Sequential Monte Carlo Estimation of DSGE Models[1]

Hao Chen

Fuqua School of Business, Duke University

100 Fuqua Drive, Box 90120, Durham, NC 27708-0120

hao.chen@duke.edu

Francesca Petralia

Department of Statistical Sciences, Duke University

214 Old Chemistry Building, Durham, NC 27708-0251

francesca.petralia@duke.edu

and

Hedibert F. Lopes

University of Chicago Booth School of Business

5807 South Woodlawn Avenue, Chicago, IL 60637-1656

hlopes@chicagobooth.edu

This Draft: June 2010

Abstract

The use of particle filters in the estimation of non-linear and/or non-normal DSGE models has focused on the evaluation of the likelihood function (see Fernández-Villaverde and Rubio-Ramírez (2005)). Under this approach, Bayesian inference on fixed model parameters is carried out through hybrid algorithms that combine Markov chain Monte Carlo with embedded sequential Monte Carlo procedures for likelihood approximation. In this paper, we use full SMC methods to estimate non-linear DSGE models in an on-line fashion. SMC algorithms allow a simultaneous filtering of time-varying state vectors and estimation of fixed parameters. We first establish empirical feasibility of the full SMC approach by comparing estimation results from both MCMC batch estimation and SMC on-line estimationa on a simple neoclassical growth model. We then estimate a large scale DSGE model for the Euro area developed in Smets and Wouters (2003) with a full SMC approach, and revisit the on-going debate between the merits of reduced form and structural models in the macroeconomics context by performing sequential model assessment between the DSGE model and various VAR/BVAR models.

Keywords: Sequential Monte Carlo Methods, On-line Bayesian Estimation, DSGE Models, Unobserved State Variables, Sequential Model Comparison

[1]Contact author information: Hedibert F. Lopes, The University of Chicago Booth School of Business. E-mail: hlopes@chicagobooth.edu.

1 *Introduction*

The formal estimation of dynamic stochastic general equilibrium models has generated a steady stream of literature over recent years. Proposed estimation methods have evolved from earlier ad hoc techniques to the latest Bayesian simulation based methods utilizing Markov chain Monte Carlo algorithms (see Fernández-Villaverde (2009) for an in-depth review on this subject). We would like to push the frontier of this line of research by presenting a full SMC approach to the estimation of dynamic equilibrium models.

DSGE models are characterised by a likelihood function for the observables, which in turn is driven by a set of structural parameters representing preferences and technology in the modelled economy. It is therefore natural to think about a likelihood-based estimation technique for these models. The challenge however, is that the likelihood of a DSGE model is a high dimensional object littered with local extrema and flat surfaces. Moreover, most DSGE models of interest do not have closed form solutions, thus depriving the researcher of an analytical form for the likelihood function. Finding better solution methods for DSGE models has always been an actively pursued area of research, although this will not be the focus of our paper. For our purposes, there already exist several classes of numerical solution methods well documented by Aruoba, Rubio-Ramírez and Fernández-Villaverde (2006) that we could use straight off the shelves.

With a proper solution method, one can find the approximated policy functions of the DSGE model of interest and construct a likelihood function from the state space representation of the model. Before Fernández-Villaverde and Rubio-Ramírez (2005) (henceforth FVRR), the typical thing to do would be to solve the linearized model with normal shocks, and then evaluate the resulting likelihood using the Kalman filter. FVRR showed that one could use a sequential Monte Carlo filter to evaluate the likelihood, which can now be constructed from a nonlinear and non-normal state space representation, resulting in improved model fit. The latest Bayesian estimation technique for DSGE models is thus to embed some type of subroutine capable of extracting latent signals inside an MCMC method, typically the Metropolis-Hastings algorithm, which is used to obtain the posterior distributions of the structural parameters.

The latest development in the SMC literature demonstrated that SMC filters could not only be used for state-filtering, but parameter learning as well. The idea of using SMC methods to estimate parameters is not new, Liu and West (2001), Storvik (2002), Fearnhead (2002) and Fearnhead and Clifford (2003) have shown the concept is viable, while Gilks and Berzuini (2001), Carvalho, Johannes, Lopes and Polson (to appear) and Johannes and Polson (2008) have proposed several practical SMC-based estimation methods. Combining several ideas from those papers, we present an SMC estimation method for joint state and parameter learning on DSGE models.

So why use SMC when MCMC is proven to work? First of all, MCMC methods rely on Markov chain convergence, which might not be geometrically ergodic for structural models such as DSGE models (see Papaspiliopoulos and Roberts, 2008). Chib and Ramamurthy

(2010) proposed an MCMC method that features random parameter clustering and taylored proposal distributions which aims to address exactly the problem of slow mixing when performing MCMC on complex models with high dimensional parameter spaces. While their method appears to be a decent remedy to slow MCMC convergence with DSGE models, it's still an MCMC method, and thus doesn't offer the benefits inherent to SMC methods. Second, by nature of SMC methods, we can obtain posterior approximations of the parameters and states at each time period, which allows us to perform on-line estimation. To obtain the same amount of information with MCMC, one would have to resort to repeated implementation of MCMC at each time period, which is more inefficient in terms of running time and computing resources. Perhaps the biggest advantage of being able to perform on-line estimation is that it allows us to compute the marginal likelihoods of the model at each time point very easily, thus making it possible to perform model comparison sequentially. Giacomini and Rossi (2007) showed that the performance of the models might be time varying in an environment characterized by instability and model misspecification. A dynamic comparison between different models could be useful in choosing sequentially through time the best model and provide important information about the data generating process. Finally, SMC methods are easiliy parallelizable without additional coding effort in contrast to MCMC based methods. As the estimation of DSGE models involves computationally intensive numerical methods to solve the model, this is helpful in that it allows practitioners to combine posterior particles from different computers to form a more accurate approximation of the parameter posteriors.

The goal of this paper is to introduce SMC based methods to the empirical macroeconomics literature as a viable alternative to MCMC estimation. We provide two examples to this end, we first compute and estimate the benchmark dynamic equilibrium economy, the stochastic neoclassical growth model. After we solve the model with a second order approximation of the policy functions, we estimate it using both the MCMC with nonlinear filter approach from Ramírez et al (2004) and a full SMC method utilizing parameter sufficient statistics. In this example, we show that an SMC estimation method could provide comparable DSGE estimation results to MCMC methods. In our second example, we first apply SMC estimation to the DSGE model for the Euro area. We then show that the marginal likelihoods for any sequentially estimated model could be obtained at each period as a by-product of the SMC estimation procedure, which allows us to compute bayes factors between the estimated DSGE model and various reduced-form models in an on-line fashion. This could be a useful tool in the on-going investigation of whether to use deep structural models based on economic theory such as the DSGE or pure statistical models such as VAR or BVAR.

The rest of the paper is organized as follows. Section 2 presents the two types of Bayesian estimation methods we wish to compare on DSGE type structural models as well as estimation results for the neoclassical growth model from both mothods. Section 3 discusses the potential usefulnessness of sequential model comparison by comparing a DSGE model in the Euro area with various BVAR models. In section 5 we present our concluding remarks and possible future work directions.

2 Macroeconomic models

2.1 A Neoclassical Growth Model

The neoclassical growth model consists of the following relationships: a single good in the ceconomy is produced according to the production function $q_t = e^{z_t} k_t^{\alpha}$, where q_t is output, k_t is aggregate capital and z_t is a stochastic process modelling technological process. The law of motion for capital is $k_{t+1} = (1-\delta)k_t + i_t$, where δ is the depreciation rate and i_t is the investment at time t. There is a representative agent in the economy, who decides how much to consume in order to maximize his expected utility function:

$$E_0 \sum_{t=1}^{\infty} \beta^t \frac{c_t^{1-\tau}}{1-\tau}$$

where β is the discount factor, E_0 is the conditional expectation operator. Let us assume that technology evolves according to a stationay (i.e. $|\rho| < 1$) AR(1) process, i.e. $z_{t+1} = \rho z_t + \epsilon_t$ where ϵ_t is from $N(0, \sigma^2)$. We are interested in making inference on $\theta = [\alpha, \beta, \delta, \rho, \sigma, \tau]$, the six structural parameters involved in the model. In order to do that we have to first solve the maximization problem that is fully characterized by the following equilibrium conditions:

$$c_t^{-\tau} = \beta E_0 c_{t+1}^{-\tau} [\alpha e^{z_{t+1}} k_{t+1}^{\alpha-1} + 1 - \delta], \tag{1}$$

$$c_t + k_{t+1} = q_t + k_t(1-\delta), \tag{2}$$

$$z_t = \rho z_{t-1} + \epsilon_t. \tag{3}$$

The first equation is the intertemporal Euler condition relating current and future marginal utility of consumption. The second equation is the resource constraint of the economy and the third equation is the law of motion for technology. Solving for the equilibrium of the above model means finding the policy functions for consumption $c(\cdot,\cdot)$ and next period's capital $k'(\cdot,\cdot)$ which gives the optimal solution to the above utility maximisation problem as functions of the state variables, namely capital and technology.

The above system of equations do not have a closed form solution, thus must be solved with numerical methods. Aruoba, Fernández-Villaverde, and Rubio-Ramírez (2006) gave a nice comparison of solution methods for dynamic equilibrium models. For our estimation purpose, we use a second order perturbation method since it offers a nice balance between accuracy, speed and ease of programing. Pioneered by Hall (1971) and Magill (1977) and extended by Judd and Guu (1993) and Gaspar and Judd (1997), this class of solution methods build a Taylor series expansion of the policy functions around the steady state of the economy and a perturbation parameter. We follow Judd and Guu (2001) and use the standard deviation of the normal shock to technology as the perturbation parameter. The following equations are the resulting second order approximation of the policy functions around the non-stochastic steady state (See

Schmitt-Groh and Uribe (2004) for more details):

$$\hat{c}_t = \alpha_1 \hat{k}_t + \alpha_2 z_t + \alpha_3 \hat{k}_t^2 + \alpha_4 z_t^2 + \alpha_5 \hat{k}_t z_t + \alpha_6 \sigma^2 \tag{4}$$

$$\hat{k}_{t+1} = \beta_1 \hat{k}_t + \beta_2 z_t + \beta_3 \hat{k}_t^2 + \beta_4 z_t^2 + \beta_5 \hat{k}_t z_t + \beta_6 \sigma^2 \tag{5}$$

where the $\hat{}$ above a variable denotes its log deviation from the steady state. To find α_i and β_i in the above second order approximation, we plug the policy functions (8) and (9) into the equilibrium conditions (5)-(7) and take successive derivatives with respect to k, z and σ and set those derivatives to zero. This generates a system of equations on the unknown coefficients α_i and β_i which is trivial to solve.

It is important to note that the state space representation of the model depends crucially on the solution method. In particular if we use a second order approximation of the policy functions, then equation 6 and 7 are the transition equations of the state variables and 8 and 9 the measurement equations of the observables.

$$\hat{k}_{t+1} = \beta_1 \hat{k}_t + \beta_2 z_t + \beta_3 \hat{k}_t^2 + \beta_4 z_t^2 + \beta_5 \hat{k}_t z_t + \beta_6 \sigma^2 \tag{6}$$

$$z_t = \rho z_{t-1} + \epsilon_t \tag{7}$$

$$q_t = e^{z_t} k_t^{\alpha} + \epsilon_q \tag{8}$$

$$i_t = k_{t+1} - k_t(1-\delta) + \epsilon_i \tag{9}$$

where ϵ_q and ϵ_i are normally distributed with mean zero and standard deviation σ_y and σ_i.

Once the equilibrium model is solved and we obtain the above state-space representation, the SMC estimation algorithm could be applied. We note that techonolgy follows a stationay AR(1) process and thus the structural parameters (ρ, σ) have a readily available conditional sufficient statistics structure to exploit. More specifically, we have

$$p(\rho | x_{1:t}, y_{1:t}, \sigma) \sim N(b, \sigma B^{-1}), \tag{10}$$

$$p((\sigma^2)^{-1} | x_{1:t}, y_{1:t}) \sim Ga((t-2)/2, R(b)/2), \tag{11}$$

where $b_t = B_t^{-1} \sum_{k=2}^t z_k z_{k-1}$, $B_t = \sum_{k=2}^t z_{k-1}^2$ and $R_t(b) = \sum_{k=2}^t (z_k - b_t z_{k-1})^2$. The proposed SMC algorithms thus apply to the neoclassical growth model with $x_t = (k_t, z_t)$, $y_t = (q_t, i_t)$, $\phi = (\alpha, \beta, \delta, \tau, \sigma_y, \sigma_i)$, $\varphi = (\rho, \sigma)$ and $s_t = (b_t, B_t, R_t)$.

2.2 DSGE model for the euro data

3 Posterior inference

The main problem related to estimating DSGE models is the evaluation of the likelihood. Before FVRR, the general approach was to consider a linearized version of the model with normal shocks, and then evaluate the resulting likelihood using Kalman filter. With the advent of SMC methods we can now deal with nonlinear and non-normal dynamic general equilibrium models. In particular FVRR used a combination of Sequential Monte Carlo and Markov Chain Monte Carlo where the former is mainly used to filter through the model states for evaluation of the likelihood function. We eliminate the MCMC step using a full SMC approach to perform simultaneous estimation of time-varying state vectors and fixed parameters. The main advantage of our method is the access to model marginal likelihood and posterior distributions over real time. In this section we first describe both the SMC filter within MCMC approach and the full SMC approach, we then estimate the stochastic neoclassical growth model with both approaches.

3.1 SMC-within-MCMC approach

General state-space models are characterized by the following observation and state evolution equations

$$\begin{aligned} y_{t+1} &\sim p(y_{t+1}|x_{t+1},\theta) \\ x_{t+1} &\sim p(x_{t+1}|x_t,\theta), \end{aligned}$$

where we observe the $\mathcal{Y}$-valued process $\{Y_t\}_{t=1}^T$,which is driven by the $\mathcal{X}$-valued latent Markov process $\{X_t\}_{t=1}^T$, and the model parameters are collected in θ. The initial state distribution and parameter prior are denoted by $p(x_0|\theta)$ and $p(\theta)$, respectively. A Metropolis-Hasting based estimation requires the evaluation of the model likelihood

$$p(y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|y_{1:t-1},\theta). \tag{12}$$

A DSGE model has nonlinear likelihood, thus two types of likelihood evaluation methods are used for MCMC based estimation methods such as Metropolis-Hasting. One could either linearize the state and observation equations and use the Kalman filter or use a bootstrap filter to get a particle approximation of the likelihood without imposing any linearity assumption on the model. FVRR compared both methods of likelihood evaluation and concluded that an SMC likelihood evaluation is superior to a Kalman filter evaluation in the sense of better model fit. We will briefly describe the MH algorithm with SMC likelihood evaluation below.

The general idea of the procedure is to use a particle filter to evaluate the likelihood function of the model for each MCMC draw of θ, thereby obtaining the following particle approximation of the likelihood,

$$p(y_{1:T}|\theta) \approx p^N(y_{1:T}|\theta) \;\; = \;\; \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p(y_t|y_{1:t-1}, x_t^{(i)}, \theta)\delta_{x_t^{(i)}}(x_t),$$

where N denotes the number of particles used and $\delta_{x_0}(x)$ is the Dirac delta mass centered at x_0. At each time step, the conditional model likelihood

$$p(y_t|y_{1:t-1}, \theta) = \int p(x_{t-1}|y_{1:t-1}, \theta)p(x_t|x_{t-1}, \theta)p(y_t|x_t, \theta)dx_{t-1:t} \tag{13}$$

is approximated by $\frac{1}{N}\sum_{i=1}^{N} p(y_t|y_{1:t-1}, x_t^{(i)}, \theta)\delta_{x_t^{(i)}}(x_t)$.

This particle approximation of the model likelihood is then used in a Metropolis-Hasting kernel to search the parameter space (after specifying some priors on the parameters). Note that the use of SMC in this case is of state filtering only since θ is considered a fixed value inside the bootstrap filter. The particle filter is essentially used to average the conditional model likelihood over the simulated state paths $\{x_{0:T}^i\}_{i=1}^N$. A slightly modified bootstrap filter from Doucet, de Freitas and Gordon (2001) could fullfill the purpose of likelihood evaluation and is given as follows:

1. *Initialisation:* For $i = 1, \cdots, N$, sample $x_0^{(i)} \sim p(x_0|\theta)$ and set $t = 1$.

2. *Importance Sampling:* For $i = 1, \cdots, N$, sample $\tilde{x}_t^{(i)} \sim p(x_t|x_{t-1}^{(i)}, \theta)$, with importance weights proportional to $\tilde{w}_t^{(i)} \propto p(y_t|\tilde{x}_t^{(i)}, \theta)$.

3. *Resampling:* Resample with replacement N particles $(x_t^{(i)}; i = 1, \cdots, N)$ from the set $(\tilde{x}_t^{(i)}; i = 1, \cdots, N)$ according to the importance weights earlier. If $t \neq T$, set $t \leftarrow t + 1$ and go to step 2.

Note that the above algorithm does not store the particle paths $x_{0:t}^{(i)}$ since we are not doing inference on $p(x_{0:t}|y_{1:t}, \theta)$, and all we're interested in is using the importance weights $\tilde{w}_t^{(i)}$ to get a particle approximation of the model likelihood

$$p(y_{1:T}|\theta) \;\; \approx \;\; \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p(y_t|\tilde{x}_t^{(i)}, \theta)\delta_{\tilde{x}_t^{(i)}}(\tilde{x}_t).$$

Another point worth mentioning is that even though we are not using the particle filter for its original purpose of finding the joint or marginal distribution of the states, the resampling step is still needed to maintain algorithm stability over time. Without resampling, the variance of the importance weights will increase exponentially over time, thus leading to unstable likelihood computation. Doucet and Johansen (2008) has an excellant tutorial on particle filters that demonstrates in detail the necessity of the resampling step in any SMC method.

3.2 Parameter and state learning

SMC methods are commonly used in the analysis of time-series data where observation and hidden Markov state evolution form a coupled dynamic process. The major statistical challenge in SMC is the estimation of fixed parameters. In order to facilitate correct traversal of the parameter space, it is necessary to introduce artificial particle dynamics. One of the methods that have been successfully applied in the literature is the modified normal kernel move of Liu and West (2001) (henceforth LW). The main feature of LW is the use of a normal mixture smoothing kernel with location shrinkage to compensate for the "loss of information" in artificial parameter evolution. LW also uses the auxilliary particle filter (APF) developed by Pitt and Shepard (1999b), which essentially does a one-step look ahead before sampling new particles so that only particles more likely to be consistent with the new arriving observation evolve to become new particles for the next time step. Another SMC method viable for DSGE estimation is the resample-move algorithm from Gilks and Berzuini (2001) (henceforth GB). Their algorithm combines MCMC moves with importance sampling/resampling and has a computational advantage over LW when applied to DSGE estimation. We chose LW and GB as the base algorithms for DSGE estimation despite numerous recent developments such as Storvik (2002), Fearnhead (2002), Polson, Stroud and Mller (2008) and Carvalho, Johannes, Lopes, and Polson (2010) for several reasons which will be made clear in a moment. We do, however, use a couple of key ideas from those papers in combination with LW to achieve successful SMC estimation of DSGE models.

DSGE models are complex structural models where many model parameters enter the observation equations in a nonlinear fashion. In addition, there is no analytical form for the state evolution equations in all but the simplest DSGE models. Typically, one has to solve a system of stochastic partial differential equations defined by the model with numerical methods to obtain tractable state evolution equations, conditional on the model parameters. In this sense, model parameters set the 'structure' of the state-space form of the model, thus the name 'structural parameters'. With the exception of LW and RW, all of the remaining papers mentioned earlier have made the assumption that the model parameters have conditional sufficient statistics available given the state trajectories and data in order to exploit the benefits of incorporating sufficient statistics into their SMC methods. Due to the complex nature of DSGE models, many parameters simply don't have a conditional sufficient statistics structure to exploit, thus preventing us from using those methods relying solely on parameter sufficient statistics. GB makes no explicit assumption about model parameters having sufficient statistics, and samples parameter particles from a Markov chain transition kernel. The disadvantage of their approach is that it suffers from the curse of dimensionality, as demonstrated in Bengtsson, Bickel and Li (2007). We are, however, only interested in estimating a fixed set of parameters that does not grow in size and this is therefore not an issue for us. Polson, Stroud and Mller (2002) proposed a general algorithm that approximates the joint posterior of the states and the parameters by a mixture of fixed lag smoothing distributions. Without sufficient statistics, this approach requires the full history $(x_{0:t}, y_{1:t})$ for inference at time t. Therefore, it also suffers from increasing computation cost over time in the absense of sufficient statistics. The particle learning algorithm proposed in Carvalho, Johannes, Lopes, and Polson (2010) is similiar to GB in spirit but requires conditional sufficient statistics to move particles using Gibbs sampling. LW and GB thus emerge as the only sequential learning methods that could be applied to a general state-space model and at the same time only requiring state and parameter particles from the last time step to make inference in the current time step.

LW with Sufficient Statistics

Extending the key idea in West (1993a,b) of using mixtures to approximate posterior distributions, LW uses the following smooth kernel density to approximate $p(\theta|y_{1:T})$ given particles $\{\theta_t^{(i)}\}_{i=1}^N$ with weights $\{w_t^{(i)}\}_{i=1}^N$:

$$p(\theta|y_{1:T}) \approx \sum_{i=1}^{N} w_t^{(i)} N(\theta|m_t^{(i)}, h^2 V_t) \tag{14}$$

where $h > 0$ is a smoothing parameter to be set by the user and V_t is the Monte Carlo posterior variance. West (1993a,b) specifies the shrinkage rule for the kernel locations to be

$$m_t^{(i)} = a\theta_t^{(i)} + (1-a)\bar{\theta}_t \tag{15}$$

where $a = \sqrt{1-h^2}$. This normal mixture smoothing kernel coupled with location shrinkage is what gives LW the ability to do effective sequential parameter learning without sufficient statistics.

LW, however, is known to suffer from particle impoverishment. This is not so much a problem with the LW algorithm but rather a problem with SMC methods in general. There are two possible ways to mitigate this problem. One way is to incorporate sufficient statistics whenever possible as numerous papers cited earlier have shown this to be a proven mechanism of alleviating particle degeneracy. Let $\theta = (\phi, \varphi)$ where ϕ and φ denote the subsets of parameters without and with sufficient statistics conditional on the states, respectively. We assume that the posterior for the parameter vector φ is dependent on $(x_{0:T}, y_{1:T})$ only through some low dimensional sufficient statistics, such that

$$p(\varphi|x_{0:T}, y_{1:T}) = p(\varphi|s_t),$$

where $s_t = S(s_{t-1}, x_t, y_t)$ is a recursively defined sufficient statistics. Then for those parameters with sufficient statistics, particle evolution boils down to sampling from the exact posterior distributions $p(\varphi|s_t)$. This is synonymous with using MCMC moves within a SMC filter to address particle decay in the pure state filtering literature. Another way to help improve particle variability is to use a more efficient resampling scheme. Variability of particles can be assessed by a quantity known as the effective sample size (ESS), ESS at time t is defined as

$$ESS(t) = \left(\sum_{i=1}^{N} (w_t^{(i)})^2 \right)^{-1}.$$

Inference based on N weighted particles is equivalent to inference based on ESS particles from the exact target distribution. One could choose to only resample when the ESS falls below a certain threshold, which is typically set at $N/2$. In addition to using the adaptive resampling approach just mentioned, one could also entertain different resampling methods. Douc, Capp and Moulines (2005) provides a comparative study on some of the most popular resampling schemes. We use the multinomial resampling scheme for our SMC implementation. We propose the following algorithm for DSGE estimation:

1. *Initialisation*: For $i = 1, \cdots, N$, sample initial particles $(\phi, \varphi)^{(i)} \sim p(\theta)$, $x_0^{(i)} \sim p(x_0|\theta)$, set $t = 0$.

2. *APF Look-ahead:* For $i = 1, \cdots, N$, compute look-ahead particles $(\mu_{t+1}^{(i)}, m_t^{(i)})$ for $(x_{t+1}^{(i)}, \phi_t^{(i)})$ by $\mu_{t+1}^{(i)} = E(x_{t+1}|x_t^{(i)}, \phi_{t,}^{(i)} \varphi_t^{(i)})$ and $m_t^{(i)} = a\phi_t^{(i)} + (1-a)\bar{\phi}_t$, where $\bar{\phi}_t$ is the mean of particles $\{\phi_t^{(i)}\}_{i=1}^N$.

3. *Auxiliary resampling:* Sample auxiliary index set $k^{1:N} \sim Multinomial(N, g_{t+1}^{1:N})$ where $g_{t+1}^{(i)} \propto w_t^{(i)} p(y_{t+1}|\mu_{t+1}^{(i)}, m_t^{(i)}, \varphi_t^{(i)})$. Use adaptive resampling if necessary when resampling state and parameter particles later.

4. *Sample ϕ:* For $k = 1, \cdots, N$, sample $\phi_{t+1}^{(k)} \sim N(\cdot|m_t^{(k)}, h^2 V_t)$, where $m_t^{(k)}$ is the k^{th} component of the smooth kernel density in (3) and V_t is the Monte Carlo variance of particles $\{\phi_t^{(i)}\}_{i=1}^N$.

5. *State evolution:* For $k = 1, \cdots, N$, sample $x_{t+1}^{(k)} \sim p_{\phi_{t+1}^{(k)}}(x_{t+1}|x_t^{(k)}, \phi_{t+1}^{(k)}, \varphi_t^{(k)})$.

6. *(Optional) Sample φ* : For $k = 1, \cdots, N$, update the sufficient statistics particles by $s_{t+1}^{(k)} = S(s_t^{(k)}, x_{t+1}^{(k)}, y_{t+1})$, and then sample $\varphi_{t+1}^{(k)} \sim p(\cdot|s_{t+1}^{(k)})$.

7. *Update posterior weights:* The new posterior weights are given by

$$w_{t+1}^{(k)} \propto p(y_{t+1}|x_{t+1}^{(k)}, \phi_{t+1}^{(k)}, \varphi_{t+1}^{(k)})/g_{t+1}^{(k)}.$$

8. If $t \neq T$, set $t \leftarrow t + 1$ and go to step 2.

When applied to the estimation of a DSGE model, the model typically have to be solved at the end of step 4 to get the new state evolution equations based on the newly sampled parameter particles. This is reflected by the notation in step 5 of the above algorithm where the state propagation density is dependent on $\phi_{t+1}^{(k)}$. If one wishes to incorporate sufficient statistics, then ideally the model should be solved again at the end of step 6 to reflect the update of those parameters with sufficient statistics. We omit this step when we estimate the neoclassical growth model as solving a DSGE model is computationally expensive and we found no significant empirical improvement in the estimation results to justify the increase in computation time. Note that this could change depending on the model so we recommend users to try both and weigh the trade-offs between run time and quality of estimation. If applied to the DSGE model for the Euro area in Smets and Wouters (2003) (henceforth SW), the LW algorithm requires the model to be solved at least twice per particle at each time period, once at the end of step 2 and once at the end of step 4. The reason is that for the SW DSGE model, not only do the state equations depend on the structural parameters, but some of the observation equations do as well. In step 3, the likelihood function depends on m_t instead of ϕ_t, thus the observation equation must be re-solved to reflect this change. Since solving the model is the computational bottleneck in DSGE estimation, we turn to the SMC method from Gilks and Berzuini (2001) to estimate the SW model.

GB with Sufficient Statistics

Incorporating Markov chain moves into a SMC method is by now a fairly established method of dealing with the issue of particle degeneracy. The basic idea in Gilks and Berzuini (2001) is essentially a generalization of the particle learning algorithm proposed in Carvalho, Johannes, Lopes, and Polson (2010). To see this, set the sequence of target distributions in GB to $p(x_t, \theta|y_{1:t})$, then the GB resample weight at time $t+1$ becomes

$$\begin{aligned} w_{t+1}^{(i)} &= \frac{p(x_t^{(i)}, \theta_t^{(i)}|y_{1:t+1})}{p(x_i^{(i)}, \theta_t^{(i)}|y_{1:t})} \propto \frac{p(y_{1:t+1}|x_t^{(i)}, \theta_t^{(i)})p(x_t^{(i)}, \theta_t^{(i)}|y_{1:t})}{p(x_i^{(i)}, \theta_t^{(i)}|y_{1:t})} \\ &\propto p(y_{1:t+1}|x_t^{(i)}, \theta_t^{(i)}), \end{aligned}$$

which is exactly the resampling weight in PL. Since the two algorithms are practically the same, incorporating the use of sufficient statistics in GB simply means using a mix of Gibbs and Metropolis Hastings moves to evolve the particles. We modify the original GB algorithm slightly to accomodate the use of sufficient statistics, and as before let $\theta = (\phi, \varphi)$ where ϕ and φ denote the subsets of parameters without and with sufficient statistics conditional on the states, respectively. The proposed algorithm for DSGE estimation is as follows:

1. *Initialisation:* For $i = 1, \cdots, N$, sample initial particles $(\phi, \varphi)^{(i)} \sim p(\theta)$, $x_0^{(i)} \sim p(x_0|\theta)$, set $t = 0$.

2. *Importance resampling:* For $i = 1, \cdots, N$, compute $\tilde{x}_{t+1}^{(i)} \sim p(x_{t+1}|x_t^{(i)}, \phi_t^{(i)}, \varphi_t^{(i)})$, then compute $w_{t+1}^{(i)} \propto p(y_{t+1}|\tilde{x}_{t+1}^{(i)}, \phi_t^{(i)}, \varphi_t^{(i)})$ and sample index set $k^{1:N} \sim Mult(N, w_{t+1}^{1:N})$.

3. *Sample ϕ:* For $k = 1, \cdots, N$, Sample $\phi_{t+1}^{(k)} \sim q_t(\phi_t^{(k)})$.

4. *State evolution:* For $k = 1, \cdots, N$, sample $x_{t+1}^{(k)} \sim p(x_{t+1}|x_t^{(k)}, \phi_{t+1}^{(k)}, \varphi_t^{(i)})$.

5. *MH accept/reject:* Accept the particle set $(\phi_{t+1}^{(k)}, x_{t+1}^{(k)})$ with the usual MH importance ratio.

6. *(Optional) Sample φ:* For $k = 1, \cdots, N$, update the sufficient statistics particles by $s_{t+1}^{(k)} = S(s_t^{(k)}, x_{t+1}^{(k)}, y_{t+1})$, and then sample $\varphi_{t+1}^{(k)} \sim p(\cdot|s_{t+1}^{(k)})$.

7. If $t \neq T$, set $t \leftarrow t+1$ and go to step 2.

With the above algorithm, any DSGE model is usually solved at the end of step 3 to obtain new state evolution conditional on newly sampled parameter particles. In the presence of sufficient statistics, the model should ideally be solved again at the end of step 6 just as in the LW algorithm earlier. According to GB, the kernel q_t in step 3 has to be neither irreducible or reversible, so a wide range of Markov kernels are possible. We use a properly tuned random walk Metropolis kernel here to sample general model parameters. To reflect the conditional dependence between model parameter and states, we perform the accept/reject step after we evolve the current states based on the newly sampled parameters. If the particle set $(\phi_{t+1}^{(k)}, x_{t+1}^{(k)})$ is rejected in step 5, we replace it with the particle set $(\phi_t^{(k)}, \tilde{x}_{t+1}^{(k)})$. Note that with this algorithm, we need only solve the DSGE model once per particle at time t since we only need to compute one set of weights. This would translate to a significantly shorter run time compared to LW when estimating the DSGE model for the Euro area.

Sequential Bayes factor

With SMC estimation, we have at each time point t the particle approximation of the model posterior $p(x_t, \theta | y_{1:t}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{(x_t,\theta)^{(i)}}$. We can use the swarm of posterior particles $\left\{ x_t^{(i)}, \theta_t^{(i)} \right\}_{i=1}^{N}$ to approximate the marginal likelihood of the model by

$$p(y_{t+1}|y_{1:t}) \approx \frac{1}{N}\sum_{i=1}^{N} p(y_{t+1}|x_t^{(i)}, \theta^{(i)}).$$

This allows Bayes factors to be computed at each time point in the dataset as a by-product of the SMC estimation procedure as the quantities $p(y_{t+1}|x_t^{(i)}, \theta^{(i)})$ are already calculated by the estimation algorithm to use as importance weights. The Bayes factor for competing models M_1 and M_2 at time t is given by

$$\frac{p(M_1|y_{1:t})}{p(M_2|y_{1:t})},$$

where $p(M_i|y_{1:t}) = \prod_{k=1}^{t} p(y_k|y_{1:k-1}, M_i)$.

Convergence Results for SMC Methods

Numerous papers in the literature have investigated convergence properties of SMC methods. Chopin (2004) and Del Moral, and Guionnet (1999) developed central limit theorems for particle filters while Crisan and Doucet (2002) gave a nice survey on the subject. GB also developed a CLT for their algorithm, although it's only for the limiting case of the number of particles, and not for the number of observations. Those convergence results, however, all apply to models with fixed state-space representations. In DSGE models, the state evolution is dependent on the parameter values and each MCMC draw or SMC particle of the parameters will lead to a different state evolution. As such, no formal convergence study has been done on estimating DSGE models with SMC methods and this is possibly something of interest to researchers with a flair in theoretical work. Note that this is not an issue with MCMC estimation of DSGE models. This is because the states are essentially integrated out by the particle filter when evaluating the likelihood function, and so standard MCMC convergence results could be applied to DSGE estimation. On a related note, Fernández-Villaverde, Rubio-Ramírez, and Santos (2006) investigated the convergence properties of the likelihood of computed dynamic equilibrium models, followed by Ackerberg, Geweke, and Hahn (2009). These two papers not only provide insight on convergence studies of SMC methods applied to dynamic equilibrium models, but also contain discussions on the effects of policy function approximation errors on model parameter estimates, which are relevant to any empirical econometrician interested in this line of work.

4 Applications

4.1 Estimation of the Neoclassical Growth Model

We simulated 80 observations of i_t and q_t from the model considering the calibration in table 2. This calibration we use is taken from FVRR and is suppose to match empirical findings on real economies. We estimate this artificial dataset with both the MCMC with embedded state filter approach and the full SMC approach.
For the MCMC estimation, we ran a Metropolis Hastings algorithm for 20,000 iterations with a burn-in of 1000 considering the prior specification showed by table 2. In particular, these priors are those used in FVRR.

Parameter	Calibration	Priors
α	0.4000	Uniform(0.00,0.50)
β	0.9896	Uniform (0.75,1.00)
δ	0.0200	Uniform (0.00,0.05)
ρ	0.9500	Uniform(0.50,1.00)
σ	0.0070	Uniform(0.00,0.10)
σ_q	0.2000	Uniform (0.00,0.25)
σ_i	0.2000	Uniform (0.00,0.25)
τ	2.0000	Uniform(0.00,5.00)

Table 1: Calibration and Prior specification

The likelihood function is evaluated with a particle filter with a population of 5000 particles and is then used in a Metropolis-Hastings algorithm to sample from the parameter space. To get a sense of how the model likelihood behaves for a typical DSGE model, we plot in Figure 1 the likelihood profiles of various parameters. The dotted line is the likelihood function and the solid line is the true value. As we can see the likelihood is well behaved for for all parameters except for σ_q and σ_i.

Figure 2 shows the posteriors from the MCMC estimation. The vertical red line denotes the artificial parameter value we simulate from. For the full SMC estimation, we used 5,000 particles and we used sufficient statistics for ρ and σ. Figure 3 shows the posteriors of the parameters at the end of the observation horizon.

Discussion. Figure 4 shows the boxplot comparison of SMC and MCMC estimation results, in each figure, the left box is from SMC and the right is MCMC. The SMC results might look a bit misleading in that some of the posterior modes deviate far from the simulated parameter value, but a close inspection of the x-axes scale will reveal that this is not the case. SMC seems to be able to pick up all the parameter locations whereas MCMC does comparably with all parameters except for the two measurement errors. SMC estimation appears to have a wider confidence interval on some of the parameters as evident in the boxplot, this is because the number of particles we used to make posterior inference is one quarter of the MCMC samples used. To run the SMC estimation with the same number of particles as MCMC samples would translate to a much longer run time as the amount of computation would be equivalent to running a rolling window MCMC. One could, however, utilize parallel computing while using SMC methods to improve the quality of posterior inference.

	SMC				MCMC			
	Mean	.05	.5	.95	Mean	.05	.5	.95
α	0.3975	0.3381	0.3975	0.4594	0.3941	0.3825	0.3946	0.4068
β	0.9924	0.9855	0.9921	0.9994	0.9908	0.9877	0.9908	0.9938
δ	0.0115	0.0019	0.0112	0.0228	0.0187	0.0148	0.0186	0.0234
ρ	0.9500	0.9289	0.9501	0.9704	0.9430	0.9306	0.9430	0.9559
σ	0.0200	0.0172	0.0200	0.0234	0.0078	0.0044	0.0076	0.0134
σ_q	0.2120	0.1464	0.2167	0.2482	0.0478	0.0344	0.0473	0.0664
σ_i	0.2141	0.1492	0.2194	0.2485	0.0467	0.0302	0.0455	0.0688
τ	1.9636	0.1282	1.6662	4.7244	1.7643	1.4540	1.7758	2.0622

Table 2: Simulation Summaries

This example shows that we could obtain comparable estimation results using SMC and MCMC methods. However, SMC estimation of DSGE models offers many benefits. Figure 5 shows how the particle approximation of the posteriors for parameters β, α and ρ change over time. To obtain the same amount of information with MCMC would be extremely computationally intensive since MCMC needs to go through the entire dataset again with each new observation added. With a full SMC estimation procedure, adding a new observation only requires going through one loop of the earlier SMC algorithms, provided that the particles from the last time step were saved, which requires only $\mathcal{O}(N)$ space for inference with N particles[2].

MCMC methods rely on Markov chain convergence, and as a result parallel computation is typically not feasible. Advanced MCMC methods such as parallel tempering requires additional implementation which is time consuming. SMC methods, on the other hand, are parallelizable without much additional coding effort. And at any time t on any one machine, the particles

[2]This only holds for sequential inference on the marginal posterior $p(\theta, x_t|y_{1:t})$, inference on the full state trajectory $p(\theta, x_{1:t}|y_{1:t})$ requires $\mathcal{O}(tN)$ space.

$\{\theta_t^{(i)}\}_{i=1,\cdots,N}$ are approximately distributed as the marginal posterior $p(\theta|y_{1:t})$. We could therefore simply group particles from individual runs on different machines together according to their final weights to get a better approximate of the target posterior.

Lastly, we can compute model marginal likelihoods from the particles at each time period very easily, thereby allowing us to perform model comparison sequentially as data arrive. This could be useful as the relative performance of different models could be time varying as demonstrated in Giacomini and Rossi (2007). We will use this trait of SMC methods to compare model performance between a popular DSGE model and various reduced-form models in the next section.

4.2 DSGE Model for the Euro Area

It is known to researchers that macroeconomic data is often affected by structural instabilities. Such an environment could lead to different relative performance levels of competing models over time, which prompted a need to be able to compare models in a sequential manner. To this end, Giacomini and Rossi (2007) proposed two statistical tests based on local performance measures such as the Kullback-Leibler information criterion. One test is used to analyze the model's relative performance over observed samples, while the other is used to monitor the model's relative performance out of sample. We will show that the SMC counterpart to the tests in Giacomini and Rossi (2007) is a lot simpler, by virtue of the fact that we have access to particle approximations of marginal likelihood at each time period. For the demonstration, we used the DSGE data for the European area from Smets and Wouter (2003).

Smets and Wouters (2003) (henceforth SW) developed a DSGE model of the European economy with sticky prices and wages and estimated the linearlized model using MCMC sampling over the period 1970:1-1999:4 on seven key macroeconomic variables: GDP, consumption, investment, prices, real wages, employment and the nominal interest rate. In their model, households maximize a utility function with goods, money and leisure over an infinite horizon subject to budget constraints, act as price-setters in the labour market and choose how to best invest their capital stock. Firms engage in monopolostic competition in the intermediate goods market and the country produces a single final good used for consumption and investment by the households. The economy is in equilibrium if supply and demand are equal in the various markets (final goods, labour, etc). The linearized model has a total of ten structural shock variables, six of which follow independent first order autoregressive processes while the rest are i.i.d independent processes. SW concluded that the DSGE model fit the data as well as BVAR models since the model marginal likelihoods of the different models have comparable magnitudes. Their model marginal likelihoods were, however, computed over the entire sample, as typical in a Bayesian MCMC framework.

Giacomini and Rossi (2007) noted that certain events related to the European economy that occurred during the sample period, the creation of the European Union, for example, could have an impact over the relative performance of DSGE and BVAR models. While SW came to the conclusion that Bayesian batch model comparison favors DSGE over BVAR models of orders 1, 2 and 3, the frequentist tests from Giacomini and Rossi (2007) suggests that DSGE performs comparably to BVAR(1) and BVAR(2) for most of the dataset but outperforms both

BVAR models in the last 4 years of the sample[3].

SMC Estimation of SW Model. We estimated the SW DSGE model using the same detrended data from their paper with the GB algorithm described earlier and we used the popular DYNARE software to solve for the model equilibriums (see the appendix for details regarding DYNARE's usage as applied to the DSGE model is SW). Particle degeneracy was not a pronouced issue in this case so we use the Metropolis-Hastings kernel on all model parameters. Figure 7 shows the estimated posteriors at the end of the sample and Table 3 provides a comparison between the end of sample SMC posteriors with the MCMC posteriors from SW. As with the earlier example, more particles are needed to obtain a tighter confidence band for SMC estimation. Figure 5 shows the effective sample size of the SMC estimation with 1000 particles resampled at every time step. Most of the parameter estimates are close to SW's MCMC estimation results. Certain parameters were estimated to be different than SW, for example, standard errors for the productivity shock, labor supply shock and interest rate shock, and the adjustment costs for investment and capital utilisation. Given the small number of particles we used, perhaps it's necessary to constrain the posterior region of those parameters to facilitate more efficient particle coverage of the true posterior distribution.

SMC estimation with 1000 particles through the whole sample takes about 12 hrs to finish on a PC with a 3Ghz Core 2 Duo processor as the procedure is quite computationally intensive. It's fortunate that computation complexity increases linearly with number of particles, so estimation with 10,000 particles through the whole dataset would probably take around a week to finish. With particles on the order of tens of thousands, the quality of SMC estimation should improve significantly and be comparable with MCMC results from SW.

Model Comparison: Structural vs Reduced Form. SMC estimation methods are Bayesian methods that provide a particle approximation of the parameter posterior at each time period in the dataset, thus we could use results from an SMC estimation to compute Bayes factors for use in model comparison between fundamentally misspecified models in a sequential manner. We demonstrate this by using SMC estimation results on the SW DSGE model to compare it against with various VAR and BVAR models. We fit VAR and BVAR models with the same seven-dimensional data from SW using rolling window MCMC with 5000 MCMC interations. The VAR models are estimated from the VAR likelihood with Jefferys prior on the covariance matrix and the BVAR models are estimated with the Minnesota prior as in Litterman (1986), model marginal likelihoods for both VAR and BVAR are approximated by their harmonic means. For BVAR models, we used different values for the two hyperparameters in the Minnesota prior to investigate the sensitivity of BVAR performace to prior choice. Those two hyperparameters are λ, the prior standard deviation for coefficients on the first lag of the dependent variables, and ω, the variance discount factor for coefficients on variables other than the dependent variables.

Figure 8 shows the log Bayes factor of the VAR models against DSGE. Figure 9 shows the log Bayes factor of the BVAR models against DSGE when fixing the hyperparameter value

[3]Their study, however, showed that the pre-processing of data in SW favors the DSGE model over the reduced-form models. When they applied the same sequential test to data that's rolling-sample detrended instead of sample detrended, they found that BVAR(2) outperforms DSGE on all but a few points in the sample.

	SMC				SW - MCMC			
	5%	50%	75%	mean	5%	50%	75%	mean
σ productivity shock	0.0005	0.0097	0.0712	0.0204	0.444	0.612	0.873	0.628
σ inflation obj. shock	0.0004	0.0056	0.0543	0.0146	0.011	0.023	0.069	0.028
σ cons. pref. shock	0.1558	0.9560	1.7404	0.9286	0.173	0.297	0.571	0.324
σ gov. spending. shock	0.0596	0.1671	0.4985	0.2140	0.290	0.329	0.378	0.331
σ labor supply shock	0.0052	0.0364	0.3729	0.1434	0.997	1.658	2.603	1.709
σ interest rate shock	0.0321	0.0776	0.1503	0.0880	0.102	0.129	0.158	0.129
σ investment shock	0.0001	0.0102	0.5036	0.2272	0.099	0.129	0.247	0.140
σ equity premium shock	0.0503	0.1047	0.2039	0.1147	0.520	0.611	0.718	0.614
σ wage mark up shock	0.0312	0.2083	0.6230	0.2463	0.246	0.285	0.331	0.286
σ price mark up shock	0.0424	0.0757	0.1627	0.0853	0.139	0.162	0.192	0.163
ρ productivity shock	0.1067	0.9835	1.0000	0.8456	0.712	0.828	0.912	0.822
ρ inflation obj. shock	0.0033	1.0000	1.0000	0.8595	0.658	0.865	0.970	0.847
ρ cons. pref. shock	0.1861	0.6185	0.9562	0.6335	0.817	0.886	0.931	0.882
ρ gov. spending shock	0.9973	0.9999	1.0000	0.9991	0.912	0.956	0.982	0.952
ρ labor supply shock	0.1067	0.9835	1.0000	0.8456	0.916	0.955	0.98	0.952
ρ investment shock	0.0014	0.9898	1.0000	0.7153	0.856	0.917	0.961	0.914
investment adj cost	-0.1644	0.3713	0.9354	0.3921	4.321	5.974	7.973	6.048
σ consumption utility	0.3834	1.4140	2.6195	1.3908	1.126	1.608	2.106	1.613
h consumption habit	0.4266	0.8630	0.9758	0.8084	0.416	0.552	0.681	0.551
σ labor utility	1.0905	3.5727	6.6870	3.8446	0.439	1.188	2.365	1.265
fixed cost	0.5880	1.3625	2.2390	1.3836	1.199	1.487	1.835	1.499
calvo employment	0.0260	0.1994	0.6370	0.2397	0.503	0.596	0.671	0.593
capital util. adj. cost	0.0242	0.2848	0.7572	0.3136	0.062	0.175	0.289	0.175
calvo wages	0.0812	0.8473	0.9977	0.7033	0.690	0.758	0.817	0.756
calvo prices	0.9153	0.9918	0.9996	0.9715	0.890	0.909	0.927	0.909
indexation wages	0.1067	0.9835	1.0000	0.8456	0.383	0.663	0.900	0.655
indexation prices	0.1067	0.9835	1.0000	0.8456	0.268	0.425	0.597	0.429
r inflation	0.0032	1.8120	3.8169	1.9552	1.537	1.661	1.821	1.668
r d(inflation)	-0.9020	0.0508	1.0218	0.0485	0.134	0.221	0.313	0.222
r lagged interest rate	0.7600	0.9450	0.9873	0.9214	0.901	0.931	0.946	0.928
r d(output)	-0.2840	0.2540	1.1162	0.3139	0.131	0.173	0.219	0.174
r output	0.1558	0.9560	1.7404	0.9286	0.079	0.143	0.215	0.144
r ϵ productivity shock	-1.1588	-0.4013	0.6628	-0.3168	0.043	0.086	0.137	0.088
r ϵ labor supply	-1.0262	-0.1567	0.4230	-0.2541	0.007	0.030	0.063	0.031

Table 3: Comparison between SW-MCMC and SMC

for the Minnesota prior in $\omega = 0.3$. Fairly similar results (not shown here) are obtained for $\omega = 0.99$ and $\omega = 0.1$. Both comparisons start from $t = 35$ and end at $t = 118$ since we start the MCMC rolling window at $t = 35$. In all of our Bayes factor plots, the Bayes factor is computed relative to the DSGE model. Figure 10 shows the particle approximation of the DSGE marginal likelihoods across the dataset is quite stable for different numbers of particles used in SMC estimation.

Our model comparison results based on sequential Bayes factors tell quite a different story from SW's Bayesian batch model comparison. First of all, it would appear from 8 that all three VAR models fit the data better than DSGE for almost all the time points considered. In addition, the higher the vector autoregressive order, the better model fit the respective model delievers. This is in contrary to SW's findings, which reported that VAR(3) performs the worst out of the VAR models. Our results indicate that this was only true for the first 10 periods, but that VAR(3) quickly catches up and dominates both VAR(1) and VAR(2) as well as the DSGE model pass period 30. With BVAR models, the main finding is that relative model performance is sensitive to the choice of the prior hyperparameter λ but not ω. In each of figure **??**, 9 and **??**, ω is fixed at 0.99, 0.3 and 0.1, respectively, and λ changes from 5 to 0.01. We can see that in each figure, all three BVAR models are chosen in favor of the DSGE model for large values of λ. As λ decreases, DSGE begins to catch up with the reduced-formed models. For small values of λ such as 0.01, DSGE is chosen in favor of the BVAR models. This makes intuitive sense since λ controls the prior variance of the VAR coefficients and small values of this parameter will shrink coefficients toward their prior mean value, which is set to the identity matrix for the first lag coefficients, and zeros for all other coefficients in the Minnesota prior we used. A small value for λ will thus greatly limit the BVAR model's explanatory power, in which case the DSGE model can have a chance of performing on par or even outperform BVAR models. Another interesting effect of decreasing λ is that it changes the relative performance between the BVAR models also. Whereas for large values of λ, BVAR models have similar relative model performance as VAR models, in that higher autoregressive order delivers higher marginal model likelihoods, for low values of λ, BVAR(1) appears to better explain the data than BVAR(2). This is also intuitive to explain. Without constraining the coefficients with a tight prior variance, BVAR(2) will explain the data better with more parameters than BVAR(1). As soon as coefficients are shrunk toward the prior mean, model uncertainty outweighs explanatory power in the BVAR(2) and thus loses to BVAR(1).

The results from both comparison showed that a-theoretical models such as VAR/BVAR perform better than DSGE across the sample. For longer samples, this discrepancy in performance is further widened. The fact that complex structural models fair no better than a-theoretical models perhaps shouldn't come as a surprise as it's well known that structural models in their highly stylized form typically can't compete with pure statistical models in data fitting. However, while reduced form models excel in data fitting, they cannot produce useful economic interpretations of the data. Structural models like the DSGE, on the other hand, have economically meaningful parameters and can provide answers to important economic questions as well as aid in policy making.

5 Concluding Remarks

In this paper we demonstrated that sequential Monte Carlo methods could be used as a viable alternative to MCMC in the estimation of complex economic structural models such as DSGE models. A successful SMC method for parameter estimation requires a mixture of importance sampling/resampling and Markov chain moves. The importance sampling/resampling mechanism induces parameter learning in the particles while Markov chain moves, be it from Metropolis type kernels or exact posteriors computed from conditional sufficient statistics, replenish particles lost in the resampling step. Those two key ideas combined is what makes SMC estimation of fixed model parameters possible. We've listed the many benefits of using SMC methods instead of MCMC methods, and provided an example that shows the simplicity an usefulness of performing sequential model comparison with SMC estimations.

While it may seem like SMC estimation is more computationally intensive than MCMC estimation, the converse is probably the case in practice. In our computation time comparisons, SMC has to generate posterior inference at each point in the data whereas MCMC only generates inference at the end of the dataset. A fair comparison would be between rolling-window MCMC and SMC, since both would then generate the same amount of posterior information. To take full advantage of an SMC estimation procedure, we recommend the following usage. First run the SMC method over all available historical data with several different settings for the number of particles. Perhaps a low, medium, and high number of particles, which would correspond to three different "calibration levels" so to speak. This would take a long time to run (around the same amount of time as rolling-window MCMC with comparable sample sizes) but only needs to be done once. Then as new data arrives, new inference could be made with quick updates based on estimation results from the last period. Each new update with say, one million particles, will be much faster than an MCMC analysis with the same number of iterations as MCMC has to traverse through the entire dataset whereas SMC doesn't.

In summary, SMC has many practical benefits over MCMC but needs additional thoretical as well as methodological development to flourish as an estimation method. That said, we believe that SMC estimation applied to economic models is an exciting area of research and hope that this paper will provide a fresh perspective in the ever growing Bayesian econometric literature.

Appendix. Linearized DSGE model in Smets and Wouters (2003)

In this section we provide the complete description of the linearized DSGE model for the Euro area from Smets and Wouters (2003), which include the system of rational expectations equations, the vector of endogenous variables, the list of exogenous shock variables, and the vector of parameters. The linearized model is characterized by the following system of linear rational expectations equations:

$$
\begin{aligned}
C_t &= \frac{h}{1+h}C_{t-1} + \frac{1}{1+h}C_{t+1} - \frac{1-h}{(1+h)\sigma_c}(R_t - \pi_{t+1}) + \frac{1-h}{(1+h)\sigma_c}(\epsilon_t^b - \epsilon_{t+1}^b) \\
I_t &= \frac{1}{1+\beta}I_{t-1} + \frac{\beta}{1+\beta}I_{t+1} + \frac{\varphi}{1+\beta}Q_t + \beta\epsilon_{t+1}^I - \epsilon_t^I \\
Q_t &= -(R_t - \pi_{t+1}) + \frac{1-\tau}{1-\tau+\bar{r}^k}Q_{t+1} + \frac{\bar{r}^k}{1-\tau+\bar{r}^k}r_{t+1}^k + \eta_t^Q \\
K_t &= (1-\tau)K_{t-1} + \tau I_{t-1} \\
\pi_t &= \frac{\beta}{1+\beta\gamma_p}\pi_{t+1} + \frac{\gamma_p}{1+\beta\gamma_p}\pi_{t-1} + \frac{1}{1+\beta\gamma_p}\frac{(1-\beta\xi_p)(1-\xi_p)}{\xi_p}\left[\alpha r_t^k + (1-\alpha)w_t - \epsilon_t^a + \eta_t^p\right] \\
w_t &= \frac{\beta}{1+\beta}w_{t+1} + \frac{1}{1+\beta}w_{t-1} + \frac{\beta}{1+\beta}\pi_{t+1} - \frac{1+\beta\gamma_w}{1+\beta}\pi_t + \frac{\gamma_w}{1+\beta}\pi_{t-1} \\
&\quad - \frac{1}{1+\beta}\frac{(1-\beta\xi_w)(1-\xi_w)}{\left(1+\frac{(1+\lambda_w)\sigma_L}{\lambda_w}\right)\xi_w}\left[w_t - \sigma_L L_t - \frac{\sigma_c}{1-h}(C_t - hC_{t-1}) - \epsilon_t^L - \eta_t^w\right] \\
L_t &= -w_t + (1+\psi)r_t^k + K_{t-1} \\
Y_t &= (1-\tau k_y - g_y)C_t + \tau k_y I_t + g_y\epsilon_t^G = \phi\epsilon_t^a + \phi\alpha K_{t-1} + \phi\alpha\psi r_t^k + \phi(1-\alpha)L_t \\
R_t &= \rho R_{t-1} + (1-\rho)\left[\bar{\pi}_t + r_\pi(\pi_{t-1} - \bar{\pi}_t) + r_Y Y_t\right] + r_{\Delta\pi}(\pi_t - \pi_{t-1}) \\
&\quad + r_{\Delta y}(Y_t - Y_{t-1}) - r_a\eta_t^a - r_L\eta_t^L + \eta_t^R,
\end{aligned}
$$

where variables dated at $t+1$ refer to their rational expecations. The nine endogenous variables are: inflation (π_t), nomial wage (w_t), capital (K_{t-1}), value of capital stock (Q_t), investment (I_t), consumption (C_t), interest rate (R_t), rental rate of capital (r_t^k), and labor (L_t). The exogenous shocks are: productivity shock (ϵ_t^a), inflation objective shock ($\bar{\pi}_t$), consumption preference shock (ϵ_t^b), government spending shock (ϵ_t^G), labor supply shock (ϵ_t^L), investment shock (ϵ_t^I), interest rate shock (η_t^R), equity premium shock (η_t^Q), price mark-up shock (η_t^p), and wage mark-up shock (η_t^w). Of those ten exogenous shock variables, the first six follow independent AR(1) processes, and the remaining four follow IID independent processes. There is a total of thirty-four model parameters, they include the shock process parameters and those involved in the previous equations.

References

[1] Aruoba, S.B., Fernández-Villaverde, J., and Rubio-Ramírez, J.F. (2006). Comparing solution methods for dynamic equilibrium economies. *Journal of Economic Dynamics & Control* 30, 2477-2508.

[2] Bengtsson, T., Bickel, P., and Li, B. (2007). Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *Technical Report.*

[3] Carvalho, C.M., Johannes, M., Lopes, H.F., and Polson, N.G. (2009). Particle Learning. *Bayesian Statistics 9,* (to appear).

[4] Chib, S., Srikanth, M. (2010). Tailored randomized block MCMC methods with application to DSGE models. *Journal of Econometrics* 155 (2010) 19-38.

[5] Chopin, N. (2004). Central limit theorem for sequential Monte Carlo and its application to Bayesian inference. *Ann. Statist.,* 32. 2385-2411.

[6] Ackerberg, D., Geweke, J., and Hahn, J. (2009). Comments on "Convergence Properties of the Likelihood of Computed Dynamic Models". *Econometrica,* Vol. 77, No. 6, 2009-2017.

[7] Del Moral, P., and Guionnet, A. (1999). A central limit theorem for non-linear filtering using interacting particle systems. *Ann. Appl. Probab.,* 9, 275-297.

[8] Doucet, A., and Johansen, A.M. (2008). A Tutorial on Particle Filtering and Smoothing: Fifteen years later. *Unpublished Manuscript.*

[9] Douc, R., Capp, O., and Moulines, E. (2005). Comparison of resampling schemes for particle filtering. In *4th International Symposium on Image and Signal Processing and Analysis.*

[10] Fearnhead, P. (2002). Markov chain Monte Carlo, sufficient statistics, and particle filters. *Journal of Computational and Graphical Statistics,* 11, 848-862.

[11] Fearnhead, P., and Clifford, P. (2003). On-line inference for hidden Markov models via particle filters. *Journal of the Royal Statistical Society, B,* 65, 887-899.

[12] Fernández-Villaverde, J. (2009). The Econometrics of DSGE models. *PIER Working Paper 09-008.*

[13] Fernández-Villaverde, J., and Rubio-Ramírez, J.F. (2005). Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood. *Journal of Applied Econometrics* **20**: 891-910.

[14] Fernández-Villaverde, J., and Rubio-Ramírez, J.F., and Santos, M.S. (2006). Convergence Properties of the Likelihood of Computed Dynamic Models. *Econometrica,* Vol. 74, No. 1, 93-119.

[15] Gasper, J., and Judd K. (1997) Solving large-scale rational expectations models. *Macroeconomic Dynamics* 1, 45-75.

[16] Giacomini, R., and Rossi, B. (2007). Model Selection and Forecast Comparison in Unstable Environments. *Working Paper.*

[17] Gilks, W. and Berzuini, C. (2001) Following a moving targat: Monte Carlo inference for dynamic Bayesian models. *Journal of Royal Statistical Society, B,* 63, 127-146.

[18] Hall, R. (1971) The dynamic effects of fiscal policy in an economy with foresight. *Review of Economic Studies* 38, 229-244.

[19] Johannes, M., and Polson, N.G. (2008). Exact particle filtering and learing. *Working Paper.*

[20] Judd, K.L., and Guu, S.M. (1993) Perturbation solution methods for economic growth model. In *Economic and Financial Modelling in Mathematica* (ed H. Varian). New York: Springer.

[21] Judd, K.L., and Guu, S.M. (2001) Asymptotic methods for asset market equilibrium analysis. *Economic Theory* 18, 127-157.

[22] Litterman, R.B. (1986) Forecasting with Bayesian vector autoregressions: Five years of experience. *Journal of Business and Economic Statistics* **4**, 25-38.

[23] Liu, J. and West, M. Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice* (eds A. Doucet, J. De Freitas and N. Gordon). New York: Springer.

[24] Magill, J.P.M. (1977) A local analysis of N-sector capital under uncertainty. *Journal of Economic Theory* 15, 219-221.

[25] Papaspiliopoulos, O., and Roberts, G. (2008). Stability of the Gibbs sampler for Bayesian hierarchical models. *Annals of Statistics,* Volume 36, Number 1, 95-117.

[26] Pitt, M., and Shephard, N. (1999b). Filtering via simulation: Auxilliary particle filters. *Journal of the American Statistical Association* **94**: 590-599.

[27] Polson, N.G., Stroud, J.R., and Mller, P. (2008). Practical filtering with sequential parameter learning. *J. R. Statist. Soc.* B, 70, Part 2, pp. 413-428.

[28] Schmitt-Groh, S., and Uribe, M. (2004). Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal Of Economic Dynamics & Control* 28, 755-775.

[29] Smets, F. and R. Wouters. (2003). An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association,* 1, 1123-1175.

[30] Storvik, G. (2002). Particle Filters for State-Space Models With the Presence of Unknown Static Parameters. *IEEE Transactions on Signal Processing* 50, 281-289.

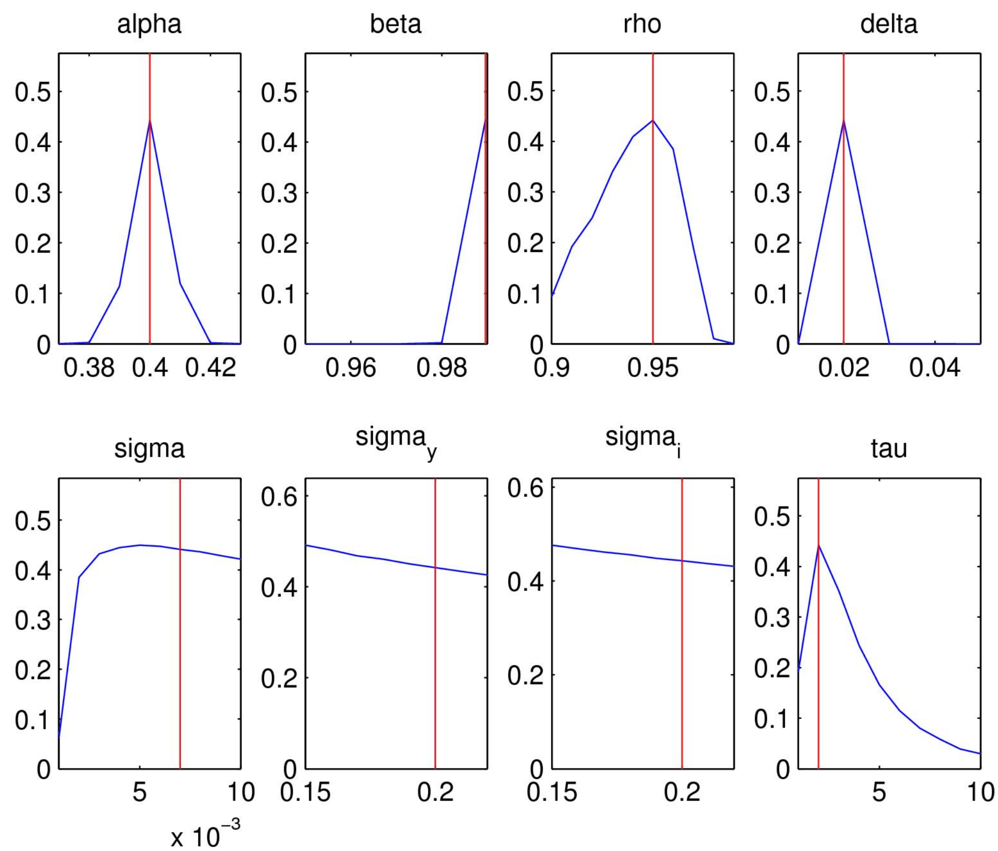

Figure 1: Profile likelihoods for the neoclassical growth model.

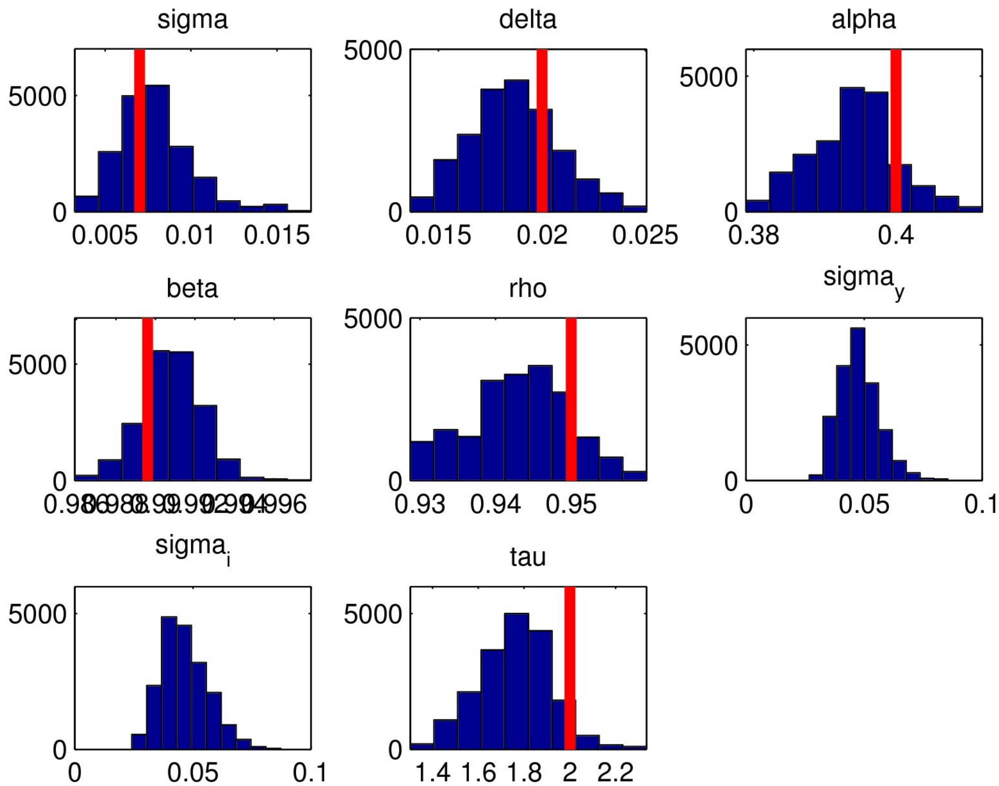

Figure 2: MCMC Histograms for the neoclassical growth model

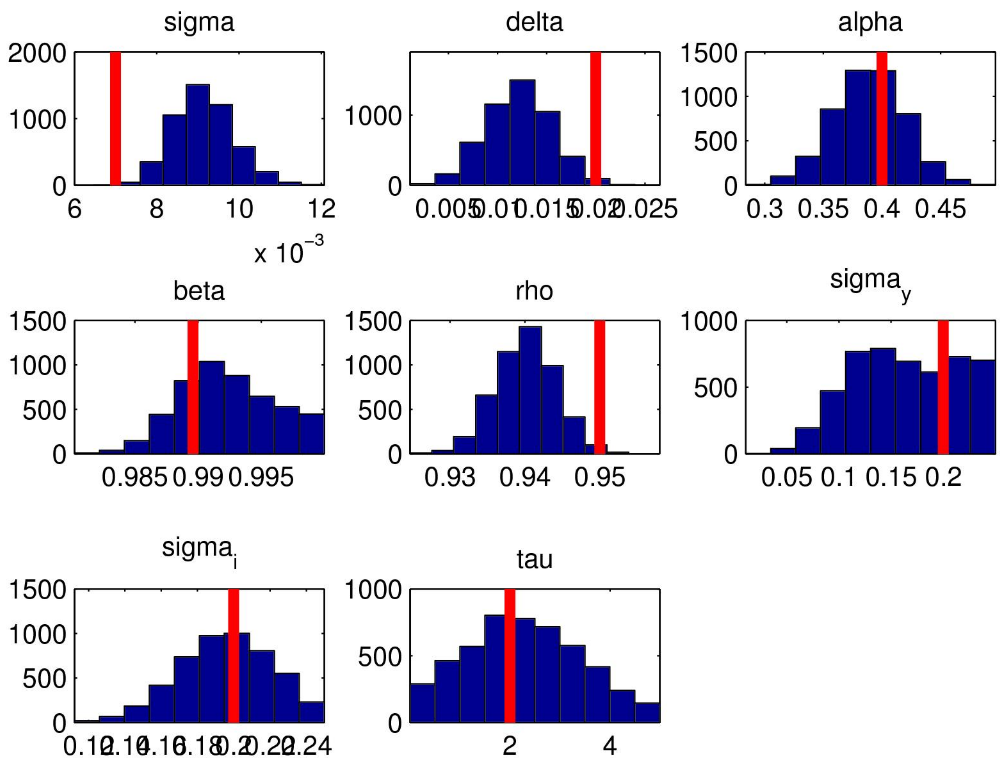

Figure 3: SMC Histograms for the neoclassical growth model

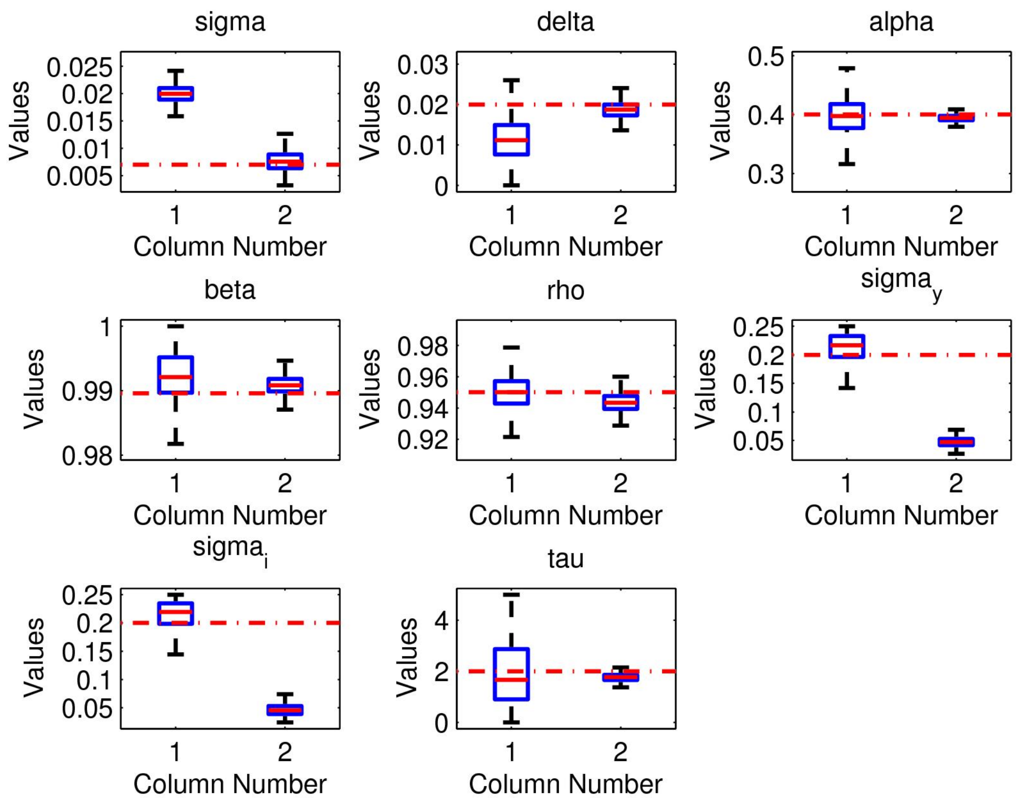

Figure 4: Boxplot Comparison (left is SMC)

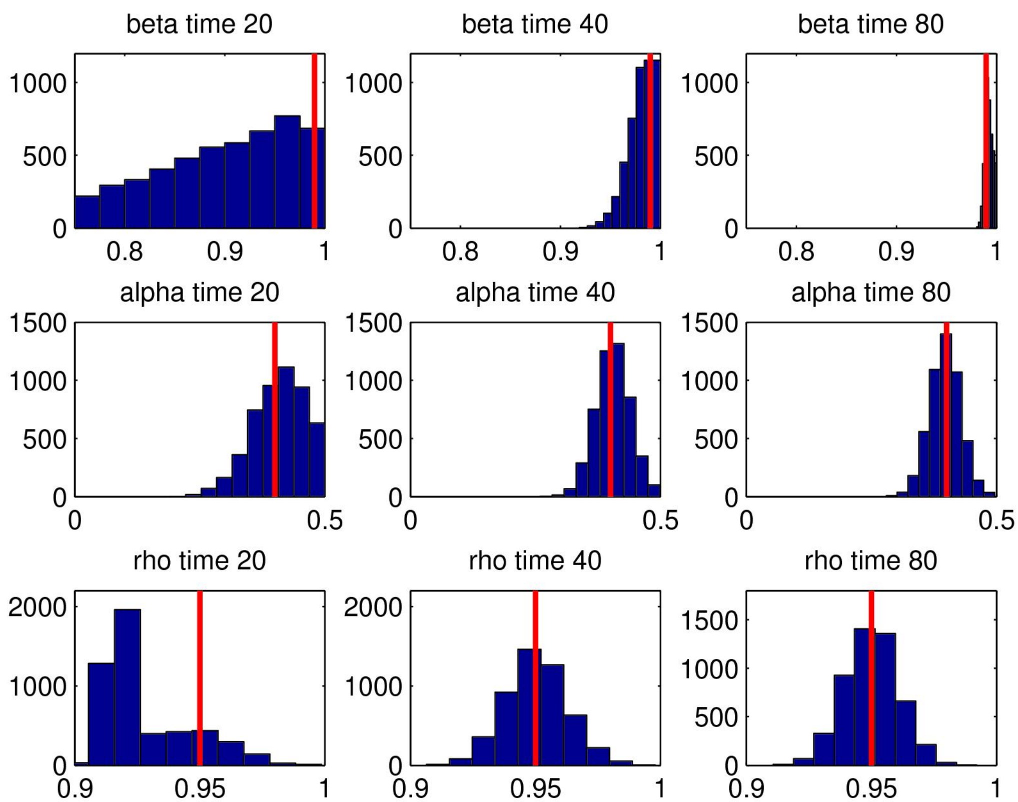

Figure 5: Evolution of parameter posteriors through time.

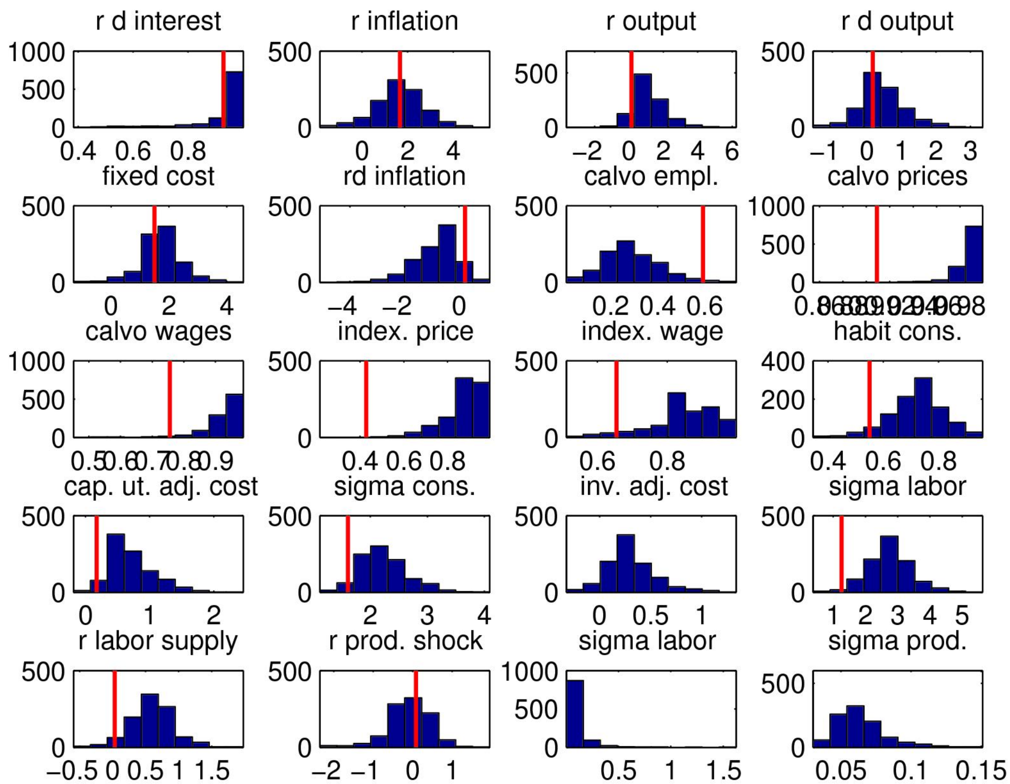

Figure 6: SMC histograms at end the of the sample for SW.

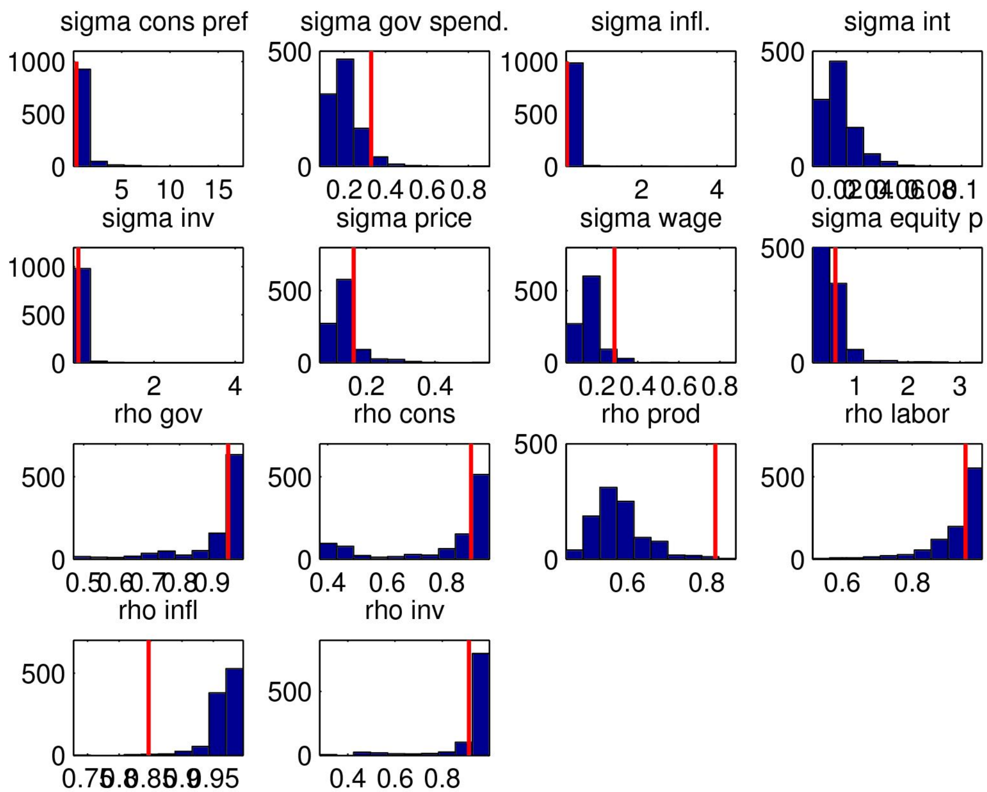

Figure 7: SMC histograms at the end of the sample for SW.

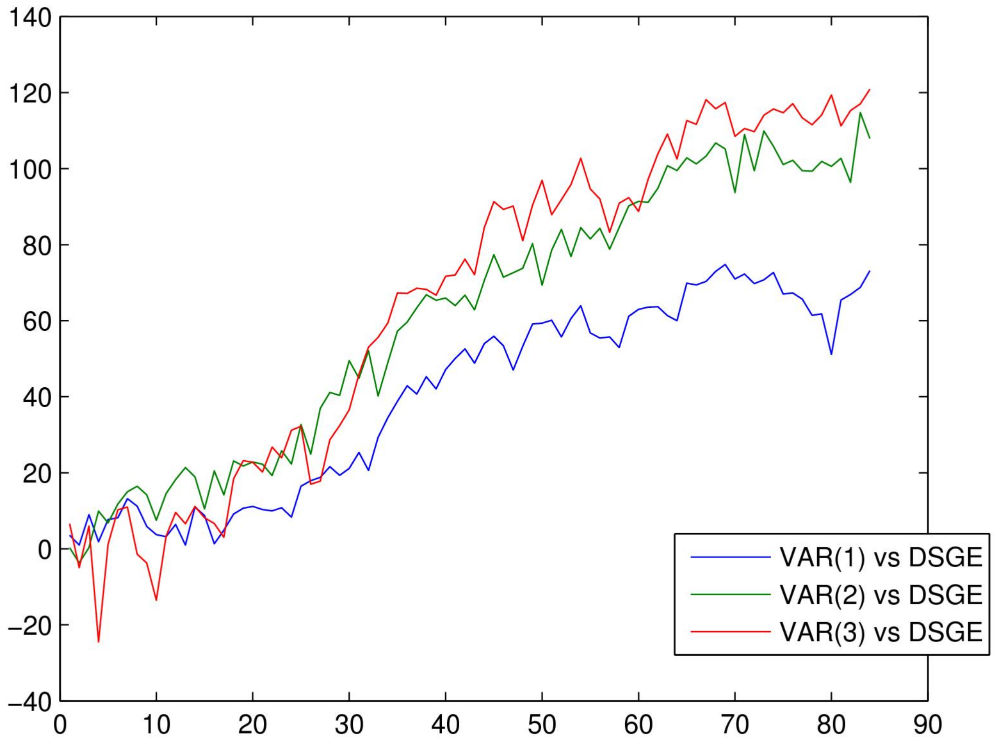

Figure 8: Sequential Log Bayes Factor: VAR vs DSGE

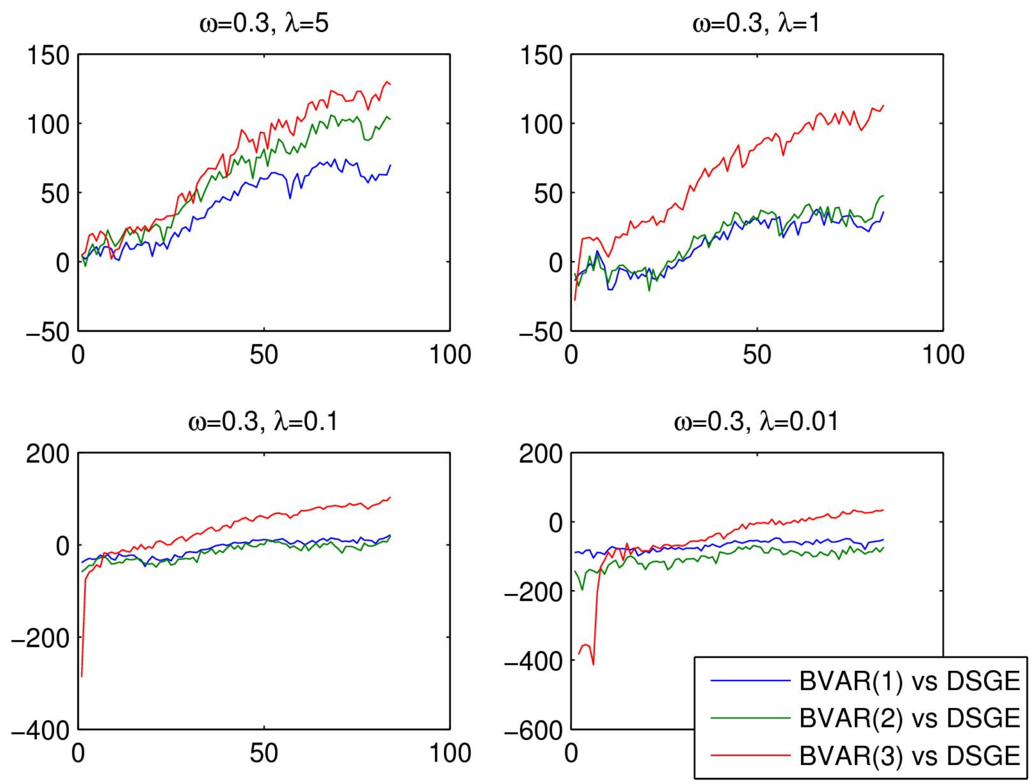

Figure 9: Sequential Log Bayes Factor (ω=0.3): BVAR vs DSGE

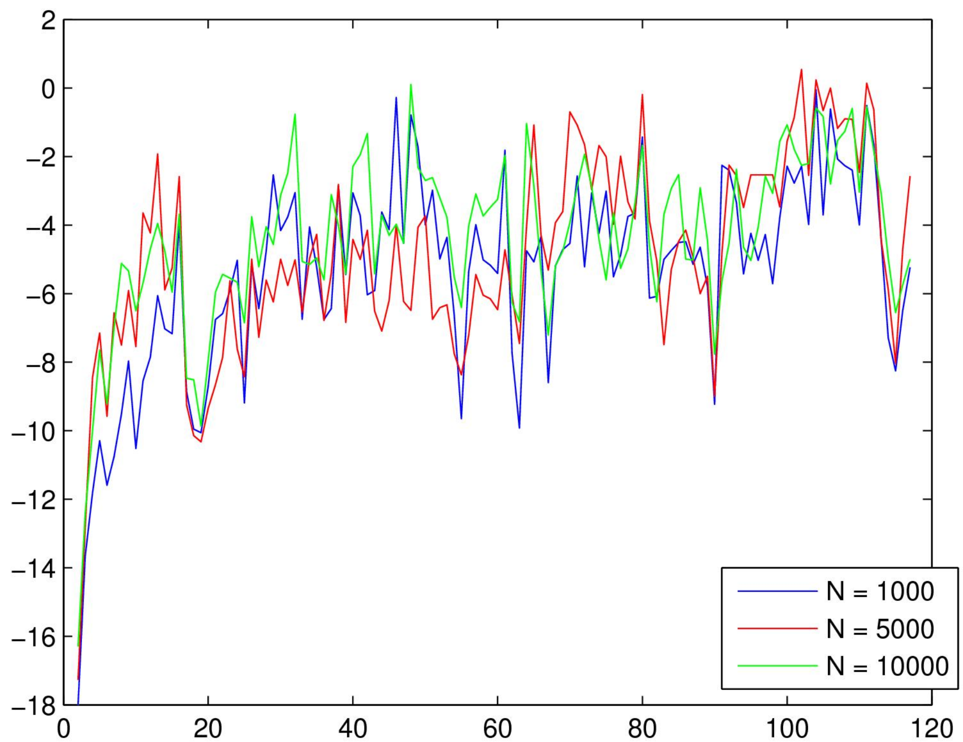

Figure 10: Stability of SMC estimation of $p(y_t|y_{t-1})$ from the SW-DSGE model

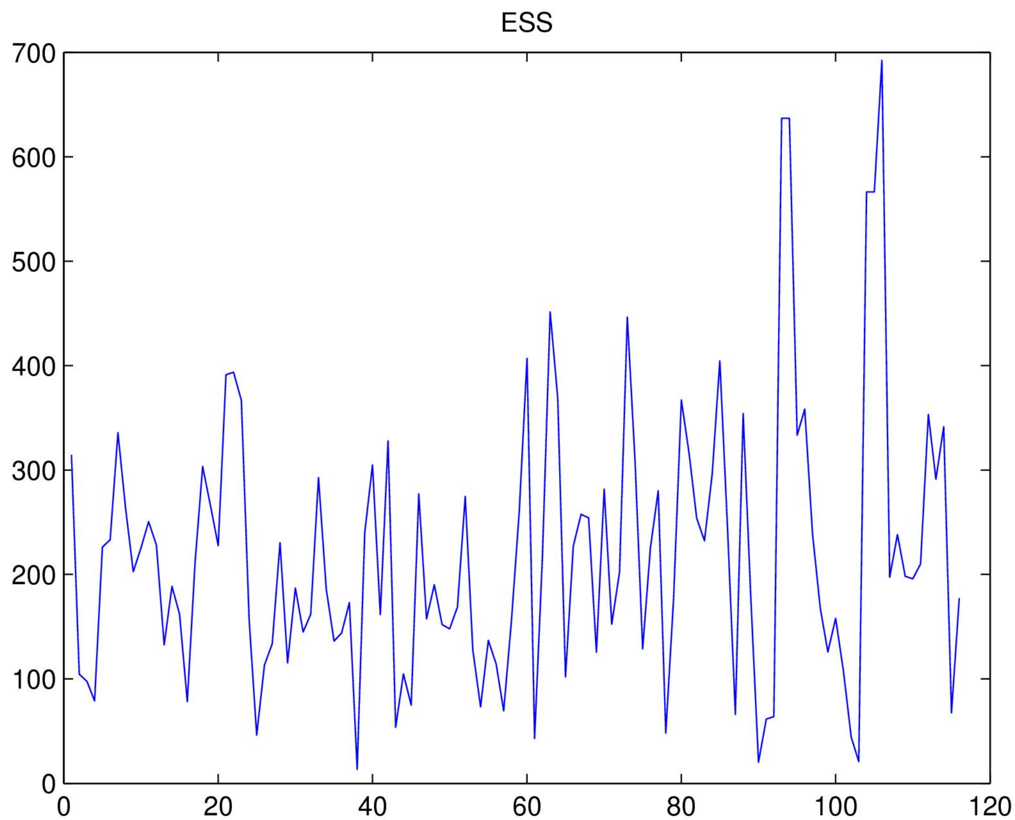

Figure 11: ESS for SW DSGE estimated with 1000 particles