–704.

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models* (2 ed.). New York: Springer-Verlag.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics 11*, 95–103.