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 University, Oxford OX1 1NF, UK
neil.shephard@nuf.ox.ac.uk

April 18, 1997

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, Gibbs sampler, Markov chain Monte Carlo, Scan sampler, Tobit model.

1 INTRODUCTION

1.1 General problem

Suppose $s \in R^{n \times p \times q}$ 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 $s_{tij}, t = 1, ..., n, i = 1, ..., p, j = 1, ..., q$. 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 and, conditional on some explanatory variables, the s_{tij} are independent over j. This very interesting literature is reviewed by Hajivassiliou and Ruud (1994). Our exposition will typically suppress the third subscript in s_{tij} as it adds nothing new to the discussion.

In this paper we will focus mainly on time series problems, although some of the methods we develop could be helpful in panel data contexts. 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 typically deliver likelihood based estimators in $O(np^4)$ 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 de Jong (1989)). 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

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

Typically $c_t, Z_t, G_t, H_t, d_t, T_t$ 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. We call the s_t the latent time series. Examples of models which can be put in this framework are VARs, structural time series, continuous time models observed irregularly and moving averages (Harvey (1989)). 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 $O(np^3)$ the following: (1) the joint density $f_S(s;\theta)$ by the Kalman filter (de Jong (1989)), (2) simulate from $\alpha|s$, or $\mu|s$ using the simulation state and signal smoothers (de Jong and Shephard (1995)), (3) simulate with replacement from $s_t|s_1,...,s_{t-1},s_{t+1},...,s_n$ using the scan sampler (de Jong (1996)). 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 scan sampler 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 $h()$ 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 structures

The univariate Tobit model is extensively studied and writes $y_t = \max(0, s_t)$.

1.2.2 Probit structures

The univariate Probit model writes $y_t = I(s_t > 0)$, while a trivariate system might have

$$y_t = \left\{ \begin{array}{l} I(s_{1t} - s_{2t} > \delta_1) \\ I(s_{1t} - s_{2t} < \delta_2) \\ s_{2t} \end{array} \right\},$$

where s_t is bivariate. Here some of s_t is observed, while the others are subject to the typical Probit cutoff. A simple case of this is where s_{1t} and s_{2t} are both nonstationary but obey a cointegrating relationship with $s_{1t} - s_{2t}$ being stationary. An interesting empirical example of this situation is that of the interest rate charged by building societies in the United Kingdom (see, for example, Anderson and Hendry (1984)) which moves only periodically and only when it becomes significantly out of line with rates charged or offered by its competitors.

1.2.3 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. Quandt (1982) reviews the econometric literature on this topic. 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 Salanie (1993) and Lee (1995a).

1.2.4 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 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 denote the unobserved implicit efficient price of a security, $\mu_t - \beta_t$ be the bid price in the absence of discreteness restrictions and $\mu_t + \alpha_t$ is the corresponding ask price. Here β_t, α_t reflects the non-negative cost of quote exposure for small trades.

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. Hasbrouck (1996) assumed μ_t follows a random walk, while α_t was assumed to be an log-autoregression of the form

$$\{\log(\alpha_t) - \gamma\} = \phi\{\log(\alpha_{t-1}) - \gamma\} + v_{\alpha t},$$

where the error term was Gaussian. A similar model was specified for β_t. Thus we have that $s_t = \{\mu_t, \log(\alpha_t), \log(\beta_t)\}'$ and

$$y_t = \left\{ \begin{array}{c} Floor(s_{1t} - \exp s_{3t}) \\ Ceiling(s_{1t} + \exp s_{2t}) \end{array} \right\}.$$

1.3 Likelihood inference

1.3.1 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 $f_S(s;\theta)$ while we will work with respect to the posterior distribution of the complete data $F_{S|Y}(s|y;\theta)$. Importantly this distribution has only positive probability at the points s such that $y = h(s)$. Then

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

Generally $F_{S|Y}(s|y;\theta)$ is intractable, however it will be seen that we can simulate from it, giving the possibility of unbiasedly estimating the score. We will write in general

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

Iterating

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

performs a series of stochastic 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 nice textbook exposition of this material is given in Tanner (1996, Ch. 4). Typically this literature considers the use of simulation as a convenient way of approximating the expectation step of the EM algorithm and so intends the applied worker to take the number of simulations R as being very large in empirical examples.

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 $\partial Q(\theta_0,\theta_0)/\partial\theta$ is zero, where θ_0 is the true value of the parameter. 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 smooth as a function of θ then[1] the estimating equation

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

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

The requirement that the simulations from $F(s|y;\theta)$ be smooth in θ is quite restrictive as it means that we cannot use rejection type methods to generate the simulations. Of course it can be removed if we were to smooth the estimated score in some way — in very early work on the use of simulation to perform inference Diggle and Gratton (1984) perform a sort of smoothing to remove the non-differentiability in their estimated likelihood function. However, this will not be discussed here as it reduces the simplicity of the approach.

1.3.2 Bayes estimator

An alternative likelihood approach is to advocate the use of a Bayes estimator, which focuses on the posterior density of $\theta|y$. This density is basically intractable when tackled directly, however

[1] This is quite a restrictive assumption, basically ruling out the use of rejection.

again the use of simulation delivers easy methods. Suppose we can simulate from $F(s|y;\theta)$ and in addition that we can simulate from $f(\theta|s)$. 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. Initialize θ.

2. Sample $s \sim F(s|y;\theta)$.

3. Sample $\theta \sim f(\theta|s)$.

4. Goto 2.

This type of algorithm will, under some rather weak regularity, converge to a draw from $(s', \theta')' |y$ using Markov chain Monte Carlo results. Averaging subsets of these simulations will lead to likelihood based Bayes estimators of the parameters. The resulting 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 inference.

Sometimes the sampling from $\theta|s$ is generically difficult (although in special cases it may not be), even though we can evaluate the likelihood $f(s;\theta)$ and can sample from the conditionals one at a time rather than the full density.

An alternative would be to add another line of simulation and so switch to the sampler

1. Initialize θ, s.

2. Sample $s|y;\theta$.

3. Sample $\alpha|s,\theta$.

4. Sample $\theta|\alpha, s$.

5. Goto 2.

This procedure's advantage is that it is usually easy to sample $\theta|\alpha, s$ and is the approach we advocate in this paper for the generic problem.

1.4 Outline of paper

In the next section we will discuss various ways of constructing general algorithms for simulating from $s|y;\theta$. A key role will be given to the scan sampler. Section 3 will then discuss parameter estimation in the context of a Tobit model, while Section 4 looks at a simple dynamic disequilibrium model. Section 5 concludes. Finally, the Appendix details various algorithms featured in the paper.

2 SIMULATING FROM $s|y;\theta$

2.1 Conditional structure

The task will be to simulate from $s|y;\theta$ where we will assume s follows the Gaussian state space form and each s_t has the deterministic constraint that $y_t = h(s_t)$. This is a particular form of the non-Gaussian SSF, where the 'measurement equation' is a deterministic and non-invertible function of a Gaussian SSF. As such, methods developed to efficiently handle non-Gaussian stochastic measurement time series, such as Shephard and Pitt (1997), will not to very helpful and there is a need for a new strategy.

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 with replacement a single s_t at a time given all the other latent points $s_{\backslash t} = (s_1, ..., s_{t-1}, s_{t+1}, ..., s_n)'$ and the observations y, repeating this operation for $t = n, ..., 1$. However, this simplifies due to the structure of the model as

$$\begin{array}{rcllll} \Pr(s_t|s_{\backslash t}, y;\theta) & \propto & \Pr(s_t|s_{\backslash t};\theta)\Pr(y_t|s_t;\theta) & = & \Pr(s_t|s_{\backslash t};\theta), & \forall\ s_t \text{ s.t. } \quad y_t = h(s_t), \\ & & & = & 0, & \text{elsewhere,} \end{array}$$

and the model sets $y_t = h(s_t)$. Hence the problem reduces to one of simply efficiently computing the Gaussian distribution of $s_t|s_{\backslash t};\theta$ and then sampling from the constrained space induced by knowing y_t.

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 and was solved by de Jong (1989). This is a mistake. The problem here is to evaluate this density and sample with replacement, carrying this out for $t = n, ..., 1$ in only $O(np^3)$. Hence, for example, 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

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

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 $O(n^2p^3)$ operation.

Luckily for us this type of problem has recently been solved by de Jong (1996) in a result he called the scan sampler. We believe this is a very important result and deserves some attention in the econometrics literature. It has the following form in order to repeatedly sample from $s_t|s_{\backslash t};\theta$ with replacement:

Scan sampler (due de Jong (1996))

1. Requires that F_t^{-1}, v_t and K_t be stored from the Kalman filter on $s_1,...,s_n$.

2. Set $t=n$, $r_n=0$ and $N_n=0$ and then run for $t=n,...,1$.

3. Compute with $L_t=T_t-K_tZ_t$

$$\begin{array}{rclrcl} e_t & = & F_t^{-1}v_t-K_t'r_t, & s_t^n & \sim & \mathsf{N}(s_t-D_t^{-1}e_t,D_t^{-1}) \\ r_{t-1} & = & Z_t'F_t^{-1}v_t+L_t'r_t-V_t'\left(s_t-s_t^n\right), & N_{t-1} & = & Z_t'F_t^{-1}Z_t+L_t'N_tL_t, \end{array} \tag{1}$$

   where

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

4. Let $t=t-1$. Goto 3 if $t>1$.

The result is a new latent time series s^n which comes about from a complete sweep of the scan sampler. The scan sampler has close similarities to both the jackknife of de Jong (1989) and the simulation signal smoother of de Jong and Shephard (1995) which draws $\mu|s;\theta$.

It is important to note that this 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 does not exist linear combinations of s_j which are perfectly predictable from $s_{\backslash j}$ for any j. If such a model did 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 amend it in a straightforward way by simply replacing the sampling step

$$s_t^n \quad \sim \quad \mathsf{N}(s_t-D_t^{-1}e_t,D_t^{-1})=f(s_t^n|s_1,...,s_{t-1},s_{t+1}^n,...,s_n^n),$$

by the more involved sampling problem of drawing from

$$\begin{array}{rcll} \Pr(s_t^n|s_1,...,s_{t-1},s_{t+1}^n,...,s_n^n,y_t;\theta) & \propto & \Pr(s_t^n|s_1,...,s_{t-1},s_{t+1}^n,...,s_n^n;\theta), & \forall\ s_t \text{ s.t. } \quad y_t=h(s_t), \\ & = & 0, & \text{elsewhere.} \end{array}$$

Hence the only issue left is simply one of carrying out this slightly more involved sampling problem: sampling from a normal distribution with a constrained support. Thus the scan sampler has completely dealt with the time series aspect of the problem of handling these models. We will return to this constrained sampling problem after the next subsection.

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 which will provide simulations from $s|y;\theta$. Although each alternative will have a geometric rate of convergence, under some weak assumptions, to $s|y;\theta$ it does not mean that they are all equally good or general. Here we highlight two alternatives which are in general inferior to the use of scan sampling but could be used in this context. A thorough empirical comparison of their relative performance for the examples given below is given in detail in Manrique (1997). Here we just mention the results.

2.3.1 Signal simulation smoothing

In some recent work, de Jong and Shephard (1995) showed how to efficiently sample in $O(np^3)$ computations without iteration or rejection from the np dimensional normal distribution of $\mu|s;\theta$. Clearly if $G_tG_t' > 0$ for all t then we might use the following scheme to sample form $s|y;\theta$.

1. Initialize s
2. Sample $\mu \sim \mu|s;\theta$.
3. Sample $s \sim s|\mu, y;\theta$
4. Goto 2.

The advantage of this scheme is that $\Pr(s|\mu, y) = \prod_{t=1}^{n} \Pr(s_t|\mu_t, y_t;\theta)$, which should be relatively easy to sample. The difficulty with this approach would come if G_tG_t' was close to zero, then knowing μ is very close to knowing s and so we might expect the sampler to converge quite slowly. In the limit as $|G_tG_t'| \to 0$ it would not converge at all. This is not a toy example, for 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. This can be compared to the situation encountered in Shephard and Pitt (1997) where the non-Gaussian measurement densities provided enough non-Gaussian error to allow this style of argument to work.

2.3.2 State sampler

The same problem occurs when we work with a MCMC sampler for the states, as suggested by Carlin and Polson (1992) in this context. 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. Initialize s and α.
2. Sample $s|\alpha, y; \theta$
3. Sample $\alpha_t|\alpha_{t-1}, \alpha_{t+1}, s_t; \theta$, for $t = 1, ..., n$
4. Goto 2.

This sampler is uniformly worse than the simulation signal smoothing algorithm just described as it performs an unnecessary Gibbs sampler at step 3. Further, it does not overcome the measurement error difficulties we mentioned that can cause problems for these types of samplers.

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 the multivariate normal distribution $s_t|s_{\backslash t} \sim N(\gamma_t, \Sigma_t)$ subject to constraints that $y_t = h(s_t)$. For some simple problems this can be carried out by directly drawing from this distribution. In general we have to rely on a Markov chain Monte Carlo technique to make suggestions for possible new samples for $s_t|s_{\backslash t}$ or for elements of that vector.

In general this is a difficult problem and has itself generated its own literature. A review of this literature is given in Hajivassiliou and Ruud (1996).

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

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

This setup covers important cases such as the multinomial Probit and Tobit models, although we cannot put disequilibrium models and the bid/ask price model discussed above 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 and Ruud (1996, p. 110-1). Notice there is no need to iterate until convergence this sampler if the scan sampler is being used as the scan sampler is itself a Gibbs sampler.

For more complicated functions $y_t = h(s_t)$ it seems difficult to be very prescriptive, rather 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

The model $y_t = \max(0, s_t)$, for a linear regression model for s_t was suggested in pioneering work by Tobin (1958). A review of much of the existing literature on this topic is given in Maddala (1983). This section will extend this work to allow for a long time series dimension in s_t.

Chib (1992) proposes a MCMC method for the Tobit regression model, where $s_t = x_t'\beta + \varepsilon_t$, where $\varepsilon_t \sim NID(0, \sigma^2)$ and the regressors are thought of as being strongly exogenous. Now $\theta = (\beta', \sigma^2)'$ and Chib studied setting $\mu_t = x_t'\beta$ and then running

- Initialize s.
- Draw $\theta|s$, and then construct μ.
- Draw $s|\mu, \theta, y$
- Goto 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, this assumption becomes less convincing in the time series context.

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

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

$$s_t|s_{\backslash t}, y_t = y_t, \quad \text{if} \quad y_t > 0,$$

and, writing $s_t|s_{\backslash t}; \theta \sim N(\gamma_t, \Sigma_t)$,

$$s_t|s_{\backslash t}, y_t; \theta \sim TN_{s_t<0}(\mu_t, \Sigma_t), \quad \text{if} \quad y_t \leq 0.$$

Here the notation $TN_{s_t<0}$ means that s_t has a truncated normal distribution and $s_t < 0$. As we can sample smoothly from truncated normal distributions, the resulting Gibbs sampler will be smooth in the parameters and will converge to $s|y;\theta$ assuming s_t is not perfectly predictable from $s_{\backslash t}$. In all our experiments we will always initialize $s_t = y_t$ if $y_t > 0$ and -0.5 otherwise.

To illustrate the relative performance of the possible MCMC samplers for this problem, we consider a simple experiment on the following model, where we assume $|\phi| < 1$,

$$\begin{array}{lcll} s_t & = & \beta + \alpha_t + \varepsilon_t, & \varepsilon_t \sim NID\left(0, \sigma_\varepsilon^2\right), \\ \alpha_{t+1} & = & \phi\alpha_t + \eta_t, & \eta_t \sim NID(0, \sigma_\eta^2), \\ \alpha_1|S_0 & \sim & N\left\{0, \sigma_\eta^2/\left(1-\phi^2\right)\right\}. & \end{array}$$

We assume ε_t and η_t are mutually uncorrelated, while S_0 denotes the information available about α_1 at time 0.

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 three different Gibbs sampler for $50,000$ iterations, discarding the first $1,000$ results, for the above set of parameter values. The three samplers are the scan sampler, multi-move and the single move algorithms discussed in Section 2. 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 single-move sampler has a higher degree of correlation than the multi-move sampler, which in turn, is more correlated than the scan sampler. In all cases the correlation is modest, but that is because we are considering a very simple model.

3.3 Estimation

3.3.1 Bayesian inference

Likelihood inference is straightforward for these truncated models as the simulator is smooth in the parameters. The Bayesian solution was spelt out in the introduction of this paper and no new issues arise. The generic algorithm takes on the form

1. Initialize θ, s,
2. Sample $s_t|s_{\backslash t}, y_t; \theta$, $t = 1, ..., n$,
3. Sample $\alpha|s, \theta$,
4. Sample $\theta|\alpha, s$,
5. Goto 2.

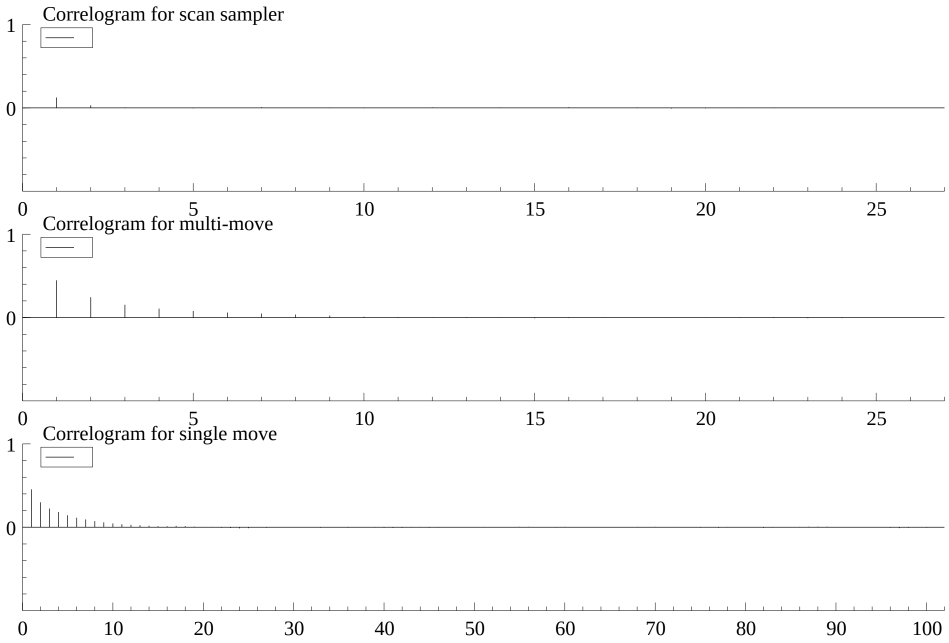

Figure 1: *Correlograms of* $y_4|Y_{50}$. *Indicates rate of convergence. From top to bottom: scan sampler, multi-move sampler and single-move sampler*

3.3.2 Quasi-maximum likelihood estimation

Although the Bayesian approach is computationally attractive, it requires the specification of a prior distribution, which may be difficult to write (and inferences may be sensitive to) in some cases. This makes the simulated EM algorithm attractive for this problem.

For censored time series models the simulated EM algorithm takes on a simple form. The log-likelihood for any latent s can be computed as $\log f(\theta; s)$ can be evaluated by the Kalman filter. Therefore, the function to be maximised, possibly numerically, in the M-step, is

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

The solution to this maximization problem will give θ^{i+1}. The process will be iterated until convergence. We further notice that $\widetilde{Q}(\theta^i, \theta^i; s)$ is an unbiased estimator of the actual score for the implicitly defined likelihood model for y.

It is important to point out that the same underlying random variates are used to simulate from the conditional distribution $s|y, \theta^i$ throughout the iterative process, and so, the sampler is smooth, i.e., is continuous on the parameters of the model θ, and hence, standard optimisation methods can be used to compute the resulting estimator.

3.4 A Monte Carlo study

In this subsection we will illustrate likelihood inference on the structure

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

with $|\phi| < 1$, ε_t and η_t mutually uncorrelated, and $\alpha_1|S_0 \sim N\left\{0, \sigma_\eta^2/\left(1-\phi^2\right)\right\}$.

Example 1 *We take $x_t = 1$ for all t in the measurement equation, the states follow an autoregressive process of order 1, and the variances of both the measurement and transition disturbances are time invariant. The artificial set of data is generated according to (2) with $\alpha_0 = 0$ for different values of the parameters values.*

Our approach is to perform both Bayesian and classical inference on the parameters of the model, which in this case are given by $\theta = (\beta, \sigma_\varepsilon^2, \sigma_\eta^2, \phi)'$.

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 distributions which we use. In particular, we use a flat Gaussian distribution for β, take $\chi_p^{-2}S_0$ for $\sigma_\varepsilon^2|y^*, \alpha, \beta$, and $\chi_q^{-2}S_1$ for $\sigma_\eta^2|\alpha, \phi$. Throughout we take $q = p = 5$ and $S_0 = p \times 1$ and $S_1 = q \times 1$.

We use $2Be(\phi_1, \phi_2) - 1$ as a prior family for ϕ, following Shephard and Pitt (1997) and Kim, Shephard, and Chib (1996). This implies $E(\phi) = \{2\phi_1/(\phi_1 + \phi_2)\} - 1$. We take $\phi_1 = 10$, $\phi_2 = 2$ so that ϕ has a prior mean of 0.66 and a standard deviation of 0.207. It could be argued that our prior should be closer to a unit root for high persistence cases. To sample ϕ we adopt the procedure used in Shephard and Pitt (1997) and Kim, Shephard, and Chib (1996), 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, as the expected proportion of censored observations in a sample is given by $n \times \Phi\left(-\beta/\sqrt{\sigma_\varepsilon^2 + \sigma_\eta^2/\left(1-\phi^2\right)}\right)$, where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal.

Figure 2 and Table 1[2] present the results of the multi-move Gibbs sampler for a censored model given in (2) with $\beta = 0.25$, $\sigma_\varepsilon^2 = 0.5$, $\sigma_\eta^2 = 0.1$, and $\phi = 0.95$ using a sample of size 200.

[2] The summary statistics of Table 1 report the inefficiency factors of the sampler. These are estimated as the variance of the sample mean from the MCMC sampling scheme relative to a hypothetical sampler which draws independent random variables from the posterior. This ratio is estimated using a Parzen window (see, for

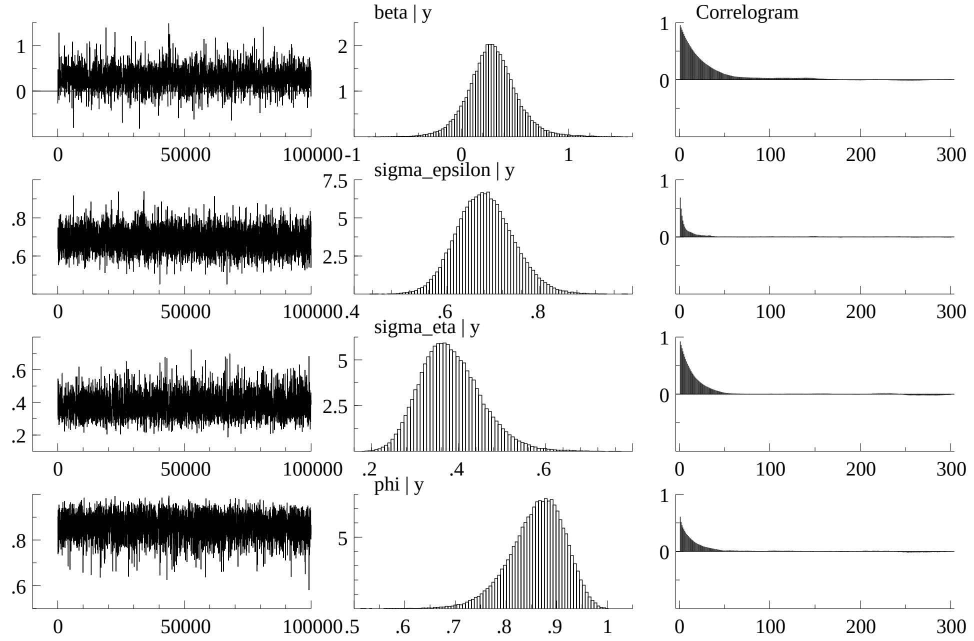

Figure 2: *Multi-move Gibbs sampler for a censored model with* $\beta = 0.25$, $\sigma_\varepsilon = 0.707$, $\sigma_\eta = 0.316$, $\phi = 0.95$. *Sample size* $n = 200$. *Left graphs: the simulation against iteration number. Middle graphs: histograms of the resulting marginal distributions and estimated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

The initial parameter values are $\beta = -0.2$, $\sigma_\varepsilon^2 = 0.4$, $\sigma_\eta^2 = 0.1$, 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 number of censored observations in the sample is 76.

The results suggest that the sampler is reasonably efficient, for inefficiency factors in the region of 30 suggest that the model can be quite precisely analyse in about 3000 iterations of the MCMC algorithm. An interesting feature of the result is that the regression coefficient β is not particularly well estimated, while ϕ is really poorly estimated.

example, Priestley (1981, Ch. 6)) with

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

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

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

	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.

Example 2 *Our second set of experiments deals with a censored model where the latent variable* s_t *is defined by equations (2) 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 iid.* $U(0,1)$ *random variables. As usual,* $\alpha_0 = 0$ *is imposed in the transition equation to generate the states.*

The true parameter values are $\beta = (0.25, 0.5)'$, $\sigma_\varepsilon^2 = 0.5$, $\sigma_\eta^2 = 0.1$, $\phi = 0.95$. Results are given only for the case where the sample size is $n = 200$.

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 number of censored observations in the sample is 62. The posterior means, covariance and correlation matrices, as well as the Monte Carlo standard errors, are presented in Table 2.

	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 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 ineffiency factor

The results suggest that the sampler does not change its behaviour very much with small changes in the model. Again this model could be reasonably analysed in 3,000 iterations, while the regression coefficients are again poorly estimated.

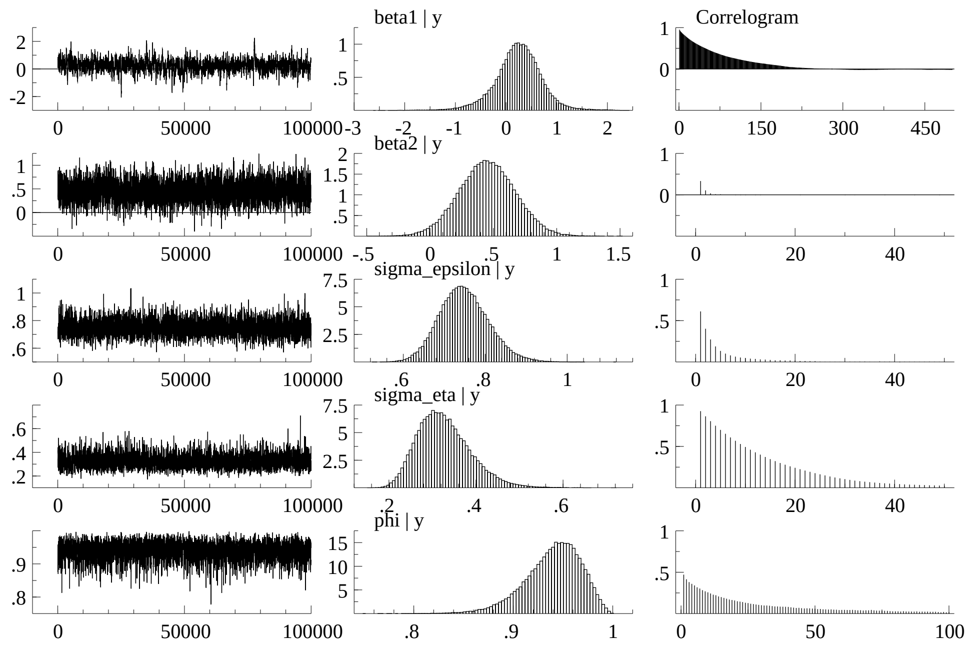

Figure 3: *Multi-move Gibbs sampler for a censored model with* $\beta = (0.25, 0.5)'$, $\sigma_\varepsilon^2 = 0.5$, $\sigma_\eta^2 = 0.1$, $\phi = 0.95$. *Sample size* $n = 200$. *Top graphs: the simulation against iteration number. Middle graphs: histograms of the resulting marginal distributions and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

3.4.2 Classical inference

To compute maximum likelihood estimators of the parameters we use the scan sampler inside a simulated EM algorithm.

Example 3 *(Special case of Example 1) 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 reparameterization $\phi = \theta_3/(1+\theta_3)$. The parameter β is unconstrained.

For each data set y_t, the initial value of θ is set to be $\theta^{(0)} = (0, 1, 1, 0)'$. We check that the results of the experiment are independent of these initial values.

The M-step of the simulated EM algorithm

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

is maximised using the BFGS method. 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 maximization routine is 250. When convergence fails after 250 iterations, the estimates are reported as equal to zero. However, this case is very rarely. Once we get the updated value of θ, $\theta^{(i)}$ the scan sampler is applied to get $s \sim s|y;\theta$. Iteration of this algorithm until convergence makes up the procedure.

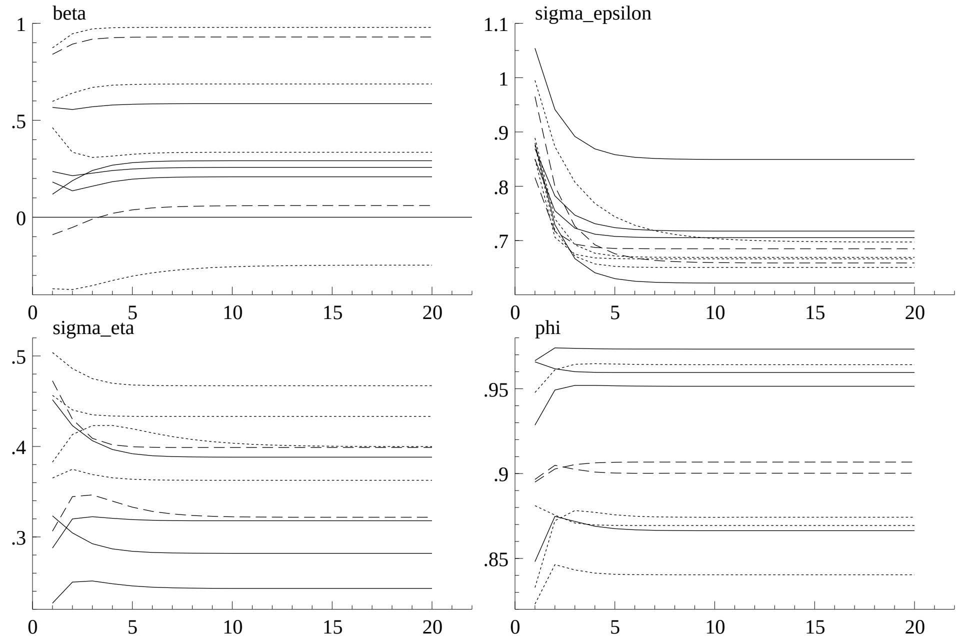

Figure 4: *EM estimates against iteration number for* 10 *samples of size* 200 *taking* $R = 5$ *scan draws. Illustrates the rapid convergence of the simulated EM algorithm*

Results for $R = 1$, $R = 5$ and $R = 10$ are presented in Table 3. We use 250 replications. We find that in this simple case there is little gain in taking R much bigger than 10. For each data set, we consider 20 iterations of the algorithm. However, as Figure 4 shows, in most cases a fairly small number is enough in the sense that after say i steps, $\theta^{(i+1)} = \theta^{(i)}$. This suggests, by using the scan sampler we have setup the EM algorithm 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 25 iterations of the algorithm, except for $n = 100$ and $\phi = 0.95$, when 50 iterations were required to get convergence for 10 samples. The simulated EM algorithm very 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.

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.

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 EM algorithm. Uses $R = 5$ scan draws. Converge denotes the percentage of optimisations which converged

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 $\theta^{(0)} = (0, 0, 1, 1, 0)'$. The sampler is performed as in Example 1 but in this case the BFGS method to maximised the log-likelihood for the 'augmented' data uses 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.

	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. Sample size $n = 200$. Uses $R = 10$ scan draws.

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 generalization to cover that problem is technically trivial.

The scan sampler can be used on this problem to deal with the time series aspect of the model leaving the only unresolved difficulty being

$$\Pr(s_t|s_{\backslash t}, y) \sim \Pr(s_t|s_{\backslash t}) \Pr(y_t|s_t),$$

which has a complicated support. Sampling from Gaussian static disequilibrium models involves simply sampling which of s_{1t} and s_{2t} is exactly equal to y_t and then sampling the other according to a truncated normal distribution. In our problem we have

$$\begin{aligned}\lambda_t &= \Pr\left[s_{1t} < s_{2t} \middle| s_{\backslash t}, \min(s_{1t}, s_{2t}) = y_t\right] \\ &= \frac{f_{s_{1t}|s_{\backslash t}}(y_{1t}) \times \Pr\left(s_{2t} > s_{1t}|s_{\backslash t}, s_{1t} = y_{1t}\right)}{f_{s_{1t}|s_{\backslash t}}(y_t) \times \Pr\left(s_{2t} > s_{1t}|s_{\backslash t}, s_{1t} = y_t\right) + f_{s_{2t}|s_{\backslash t}}(y_t) \times \Pr\left(s_{1t} > s_{2t}|s_{\backslash t}, s_{2t} = y_t\right)}\end{aligned}$$

and then conditional on $s_{1t} < s_{2t}$

$$s_{2t}|s_{\backslash t}, s_{1t} = y_t \sim TN_{s_{2t} > s_{1t}}(\mu_t, \Sigma_t).$$

An important point is that this simulator will not be continuous in the parameters, as λ_t depends on the parameters under estimation.

4.2 Previous work

Due to computational complexity and to the lack of computationally tractable statistical methods, disequilibrium models have mostly been specified and estimated without dynamic structures (with the exceptions of Laroque and Salanie (1993), Lee (1995a) and Hajivassiliou and McFadden (1996)). 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 ML parameter estimates which occur on boundaries, while there are often multiple maximums in the likelihood function.

Simulation methods can be devised to avoid the need for high dimensional integration and so 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. Since maximum likelihood estimation is asymptotically efficient, simulated scores should be used for the class of models for which it yields fast simulators that are smooth with respect to the parameters. However, this is not possible here. For dynamic disequilibrium models, simulated pseudo maximum likelihood methods have been suggested in Laroque and Salanie (1993), and some simulated likelihood estimators based on sequential evaluations of the likelihood as proposed by Hendry and Richard (1992) have been considered in Lee (1995b).

4.3 Illustration

To illustrate these points, we will apply the sampler to the following disequilibrium model:

$$\begin{aligned} s_{1t} &= a_1 x_{1t} + \phi_1 s_{1,t-1} + \sigma_1 \varepsilon_{1t}, \\ s_{2t} &= a_2 x_{2t} + \sigma_2 \varepsilon_{2t}, \\ y_t &= \min\left(s_{1t}, s_{2t}\right), \end{aligned} \tag{3}$$

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, and $|\phi_1| < 1$ to ensure stationarity. Finally, $x_t = (x_{1t}, x_{2t})'$ is strongly exogenous, which will be taken to be

$$\begin{aligned} x_{1t} &= 2.5\left(1 + \nu_t\right), \\ x_{2t} &= 5 \end{aligned} \tag{4}$$

with $\nu_t \sim N(0,1)$ and uncorrelated with $(\varepsilon_{1t}, \varepsilon_{2t})$. This is precisely the model considered by Laroque and Salanie (1993). We will interpret s_{1t} as latent demand. Demand depends on its lagged value. The true parameters are set to $a_1 = 1$, $a_2 = 1$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$ and $\phi_1 = 0.5$. This set-up ensures that $E(s_{1t}) = E(s_{2t})$, so that there is a roughly equal mix of regimes. Finally,

s_{10} was drawn from the stationary distribution for s_{1t}, i.e.,[3]

$$s_{10} \sim N\left(\frac{2.5a_1}{1-\phi_1}, \frac{\sigma_1^2}{1-\phi_1^2}\right). \tag{5}$$

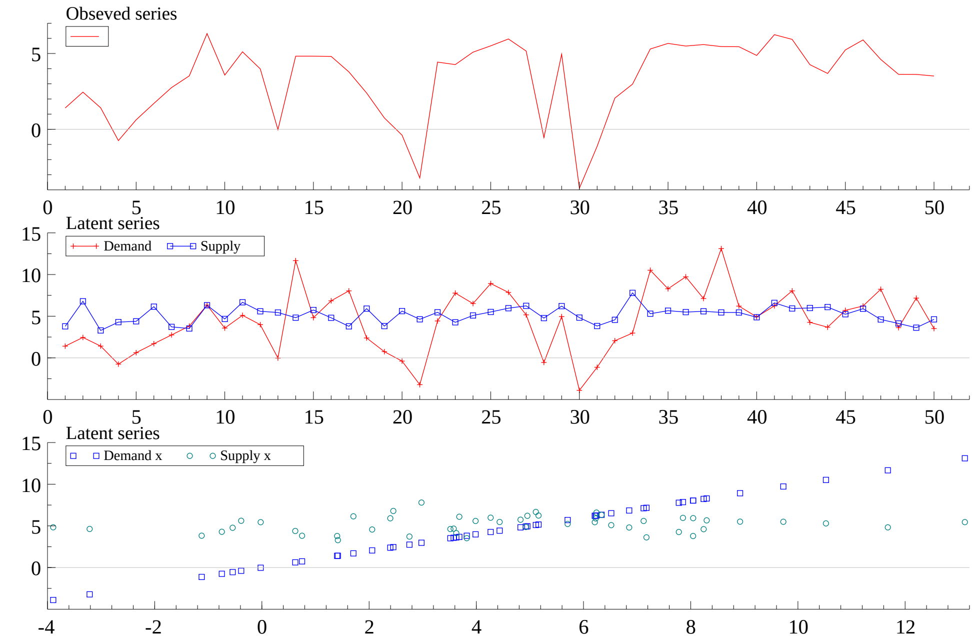

Figure 5: *Observed and latent series*

Figure 5 gives a plot of the simulated series with $n = 50$, which will be used in our illustration. For this series the minimum was equal to supply 25 times. The top graph plots the observed series, $y_t = \min(s_{1t}, s_{2t})$. The middle graph plots the unobserved latent series s_{1t} (demand) and s_{2t} (supply). The bottom graph draws demand and supply against the demand variable. The points in the bisecting line correspond to the 50 values of the demand variable, although they are not time ordered. At any time t, s_{1t} is the point on the bisecting line and s_{2t} is the point on the vertical line through s_{1t}. By looking at this graph, we see that the observed variable is equal to demand for observations on the left side of the graph, where $s_{2t} > s_{1t}$, and is equal to supply for observation on the right side, where $s_{2t} < s_{1t}$.

The sampler[4] is iterated $I = 100$ times in order to estimate the regime probabilities and the excess supply series. Results are reported in Figure 6. The top left graph plots the true values of the regime probabilities,

$$\begin{array}{ll} 1 & \text{if } y_t = s_{2t}, \\ 0 & \text{if } y_t = s_{1t}. \end{array}$$

[3]In Laroque and Salanie's paper, s_{10} is generated as $(1-\phi_1)^{-1}[a_1 x_{10} + \sigma_1 \varepsilon_{10}]$.

[4]Initial conditions for the Kalman filter are set to be the unconditional mean and unconditional variance, respectively although it would equally have been possible to use a diffuse initial condition here.

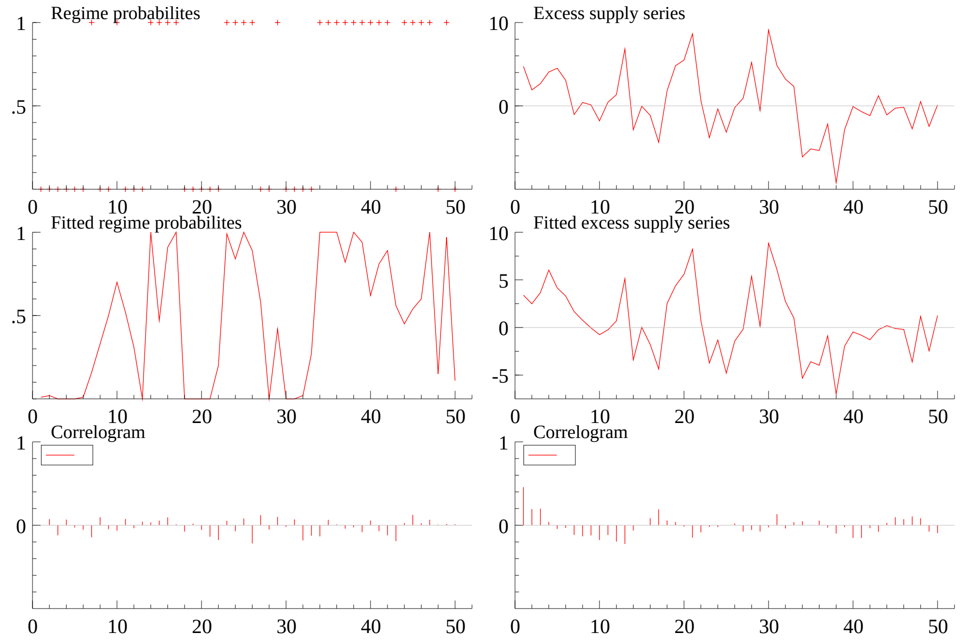

Figure 6: *True and fitted regime probabilities and excess supply series*

The simulated regime probabilities are plotted in the bottom left graph. They are obtained as the empirical probabilities, i.e., $\Pr(s_{2t} > s_{1t}|y)$ is estimated as $\Pr(\widehat{s_{2t} > s_{1t}}|y) = n_t/I$, where n_t denotes the number of iterations where $s_{2t} > s_{1t}$. The top right graph plots the true excess supply values $(s_{2t} - s_{1t})$. The bottom right graph plots the fitted excess supply values which are obtained as

$$\frac{1}{I}\sum_{i=1}^{I}\left\{s_{2t}^{(i)} - s_{1t}^{(i)}\right\}$$

where $s_{1t}^{(i)}$,$s_{2t}^{(i)}$ denotes the value of s_{1t}, s_{2t} at the i-th iteration. As we have said before, either $s_{1t}^{(i)}$ or $s_{2t}^{(i)}$ must be equal to the observed y_t.

In most cases, the estimated probabilities are meaningful in the sense that they are quite close to unity when the true probability was 1 and close to zero when the true probability was 0. However, the seventh observation is clearly misclassified $\Pr(\widehat{s_{27} > s_{17}}|y) = 0.16$, and in 18 per cent of the observations, the simulated probabilities belong to $[0.4, 0.6]$ which makes it difficult to distinguish between demand and supply observations. With respect to the fitted excess supply series, we see that it picks out reasonably well the behaviour of the actual series, particularly the peaks and lows after the first 12 observations. These results are quite encouraging as they will allow us to forecast some economic indicators, although we need to bear in mind that we have used the true parameter values.

The same experiment was carried out with different number of iterations. The results do not vary substantially, the simulated probabilities and excess supply series differ from the ones reported here by less than 0.05 and 0.25 respectively. Table 6 presents the percentages of observations correctly classified for different number of iterations ($I = 100,\ 200,\ 500$). Approximately, 85% of the demand observations are classified on the demand curve, 80% of the supply observations are classified on the supply curve, and 83% of the total observations are correctly classified.

100 scan iterations	
Demand	85.64
Supply	80.08
Total	82.86
200 scan iterations	
Demand	85.80
Supply	80.64
Total	83.22
500 scan iterations	
Demand	85.83
Supply	80.46
Total	83.14

Table 6: Percentages of observations correctly classified

We will return to this example at the end of the next subsection.

4.4 Parameter estimation

The first paper to use simulation techniques for the estimation of dynamic disequilibrium models is Laroque and Salanie (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 maximizing the resulting pseudo likelihood function to estimate the parameters.[5] It requires the exogenous variables to be strongly exogenously for the parameters of interest.

An alternative technique has been proposed by Lee (1995b) who considers the use of simulated likelihood methods based on factorizations of the sequential joint density of the observables and latent dependent variables, as suggested by Hendry and Richard (1992). We will not report this method here, although it has some advantages in this context.

We start this section by revising the Laroque and Salanie (1993) approach. Then, we suggest

[5]This technique is therefore related to the method of simulated moments developed by McFadden (1989) and Pakes and Pollard (1989).

how to perform Bayesian inference using the sampler introduced above. The next subsection will compare the two approaches in a Monte Carlo experiment.

Throughout this section, for any variable y, y^t will denote $(y_1, ..., y_t)$.

4.4.1 Dynamic Simulated Pseudo Maximum Likelihood Method

Laroque and Salanie (1993) propose a dynamic simulated pseudo maximum likelihood method to deal with a very general class of dynamic non linear models, which encompasses LDV models with lagged endogenous variables, both observed and latent, with or without serial correlation. This paper extends some earlier work for static disequilibrium models (Laroque and Salanie (1989)).

In short, dynamic pseudo-maximum likelihood (PML) methods consist in, 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 iid with known distribution. The parameter $\theta_0 = (b_0, R_0)$ fully describes the data generation process.

Laroque and Salanie (1993) focus on the so-called dynamic PML2 estimator which minimizes

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

over θ, where (g_t) is a sequence of functions which represent 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,[6]

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

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,

$$\begin{aligned} F_t(x^t, \theta) &= E\left\{ G_t(x^t, u^t, \theta) | x^t \right\} \\ V_t(x^t, \theta) &= Var\left\{ G_t(x^t, u^t, \theta) | x^t \right\}. \end{aligned}$$

Notice 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.

[6] The function g_t takes as arguments the path of exogenous variables x^t and of innovations u^t as well as the initial values (y_0, y^*_0, η_0).

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 minimization of the criterion function. For each $h = 1, ..., H$, we then compute recursively a dynamic simulation path for the endogenous variables y_t^h. The dynamic SPML2 estimator minimizes

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

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

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 Salanie (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 Salanie (1993) test the dynamic simulated pseudo maximum likelihood method on Monte Carlo generated data for the disequilibrium model defined by equations (3). They report estimation results from 200 randomly generated 50-observations samples, where the first two moments are approximated using 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).

4.4.2 Bayesian estimation

The unknown parameters in the model, say θ, can be estimated with Gibbs sampling by cycling over the steps

$$\theta \sim \theta|s, y \quad \text{and} \quad s \sim s|y, \theta.$$

We use the scan sampler to draw from $y|s, \theta$. After assuming a known prior for θ, the Gibbs sampling algorithm (with data augmentation) proceeds as follows:

- 1. Initialize θ.
- 2. For $t = n, ..., 1$ draw $s_t|s_{\backslash t}, y_t; \theta$ using the scan sampler described in Section 2.
- 3. Draw θ from $\theta|s$.
- 4. Goto 2.

4.5 A Monte Carlo study

4.5.1 Monte Carlo design

For sake of simplicity, we focus on the simple disequilibrium model given in (3), (4) and (5), where $x_t = (x_{1t}, x_{2t})'$ is strongly exogenous. In this framework $\theta = (a_1, a_2, \phi_1, \sigma_1^2, \sigma_2^2)'$. In the Bayes estimators the updating of the parameters $\theta|y, s$ follows exactly the same pattern as given in the section on Tobit models.

Our experiment deals with the model defined by equations (3), which is precisely the model analysed by Laroque and Salanie (1993). The true parameter values are $a_1 = 1$, $a_2 = 1$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$ and $\phi_1 = 0.5$. The number of periods n is equal to 50 while we will employ 500 replications throughout. The aim of this section will be to replicate the results of Laroque and Salanie (1993) and then compare them, in a simple Monte Carlo experiment, to the Bayes 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, a_2, σ_1 and σ_2, and from the uniform distribution on $[0.25, 0.75]$ for ϕ_1.

4.5.2 Laroque and Salanie (1993) method

For small samples $l^{HT}(\theta)$ is not particularly well behaved. Laroque and Salanie (1993) identify three possible source 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 not infrequently in the Monte Carlo experiment.

To try to be as kind as we could possibly be to the Laroque and Salanie (1993) procedure we tried to eliminate these problems as much as we could[7]. 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 the percentage of non-converging simulation experiments to be much lower than that reported in Laroque and Salanie (1993).

Using 500 replications we get the following results reported in Figure 7 and Table 7. 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.

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 7: Summaries of Figure 7. 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 calculates the proportion of replications for which the procedure converged.

4.5.3 Bayes 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 fixing both demand and supply equal to their observed minimum value.

After i iterations of the Gibbs sampler, a sequence of parameter values θ^1, θ^2, ...,θ^i is generated using each of the conditional posterior distributions listed above.

[7] Our Ox code for this experiment is available upon request from Aurora Manrique.

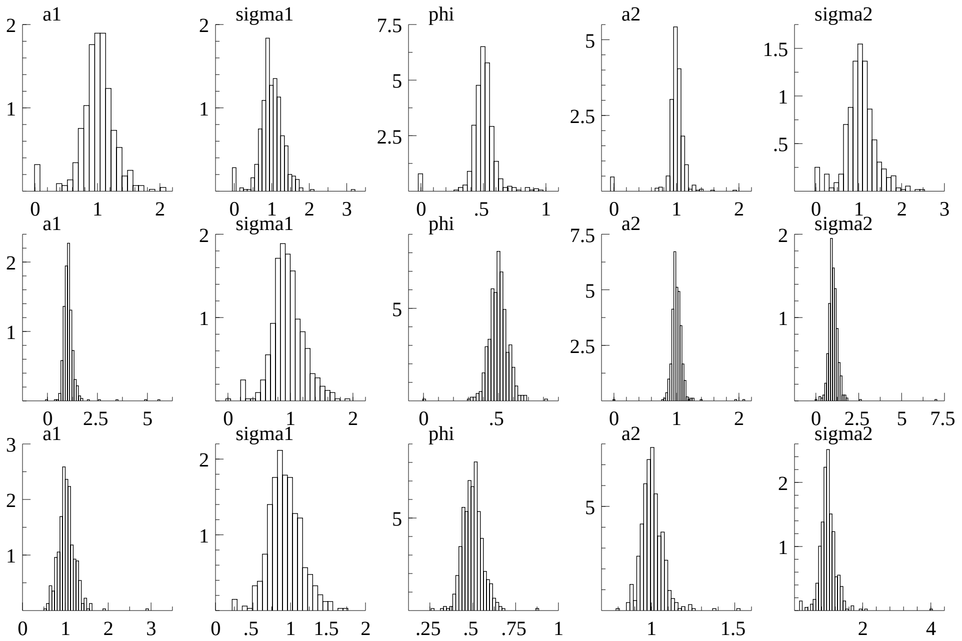

Figure 7: *Histograms of the 500 replications of the Laroque and Salanie's estimator of the disequilibrium model using* $n = 50$. *Top graphs have* $H = 10$, *middle* $H = 20$ *and bottom* $H = 50$. *The true parameters are 1.0, 1.0, 0.5, 1.0 and 1.0 as we go from the left to the right.*

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 8 and Table 8 give the results. We iterated the sampler on the demand and supply series for 500 iterations and then on the parameters and demand and supply series for $1,000$ more iterations before recording and answers. The next $10,000$ iterations are recorded. 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.

When analysing the results of the experiment for the disequilibrium model, we see that most of the estimates of a_2 is bigger than unity, while most of the estimates of ϕ_1, σ_2 are lower than the true values. The histogram for σ_1 is centered around 1. The correlograms die out very quickly. The sampler yields quite big posterior variances.

Once the parameters in both the demand and supply equations have been estimated, one may like to estimate the probability of each observation being on the demand or on the supply curve. We perform 100 iterations of the scan sampler to simulate both the regime probabilities $\Pr(s_{2t} > s_{1t}|y)$ and the excess supply series $(s_{2t} - s_{1t})$ setting the parameters equal to their estimates after $5,000$ and $10,000$ iterations of the sampler. Results are presented in Figure

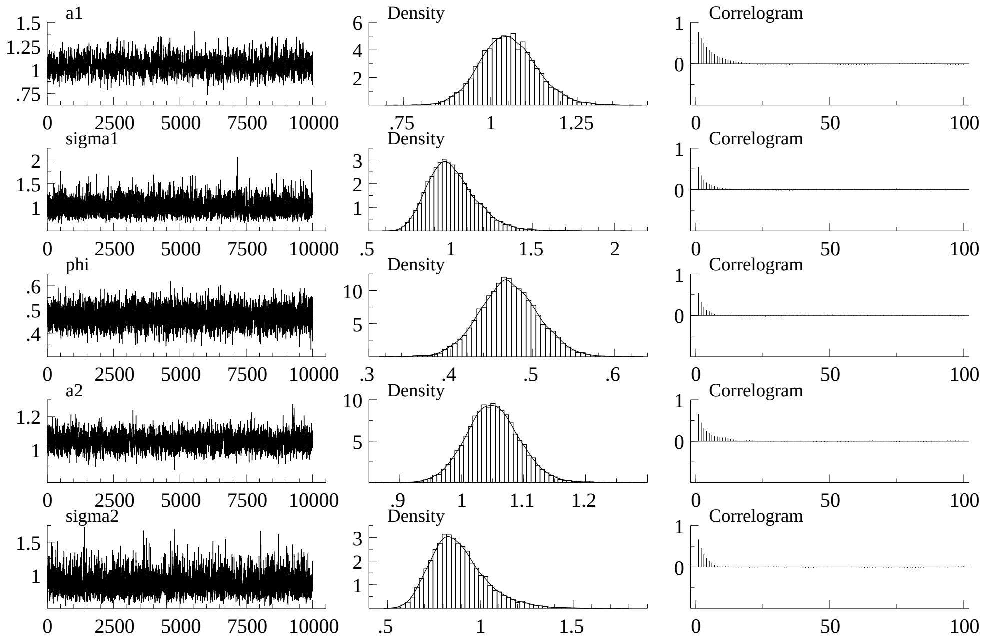

Figure 8: *Gibbs sampler for a dynamic disequilibrium model with* $a_1 = 1$, $\sigma_1 = 1$, $\phi_1 = 0.5$, $a_2 = 1$ *and* $\sigma_2 = 1$. *Left graphs: the simulation against iteration number. Middle graphs: histograms of the resulting marginal distributions and estimated densities. Right graphs: the corresponding correlograms for the iterations*

9. The actual series is the same as used in Figure 6. Using the parameter estimates after $5,000$ iterations the graphs of simulated values very much replicate the ones obtained with the true parameter values. Using the parameter estimates after $10,000$ iterations, nothing changes substantially. Despite the flat part of the bottom left graph at $t = 9$, 10, 11 compared to the small peak in the middle left graph, the simulated excess supply series follow the same pattern for all the simulated probabilities at those points in time were quite close to 0.5. As before, the simulated excess supply series fits the actual series quite well.

Sampling behaviour of Bayes estimators Now we repeat the experiment 500 times to see the sampling behaviour of the parameter estimates. We use exactly the same data as used in the previous section on the Laroque and Salanie (1993) 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.

Table 9 and Figures 10 give the results from the experiment. The results are exactly comparable with the earlier Table 7 reported for the Laroque and Salanie estimator. The results indicate that the Bayes estimator is much more precise. A general figure seems to be about 2

	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 8: Summaries of Figure 8. 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

or 3 times as efficient. Further, the Bayes estimator was computed without any problems of convergence.

	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 9: Summaries of Figure 10. Bayes estimator using 1,000 iterations. Throughout we use $n = 50$ and perform 500 replications. Figures in brackets are the estimated standard errors of the method, computed using the replications.

5 PANEL DATA MODELS

In principle the time series models we have studied in this paper can be extended to panel data contexts where we analyse many conditionally independent time series. The structure of the model becomes $y_{it} = h(s_{it})$, where

$$\begin{aligned} s_{it} &= c_{it} + Z_{it}\alpha_{it} + G_{it}u_{it}, \quad t = 1, \cdots, n, i = 1, ..., N, \\ \alpha_{it+1} &= d_{it} + T_{it}\alpha_{it} + H_{it}u_{it}, \quad u_{it} \sim NID(0, \sigma_i^2 I), \\ \alpha_{i0} &= 0, \end{aligned}$$

where $c_{it}, Z_{it}, G_{it}, H_{it}, d_{it}, T_{it}, \sigma^2$ and H_{it} will be assumed to be deterministic functions of the small dimensional panel random variable θ_i. Then $s_i = (s_{i1}, ..., s_{in})|\theta_i$ will be assumed to be independent time series. We will learn about the individual series effects θ_i by making

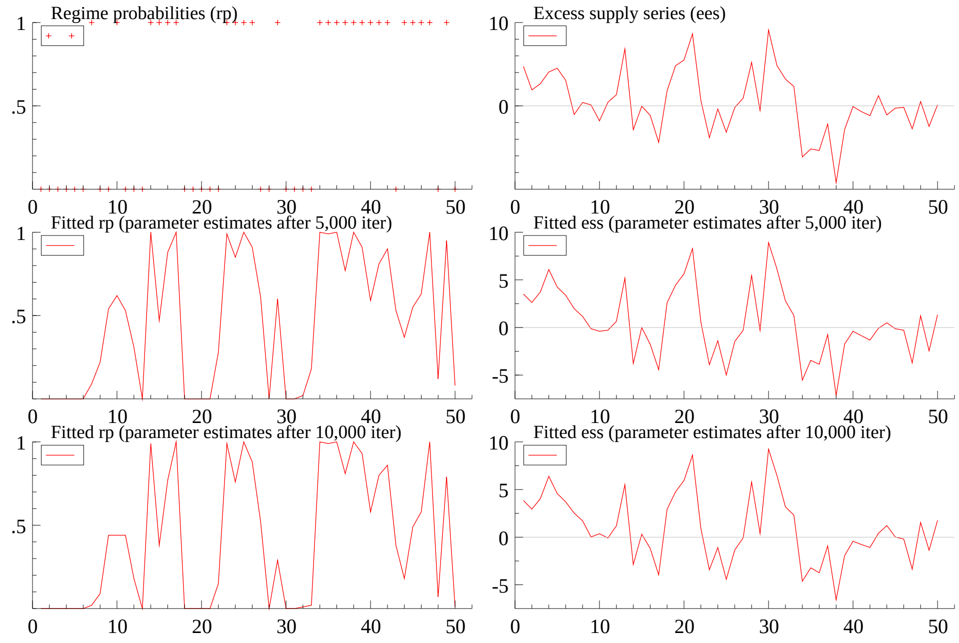

Figure 9: *Regime probabilities and excess supply series. Top graphs: true values. Simulated series with parameter values equal to their estimates after* $5,000$ *iterations in middle graphs and after* $10,000$ *iterations in bottom graphs*

an assumption that the θ_i are independent draws from some density $f(\theta_i|\theta)$. We call θ the parameters of the model.

Given θ_i the model raises no new issues for we can simulate from $s_i|y_i;\theta_i$ using the methods we developed in previous sections. Likewise we can simulate from $\theta_i|s_i$ or possibly from $\theta_i|\alpha_i, s_i$ if this is particularly convenient by simply taking into account the prior density $f(\theta_i|\theta)$.

6 CONCLUSION

This paper has set out a likelihood analysis of some Gaussian limited dependent processes. The analysis is rather simple, based around a result called the scan sampler. This is completely natural for these types of models.

Although the Gaussian assumption looks restrictive, it can in fact be generalized by allowing the introduction of mixing type effects which allow heavy tailed models and level shifts. This follows in spirit the work of, for example, Carter and Kohn (1994) and Shephard (1994).

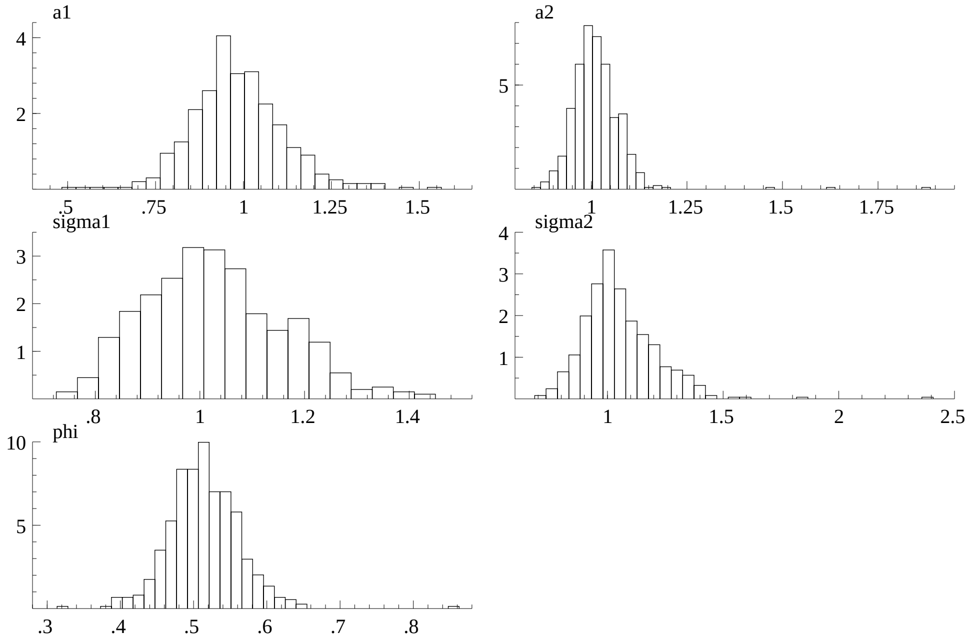

Figure 10: *Histogram of Bayes estimator using 1,000 iterations. Throughout we use* $n = 50$ *and perform 500 replications. Figures in brackets are the estimated standard errors of the method, computed using the replications.*

7 COMPUTATIONAL APPENDIX

This section details the algorithms given in this paper for general Gaussian state space form for the time series s_t :

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

Typically $c_t, Z_t, G_t, H_t, d_t, T_t$ 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.

7.1 Kalman filter

The Kalman filter has the following form

Kalman filter (see, for example, de Jong (1989)).

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

2. Compute

$$\begin{aligned} v_t &= s_t - Z_t a_{t|t-1}, & F_t &= Z_t P_{t|t-1} Z_t' + G_t G_t', \\ & & K_t &= T_t P_{t|t-1} Z_t' F_t^{-1}, \\ a_{t+1|t} &= d_t + T_t a_{t|t-1} + K_t v_t, & P_{t+1|t} &= T_t P_{t|t-1} T_t' - K_t F_t K_t' + H_t H_t'. \end{aligned} \tag{6}$$

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

Here K_t is called the Kalman gain, while v_t and $\sigma^2 F_t$ are the one-step ahead prediction error (or innovation) and its mean square error, respectively. The scaled innovations $F_t^{-\frac{1}{2}} v_t$ (or generalised least squares residuals) are NID with zero mean and constant covariance matrix in a correctly specified model.

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

$$\log f(s_1, ..., s_n; \varphi) = \sum_{t=1}^{n} \log f(s_t | s_1, ..., s_{t-1}; \varphi) = -\frac{1}{2} \sum_{t=1}^{n} \log |F_t| - \frac{1}{2} \sum_{t=1}^{n} v_t' F_t^{-1} v_t. \qquad (7)$$

7.2 Simulation state smoother

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

$$z_t = \begin{pmatrix} 1 & 0 \end{pmatrix}, G_t = \begin{pmatrix} \sigma_\varepsilon & 0 \end{pmatrix}, T_t = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix}, H_t = \begin{pmatrix} 0 & \sigma_\eta \\ 0 & 0 \end{pmatrix}, R_t = \begin{pmatrix} 0 & \sigma_\eta \end{pmatrix}.$$

Then the simulation state smoother draws from $\alpha | s$. It has the form:

Simulation state smoother (due de Jong and Shephard (1995)).

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 $L_t = T_t - K_t Z_t$

$$\begin{array}{rclrcl} C_t & = & R_t R_t' \left(I - N_t \right) R_t R_t', & \varepsilon_t & \sim & \mathsf{N}(0, C_t), \\ r_{t-1} & = & Z_t' D_t^{-1} v_t + L_t' r_t - V_t' C_t^{-1} \varepsilon_t, & N_{t-1} & = & Z_t' F_t^{-1} Z_t + L_t' N_t L_t + V_t' C_t^{-1} V_t, \end{array} \qquad (8)$$

where

$$V_t = R_t R_t' N_t L_t.$$

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

The resulting $\eta_1, ..., \eta_n$ can the be padded out to produce a draw from the $\widetilde{\eta_1} = H_1 u_1, ..., \widetilde{\eta_n} = H_n u_n | s$.

Once the disturbances of the states are drawn it is possible to reconstruct the state via the running of the forward recursion

$$a_{t+1|n} = d_t + T_t a_{t|n} + \widetilde{\eta}_t.$$

7.3 Simulation signal smoother

The simulation signal smoother (de Jong and Shephard (1995)) draws from

$$\left(c_1 + Z_1\gamma_1, ..., c_n + Z_n\gamma_n\right)|s.$$

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

$$\begin{array}{rcl} C_t &=& \Sigma_t - \Sigma_t D_t \Sigma_t, \qquad \kappa_t \sim \mathsf{N}(0, C_t), \\ r_{t-1} &=& Z_t' F_t^{-1} v_t + L_t' r_t - V_t' C_t^{-1} \kappa_t, \qquad V_t = \Sigma_t \left(D_t Z_t - K_t' N_t T_t\right), \\ N_{t-1} &=& Z_t' F_t^{-1} Z_t + L_t' N_t L_t + V_t' C_t^{-1} V_t. \end{array} \tag{9}$$

Then $s_t - \Sigma_t n_t - \kappa_t$ is a draw from the signal $c_t + Z_t\gamma_t|s$.

ACKNOWLEDGEMENTS

We would like to thank the ESRC for their financial support through the project 'Estimation via Simulation in Econometrics' and the EU through their grant 'Econometric inference using Simulation Techniques'. All the computations reported in this paper were carried out using the Ox language of Doornik (1996). Neil Shephard would like to thank Jurgen Doornik for computational advice.

References

Anderson, G. J. and D. F. Hendry (1984). An econometric model of United Kingdom building societies. *Oxford Bulletin of Economics and Statistics 46*, 185–210.

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

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. Am. Statist. Assoc. 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. Am. Statist. Assoc. 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.

de Jong, P. (1989). Smoothing and interpolation with the state space model. *J. Am. Statist. Assoc. 84*, 1085–8.

de Jong, P. (1996). The scan sampler. Unpublished paper: Statistics Department, LSE.

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

Diggle, P. J. and R. J. Gratton (1984). Monte Carlo methods of inference for implicit statistical models (with discussion). *J. R. Statist. Soc. B 46*, 193–227.

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

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.

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

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. 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 dicrete bid and ask quotes. Unpublished paper: Stern Business School, New York University.

Hendry, D. F. and J. F. Richard (1992). Likelihood evaluation for dynamic latent variables models. In D. A. B. Amman, H M and L. F. Pau (Eds.), *Computational Economics and Econometrics*, pp. 3–17. Amsterdam: Kluwer.

Kim, S., N. Shephard, and S. Chib (1996). Stochastic volatility: Optimal likelihood inference and comparison with ARCH models. Unpublished paper: Nuffield College, Oxford. Available at `http://www.nuff.ox.ac.uk`.

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

Laroque, G. and B. Salanie (1993). Simulation-based estimation of models with lagged latent variables. *J. Appl. Econometrics 8*, S119–S133. Reprinted as pp. 135–150 in van Dijk, H.K., A. Monfort and B.W. Brown(1995), *Econometric Inference Using Simulation Techniques*, Chichester: Wiley.

Lee, L. (1995a). Simulation estimation of dynamic switching regesion and dynamic disequilibrium models - some Monte Carlo results. Unpublished paper: Department of Economics, The Hong Kong University of Science and Technology.

Lee, L.-f. (1995b). Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory 11*, 437–483.

Louis, T. A. (1982). Finding observed information using the EM algorithm. *J. R. Statist. Soc. 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). Truncated time series. Unpublished, D.Phil. Thesis, Oxford University.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica 57*, 995–1026.

Pakes, A. and D. Pollard (1989). Simulation and asymptotics of optimised estimators. *Econometrica 57*, 1027–1057.

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. *Phil. Trans. R. Soc. Lond. A. 337*, 407–28.

Quandt, R. E. (1982). Econometric disequilibrium models. *Econometric Reviews 1*, 1–63.

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

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

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*. Forthcoming.

Tanner, M. A. (1996). *Tools for Statistical Inference: methods for exploration of posterior distributions and likelihood functions* (3 ed.). New York: Springer-Verlag.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica 26*, 24–36.

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. Am. Statist. Assoc. 85*, 699–704.