

Bayesian inference based only on simulated likelihood:
particle filter analysis of dynamic economic models*

THOMAS FLURY

*Oxford-Man Institute, University of Oxford,
Blue Boar Court, Alfred Road, Oxford OX1 4EH, UK*
& *Department of Economics, University of Oxford*
thomas.flury@economics.ox.ac.uk

NEIL SHEPHARD

*Oxford-Man Institute, University of Oxford,
Blue Boar Court, Alfred Road, Oxford OX1 4EH, UK*
& *Department of Economics, University of Oxford*
neil.shephard@economics.ox.ac.uk

December 1, 2008

Abstract

Suppose we wish to carry out likelihood based inference but we solely have an unbiased simulation based estimator of the likelihood. We note that unbiasedness is enough when the estimated likelihood is used inside a Metropolis-Hastings algorithm. This result has recently been introduced in statistics literature by Andrieu, Doucet, and Holenstein (2007) and is perhaps surprising given the celebrated results on maximum simulated likelihood estimation.

Bayesian inference based on simulated likelihood can be widely applied in microeconomics, macroeconomics and financial econometrics. One way of generating unbiased estimates of the likelihood is by the use of a particle filter. We illustrate these methods on four problems in econometrics, producing rather generic methods. Taken together, these methods imply that if we can simulate from an economic model we can carry out likelihood based inference using its simulations.

Keywords: Dynamic stochastic general equilibrium models, inference, likelihood, MCMC, Metropolis-Hastings, particle filter, state space models, stochastic volatility

J.E.L. codes: C11, C13, C15, C32, E32.

*We thank Arnaud Doucet for various stimulating conversations on this topic and Martin Ellison for his advice.

1 Introduction

Simulation likelihoods are simulation based estimators of the likelihood. They are often quite straightforward to compute — even for rather complicated models. In this paper we show how to carry out Bayesian analysis using a simulated likelihood, often computed using a particle filter, for a number of well known static and dynamic economic models.

Andrieu, Doucet, and Holenstein (2007) and Andrieu, Doucet, and Roberts (2007) have recently shown the surprising result that when an unbiasedly estimated likelihood is used inside a Metropolis-Hastings algorithm then the estimation error makes no difference to the equilibrium distribution of the algorithm — allowing exact inference even when the likelihood is only estimated. Our paper will be centred around using their insights. The proof of the validity of this *expected auxiliary variable method* is simple and will be explained in Section 2. This simplicity heavily contrasts with the fragility of the maximum simulated likelihood (MSL) estimator of Lerman and Manski (1981) and Diggle and Gratton (1984), which needs the accuracy of the simulator to increase faster than the sample size to produce even consistency.

Our focus in this paper is on four models. To illustrate the simplicity of the approach we look at a standard cross-sectional Probit model. Our other examples come from time series problems, where we will employ a particle filter to compute the simulated likelihood. The first time series model we analyse is a standard linear, Gaussian model, where our estimates can be compared to that obtained using a Kalman filter (which provides the exact likelihood). The second is a simple stochastic volatility model. The third is a stochastic general equilibrium model. This last problem is particularly interesting as it will show the unique power of combining the particle filter with the Metropolis-Hastings algorithm. We believe this combination will have wide application in economics where standard MCMC algorithms are difficult to apply due to the forward optimising behaviour which is common in economic models.

1.1 MCMC and particle filters

The use of the Metropolis-Hastings algorithm to produce Markov chain Monte Carlo draws from the posterior distribution of parameters given the data, is now well established in econometrics. Chib (2001) provides an extensive discussion of the methods and the corresponding literature. A difficulty with the application of these methods in some dynamic models is the evaluation of the likelihood function, even including a variety of possible auxiliary variables. This has held back the application of these methods for some problems in macroeconomics and microeconomics, where agents have to perform sophisticated optimisation type calculations in order to update their decision making. These forward looking problems typically have no explicit likelihood, but can be simulated

from. As a result they are challenging for MCMC methods.

Instead of having to evaluate the likelihood exactly, this article advocates the use of a simulated likelihood, often carried out using a particle filter. Not all readers will be familiar with particle filters and so we will briefly discuss their background here. A particle filter is a sequential simulation device to carry out on-line filtering for non-Gaussian, non-linear state space models (e.g. Gordon, Salmund, and Smith (1993), Pitt and Shephard (1999) and Doucet, de Freitas, and Gordon (2001)). It can be thought of as a modern generalisation of the Kalman filter, for the Kalman filter is able to analyse only linear, Gaussian state space models. The particle filter has minimal requirements: the ability to simulate from the dynamics of the state and being able to evaluate the measurement density.

Both the particle and Kalman filters produce filtered estimates of unobserved states given past data and estimates of the one-step ahead density, which delivers the likelihood via the prediction decomposition. In the Kalman case all these quantities are exact, in the particle filter case they are simulation based estimates. Importantly the particle filter based estimate of the likelihood is non-continuous with respect to the parameters (see the smoothed version by Pitt (2001) in the univariate state case and Flury and Shephard (2008) in the multivariate case) — even if common random numbers are used.

Inside an MCMC algorithm the Kalman filter has been used to compute the likelihood function in many papers, in vastly more papers it has been used to carry out maximum likelihood estimation (e.g. Harvey (1989) and Durbin and Koopman (2001)). The particle filter has been used to estimate the likelihood by Kim, Shephard, and Chib (1998) for the purpose of model comparison. Pitt (2001) carries out MSL estimation based on a smoothed version of a particle filter, while Durbin and Koopman (2001) discuss extensively alternative importance samplers as the basis of the same estimation methods.

Particle filters are increasingly used in economics. Kim, Shephard, and Chib (1998) and Pitt and Shephard (1999) were the first to use particle filters in economics, where they extracted the filtered estimate of the current volatility from a stochastic volatility model. Later work on this, including developments to try to learn parameters on-line using particle filters, can be seen in Johannes, Polson, and Stroud (2008) and Johannes and Polson (2008).

An interesting application in microeconomics is Gallant, Hong, and Khwaja (2008), who look at a dynamic game theory model with serially correlated production costs.

In macroeconomics Fernandez-Villaverde, Rudio-Ramirez, and Santos (2006) and Fernandez-Villaverde and Rudio-Ramirez (2007) used particle filters as a basis for estimating the likelihoods of some stochastic general equilibrium models, which they then maximised with respect to the

parameters. This is an application of MSL estimation based on a particle filter. They also mentioned the possibility of placing the estimated likelihood inside a Metropolis-Hastings algorithm, but had no rationale for this.

Although our focus has been on combining particle filters with MCMC, other importance sampling estimators of the likelihood can also be combined with MCMC to produce valid samples from the correct posterior distribution. The work reviewed in Durbin and Koopman (2001) provides a MSL estimator of many statistical state space models. All of that work can be immediately ported over to the Bayesian environment, but without any distortions created by the simulation errors.

1.2 Maximum simulated likelihood estimation

Inference using maximum simulated likelihood estimation goes back at least to Lerman and Manski (1981) and Diggle and Gratton (1984). In the simplest applications of this technique simulation is used to unbiasedly estimate the likelihood using M independent and identically distributed (i.i.d.) draws. The estimate is then numerically maximised with respect to the parameters. Based on a sample of size T , for i.i.d. data we need that $T/M \rightarrow 0$ for this MSL estimator to be consistent and $T/\sqrt{M} \rightarrow 0$ to have the same distribution as the maximum likelihood estimator.

This is a deeply disappointing result, for it means in a data rich environment the simulation demands of the MSL method could be dramatic. Further, it is often difficult to know if T/\sqrt{M} is close to zero. These observations lead to the development of new simulation based inference strategies in the late 1980s and early 1990s. These included simulated scores, the stochastic EM algorithm, indirect inference and efficient method of moments. Discussions of these topics can be found in, for example, Hajivassiliou and McFadden (1998), Gouriéroux, Monfort, and Renault (1993), Smith (1993), Gallant and Tauchen (1996) and Gouriéroux and Monfort (1996).

1.3 Outline of the paper

This paper is structured as follows. In Section 2 we provide a self-contained discussion of why using an estimated likelihood makes no difference to the equilibrium distribution of a Metropolis-Hastings algorithm. We compare this to the usual MSL approach. In Section 3 we give an example of a static model. Section 4 treats dynamic models and introduces particle filters as a convenient tool to obtain unbiased likelihood estimates. In Sections 5-7 we provide some examples from macroeconomics and finance to demonstrate the performance of this algorithm in dynamic models. Section 8 gives some conclusions we draw from our paper.

2 Expected auxiliary variable method

2.1 An auxiliary variable interpretation of importance sampling

The basics of this argument can be expressed simply. Suppose we wish to carry out inference by sampling from

$$f(\theta|y) \propto f(y|\theta)f(\theta),$$

where θ are the parameters, y the data, $f(\theta)$ is a prior and $f(y|\theta)$ a likelihood. Here f denotes the density or probability function for continuous or discrete variables.

This typically requires us to calculate the likelihood $f(y|\theta)$, but here we assume that all we have is a simulation based estimator

$$\hat{f}_u(y|\theta),$$

which is unbiased

$$E_u \left\{ \hat{f}_u(y|\theta) \right\} = f(y|\theta),$$

where we average over the simulation denoted by the multivariate u . We assume this estimator is itself a density function in y . Then we can think of the simulation estimator \hat{f} as being based on an auxiliary variable:

$$\hat{f}_u(y|\theta) = g(y, u|\theta),$$

that is g is a joint density which, when marginalised over u , delivers $f(y|\theta)$.

2.2 Bayesian analysis

This simple insight has significant implications econometrically, because now we can carry out inference by sampling from

$$g(u, \theta|y) \propto g(y, u|\theta)f(\theta),$$

This simulation based Bayesian method will deliver draws

$$\left(u^{(1)}, \theta^{(1)}\right), \left(u^{(2)}, \theta^{(2)}\right), \dots, \left(u^{(N)}, \theta^{(N)}\right)$$

so throwing away the u samples leaves us with

$$\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)}$$

which are from $f(\theta|y)$. These samples can be used to approximate the entire posterior distribution to arbitrary levels of accuracy, where in turn the statistical accuracy of the simulation-based approximation can be measured. Further, it can be used to approximate the posterior median, which is an efficient estimator in the classical sense of θ by the Bernstein- von Mises Theorem.

The sampling can be carried out using generic Markov chain Monte Carlo (MCMC) algorithms. Sample $\theta^{(i)}$ from a proposal $dQ(\theta^{(i)} | \theta^{(i-1)})$, draw the uniformly distributed u and compute

$$\widehat{L}^{(i)} = \widehat{f}_u^{(i)}(y|\theta^{(i)})$$

The acceptance probability is given by

$$q = \min \left[\frac{\widehat{L}^{(i)} f(\theta^{(i)})}{\widehat{L}^{(i-1)} f(\theta^{(i-1)})} \frac{dQ(\theta^{(i-1)}|\theta^{(i)})}{dQ(\theta^{(i)}|\theta^{(i-1)})}, 1 \right].$$

Draw $V \sim U(0, 1)$ and if $V > q$ set

$$(\widehat{L}^{(i)}, \theta^{(i)}) = (\widehat{L}^{(i-1)}, \theta^{(i-1)}).$$

Under very weak conditions (e.g. Chib (2001)) the sequence $\{\theta^{(i)}\}$ for $i = 1, \dots, N$ converges to samples from $f(\theta|y)$ as $N \rightarrow \infty$.

An important practical observation is that the choice for the precision parameter M of the likelihood approximation does of course play a role in the algorithm: it can influence the rejection rate. If M is very small then \widehat{f} will be a very jittery estimator of f which will increase the chance the algorithm gets stuck. Hence increasing M will improve the mixing of the MCMC chain, so increasing the incremental information in each new $\theta^{(i)}$.

In this paper we suggest that for many economic models the issue of needing M to be large can be entirely sidestepped — while still insisting on efficiency.

We saw this style of argument first in the context of dynamic models in a paper in statistical theory by Andrieu, Doucet, and Holenstein (2007). The framework we use is vastly simpler and gets immediately to the heart of the issue. We also know of an unpublished note by Andrieu, Doucet, and Roberts (2007) who call this approach the Expected Auxiliary Variable method and we use that phrase here. We make our contribution by illustrating these themes on some core econometric problems.

2.3 Simple example

An example of this is the very simplest discrete choice model (e.g. Train (2003)) where y_t is binary

$$\Pr(y_t = 1|x_t, \beta, \psi) = p_t = \Pr(x_t' \beta + \varepsilon_t \geq 0), \quad \varepsilon_t|x_t \sim F_t(\psi),$$

where we assume $y_t|x_t$ are independent over t . Assume we can simulate from $F_t(\psi)$.¹ Write these simulations as $\varepsilon_t^{(1)}, \dots, \varepsilon_t^{(M)}$. Then the simplest simulation based estimator of p_t is

$$\widehat{p}_t = \frac{1}{M} \sum_{j=1}^M 1_{x_t' \beta + \varepsilon_t^{(j)} \geq 0},$$

delivering the simulated likelihood function

$$\widehat{f}(y|\beta, \psi) = \prod_{t=1}^T \widehat{p}_t^{y_t} (1 - \widehat{p}_t)^{1-y_t}.$$

It is easy to see that this is an unbiased estimator of the true likelihood

$$f(y|\beta, \psi) = \prod_{t=1}^T p_t^{y_t} (1 - p_t)^{1-y_t},$$

but of course the score is biased. It is this bias which drives the fact that the MSL estimator of $\theta = (\beta', \psi)'$ behaves poorly asymptotically unless $M \rightarrow \infty$.

2.4 Comparison to MSL estimation

The above algorithm is fundamentally different from common econometric practice, which is to compute \widehat{f} with common uniform random variables, smoothing out the simulation error of the estimator of the likelihood as θ varies. Here new uniforms are drawn at each iteration. This is crucial.

When one uses common random numbers \bar{u} , which are drawn once from a large dimensional uniform distribution and held fixed as we vary θ , then

$$\widehat{f}_{\bar{u}}(y|\theta) = g(y, \bar{u}|\theta),$$

such that we converge to the wrong posterior, which is proportional to

$$\widehat{f}_{\bar{u}}(y|\theta) f(\theta)$$

if we use this inside the Metropolis-Hastings algorithm. This posterior then needs M to increase quickly in order for it not to have an impact on the solution.

As mentioned before, the MSL estimator has the inconvenient property that for a sample of size T , for i.i.d. data we need that $T/M \rightarrow 0$ for it to be consistent and $T/\sqrt{M} \rightarrow 0$ to have the same distribution as the maximum likelihood estimator. In a data rich environment the simulation demands of the MSL method could be dramatic and it is often difficult to know if T/\sqrt{M} is close to zero.

¹It is interesting to note there is no requirement to be able to compute $F_t(\psi)$.

In the approach considered in this paper, our only concern for the choice of M is that the likelihood estimate is not too jittery. The choice of M is quite easy in practice and unless we have an extremely large number of observations, or a very spiky likelihood, a moderate M value is often found to be sufficient.

3 Example: binary choice models

3.1 The model and prior

To start off we illustrate the workings of the algorithm in a rather simple static setup: a binary choice model. We use the classical data set from Mroz (1987) to study the labour force participation of $T = 753$ women. We perform inference on the parameters of the binary choice model by using an unbiased estimate of the likelihood inside an MCMC algorithm. We will make sufficiently strong assumptions that in this case an analytic likelihood is available, which allows us to cleanly assess the effectiveness of the expected auxiliary variable method.

In this model we have a binary variable y_t which takes the value 1 if a woman works and 0 otherwise. We assume that $y_t = 0$ if $y_t^* \leq 0$ and $y_t = 1$ if $y_t^* > 0$, where

$$y_t^* = \beta_0 + \beta_1 \text{nwifinc}_t + \beta_2 \text{educ}_t + \beta_3 \text{exper}_t + \beta_4 \text{exper}_t^2 + \beta_5 \text{age}_t \\ + \beta_6 \text{kidslt6}_t + \beta_7 \text{kidsge6}_t + \varepsilon_t$$

The explanatory variables are non-wife income, education, experience, experience squared, age, number of children less than six years old, and number of children between the ages of 6 and 18. We write

$$p_t = \Pr(y_t = 1 | x_t, \beta, \psi) = \Pr(x_t' \beta + \varepsilon_t \geq 0) = \Pr(-\varepsilon_t \leq x_t' \beta) = F(x_t' \beta | \psi)$$

where $\beta = (\beta_0, \dots, \beta_7)'$. Here we choose the normal distribution for $F(x_t' \beta | \psi)$. We assume a Gaussian prior given by $\beta \sim N(\beta^0, I_8)$ where²

$$\beta^0 = (0.5855, -0.0034, 0.0380, 0.0395, -0.0006, -0.0161, -0.2618, 0.0130)'$$

Given the Gaussian errors we can easily compute the true likelihood and thus use the results from this exact likelihood model as a benchmark³.

²The prior mean β^0 was set as the least squares estimator of the corresponding linear probability model.

³Of course much Bayesian econometric work on discrete choice models has been carried out based on MCMC methods — see, for example, Chib and Greenberg (1998). These use different auxiliary variables, chosen to reflect the particular problem at hand, rather than generic simulated likelihood methods. It should be expected that these tailored methods will be more computationally efficient than the methods we are advocating here.

3.2 Simulated likelihood

For the simulation based estimator of p_t we draw

$$\varepsilon_t^{(j)} \stackrel{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2), \quad j = 1, \dots, M,$$

and compute the simulation based estimators of the probabilities as

$$\hat{p}_t = \frac{1}{M} \sum_{j=1}^M 1_{x'_t \beta + \varepsilon_t^{(j)} \geq 0}.$$

The estimate of the likelihood is given by

$$\hat{f}(y|\beta, \sigma_\varepsilon^2) = \prod_{t=1}^T \hat{p}_t^{y_t} (1 - \hat{p}_t)^{1-y_t}.$$

We then use this estimator inside an MCMC algorithm as discussed above in order to make inference on β .

We recall here that in the usual Probit model the variance has to be normalized, for it is impossible to estimate both β and σ_ε^2 as

$$\Pr(\varepsilon_t \leq x'_t \beta) = \Pr\left(\frac{\varepsilon_t}{\sigma_\varepsilon} \leq \frac{x'_t \beta}{\sigma_\varepsilon}\right)$$

In Probit models one usually sets $\sigma_\varepsilon = 1$. In our setup the choice of σ_ε can matter for the performance of the algorithm. If we fix σ_ε too small we can end up with a pair $(y_t = 1, \hat{p}_t = 0)$ which results in $\hat{f}(y|\beta, \psi) = 0$. We suggest setting $\sigma_\varepsilon = 1$ by default and if this causes problems to tune it such that we just avoid this undesirable outcome.

3.3 The sampling algorithm

Throughout we will use a Metropolis-Hastings algorithm to sample from the posterior distribution, using the estimated likelihood. Instead of making proposals to update all the parameters at once, we break the updating into a series of univariate proposals.

Throughout we have chosen to use random walk proposals for the parameters, each applied one at a time:

$$\begin{aligned} \beta_{0,i} &= \beta_{0,i-1} + 0.1326\nu_{1,i}, & \beta_{4,i} &= \beta_{4,i-1} + 0.0005\nu_{5,i}, \\ \beta_{1,i} &= \beta_{1,i-1} + 0.0058\nu_{2,i}, & \beta_{5,i} &= \beta_{5,i-1} + 0.0031\nu_{6,i}, \\ \beta_{2,i} &= \beta_{2,i-1} + 0.0109\nu_{3,i}, & \beta_{6,i} &= \beta_{6,i-1} + 0.2317\nu_{7,i}, \\ \beta_{3,i} &= \beta_{3,i-1} + 0.0108\nu_{4,i}, & \beta_{7,i} &= \beta_{7,i-1} + 0.0703\nu_{8,i}, \end{aligned}$$

where $\nu_{j,i} \stackrel{i.i.d.}{\sim} N(0, 1)$ for $j = 1, \dots, 8$ and $i = 1, \dots, N$. The variances in the random walk proposals were chosen to aim for a 40% acceptance probability for each parameter, see e.g. Gelman,

	exact likelihood				$M = 1,000$				$M = 2,000$				$M = 4,000$			
	mean	mcse	Pr	inef	mean	mcse	Pr	inef	mean	mcse	Pr	inef	mean	mcse	Pr	inef
β_0	0.295	.033	.42	364	0.313	.039	.28	593	0.349	.035	.33	462	0.376	.030	.36	420
β_1	-0.012	.000	.41	31	-0.012	.000	.28	57	-0.012	.000	.33	35	-0.012	.000	.36	42
β_2	0.130	.001	.41	143	0.130	.002	.27	328	0.130	.001	.33	209	0.126	.001	.36	168
β_3	0.124	.001	.41	149	0.125	.001	.27	149	0.123	.001	.32	160	0.124	.001	.36	197
β_4	-0.002	.000	.41	130	-0.002	.000	.28	135	-0.002	.000	.33	162	-0.002	.000	.36	189
β_5	-0.053	.001	.41	299	-0.054	.001	.28	479	-0.054	.001	.33	458	-0.054	.001	.37	321
β_6	-0.868	.004	.43	70	-0.871	.004	.29	100	-0.876	.004	.34	128	-0.871	.004	.37	67
β_7	0.035	.001	.41	81	0.034	.002	.28	125	0.032	.002	.33	94	0.032	.001	.36	77

Table 1: Results from MCMC for labour force participation using $N = 100,000$, $M = 1,000$, $M = 2,000$ and $M = 4,000$. mcse denotes the estimate of the Monte Carlo standard error of the posterior expectation, Pr denotes the probability of accepting a Metropolis-Hastings proposal for that parameter and inef denotes the rate of inefficiency of the MCMC sampler.

Carlin, Stern, and Rubin (2003). We loop through the parameters to make a proposal for each one individually and accept or reject it. The proposal variances were tuned on the exact likelihood model. This will allow us to see how fast the estimated likelihood comes close to the truth as we increase M . For the algorithm we set $M = 1,000$, $M = 2,000$, $M = 4,000$ and $N = 100,000$.

3.4 The results

Table 1 shows the following statistics for the MCMC algorithm: The arithmetical mean, the Monte Carlo standard error, the acceptance probability, and the inefficiency of the MCMC algorithm compared to a hypothetical sampler which produces i.i.d. draws from the posterior distribution.⁴ We discard the first half of the MCMC sample, to reduce the impact of burn-in, so that all statistics are based on the second half and so roughly represent draws from the posterior distribution.

	posterior covariance and correlation															
	Exact likelihood							$M = 4,000$								
β_0	.20	.07	-.54	-.15	.18	-.75	-.30	-.41	.16	.02	-.43	-.21	.23	-.71	-.29	-.40
β_1	.00	.00	-.34	.05	.02	-.13	.04	-.04	.00	.00	-.34	.03	.05	-.09	.08	-.01
β_2	-.01	.00	.00	-.06	.04	.00	-.17	.05	.00	.00	.00	-.03	.03	-.16	-.25	-.04
β_3	.00	.00	.00	.00	-.92	-.07	-.03	.06	.00	.00	.00	.00	-.92	-.06	-.01	.09
β_4	.00	.00	.00	.00	.00	-.06	.01	.02	.00	.00	.00	.00	.00	-.07	.01	-.01
β_5	.00	.00	.00	.00	.00	.00	.44	.29	.00	.00	.00	.00	.00	.00	.44	.29
β_6	-.02	.00	.00	.00	.00	.00	.01	.10	-.01	.00	.00	.00	.00	.00	.01	.10
β_7	-.01	.00	.00	.00	.00	.00	.00	.00	-.01	.00	.00	.00	.00	.00	.00	.00

Table 2: Results from MCMC for labour force participation; exact likelihood model with $N = 100,000$ and approximate analysis using $M = 4,000$; posterior covariance (lower triangle) and correlation (upper triangle) matrix.

The acceptance probabilities increase with M , so that with $M = 4,000$ we have almost the 40% acceptance rate as the proposal variances were tuned for the exact likelihood model. The inefficiency decreases with M but not monotonically so. Almost all of the posterior means based on the estimated likelihood models are not significantly different from the exact likelihood model.

⁴Inefficiency is defined as $1 + 2 \sum_{l=1}^{500} (1 - \frac{l}{500}) \rho_l$ where ρ_l denotes autocorrelation at lag l .

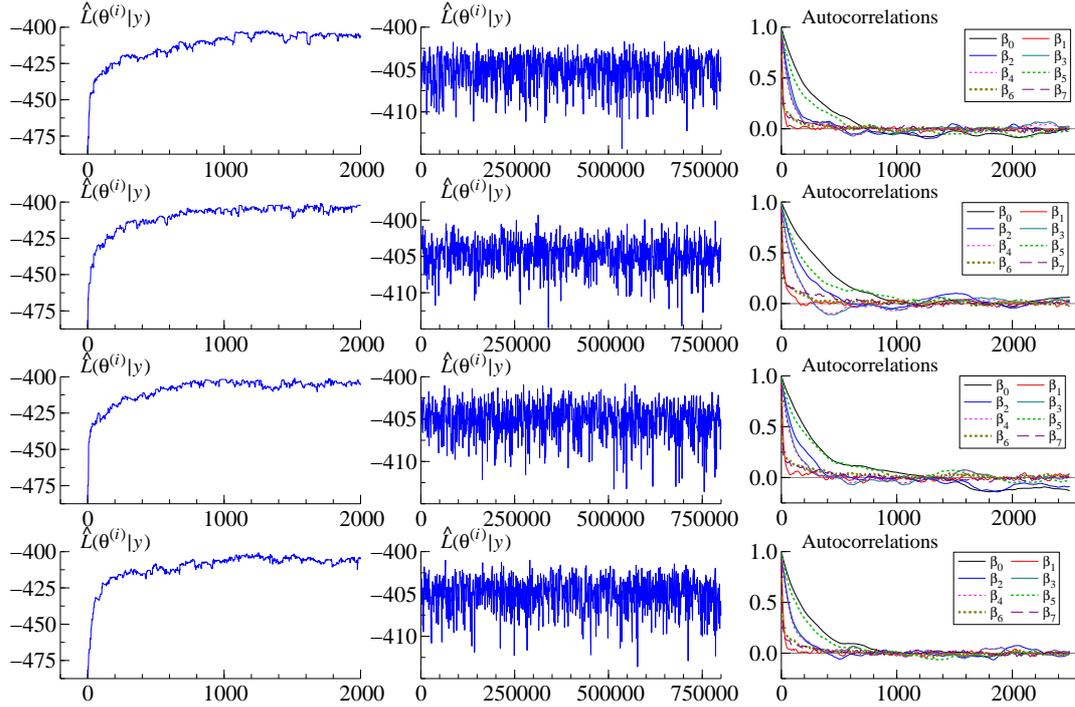


Figure 1: Labour force participation model; Likelihoods and ACFs of parameters. First column: $\hat{L}(\theta^{(i)})$ for $i = 1, \dots, 2000$; second column: $\hat{L}(\theta^{(i)})$ for $i = 2001, \dots, 8N$; third column: ACF of $\theta^{(i)}$. First row: exact likelihood model; second row: estimated likelihood with $M = 1000$; third row: $M = 2000$; fourth row: $M = 4000$.

Our results are very close to those from an ordinary Probit regression. All parameter estimates have the expected sign, except β_7 .

Table 2 shows the posterior covariance (lower triangle) and correlation (upper triangle) matrix of the parameters. For brevity we only report the covariance matrices of the exact likelihood and the $M = 4,000$ estimated likelihood model.

Since the $\theta^{(i)}$ are highly correlated we use the Newey-West Heteroskedasticity and Autocorrelation Consistent (HAC) estimator for the computation of the Monte Carlo standard error (mcse). The choice of the lag length B for this estimator is not obvious. The results reported are based on a choice of $B = 500$.

Figure 1 compares the likelihood estimates and the autocorrelation functions (ACFs) for $N = 100,000$ between the exact likelihood model, and the ones with estimated likelihoods with $M = 1000$, $M = 2000$ and $M = 4000$. Figure 2 compares the estimated posterior densities of the parameters. We note that the performance improves as M increases.

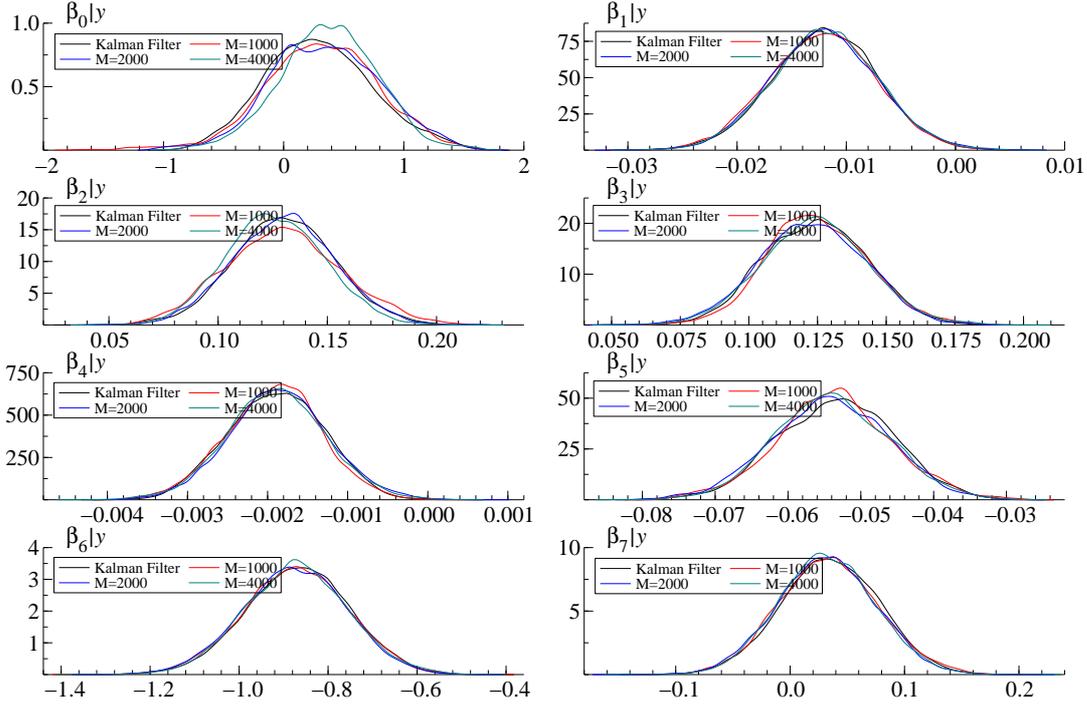


Figure 2: Labour force participation model; non-parametric kernel estimator of posterior density for $i = 50,000, \dots, 100,000$. Black: exact likelihood model; red: estimated likelihood with $M = 1,000$; blue: $M = 2,000$; green: $M = 4,000$.

4 Dynamic economic models

We now move on to treat parameter estimation in dynamic economic models. First, we provide the very general assumptions on the models we consider here. Then, we describe the MCMC algorithm and particle filter we will use.

4.1 Assumptions

We assume we have some observations

$$y = (y_1, y_2, \dots, y_T)$$

and wish to make Bayesian inference on some unknown parameters θ . We consider an underlying non-linear and non-Gaussian state-space model of the following type.

Assumption 1 *The simulation state space (SSS) model:*

1. We can compute the measurement density

$$f(y_t | \alpha_t, \mathcal{F}_{t-1}, \theta), \quad t = 1, 2, \dots, T,$$

where α_t is the unobserved state and $\mathcal{F}_{t-1} = y_1, y_2, \dots, y_{t-1}$ is the natural filtration.

2. We can simulate from the random variable

$$\alpha_t | \alpha_{t-1}, \mathcal{F}_{t-1}, \theta, \quad t = 1, 2, \dots, T,$$

where we assumed that we can also draw from the initial condition $\alpha_0 | \mathcal{F}_0, \theta$.

3. We can compute the prior $f(\theta)$.

At no point do we assume we can compute $f(\alpha_{t+1} | \alpha_t, \mathcal{F}_t, \theta)$. Computing this distribution is often hard in models encountered in economics and finance, but we can often simulate from it. We do not assume that such simulations are continuous with respect to θ with common random numbers (which allows the use of rejection in the simulation). Continuity in θ plays no role at all in our analysis. We will argue in Sections 6 and 7 that a large number of intractable econometric models are of this simulation state space form. Leading examples are DSGE models, some (continuous time) stochastic volatility models and some models in industrial organisation.

4.2 Predictive decomposition

For dynamic models we can always decompose the likelihood using the predictive decomposition

$$f(y | \mathcal{F}_0, \theta) = \prod_{t=1}^T f(y_t | \mathcal{F}_{t-1}, \theta).$$

This is key to the success of the Kalman filter and the use of hidden Markov models (e.g. Durbin and Koopman (2001)), where the predictive distributions $f(y_t | \mathcal{F}_{t-1}, \theta)$ can be computed exactly using recursive formulae.

In more general models the predictive distributions can only be approximated. Here we will use simulation to unbiasedly estimate $f(y_t | \mathcal{F}_{t-1}, \theta)$. This will be carried out using a particle filter, whose recursive structure will allow us to calculate an unbiased estimator of $f(y | \mathcal{F}_0, \theta)$. This can then be used as the basis for inference using an MCMC algorithm analogous to the above strategy.

4.3 Particle filter estimator of the likelihood

The modern statistical literature on particle filters started with Gordon, Salmond, and Smith (1993), while a book length review is given in Doucet, de Freitas, and Gordon (2001). Kim, Shephard, and Chib (1998) and Pitt and Shephard (1999) introduced particle filters into economics and estimated the likelihood function $f(y | \theta)$ as a by-product in order to do model comparison via marginal likelihoods. Particle filters have recently received some attention in macroeconomics due to the work of, for example, Fernandez-Villaverde and Rubio-Ramirez (2007) and Aruoba, Fernandez-Villaverde, and Rubio-Ramirez (2006).

Here we give a very simple particle filter, which can be generically coded for simulation state space models.

Particle filter

1. Draw $\alpha_1^{(1)}, \dots, \alpha_1^{(M)}$ from $\alpha_1 | \mathcal{F}_0, \theta$. Set $t = 1$ and $l_0 = 0$.
2. Compute the weights

$$w_t^{(j)} = f(y_t | \alpha_t^{(j)}, \mathcal{F}_{t-1}, \theta), \quad j = 1, \dots, M$$

and the normalised weights so that

$$W_t^{(j)} = \frac{w_t^{(j)}}{\sum_{k=1}^M w_t^{(k)}} \quad j = 1, \dots, M$$

3. Resample by drawing $u \sim U(0, 1)$ and let

$$u^{(j)} = \frac{u}{M} + \frac{j-1}{M}$$

for $j = 1, \dots, M$. Find the indices i^1, \dots, i^M such that

$$\sum_{k=1}^{i^j-1} W_t^{(k)} < u^{(j)} \leq \sum_{k=1}^{i^j} W_t^{(k)} \quad (1)$$

4. Sample

$$\alpha_{t+1}^{(j)} \sim \alpha_{t+1} | \alpha_t^{(i^j)}, \mathcal{F}_t, \theta$$

for $j = 1, \dots, M$.

5. Record

$$l_t(\theta) = l_{t-1}(\theta) + \log \left\{ \frac{1}{M} \sum_{j=1}^M w_t^{(j)} \right\}$$

Let $t = t + 1$ and goto 2.

Then it is relatively easy to show that

$$\exp(l_T(\theta)) \xrightarrow{a.s.} f(y|\theta)$$

as $M \rightarrow \infty$ (see for example Del-Moral (2004))⁵, while crucially for us

$$E[\exp(l_T(\theta))] = f(y|\theta).$$

⁵An important aspect of the convergence is that it is pointwise: particle filters are not continuous with respect to θ so neither is $l_T(\theta)$. This is a fundamental property of particle filters and it is caused by the resampling step (1), which is very hard to overcome unless α_t is univariate — see Pitt (2001). Discontinuous estimates of the likelihood function cause problems for maximum simulated likelihood techniques such as those proposed by Fernandez-Villaverde, Rudio-Ramirez, and Santos (2006) and Fernandez-Villaverde and Rudio-Ramirez (2007). To our knowledge only Flury and Shephard (2008) offer a smooth particle filter which circumvents this problem in the multivariate state case.

See also Del-Moral (2004).

In order to carry out MCMC we proceed by replacing $f(y|\theta)$ by its particle filter estimate

$$\widehat{L}(\theta) = \exp \{l_T(\theta)\}$$

which we are allowed to do as the particle filter indeed provides an unbiased estimate of $f(y|\theta)$, which can be used inside an MCMC algorithm as discussed in Section 2.2. This has the same structure as before, delivering simulations from the posterior distribution.

We now move on to provide some examples in economics to demonstrate the performance of this algorithm.

5 Example: Gaussian linear model

We start off by analysing a simple linear Gaussian model where an analytical expression for $f(y|\theta)$ is readily available from the Kalman filter in order to be able to evaluate the particle MCMC algorithm.

We consider the Gaussian linear model (see e.g. Harvey (1989) and Durbin and Koopman (2001))

$$\begin{aligned} y_t &= \mu + \alpha_t + \sigma_\epsilon \varepsilon_t, \\ \alpha_{t+1} &= \phi \alpha_t + \sigma_\eta \eta_t, \end{aligned} \quad \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \stackrel{i.i.d.}{\sim} N(0, I),$$

where $\alpha_0 \sim N(0, \sigma_\eta^2 / (1 - \phi^2))$. In order to guarantee positive variances we parameterise the log of the variances, while we impose that $\phi \in (-1, 1)$ by allowing no prior probability outside that region. We take $\theta = (\mu, \log \sigma_\epsilon^2, \phi, \log \sigma_\eta^2)'$, where μ controls the unconditional mean of y_t , ϕ the autocorrelation and σ_η^2 the variance of the latent process. The likelihood can be computed using the Kalman filter and this will serve us as a benchmark. For our simulation study we generate $T = 1,000$ observations from this DGP with parameterization $\theta^* = (0.5, \log 1, 0.825, \log 0.75)'$. We assume a Gaussian prior given by $\theta \sim N(\theta_0, I_4)$ where $\theta_0 = (0.25, \log 1.5, 0.475, \log 0.475)'$.

Any proposals for $\phi \notin (-1, 1)$ are automatically rejected. We are using the following four random walk proposals for the transformed parameters, each applied one at a time:

$$\begin{aligned} \mu_i &= \mu_{i-1} + 0.3298\nu_{1,i} & \phi_i &= \phi_{i-1} + 0.0671\nu_{3,i} \\ \log \sigma_{\epsilon,i}^2 &= \log \sigma_{\epsilon,i-1}^2 + 0.1866\nu_{2,i} & \log \sigma_{\eta,i}^2 &= \log \sigma_{\eta,i-1}^2 + 0.2676\nu_{4,i} \end{aligned}$$

where $\nu_{j,i} \sim i.i.N(0, 1)$ for $j = 1, \dots, 4$ and $i = 1, \dots, N$. The variances in the random walk proposals were chosen to aim for a 40% acceptance probability for each parameter. We loop through the parameters to make a proposal for each one individually and accept or reject it. Generally we run the MCMC algorithm with N iterations and allow for a burn-in period of $\frac{N}{2}$, such that all of the below statistics are computed with the second half of the draws.

We start by running the Kalman filter for a chain length of $N = 100,000$. The last column in Figure 4 shows the Markov chain output for $\mu|y$ and $\sigma_\epsilon^2|y$ and the histograms, which are only based on the second half of the chains (the other parameters have similar types of sample paths). Next we ran the particle filter at $M = 100$ and $N = 100,000$. As we can see in the first column in Figure 4, using only $M = 100$ particles appears to be insufficient for the length of the chain considered here. The chain gets stuck on specific parameter values for a considerable time. To see what happens in the very long run we let the MCMC algorithm run up to $N = 1,000,000$ and the second column in Figure 4 shows the parameter chains and histograms in this scenario. We see that even though the number of particles is too small at $M = 100$, we can make up for it by having the Markov chain run longer. The chain still gets stuck but the histograms start to look better. We now look at the impact of increasing the number of particles from only $M = 100$ to $M = 1,000$. The third column in Figure 4 shows the drastic improvement in the performance of the algorithm from this increase. Both parameter chains and histograms start looking very similar to the Kalman filter results.

Table 3 compares the output of the MCMC algorithm from the Kalman Filter, with that from the $M = 100$, $N = 1,000,000$ and that from the $M = 1,000$, $N = 100,000$ particle filter. The table shows the following statistics: The arithmetical mean, the Monte Carlo standard error, the acceptance probability, the posterior variance and the inefficiency. We see how the acceptance probabilities improve with M . We do not report it here for brevity, but with $M = 2,000$ we achieve an acceptance probability of 30%. To see that effect more clearly Figure 3 compares the evolution of the ACFs and likelihoods as we run through the setups $M = 100$ with $N = 100,000$, $M = 100$ with $N = 1,000,000$ and $M = 1,000$ with $N = 100,000$ with the Kalman filter output.

	exact likelihood					$M = 100, N = 10^6$					$M = 1,000$				
	mean	mcese	Pr	var	inef	mean	mcese	Pr	var	inef	mean	mcese	Pr	var	inef
μ	0.562	.001	.40	.015	7.2	0.553	.002	.02	.013	531	0.562	.001	.25	.014	7.8
σ_ϵ^2	1.030	.003	.40	.012	39	1.033	.003	.02	.010	855	1.031	.003	.25	.012	29
ρ	0.783	.001	.39	.001	37	0.785	.001	.02	.001	775	0.783	.001	.25	.001	30
σ_η^2	0.621	.004	.40	.015	48	0.607	.003	.02	.011	784	0.618	.003	.26	.015	36

Table 3: Results from MCMC for the linear state space model using $N = 100,000$ for the exact likelihood (computed by the Kalman filter) and the simulation likelihood based on $M = 1,000$. When $M = 100$ the MCMC algorithm became very inefficient and we needed to use $N = 1,000,000$. mcese denotes the estimate of the Monte Carlo standard error of the posterior expectation, Pr denotes the probability of accepting a Metropolis-Hastings proposal for that parameter and inef denotes the rate of inefficiency of the MCMC sampler.

As a final note we would like to comment on how we suggest choosing M and the variances of the random-walk proposals. We find that a good indication whether one has reached a sufficient number of particles – sufficient in the sense of achieving a likelihood estimate that is not too jittery – is when the speed with which the acceptance probabilities increase with M starts to slow down and improvements become only marginal. Once this point has been reached we recommend tuning the

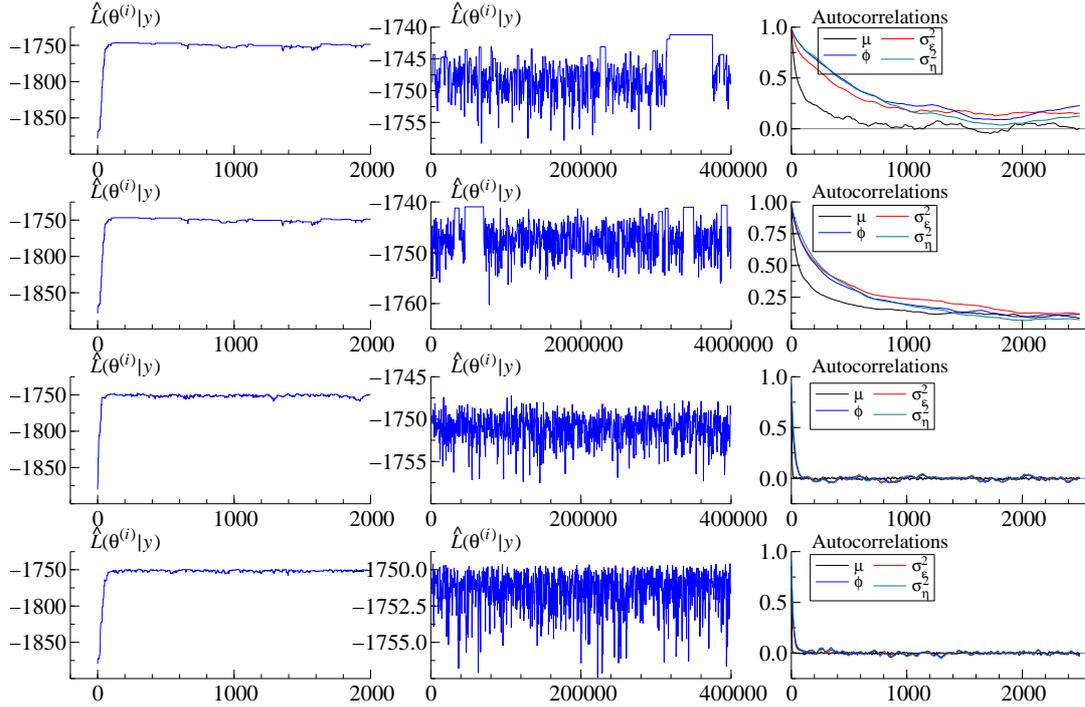


Figure 3: Random walk with noise model; first row: likelihoods and ACFs from $M = 100$ with $N = 100,000$ particle filter; second row: $M = 100$ with $N = 1,000,000$; third row: $M = 1,000$ with $N = 100,000$; fourth row: Kalman filter with $N = 100,000$.

proposal variances to get the desired levels for the acceptance probabilities. If one ends up having to decrease variances by a lot to get acceptance probabilities of around 40% for long chains this is an indication that M is not sufficiently large. It is helpful to always keep an eye on the ACFs. If one has to use small proposal variances to get acceptance probabilities of 40% and observes highly autocorrelated chains at the same time this is another strong indicator that M is too small.

6 Example: discrete time Gaussian SV model

We now turn to a simple real life problem and estimate the Gaussian discrete time stochastic volatility model (e.g. the reviews in Ghysels, Harvey, and Renault (1996) and Shephard (2005)).

The stock returns are assumed to follow the process

$$y_t = \mu + \exp\{\beta_0 + \beta_1 \alpha_t\} \varepsilon_t$$

and the stochastic volatility factor

$$\alpha_{t+1} = \phi \alpha_t + \eta_t,$$

where

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \stackrel{i.i.d.}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \quad \alpha_0 \sim N\left(0, \frac{1}{1 - \phi^2}\right).$$

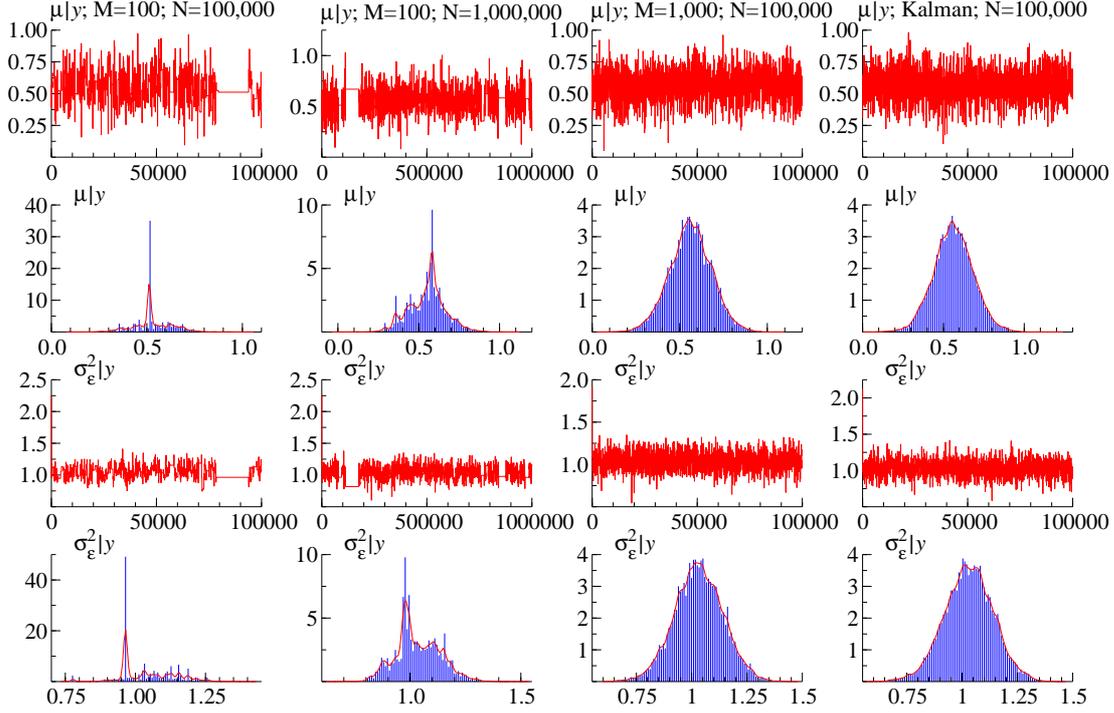


Figure 4: Random walk with noise model; particle chains and histograms for parameters μ and σ_ϵ^2 ; first column: particle filter with $M = 100$ and $N = 100,000$; second column: $M = 100$ and $N = 1,000,000$; third column: $M = 1,000$ and $N = 100,000$; fourth column: Kalman filter with $N = 100,000$.

Now we have $\theta = (\mu, \beta_0, \beta_1, \phi, \rho)'$. For this model the likelihood is not available. Researchers have used MCMC, MSLE, method of moments and indirect inference to estimate this type of model. See the reviews above for a discussion of the literature on this.

We assume a Gaussian prior given by $\theta \sim N(\theta_0, I_5)$ where $\theta_0 = (0.036, -0.286, 0.077, 0.984, -0.794)'$. Any proposals for $\phi, \rho \notin (-1, 1)$ are automatically rejected. We are using the following random walk proposals

$$\begin{aligned} \mu_i &= \mu_{i-1} + 0.017\nu_{1,i}, & \phi_i &= \phi_{i-1} + 0.004\nu_{4,i}, \\ \beta_{0,i} &= \beta_{0,i-1} + 0.104\nu_{2,i}, & \rho_i &= \rho_{i-1} + 0.067\nu_{5,i}, \\ \beta_{1,i} &= \beta_{1,i-1} + 0.010\nu_{3,i}, \end{aligned}$$

where $\nu_{j,i} \stackrel{i.i.d.}{\sim} N(0, 1)$ for $j = 1, \dots, 5$ and $i = 1, \dots, N$. We loop through the parameters to make a proposal for each one individually and accept or reject it.

The data we use for this study is the end-of-day level of the SNP500 Composite Index (NYSE/AMEX only) from CRSP. We use 3271 daily observations from 03.01.1995 until 31.12.2007. The daily returns are defined as

$$y_t = 100 (\log \text{SNP500}_t - \log \text{SNP500}_{t-1})$$

The statistics are displayed in Table 4. The most interesting aspect of the estimated parameters

	mean	mcse	Pr	posterior covariance and correlation					inef
μ	0.042	0.000	0.410	0.000	-0.686	-0.137	0.304	0.145	15
β_0	-0.141	0.001	0.395	-0.001	0.006	0.061	-0.222	-0.044	12
β_1	0.080	0.000	0.398	0.000	0.000	0.000	-0.688	-0.015	18
ϕ	0.982	0.000	0.424	0.000	0.000	0.000	0.000	-0.054	16
ρ	-0.742	0.000	0.427	0.000	0.000	0.000	0.000	0.002	6.4

Table 4: Results from the particle MCMC algorithm for the stochastic volatility model; $N = 100,000$, $M = 2,000$. The correlations are above the leading diagonal. Inef denotes the rate of inefficiency of the MCMC sampler, Pr denotes the probability of accepting a Metropolis-Hastings proposal for that parameter and mcse denotes our estimate of the Monte Carlo standard error of the posterior expectation.

is the extremely strong negative statistical leverage parameter ρ and the high negative posterior correlation between μ and β_0 and β_1 and ϕ . It is interesting to observe that ρ is not importantly correlated with the other parameters in the model.

The data together with the filtered time-series for daily volatility $E\{\exp(\beta_0 + \beta_1\alpha_t) | \mathcal{F}_t\}$, estimated by using the posterior means of the parameters, is plotted in Figure 5. Figure 6 depicts likelihoods and ACFs from the MCMC output.

This analysis is based on $M = 2,000$ and $N = 100,000$. Convergence looks pretty fast for this algorithm, with very modest inefficiency factors. Of course it is computationally demanding as each simulated likelihood evaluation is quite expensive in terms of time, but the coding effort is very modest indeed. The implementation of this method is quite simple and seems competitive with other procedures put forward in the literature.

7 Example: a dynamic stochastic equilibrium model

We now estimate a simple DSGE model. An and Schorfheide (2007) were the first to consider Bayesian inference for DSGE models. Fernandez-Villaverde and Rudio-Ramirez (2007) were the first to consider using particle filters to perform parameter inference. In their paper however they only report maximum simulated likelihood estimates and indicate the possibility of using this present algorithm but provide no rationale for why it would work. As mentioned earlier the kind of maximum simulated likelihood approach they use runs into difficulties incurred because of the pointwise convergence of the particle filter estimate of the likelihood. Amisano and Tristani (2007) seem to use this particle filter in MCMC algorithm to estimate a DSGE model, but without any further justification.

The model we consider here is a simple version of a DSGE model. There is a representative household maximizing its lifetime utility, given by

$$\max_{\{C_t, L_t\}_{t=0}^{\infty}} E_0 \left[\sum_{t=0}^{\infty} \beta^t \{ \log(C_t) + \psi \log(1 - L_t) \} \right], \quad \beta \in (0, 1), \psi > 0,$$

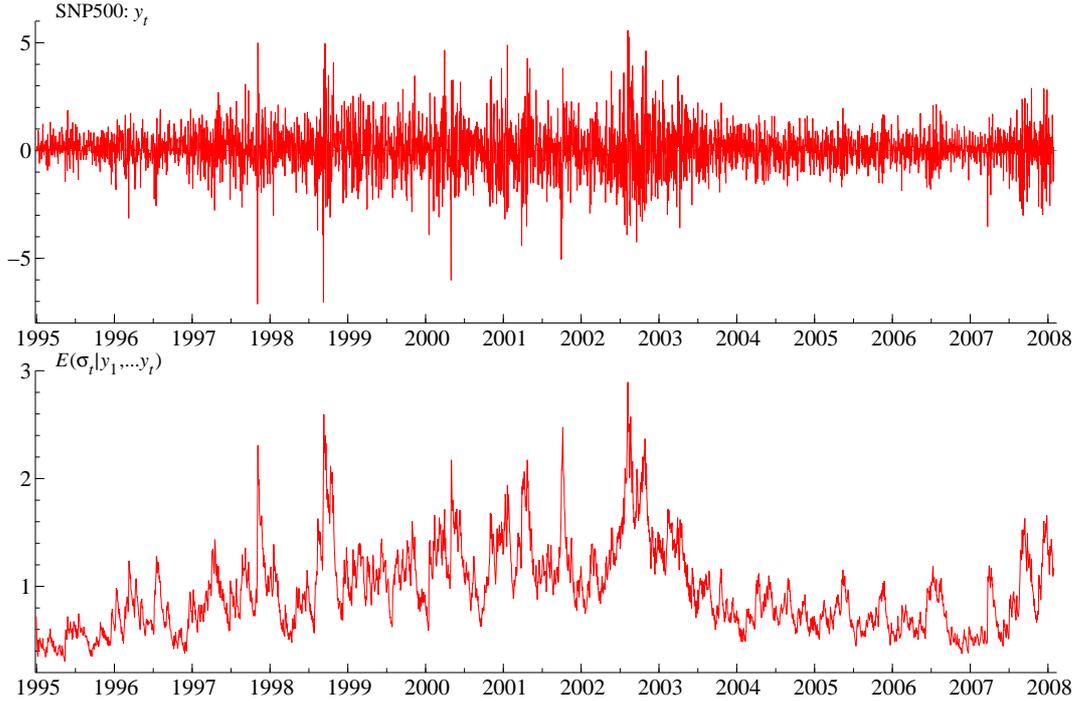


Figure 5: Stochastic volatility model; Top: log-returns on end-of-day level of the SNP500 Composite Index (NYSE/AMEX only) from 03.01.1995 until 31.12.2007; bottom: $E\{\exp(\beta_0 + \beta_1 \alpha_t | \mathcal{F}_t)\}$ based on posterior means of the parameters.

where C_t is consumption, $L_t \in [0, 1)$ labour, β the discount factor and ψ determines labour supply.

In this economy there is one single good which is produced according to

$$Y_t = A_t K_t^\alpha L_t^{1-\alpha},$$

where K_t is the stock of capital and A_t technology. The stock of capital evolves according to

$$K_{t+1} = (1 - \delta) K_t + U_t I_t,$$

where I_t is investment, U_t investment technology and δ the depreciation rate.

We further assume this to be a closed economy without government, such that the aggregate resource constraint is given by

$$C_t + I_t = Y_t.$$

In the discussion that follows E_t stands for the expectation of functions of future (A_t, U_t) given information available at time t .

We assume the following model for the technologies

$$\log A_t = \rho_a \log A_{t-1} + \sigma_a \eta_{a,t},$$

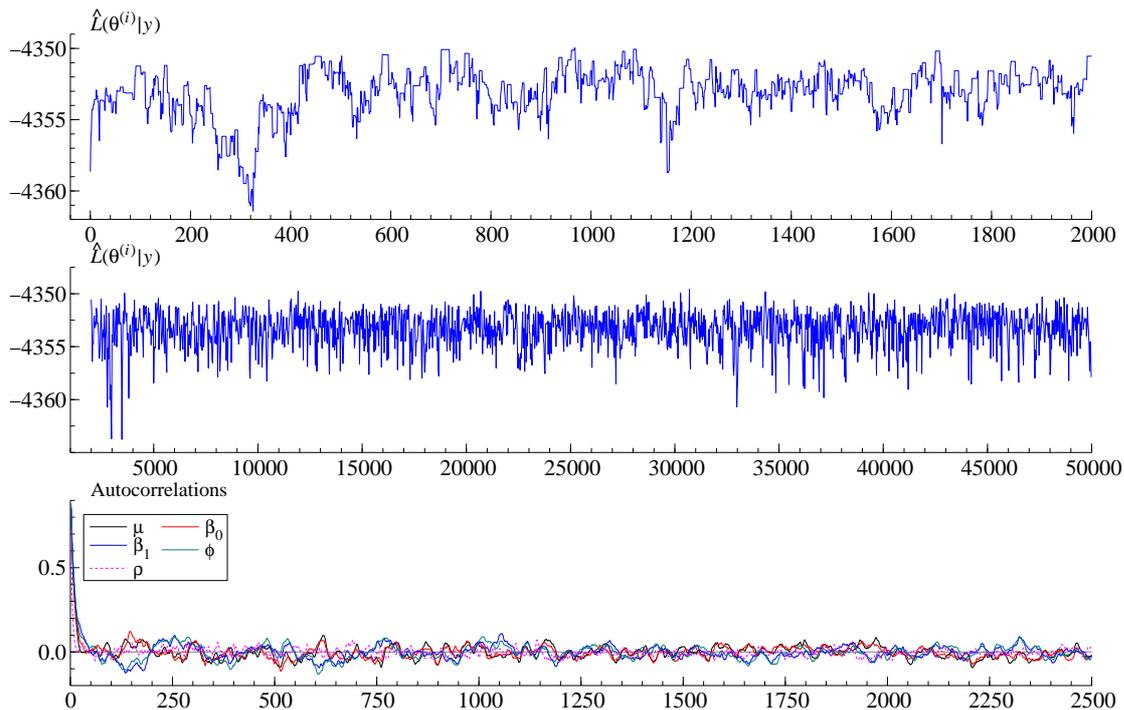


Figure 6: Stochastic volatility model; particle filter with $M = 2000$: Likelihoods and ACFs of parameters. Top: $\hat{L}(\theta^{(i)})$ for $i = 1, \dots, 2000$; middle: $\hat{L}(\theta^{(i)})$ for $i = 2001, \dots, 5N$; bottom: ACF of $\theta^{(i)}$.

$$\log U_t = \rho_u \log U_{t-1} + \sigma_u \eta_{u,t},$$

where $\eta_{a,t}, \eta_{u,t} \stackrel{i.i.d.}{\sim} N(0, 1)$.

It is supposed that the econometrician is given time series of Y_t and L_t observed with noise and that the task is to carry out inference on

$$\theta = (\alpha, \beta, \delta, \psi, \rho_a, \rho_u, \sigma_{\epsilon_y}, \sigma_{\epsilon_l}, \sigma_a, \sigma_u)',$$

where $\sigma_{\epsilon_y}, \sigma_{\epsilon_l}$ are the standard errors of the measurement error for the observations. Their functional form will be explained in more detail in a minute.

In this economy the central planner and the competitive equilibrium coincide. We decide to solve the central planner's problem

$$\max_{\{K_{t+1}, L_t\}_{t=0}^{\infty}} \left\{ \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left(\log \left\{ A_t K_t^\alpha L_t^{1-\alpha} + \frac{1}{U_t} ((1-\delta) K_t - K_{t+1}) \right\} + \psi \log \{1 - L_t\} \right) \right] \right\}$$

The first order equilibrium conditions

$$\frac{1}{C_t} = U_t \beta \mathbb{E}_t \left[\frac{1}{C_{t+1}} \left\{ \alpha A_{t+1} K_{t+1}^{\alpha-1} L_{t+1}^{1-\alpha} + \frac{1}{U_{t+1}} (1-\delta) \right\} \right],$$

$$\psi \frac{1}{1-L_t} = \frac{1}{C_t} (1-\alpha) A_t K_t^\alpha L_t^{-\alpha},$$

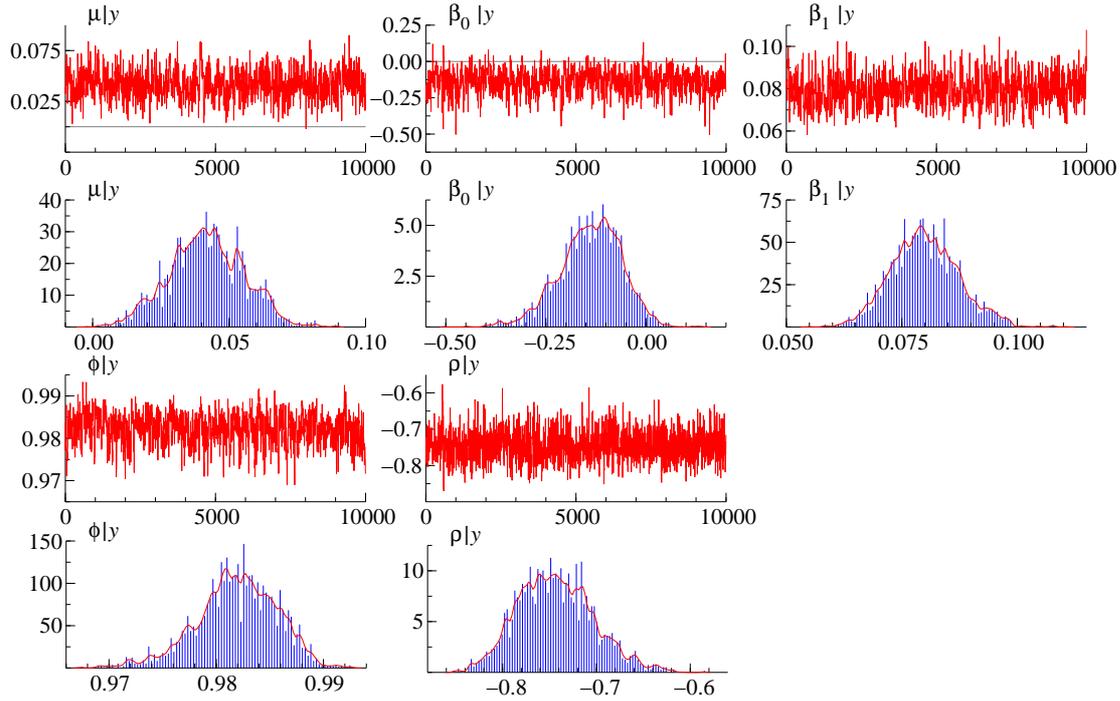


Figure 7: Stochastic volatility model; particle filter with $M = 2,000$: $\theta^{(i)}$ for $i = 1, \dots, 100,000$ and histogram of parameters for $i = 50,000, \dots, 100,000$.

together with the resource constraint

$$C_t = A_t K_t^\alpha L_t^{1-\alpha} + \frac{1}{U_t} ((1 - \delta) K_t - K_{t+1}),$$

and the technology processes fully characterize the solution to the problem.

Solving this system of non-linear expectational difference equations involves finding policy functions g and h such that

$$\begin{aligned} (C_t, L_t)' &= g(K_t, A_t, U_t), \\ (K_{t+1}, A_{t+1}, U_{t+1})' &= h(K_t, A_t, U_t) + \sigma \Omega \eta_{t+1}, \end{aligned}$$

where σ is the perturbation parameter and

$$\Omega = \begin{pmatrix} 0 & 0 \\ \sigma_a & 0 \\ 0 & \sigma_u \end{pmatrix}, \quad \eta_{t+1} = \begin{pmatrix} \eta_{a,t+1} \\ \eta_{u,t+1} \end{pmatrix}.$$

We find a second order approximation to these policy functions by perturbation methods. We solve the system in terms of log-deviations from a non-stochastic steady-state and use notation $\hat{c}_t = \log C_t / C_{ss}$, where C_{ss} denotes the non-stochastic steady-state.

Now unify notation by letting \hat{x}_t denote the state variables $\hat{x}_t = (\hat{k}_t, \hat{a}_t, \hat{u}_t)'$. The solution will

be of the state space form

$$\begin{aligned}\widehat{k}_{t+1} &= h_{x,1}\widehat{x}_t + \frac{1}{2}\widehat{x}'_t h_{xx,1}\widehat{x}_t + \frac{1}{2}h_{\sigma\sigma,1}\sigma^2, \\ \widehat{a}_t &= \rho_a\widehat{a}_{t-1} + \sigma_a\eta_{a,t}, \\ \widehat{u}_t &= \rho_u\widehat{u}_{t-1} + \sigma_u\eta_{u,t},\end{aligned}\tag{2}$$

and

$$\begin{aligned}\widehat{c}_t &= g_{x,1}\widehat{x}_t + \frac{1}{2}\widehat{x}'_t g_{xx,1}\widehat{x}_t + \frac{1}{2}g_{\sigma\sigma,1}\sigma^2, \\ \widehat{l}_t &= g_{x,2}\widehat{x}_t + \frac{1}{2}\widehat{x}'_t g_{xx,2}\widehat{x}_t + \frac{1}{2}g_{\sigma\sigma,2}\sigma^2.\end{aligned}$$

We rely on code from Schmitt-Grohe and Uribe (2004) and Klein (2000) to solve for the unknown derivatives $h_x, g_x, h_{xx}, g_{xx}, h_{\sigma\sigma}, g_{\sigma\sigma}$.

We make the simple assumption that the observable variables are given by

$$\begin{aligned}\widehat{GDP}_t &= \widehat{y}_t + \sigma_{\epsilon_y}\epsilon_{y,t}, \\ \widehat{H}_t &= \widehat{l}_t + \sigma_{\epsilon_l}\epsilon_{l,t},\end{aligned}\tag{3}$$

where \widehat{GDP}_t represents detrended real GDP per capita and \widehat{H}_t represents hours worked per capita.

From our model we compute these as

$$\begin{aligned}\widehat{y}_t &= \widehat{a}_t + \alpha\widehat{k}_t + (1 - \alpha)\widehat{l}_t, \\ \widehat{l}_t &= g_{x,2}\widehat{x}_t + \frac{1}{2}\widehat{x}'_t g_{xx,2}\widehat{x}_t + \frac{1}{2}g_{\sigma\sigma,2}\sigma^2.\end{aligned}$$

Equations (2) together with the observation equations (3) specify a non-linear state-space system, from which we can easily simulate and hence use the particle filter to evaluate the likelihood of the model $\widehat{L}(\theta)$. The parameters of the model are

$$\theta = (\alpha, \beta, \delta, \psi, \rho_a, \rho_u, \sigma_{\epsilon_y}, \sigma_{\epsilon_l}, \sigma_a, \sigma_u)'$$

As a demonstration of the model fitting exercise we limit ourselves to estimating a subset of the parameters. In particular we take α, β, δ and ψ as fixed and fit the remaining parameters. We use the particle filter in MCMC algorithm. The algorithm to obtain $\widehat{L}(\theta)$ inside the MCMC algorithm works as follows

1. Given $\theta^{(i)}$, compute $K_{ss}, A_{ss}, U_{ss}, C_{ss}, L_{ss}$.
2. Use perturbation methods to find numerical values for $h_x, g_x, h_{xx}, g_{xx}, h_{\sigma\sigma}, g_{\sigma\sigma}$.
3. Run the particle filter on the state-space system (3) and (2) to obtain $\widehat{L}(\theta^{(i)})$.

We use this model to simulate a realization of the economy. We generate $T = 200$ observations with the parameterization

α	β	δ	ψ	ρ_a	ρ_u	$\log \sigma_{\epsilon_y}$	$\log \sigma_{\epsilon_l}$	$\log \sigma_a$	$\log \sigma_u$
0.37	.9992	.0154	1.956	.98	.96	$\log 0.002457$	$\log 0.004877$	$\log 0.005$	$\log 0.042$

The values chosen are inspired by the findings in Fernandez-Villaverde and Rudio-Ramirez (2007).

In order to guarantee positive standard errors in the MCMC algorithm we parameterise the log of the standard errors. We assume a Gaussian prior with mean θ_0 , given by

ρ_a	ρ_u	$\log \sigma_{\epsilon_y}$	$\log \sigma_{\epsilon_l}$	$\log \sigma_a$	$\log \sigma_u$
.95	.90	$\log 0.003457$	$\log 0.005877$	$\log 0.006$	$\log 0.052$

Any proposals for $\rho_a, \rho_u \notin (-1, 1)$ are automatically rejected. We are using the following random walk proposals

$$\begin{aligned} \rho_{a,i} &= \rho_{a,i-1} + 0.0545\nu_{1,i}, & \log \sigma_{\epsilon_l,i} &= \log \sigma_{\epsilon_l,i-1} + 0.159\nu_{4,i}, \\ \rho_{u,i} &= \rho_{u,i-1} + 0.0077\nu_{2,i}, & \log \sigma_{a,i} &= \log \sigma_{a,i-1} + 0.1125\nu_{5,i}, \\ \log \sigma_{\epsilon_y,i} &= \log \sigma_{\epsilon_y,i-1} + 0.251\nu_{3,i}, & \log \sigma_{u,i} &= \log \sigma_{u,i-1} + 0.06\nu_{6,i}, \end{aligned}$$

where $\nu_{j,i} \sim i.i.N(0, 1)$ for $j = 1, \dots, 6$ and $i = 1, \dots, N$. We loop through the parameters to make a proposal for each one individually and accept or reject it.

For the particle filter we chose $M = 60,000$ particles. We have to use such a large number of particles because \widehat{k}_t does not have a stochastic noise term itself. This makes it a slowly mixing process and this gives the particle filter a hard time. It will take a long period of time to “forget” a small mistake made at $t = 0$, so the initial draw \widehat{k}_0 is very important for the performance of the DSGE estimation. The small variances in the observation processes make it a very spiky likelihood and this results in very few particles $\widehat{x}_0^{(i)}$ having very large weights. This results in the MCMC algorithm being stuck for a long time if we do not have a rather large number of particles.

We produce results for $N = 10,000$. As usual we display the likelihood and the ACFs in Figure 8 and the parameter histories and their histograms in Figure 9.

Table 5 shows the usual statistics and the variance-covariance matrix. The posterior means take reasonable values. The estimates for ρ_a and ρ_u seem a bit far away from their true values.

	mean	mcse	Pr	inef	posterior covariance and correlation					
ρ_a	0.9512	.001	.396	18	5.26E-04	0.034	0.100	0.029	-0.209	0.047
ρ_u	0.9353	.002	.487	78	0.17E-04	4.49E-04	-0.137	0.111	0.001	0.668
σ_{ϵ_y}	0.0020	.000	.448	103	0.02E-04	-0.03E-04	0.01E-04	-0.614	-0.291	-0.057
σ_{ϵ_l}	0.0053	.000	.424	22	0.01E-04	0.03E-04	-0.01E-04	0.01	-0.198	0.012
σ_a	0.0055	.000	.386	32	-0.03E-04	0.00E-04	0.00E-04	0.00E-04	0.00E-04	-0.004
σ_u	0.0394	.000	.421	54	0.04E-04	0.46E-04	0.00E-04	0.00E-04	0.00E-04	0.11E-04

Table 5: Results from MCMC with particle filter; $N = 10,000$, $M = 60,000$.

Fitting the DSGE model here is done in a spirit of demonstrating the workings and capabilities of the algorithm rather than gaining any new insight on model parameters. This concludes the dynamic model example section.

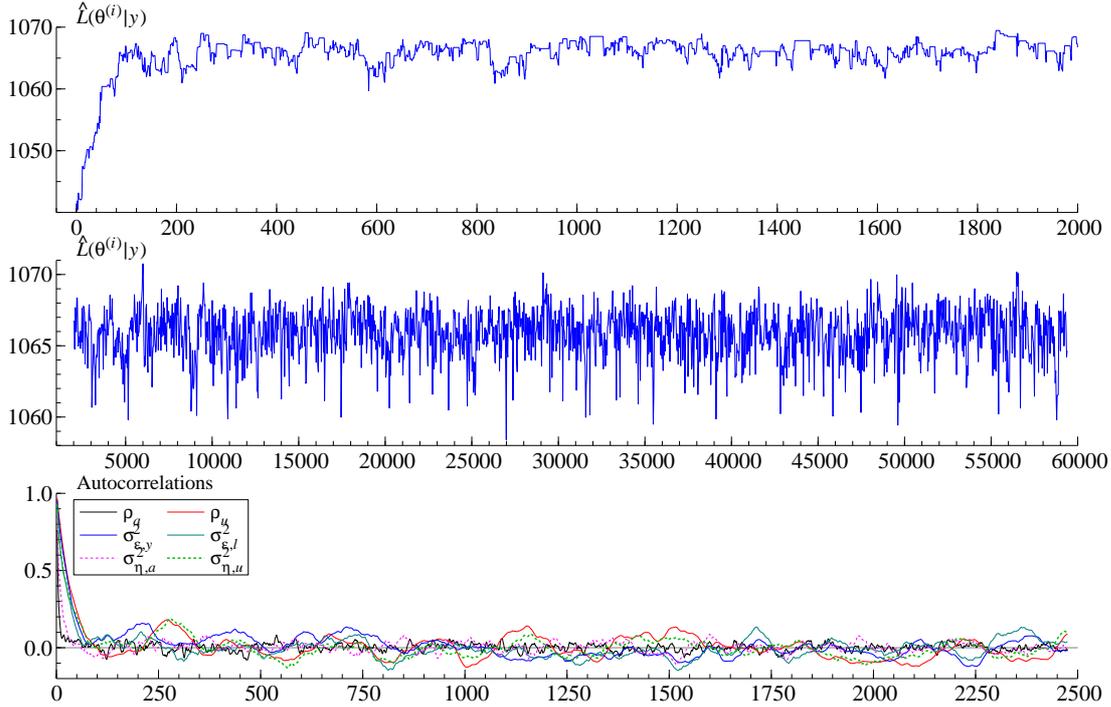


Figure 8: DSGE model; particle filter with $M = 60,000$: Likelihoods and ACFs of parameters. Top: $\hat{L}(\theta^{(i)})$ for $i = 1, \dots, 2000$; middle: $\hat{L}(\theta^{(i)})$ for $i = 2001, \dots, 6N$; bottom: ACF of $\theta^{(i)}$.

8 Conclusion

In the econometric literature estimated likelihoods are sometimes used as the basis for approximate maximum likelihood estimation. Such maximum simulated likelihood estimators have a deep theoretical flaw, as emphasised in the literature. In this paper we note that the effect of estimation can be removed by replacing the maximisation of the likelihood by placing the estimated likelihood inside an MCMC algorithm. The theory of this is very simple.

In this paper we show the power of this approach, providing examples drawn from microeconomics, financial econometrics and macroeconomics. When we use these methods on dynamic models it is convenient to use a particle filter to deliver an unbiased estimator of the likelihood. Such estimators are pretty general as they just need one to be able to simulate from the dynamics of the model to be able to implement it.

Particle filter based estimation of dynamic models is not fast, but it is conceptionally simple and applies very widely in economics. We believe that it provides a simple generalisation of the Kalman filter and can now handle many problems where estimation was previously cumbersome.

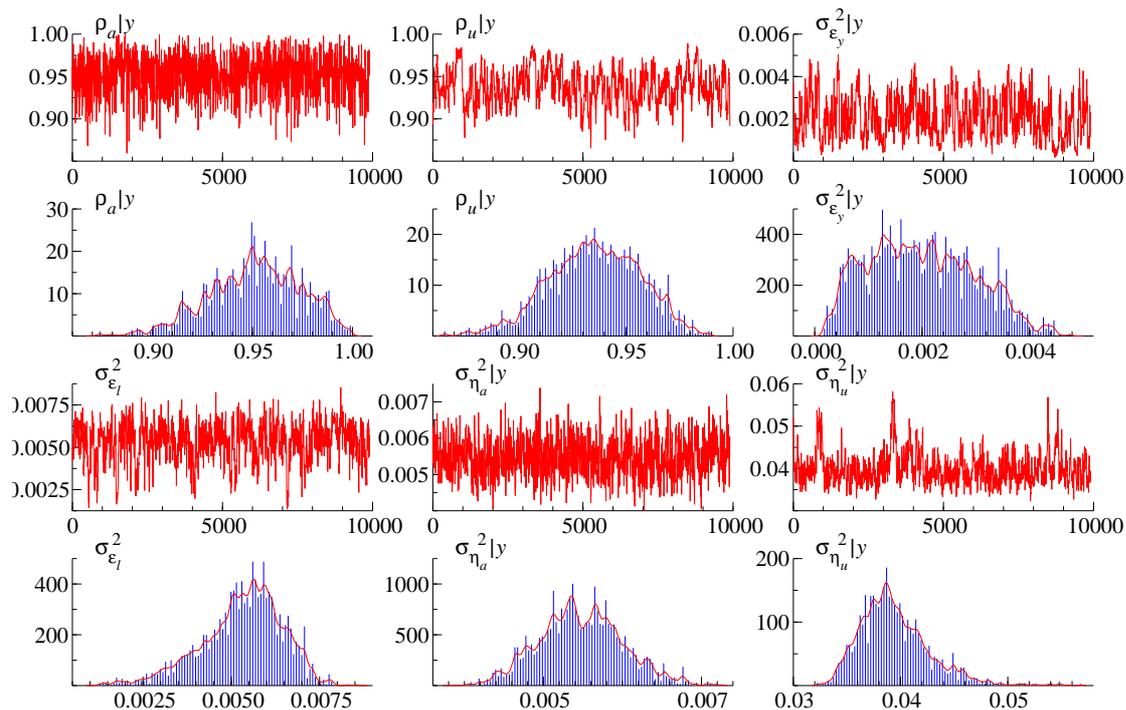


Figure 9: DSGE model; particle filter with $M = 60,000$: $\theta^{(i)}$ for $i = 1, \dots, 10,000$ and histogram of parameters for $i = 5,000, \dots, 10,000$.

References

- Amisano, G. and O. Tristani (2007). Euro area inflation persistence in an estimated nonlinear DSGE model. *ECB Working Paper Series* (754).
- An, S. and F. Schorfheide (2007). Bayesian Analysis of DSGE Models. *Econometric Reviews* 26(2-4), 113–172.
- Andrieu, C., A. Doucet, and R. Holenstein (2007). Particle Markov chain Monte Carlo. Unpublished paper: Department of Mathematics, University of Bristol.
- Andrieu, C., A. Doucet, and G. O. Roberts (2007). The expected auxiliary variable method for Monte Carlo simulation. Unpublished paper.
- Aruoba, S. B., J. Fernandez-Villaverde, and J. F. Rubio-Ramirez (2006). Comparing solution methods for dynamic equilibrium economies. *Journal of Economic Dynamics & Control* 30, 2477–2508.
- Chib, S. (2001). Markov chain Monte Carlo methods: computation and inference. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, pp. 3569–3649. Amsterdam: North-Holland.
- Chib, S. and E. Greenberg (1998). Analysis of multivariate Probit models. *Biometrika* 85, 347–61.
- Del-Moral, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. New York: Springer.
- Diggle, P. J. and R. J. Gratton (1984). Monte Carlo methods of inference for implicit statistical models (with discussion). *Journal of the Royal Statistical Society, Series B* 46, 193–227.
- Doucet, A., N. de Freitas, and N. J. Gordon (Eds.) (2001). *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.
- Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

- Fernandez-Villaverde, J. and J. F. Rubio-Ramirez (2007). Estimating macroeconomics models: a likelihood approach. *Review of Economic Studies* 74, 1059–1087.
- Fernandez-Villaverde, J., J. F. Rubio-Ramirez, and M. S. Santos (2006). Convergence properties of the likelihood of computed dynamic models. *Econometrica* 74, 93–119.
- Flury, T. and N. Shephard (2008). Estimating time varying volatility with a smooth particle filter. Unpublished paper.
- Gallant, A. R., H. Hong, and A. Khwaja (2008). Estimating a dynamic oligopolistic game with serially correlated unobserved production costs. Working paper, Fuqua School of Business, Duke University.
- Gallant, A. R. and G. Tauchen (1996). Which moments to match. *Econometric Theory* 12, 657–81.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003). *Bayesian Data Analysis* (2 ed.). London: Chapman & Hall.
- Ghysels, E., A. C. Harvey, and E. Renault (1996). Stochastic volatility. In C. R. Rao and G. S. Maddala (Eds.), *Statistical Methods in Finance*, pp. 119–191. Amsterdam: North-Holland.
- Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE-Proceedings F* 140, 107–113.
- Gourieroux, C. and A. Monfort (1996). *Simulation Based Econometric Methods*. Oxford: Oxford University Press.
- Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics* 8, S85–S118.
- Hajivassiliou, V. and D. McFadden (1998). The method of simulated scores for the estimation of LDV models. *Econometrica* 66, 863–896.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Johannes, M. and N. Polson (2008). Particle filtering. In T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*. Springer. Forthcoming.
- Johannes, M., N. Polson, and J. Stroud (2008). Learning about jumps and stochastic volatility: Filtering stochastic differential equations with jumps. *Review of Financial Studies*. Forthcoming.
- Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–393.
- Klein, P. (2000). Using the generalized Schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics & Control* 24, 1405–1423.
- Lerman, S. and C. Manski (1981). On the use of simulated frequencies to approximate choice probabilities. In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 305–319. Cambridge: M.I.T. Press.
- Mroz, T. A. (1987). The Sensitivity of an Empirical Model of Married Women’s Hours of Work to Economic and Statistical Assumptions. *Econometrica* 55(4), 765–799.
- Pitt, M. K. (2001). Smooth particle filters for likelihood maximisation. Unpublished paper: Department of Economics, Warwick University.
- Pitt, M. K. and N. Shephard (1999). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association* 94, 590–599.
- Schmitt-Grohe, S. and M. Uribe (2004). Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics & Control* 28, 755–775.
- Shephard, N. (Ed.) (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.
- Smith, A. A. (1993). Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics* 8, S63–S84.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.