Econometric Theory, **27**, 2011, 933–956.
doi:10.1017/S0266466610000599

BAYESIAN INFERENCE BASED ONLY ON SIMULATED LIKELIHOOD: PARTICLE FILTER ANALYSIS OF DYNAMIC ECONOMIC MODELS

THOMAS FLURY AND NEIL SHEPHARD
University of Oxford

We note that likelihood inference can be based on an unbiased simulation-based estimator of the likelihood when it is used inside a Metropolis–Hastings algorithm. This result has recently been introduced in statistics literature by Andrieu, Doucet, and Holenstein (2010, *Journal of the Royal Statistical Society, Series B*, 72, 269–342) and is perhaps surprising given the 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 through a particle filter. We illustrate these methods on four problems, 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.

1. INTRODUCTION

Simulation-based estimators of the likelihood 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 (2010) 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 centered 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 simulated maximum likelihood (SML) estimator of Lerman and Manski (1981) and Diggle and Gratton (1984).

We thank Arnaud Doucet for various stimulating conversations on this topic, Martin Ellison for his advice, and the editor and referees for their comments. Address correspondence to Thomas Flury, Oxford-Man Institute, University of Oxford, Eagle House, Walton Well Road, Oxford OX2 6EE, United Kingdom; e-mail: thomas.flury@oxford-man.ox.ac.uk.

Our focus in this paper is on four models. We estimate a standard cross-sectional probit model and a linear, Gaussian time series model. In these two models the exact likelihood is available as a point of comparison. The third model is a simple stochastic volatility model. The fourth is a stochastic general equilibrium model, which 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 Markov chain Monte Carlo (MCMC) algorithms are difficult to apply because of the forward optimizing behavior, which is common in economic models.

1.1. MCMC and Particle Filters

The use of the Metropolis–Hastings algorithm to produce MCMC draws from the posterior distribution of parameters given the data is now well established in econometrics (e.g., Chib, 2001). 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 forward looking problems typically have no explicit likelihood.

Instead of having to evaluate the likelihood exactly, this paper advocates the use of a simulated likelihood, often carried out using a particle filter, which is a sequential simulation device for online filtering for non-Gaussian, nonlinear state space models (e.g., Gordon, Salmond, and Smith, 1993; Pitt and Shephard, 1999; and Doucet, de Freitas, and Gordon, 2001). It can be thought of as a modern generalization of the Kalman filter, which is only able to analyze linear, Gaussian state space models. The particle filter has minimal requirements: the ability to simulate from the dynamics of the state and 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.

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; Durbin and Koopman, 2001). Kim, Shephard, and Chib (1998) have used the particle filter to estimate the likelihood for the purpose of model comparison. Pitt (2001) carries out SML estimation based on a smoothed version of a particle filter, whereas Durbin and Koopman (2001) discuss importance samplers as the basis of the same estimation methods.

Particle filters are increasingly used in economics, following their introduction to that subject by Kim et al. (1998) and Pitt and Shephard (1999), who extracted the filtered estimate of the current volatility from a stochastic volatility model. Later work on this, including developments to try to learn parameters online using particle filters, can be seen in Andrieu, de Freitas, and Doucet

(1999), Fearnhead (2002), Storvik (2002), Johannes, Polson, and Stroud (2008), and Johannes and Polson (2009). An interesting application in microeconomics is the work of 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 for maximum likelihood estimation of stochastic general equilibrium models. This is an application of SML 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 (e.g., Durbin and Koopman, 2001) can also be combined with MCMC to produce valid samples from the correct posterior distribution. All of that work can be immediately ported over to the Bayesian environment but without any distortions created by the simulation errors.

1.2. 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 SML approach. In Section 3 we give an example of a static model. Section 4 treats dynamic models and introduces particle filters. In Sections 5–7 we provide examples from macroeconomics and finance to demonstrate the performance of this algorithm in dynamic models. Section 8 concludes.

2. EXPECTED AUXILIARY VARIABLE METHOD

2.1. An Auxiliary Variable Interpretation of Importance Sampling

Write θ as the parameters and use y to denote the data. Then we wish to carry out inference by sampling from

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

where $f(\theta)$ is a prior and $f(y|\theta)$ a likelihood. Here f denotes the density or probability function.

We assume that we have an unbiased simulation-based estimator

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

where we average over the simulation denoted by u. The auxiliary variable u might not have any physical interpretation, but in practice they are usually independent and identically distributed (i.i.d.) uniforms, which are transformed to model specific random variables. We assume $\hat{f}$ is itself a density function in y. Then we can think of $\hat{f}$ as being based on an auxiliary variable:

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

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

2.2. Bayesian Analysis

This insight has implications econometrically; for 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 $\left(u^{(1)},\theta^{(1)}\right)$, $\left(u^{(2)},\theta^{(2)}\right),\ldots,\left(u^{(N)},\theta^{(N)}\right)$, and so throwing away the u samples leaves us with $\theta^{(1)}$, $\theta^{(2)},\ldots,\theta^{(N)}$ that 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.

We can sample using generic MCMC algorithms. Sample θ^* from a proposal density $Q\left(\theta^*|\theta^{(i-1)}\right)$, draw the u, and compute $\hat{L}^* = \hat{f}_u^*(y|\theta^*)$. The acceptance probability is given by

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

Draw $V \sim U(0,1)$ and if $V \leq q$ set $\left(\hat{L}^{(i)},\theta^{(i)}\right) = \left(\hat{L}^*,\theta^*\right)$; otherwise $\left(\hat{L}^{(i)},\theta^{(i)}\right) = \left(\hat{L}^{(i-1)},\theta^{(i-1)}\right)$. Under very weak conditions (e.g., Chib, 2001) the sequence $\left\{\theta^{(i)}\right\}$ for $i = 1,\ldots,N$ converges to samples from $f(\theta|y)$ as $N \to \infty$.

We decide how much simulation effort to put in, which influences the variability of $\hat{f} - f$. This does not affect the equilibrium distribution of the MCMC, but it does affect the rejection rate. If the estimator is poor then $\hat{f}$ will be a very jittery estimator of f, which will increase the chance that the algorithm gets stuck. Hence increasing the amount of simulation will improve the mixing of the MCMC chain.

We saw this style of argument first in the context of dynamic models in Andrieu et al. (2010), and we know of an unpublished note by Andrieu et al. (2007), who call this approach the Expected Auxiliary Variable method; 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

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

where we assume that $y_t|x_t$ are independent over t. Assume that we can simulate from $F_t(\psi)$. Write these simulations as $\varepsilon_t^{(1)}, \ldots, \varepsilon_t^{(M)}$. Then the simplest simulation-based estimator of p_t is $\hat{p}_t = 1/M \sum_{j=1}^{M} 1_{x_t'\beta + \varepsilon_t^{(j)} \geq 0}$ delivering the simulated likelihood function

$$\hat{f}(y|\beta, \psi) = \prod_{t=1}^{T} \hat{p}_t^{\,y_t}(1 - \hat{p}_t)^{1-y_t}, \quad \text{estimating } f(y|\beta, \psi) = \prod_{t=1}^{T} p_t^{y_t}(1 - p_t)^{1-y_t}.$$

This estimator is unbiased, but the score is biased. It is this bias that drives the fact that the SML estimator of $\theta = \left(\beta', \psi'\right)'$ behaves poorly asymptotically unless $M \to \infty$.

2.4. Comparison to SML Estimation

Inference using SML estimation goes back at least to Lerman and Manski (1981) and Diggle and Gratton (1984). Simulation is used to unbiasedly estimate the likelihood using M i.i.d. draws, which are kept common as θ varies. The estimated likelihood is numerically maximized with respect to θ.

Based on a sample of size T, for i.i.d. data we need that $T, M \to \infty$ for this SML estimator to be consistent and $\sqrt{T}/M \to 0$ to have the same distribution as the maximum likelihood estimator. This is a disappointing result, for it means that in a data-rich environment the simulation demands of the SML method could be rather large. Further, it is often difficult to know if $\sqrt{T}/M$ is close to zero. These observations led to the development of new simulation-based inference strategies in the late 1980s and early 1990s. These included simulated scores, the stochastic expectation-maximization algorithm, indirect inference, and efficient method of moments. Discussions of these topics can be found in, for example, Hajivassiliou and McFadden (1998), Gourieroux et al. (1993), Smith (1993), Gallant and Tauchen (1996), and Gourieroux and Monfort (1996).

Another disadvantage of SML is that in practice it requires continuity in the estimated log-likelihood and hence is not even applicable in the preceding simple example.

The MCMC algorithm is fundamentally different from SML. Here new uniforms are drawn at each iteration. This is crucial. When one uses common random numbers $\overline{u}$, which are held fixed as we vary θ, then $\hat{f}_{\overline{u}}(y|\theta) = g(y, \overline{u}|\theta)$, such that we converge to the wrong posterior, which is proportional to $\hat{f}_{\overline{u}}(y|\theta)f(\theta)$ if we use this inside the Metropolis–Hastings algorithm. This posterior then needs M to increase quickly for it not to have an impact on the solution.

Here our only concern is the choice of M to ensure 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. An interesting suggestion came from one referee, who suggested choosing M such as to aim for the variance of the estimated log-likelihood—for a given θ—to be around one on the log-likelihood scale.

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, returning to Section 2.3. We use the classical data set from Mroz (1987) to study the labor 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 we can cleanly assess the effectiveness of the expected auxiliary variable method by comparing it to classical MCMC.

In this model we have a binary variable y_t that 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

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

The explanatory variables are nonwife income, education, experience, experience squared, age, number of children less than six years old, and number of children between the ages of six and 18. We write $p_t = \Pr(y_t = 1|x_t, \beta, \psi)$, where $\beta = (\beta_0, \ldots, \beta_7)'$. Here we choose the normal distribution for ε_t. We assume a Gaussian prior given by $\beta \sim N\left(\beta^0, I_8\right)$ where[1] $\beta^0 = (0.5855, -0.0034, 0.0380, 0.0395, -0.0006, -0.0161, -0.2618, 0.0130)'$. Given the Gaussian errors we can easily compute the true likelihood and thus use the results from this exact likelihood model as a benchmark.[2]

For the simulation-based estimator of p_t we draw $\varepsilon_t^{(j)} \overset{i.i.d.}{\sim} N\left(0, \sigma_\varepsilon^2\right)$, $j = 1, \ldots, M$, and compute the simulation-based estimators of the probabilities as outlined in Section 2.3.

In the usual probit model the variance has to be normalized (usually one sets $\sigma_\varepsilon = 1$), for it is impossible to estimate both β and σ_ε^2. 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 $\left(y_t = 1, \hat{p}_t = 0\right)$ that 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]

3.2. The Sampling Algorithm

Throughout we will use a Metropolis–Hastings algorithm to sample from the posterior distribution, using the estimated likelihood. Throughout we use random walk proposals for the parameters, each applied one at a time: $\Delta\beta_{0,i} = 0.1326\nu_{1,i}$, $\Delta\beta_{1,i} = 0.0058\nu_{2,i}$, $\Delta\beta_{2,i} = 0.0109\nu_{3,i}$, $\Delta\beta_{3,i} = 0.0108\nu_{4,i}$, $\Delta\beta_{4,i} = 0.0005\nu_{5,i}$, $\Delta\beta_{5,i} = 0.0031\nu_{6,i}$, $\Delta\beta_{6,i} = 0.2317\nu_{7,i}$, $\Delta\beta_{7,i} = 0.0703\nu_{8,i}$, where Δ denotes the difference operator and $\nu_{j,i} \overset{i.i.d.}{\sim} N(0,1)$ 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, for example, Gelman, Carlin, Stern, and Rubin (2003). The proposal variances were tuned on the exact likelihood MCMC. This will allow us to see how fast the estimated likelihood comes close to the truth as we increase M. We set $M = 1{,}000$, $M = 2{,}000$, $M = 4{,}000$, and $N = 100{,}000$.

3.3. The Results

Table 1 shows the following statistics for the MCMC algorithm: the arithmetical mean, the Monte Carlo standard error,[4] the acceptance probability, and the inefficiency of the MCMC algorithm compared to a hypothetical sampler that produces i.i.d. draws from the posterior distribution.[5] In all examples in this paper we discard the first half of the MCMC samples to reduce the impact of burn-in, so that all statistics are based on the second half and thus roughly represent draws from the posterior distribution.

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 MCMC 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. 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 standard deviations (diagonal) and correlation (upper triangle) matrix of the parameters. For brevity we only report this for the exact likelihood and the $M = 4{,}000$ estimated likelihood model.

Figure 1 compares the likelihood estimates and the autocorrelation functions (ACFs) for $N = 100{,}000$ between the exact likelihood MCMC and the ones with estimated likelihoods with $M = 1{,}000$, $M = 2{,}000$, and $M = 4{,}000$. Figure 2 compares the estimated posterior densities of the parameters. We note that the performance improves as M increases.

All simulations in this paper were carried out in OxMetrics 5 on an Intel Core 2 CPU with 2.66 GHz and 4 GB RAM. Computer time for the exact likelihood MCMC was roughly 600 seconds; for the estimated likelihood with $M = 1{,}000$ it was roughly 20,000 seconds; for $M = 2{,}000$ it was 37,000 seconds, and for $M = 4{,}000$ it was 70,000 seconds.

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.

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 nonlinear and non-Gaussian state space model of the following type.

TABLE 1. Results from MCMC for labor force participation using $N = 100{,}000$, $M = 1{,}000$, $M = 2{,}000$, and $M = 4{,}000$. Here μ_{mc} denotes the arithmetical mean, σ_{mc} 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.

	Exact likelihood MCMC				$M = 1{,}000$				$M = 2{,}000$				$M = 4{,}000$			
	μ_{mc}	σ_{mc}	Pr	inef	μ_{mc}	σ_{mc}	Pr	inef	μ_{mc}	σ_{mc}	Pr	inef	μ_{mc}	σ_{mc}	Pr	inef
β_0	0.295	0.033	0.42	364	0.313	0.039	0.28	593	0.349	0.035	0.33	462	0.376	0.030	0.36	420
β_1	−0.012	0.000	0.41	31	−0.012	0.000	0.28	57	−0.012	0.000	0.33	35	−0.012	0.000	0.36	42
β_2	0.130	0.001	0.41	143	0.130	0.002	0.27	328	0.130	0.001	0.33	209	0.126	0.001	0.36	168
β_3	0.124	0.001	0.41	149	0.125	0.001	0.27	149	0.123	0.001	0.32	160	0.124	0.001	0.36	197
β_4	−0.002	0.000	0.41	130	−0.002	0.000	0.28	135	−0.002	0.000	0.33	162	−0.002	0.000	0.36	189
β_5	−0.053	0.001	0.41	299	−0.054	0.001	0.28	479	−0.054	0.001	0.33	458	−0.054	0.001	0.37	321
β_6	−0.868	0.004	0.43	70	−0.871	0.004	0.29	100	−0.876	0.004	0.34	128	−0.871	0.004	0.37	67
β_7	0.035	0.001	0.41	81	0.034	0.002	0.28	125	0.032	0.002	0.33	94	0.032	0.001	0.36	77

TABLE 2. Results from MCMC for labor force participation; exact likelihood MCMC with $N = 100{,}000$ and approximate analysis using $M = 4{,}000$; posterior standard deviations (diagonal) and correlation (upper triangle) matrix

	Posterior standard deviations and correlation															
	Exact likelihood MCMC								$M = 4{,}000$							
β_0	0.44	0.07	−0.54	−0.15	0.18	−0.75	−0.30	−0.41	0.40	0.02	−0.43	−0.21	0.23	−0.71	−0.29	−0.40
β_1		0.00	−0.34	0.05	0.02	−0.13	0.04	−0.04		0.00	−0.34	0.03	0.05	−0.09	0.08	−0.01
β_2			0.02	−0.06	0.04	0.00	−0.17	0.05			0.02	−0.03	0.03	−0.16	−0.25	−0.04
β_3				0.02	−0.92	−0.07	−0.03	0.06				0.02	−0.92	−0.06	−0.01	0.09
β_4					0.00	−0.06	0.01	0.02					0.00	−0.07	0.01	−0.01
β_5						0.01	0.44	0.29						0.01	0.44	0.29
β_6							0.12	0.10							0.12	0.10
β_7								0.04								0.04

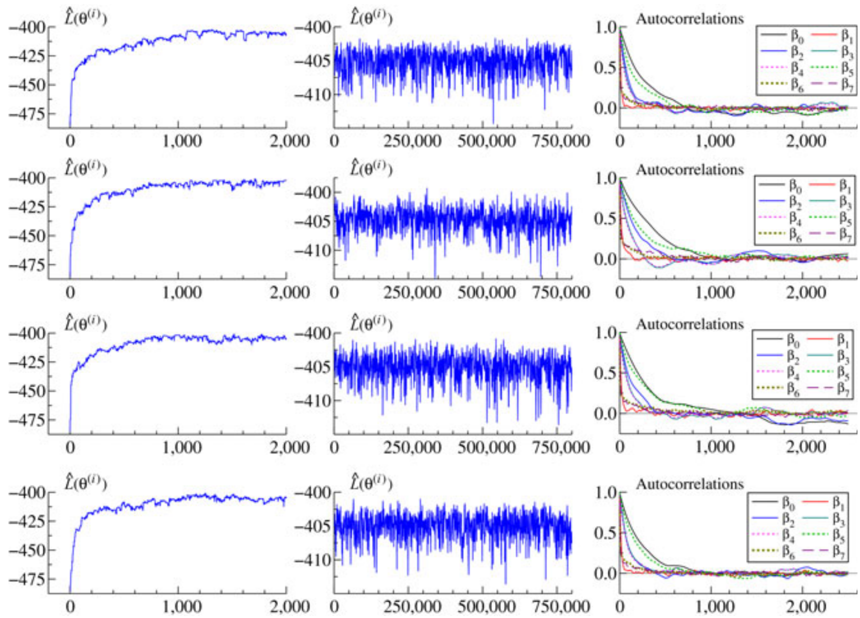

FIGURE 1. Labor force participation model: likelihoods and ACFs of parameters. First column: $\hat{L}(\theta^{(i)})$ for $i = 1,\ldots,2{,}000$; second column: $\hat{L}(\theta^{(i)})$ for $i = 2{,}001,\ldots,8N$; third column: ACF of $\theta^{(i)}$. First row: exact likelihood MCMC; second row: estimated likelihood with $M = 1{,}000$; third row: $M = 2{,}000$; fourth row: $M = 4{,}000$.

Assumption 1. For the simulation state space (SSS) model:

1. We can compute the measurement density $f(y_t|\alpha_t,\mathcal{F}_{t-1},\theta)$, $t = 1,2,\ldots,T$, where α_t is the unobserved state and $\mathcal{F}_{t-1} = \sigma(y_1,y_2,\ldots,y_{t-1})$ is the natural filtration.
2. We can simulate from the random variable $\alpha_t|\alpha_{t-1},\mathcal{F}_{t-1},\theta$, $t = 1,2,\ldots,T$, where we assumed that we can also draw from the initial condition $\alpha_0|\mathcal{F}_0,\theta$.
3. We can compute the prior $f(\theta)$.

Remark 1. 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 most of the time we can 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. A large number of intractable econometric models are of this SSS form. Leading examples are dynamic stochastic general equilibrium (DSGE) models, some (continuous time) stochastic volatility models, and some models in industrial organization.

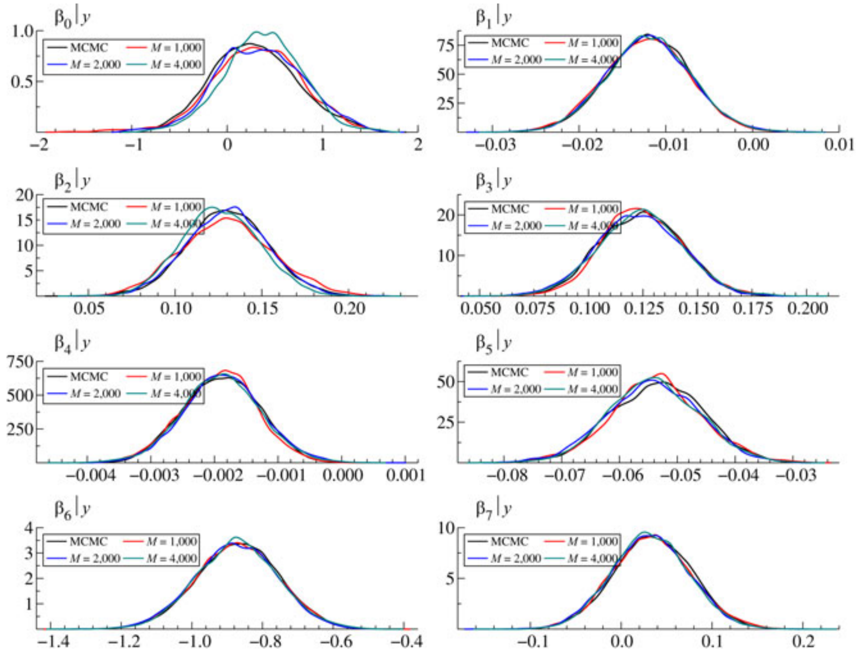

FIGURE 2. Labor force participation model; nonparametric kernel estimator of posterior density for $i = 50{,}000, \ldots, 100{,}000$. Black: exact likelihood MCMC; red: estimated likelihood with $M = 1{,}000$; blue: $M = 2{,}000$; green: $M = 4{,}000$.

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 formulas.

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 preceding strategy.

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

Here we give a simple particle filter, generically coded for SSS models.

Particle filter

1. Draw $\alpha_1^{(1)}, \ldots, \alpha_1^{(M)}$ from $\alpha_1 | \mathcal{F}_0, \theta$. Set $t = 1$ and $l_0 = 0$.
2. Compute $w_t^{(j)} = f(y_t | \alpha_t^{(j)}, \mathcal{F}_{t-1}, \theta)$ and $W_t^{(j)} = w_t^{(j)} / \left(\sum_{k=1}^{M} w_t^{(k)} \right)$ for $j = 1, \ldots, M$.
3. Resample by drawing $u \sim U(0,1)$ and let $u^{(j)} = \frac{u}{M} + (j-1)/M$ for $j = 1, \ldots, M$. Find the indexes $i^1, \ldots, i^M$ such that $\sum_{k=1}^{i^j - 1} W_t^{(k)} < u^{(j)} \leq \sum_{k=1}^{i^j} W_t^{(k)}$.
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 $l_t(\theta) = l_{t-1}(\theta) + \log \left\{ M^{-1} \sum_{j=1}^{M} w_t^{(j)} \right\}$. Let $t = t + 1$ and goto 2.

Then it is relatively easy to show that $\exp(l_T(\theta)) \overset{a.s.}{\to} f(y|\theta)$ as $M \to \infty$ (see, e.g., Del-Moral (2004)),[6] whereas crucially for us $\mathrm{E}\left[\exp(l_T(\theta))\right] = f(y|\theta)$. See Del-Moral (2004, Thm. 7.4.2) for a proof. It is worthwhile to note that the unbiasedness property is valid for more sophisticated particle filters.[7] Yet another nice property of particle filters, which is not essential but extremely useful for this approach, is that under mixing assumptions the relative variance of the estimate of the likelihood only increases linearly in time (see Cérou, Del Moral, and Guyader, 2008), and so we expect them to work well in an MCMC algorithm, even for large data sets.

To carry out MCMC we proceed by replacing $f(y|\theta)$ by its particle filter estimate $\hat{L}(\theta) = \exp\{l_T(\theta)\}$, which we are allowed to do as the particle filter indeed provides an unbiased estimate of $f(y \mid \theta)$, 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.

5. EXAMPLE: GAUSSIAN LINEAR MODEL

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

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

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

where $\alpha_0 \sim N\left(0, \sigma_\eta^2 / \left(1 - \phi^2\right)\right)$. To guarantee positive variances we parameterize the log of the variances, and we impose that $\phi \in (-1, 1)$ by allowing no prior probability outside that region. We take $\theta = \left(\mu, \log\sigma_\epsilon^2, \phi, \log\sigma_\eta^2\right)'$, 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 data generating process 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: $\Delta\mu_i = 0.3298\nu_{1,i}$, $\Delta\log\sigma_{\epsilon,i}^2 = 0.1866\nu_{2,i}$, $\Delta\phi_i = 0.0671\nu_{3,i}$, $\Delta\log\sigma_{\eta,i}^2 = 0.2676\nu_{4,i}$, where $\nu_{j,i} \overset{i.i.d.}{\sim} N(0,1)$ 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.

In Figure 3 we report the Markov chain output for $\mu|y$ and $\sigma_\epsilon^2|y$ and the histograms (the other parameters have similar types of sample paths). All histograms plotted in this paper are only based on the second half of the chains. The last column shows the output from the Kalman filtered exact likelihood MCMC for $N = 100{,}000$. We first ran the particle MCMC at $M = 100$ and $N = 100{,}000$.

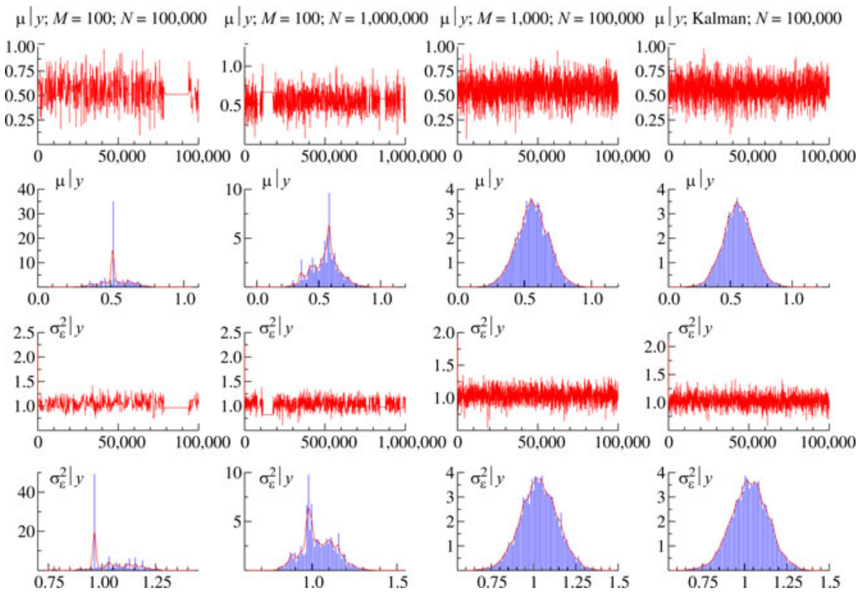

FIGURE 3. Gaussian linear model; particle chains and histograms for parameters μ and σ_ϵ^2; first column: particle filter with $M = 100$ and $N = 100{,}000$; second column: $M = 100$ and $N = 1{,}000{,}000$; third column: $M = 1{,}000$ and $N = 100{,}000$; fourth column: exact likelihood MCMC with $N = 100{,}000$.

TABLE 3. Results from MCMC for the Gaussian linear model using $N = 100{,}000$ for the exact likelihood (Kalman filter) MCMC and the particle MCMC based on $M = 1{,}000$. When $M = 100$ the MCMC algorithm was very inefficient, and we set $N = 1{,}000{,}000$. Here μ_{mc} denotes the arithmetical mean, σ_{mc} the estimate of the Monte Carlo standard error of the posterior expectation, Pr the acceptance probability for that parameter, and inef the rate of inefficiency of the MCMC sampler.

	Exact likelihood MCMC					$M = 100$, $N = 10^6$					$M = 1,000$				
	μ_{mc}	σ_{mc}	Pr	var	inef	μ_{mc}	σ_{mc}	Pr	var	inef	μ_{mc}	σ_{mc}	Pr	var	inef
μ	0.562	0.001	0.40	0.015	7.2	0.553	0.002	0.02	0.013	531	0.562	0.001	0.25	0.014	7.8
σ_ϵ^2	1.030	0.003	0.40	0.012	39	1.033	0.003	0.02	0.010	855	1.031	0.003	0.25	0.012	29
ρ	0.783	0.001	0.39	0.001	37	0.785	0.001	0.02	0.001	775	0.783	0.001	0.25	0.001	30
σ_η^2	0.621	0.004	0.40	0.015	48	0.607	0.003	0.02	0.011	784	0.618	0.003	0.26	0.015	36

From the first column in Figure 3 we see that using only $M = 100$ particles appears to be insufficient for the length of the chain. 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 3 shows the results in this scenario. 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. It still gets stuck but the histograms start to look better. In the third column of Figure 3 we see the drastic improvement in the performance of the algorithm from increasing the number of particles to $M = 1{,}000$. Both parameter chains and histograms start looking very similar to the exact likelihood MCMC results.

Table 3 compares the output of the exact likelihood MCMC algorithm with that from the $M = 100$, $N = 1{,}000{,}000$ and that from the $M = 1{,}000$, $N = 100{,}000$ particle MCMC algorithm. 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.[8] To demonstrate that effect more clearly Figure 4 compares the evolution of the ACFs and likelihoods as we run through the setups $M = 100$ with

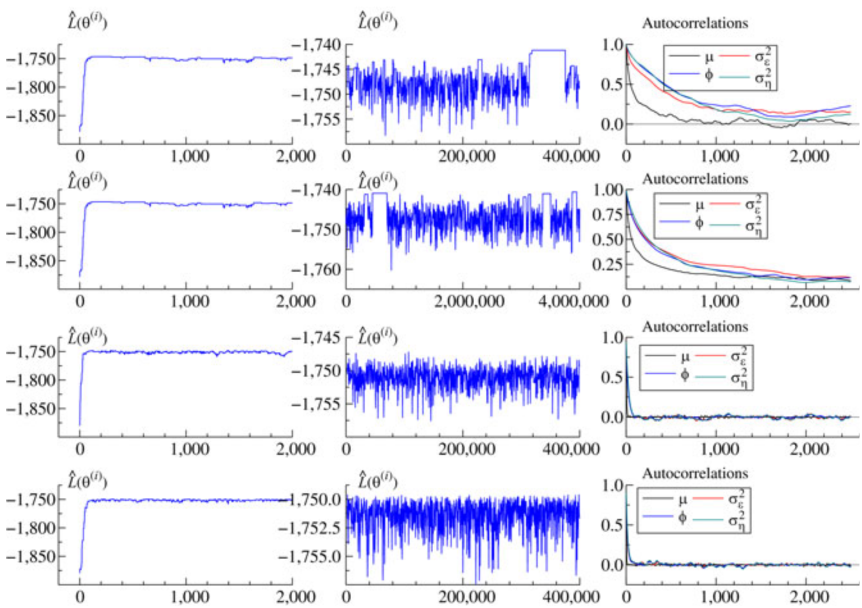

FIGURE 4. Gaussian linear model; first column: $\hat{L}(\theta^{(i)})$ for $i = 1, \ldots, 2{,}000$; second column: $\hat{L}(\theta^{(i)})$ for $i = 2{,}001, \ldots, 4N$; third column: ACF of $\theta^{(i)}$; first row: $M = 100$ with $N = 100{,}000$ particle filter; second row: $M = 100$ with $N = 1{,}000{,}000$; third row: $M = 1{,}000$ with $N = 100{,}000$; fourth row: exact likelihood MCMC with $N = 100{,}000$.

$N = 100{,}000$, $M = 100$ with $N = 1{,}000{,}000$ and $M = 1{,}000$ with $N = 100{,}000$ with the exact likelihood MCMC output.

A short comment on how we suggest choosing M and the variances of the random walk proposals. We find that a good indication of 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.[9] 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 great deal 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.

Computer time for the exact likelihood MCMC was roughly 2,100 seconds, for the estimated likelihood with $M = 100$ it was roughly 5,300 seconds, for $M = 1{,}000$, 40,000 seconds, and for $M = 100$ with $N = 10^6$, 52,400 seconds.

6. EXAMPLE: DISCRETE TIME GAUSSIAN STOCHASTIC VOLATILITY MODEL

We now turn to a simple real life problem and estimate the Gaussian discrete time stochastic volatility model (see, e.g., the reviews in Ghysels, Harvey, and Renault, 1996; Shephard, 2005). The stock returns and stochastic volatility factor are assumed to follow the processes

$$\begin{aligned} y_t &= \mu + \exp\{\beta_0 + \beta_1 \alpha_t\}\varepsilon_t, \\ \alpha_{t+1} &= \phi\alpha_t + \eta_t \end{aligned} \qquad \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \overset{i.i.d.}{\sim} N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

where $\alpha_0 \sim N\left(0, \left(1-\phi^2\right)^{-1}\right)$. Now we have $\theta = (\mu, \beta_0, \beta_1, \phi, \rho)'$. For this model the likelihood is not available. Researchers have used MCMC, SML estimation, method of moments, and indirect inference to estimate this type of model. See the reviews mentioned previously 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: $\Delta\mu_i = 0.017\nu_{1,i}$, $\Delta\beta_{0,i} = 0.104\nu_{2,i}$, $\Delta\beta_{1,i} = 0.010\nu_{3,i}$, $\Delta\phi_i = 0.004\nu_{4,i}$, $\Delta\rho_i = 0.067\nu_{5,i}$, where $\nu_{j,i} \overset{i.i.d.}{\sim} N(0,1)$ for $j = 1, \ldots, 5$ and $i = 1, \ldots, N$.

We use 3,271 daily observations from 03.01.1995 until 31.12.2007 of the end-of-day level of the SNP500 Composite Index (NYSE/AMEX only) from the Center for Research in Security Prices. Daily returns are defined as $y_t = 100\left(\log \mathrm{SNP500}_t - \log \mathrm{SNP500}_{t-1}\right)$. The statistics are displayed in Table 4. The most interesting aspect of the estimated parameters is the extremely strong negative

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. Here inef denotes the rate of inefficiency of the MCMC sampler, Pr denotes the acceptance probability for that parameter, and σ_{mc} denotes the estimate of the Monte Carlo standard error of the posterior expectation.

	μ_{mc}	σ_{mc}	Pr	Posterior standard deviations and correlation					Inef
μ	0.042	0.000	0.410	0.014	−0.686	−0.137	0.304	0.145	15
β_0	−0.141	0.001	0.395		0.079	0.061	−0.222	−0.044	12
β_1	0.080	0.000	0.398			0.007	−0.688	−0.015	18
ϕ	0.982	0.000	0.424				0.004	−0.054	16
ρ	−0.742	0.000	0.427					0.042	6.4

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.

Both y_t and the filtered time series for daily volatility $\mathrm{E}\{\exp(\beta_0 + \beta_1 \alpha_t) \mid \mathcal{F}_t\}$, estimated by using the parameters' posterior means, and the parameter histograms are plotted in Figure 5.

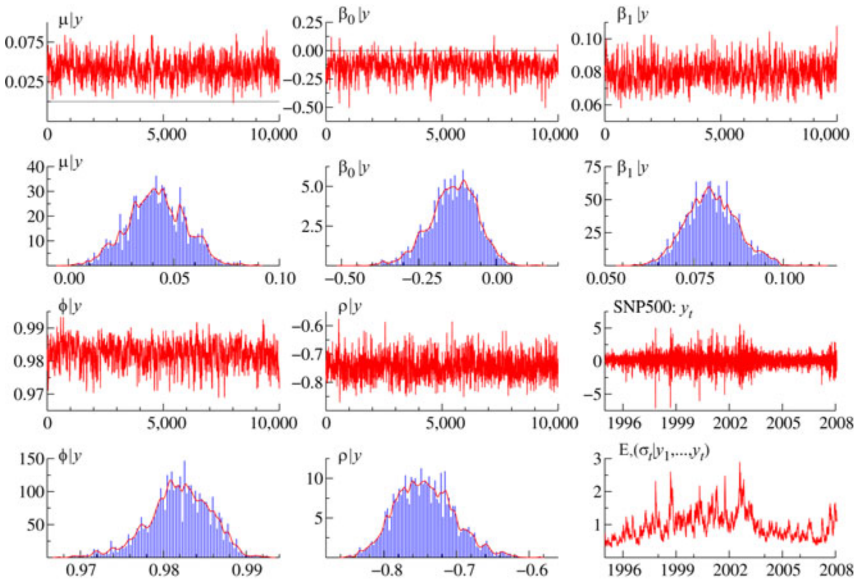

FIGURE 5. Stochastic volatility model; particle filter with $M = 2{,}000$: $\theta^{(i)}$ for $i = 1, \ldots, 100{,}000$ and histogram of parameters for $i = 50{,}000, \ldots, 100{,}000$. Bottom right corner: log-returns on the SNP500 Composite Index and $\mathrm{E}\{\exp(\beta_0 + \beta_1 \alpha_t | \mathcal{F}_t)\}$ based on posterior means of the parameters.

This analysis is based on $M = 2{,}000$ and $N = 100{,}000$. Convergence looks quite 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.[10] 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.[11] It is important to notice that the particle MCMC approach is currently the only feasible approach to estimating parameters of this type of model.

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

$$\max_{\{C_t,L_t\}_{t=0}^{\infty}} \mathrm{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left\{ \log(C_t) + \psi \log(1 - L_t) \right\} \right], \qquad \beta \in (0,1), \qquad \psi > 0,$$

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

There is one single good that 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 assume this to be a closed economy without government and hence with aggregate resource constraint $C_t + I_t = Y_t$. 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: $\log A_t = \rho_a \log A_{t-1} + \sigma_a \eta_{a,t}$ and $\log U_t = \rho_u \log U_{t-1} + \sigma_u \eta_{u,t}$, where $\eta_{a,t}, \eta_{u,t} \overset{i.i.d.}{\sim} N(0,1)$.

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:

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

subject to the resource constraints. The first-order equilibrium conditions

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

together with the resource constraint $C_t = A_t K_t^{\alpha} L_t^{1-\alpha} + \frac{1}{U_t}((1-\delta) K_t - K_{t+1})$ and the technology processes, fully characterize the solution to the problem.

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

$$(C_t, L_t)' = g(K_t, A_t, U_t),$$
$$(K_{t+1}, A_{t+1}, U_{t+1})' = h(K_t, A_t, U_t) + \sigma \begin{pmatrix} 0 & 0 \\ \sigma_a & 0 \\ 0 & \sigma_u \end{pmatrix} \eta_{t+1}, \qquad \eta_{t+1} = \begin{pmatrix} \eta_{a,t+1} \\ \eta_{u,t+1} \end{pmatrix},$$

where σ is the perturbation parameter. We find a second-order approximation to these policy functions by perturbation methods; see, for example, Judd (1998). We solve the system in terms of log-deviations from a nonstochastic steady state and use notation $\hat{c}_t = \log C_t / C_{ss}$, where C_{ss} denotes the nonstochastic steady state.

Now unify notation by letting $\hat{x}_t = \left(\hat{k}_t, \hat{a}_t, \hat{u}_t\right)'$. The solution will be of the state space form

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

and

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

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 variables are given by

$$G\hat{D}P_t = \hat{y}_t + \sigma_{\epsilon_y}\epsilon_{y,t}, \qquad \hat{H}_t = \hat{l}_t + \sigma_{\epsilon_l}\epsilon_{l,t}, \tag{2}$$

where $G\hat{D}P_t$ represents detrended real gross domestic product per capita and $\hat{H}_t$ represents hours worked per capita. From our model we compute these as

$$\hat{y}_t = \hat{a}_t + \alpha \hat{k}_t + (1-\alpha)\hat{l}_t, \qquad \hat{l}_t = g_{x,2}\hat{x}_t + \frac{1}{2}\hat{x}_t' g_{xx,2}\hat{x}_t + \frac{1}{2} g_{\sigma\sigma,2}\sigma^2.$$

Equations (1) together with the observation equations (2) specify a nonlinear state space system, from which we can easily simulate and hence use the particle filter to evaluate the likelihood of the model $\hat{L}(\theta)$. 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 MCMC algorithm. The algorithm to obtain $\hat{L}(\theta)$ inside the MCMC algorithm works by first computing $K_{ss}, A_{ss}, U_{ss}, C_{ss}, L_{ss}$ given $\theta^{(i)}$ and then using perturbation methods to find numerical values for $h_x, g_x, h_{xx}, g_{xx}, h_{\sigma\sigma}, g_{\sigma\sigma}$ and finally running the particle filter on the state space system (2) and (1) to obtain $\hat{L}\left(\theta^{(i)}\right)$.

We use this model to simulate the economy, with $T = 200$ and 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).

To guarantee positive standard errors in the MCMC algorithm we parameterize the log of the standard errors. We assume a Gaussian prior with mean θ_0, given by

ρ_a	ρ_u	$\log\sigma_{\epsilon_y}$	$\log\sigma_{\epsilon_l}$	$\log\sigma_a$	$\log\sigma_u$
0.95	0.90	$\log 0.003457$	$\log 0.005877$	$\log 0.006$	$\log 0.052$.

Any proposals for $\rho_a, \rho_u \notin (-1, 1)$ are automatically rejected. We use the random walk proposals $\Delta\rho_{a,i} = 0.0545\nu_{1,i}$, $\Delta\rho_{u,i} = 0.0077\nu_{2,i}$, $\Delta\log\sigma_{\epsilon_y,i} = 0.251\nu_{3,i}$, $\Delta\log\sigma_{\epsilon_l,i} = 0.159\nu_{4,i}$, $\Delta\log\sigma_{a,i} = 0.1125\nu_{5,i}$, $\Delta\log\sigma_{u,i} = 0.06\nu_{6,i}$ where $\nu_{j,i} \sim i.i.d.N(0,1)$ for $j = 1, \ldots, 6$ and $i = 1, \ldots, N$.

For the particle filter we chose $M = 60{,}000$ particles. We have to use such a large number of particles because $\hat{k}_t$ appears to be a slowly mixing process, which gives the particle filter a hard time. It will take a long period of time to "forget" a small mistake made at $t = 0$, and so the initial draw $\hat{k}_0$ 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 $\hat{x}_0^{(i)}$ 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.[12]

We produce results for $N = 10{,}000$. As usual we display the parameter histories with their histograms and the likelihood and the ACFs in Figure 6.

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

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.

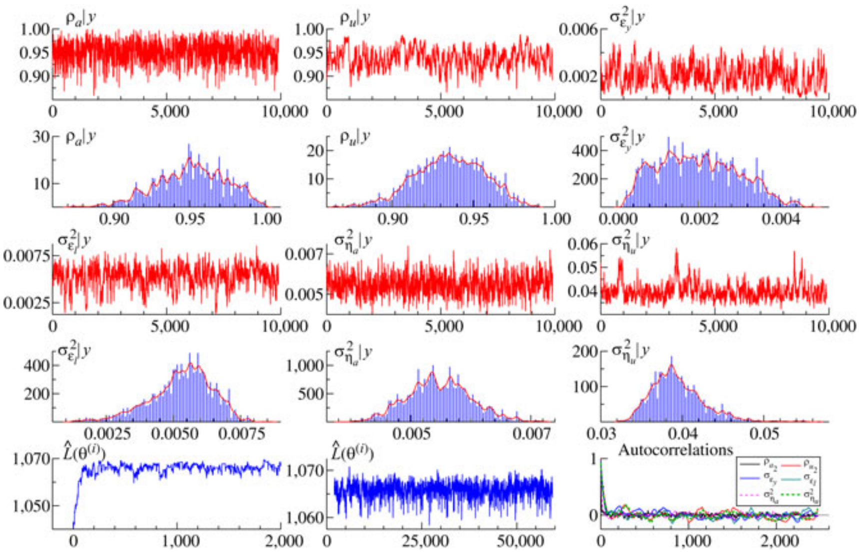

FIGURE 6. DSGE model; particle filter with $M = 60,000$: $\theta^{(i)}$ for $i = 1, \ldots, 10,000$ and histogram of parameters for $i = 5{,}000, \ldots, 10{,}000$. Last row: likelihoods and ACFs of parameters; left: $\hat{L}(\theta^{(i)})$ for $i = 1, \ldots, 2{,}000$; middle: $\hat{L}(\theta^{(i)})$ for $i = 2{,}001, \ldots, 6N$; right: ACF of $\theta^{(i)}$.

TABLE 5. Results from MCMC with particle filter; $N = 10{,}000$, $M = 60{,}000$

	μ_{mc}	σ_{mc}	Pr	inef	Posterior standard deviations and correlation					
ρ_a	0.9512	0.001	0.396	18	0.023	0.034	0.100	0.029	−0.209	0.047
ρ_u	0.9353	0.002	0.487	78		0.021	−0.137	0.111	0.001	0.668
σ_{ϵ_y}	0.0020	0.000	0.448	103			0.001	−0.614	−0.291	−0.057
σ_{ϵ_l}	0.0053	0.000	0.424	22				0.001	−0.198	0.012
σ	0.0055	0.000	0.386	32					0.001	−0.004
σ_u	0.0394	0.000	0.421	54						0.003

8. CONCLUSION

In the econometric literature estimated likelihoods are sometimes used as the basis for approximate maximum likelihood estimation. Such SML estimators have a number of shortcomings, as emphasized in the literature. In this paper we note that the effect of estimation can be removed by replacing the maximization of the likelihood by placing the simulation-based 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 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 quite 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 MCMC of dynamic models is not fast, but it is conceptionally simple and applies very widely in economics. We believe that it provides a simple generalization of the Kalman filter and can now handle many problems where estimation was previously cumbersome.

NOTES

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

2. Much Bayesian econometric work on discrete choice models has been carried out based on MCMC methods (e.g., Chib and Greenberg, 1998). These use different auxiliary variables, chosen to reflect the problem at hand. Typically they will be more computationally efficient than the methods we are advocating here.

3. An altogether different approach, which has been suggested by one referee, avoids changing σ_ε by estimating p_t using importance sampling. This potentially improves the performance of the algorithm for fixed computational complexity.

4. Because the $\theta^{(i)}$ are highly correlated we use the Newey–West heteroskedasticity and autocorrelation consistent estimator for the computation of the Monte Carlo standard error (σ_{mc}). The choice of the lag length B for this estimator is not obvious. The results reported are based on a choice of $B = 500$.

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

6. The convergence is pointwise: particle filters are not continuous with respect to θ, and so neither is $l_T(\theta)$. This is caused by the resampling step 3, which is very hard to overcome (Pitt, 2001; Flury, 2009).

7. This includes the resampling algorithm from Kitagawa (1996) that we use in this paper.

8. We do not report it here for brevity, but with $M = 2{,}000$ we achieve an acceptance probability of 30%.

9. As suggested by one referee, an alternative approach would be to set M to obtain a specified level of the variance of the log-likelihood estimate for a given θ. From practice however we notice that the variance can change quite a bit in the function of θ. For this reason we decided to base our approach to choosing M on the behavior of the Markov chain as it explores the parameter space.

10. The computer time for this model was roughly 540,000 seconds. Given recent developments in running particle filters on graphics cards (see Lee, Yau, Giles, Doucet, and Holmes, 2009) we believe these methods to be not too costly for even a simple desktop PC.

11. An and Schorfheide (2007) considered Bayesian inference for DSGE models. Fernandez-Villaverde and Rudio-Ramirez (2007) used particle filters to perform parameter inference. They only report SML estimates and indicate the possibility of using this present algorithm but provide no rationale for why it would work. Amisano and Tristani (2007) seem to use this particle MCMC algorithm to estimate a DSGE model but without any further justification.

12. We used Matlab for this model, and so computing time is not comparable with previous models. It took almost a week on our computer.

REFERENCES

Amisano, G. & O. Tristani (2007) Euro Area Inflation Persistence in an Estimated Nonlinear DSGE Model. ECB Working paper 754.

An, S. & F. Schorfheide (2007) Bayesian analysis of DSGE models. *Econometric Reviews* 26, 113–172.

Andrieu, C., N. de Freitas, & A. Doucet (1999) Sequential MCMC for Bayesian model selection. In *IEEE Higher Order Statistics Workshop, Ceasarea*, pp. 130–134.

Andrieu, C., A. Doucet, & R. Holenstein (2010) Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*. 72, 269–342.

Andrieu, C., A. Doucet, & G.O. Roberts (2007) The Expected Auxiliary Variable Method for Monte Carlo Simulation. Manuscript.

Aruoba, S.B., J. Fernandez-Villaverde, & J.F. Rubio-Ramirez (2006) Comparing solution methods for dynamic equilibrium economies. *Journal of Economic Dynamics & Control* 30, 2477–2508.

Cérou, F., P. Del Moral, & A. Guyader (2008) A Non Asymptotic Variance Theorem for Unnormalized Feynman-Kac Particle Models. Technical report HAL-INRIA RR-6716, Institut national de recherche en informatique et en automatique.

Chib, S. (2001) Markov chain Monte Carlo methods: Computation and inference. In J.J. Heckman & E. Leamer (eds.), *Handbook of Econometrics*, vol. 5, pp. 3569–3649. North-Holland.

Chib, S. & E. Greenberg (1998) Analysis of multivariate probit models. *Biometrika* 85, 347–361.

Del-Moral, P. (2004) *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer-Verlag.

Diggle, P. J. & 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, & N. J. Gordon (eds.) (2001) *Sequential Monte Carlo Methods in Practice*. Springer-Verlag.

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

Fearnhead, P. (2002) MCMC, sufficient statistics and particle filter. *Journal of Computational and Graphical Statistics* 11, 848–862.

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

Fernandez-Villaverde, J., J. F. Rudio-Ramirez, & M. S. Santos (2006) Convergence properties of the likelihood of computed dynamic models. *Econometrica* 74, 93–119.

Flury, T. (2009) Smooth Particle Filters and Their Use in Economics. Manuscript, University of Oxford.

Gallant, A.R., H. Hong, & 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. & G. Tauchen (1996) Which moments to match. *Econometric Theory* 12, 657–681.

Gelman, A., J.B. Carlin, H. S. Stern, & D. B. Rubin (2003) *Bayesian Data Analysis*, 2nd ed. Chapman and Hall.

Ghysels, E., A.C. Harvey, & E. Renault (1996) Stochastic volatility. In C.R. Rao and G.S. Maddala (eds.), *Statistical Methods in Finance*, pp. 119–191. North-Holland.

Gordon, N. J., D. J. Salmond, & A. F. M. Smith (1993) A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE-Proceedings F* 140, 107–113.

Gourieroux, C. & A. Monfort (1996) *Simulation Based Econometric Methods*. Oxford University Press.

Gourieroux, C., A. Monfort, & E. Renault (1993) Indirect inference. *Journal of Applied Econometrics* 8, S85–S118.

Hajivassiliou, V. & 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 University Press.

Johannes, M. & N. Polson (2009) Particle filtering. In T.G. Andersen, R.A. Davis, J.-P. Kreiss, and T. Mikosch (eds.), *Handbook of Financial Time Series*, pp. 1015–1029. Springer-Verlag.

Johannes, M., N. Polson, & J. Stroud (2008) Learning about jumps and stochastic volatility: Filtering stochastic differential equations with jumps. *Review of Financial Studies*. Forthcoming.

Judd, K. (1998) *Numerical Methods in Economics*. MIT Press.

Kim, S., N. Shephard, & S. Chib (1998) Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–393.

Kitagawa, G. (1996) Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics* 5, 1–25.

Klein, P. (2000) Using the generalized Schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics & Control* 24, 1405–1423.

Lee, A., C. Yau, M. B. Giles, A. Doucet, & C. C. Holmes (2009) On the Utility of Graphics Cards to Perform Massively Parallel Simulation with Advanced Monte Carlo Methods. Working paper, Oxford-Man Institute of Quantitative Finance, University of Oxford.

Lerman, S. & C. Manski (1981) On the use of simulated frequencies to approximate choice probabilities. In C. Manski & D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 305–319. MIT 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, 765–799.

Pitt, M. K. (2001) Smooth Particle Filters for Likelihood Maximisation. Manuscript, Department of Economics, Warwick University.

Pitt, M. K. & N. Shephard (1999) Filtering via simulation: Auxiliary particle filter. *Journal of the American Statistical Association* 94, 590–599.

Schmitt-Grohe, S. & 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 University Press.

Smith, A.A. (1993) Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics* 8, S63–S84.

Storvik, G. (2002) Particle filters in state space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing* 50, 281–289.

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