Efficient econometric inference based on estimated likelihoods

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

September 9, 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. It 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 three 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: Particle filter, MCMC, Metropolis-Hastings, likelihood, inference, state-space models, DSGE, stochastic volatility

1 Introduction

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 log of this estimate is then numerically maximised with respect to the parameters. Based on a sample of size T, for i.i.d. data the theory of this, for example discussed in Gourieroux and Monfort (1996, Ch. 3), needs that M goes to infinity for this maximum simulation likelihood estimator to be consistent and $\sqrt{T}/M \to 0$ to have the same distribution as the maximum likelihood estimator.[1]

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

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

where we assume $y_t|x_t$ are independent over t. Assume we can simulate from $F_t(\psi)$.[2] Write these simulations as $\varepsilon_t^{(1)}, ..., \varepsilon_t^{(M)}$. 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

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

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

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

but the score is biased and it is this bias which drives the fact that the maximum simulated likelihood estimator of $\theta = \left(\beta', \psi'\right)'$ behaves poorly asymptotically unless $M \to \infty$.

Here 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 argument first in the context of dynamic models in a paper in statistical theory by Andrieu, Doucet, and Holenstein (2007, Theorem 5, Section 5.2). The framework we use is vastly simpler and gets immediately to the heart

[1] Alternatives to maximum simulated likelihood include simulated scores, the stochastic EM algorithm, indirect inference, efficient method of moments and Markov chain Monte Carlo. Discussions of these topics can be found in, for example, Hajivassiliou and McFadden (1998), Chib (2001), Gourieroux, Monfort, and Renault (1993), Smith (1993), Gallant and Tauchen (1996) and Gourieroux and Monfort (1996).

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

of the issue. We make our contribution by illustrating these themes on some core econometric problems.

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

$$\mathrm{d}F(\theta|y) \propto \mathrm{d}F(y|\theta)\mathrm{d}F(\theta),$$

where $\mathrm{d}F(\theta)$ is a prior. This typically requires us to calculate the likelihood $\mathrm{d}F(y|\theta)$, but here we assume that all we have is a simulation based estimator

$$\mathrm{d}\widehat{F}_u(y|\theta),$$

which is unbiased

$$\mathrm{E}_u\left\{\mathrm{d}\widehat{F}_u(y|\theta)\right\} = \mathrm{d}F(y|\theta),$$

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

$$\mathrm{d}\widehat{F}_u(y|\theta) = \mathrm{d}G(y, u|\theta),$$

that is $\mathrm{d}G$ is a joint density which, when marginalised over u, delivers $\mathrm{d}F(y|\theta)$.

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

$$\mathrm{d}G(u, \theta|y) \propto \mathrm{d}G(y, u|\theta)\mathrm{d}F(\theta),$$

This simulation based Bayesian method will deliver draws

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

so throwing away the u samples leaves us with

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

which are from $\mathrm{d}F(\theta|y)$. These samples 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 $\mathrm{d}Q\left(\theta^{(i)} \mid \theta^{(i-1)}\right)$, draw the uniformly distributed u and compute[3]

$$\widehat{L}^{(i)} = \mathrm{d}\widehat{F}_u^{(i)}(y|\theta^{(i)})$$

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

The acceptance probability is given by

$$q = \min\left[\frac{\widehat{L}^{(i)}}{\widehat{L}^{(i-1)}}\frac{\mathrm{d}F(\theta^{(i)})}{\mathrm{d}F(\theta^{(i-1)})}\frac{\mathrm{d}Q\left(\theta^{(i-1)}|\theta^{(i)}\right)}{\mathrm{d}Q\left(\theta^{(i)}|\theta^{(i-1)}\right)}, 1\right], \qquad V \sim U(0,1).$$

If $V > q$ set

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

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

An important practical observation is that M does of course play a role in the algorithm: it can influence the rejection rate. If M is very small then $\mathrm{d}\widehat{F}$ will be a very jittery estimator of $\mathrm{d}F$ 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 $\theta^{(i)}$.

The above argument opens up the possibility of carrying out likelihood inference for wide classes of models in economics. All we need is an unbiased estimator of the likelihood. It turns out that for dynamic models particle filters deliver such an estimator for rather general state space models. This is the topic of Andrieu, Doucet, and Holenstein (2007, Theorem 5, Section 5.2). The common theme of the dynamics of these models is that they can be simulated, which is all we need to perform inference.

In this paper we will detail three application areas of these methods: individual choice, stochastic volatility and dynamic stochastic general equilibrium (DSGE) models. Hence the paper covers problems in microeconomics, financial econometrics and macroeconomics. The method we outline below can also be applied in other contexts, for example in models whose dynamics are non-stochastic (e.g. chaotic models or weather models driven by differential equations).

This paper is structured as follows: In section 2 we give an example of a static model. Section 3 treats dynamic models and introduces particle filters as a convenient tool to obtain unbiased likelihood estimates. In section 4 we provide some examples from macroeconomics and finance to demonstrate the performance of this algorithm in dynamic models.

2 Static models

Here we illustrate the workings of the algorithm in static models by the example of a binary choice model. We use the classical data set from Mroz (1987) to study the labour force participation of $T = 753$ women. We posit a simple binary choice model and perform inference on its parameters by using an unbiased estimate of the likelihood inside an MCMC algorithm.

The binary variable y_t 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

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

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

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

where

$$\beta = (\beta_0, \ldots, \beta_7)'$$

Here we choose the normal distribution for $F\left(x_t'\beta \mid \psi\right)$. For the simulation based estimator of p_t we draw

$$\varepsilon_t^{(j)} \sim i.i.N\left(0, \sigma_\varepsilon^2\right) \qquad j = 1, \ldots, M$$

and compute

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

The estimate of the likelihood is given by

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

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

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

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

In Probit models one usually sets $\sigma_\varepsilon = 1$ and in Logit models the variance is given by $\frac{\pi^2}{3}$. 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 $(y_t = 1, \widehat{p}_t = 0)$ which results in $\mathrm{d}\widehat{F}(y|\beta, \psi) = 0$. We suggest setting $\sigma_\varepsilon = 1$

by default and if this causes problems to tune it such that we just avoid this undesirable outcome. We assume a Gaussian prior given by $\beta \sim N(\beta_0, I_8)$ where

$$\beta_0 = (0.5855, -0.0034, 0.0380, 0.0395, -0.0006, -0.0161, -0.2618, 0.0130)'$$

We are using the following random walk proposals for the parameters, each applied one at a time:

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

where $\nu_{j,i} \sim i.i.N(0,1)$ for $j = 1, \ldots, 8$ and $i = 1, \ldots, N$. The variances in the random walk proposals were chosen to aim for a 40% acceptance probability for each parameter, see e.g. Gelman, Carlin, Stern, and Rubin (2003). We loop through the parameters to make a proposal for each one individually and accept or reject it. Given the Gaussian errors we assumed, we can easily compute the true likelihood and thus use this exact likelihood model as a benchmark. The proposal variances were tuned on this 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 $M = 1000, 2000, 4000$ and $N = 100000$.

Tables 1 and 2 show the following statistics for the MCMC algorithm: The arithmetical mean, the Monte Carlo standard error, the acceptance probability, and the inefficiency.[4] We discard the first half of the sample so that all statistics are based on the second half. The acceptance

	exact likelihood				$M = 1000$			
	mean	MC s.e.	P(accept)	inefficiency	mean	MC s.e.	P(accept)	inefficiency
β_0	0.295	0.033	0.418	364	0.313	0.039	0.283	593
β_1	-0.012	0.000	0.409	31	-0.012	0.000	0.277	57
β_2	0.130	0.001	0.413	143	0.130	0.002	0.274	328
β_3	0.124	0.001	0.406	149	0.125	0.001	0.272	149
β_4	-0.002	0.000	0.413	130	-0.002	0.000	0.276	136
β_5	-0.053	0.001	0.414	299	-0.054	0.001	0.278	479
β_6	-0.868	0.004	0.427	70	-0.871	0.004	0.286	100
β_7	0.035	0.001	0.411	81	0.034	0.002	0.277	125

Table 1: Results from MCMC for labour force participation; exact likelihood and $M = 1000$ estimated likelihood model with $N = 100000$.

probabilities increase with M, so that with $M = 4000$ 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

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

	$M = 2000$				$M = 4000$			
	mean	MC s.e.	P(accept)	inefficiency	mean	MC s.e.	P(accept)	inefficiency
β_0	0.349	0.035	0.333	462	0.376	0.030	0.365	420
β_1	-0.012	0.000	0.332	35	-0.012	0.000	0.361	42
β_2	0.130	0.001	0.330	209	0.126	0.001	0.361	168
β_3	0.123	0.001	0.321	160	0.124	0.001	0.355	197
β_4	-0.002	0.000	0.332	162	-0.002	0.000	0.362	189
β_5	-0.054	0.001	0.334	458	-0.054	0.001	0.366	321
β_6	-0.876	0.004	0.338	128	-0.871	0.004	0.374	67
β_7	0.032	0.002	0.328	94	0.032	0.001	0.360	77

Table 2: Results from MCMC for labour force participation; $N = 100000$, $M = 2000$ and $M = 4000$.

models are not significantly different from the exact likelihood model. Our results are very close to those from an ordinary MC estimated Probit regression. All parameter estimates have the expected sign, except β_7.

Tables 3 and 4 show the 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 $M = 4000$ estimated likelihood model. Since the θ_i are highly correlated we use the

	covariance and correlation							
β_0	55.401	0.367	-0.689	-0.196	0.285	-0.884	-0.723	-0.902
β_1	0.074	0.001	-0.558	0.116	-0.049	-0.277	-0.084	-0.216
β_2	-1.476	-0.004	0.083	0.017	-0.036	0.325	0.073	0.485
β_3	-0.318	0.001	0.001	0.047	-0.986	0.021	0.005	0.171
β_4	0.014	0.000	0.000	-0.001	0.000	-0.144	-0.112	-0.233
β_5	-0.786	-0.001	0.011	0.001	0.000	0.014	0.923	0.853
β_6	-4.480	-0.002	0.017	0.001	-0.001	0.092	0.693	0.773
β_7	-2.123	-0.002	0.044	0.012	0.000	0.032	0.203	0.100

Table 3: Results from MCMC for labour force participation; exact likelihood model with $N = 100000$; covariance (lower triangle) and correlation (upper triangle) matrix.

	covariance and correlation							
β_0	43.974	0.182	-0.494	-0.251	0.326	-0.845	-0.706	-0.867
β_1	0.036	0.001	-0.510	0.135	-0.066	-0.081	0.150	0.023
β_2	-0.907	-0.004	0.077	0.037	-0.030	0.023	-0.207	0.174
β_3	-0.396	0.001	0.002	0.057	-0.988	0.021	0.053	0.205
β_4	0.016	0.000	0.000	-0.002	0.000	-0.133	-0.142	-0.264
β_5	-0.670	0.000	0.001	0.001	0.000	0.014	0.932	0.850
β_6	-4.092	0.004	-0.050	0.011	-0.001	0.097	0.765	0.793
β_7	-1.724	0.000	0.014	0.015	-0.001	0.030	0.208	0.090

Table 4: Results from MCMC for labour force participation; $N = 100000$, $M = 4000$; covariance (lower triangle) and correlation (upper triangle) matrix.

Newey-West Heteroskedasticity and Autocorrelation Consistent (HAC) estimator of the variance-covariance matrix. 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 $N =$

100000 between the exact likelihood model, and the ones with estimated likelihoods with $M = 1000$, $M = 2000$ and $M = 4000$. Figures 2 and 3 compare the parameter histograms. We note that the performance improves as M increases.

3 Dynamic models

We now move on to treat parameter estimation in dynamic 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.

3.1 Assumptions

We assume we have some observations

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

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 model:*

1. *We can compute the measurement density*

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

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

2. *We can simulate from the random variable*

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

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

3. *We can compute the prior $\mathrm{d}F(\theta)$.*

At no point do we assume we can compute $\mathrm{d}F(\alpha_{t+1}|\alpha_t, \mathcal{F}_t, \theta)$. 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 Section 4 that a large number of intractable econometric models are of this form. Leading examples are DSGE models, some (continuous time) stochastic volatility models and some models in industrial organisation.

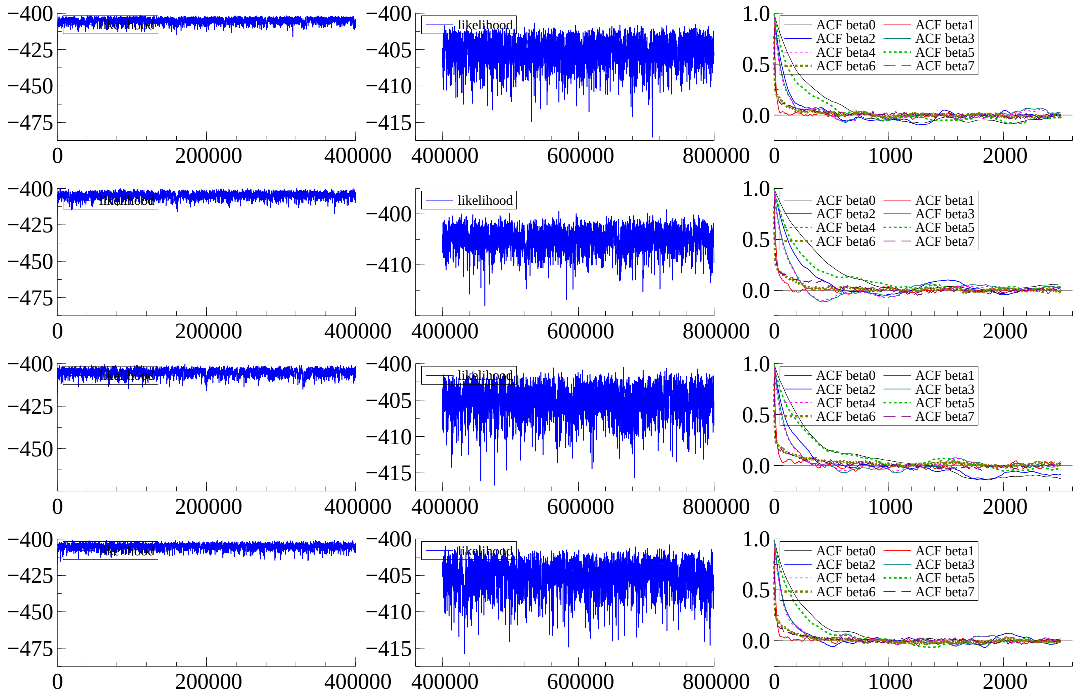

Figure 1: Labour force participation model; Likelihoods and ACFs of parameters. Fist column: $\widehat{L}\left(\theta^{(i)}\right)$ for $i = 1, \ldots, \frac{N}{2}$; second column: $\widehat{L}\left(\theta^{(i)}\right)$ for $i = \frac{N}{2}, \ldots, N$; third column: ACF of $\theta^{(i)}$. First row: exact likelihood model; second row: estimated likelihood with $M = 1000$; third row: $M = 2000$; fourth row: $M = 4000$.

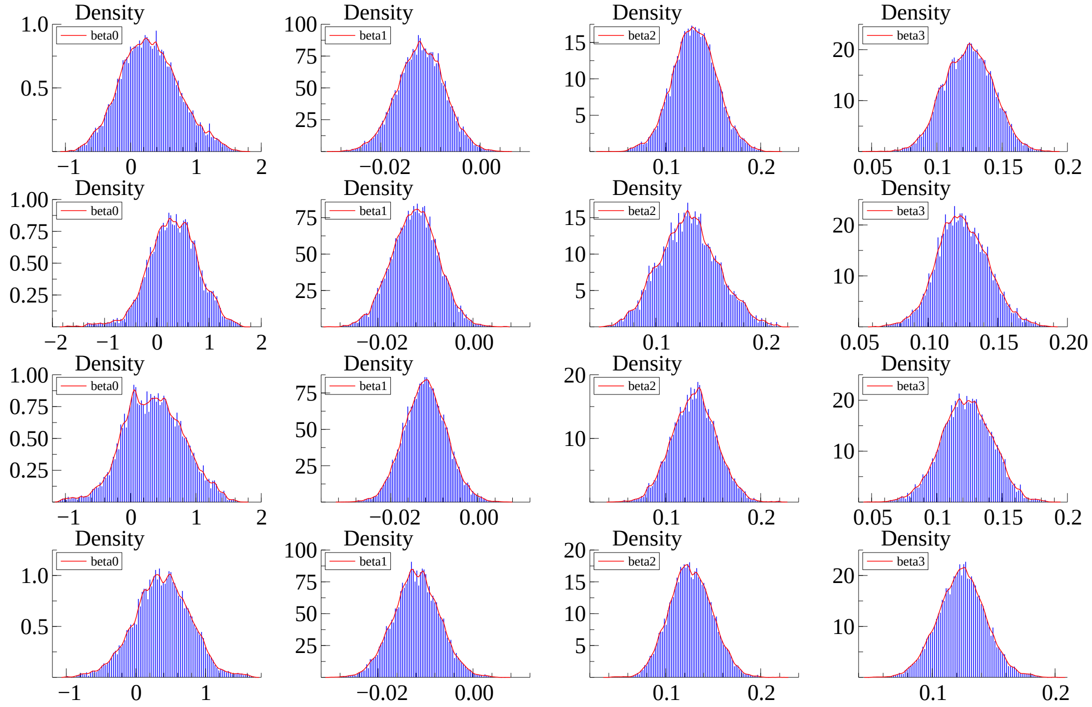

Figure 2: Labour force participation model histogram of parameters for $i = 50000, \ldots, 100000$. First row: exact likelihood model; second row: estimated likelihood with $M = 1000$; third row: $M = 2000$; fourth row: $M = 4000$.

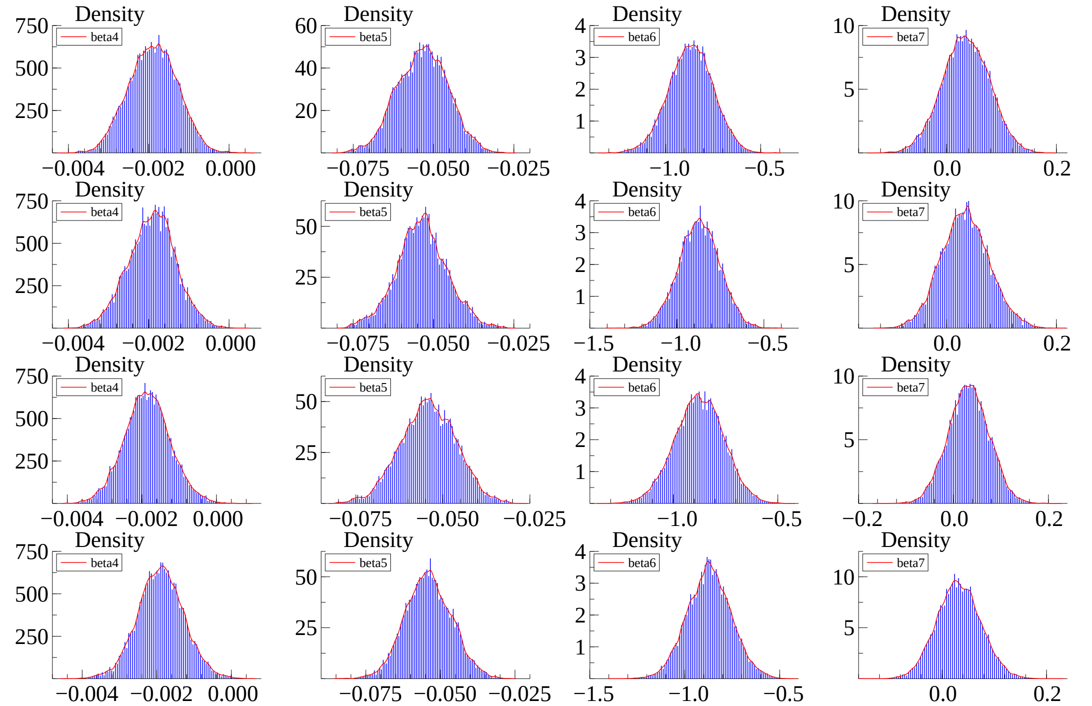

Figure 3: Labour force participation model histogram of parameters for $i = 50000, \ldots, 100000$. First row: exact likelihood model; second row: estimated likelihood with $M = 1000$; third row: $M = 2000$; fourth row: $M = 4000$.

3.2 Predictive decomposition

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

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

It 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 $\mathrm{d}F(y_t|\mathcal{F}_{t-1},\theta)$ 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 $\mathrm{d}F(y_t|\mathcal{F}_{t-1},\theta)$. This will be carried out using a particle filter, whose recursive structure will allow us to calculate an unbiased estimator of $\mathrm{d}F(y|\mathcal{F}_0,\theta)$. This can then be used as the basis for inference using an MCMC algorithm analogous to the above strategy.

3.3 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 $\mathrm{d}F(y|\theta)$ as a by-product in order to do model comparison via marginal likelihoods. They have recently received some attention in macroeconomics due to the work of, for example, Fernanzed-Villaverde and Rudio-Ramirez (2007) and Aruoba, Fernandez-Villaverde, and Rubio-Ramirez (2006).

Here we give a very simple particle filter, which can be generically coded and just needs the ability to evaluate $\mathrm{d}F(y_t|\alpha_t,\mathcal{F}_{t-1},\theta)$ and to simulate from $\alpha_{t+1}|\alpha_t,\mathcal{F}_t,\theta$.

Particle filter

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

2. Compute the weights

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

   and the normalised weights so that

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

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

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

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

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

4. Sample

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

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

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

$$\exp\left(l_T(\theta)\right) \overset{a.s.}{\to} \mathrm{d}F(y|\theta)$$

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

$$\mathrm{E}\left[\exp(l_T(\theta))\right] = \mathrm{d}F(y|\theta).$$

A proof of this unbiasedness is provided in the Appendix[5].

In order to carry out MCMC we proceed by replacing $\mathrm{d}F(y|\theta)$ by its particle filter estimate

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

which we are allowed to do as the particle filter indeed provides an unbiased estimate of $\mathrm{d}F(y \mid \theta)$. The resulting algorithm is

MCMC Algorithm with Estimated Likelihood

1. Initialise $\theta^{(0)}$, $i = 1$.

[5] An 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 Fernanzed-Villaverde, Rudio-Ramirez, and Santos (2006) and Fernanzed-Villaverde and Rudio-Ramirez (2007). To our knowledge only Flury and Shephard (2008) offer a smooth particle filter which circumvents this problem.

2. Propose

$$\theta^{(i)} \sim \mathrm{d}Q\left(\theta|\theta^{(i-1)}\right)$$

where we can evaluate $\mathrm{d}Q$. Obtain $L^{(i)} = \widehat{L}(\theta^{(i)})$.

3. Draw $V \sim U(0,1)$ and if

$$V > \min\left[\frac{L^{(i)}}{L^{(i-1)}}\frac{\mathrm{d}F(\theta^{(i)})}{\mathrm{d}F(\theta^{(i-1)})}\frac{\mathrm{d}Q\left(\theta^{(i-1)}|\theta^{(i)}\right)}{\mathrm{d}Q\left(\theta^{(i)}|\theta^{(i-1)}\right)}, 1\right], \tag{2}$$

then write $\theta^{(i)} = \theta^{(i-1)}$, or else retain the proposed $\theta^{(i)}$.

4. Set $i = i + 1$ and go to 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.

4 Examples

This section aims at illustrating the performance of the particle filter in MCMC algorithm. We start off by analyzing a simple linear Gaussian model where an analytical expression for $\mathrm{d}F(y \mid \theta)$ is readily available from the Kalman filter in order to be able to evaluate the new algorithm. We then move on to a discrete time Gaussian SV model and finally conclude with the analysis of a DSGE model.

4.1 Gaussian linear model

It is useful to start the analysis with a model where analytic solutions are available. We consider the Gaussian, linear model (see e.g. Harvey (1989) and Durbin and Koopman (2001))

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

where $\alpha_0 \sim N\left(0, \sigma_\eta^2/\left(1-\phi^2\right)\right)$. 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 $\theta = (\mu, \log\sigma_\epsilon, \phi, \log\sigma_\eta)'$, 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 parametrization $\theta^* = (0.5, \log 1, 0.825, \log 0.75)'$. 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:

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

where $\nu_{j,i} \sim i.i.N(0,1)$ for $j = 1, \ldots, 4$ and $i = 1, \ldots, N$. 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 presenting the results obtained from using the Kalman filter. Figure 5 shows the Markov chain for the parameters where $N = 100000$ and the histograms, which are based on the second half of the chains. Figure 4 depicts the likelihood and the autocorrelation functions (ACFs) of the parameters for $N = 100000$. We do not report it here but for the Kalman filter the Markov chain seems to have stabilized already with only $N = 10000$ iterations.

Figures 6 and 7 depict the same for the particle filter run at $M = 100$ particles. Using only $M = 100$ particles appears to be insufficient for the lengths 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 $N = 1000000$ and figures 8 and 9 show the same for 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 $M = 1000$. Figures 10 and 11 show the drastic improvement in the performance of the algorithm from this increase.

Table 5 shows the following statistics for the MCMC algorithm with the Kalman Filter: The arithmetical mean, the Monte Carlo standard error, the acceptance probability, the covariance (lower triangle) and correlation (upper triangle) matrix, and the inefficiency.

Tables 6 and 8 show the MCMC statistics for $M = 100$ and $M = 1000$ particles, and table 7 shows the statistics for the $N = 1000000$ chain with $M = 100$. We see how the acceptance probabilities improve with M. We do not report it here for brevity, but with $M = 2000$ we achieve an acceptance probability of 30%. To see that effect more clearly figure 12 compares the evolution

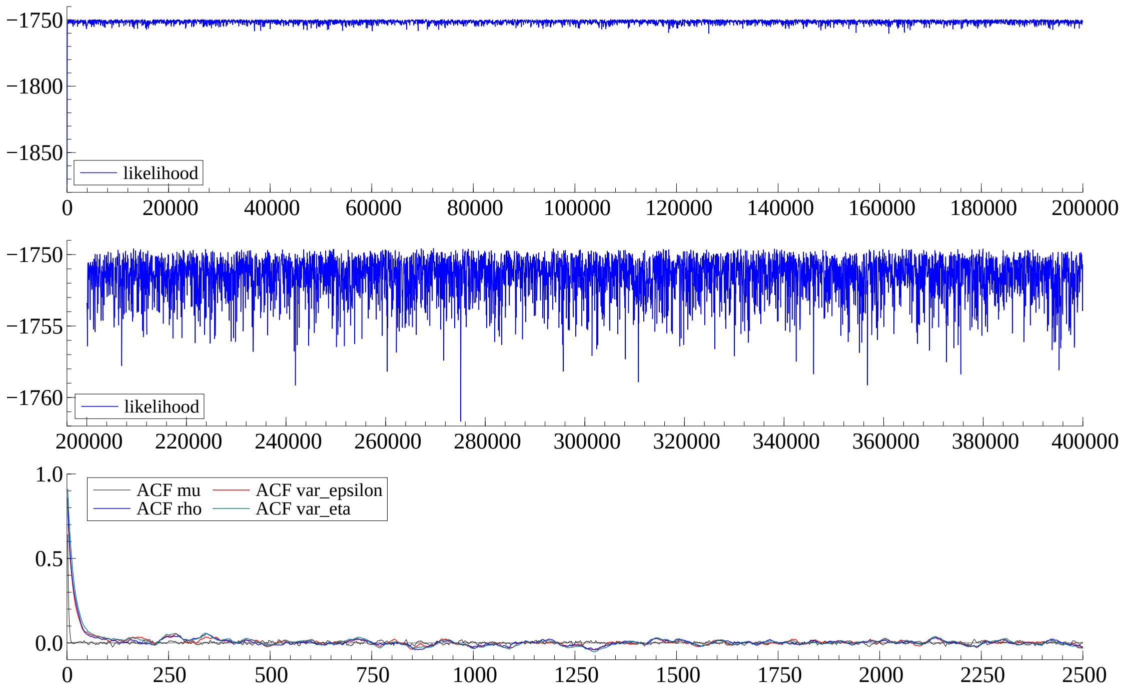

Figure 4: Random walk with noise model; Kalman filter: Likelihood and ACFs of parameters. Top: $\widehat{L}\left(\theta^{(i)}\right)$ for $i = 1, \ldots, \frac{N}{2}$; middle: $\widehat{L}\left(\theta^{(i)}\right)$ for $i = \frac{N}{2}, \ldots, N$; bottom: ACF of $\theta^{(i)}$.

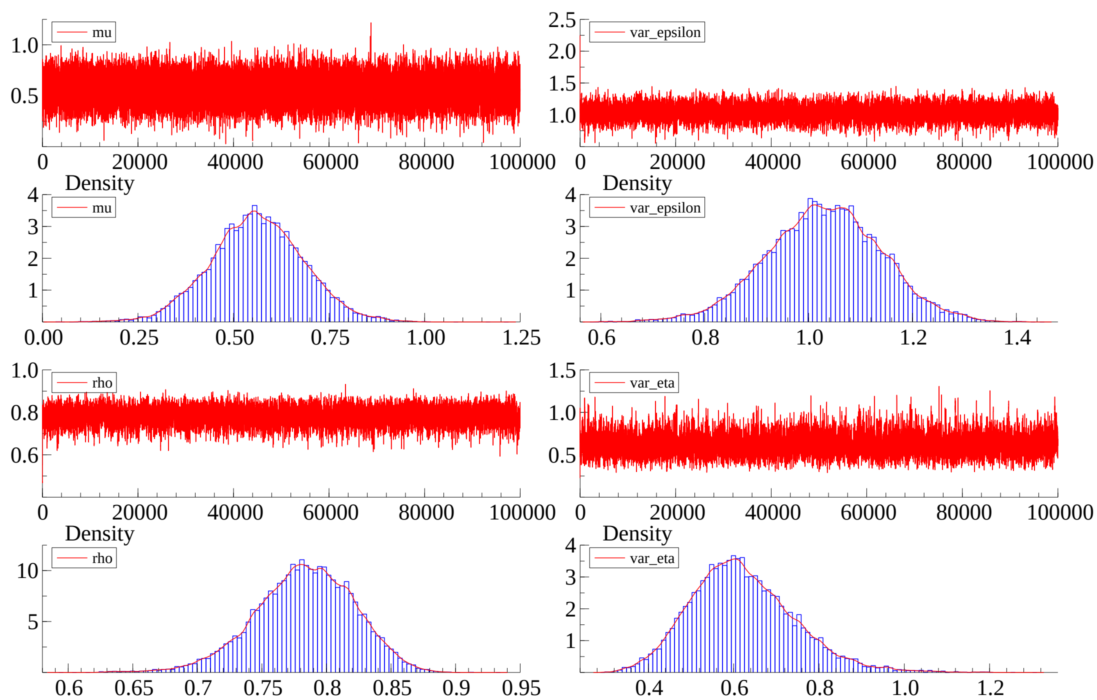

Figure 5: Random walk with noise model; Kalman filter: $\theta^{(i)}$ for $i = 1, \ldots, 100000$ and histogram of parameters for $i = 50000, \ldots, 100000$.

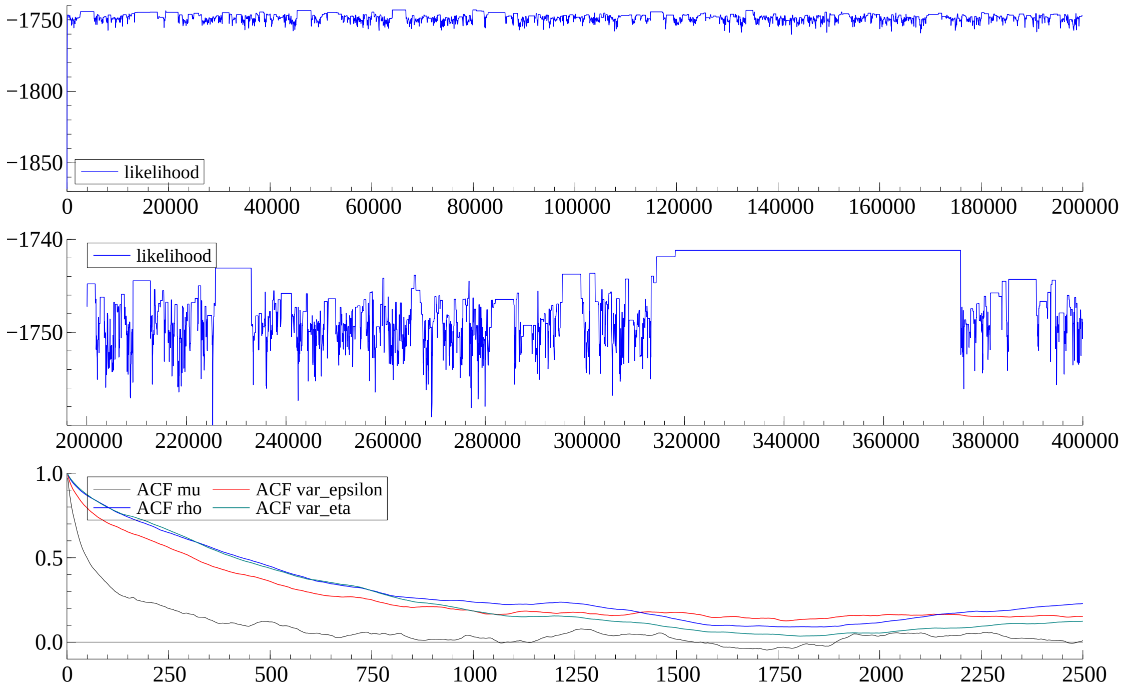

Figure 6: Random walk with noise model; particle filter with $M = 100$: Likelihood and ACFs of parameters. Top: $\widehat{L}\left(\theta^{(i)}\right)$ for $i = 1, \ldots, \frac{N}{2}$; middle: $\widehat{L}\left(\theta^{(i)}\right)$ for $i = \frac{N}{2}, \ldots, N$; bottom: ACF of $\theta^{(i)}$.

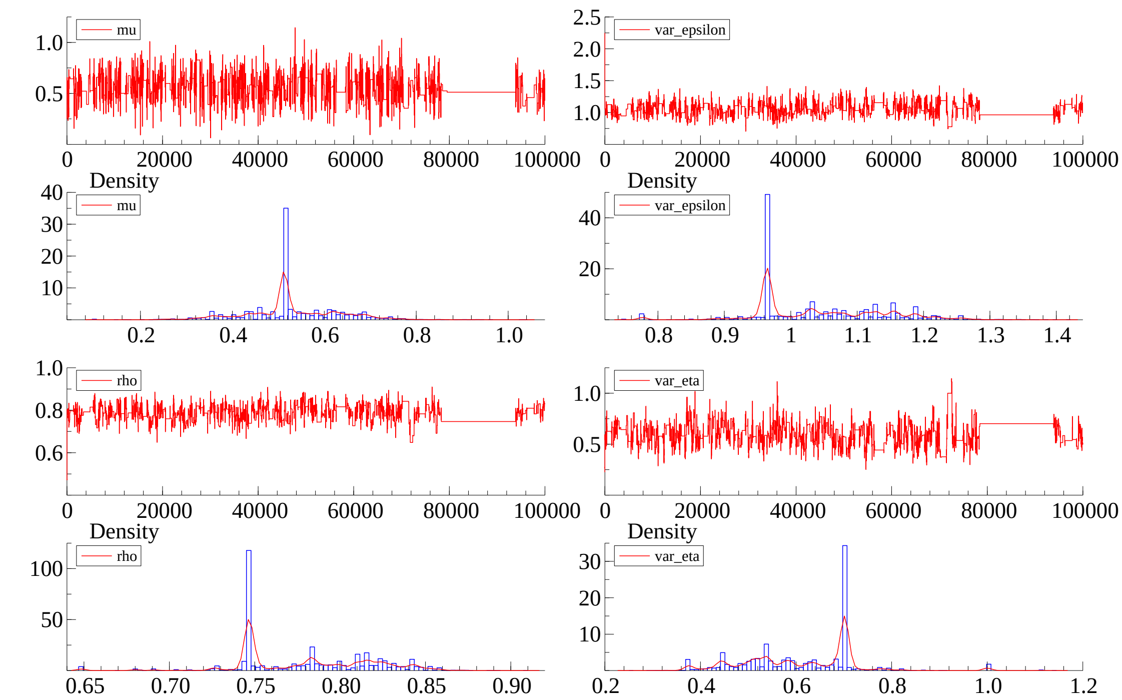

Figure 7: Random walk with noise model; particle filter with $M = 100$: $\theta^{(i)}$ for $i = 1, \ldots, 100000$ and histogram of parameters for $i = 50000, \ldots, 100000$.

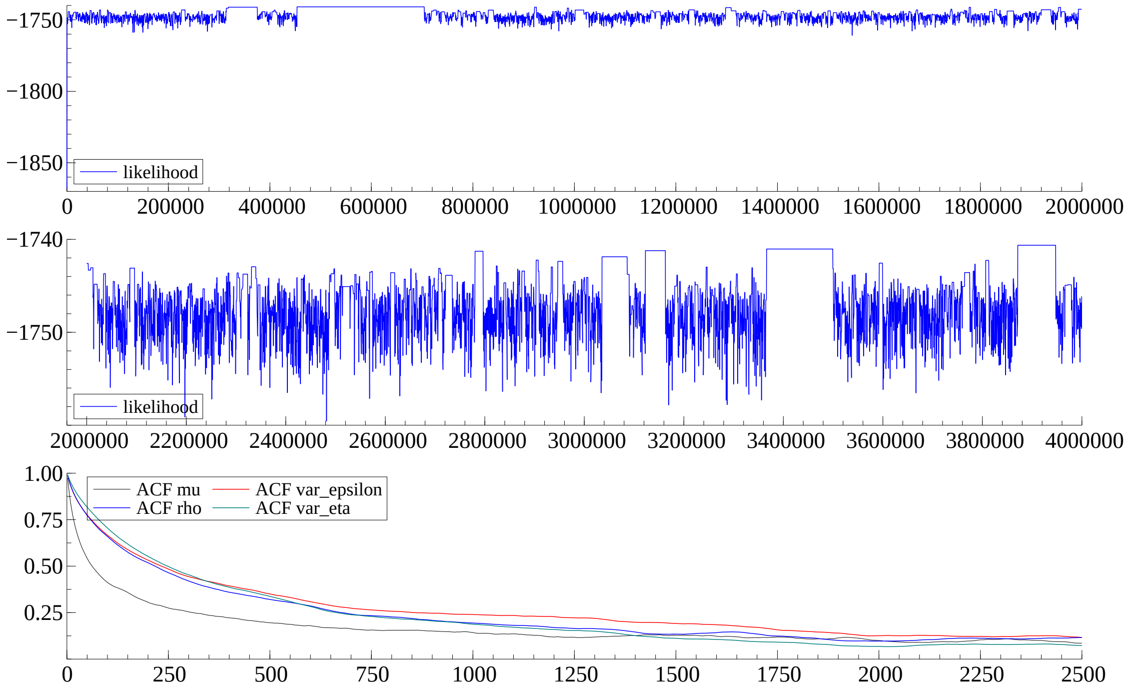

Figure 8: Random walk with noise model; particle filter with $M = 100$: Likelihood and ACFs of parameters. Top: $\widehat{L}\left(\theta^{(i)}\right)$ for $i = 1, \dots, \frac{N}{2}$; middle: $\widehat{L}\left(\theta^{(i)}\right)$ for $i = \frac{N}{2}, \dots, N$; bottom: ACF of $\theta^{(i)}$.

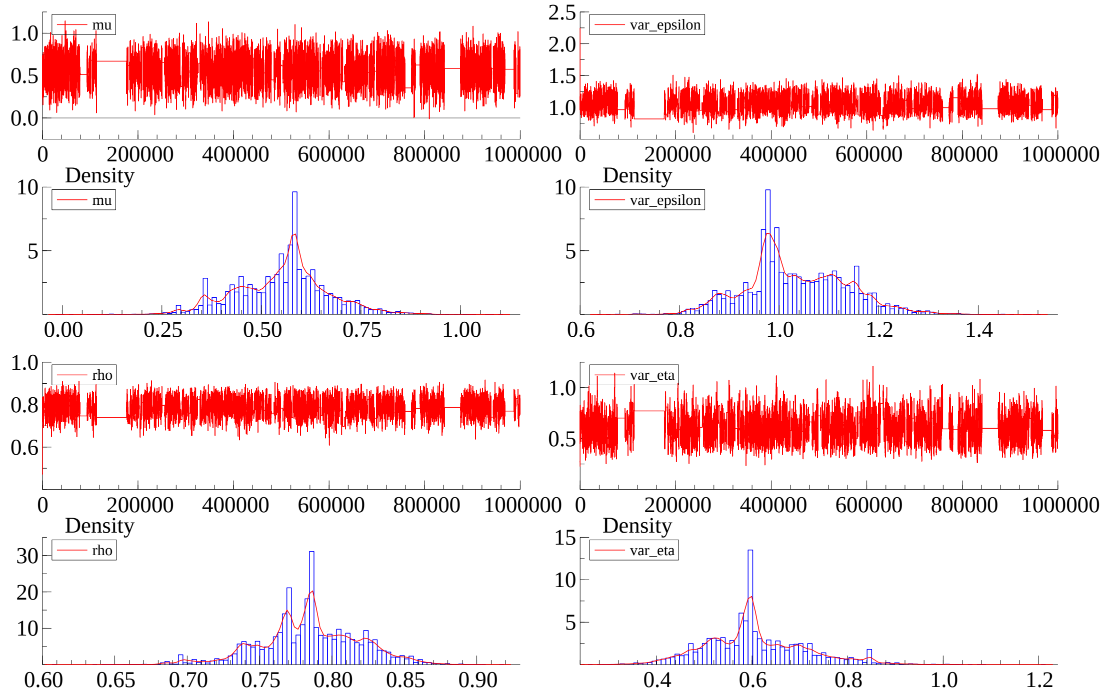

Figure 9: Random walk with noise model; particle filter with $M = 100$: $\theta^{(i)}$ for $i = 1, \dots, 1000000$ and histogram of parameters for $i = 500000, \dots, 1000000$.

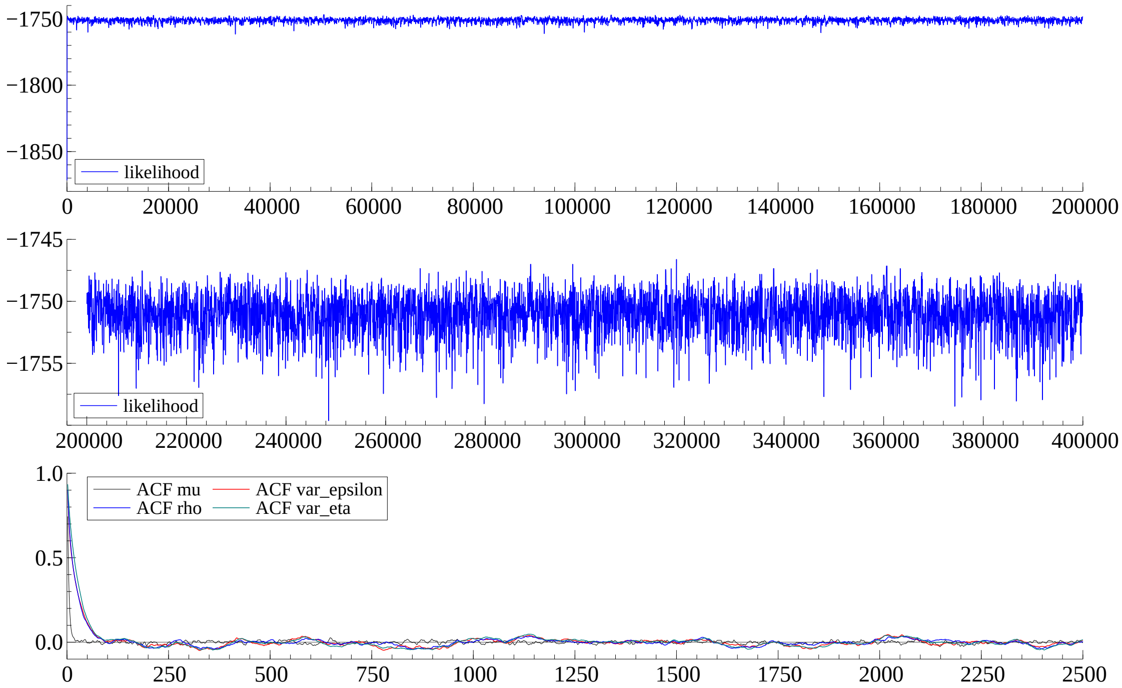

Figure 10: Random walk with noise model; particle filter with $M = 1000$: Likelihood and ACFs of parameters. Top $\widehat{L}\left(\theta^{(i)}\right)$ for $i = 1, \ldots, \frac{N}{2}$; middle: $\widehat{L}\left(\theta^{(i)}\right)$ for $i = \frac{N}{2}, \ldots, N$; bottom: ACF of $\theta^{(i)}$.

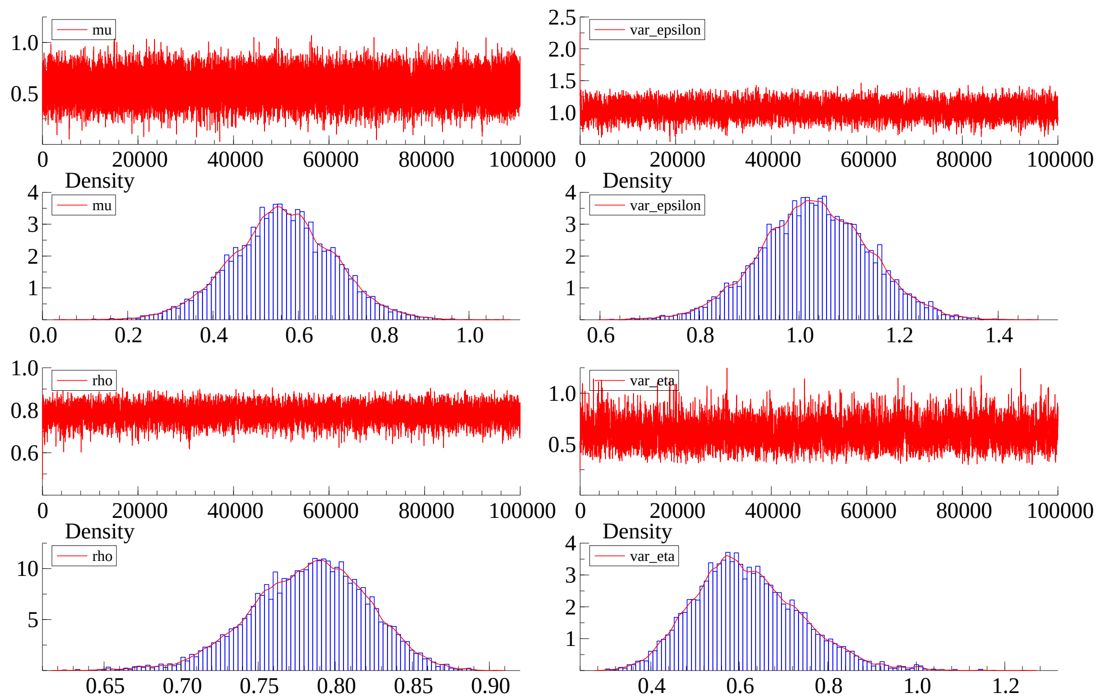

Figure 11: Random walk with noise model; particle filter with $M = 1000$: $\theta^{(i)}$ for $i = 1, \ldots, 100000$ and histogram of parameters for $i = 50000, \ldots, 100000$.

of the ACFs and likelihoods as we increase $M = 100, 500, 1000, 2000$ with the Kalman filter output, for $N = 100000$. Figure 13 compares the corresponding histograms.

	mean	MC s.e.	P(accept)	covariance and correlation				inefficiency
μ	0.562	0.001	0.398	0.071	-0.092	-0.069	0.103	7.2
σ_ϵ	1.030	0.003	0.399	-0.016	0.416	0.958	-0.983	39
ρ	0.783	0.001	0.391	-0.004	0.136	0.049	-0.978	37
σ_η	0.621	0.004	0.400	0.023	-0.498	-0.169	0.616	48

Table 5: Results from MCMC with Kalman filter; $N = 100000$.

	mean	MC s.e.	P(accept)	covariance and correlation				inefficiency
μ	0.531	0.006	0.021	1.816	-0.052	-0.126	0.154	288
σ_ϵ	1.041	0.008	0.021	-0.127	3.258	0.896	-0.942	852
ρ	0.780	0.003	0.020	-0.130	1.238	0.586	-0.944	981
σ_η	0.611	0.010	0.022	0.478	-3.916	-1.664	5.305	919

Table 6: Results from MCMC with particle filter; $N = 100000$, $M = 100$.

	mean	MC s.e.	P(accept)	covariance and correlation				inefficiency
μ	0.553	0.002	0.018	2.615	0.016	0.015	0.002	532
σ_ϵ	1.033	0.003	0.018	0.045	3.134	0.722	-0.790	855
ρ	0.785	0.001	0.018	0.014	0.752	0.346	-0.838	776
σ_η	0.607	0.003	0.019	0.007	-2.622	-0.923	3.512	785

Table 7: Results from MCMC with particle filter; $N = 1000000$, $M = 100$.

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 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.

4.2 Discrete time Gaussian SV model

We now turn to a simple real life problem and estimate the the Gaussian discrete time stochastic volatility model (e.g. Ghysels, Harvey, and Renault (1996) and Kim, Shephard, and Chib (1998)).

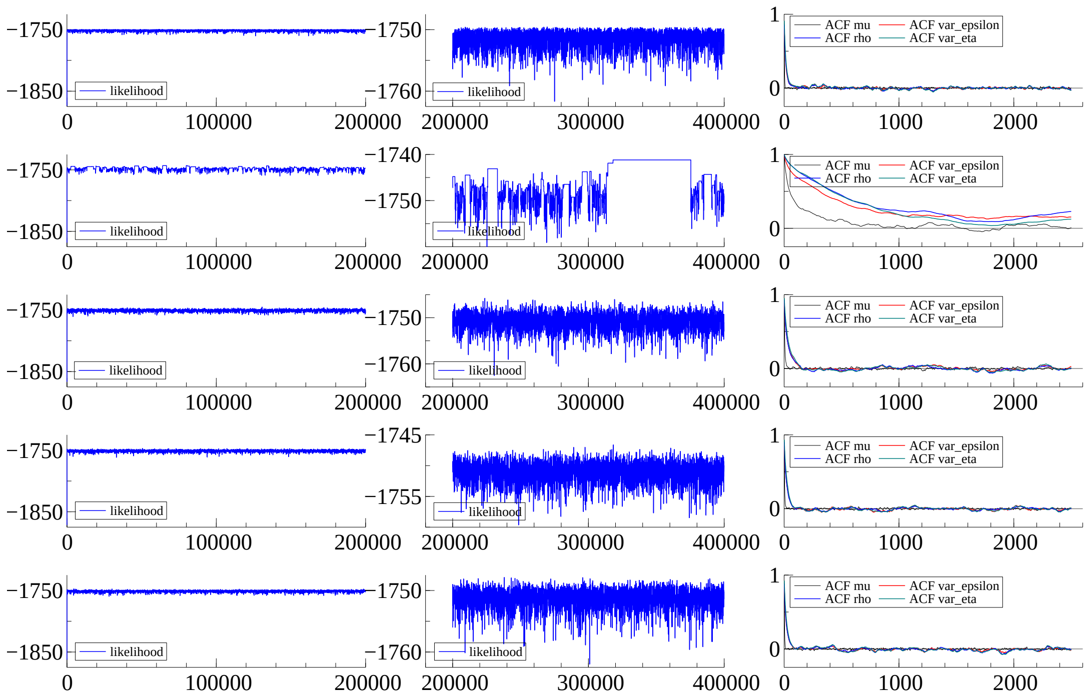

Figure 12: Random walk with noise model; $N = 100000$; first row: likelihoods and ACFs from Kalman filter, second row: particle filter with $M = 100$, third row: $M = 500$, fourth row: $M = 1000$; last row: $M = 2000$.

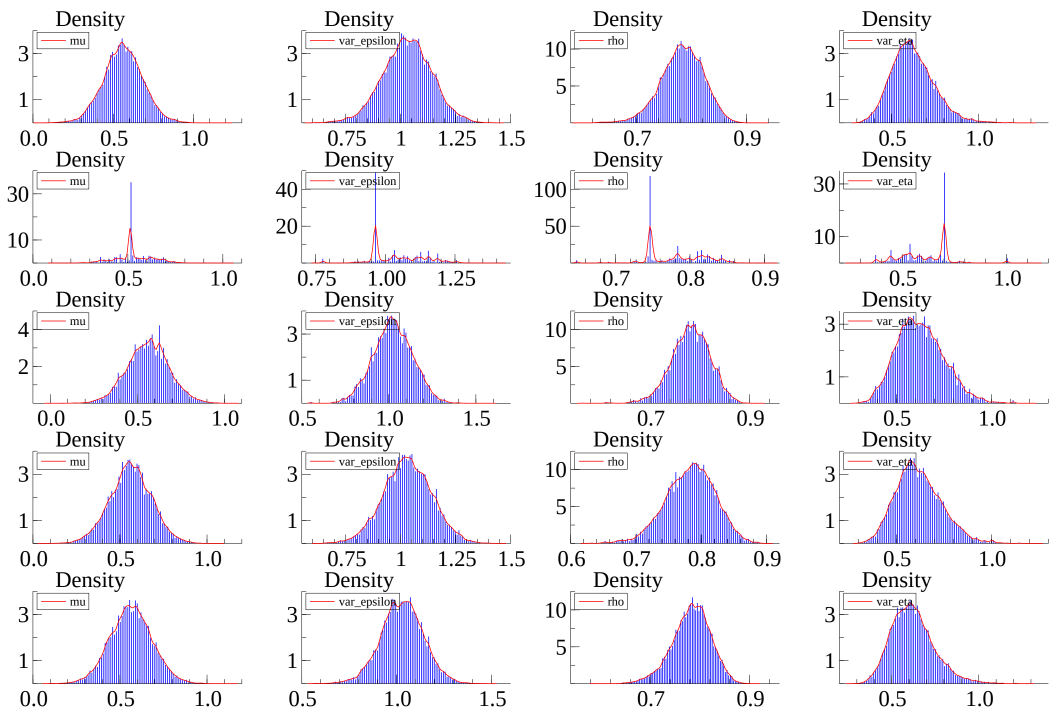

Figure 13: Random walk with noise model; $N = 100000$; first row: parameter histograms from Kalman filter, second row: particle filter with $M = 100$, third row: $M = 500$, fourth row: $M = 1000$; last row: $M = 2000$.

	mean	MC s.e.	P(accept)	covariance and correlation				inefficiency
μ	0.562	0.001	0.252	0.097	-0.085	-0.100	0.071	7.8
σ_ϵ	1.031	0.003	0.250	-0.017	0.414	0.914	-0.968	29
ρ	0.783	0.001	0.245	-0.007	0.130	0.049	-0.959	30
σ_η	0.618	0.004	0.256	0.017	-0.485	-0.165	0.606	36

Table 8: Results from MCMC with particle filter; $N = 100000$, $M = 1000$.

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

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

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 review in Shephard (2005). We assume a Gaussian prior given by $\theta \sim N(\theta_0, I_5)$ where θ_0 is

μ	β_0	β_1	ϕ	ρ
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

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

where $\nu_{j,i} \sim i.i.N(0,1)$ for $j = 1, \ldots, 5$ and $i = 1, \ldots, N$. 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 returns are defined as

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

	mean	MC s.e.	P(accept)	covariance and correlation					inefficiency
μ	0.042	0.000	0.411	0.003	-0.895	-0.340	0.511	0.310	15
β_0	-0.141	0.001	0.395	-0.015	0.091	0.217	-0.402	-0.184	12
β_1	0.080	0.000	0.398	-0.001	0.002	0.001	-0.926	-0.015	18
ϕ	0.982	0.000	0.424	0.001	-0.002	-0.001	0.000	0.010	16
ρ	-0.742	0.000	0.427	0.002	-0.005	0.0000	0.0000	0.010	6.4

Table 9: Results from MCMC with particle filter; $N = 100000$, $M = 2000$.

The statistics are displayed in table 9. The data together with the time-series for α_t and volatility $\exp\{\beta_0 + \beta_1 \alpha_t\}$, estimated with the posterior means of the parameters is plotted in Figure 14. Figures 15 and 16 depict likelihoods, ACFs, parameters and their histograms.

4.3 A DSGE model

We now estimate a simple DSGE model. An and Schorfheide (2007) were the first to consider Bayesian inference for DSGE models.

Fernanzed-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 likelihood estimates and indicate the possibility of using this present algorithm. As mentioned earlier this kind of maximum simulated likelihood approach 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 simple version of a DSGE model. There is a representative household maximizing its lifetime utility, given by

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

where C_t is consumption, L_t 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

$$C_t + I_t = Y_t$$

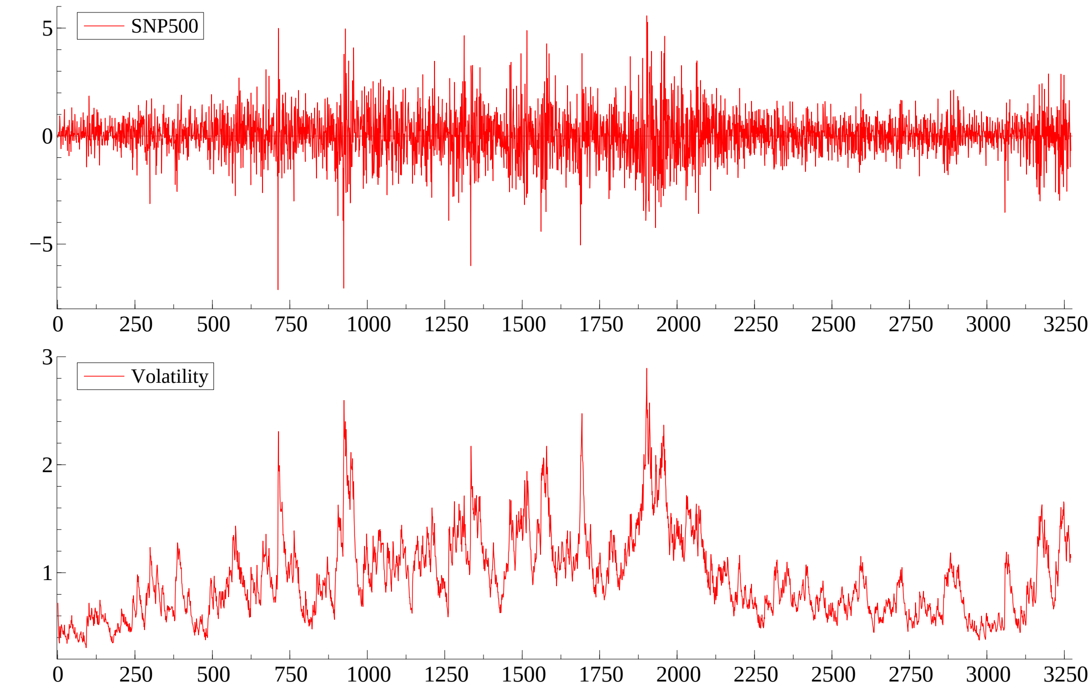

Figure 14: Stochastic volatility model; Top: log-returns on end-of-day level of the SNP500 Composite Index (NYSE/AMEX only) from 03.01.1995 until 31.12.2007; bottom: $\exp\{\beta_0 + \beta_1 \alpha_t\}$ based on posterior means of the parameters.

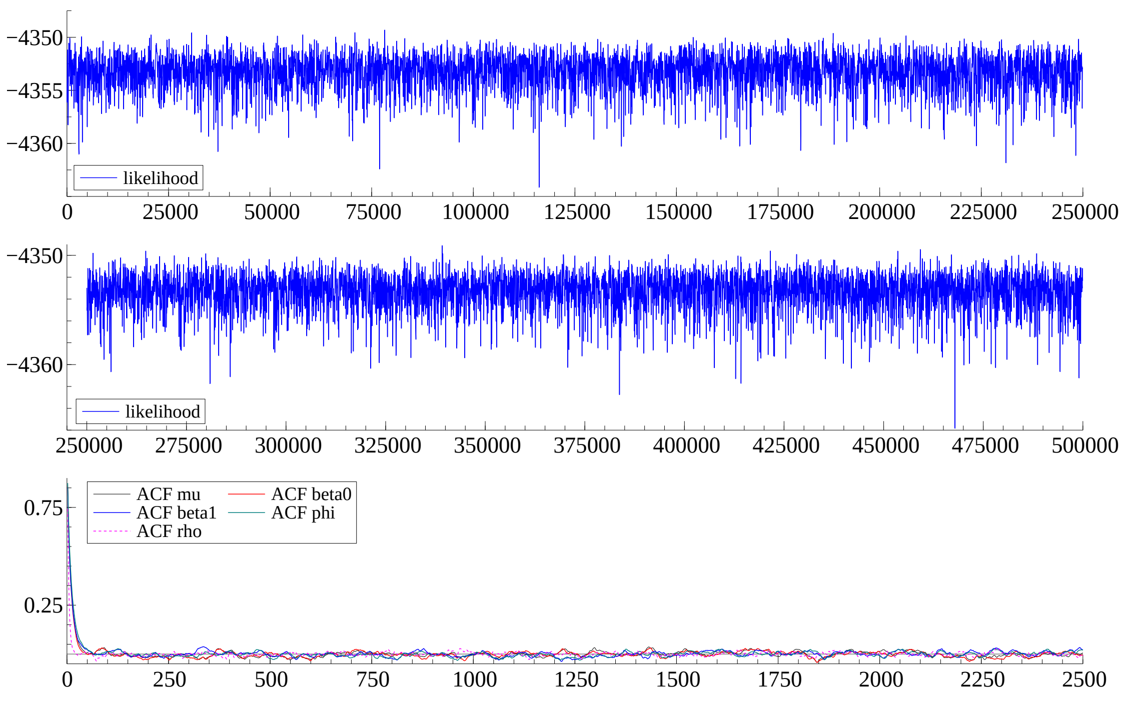

Figure 15: Stochastic volatility model; particle filter with $M = 2000$: Likelihoods and ACFs of parameters. Top: $\widehat{L}\left(\theta^{(i)}\right)$ for $i = 1, \ldots, \frac{N}{2}$; middle: $\widehat{L}\left(\theta^{(i)}\right)$ for $i = \frac{N}{2}, \ldots, N$; bottom: ACF of $\theta^{(i)}$.

We assume the following laws of motion for the technologies

$$\log A_t = \rho_a \log A_{t-1} + \sigma \eta_{a,t}$$
$$\log U_t = \rho_u \log U_{t-1} + \sigma_u \eta_{u,t}$$

where $\eta_{a,t}, \eta_{u,t} \sim i.i.N(0,1)$. In this economy the central planner and the competitive equilibrium coincide. We decide to solve the central planner's problem

$$\max_{\{K_{t+1},L_t\}_{t=0}^{\infty}} \left\{ E_0 \left[\sum_{t=0}^{\infty} \beta^t \left(\log \left\{ A_t K_t^{\alpha} L_t^{1-\alpha} + \frac{1}{U_t} \left((1-\delta) K_t - K_{t+1} \right) \right\} + \psi \log \{1 - L_t\} \right) \right] \right\}$$

The first order equilibrium conditions

$$\frac{1}{C_t} = U_t \beta E_t \left[\frac{1}{C_{t+1}} \left(\alpha A_{t+1} K_{t+1}^{\alpha-1} L_{t+1}^{1-\alpha} + \frac{1}{U_{t+1}} (1-\delta) \right) \right]$$
$$\psi \frac{1}{1-L_t} = \frac{1}{C_t} (1-\alpha) A_t K_t^{\alpha} L_t^{-\alpha}$$

together with the resource constraint

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

and the technology processes fully characterize the solution to the problem. Now we can solve for the equilibrium of this economy. Solving the system of non-linear expectational difference equations involves finding policy functions g and h such that

$$(C_t, L_t) = g(K_t, A_t, U_t)$$
$$\begin{bmatrix} K_{t+1} & A_{t+1} & U_{t+1} \end{bmatrix} = h(K_t, A_t, U_t) + \sigma \Omega \eta_{t+1}$$

where

$$\Omega = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & \frac{\sigma_u^2}{\sigma} \end{pmatrix}, \quad \eta_{t+1} = \begin{pmatrix} \eta_{a,t+1} \\ \eta_{u,t+1} \end{pmatrix}$$

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 $\widehat{c}_t = \log \frac{C_t}{C_{ss}}$, where C_{ss} denotes the non-stochastic steady-state. We now unify notation by letting $\widehat{x}_t$ denote the state variables $\widehat{x}_t = \begin{pmatrix} \widehat{k}_t & \widehat{a}_t & \widehat{u}_t \end{pmatrix}'$. The solution will be of the form

$$\widehat{k}_{t+1} = h_{x,1} \widehat{x}_t + \frac{1}{2} \widehat{x}_t' h_{xx,1} \widehat{x}_t + \frac{1}{2} h_{\sigma\sigma,1} \sigma^2 \tag{3}$$
$$\widehat{a}_t = \rho_a \widehat{a}_{t-1} + \sigma \eta_{a,t} \tag{4}$$
$$\widehat{u}_t = \rho_u \widehat{u}_{t-1} + \sigma_u \eta_{u,t} \tag{5}$$

and

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

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 variable is given by

$$\widehat{GDP}_t = \widehat{y}_t + \sigma_\epsilon \epsilon_t \tag{6}$$

where $\widehat{GDP}_t$ is detrended real GDP per capita and we compute

$$\widehat{y}_t = \widehat{a}_t + \alpha \widehat{k}_t + (1-\alpha)\,\widehat{l}_t$$

Equations (3)-(5) together with the observation equation (6) 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 we need to estimate are

$$\theta = \begin{pmatrix} \alpha & \beta & \delta & \psi & \rho_a & \rho_u & \sigma_\epsilon & \sigma & \sigma_u \end{pmatrix}'$$

and we use the particle filter in MCMC algorithm. The algorithm to obtain $\widehat{L}(\theta)$ inside the MCMC algorithm works as follows

1. Given θ_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 (6) and (3)-(5) to obtain $\widehat{L}(\theta_i)$.

We use quarterly data on detrended real US GDP per capita from $1960Q1$ to $2008Q2$. We use quarterly GDP in chained 2000 dollars from the Bureau of Economic Analysis and data on monthly population estimates from the U.S. Census Bureau. We divide the GDP by the quarterly average population to generate the real GDP per capita series. The log of this series is then detrended to obtain $\widehat{GDP}_t$.

In order to guarantee positive variances and $\psi > 0$ we parameterise the log of the variances and ψ. We assume a Gaussian prior given by $\theta \sim N(\theta_0, I_9)$ where θ_0 is

α	β	δ	$\log\psi$	ρ_a	ρ_u	$\log\sigma_\epsilon$	$\log\sigma$	$\log\sigma_u$
0.37	0.99	0.0154	$\log 1.956$	0.98	0.96	$\log 0.0036$	$\log 0.005$	$\log 0.0042$

Any proposals for $\alpha, \beta, \delta \notin (0,1)$, and for $\rho_a, \rho_u \notin (-1,1)$ are automatically rejected. We are using the following random walk proposals

$$
\begin{aligned}
\alpha_i &= \alpha_{i-1} + 0.033\nu_{1,i} \\
\beta_i &= \beta_{i-1} + 0.042\nu_{2,i} \\
\delta_i &= \delta_{1,i-1} + 0.034\nu_{3,i} \\
\log \psi_i &= \log \psi_{i-1} + 0.201\nu_{4,i} \\
\rho_{a,i} &= \rho_{a,i-1} + 0.078\nu_{5,i} \\
\rho_{u,i} &= \rho_{u,i-1} + 0.049\nu_{6,i} \\
\log \sigma_{\epsilon,i} &= \log \sigma_{\epsilon,i-1} + 0.249\nu_{9,i} \\
\log \sigma_i &= \log \sigma_{i-1} + 0.091\nu_{7,i} \\
\log \sigma_{u,i} &= \log \sigma_{u,i-1} + 0.096\nu_{8,i}
\end{aligned}
$$

where $\nu_{j,i} \sim i.i.N(0,1)$ for $j = 1, \ldots, 9$ and $i = 1, \ldots, N$. We loop through the parameters to make a proposal for each one individually and accept or reject it.

Currently we use $M = 24000$ and $N = 10000$. As usual we display the likelihood and the ACFs in figure 17 and the parameter histories and their histograms in figure 18. From the parameter histories and histograms we see that the parameters of the model are not well identified. This could be due to the chain being too short or the number of particles being insufficient.

Tables 10 and 11 show the usual statistics and the variance-covariance matrix. The posterior means take reasonable values. The estimates for β, ρ_a, and ρ_u seem rather far away from 1 and δ seems rather large.

	mean	MC s.e.	P(accept)	inefficiency
α	0.344	0.055	0.391	983
β	0.804	0.026	0.348	386
δ	0.127	0.025	0.372	346
ψ	3.292	0.767	0.394	63
ρ_a	0.731	0.074	0.411	565
ρ_u	0.827	0.041	0.377	836
σ_ε	0.001	0.000	0.413	31
σ	0.003	0.001	0.402	1022
σ_u	0.035	0.006	0.391	1039

Table 10: Results from MCMC with particle filter; $N = 10000$, $M = 24000$.

Fitting the DSGE model here is done more in a spirit of demonstrating the workings and capabilities of the algorithm than gaining any new insight on model parameters. We think that using more observation equations will drive the posterior means to more credible values. Also we deliberately decided not to fix $\delta = 0.0145$ or $\beta = 0.99$ as often done in practice. This concludes the dynamic model example section.

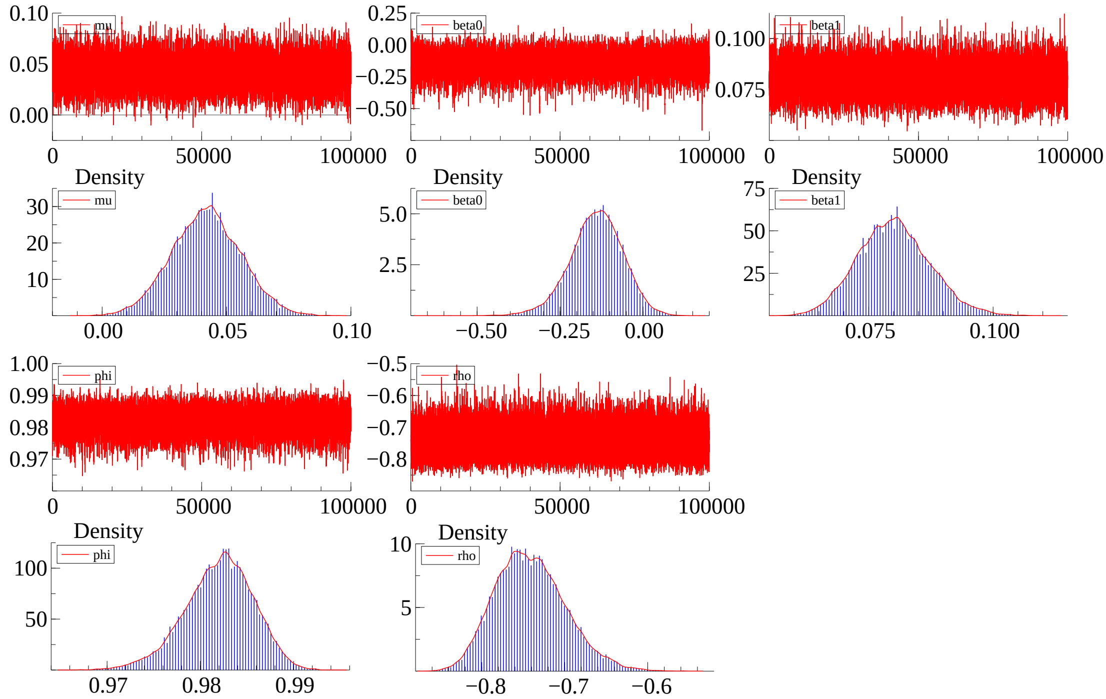

Figure 16: Stochastic volatility model; particle filter with $M = 2000$: $\theta^{(i)}$ for $i = 1, \ldots, 100000$ and histogram of parameters for $i = 50000, \ldots, 100000$.

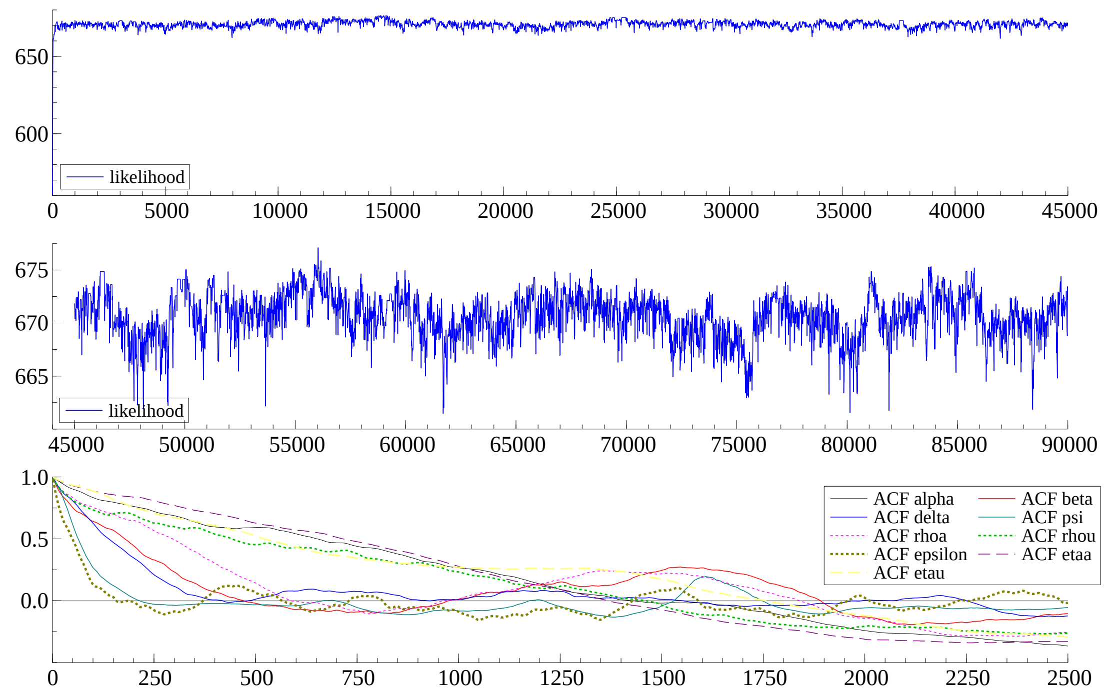

Figure 17: DSGE model; particle filter with $M = 24000$: Likelihoods and ACFs of parameters. Top: $\widehat{L}\left(\theta^{(i)}\right)$ for $i = 1, \ldots, \frac{N}{2}$; middle: $\widehat{L}\left(\theta^{(i)}\right)$ for $i = \frac{N}{2}, \ldots, N$; bottom: ACF of $\theta^{(i)}$.

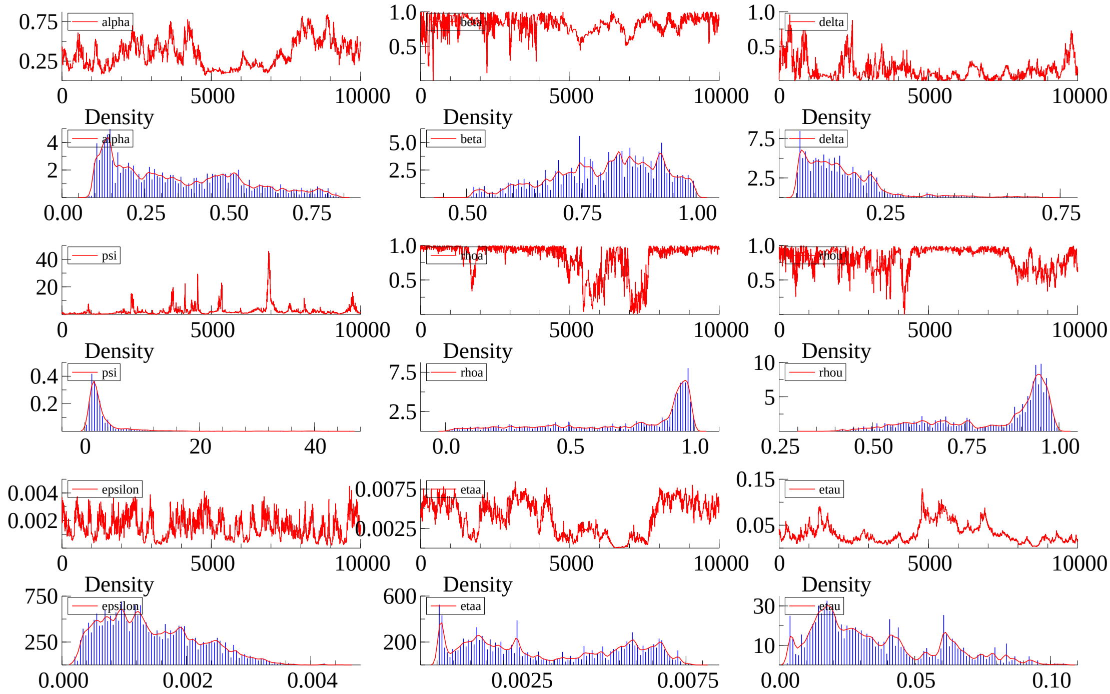

Figure 18: DSGE model; particle filter with $M = 24000$: $\theta^{(i)}$ for $i = 1, \ldots, 10000$ and histogram of parameters for $i = 5000, \ldots, 10000$.

	covariance and correlation								
α	15.228	0.599	0.286	-0.400	0.689	-0.932	-0.331	0.889	-0.917
β	4.343	3.452	0.582	-0.557	0.584	-0.583	-0.149	0.467	-0.777
δ	1.980	1.916	3.137	-0.058	0.646	-0.272	0.472	0.386	-0.467
ψ	-84.761	-56.154	-5.546	2944.000	-0.275	0.445	0.695	-0.469	0.348
ρ_a	14.036	5.662	5.971	-77.791	27.245	-0.751	-0.092	0.723	-0.648
ρ_u	-10.543	-3.140	-1.397	69.945	-11.357	8.404	0.452	-0.909	0.808
σ_ε	-0.009	-0.002	0.006	0.273	-0.003	0.009	0.000	-0.364	0.144
σ	0.167	0.042	0.033	-1.224	0.182	-0.127	0.000	0.002	-0.753
σ_u	-1.633	-0.659	-0.378	8.616	-1.543	1.070	0.000	-0.017	0.208

Table 11: Results from MCMC with particle filter; $N = 10000$, $M = 24000$; covariance (lower triangle) and correlation (upper triangle) matrix.

5 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 a MCMC algorithm. The theory of this is very simple.

In this paper we show the power of this approach, providing examples drawn from microeconometrics, 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.

A Unbiasedness of particle filter

We now provide a proof of the claim that the likelihood approximation from the particle filter is unbiased. The proof closely follows Del-Moral (2004) and has no originality in it, but is accessible. For the sake of notational simplicity we drop the reference to the parameters. We first introduce extra notation and highlight an important property of the particle filter described in section 3.3. The proof then works by providing the martingale decomposition of the estimation error of the particle filter.

A.1 The particle filter

At each point in time the particle cloud of the particle filter provides a random measure defined by

$$\eta_t^M = \frac{1}{M}\sum_{j=1}^{M}\delta_{\alpha_t^{(j)}}$$

where $\delta_{\alpha_t^{(j)}}$ is the dirac delta, for the prediction density

$$\eta_t = \mathrm{d}F\left(\alpha_t \mid \mathcal{F}_{t-1}\right)$$

For any function h_t we get an approximation for

$$\eta_t\left(h_t\right) = \mathrm{E}\left[h_t\left(\alpha_t\right) \mid \mathcal{F}_{t-1}\right]$$

given by

$$\eta_t^M\left(h_t\right) = \frac{1}{M}\sum_{j=1}^{M} h_t\left(\alpha_t^{(j)}\right)$$

We can write the likelihood approximation from the particle filter as

$$\widehat{f}(y) = \prod_{t=0}^{T} \eta_t^M (G_t) = \prod_{t=0}^{T} \frac{1}{M} \sum_{j=1}^{M} G_t \left(\alpha_t^{(j)} \right)$$

where we use notation $G_t(\alpha_t) = \mathrm{d}F(y_t \mid \alpha_t)$ for the measurement density. The resampling step can be understood as drawing indices from a multinomial distribution. Denote by $n_t^{(j)}$ the number of times particle $\alpha_t^{(j)}$ has been resampled. The $n_t^{(j)}$ follow a multinomial distribution

$$\left(n_t^{(1)}, \ldots, n_t^{(M)} \right) \sim \text{Multinomial} \left(M, W_t^{(1)}, \ldots, W_t^{(M)} \right)$$

where $W_t^{(j)}$ are the normalized weights. An important property of this resampling algorithm is that

$$E\left[n_t^{(j)} \mid \overline{\alpha}_t \right] = \frac{G_t\left(\alpha_t^{(j)}\right)}{\frac{1}{M}\sum_{k=1}^{M} G_t\left(\alpha_t^{(k)}\right)}$$

where $\overline{\alpha}_t = \left\{ \alpha_t^{(1)}, \ldots, \alpha_t^{(M)} \right\}$. Hence

$$E\left[\frac{1}{M} \sum_{j=1}^{M} n_t^{(j)} \delta_{\alpha_t^{(j)}} \mid \overline{\alpha}_t \right] = \sum_{j=1}^{M} \frac{G_t\left(\alpha_t^{(j)}\right)}{\sum_{k=1}^{M} G_t\left(\alpha_t^{(k)}\right)} \delta_{\alpha_t^{(j)}}$$

After resampling we draw predictive particles

$$\alpha_{t+1}^{(j)} \sim \mathrm{d}F\left(\alpha_{t+1} \mid \alpha_t^{(i^j)} \right) \quad j = 1, \ldots, M$$

These particles now provide an approximation for η_{t+1}, given by

$$\eta_{t+1}^M = \frac{1}{M} \sum_{j=1}^{M} \delta_{\alpha_{t+1}^{(j)}}$$

We conclude by pointing out an observation which will be crucial for the proof. Note that

$$\begin{aligned} E\left[\eta_{t+1}^M (G_{t+1}) \mid \overline{\alpha}'_t \right] &= \frac{1}{M} \sum_{j=1}^{M} \int \mathrm{d}F\left(\alpha_{t+1} \mid \alpha_t^{(i^j)} \right) G_{t+1}(\alpha_{t+1}) \, \mathrm{d}\alpha_{t+1} \\ &= \frac{1}{M} \sum_{j=1}^{M} n_t^{(j)} \int \mathrm{d}F\left(\alpha_{t+1} \mid \alpha_t^{(j)} \right) G_{t+1}(\alpha_{t+1}) \, \mathrm{d}\alpha_{t+1} \end{aligned}$$

where $\overline{\alpha}'_t = \left\{ \alpha_t^{(i^1)}, \ldots, \alpha_t^{(i^M)} \right\}$ and hence

$$\begin{aligned} E\left[\eta_{t+1}^M (G_{t+1}) \mid \overline{\alpha}_t \right] &= E\left[\frac{1}{M} \sum_{j=1}^{M} n_t^{(j)} \int \mathrm{d}F\left(\alpha_{t+1} \mid \alpha_t^{(j)} \right) G_{t+1}(\alpha_{t+1}) \, \mathrm{d}\alpha_{t+1} \mid \overline{\alpha}_t \right] \\ &= \sum_{j=1}^{M} \frac{G_t\left(\alpha_t^{(j)}\right)}{\sum_{k=1}^{M} G_t\left(\alpha_t^{(k)}\right)} \int \mathrm{d}F\left(\alpha_{t+1} \mid \alpha_t^{(j)} \right) G_{t+1}(\alpha_{t+1}) \, \mathrm{d}\alpha_{t+1} \end{aligned} \quad (7)$$

Now we proceed by using these results in a martingale decomposition of the estimation error to show unbiasedness.

A.2 Unbiasedness

We now produce a martingale decomposition of the difference between the particle approximation and the true likelihood. We derive the true likelihood of the model by

$$\begin{aligned} \mathrm{d}F(y) &= \int \mathrm{d}F(y \mid \alpha)\, \mathrm{d}F(\alpha)\, \mathrm{d}\alpha \\ &= \int \prod_{t=0}^{T} G_t(\alpha_t) \prod_{t=0}^{T} \mathrm{d}F(\alpha_t \mid \alpha_{t-1})\, \mathrm{d}\alpha \\ &= E\left[\prod_{t=0}^{T} G(\alpha_t)\right] \end{aligned}$$

where $\alpha = (\alpha_1, \ldots, \alpha_T)$. Note that the expectation is with respect to the distribution of the Markov chain α. Write the prediction error as

$$\begin{aligned} \Gamma_T^M(G_T) &= \mathrm{d}\widehat{F}(y) - \mathrm{d}F(y) \\ &= \prod_{t=0}^{T} \frac{1}{M} \sum_{j=1}^{M} G_t\left(\alpha_t^{(j)}\right) - E\left[\prod_{t=0}^{T} G_t(\alpha_t)\right] \end{aligned}$$

For any function h_{t+1} define the operators

$$\begin{aligned} Q_{t+1}(h_{t+1})(\alpha_t) &= G_t(\alpha_t) \int f(\alpha_{t+1} \mid \alpha_t)\, h_{t+1}(\alpha_{t+1})\, \mathrm{d}\alpha_{t+1} \\ Q_{t,T} &= Q_{t+1} Q_{t+2} \cdots Q_T \end{aligned}$$

We understand then that $Q_{t+1}(G_{t+1})(\alpha_t)$ moves the likelihood from period t to $t+1$ and hence $Q_{t,T}$ moves it from t to T. This operator is crucial for the proof. Q_{t+1} operates on measures as well as particle approximations to it. Now

$$\begin{aligned} \Gamma_T^M(G_T) &= \prod_{t=0}^{T} \frac{1}{M} \sum_{j=1}^{M} G_t\left(\alpha_t^{(j)}\right) - E\left[\prod_{t=0}^{T} G_t(\alpha_t)\right] \\ &= \prod_{t=0}^{T} \frac{1}{M} \sum_{j=1}^{M} G_t\left(\alpha_t^{(j)}\right) - E\left[Q_{0,T}(G_T)(\alpha_0)\right] \\ &= \gamma_T^M \frac{1}{M} \sum_{j=1}^{M} G_T\left(\alpha_T^{(j)}\right) - E\left[Q_{0,T}(G_T)(\alpha_0)\right] \end{aligned}$$

To make the formulae more readable we used notation

$$\gamma_T^M = \prod_{t=0}^{T-1} \eta_t^M\left(G_t(\alpha_t)\right)$$

Then we write

$$\Gamma_T^M(G_T) = \gamma_T^M \frac{1}{M} \sum_{j=1}^{M} G_T\left(\alpha_T^{(j)}\right) - \gamma_{T-1}^M \frac{1}{M} \sum_{j=1}^{M} Q_{T-1,T}(G_T)\left(\alpha_{T-1}^{(j)}\right)$$

$$+\gamma_{T-1}^{M}\frac{1}{M}\sum_{j=1}^{M}Q_{T-1,T}(G_T)\left(\alpha_{T-1}^{(j)}\right)-\gamma_{T-2}^{M}\frac{1}{M}\sum_{i=1}^{M}Q_{T-2,T}(G_T)\left(\alpha_{T-2}^{(j)}\right)$$

$$\ldots$$

$$+\frac{1}{M}\sum_{j=1}^{M}Q_{0,T}(G_T)\left(\alpha_0^{(j)}\right)-E\left[Q_{0,T}(G_T)\left(\alpha_0\right)\right]$$

$$=\sum_{t=0}^{T}\left(\gamma_t^{M}\frac{1}{M}\sum_{j=1}^{M}Q_{t,T}(G_T)\left(\alpha_t^{(j)}\right)-\gamma_{t-1}^{M}\frac{1}{M}\sum_{i=1}^{M}Q_{t-1,T}(G_T)\left(\alpha_{t-1}^{(j)}\right)\right)$$

where we use the conventions $Q_{T,T}=Id$ and $\gamma_{-1}^{M}\frac{1}{M}\sum_{j=1}^{M}Q_{-1,T}(G_T)\left(\alpha_{-1}^{(j)}\right)=E\left[Q_{0,T}(G_T)\left(\alpha_0\right)\right]$. Define the filtration of the particle history $\mathcal{F}_{t-1}^{M}=(\overline{\alpha}_0,\ldots,\overline{\alpha}_{t-1})$ and take the expectation

$$E\left[\Gamma_T^M\left(G_T\right)\mid\mathcal{F}_{T-1}^{M}\right]=\sum_{t=0}^{T}\gamma_t^{M}E\left[\frac{1}{M}\sum_{j=1}^{M}Q_{t,T}(G_T)\left(\alpha_t^{(j)}\right)\mid\mathcal{F}_{t-1}^{M}\right]$$
$$-\sum_{t=0}^{T}\gamma_{t-1}^{M}\frac{1}{M}\sum_{j=1}^{M}Q_{t-1,T}(G_T)\left(\alpha_{t-1}^{(j)}\right)$$

Using (7), the first term is

$$E\left[\frac{1}{M}\sum_{j=1}^{M}Q_{t,T}(G_T)\left(\alpha_t^{(j)}\right)\mid\mathcal{F}_{t-1}^{M}\right]=\sum_{j=1}^{M}\frac{G_{t-1}\left(\alpha_{t-1}^{(j)}\right)}{\sum_{k=1}^{M}G_{t-1}\left(\alpha_{t-1}^{(k)}\right)}\int\mathrm{d}F\left(\alpha_t\mid\alpha_{t-1}^{(j)}\right)Q_{t,T}(G_T)\left(\alpha_t\right)\mathrm{d}\alpha_t$$

For the second term we have

$$\gamma_{t-1}^{M}\frac{1}{M}\sum_{j=1}^{M}Q_{t-1,T}(G_T)\left(\alpha_{t-1}^{(j)}\right)=\gamma_t^{M}\frac{1}{\eta_{t-1}^{M}\left(G_{t-1}\left(\alpha_{t-1}\right)\right)}\frac{1}{M}\sum_{j=1}^{M}Q_{t-1,T}(G_T)\left(\alpha_{t-1}^{(j)}\right)$$
$$=\gamma_t^{M}\frac{1}{\eta_{t-1}^{M}\left(G_{t-1}\left(\alpha_{t-1}\right)\right)}\frac{1}{M}\sum_{j=1}^{M}G_{t-1}\left(\alpha_{t-1}^{(j)}\right)\int\mathrm{d}F\left(\alpha_t\mid\alpha_{t-1}^{(j)}\right)Q_{t,T}(G_T)\left(\alpha_t\right)\mathrm{d}\alpha_t$$
$$=\gamma_t^{M}\sum_{j=1}^{M}\frac{G_{t-1}\left(\alpha_{t-1}^{(j)}\right)}{\sum_{k=1}^{M}G_{t-1}\left(\alpha_{t-1}^{(k)}\right)}\int\mathrm{d}F\left(\alpha_t\mid\alpha_{t-1}^{(i)}\right)Q_{t,T}(G_t)\left(\alpha_t\right)\mathrm{d}\alpha_t$$

Thus

$$E\left[\Gamma_T^M\left(G_T\right)\mid\mathcal{F}_{T-1}^{M}\right]=0$$

Interpretation: We have decomposed the estimation error $\Gamma_T^M\left(G_T\right)$ into a sum of differences. When we use a particle filter and propagate the particle cloud through time we accumulate approximation errors: At the beginning of time we use approximation η_0^M for η_0 and hence make a sampling error. Then we would like to move this measure forward one period according to the updating and prediction equations (which we cannot solve for analytically). Denote this hypothetically forward-moved measure by $\Phi_1\left(\eta_0^M\right)$, where Φ_1 is the "move forward" operator. Ideally we would use

$\Phi_1\left(\eta_0^M\right)$ as an approximation for η_1. As we cannot do this, we approximate this quantity with η_1^M, and make yet another sampling error. Again, as we cannot move this forward "directly" and use $\Phi_2\left(\eta_1^M\right)$, we approximate it with η_2^M making yet another sampling error. Running this reasoning up to time T we can write the difference between the particle approximation and the truth as a sum of one-step prediction errors. One then shows that the conditional expectation of η_t^M given particles at time $t-1$ is the same as $\Phi_t\left(\eta_{t-1}^M\right)$. See also Del-Moral (2004) Ch. 7 for more on this.

References

Amisano, G. and O. Tristani (2007). Euro area inflation persistence in an estimated nonlinear DSGE model. *ECB Working Waper 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.

Aruoba, S. B., J. Fernandez-Villaverde, and J. F. Rubio-Ramirez (2006). Comparing soultion 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.

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.

Fernanzed-Villaverde, J. and J. F. Rudio-Ramirez (2007). Estimating macroeconomics models: a likelihood approach. *Review of Economic Studies 74*, 1059–1087.

Fernanzed-Villaverde, J., J. F. Rudio-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. 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.

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 expecatations 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.

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. (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.