Structural Estimation Using Sequential Monte Carlo Methods

by

Hao Chen

Department of Business Administration
Duke University

Date: _____

Approved:

_____
A. Ronald Gallant, Supervisor

_____
Hedibert Freitas Lopes

_____
Juan Rubio-Ramirez

_____
Alexandre Belloni

_____
Giuseppe Lopomo

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Business Administration
in the Graduate School of Duke University
2011

Abstract

(0501, 0463)

Structural Estimation Using Sequential Monte Carlo Methods

by

Hao Chen

Department of Business Administration
Duke University

Date: _____

Approved:

_____
A. Ronald Gallant, Supervisor

_____
Hedibert Freitas Lopes

_____
Juan Rubio-Ramirez

_____
Alexandre Belloni

_____
Giuseppe Lopomo

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Business Administration in the Graduate School of Duke University
2011

Abstract

This dissertation aims to introduce a new sequential Monte Carlo (SMC) based estimation framework for structural models used in macroeconomics and industrial organization. Current Markov chain Monte Carlo (MCMC) estimation methods for structural models suffer from slow Markov chain convergence, which means parameter and state spaces of interest might not be properly explored unless huge numbers of samples are simulated. This could lead to insurmountable computational burdens for the estimation of those structural models that are expensive to solve. In contrast, SMC methods rely on the principle of sequential importance sampling to jointly evolve simulated particles, thus bypassing the dependence on Markov chain convergence altogether. This dissertation will explore the feasibility and the potential benefits to estimating structural models using SMC based methods.

Chapter 1 casts the structural estimation problem in the form of inference of hidden Markov models and demonstrates with a simple growth model.

Chapter 2 presents the key ingredients, both conceptual and theoretical, to successful SMC parameter estimation strategies in the context of structural economic models.

Chapter 3, based on Chen, Petralia and Lopes (2010), develops SMC estimation methods for dynamic stochastic general equilibrium (DSGE) models. 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 online estimation 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.

Chapter 4 proposes an SMC estimation procedure and show that it readily applies to the estimation of dynamic discrete games with serially correlated endogenous state variables. I apply this estimation procedure to a dynamic oligopolistic game of entry using data from the generic pharmaceutical industry and demonstrate that the proposed SMC method can potentially better explore the parameter posterior space while being more computationally efficient than MCMC estimation. In addition, I show how the unobserved endogenous cost paths could be recovered using particle smoothing, both with and without parameter uncertainty. Parameter estimates obtained using this SMC based method largely concur with earlier findings that spillover effect from market entry is significant and plays an important role in the generic drug industry, but that it might not be as high as previously thought when full model uncertainty is taken into account during estimation.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my advisor, A. Ronald Gallant, for providing me with the computational resources that made parts of this dissertation possible, and for his guidance and support during the development of this dissertation in general. I would like to thank Francesca Petralia for the many hours we spent discussing and working on a project that eventually evolved into a chapter of this dissertation. I would also like to thank Hedibert Lopes for his guidance and encouragement during the past two years.

A special thanks goes to my good friend and colleague Howard Kung, for being a constant source of encouragement and inspiration throughout the years, and a great basketball coach.

Finally, I'd like to thank my family for their unwavering support throughout my Ph.D. career. To them I dedicate this work.

1

Introduction

In recent years, structural models have become the standard tool to address research questions in several major fields of economics such as macroeconomics and industrial organization. The ability to capture underlying processes driving an economic phenomenon using highly interpretable structural parameters can offer economists new insights that's not attainable with a purely statistical reduced-form model.

Structural models, however, are difficult to estimate for a number of reasons. First of all, structural models often incorporate nonlinearity and/or non-normality in their model assumption. Secondly, structural models often feature unobserved latent processes, which present their own challenges in terms of model inference. Third, it's often computationally intensive to put a structural model into reduced-form, which involves solving the model using numerical methods. These reasons make classical inference for structural models all but extremely difficult. In contrast, Bayesian estimation methods are straightforward to apply and provide likelihood-based inference. Recent empirical literature in macroeconomics and industrial organization have adopted Bayesian simulation-based strategies with success, taking full advantage of today's readily available computing resources. The most common Bayesian

structural estimation method currently used is an MCMC-based algorithm, typically built around a Metropolis-Hastings random walk kernel to address non-conjugate posterior forms. To deal with the latent processes involved, any unobserved state variables are integrated out of the model likelihood with a particle filter. Applying MCMC to structural estimation has a number of drawbacks, however, the biggest of which is the well documented fact that MCMC sample chains typically suffer from slow Markov convergence. Although advanced MCMC methods, such as Chib and Ramamurthy (2010), have been proposed for DSGE models to address this issue, those methods most likely will require significant tuning to work on other structural models. There is thus a need for a general structural estimation framework that can address this slow mixing problem.

Sequential Monte Carlo (SMC) methods, otherwise known as particle filters in a pure state filtering context, are a class of numerical methods designed to perform inference in general state space hidden Markov models. Unlike Kalman filters, SMC methods don't rely on local linearization or normal shock assumptions and can thus better capture any inherent nonlinearities present in models of interest. This flexibility is the main reason behind the ever growing popularity of SMC methods in a diverse array of fields ranging from financial econometrics to robotics. One of the primary problems SMC methods are designed for is a filtering problem where the goal is to obtain posterior inference on the latent state trajectories given some noisy observation. When embedded in an MCMC algorithm for structural parameter estimation, as done in Fernndez-Villaverde and Rubio-Ramrez (2005) and Gallant, Hong and Khwaja (2010), the importance weights of the filtered particles are used to obtain an approximation of likelihood function. In this sense, particle filters are essentially used as a Monte Carlo integrator to marginalize out the unobserved state variables. Recent developments have shown that SMC methods are capable of estimating fixed parameters in addition to state filtering. Practical SMC-based methods were pro-

posed by Liu and West (2001), Storvik (2002), Fearnhead (2002), Gilks and Berzuini (2001), and Carvalho, Johannes, Lopes and Polson (2010) among others. SMC based estimation methods have several advantages over traditional MCMC based Bayesian methods when applied to structural models such as DSGE models. First, SMC methods don't rely on Markov chain convergence and instead work on the principle of importance sampling. For complex structural models, Markov convergence might not be geometrically ergodic (see Papaspiliopoulos and Roberts (2008). Secondly, it might be difficult for an MCMC algorithm to fully explore model parameter spaces either for the slow mixing problem inherent to MCMC strategies or for model specific reasons such as multimodel posteriors. In this scenario, a well designed SMC-based procedure can visit parts of the posterior space that an MCMC algorithm is unlikely to have visited. Lastly, SMC methods developed in this dissertation enjoy a level of computational efficiency over MCMC methods by combining state filtering and parameter learning. Depending on the application, this efficiency could translate into significant savings in computational resources required.

This chapter will present the key concepts in the SMC literature as well as introducing parameter estimation using SMC methods.

1.1 Hidden Markov Models

General state-space hidden Markov models (HMM) are a very powerful class of models that could be adapted to a wide array of scenarios. General state-space models are characterized by the following observation and state evolution equations

$$y_{t+1} \sim p(y_{t+1}|x_{t+1}, \theta) \tag{1.1}$$

$$x_{t+1} \sim p(x_{t+1}|x_t, \theta), \tag{1.2}$$

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. In a pure state-filtering context where the fixed parameter is known, θ is suppressed in the above equations. Given observations $y_{1:T}$, a Bayesian model defined by equations (1.1) and (1.2) has the following likelihood function

$$p(y_{1:T}|x_{0:T}, \theta) = \prod_{t=1}^{T} p(y_t|x_t, \theta) \tag{1.3}$$

and prior distribution

$$p(x_{0:T}, \theta) = p(x_0|\theta)p(\theta) \prod_{t=1}^{T} p(x_t|x_{t-1}, \theta). \tag{1.4}$$

The goal of inference is to obtain the joint state and parameter posterior distribution

$$p(x_{0:T}, \theta|y_{1:T}) = \frac{p(x_{0:T}, \theta, y_{1:T})}{\int p(x_{0:T}, \theta, y_{1:T}) dx_{0:T} d\theta}. \tag{1.5}$$

In certain situations, the marginal likelihood

$$p(y_{1:T}) = \int p(x_{0:T}, \theta, y_{1:T}) dx_{0:T} d\theta \tag{1.6}$$

is also a quantity of interest.

1.2 Example: A Simple Growth Model

The above specification captures a great many models of interest in economics. The following is a simple growth model where labor is supplied inelastically and fixed at unity. A representative agent seeks to maximize the expected value of lifetime utility

$$E_0 \sum_{t=1}^{\infty} \beta^t \ln(c_t),$$

subject to

$$q_t = e^{z_t} k_t^{\alpha} \tag{1.7}$$

$$q_t = c_t + i_t \tag{1.8}$$

$$k_{t+1} = i_t + (1-\delta)k_t \tag{1.9}$$

$$z_{t+1} = \rho z_t + \epsilon_t, \tag{1.10}$$

where q_t, k_t, c_t, i_t and z_t denote output, capital, consumption, investment and technology. The structural parameters are $(\alpha, \beta, \delta, \rho, \sigma)$, which denote capital's share of output, discount factor, capital depreciation, persistence to technology evolution, and volatility in technology shock $\epsilon_t \sim N(0, \sigma^2)$. To convert the above model into state-space form, numerical methods would be needed to find an approximate to the policy function for consumption in the form of $c(z_t, k_t)$ as no close-form solution is available. If we look at the full depreciation case where $\delta = 1$, however, a close-form solution for consumption is given by $c(z_t, k_t) = (1 - \alpha\beta)e^{z_t} k_t^{\alpha}$. Assume that we observe output and investment with measurement errors, the state-space form of the above growth model is characterized by observation equations

$$q_t = e^{z_t} k_t^{\alpha} + \epsilon_{q_t} \tag{1.11}$$

$$i_t = \alpha\beta e^{z_t} k_t^{\alpha} + \epsilon_{i_t}, \tag{1.12}$$

and state equations (1.6) and

$$k_t = \alpha\beta e^{z_t} k_t^{\alpha}. \tag{1.13}$$

2

SMC Methods for Parameter Estimation

Computing the posterior for a general HMM in close-form is only possible in but a very limited number of cases. In most situations, one must resort to finding an approximate posterior with numerical methods. SMC methods are powerful simulation methods designed to produce samples approximately distributed according to the true posterior. There has been a rich development in the SMC literature in recent years[1] and several key ideas form the basis of the estimation strategies proposed in this dissertation.

2.1 State Filtering

State filtering is the central problem of interest in the SMC literature and the subject of the vast majority of applications for SMC methods. As such we will first consider the standard filtering problem without parameter uncertainty in the following state-

[1] See Doucet and Johansen (2008) for an excellent tutorial on particle filtering and smoothing. See Doucet et al. (2001) for a complete reference of recent developments in SMC methods.

space setup

$$y_{t+1} \sim p(y_{t+1}|x_{t+1}) \tag{2.1}$$

$$x_{t+1} \sim p(x_{t+1}|x_t), \tag{2.2}$$

where the unobserved Markov process $\{x_t\}_{t\geq 1}$ has initial distribution $p(x_0)$, and the observations $\{y_t\}_{t\geq 1}$ are conditionally independent given $\{x_t\}_{t\geq 0}$. In state filtering, one wishes to make sequential inference on the marginal filtering distribution $p(x_{t+1}|y_{1:t+1})$. Note that we are not making inference on the joint filtering posterior $p(x_{1:t}|y_{1:t})$. This is because it's inherently impossible to obtain a reliable SMC estimate of $p(x_{1:t}|y_{1:t})$ due to the sample impoverishment problem caused by the resampling step in SMC methods. With only a fixed number of particles, the particle approximation of $p(x_{1:t}|y_{1:t})$ will eventually collapse to a single point for a big enough t (a rigorous discussion on this problem can be found in Doucet and Johansen (2008)). We thus focus on the marginal filtering distribution only, which can be consistently estimated under regularity conditions.

We would like to approximate the following filtering recursions,

$$p(x_{t+1}|y_{1:t+1}) = \frac{p(y_{t+1}|x_{t+1})p(x_{t+1}|y_{1:t})}{p(y_{t+1}|y_{1:t})} \tag{2.3}$$

$$p(x_{t+1}|y_{1:t}) = \int p(x_t|y_{1:t})p(x_{t+1}|x_t)dx_t. \tag{2.4}$$

At the core of any SMC method is importance sampling, an importance function of the following decomposable form is used to facilitate fixed-cost sequential estimation,

$$q_t(x_{1:t+1}) = q_t(x_{1:t})q_{t+1}(x_{t+1}|x_{1:t}) \tag{2.5}$$

$$= q_1(x_1)\prod_{k=2}^{t+1} q_k(x_k|x_{1:k-1}). \tag{2.6}$$

This means we can obtain approximate samples from $p(x_{t+1}|y_{1:t+1})$ by sampling particles $x_{t+1}^{(i)}$ from $q_{t+1}(x_{t+1}|x_{1:t}^{(i)})$ with unnormalized weights $w_{t+1}(x_{1:t+1})$. The weights

$w_{t+1}(x_{1:t+1})$ can also be computed in the following recursive way,

$$w_{t+1}(x_{1:t+1}) = w_t(x_{1:t})\alpha_{t+1}(x_{1:t+1}) \tag{2.7}$$

$$= w_1(x_1)\prod_{k=2}^{t+1}\alpha_{t+1}(x_{1:t+1}), \tag{2.8}$$

where the incremental weights are

$$\alpha_{t+1}(x_{1:t+1}) = \frac{p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)}{q_{t+1}(x_{t+1}|x_{1:t})}. \tag{2.9}$$

Many different SMC algorithms simply differ in the way the importance function $q_{t+1}(x_{t+1}|x_{1:t})$ is chosen. For example, in the bootstrap state filter used in Fernández-Villaverde and Rubio-Ramírez (2005) and Gallant et al. (2010), $q_{t+1}(x_{t+1}|x_{1:t}) = p(x_{t+1}|x_t)$. In other words, the importance function is chosen to be the prior density of the latent state. It can be shown that the variance of the importance weights $w_{t+1}^{(i)}$ is minimized by choosing an importance function of the form

$$q_{t+1}^{opt}(x_{t+1}|x_{1:t}) = \frac{p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)}{p(y_{t+1}|x_t)} \tag{2.10}$$

$$= p(x_{t+1}|y_{t+1}, x_t), \tag{2.11}$$

in which case the incremental weights $\alpha_{t+1}(x_{1:t+1})$ would reduce to $p(y_{t+1}|x_t)$. This inclusion of information y_{t+1} in the importance function at time $t+1$ is referred to as perfect adaption in Pitt and Shephard (1999) and will yield optimal SMC performance. For structural models under consideration in this dissertation, however, $q_{t+1}^{opt}(x_{t+1}|x_{1:t})$ is usually not available in close-form. The auxiliary particle filter developed in Pitt and Shephard (1999) allows us to make use of information y_{t+1} in a different manner and offers better performance than the simple bootstrap filter. Auxiliary particle filtering is an example of the resample-sample class of SMC methods, which reverses the sample then resample logic in the bootstrap filter. This simple change will result in a greater number of distinct particles representing the

target distribution as resampling will cause particle degeneracy. Auxiliary particle filtering also makes use of information at $t+1$ by first looking ahead to resample state particles that will best rationalize the data at $t+1$ before sampling posterior state particles at time $t+1$. The auxiliary particle filter is an example of what is known as a resample-sample method, which is designed to alleviate particle degeneracy by changing the order of the sampling and resampling step in a standard SMC method. For a resample-sample method, the incremental importance weights are

$$\alpha_{t+1}(x_{1:t+1}) = \frac{p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)}{q^1_{t+1}(x_t|y_{t+1})q^2_{t+1}(x_{t+1}|x_{1:t})}, \tag{2.12}$$

where particles are first resampled according to weights $q^1_{t+1}(x_t|y_{t+1})$ and then sampled from the importance function $q^2_{t+1}(x_{t+1}|x_{1:t})$. When the optimal importance function (2.11) is not available, local approximations are used, typically setting $q^1_{t+1}(x_t|y_{t+1}) = p(y_{t+1}|E(x_{t+1}|x_t))$ and $q^2_{t+1}(x_{t+1}|x_{1:t}) = p(x_{t+1}|x_t)$.

2.2 Joint State Filtering and Parameter Learning

In the presence of parameter uncertainty, we are now working with the state-space system (1.1) and (1.2) and would like to make sequential inference on $p(x_{t+1}, \theta|y_{1:t+1})$. This means we want to build an empirical measure of $p(x_{t+1}, \theta|y_{1:t+1})$ with a swarm of weighted particles,

$$\hat{p}(x_{t+1}, \theta|y_{1:t+1}) = \frac{1}{N}\sum_{i=1}^{N} W^i_{t+1}\delta_{(x^i_{t+1},\theta^i_{t+1})}(x_{t+1}, \theta_{t+1}), \tag{2.13}$$

where

$$W^i_{t+1} = \frac{w_{t+1}(x^i_{t+1}, \theta^i_{t+1})}{\sum_{j=1}^{N} w_{t+1}(x^j_{t+1}, \theta^j_{t+1})}. \tag{2.14}$$

Most recent developments in SMC parameter estimation make use of parameter sufficient statistics for efficient computation, which would require posterior distributions

of model parameters to be in analytical form. Examples of this line of work include Storvik (2002), Fearnhead (2002) and Carvalho et al. (2010), to name a few. As parameters in most structural models often don't have closed-form posteriors available, a successful SMC estimation algorithm for structural models calls for a base algorithm that doesn't require the use of sufficient statistics.

One such method is the modified normal kernel move in Liu and West (2001) (LW). The main strategy in this method is to evolve parameters using an artificial law of motion as if they are additional state variables in the system and adjusting for the loss of information incurred through kernel shrinkage locations. Specifically, the LW method uses the following normal mixture smoothing kernel with location shrinkage to facilitate parameter learning

$$p(\theta|y_{1:t+1}) \approx \sum_{i=1}^{N} w_t^{(i)} N(\theta|a\theta_t^{(i)} + (1-a)\bar{\theta}_t, (1-a^2)V_t), \tag{2.15}$$

where $\theta_t^{(i)}$ denote the posterior samples of θ a time t, V_t denote the posterior variance of $\theta_t^{(i)}$, and $a = (3\delta - 1)/2\delta$ controls the degree of location shrinkage through the discount factor δ. Empirical evidence show that setting δ to values in the range of [0.9, 0.95] works best parameter estimation, as higher values will contribute to particle degeneracy. The shrinkage pattern of the kernel locations, coupled with the variance correction term $1 - a^2$ gives LW the ability to traverse the parameter space using artificial evolutions without Monte Carlo variance degradation. In the resample-sample framework, the LW algorithm is essentially an SMC filter with the following incremental importance weights

$$\alpha_{t+1}(x_{1:t+1}) = \frac{p(y_{t+1}|x_{t+1}, \theta_{t+1})p(x_{t+1}, \theta_{t+1}|x_t, \theta_t)}{q_{t+1}^1(x_t, \theta_t|y_{t+1})q_{t+1}^2(x_{t+1}, \theta_{t+1}|x_t, \theta_t, y_{t+1})} \tag{2.16}$$

$$= \frac{p(y_{t+1}|x_{t+1}, \theta_{t+1})}{p(y_{t+1}|g(x_t), m(\theta_t))}, \tag{2.17}$$

where θ_t denote the time t particle estimates of the posterior $p(\theta|y_{1:t})$. For the

resampling step, LW follows the idea in Pitt and Shephard (1999) and uses particles in period t to look ahead before sampling for period $t+1$. In particular, $g(x_t) = E(x_{t+1}|x_t, m(\theta_t))$ and $m(\theta_t) = a\theta_t + (1-a)\bar{\theta}_t$. Going from (2.14) to (2.15), $p(x_{t+1}, \theta_{t+1}|x_t, \theta_t)$ cancels with q^2_{t+1} as both decomposes into $p(x_{t+1}|\theta_t, x_t)p(\theta_{t+1}|x_t, \theta_t)$, where $p(\theta_{t+1}|x_t, \theta_t)$ is given by the LW mixture kernel (2.15).

Introduced by Gilks and Berzuini (2001), another way to explore the parameter space in the SMC setting is through the incorporation of Markov Chain moves that target parameter posteriors. In the state filtering literature, the use of MCMC moves is a tested method of alleviating sample degeneracy caused by successive resampling. With a properly designed Markov kernel, samples from an SMC method can be "jittered" to replenish variability and thus improve the quality of the posterior approximation. In a parameter estimation context, the key principle is to design at time t a Markov kernel $K_t(\theta'|\theta_t)$ whose invariant distribution is the parameter posterior $p(\theta|x_t, y_{1:t})$,

$$\int p(\theta|x_t, y_{1:t}) K_t(\theta'|\theta_t) d\theta_t = p(\theta'|x_t, y_{1:t}). \tag{2.18}$$

Constructing such Markov kernels is straightforward as all the usual MCMC theory applies. For example, if every element in $\theta = (\theta_1, \cdots, \theta_M)$ has a close-form posterior distribution, we can use the following Markov kernel based on the Gibbs sampler,

$$K_t(\theta'|\theta_t) = p(\theta'_{1:L}|x_{1:t}, y_{1:t}) \prod_{m=L+1}^{M} p(\theta'_m|\theta'_{1:m-1}, \theta_{m+1:M}, x_{1:t}, y_{1:t}). \tag{2.19}$$

In other words, we first sample $\theta'_{1:L} \sim p(\theta'_{1:L}|x_{1:t}, y_{1:t})$, then sample $\theta'_{L+1} \sim p(\theta'_{L+1}|\theta'_{1:L}, \theta_{L+2:M}, x_{1:t}, y_{1:t})$, then sample $\theta'_{L+2} \sim p(\theta'_{L+2}|\theta'_{1:L+1}, \theta_{L+3:M}, x_{1:t}, y_{1:t})$ and so on.

If $p(\theta'_m|\theta'_{1:m-1}, \theta_{m+1:M}, x_{1:t}, y_{1:t})$ is not available in close-form, one can sample from a properly tuned Metropolis-Hastings kernel $q(\theta'_m|\theta'_{1:m-1}, \theta_{m+1:M}, x_{1:t}, y_{1:t})$ instead,

and accept θ'_m with the Metropolis-Hastings acceptance probability

$$\min\left(1, \frac{p(\theta'|y_{1:t}, x_{1:t})}{p(\theta_{t-1}|y_{1:t}, x_{1:t})} \frac{q(\theta_{t-1}|\theta', y_{1:t}, x_{1:t})}{q(\theta'|\theta_{t-1}, y_{1:t}, x_{1:t})}\right) \tag{2.20}$$

$$= \min\left(1, \frac{p(y_t|x_t, \theta')p(x_t|x_{t-1}, \theta')}{p(y_t|x_t, \theta_{t-1})p(x_t|x_{t-1}, \theta_{t-1})} \frac{q(\theta_{t-1}|\theta', y_{1:t}, x_{1:t})}{q(\theta'|\theta_{t-1}, y_{1:t}, x_{1:t})}.\right) \tag{2.21}$$

2.3 Use of Sufficient Statistics

Parameter learning filters such as Liu and West (2001) have known sample degeneracy issues, the complexity of most structural models can only exacerbate this problem as the importance densities may be very different from the posterior distributions. This will lead to the importance weights in (2.17) having high variance, and thus the resulting particle approximation of the parameter posteriors will be unreliable. One strategy to remedy this problem is to use conditional sufficient statistics to create what are called Rao-Blackwellized filters, which can reduce the variance in the importance weights, thereby providing more efficient particle approximation of the target densities. Storvik (2002), Fearnhead (2002) and Carvalho et al. (2010) are some recent examples of SMC parameter estimation strategies built entirely around the efficient use of parameter sufficient statistics. Supplementing general parameter space traversal mechanisms such as the normal mixture shrinkage kernel in LW with conditional sufficient statistics will be the central parameter estimation strategy explored in this dissertation.

Let the set of model parameters be grouped into $\theta = (\phi, \varphi)$, where φ collects those parameters with conditional sufficient statistics and ϕ contains those without. Then for φ, we have

$$p(\varphi|x_t, y_t) = p(\varphi|s_t), \tag{2.22}$$

where $s_t = S(s_{t-1}, x_t, y_t)$ is a recursively defined lower dimensional sufficient statistics. Sampling new particles for φ reduces to sampling from their exact posterior

Table 2.1: Simple growth model: calibrated parameter values for simulated data and the choice of prior $p(\theta)$ for SMC estimation.

θ	calibration	$p(\theta)$
α	0.33	Uniform(0,1)
β	0.96	Uniform(0.75,1)
ρ	0.8	Uniform(0,1)
σ	0.05	Uniform(0,0.1)
σ^q	0.014	Uniform(0,0.1)
σ^i	0.02	Uniform(0,0.1)

distributions. Even in situations where complete Blackwellisation is not possible, the ability to sample from $p(\varphi|s_t)$ can enhance the overall estimation quality of θ by rejuvenating θ particles lost due to resampling. As a quick demonstration, I estimate the simple growth model in section 1.2 with 200 simulated observations (see Table 2.1 for calibration used). The parameters to be estimated are $\theta = (\alpha, \beta, \rho, \sigma, \sigma^q, \sigma^i)$, where σ^q and σ^i are the volatilities in the normal observational errors in equations (1.11) and (1.12), respectively. The model is estimated using both the exact algorithm proposed in Liu and West (2001) and a hybrid algorithm that uses the LW kernel on parameters $(\alpha, \beta, \sigma_q, \sigma_i)$ and draws from exact posteriors based on sufficient statistics for parameters (ρ, σ). Figures 2.1 and 2.2 show that despite having the same prior $p(\theta)$, the two algorithms performed quite differently. Note that while the two algorithms both tracked ρ and σ to their true values, the LW algorithm appears to have difficulties learning about the other parameters, in particular the nonlinear α and β. The hybrid algorithm, on the other hand, does markedly better than LW on all parameters despite only taking advantage of the conditional sufficient statistics structure on ρ and σ.

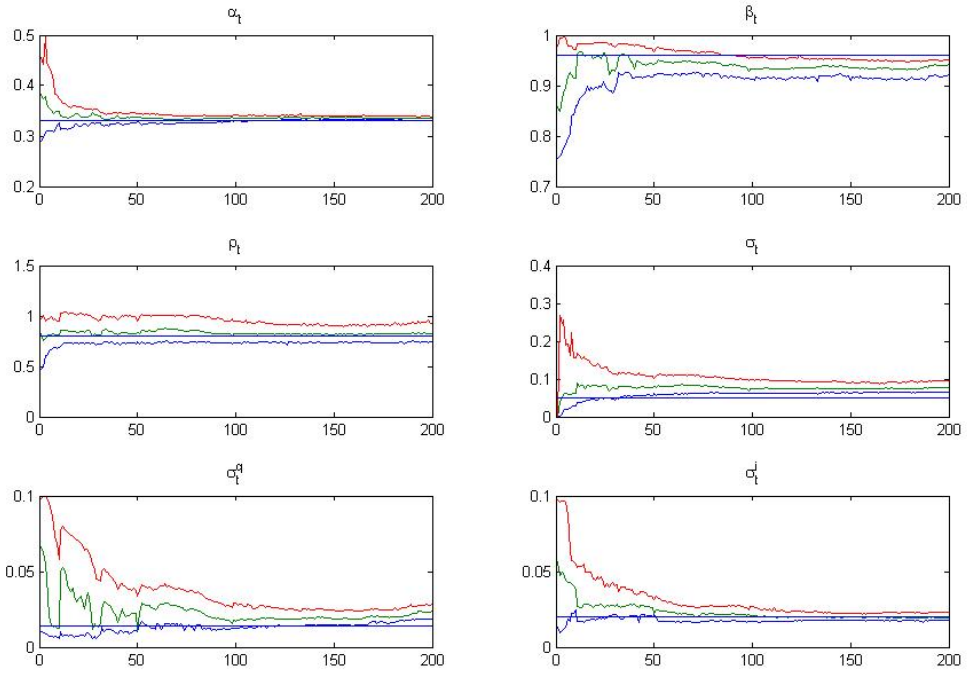

FIGURE 2.1: SMC estimation using LW supplemented with sufficient statistics on (ρ, σ). The calibrated parameter values are marked by the horizontal line. The red, green and blue lines are the estimated 2.5%, 50% and 97.5% posterior quantiles.

2.4 Theoretical Survey

Before proceeding to develop SMC methods for structural estimation, it's necessary to conduct a survey on the current state of theoretical development in the SMC literature. Generally speaking, the current SMC theoretical development is slightly lagged behind the empirical literature. The majority of the available SMC convergence theories apply to simpler settings, such as pure state filtering, or fixed parameter estimation with no latent state variables involved. A formal convergence study for the estimation problem in this dissertation would involve fixed parameters in a hidden Markov model without close-form state-space representations, which is very difficult if at all feasible. Nevertheless, a number of those theory papers not only

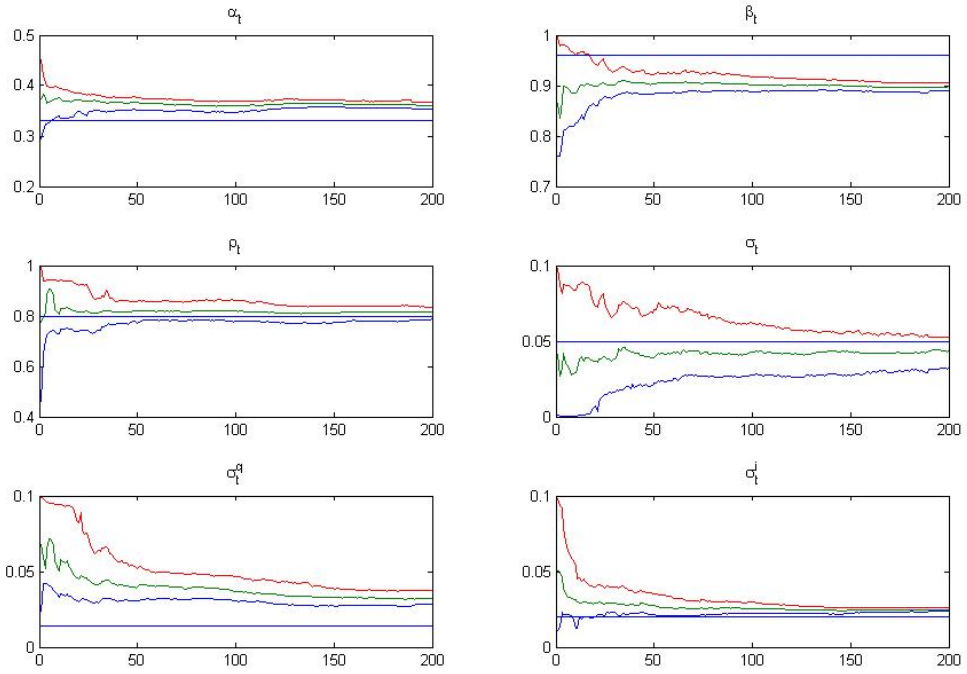

FIGURE 2.2: SMC estimation using LW only. The calibrated parameter values are marked by the horizontal line. The red, green and blue lines are the estimated 2.5%, 50% and 97.5% posterior quantiles.

provide starting points for a formal theoretical study of structural estimation using SMC methods, but also shed light on the possible regularity conditions under which they work as well as any performance limitations of this estimation framework. It's also worth mentioning that even though the current MCMC-based structural estimation methodology such as that employed in Fernández-Villaverde and Rubio-Ramírez (2005) and Gallant et al. (2010) incorporates particle filters for likelihood evaluation, parameter space traversal is still carried out with MCMC traversal mechanisms. The particle filters in those methods are essentially used as numerical integrators to integrate the latent state variables out of the likelihood function. As such, MCMC convergence theory still applies to those methods in terms of parameter estimation. One might be interested, however, to apply available SMC convergence theories to

investigate the convergence properties of the approximated likelihood functions in the aforementioned MCMC structural estimation framework. This is a related but different issue already visited by Fernández-Villaverde et al. (2006) and Ackerberg et al. (2009).

In the current probability literature, a number of papers have established central limit theorems for standard SMC methods (i.e., sequential importance sampling with a resampling scheme) under different conditions. As any SMC method will suffer from severe particle degeneracy problems for a fixed number of particles and a large enough T, the standard mode of analysis in developing those CLTs is to fix T and investigate the limiting properties of the empirical measure of the target densities as the particle size goes to infinity. To cite a few notable ones, Chopin (2004) developed a CLT for particle filters under minimal assumptions (target particles must satisfy some finite variance condition in the Euclidean norm). Based on this CLT, a theoretical framework is then provided to study the rate of divergence of asymptotic variances of different particle filtering methods. The filtering scenarios examined in the paper, however, does not include HMMs with fixed parameters due to model complexity. Künsch (2005) developed a CLT in the L1-norm, assuming only that the state transition equation is continuous and the observation equation is positive and upper bounded. This paper also established the conditions under which uniform convergence of the particle filter could be achieved at all t. The results in this paper, however, only apply to the pure state filtering problem. Gilks and Berzuini (2001) established a CLT for their resample-move algorithm, which was originally designed to perform fixed parameter estimation with no latent state variables. In addition, Shiga and Tanaka (1985) and Moral and Guionnet (1999) both developed CLTs using interacting particle theory.

Due to the complexity of the structural models considered, it is rather difficult to check the conditions required to quote the CLTs cited in the last paragraph for

structural estimation. Crisan and Doucet (2002), on the other hand, has established several convergence results in the weak topology without developing a CLT, and could potentially be applied to the methods developed in this dissertation. *Theorem 1* from Crisan and Doucet (2002) establishes almost sure convergence, requiring only that the state transition kernel be Feller and the likelihood function be bounded, continuous and strictly positive. If we treat fixed parameters as another state variable, then the mixture kernel in Liu and West (2001) and the random-walk Metropolis kernel in a Markov-move method could be both seen as the transition kernels for θ. Both clearly satisfy the Feller property, which essentially states that the result of applying the transition kernel to a bounded continuous function is again a bounded continuous function. The second condition for achieving almost sure convergence is satisfied by the DSGE likelihoods considered in chapter 3, although not by the game model in chapter 4 as the dynamic game likelihood is not continuous. *Theorem 2* of Crisan and Doucet (2002) establishes convergence in the mean square sense, requiring only the importance weights be upper bounded and a standard resampling scheme. Since all the methods developed in later chapters use the standard multinomial resampling scheme, and the importance weights appear to be well behaved empirically, this theorem could potentially apply to structural estimation as well. Another useful albeit negative result from Crisan and Doucet (2002) is that uniform convergence is not possible when fixed parameters are involved as the model is not ergodic anymore. Reassuringly, one of the conditions for uniform convergence in the L1-norm in Künsch (2005) is that the state transition densities be the same for all t, this is clearly violated by DSGE models since the computed law of motion is dependent on the parameter proposal at each t.

3

SMC Estimation of DSGE Models

The formal estimation of dynamic stochastic general equilibrium (DSGE) 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 characterized by a likelihood function for the observables, which in turn is driven by a set of structural parameters representing preferences and technology in the modeled 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 hard to evaluate. The likelihood function is usually a high dimensional object that exhibits nonlinearity in both parameters and state variables and often also features non-normal shocks. As such 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 et al. (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. The standard thing to do is to solve the linearized model with normal shocks, and then evaluate the resulting likelihood using the Kalman filter. Fernández-Villaverde and Rubio-Ramírez (2005) (henceforth FVRR) showed that better model fit could be achieved by using a sequential Monte Carlo filter to evaluate the likelihood, as the SMC filter allows the likelihood to be constructed from a nonlinear and non-normal state space representation. The latest Bayesian estimation technique for DSGE models is thus to embed a nonlinear state filter inside an MCMC routine such as a Metropolis-Hastings kernel, the outer MCMC loop will explore the parameter space while the embedded state filter constructs the likelihood function with the latent state variables integrated out.

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 to be viable, while Gilks and Berzuini (2001), Carvalho et al. (2010) 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 Srikanth (2010) proposed an MCMC method that features random parameter clustering and tailored 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, parallelization is straightforward when implementing SMC methods 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 chapter 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 filter-embedded MCMC approach from Fernández-Villaverde and Rubio-Ramírez (2005) and a full SMC approach utilizing parameter sufficient statistics. In this example, we show that an SMC estimation method could provide comparable DSGE estimation results to MCMC methods. We then apply SMC estimation to a large scale DSGE model built for the Euro area by Smets and Wouters (2003). We 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 vector autoregressions (VAR) or Bayesian vector autoregressions (BVAR).

3.1 Two Bayesian Estimation Methods

The main problem related to estimating DSGE models is the evaluation of the likelihood. 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 SMC filter is used to integrate out 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.

3.1.1 MCMC with Embedded State Filter

An MCMC-based estimation requires the evaluation of the model likelihood

$$p(y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|y_{1:t-1}, \theta). \tag{3.1}$$

A DSGE model has a likelihood that's nonlinear in terms of state variables and fixed parameters, thus two types of likelihood evaluation methods are used for MCMC based estimation methods. 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. 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,

$$p(y_{1:T}|\theta) \approx p^N(y_{1:T}|\theta) = \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p(y_t|y_{1:t-1}, x_t^{(i)}, \theta)\delta_{x_t^{(i)}}(x_t), \tag{3.2}$$

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

$$p(y_t|y_{1:t-1}, \theta) = \int p(x_{t-1}|y_{1:t-1}, \theta)p(x_t|x_{t-1}, \theta)p(y_t|x_t, \theta)dx_{t-1:t} \tag{3.3}$$

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

$\{x^i_{0:T}\}_{i=1}^N$. A slightly modified bootstrap filter from Doucet et al. (2001) could fullfill the purpose of likelihood evaluation and is given as follows:

1. *Initialization:* 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 $p(y_{1:T}|\theta)$. 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) provide an excellent tutorial on particle filters that demonstrates in detail the necessity of the resampling step in any SMC method.

3.1.2 SMC Joint State and Parameter 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 auxiliary particle filter developed by Pitt and Shephard (1999), 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 et al. (2008) and Carvalho et al. (2010) for the main reason that a general algorithm that doesn't rely on exact parameter posteriors is needed for DSGE estimation. 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 GB, 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 et al. (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 et al. (2008) 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 absence of sufficient statistics. The particle learning algorithm proposed in Carvalho et al. (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.

Algorithm 1: LW with Sufficient Statistics

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

$$p(\theta|y_{1:T}) \approx \sum_{i=1}^{N} w_t^{(i)} N(\theta|m_t^{(i)}, h^2 V_t) \tag{3.4}$$

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

locations to be

$$m_t^{(i)} = a\theta_t^{(i)} + (1-a)\bar{\theta}_t \tag{3.5}$$

where $a = \sqrt{1-h^2}$. We 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. For those model parameters with sufficient statistics, particle evolution boils down to sampling from their exact posterior distributions $p(\varphi|s_t)$, where $s_t = S(s_{t-1}, x_t, y_t)$ is a recursively defined sufficient statistics. 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

$$ESS(t) = \left(\sum_{i=1}^{N} (w_t^{(i)})^2 \right)^{-1}. \tag{3.6}$$

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 et al. (2005) provides a comparative study on some of the most popular resampling schemes. We use the multinomial resampling scheme for our SMC implementation. Group the model parameters into $\theta = (\phi, \varphi)$, where ϕ and φ denote the subsets of parameters without and with sufficient statistics conditional on the states, respectively. We propose the following algorithm for DSGE estimation:

1. *Initialization:* 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. *One step prediction:* 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 $i = 1, \cdots, N$, 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 (3.4) and V_t is the Monte Carlo variance of particles $\{\phi_t^{(i)}\}_{i=1}^N$.

5. *Evolve states:* For $i = 1, \cdots, N$, 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. *Sample φ:* For $i = 1, \cdots, N$, 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 importance 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. Therefore before building the weights $g_{t+1}^{(i)}$ in step 3, the observation equations must be re-solved to reflect the newly sampled $m_t^{(i)}$ particles. Since solving the model is the computational bottleneck in DSGE estimation, we turn to the SMC method from Gilks and Berzuini (2001) (henceforth GB) to estimate the SW model.

Algorithm 2: 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 in some way a generalization of the particle learning algorithm proposed in Carvalho et al. (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

$$w_{t+1}^{(i)} = \frac{p(x_t^{(i)}, \theta_t^{(i)}|y_{1:t+1})}{p(x_t^{(i)}, \theta_t^{(i)}|y_{1:t})} \propto \frac{p(y_{1:t+1}|x_t^{(i)}, \theta_t^{(i)})p(x_t^{(i)}, \theta_t^{(i)}|y_{1:t})}{p(x_t^{(i)}, \theta_t^{(i)}|y_{1:t})} \tag{3.7}$$

$$\propto p(y_{1:t+1}|x_t^{(i)}, \theta_t^{(i)}), \tag{3.8}$$

which is exactly the resampling weight in Carvalho et al. (2010). Incorporating the use of sufficient statistics in GB then simply means using a mix of Gibbs and Metropolis Hastings moves to evolve the particles. We modify the original GB algorithm slightly to accommodate the use of sufficient statistics, and propose the following algorithm:

1. *Initialization:* 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 Sampling:* For $i = 1, \cdots, N$, sample $\tilde{x}_{t+1}^{(i)} \sim p(x_{t+1}|x_t^{(i)}, \phi_t^{(i)}, \varphi_t^{(i)})$, 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 $k^{1:N} \sim Multinomial(N, w_{t+1}^{1:N})$.

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

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

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

6. *Sample φ:* For $i = 1, \cdots, N$, 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 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.

3.2 Estimation of a Neoclassical Growth Model

We test the above modified LW algorithm on a stochastic neoclassical growth model without leisure. Adding leisure to the model would only add the intratemporal first order condition to the list of equilibrium conditions in solving the model and would not otherwise change the estimation algorithm. The neoclassical growth model consists of the following relationships: a single good in the economy 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 modeling 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:

$$E_0 \sum_{t=1}^{\infty} \beta^t \frac{c_t^{1-\tau}}{1-\tau} \tag{3.9}$$

where β is the discount factor, E_0 is the conditional expectation operator. Let us assume that technology evolves according to a stationary AR(1) process, i.e. $z_{t+1} = \rho z_t + \epsilon_t$ where $\epsilon_t \sim N(0, \sigma^2)$ and $|\rho| < 1$. 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:

$$c_t^{-\tau} = \beta E_0 c_{t+1}^{-\tau} [\alpha e^{z_{t+1}} k_{t+1}^{\alpha-1} + 1 - \delta], \tag{3.10}$$

$$c_t + k_{t+1} = q_t + k_t(1-\delta), \tag{3.11}$$

$$z_t = \rho z_{t-1} + \epsilon_t. \tag{3.12}$$

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 maximization problem as functions of the state variables, namely capital and technology.

The above system of equations does not have a closed form solution, thus must be solved with numerical methods. Aruoba et al. (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 programming. Pioneered by Hall (1971) and Magill (1977) and extended by Judd and Guu (1993) and Gasper 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):

$$\hat{c}_t = \alpha_1 \hat{k}_t + \alpha_2 z_t + \alpha_3 \hat{k}_t^2 + \alpha_4 z_t^2 + \alpha_5 \hat{k}_t z_t + \alpha_6 \sigma^2, \tag{3.13}$$

$$\hat{k}_{t+1} = \beta_1 \hat{k}_t + \beta_2 z_t + \beta_3 \hat{k}_t^2 + \beta_4 z_t^2 + \beta_5 \hat{k}_t z_t + \beta_6 \sigma^2, \tag{3.14}$$

where the $\hat{}$ above a variable denotes its log deviation from the steady state. To find the α and β coefficients in the above second order approximation, we plug the policy functions (3.13) and (3.14) into the equilibrium conditions (3.10)-(3.12) 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, 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 (3.12) and (3.14) are the transition equations of the state variables and the measurement equations are given by

$$q_t = e^{z_t} k_t^{\alpha} + \epsilon_q \tag{3.15}$$

$$i_t = k_{t+1} - k_t(1-\delta) + \epsilon_i, \tag{3.16}$$

where $\epsilon_q \sim N(0, \sigma_q^2)$ and $\epsilon_i \sim N(0, \sigma_i^2)$.

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

$$p(\rho|x_{1:t}, y_{1:t}, \sigma) \sim N(b_t, \sigma B_t^{-1}) \tag{3.17}$$

$$p((\sigma^2)^{-1}|x_{1:t}, y_{1:t}) \sim Ga((t-2)/2, R_t(b)/2), \tag{3.18}$$

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 $s_t = (b_t, B_t, R_t)$.

Estimation Results

We simulated 80 observations of investment and output from the model with the calibration in table 3.1. This calibration mimics that from FVRR and reflects empirical findings on real economies. We estimate this artificial data set 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

Table 3.1: Neoclassical growth model: calibrated parameter values for simulated data and the choice of prior $p(\theta)$ for SMC and MCMC estimation.

θ	calibration	$p(\theta)$
α	0.4	Uniform(0,0.5)
β	0.9896	Uniform(0.75,1)
δ	0.02	Uniform(0,0.05)
ρ	0.95	Uniform(0.5,1)
σ	0.007	Uniform(0,0.1)
σ^q	0.2	Uniform(0,0.25)
σ^i	0.2	Uniform(0,0.25)
τ	2	Uniform(0,5)

with a burn-in of 1000 considering the prior specification showed by table 3.1. The same prior information is used in SMC estimation.

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 3.1 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 3.2 and 3.3 show the estimated posteriors using both estimation methods. The vertical red lines denote the artificial parameter values we simulate from.

Discussion

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 on figures 3.2 and 3.3 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 figure 3.4,

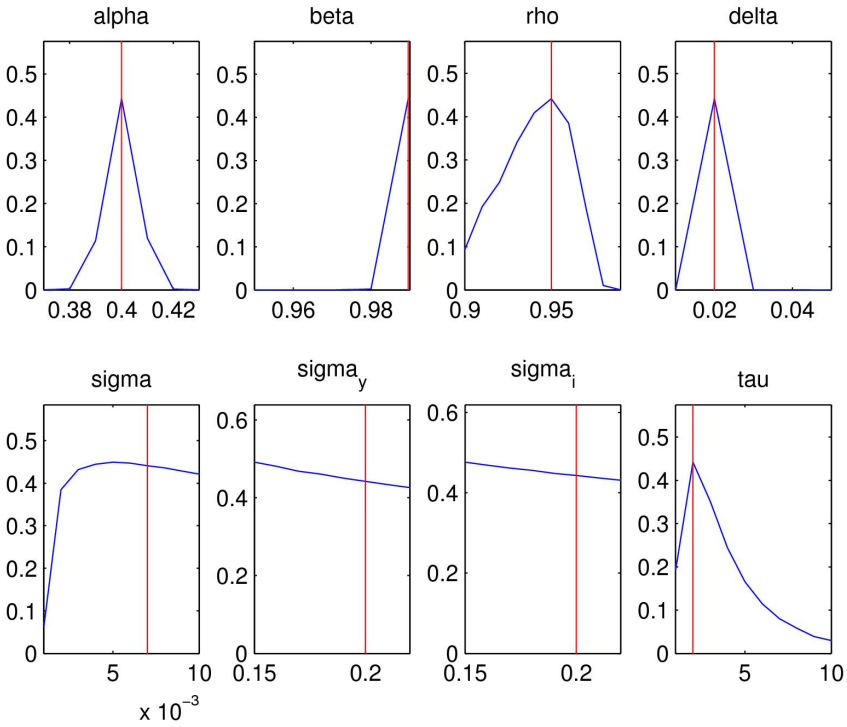

FIGURE 3.1: Profile likelihoods of the neoclassical growth model. The vertical red lines mark the parameter calibration used for data simulation.

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.

To investigate the stability of the proposed SMC algorithm for DSGE estimation, the estimation is repeated over the same data set using different random seeds. To speed up this study, a linear approximation to the policy function is used instead of the second-order approximation, meaning all the second-order terms involving $\hat{k}_t$ and z_t in (3.13) and (3.14) are now dropped from the policy function approximation.

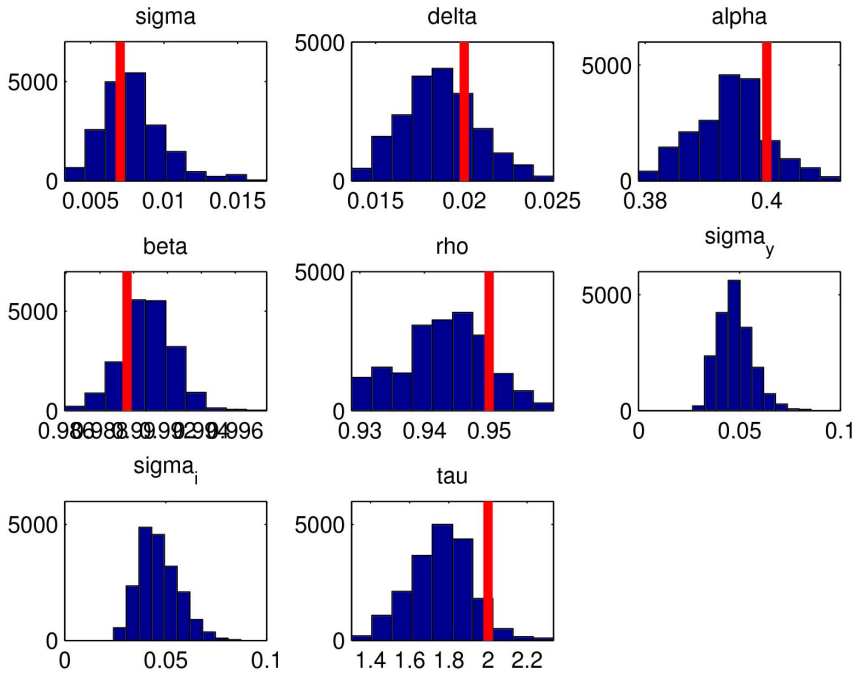

FIGURE 3.2: Posterior histograms from MCMC estimation using random-walk Metropolis-Hastings kernel. MCMC sample size is 20,000 with 1,000 burn-in. The vertical red lines mark the parameter calibration used for data simulation.

The estimation quality will suffer as a result, but will not alter our interpretation of the stability of the proposed SMC algorithm. The estimation is repeated five times, each using a different seed. From figures 3.6 and 3.7, we can see that parameter estimates for σ, β, and σ_i are fairly stable across different seeds. On the other hand, Parameters such as δ, ρ and τ appear to be very sensitive to initial seeds. These results suggest that SMC algorithms could suffer from instability for certain nonlinear parameters in the model. On the other hand, these results were all from estimation runs using 5000 particles, which is a fairly small number of posterior samples. More in-depth stability studies, using more particles, and across data sets of varying lengths should be looked into before final conclusions can be drawn on the stability of SMC estimation of DSGE models.

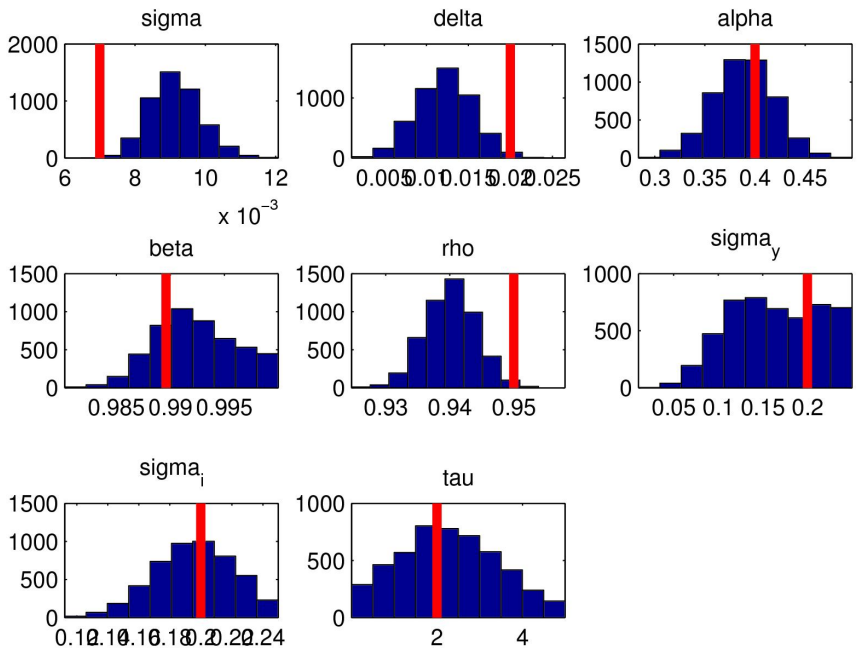

FIGURE 3.3: Posterior histograms from SMC estimation at the last data point. SMC sample size is 5,000. The vertical red lines mark the parameter calibration used for data simulation.

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 3.4 shows how the particle approximation of the posteriors for parameters β, α and ρ change over time. Having access to posteriors at each data point allows one to understand the extent particular observations influence parameter inference. 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

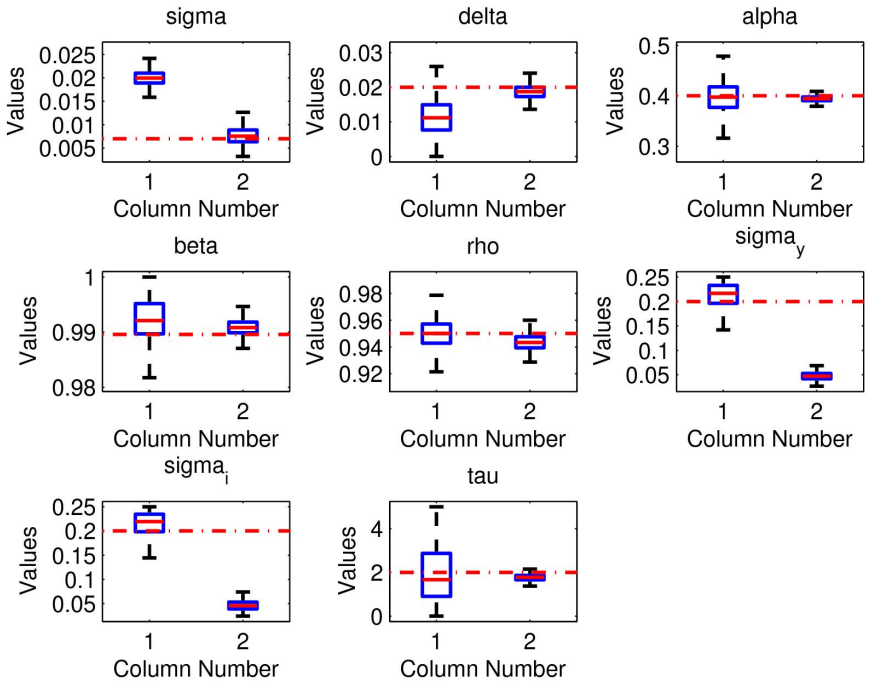

FIGURE 3.4: Boxplot comparison of estimation quality between typical MCMC and SMC runs using the same settings described (SMC is on the left).

$\mathcal{O}(N)$ space for inference with N particles[1].

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 $\{\theta_t^{(i)}\}_{i=1,\cdots,N}$ 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

[1] 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.

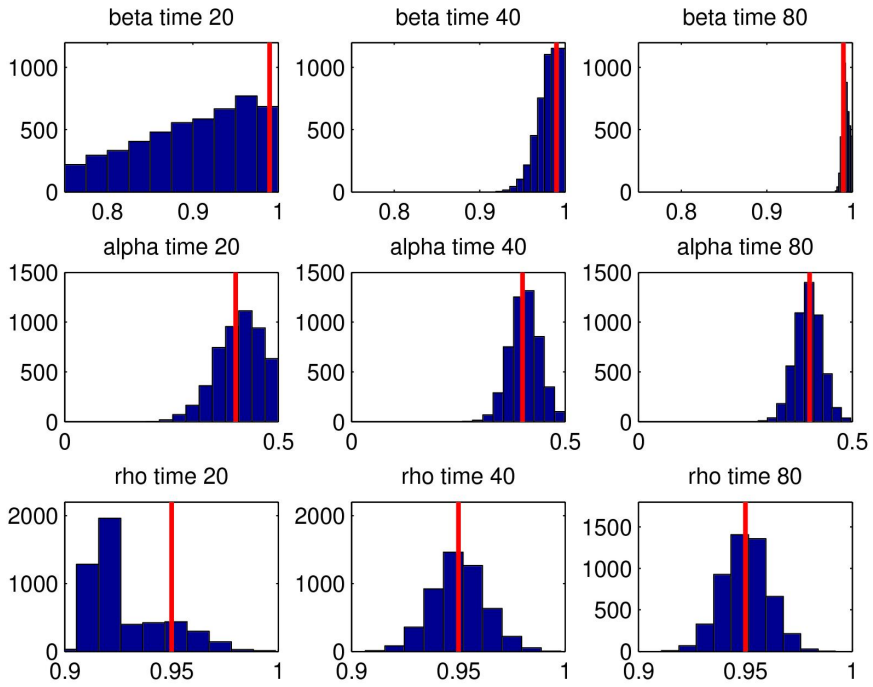

FIGURE 3.5: Evolution of parameter posteriors for (β, α, ρ) as an SMC algorithm progresses through the data set.

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).

3.3 Structural vs Reduced-form

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

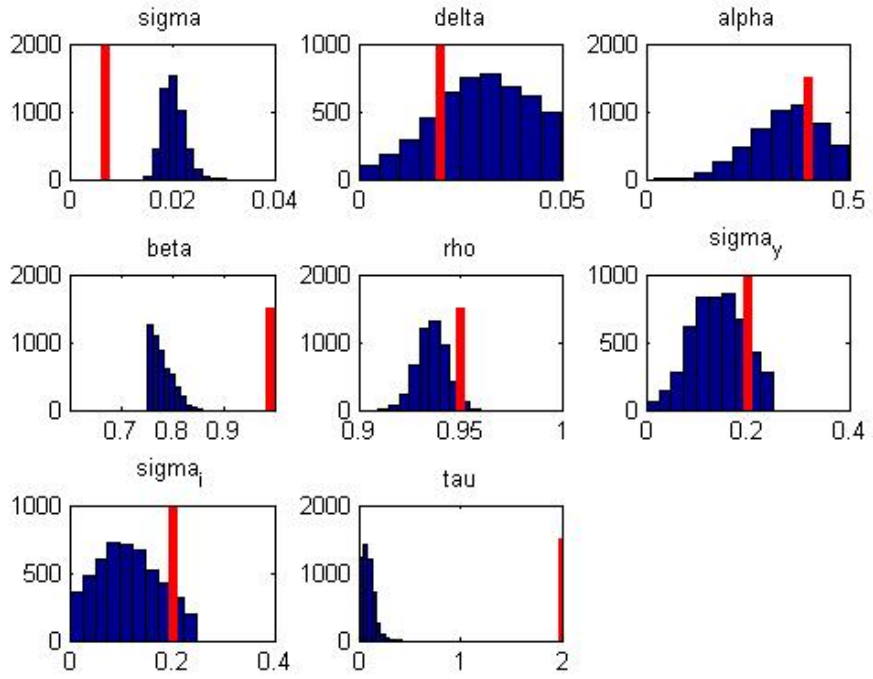

FIGURE 3.6: Stability study: one typical estimation run with 5000 particles on the same simulated data set using linear policy function approximation.

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 Wouters (2003).

DSGE Model for the Euro Area

Smets and Wouters (2003) (henceforth SW) developed a DSGE model of the European economy with sticky prices and wages and estimated the linearized 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

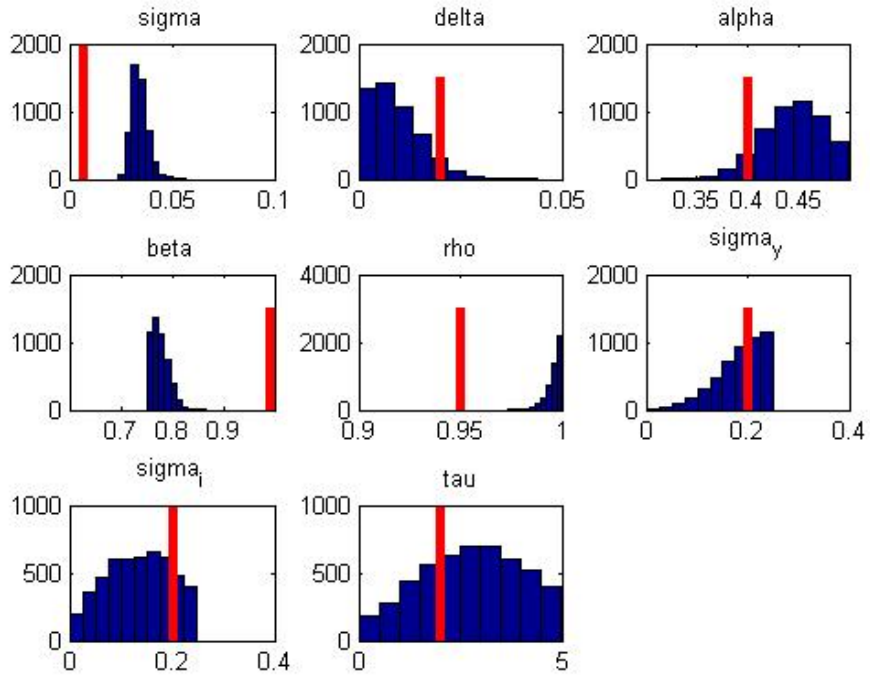

FIGURE 3.7: Stability study: another typical estimation run with 5000 particles on the same simulated data set using linear policy function approximation.

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 labor market and choose how to best invest their capital stock. Firms engage in monopolistic 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, labor, 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.

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 pronounced issue in this case so we use the Metropolis-Hastings kernel on all model parameters. Table 3.2 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 [fig13] shows that the average effective sample size of SMC estimation is around 30% of the original SMC samples simulated for a large scale DSGE model such as the SW model. 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 utilization. All of this is evidence that a much larger number of particles is needed to properly visit the parameter posteriors.

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 data set would probably take around a week to finish.

Bayesian Sequential Model Comparison

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. The frequentist tests from

Table 3.2: Comparison of posterior quantiles between SMC results from the GB algorithm and MCMC results from Smets and Wouters (2003).

	SMC			MCMC		
	5%	50%	75%	5%	50%	75%
σ productivity	0.0005	0.0097	0.0712	0.444	0.612	0.873
σ inflation obj.	0.0004	0.0056	0.0543	0.011	0.023	0.069
σ cons. pref.	0.1558	0.9560	1.7404	0.173	0.297	0.571
σ gov. spending.	0.0596	0.1671	0.4985	0.290	0.329	0.378
σ labor supply	0.0052	0.0364	0.3729	0.997	1.658	2.603
σ interest rate	0.0321	0.0776	0.1503	0.102	0.129	0.158
σ investment	0.0001	0.0102	0.5036	0.099	0.129	0.247
σ equity premium	0.0503	0.1047	0.2039	0.520	0.611	0.718
σ wage mark up	0.0312	0.2083	0.6230	0.246	0.285	0.331
σ price mark up	0.0424	0.0757	0.1627	0.139	0.162	0.192
ρ productivity	0.1067	0.9835	1.0000	0.712	0.828	0.912
ρ inflation obj.	0.0033	1.0000	1.0000	0.658	0.865	0.970
ρ cons. pref.	0.1861	0.6185	0.9562	0.817	0.886	0.931
ρ gov. spending	0.9973	0.9999	1.0000	0.912	0.956	0.982
ρ labor supply	0.1067	0.9835	1.0000	0.916	0.955	0.98
ρ investment	0.0014	0.9898	1.0000	0.856	0.917	0.961
investment adj cost	-0.1644	0.3713	0.9354	4.321	5.974	7.973
σ consump. util.	0.3834	1.4140	2.6195	1.126	1.608	2.106
h consump. habit	0.4266	0.8630	0.9758	0.416	0.552	0.681
σ labor util.	1.0905	3.5727	6.6870	0.439	1.188	2.365
fixed cost	0.5880	1.3625	2.2390	1.199	1.487	1.835
calvo employment	0.0260	0.1994	0.6370	0.503	0.596	0.671
capt. util. adj. cost	0.0242	0.2848	0.7572	0.062	0.175	0.289
calvo wages	0.0812	0.8473	0.9977	0.690	0.758	0.817
calvo prices	0.9153	0.9918	0.9996	0.890	0.909	0.927
indexation wages	0.1067	0.9835	1.0000	0.383	0.663	0.900
indexation prices	0.1067	0.9835	1.0000	0.268	0.425	0.597
r inflation	0.0032	1.8120	3.8169	1.537	1.661	1.821
r d(inflation)	-0.9020	0.0508	1.0218	0.134	0.221	0.313
r lagged int. rate	0.7600	0.9450	0.9873	0.901	0.931	0.946
r d(output)	-0.2840	0.2540	1.1162	0.131	0.173	0.219
r output	0.1558	0.9560	1.7404	0.079	0.143	0.215
r ϵ productivity	-1.1588	-0.4013	0.6628	0.043	0.086	0.137
r ϵ labor	-1.0262	-0.1567	0.4230	0.007	0.030	0.063

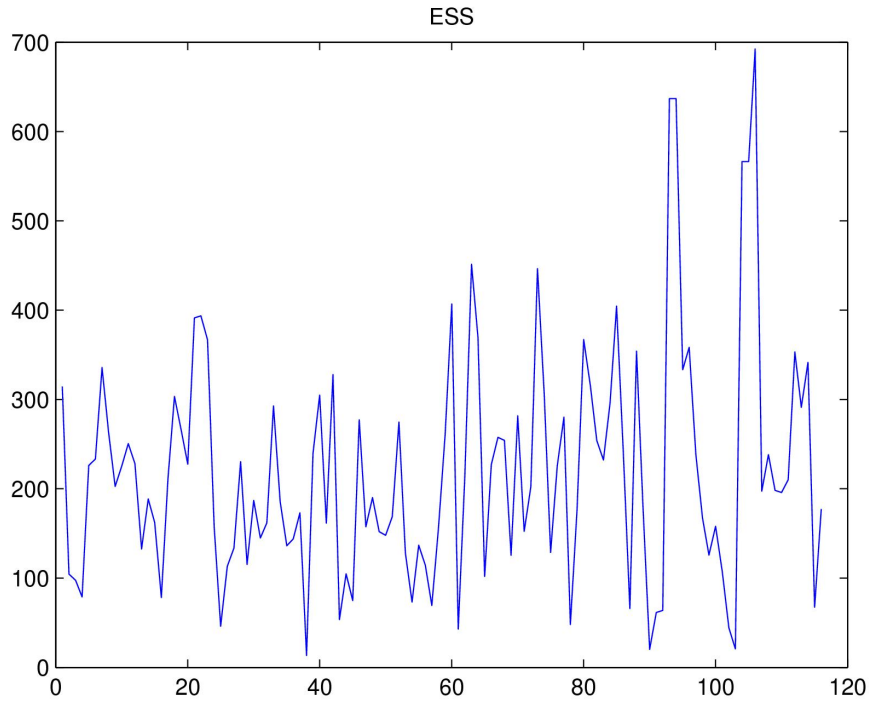

FIGURE 3.8: Effective sample size from SMC estimation of the SW DSGE model using the proposed GB algorithm with 1000 particles.

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. 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.

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 $\left\{x_t^{(i)}, \theta_t^{(i)}\right\}$ to approximate the marginal likelihood of the model by

$$p(y_{t+1}|y_{1:t}) \approx \frac{1}{N}\sum_{i=1}^{N} p(y_{t+1}|x_t^{(i)}, \theta^{(i)}). \tag{3.19}$$

This allows Bayes factors to be computed at each time point in the data set 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 and at time is given by

$$\frac{p(M_1|y_{1:t})}{p(M_2|y_{1:t})}, \tag{3.20}$$

where $p(M_i|y_{1:t}) = \prod_{k=1}^{t} p(y_k|y_{1:k-1}, M_i)$. Those sequentially computed Bayes factors allow us to perform a Bayesian version of the sequential tests in Giacomini and Rossi (2007).

We used 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 iterations. 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 performance 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.

All model comparisons exercises 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

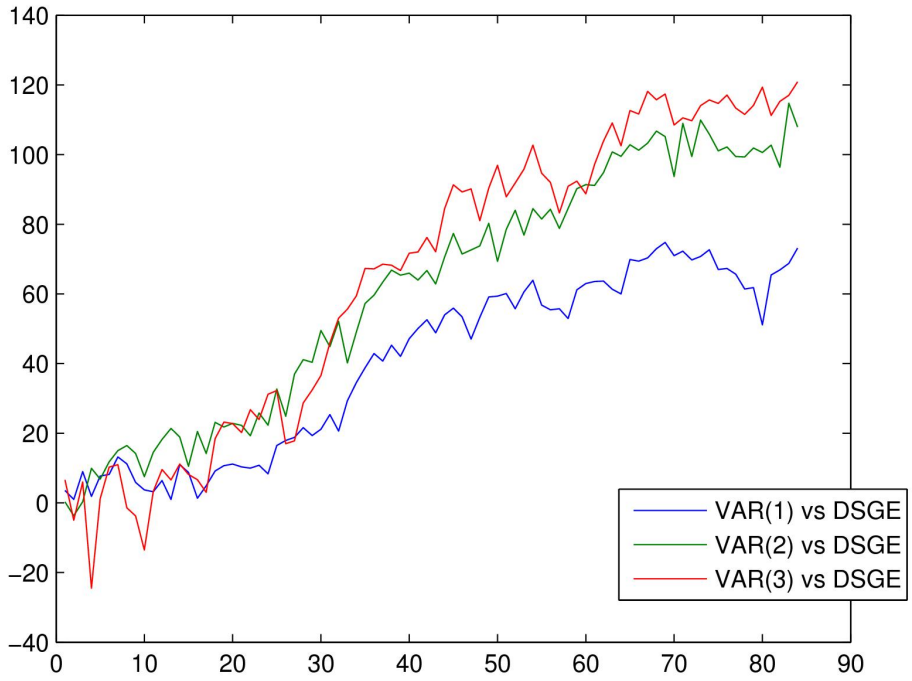

FIGURE 3.9: Sequential Log Bayes Factor: VAR vs SW DSGE.

Bayes factor is computed relative to the DSGE model. Figure 3.11 shows the particle approximation of the DSGE marginal likelihoods across the data set 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 3.7 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 delivers. 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 perfor-

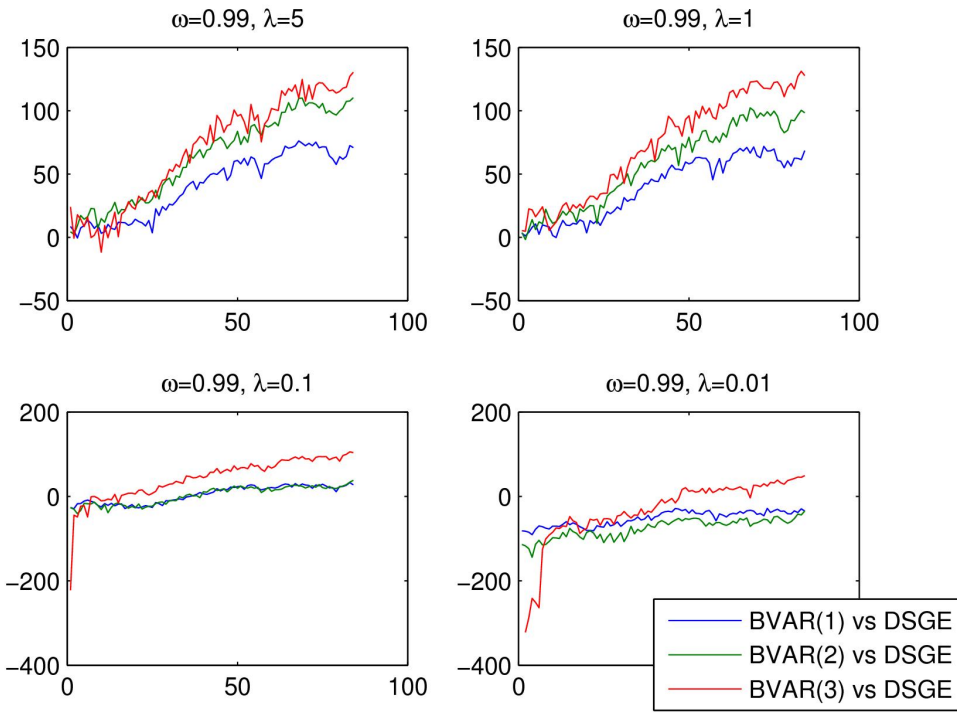

FIGURE 3.10: Sequential Log Bayes Factor (ω=0.99): BVAR vs DSGE.

mance is sensitive to the choice of the prior hyperparameter λ but not ω. In each of figure 3.8, 3.9 and 3.10, 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

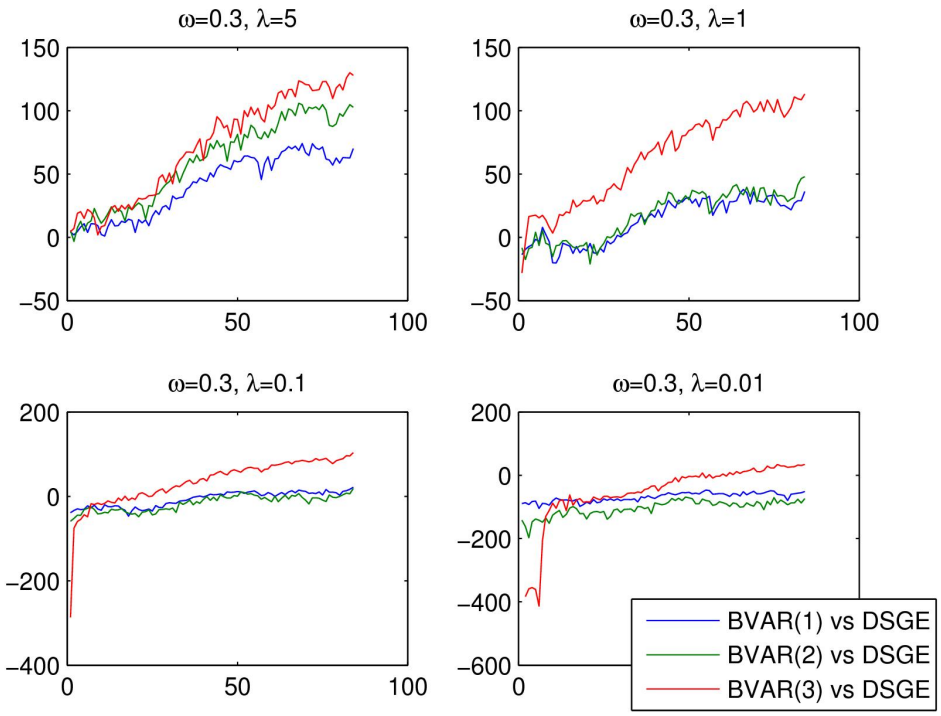

FIGURE 3.11: Sequential Log Bayes Factor (ω=0.3): BVAR vs DSGE.

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

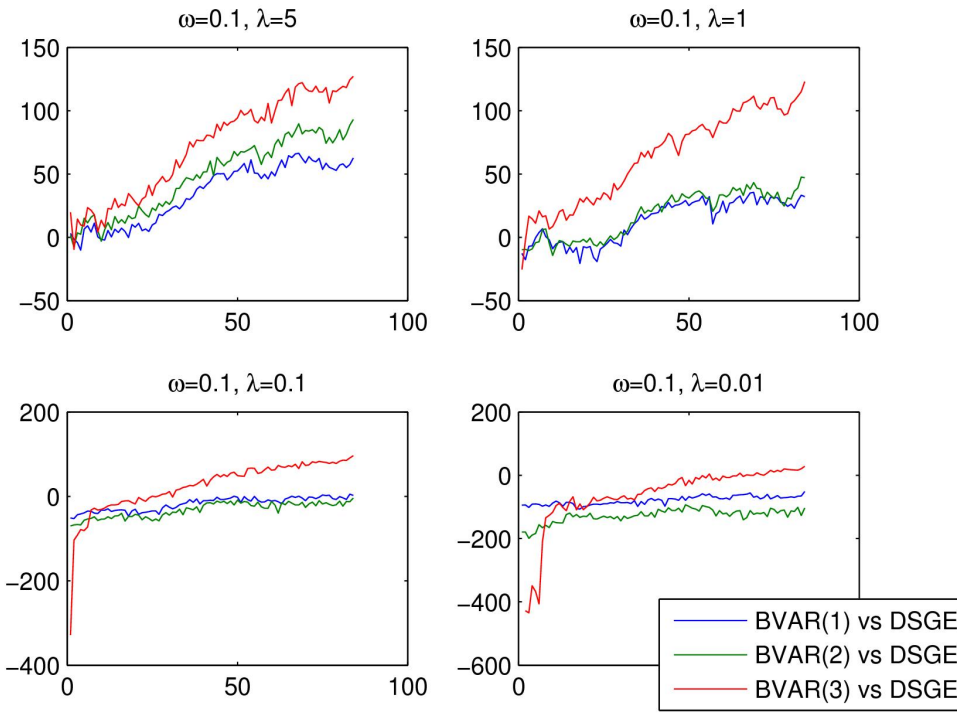

FIGURE 3.12: Sequential Log Bayes Factor (ω=0.1): BVAR vs DSGE.

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.

3.4 Concluding Discussion

In this chapter 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.

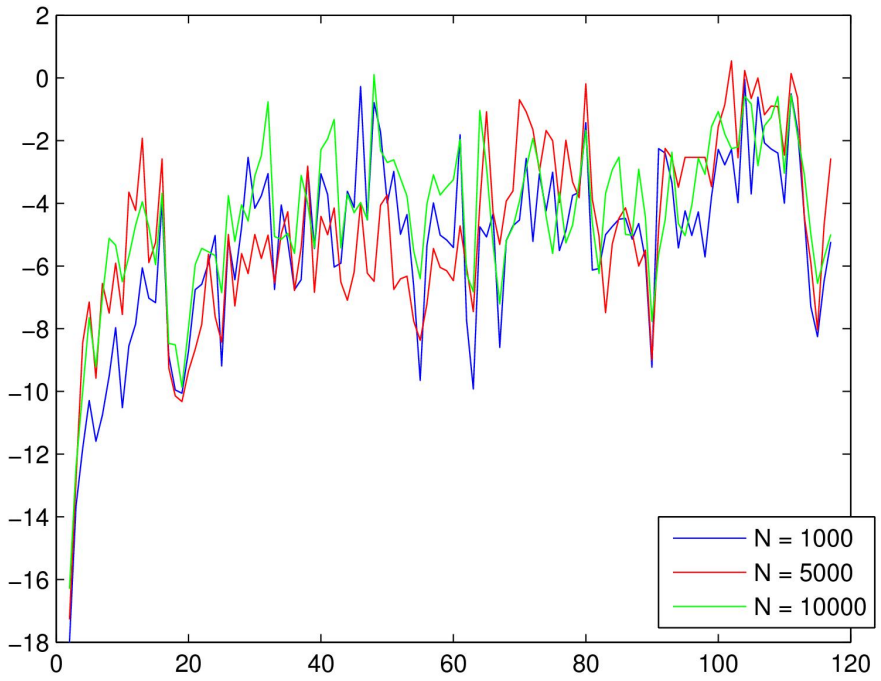

FIGURE 3.13: Stability of SMC estimation of $p(y_t|y_{t-1})$ from the SW-DSGE model.

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 reverse 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 data set. 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 data set whereas SMC doesn't.

In summary, SMC has many practical benefits over MCMC but needs additional theoretical 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 work will provide a fresh perspective in the ever growing Bayesian econometric literature.

4

SMC Estimation of a Dynamic Game

Due to the pivotal role game theory plays in fields such as industrial organization (IO), there has been a constant stream of literature pertaining to the structural estimation of game theoretic models over the last two decades. For an excellent up-to-date survey on this subject, see Aguirregabiria and Mira (2004). Much of the existing work in the area focuses on games under the incomplete information assumption. Under the complete information setting, static games have been studied by Bresnahan and Reiss (1990), Berry (1992), Tamer (2003), Bajari et al. (2004), and Ciliberto and Tamer (2009). Other than Gallant et al. (2010), there appears to be no other work on the estimation of dynamic games of complete information. This disproportional spread in the game estimation literature can be attributed to the relative simplicity of the two-step estimation strategy that is commonly applied to games of incomplete information and the high computational cost of obtaining game equilibria under the complete information setting.

Though a challenging task both conceptually and computationally, the estimation of dynamic games of complete information is critical to the development of the empirical IO literature as this class of game theoretic models better capture

the strategic interaction between firms in market competition by allowing for unobserved firm level heterogeneity. In this paper, I propose an alternative Bayesian approach for the estimation of dynamic games of complete information that allows for serially correlated unobserved endogenous state variables. I demonstrate through application of this estimation method on a dynamic oligopolistic game of entry that it's more computationally efficient than the MCMC based approach used in Gallant et al. (2010) (henceforth GHK).

Currently the most popular method for game estimation is a two-step estimation strategy based on the CCP method from Hotz and Miller (1993). In the first stage the econometrician backs out the conditional choice probabilities, that is the probabilities that actions from some finite action set is played conditional on some finite vector of state variables, from the reduced form of the model. In a second stage, structural parameters are recovered by finding parameter values that best rationalize the choice specific value functions from the first stage. The CCP method's main attraction is its computational advantage over a full solution method such as the nested fixed point algorithm used on the engine replacement model in Rust (1987). Unlike a full solution method, the two stage estimation procedure avoids solving repeated DPs, thereby greatly reducing computational cost. There have been several important extensions over the original CCP method. For example, Hotz et al. (1994) proposed a simulation-based CCP estimator that can deal with very large state spaces, Aguirregabiria and Mira (2002) proposed a recursive CCP method that significantly reduces finite sample bias, and Aguirregabiria and Mira (2004) showed that the recursive CCP method also handles discrete unobserved heterogeneity. The main drawback to the CCP class of estimators is that it's difficult to incorporate unobserved endogenous heterogeneity with serial correlation. One way to deal with unobserved heterogeneity is to adopt a finite mixture model for the unobservables. Arcidiacono and Miller (2008) uses the EM algorithm along with a finite mixture

model to allow for serially correlated unobserved state variables. The well-known Geweke-Hajivassilion-Keane estimator uses importance sampling simulation to address normally distributed unobservables in discrete choice models. To the best of my knowledge, GHK is the only work that addresses endogenous unobservables with serial correlation in a dynamic game setting.

This chapter primarily contributes to the currently small but growing Bayesian literature on the estimation of dynamic discrete choice models. Bayesian estimation approaches for this class of models usually require prohibitively high computational effort to evaluate the likelihood function for posterior inference. With an MCMC algorithm, each parameter proposal requires the likelihood function to be re-evaluated, and thus the dynamic program associated with the problem to be solved. Combine that with a full-solution DP solver and unobserved state variables that need to be integrated out of the likelihood, Bayesian estimation quickly becomes borderline infeasible. Bayesian simulations-based methods, however, are natural candidates for the estimation of models with serially correlated unobservables. The main issue in estimating this class of models is the high dimensional integration over the histories of unobserved state variables in extracting either the conditional choice probabilities in CCP methods, or the model posterior distributions in Bayesian methods. Those high dimensional integrals can be readily approximated using Monte Carlo integration techniques. For this reason, most recent work in the estimation of dynamic discrete choice models featuring serially correlated unobservables are Bayesian in nature. Imai et al. (2005) and Norets (2009) use MCMC algorithms to integrate out unobserved state variables in single agent problems. Both papers also avoid obtaining full DP solutions by updating the Bellman equation once with each parameter proposal, thus greatly reducing computational cost. GHK uses a particle filter to integrate out unobserved state variables in a dynamic game of entry and does posterior inference with a random walk Metropolis-Hastings kernel, an approach sim-

ilar to that used in Fernández-Villaverde and Rubio-Ramírez (2005) in estimating a dynamic stochastic general equilibrium model. This paper utilizes an alternative Bayesian estimation procedure based on recent developments in sequential Monte Carlo methods to perform filtering of the latent states and parameter estimation jointly. Chen et al. (2010) have applied this class of methods to the estimation of dynamic stochastic equilibrium models with success. Empirical results indicate that this is a viable estimation method for dynamic discrete choice models with a full solution DP solver embedded. It's possible to embed approximation-based solution methods instead for faster computation, section 3 contains more details on this subject.

SMC based methods, otherwise known as particle filters in a pure state filtering context, are a class of numerical methods designed to perform inference in general state space hidden Markov models. Unlike Kalman filters, SMC methods don't rely on local linearization or normal shock assumptions and require only that the model has a state space form. This flexibility is the main reason behind the ever growing popularity of SMC methods in a diverse array of fields ranging from financial econometrics to robotics. One of the primary problems SMC methods are designed for is a filtering problem where the goal is to obtain posterior inference on the latent state trajectories given some noisy observation. When embedded in an MCMC algorithm for parameter estimation, as is the case in GHK, the importance weights of the filtered particles are used to obtain an approximation of likelihood function. In this sense, particle filters are essentially used as a Monte Carlo integrator to marginalize out the unobserved state variables. Recent developments have shown that SMC methods are capable of estimating fixed parameters in addition to state filtering. Practical SMC-based methods were proposed by Liu and West (2001), Storvik (2002), Fearnhead (2002), Gilks and Berzuini (2001), and Carvalho et al. (2010) among others. SMC based estimation methods have several advantages over traditional MCMC based

Bayesian methods when applied to dynamic discrete game estimation. First, SMC methods don't rely on Markov chain convergence and instead work on the principle of importance sampling. For complex structural models such as dynamic games, Markov convergence might not be geometrically ergodic (see Papaspiliopoulos and Roberts (2008)). Indeed the MCMC chain generated using a one-move-at-a-time scheme in GHK converges slowly and model parameters exhibit high sample auto-correlation even at lag lengths of over 300. Second, the profile likelihood surfaces for a dynamic game model are littered with a combination of plateaus and local hills. In other words, it's hard for an MCMC algorithm to explore the posterior parameter space well. In this scenario, a well designed SMC-based procedure can visit parts of the posterior space that an MCMC algorithm is unlikely to have visited. Designing such an SMC algorithm is, however, a no less challenging task then coming up with an efficient MCMC algorithm for the same class of models. Lastly, by combining the filtering of the latent state trajectories and parameter learning in one SMC procedure, the computational effort involved in estimating the structural parameters are greatly reduced than that in an MCMC estimation routine. For the same parameter posterior sample size N inferred from the same data set of length T, the number of games that need to be solved in an SMC estimation procedure is of $\mathcal{O}(NT)$. The number of games that need to be solved in an MCMC estimation is of $\mathcal{O}(NMT)$, where M is the number of particles in the embedded particle filter for pure state filtering. As the estimation of the dynamic game of entry employs full DP solutions, which is computationally intensive to obtain, an SMC estimation can finish in a mere fraction of the time it takes an MCMC estimation to complete.

The dynamic game of entry from GHK is an example of the type of games SMC algorithms are capable of estimating. This entry model features cross-product spillovers on future costs in a dynamic oligopolistic setting, which is the first of its

kind[1]. In this model setting, firms may chose to enter a particular market even if it will lead to a loss in that market period provided the cost spillovers from the entry will lead to better long-run profitability in future markets. Heterogeneity of firms that participate in this entry game is induced by serially correlated firm specific costs that are endogenous to past entry decisions. These salient features make this entry game a challenging model to estimate and thereby a perfect test bed for an SMC based estimation procedure. The proposed algorithm could be applied to other dynamic discrete games of equal or lesser complexity as long as the model could be put into standard state-space form and a reliable solution method for the game could be implemented.

4.1 A Dynamic Model of Entry Decisions

4.1.1 Data

For the estimation of the dynamic entry model, I use the data set from Scott-Morton (1999). This data set contains the entry decisions of generic drug manufacturers between 1984 and 1994. During this period, firms can enter a market by submitting an Abbreviated New Drug Application (ANDA). In estimating the model, the only information needed from the data set is total market revenues and the entry decisions of potential firms at each market opportunity. As in GHK, I only use a sub-sample of the data covering the period between 1990 and 1994 for estimation due to the FDA bribery scandal in 1989. Only ANDAs for generic drugs in the form of orally ingested solids during this period are considered. This means the actual data used for estimation consists of 40 market opportunities and 51 potential entrants. The pre-scandal data, however, is used to recover parts of the cost structure as will be

[1] Static entry games have been studied among others by Bresnahan and Reiss (1990), Bresnahan and Reiss (1991), Berry (1992), Scott-Morton (1999), Mazzeo (2002), Seim (2006), Orhun (2006), Zhu and Singh (2006). Dynamic entry competition have been studies among others by Hitsch (2006), Ching (2009) and Shen (2010).

made clear later. The ANDA openings are ordered by calendar dates in the data set. The top four firms in descending order are Mylan, Novopharm, Lemmon and Geneva, who entered 45%, 27.5%, 25% and 25% of the markets, respectively. Mean market revenue in thousands of dollars is 126,901, with a standard deviation of 161,580, a minimum of 72, and a maximum of 614,593. See Table 4.1 and 4.2 for more details regarding this data set.

4.1.2 Model

There are I profit maximizing firms that operate over an infinite horizon $t = 1, \cdots, \infty$. The time index t denotes market openings, i.e., when a branded drug's patent expires. As such, the actual elapsed time between t and $t+1$ could be, and most likely will be, different from that between $t+4$ and $t+5$. GHK adopts this timing convention to mainly avoid the computational burden of identifying the exact entry dates from data. Interested readers can see GHK for detailed reasons of this timing convention. In the remainder, will be used to refer to market opening and calendar time interchangeably as market openings correspond to distinct calendar dates. Let A_{it} denote the entry decision of firm i at market opening t, then $A_{i,t} \in \{0, 1\}$, where $A_{i,t} = 1$ denotes decision to enter the market. Let $N_t = \sum_{i=1}^{I} A_{i,t}$ denote the total number of entering firms at time t. The firm specific costs are endogenous to past entry decisions and random shocks. Log cost for firm i at time t evolves as follows,

$$c_{it} = \mu_c + \rho_c(c_{i,t-1} - \mu_c) - \kappa_c A_{i,t-1} + e^c_{it} \tag{4.1}$$

where $e^c_{it} \sim N(0, \sigma^2_c)$. The long run mean of log cost is captured by μ_c, ρ_c captures the persistence level of the cost dynamics. The spillover effect is the increase in profitability due to experience gained from market entries in the past. In this model, this effect is captured by $0 < \kappa_c < 1$, in the form of a reduction of firm specific cost due to entry in the last market opening. Note that those parameters are not firm specific, in other words, heterogeneity of firms arises endogenously from past entry decisions

Table 4.1: Shown here is the post-scandal data used in the study. Entry decisions are indicated by 1 for entry and 0 for no-entry. Total entrants show how many of the fifty-one potential entrants entered, including the four dominant firms. Revenue of the branded drug is that in the year before patent expiration. .

	Dominant Firms (enter = 1, not enter = 0)					
Drug	Mylan	Novopharm	Lemmon	Geneva	Total Entrants	Revenue ($'000s)
SULIN.	1	0	1	1	7	189010
ERYTH.	0	0	0	0	1	13997
ATENO.	1	0	0	0	4	69802
NIFED.	0	1	0	0	5	302983
MINOC.	0	0	0	0	3	55491
METHO.	1	0	0	0	3	24848
PYRID.	0	0	0	0	1	2113
ESTRO.	0	0	0	0	2	6820
LOPER.	1	1	1	1	5	31713
PHEND.	0	0	0	0	1	1269
TOLME.	1	1	1	1	7	59108
CLEMA.	0	0	1	0	1	9077
CINOX.	0	0	0	0	1	6281
DILTI.	1	1	0	0	5	439125
NORTR.	1	0	0	1	3	187683
TRIAM.	0	0	0	1	2	22092
PIROX.	1	1	1	0	9	309756
GRISE.	0	0	0	0	1	11727
PYRAZ.	0	0	0	0	1	306
DIFLU.	0	0	1	0	2	96488
CARBI.	0	0	1	0	4	117233
PINDO.	1	1	0	1	7	37648
KETOP.	0	0	0	0	2	107047
GEMFI.	1	0	1	0	5	330539
BENZO.	0	0	0	0	1	2597
METHA.	0	0	0	0	1	1858
METHA.	0	0	0	1	3	4792
ALPRA.	1	1	0	0	7	614593

Table 4.2: Continue from Table 4.1

Drug	Dominant Firms (enter = 1, not enter = 0)				Total Entrants	Revenue ($'000s)
	Mylan	Novopharm	Lemmon	Geneva		
NADOL.	1	0	0	0	2	125379
LEVON.	0	0	0	0	1	47836
METOP.	1	1	0	1	9	235625
NAPRO.	1	1	1	1	8	456191
NAPRO.	1	1	1	1	7	164771
GUANA.	0	0	0	0	2	18120
TRIAZ.	0	0	0	0	2	71282
GLIPI.	1	0	0	0	1	189717
CIMET.	1	1	0	0	3	547218
FLURB.	1	0	0	0	1	155329
SULFA.	0	0	0	0	1	72
HYDRO.	0	0	0	0	1	8492
Mean	0.45	0.28	0.25	0.25	3.3	126901

and the stochastic dynamics of costs. As this is a game of complete information, all firms observe each others' costs. From the econometrician's perspective, cost can be decomposed into an observable component and an unobservable centered AR(1) process:

$$c_{i,t} = c_{u,i,t} + c_{k,i,t} \tag{4.2}$$

$$c_{u,i,t} = \mu_c + \rho_c(c_{u,i,t-1} - \mu_c) + e^c_{it} \tag{4.3}$$

$$c_{k,i,t} = \rho_c c_{k,i,t-1} - \kappa_c A_{i,t-1}. \tag{4.4}$$

Iterating (4.4), we can see that the cumulative cost impact of past entries at market opening t is $-\sum_{j=0}^{\infty} \rho_c^j \kappa_c A_{i,t-j-1}$.

The total log revenue amongst firms that enter at time t is r_t, and follows the following dynamics,

$$r_t = \mu_r + e^r_t, \tag{4.5}$$

where $e_t^r \sim N(0, \sigma_r^2)$. Therefore, total log revenue is normally distributed around the average total log revenue across firms and market openings. Strategic interactions between firms are modeled for the top three[2] dominant firms as equilibria computation for more firms would be prohibitively expensive with a full solution DP algorithm. Since the top three firms account for more than half of all market entries within the data period considered, a three firm model can still capture significant market dynamics within the generic drug industry. To account for the non-dominant firms, the total revenue available to the dominant firms is R_t^γ instead of R_t, where $r_t = log(R_t)$. This convention of using lower case to denote log-scale variables and upper case to denote real-scale variables also applies to the cost variables. The parameter γ controls for the market share of non-dominant firms and related literature suggests a reasonable range of between 0.908 and 1. Thus, dominant firm i's per period profit at time t is given by

$$A_{it}(R_t^\gamma / N_t - C_{it}). \tag{4.6}$$

Dominant firm i's objective as it participates in the entry game is to maximize the discounted total profit over the infinite horizon, which at time t is given by

$$\sum_{j=0}^{\infty} \beta^j A_{i,t+j}(R_{t+j}^\gamma / N_{t+j} - C_{i,t+j}). \tag{4.7}$$

Firm i would thus need to solve a dynamic program[3] with the following Bellman equation on its choice specific value function,

$$V_i(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t) = A_{it}\left(R_t^\gamma / N_t - C_{it}\right) + \tag{4.8}$$

$$\beta \mathcal{E}\left[V_i(A_{i,t+1}^E, A_{-i,t+1}^E, C_{i,t+1}, C_{-i,t+1}, R_{t+1}) | A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t\right] \tag{4.9}$$

where $-i$ denotes firms other than firm i.

[2] GHK estimates the model with both three and four dominant firms, and obtained very similar parameter estimates.

[3] See Rust (2006) for a discussion on how to use dynamic programming to solve dynamic games. A more detailed treatment of the subject is given in Ljungqvist and Sargent (2000).

The choice specific value function $V_i(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t)$ represents firm i's total discounted profit from making an entry decision A_{it} at time t, conditional on entry decisions made by other firms at time t and the expectation that all firms would be making subsequent entry decisions in equilibrium based on decisions made at t. The expectation is over the distribution of the state variables $(C_{i,t+1}, C_{-i,t+1}, R_{t+1})$ conditional on the realizations of state variables and actions profiles at time t. A stationary pure strategy Markov perfect equilibrium of the dynamic entry game is a best response strategy profile $(A^E_{i,t}, A^E_{-i,t})$ that satisfies the following optimal condition

$$V_i(A^E_{i,t}, A^E_{-i,t}, C_{i,t}, C_{-i,t}, R_t) \geq V_i(A_{i,t}, A^E_{-i,t}, C_{i,t}, C_{-i,t}, R_t),\ \forall i, t. \tag{4.10}$$

In words, no other action can give rise to a strictly higher cumulative profit for any firm other than that firm's equilibrium strategy, given that all competing firms play at equilibrium as well.

As all state variables are bounded and the action sets of all players finite, Theorem 5.1 of Dutta and Sundaram (1998) guarantees the existence of equilibria for this entry game. The equilibrium profile at market t can be solved given realizations of the state variables $(C_{i,t}, C_{-i,t}, R_t)$. In solving the game, we would make use of the ex ante value function, which is defined as the choice specific value function from (4.8) fixed at the equilibrium profile $A^E_{i,t}, A^E_{-i,t})$. The ex ante value function $V_i(C_{i,t}, C_{-i,t}, R_t)$ satisfies the following Bellman equation,

$$V_i(C_{i,t}, C_{-i,t}, R_t) = A^E_{i,t}(R^\gamma_t / N^E_t - C_{i,t}) + \tag{4.11}$$

$$\beta \mathcal{E}\left[V_i(C_{i,t+1}, C_{-i,t+1}, R_{t+1}) | A^E_{i,t}, A^E_{-i,t}, C_{i,t}, C_{-i,t}, R_t\right],\ \forall i, t, \tag{4.12}$$

where N^E_t is the total number of entering firms at equilibrium. The ex ante value function represents firm i's total discounted profit from making an optimal entry decision at time t, conditional on optimal entry decisions of other firms at time t and the expectation that all firms would be making subsequent choices in equilibrium.

4.1.3 State-Space Formulation

To apply SMC based methods to the estimation of the dynamic entry game, the game model must be put into state-space form. Using notations from chapter 1, we have the latent state vector $x_t = (c_{u,1,t}, \cdots, c_{u,I,t})$, and observables $y_t = (A_{1,t}, \cdots, A_{I,t}, c_{k,1,t}, \cdots, c_{k,I,t}, r_t)$. Recall that the data set used for estimation only contains the entry decisions $A_t = (A_{1,t}, \cdots, A_{I,t})$ and log market revenue r_t. Even though the observable component of the log cost $(c_{k,1,t}, \cdots, c_{k,I,t})$ is not directly observed, we can infer their values from the observed entry choices using (4.4) as it's a deterministic function of the observable cost and entry choice from the last time period. The pre-scandal data $\{A_{i,t}\}_{t=-n}^{0}$ is used in exactly this way to generate $(c_{k,1,0}, \cdots, c_{k,I,0})$ for future deterministic evolution during estimation[4]. The observed firm actions are deterministic conditional on $(x_t, r_t, y_{t-1}, \theta)$, this means the conditional likelihood $p(A_t|r_t, y_{t-1}, x_t, \theta)$ is degenerate at either 0 or 1. GHK assumes measurement error on A_t which is standard practice for likelihood based inference with discrete observations. The observation equation of the dynamic game is thus a function of A_t and r_t and has the following decompositional form,

$$p(y_t|x_t, y_{t-1}, \theta) = p(A_t|r_t, y_{t-1}, x_t, \theta)p(r_t|y_{t-1}, x_t, \theta), \tag{4.13}$$

where $p(r_t|y_{t-1}, x_t, \theta)$ is a normal density with mean μ_r and standard deviation σ_r from equation (4.5), and

$$p(A_t|r_t, y_{t-1}, x_t, \theta) = \prod_{i=1}^{I} (p_a)^{I(A_{it}=A_{it}^E)}(1-p_a)^{I(A_{it}\neq A_{it}^E)}, \tag{4.14}$$

where $1 - p_a$ is the misclassification probability. One way to rationalize this measurement error is by making the assumption that there is a small probability a firm's entry decision will not be carried out. The state equations are already given in equations (4.2)-(4.4), with the initial state distribution $p(x_0|\theta)$ being normal

[4] In GHK, (4.4) is iterated over the pre-scandal data with set to zero.

around mean μ_c and standard deviation $\sigma_c/\sqrt{1-\rho_c^2}$. The model parameters are $\theta = (\mu_c, \rho_c, \sigma_c, \kappa_c, \mu_r, \sigma_r, \gamma, \beta, p_a)$.

The observation equation (4.13) is nonlinear, non-normal and embeds game theoretic interactions amongst the participating firms in the form of the Markov perfect equilibrium solution A_{it}^E. Given $(x_t, r_t, y_{t-1}, \theta)$, A_{it}^E could be obtained by solving the Bellman equation (4.11), the details of which will be provided shortly. Of the model parameters, (γ, β) doesn't appear in the state-space form but affects the system through the game solution method. Similarly in equation (4.14), the conditioning on the latent state x_t is not explicit but through the game solver. The game theoretic component of the model essentially adds an extra layer of dependence structure between the observed entry decisions and the latent firm costs. The resulting likelihood surface mostly resembles a step function, as different values of the latent states and parameters can lead to the same game equilibria, and is the root cause for the slow convergence of MCMC algorithms.

4.2 An SMC Approach to Parameter Estimation

Sequential Monte Carlo methods are flexible simulation methods designed to produce samples from a sequence of target densities of increasing dimension, which are better known as particle filters when applied to the filtering problem in hidden Markov models. These methods make it possible to solve sequential inference problems without making normality and/or linearity assumptions, thereby allowing researchers to build more realistic models to capture the underlying dynamics of complex data sets. There has been a rich development on SMC related literature in statistics and engineering and it would be outside the scope of this paper to provide a detailed review on the subject. I will start this section by first presenting a pure state filtering algorithm for the entry game based on the auxiliary particle filter (APF), and then incorporate parameter learning on top of that filtering algorithm for efficient SMC

estimation. Smoothing algorithms are also presented, which allows researchers to make inferences on the hidden state trajectories conditional on all available information in the data. All presented algorithms are adaptable to other game models given the model can be put into state-space form and a solution method for finding game equilibria is provided.

4.2.1 State Filtering with Fixed Parameters

It's clear that the optimal importance function (2.10) is not feasible for the entry game for two reasons. First, the predictive function $p(y_{t+1}|x_t)$ in $q_{t+1}^{opt}(x_{t+1}|x_{1:t})$ is not available in close-form for the entry game. Second, approximating $p(y_{t+1}|x_t) = \int p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)dx_{t+1}$ for the dynamic entry game using a Monte Carlo integral would involve repeated evaluations of (4.14), which in turn means more DPs to solve. Implementing an approximate $q_{t+1}^{opt}(x_{t+1}|x_{1:t})$ for the dynamic game would thus result in an insurmountable computational task. Even though use of the optimal importance function is out of the question for the entry game, we can still make use of the information when making inference at time $t+1$ by avoiding the bootstrap filter.

The auxiliary filter from Pitt and Shephard (1999) adopted to the filtering of the unobserved cost paths for the dynamic entry game is given as follows,

1. For $i = 1, \cdots, N$, sample $x_0^{(i)} \sim p(x_0)$ and set $t = 0$.

2. For $i = 1, \cdots, N$, compute look-ahead particles by $\mu_{t+1}^{(i)} = E(x_{t+1}|x_t^{(i)})$.

3. Solve the game at market opening $t+1$ with the predicted latent costs $\mu_{t+1}^{(i)}$ and resample $\left\{x_t^{(i)}\right\}_{i=1}^{N}$ with the now available predictive weights $g_{t+1}^{(i)} \propto w_t^{(i)} p(y_{t+1}|\mu_{t+1}^{(i)})$, the resampled particles will be indexed by k.

4. For $i = 1, \cdots, N$, sample new particles for the latent state $x_{t+1}^{(k)} \sim p(x_{t+1}|x_t^{(k)})$.

5. Solve the game at market opening $t+1$ with the newly updated latent costs $x_{t+1}^{(k)}$. Compute the posterior weights $w_{t+1}^{(k)} \propto p(y_{t+1}|x_{t+1}^{(k)})/p(y_{t+1}|\mu_{t+1}^{(k)})$, $\left\{x_{t+1}^{(k)}\right\}_{k=1}^{N}$ weighted by $\left\{w_{t+1}^{(k)}\right\}_{k=1}^{N}$ is an approximate sample from $p(x_{t+1}|y_{1:t+1})$.

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

The above filtering algorithm should provide better performance when embedded in an MCMC algorithm in place of the bootstrap filter. In the context of SMC methods, this filtering algorithm can constitute the forward filtering component of the often used forward-filtering-backward-sampling scheme in generating realizations from the full smoothing distribution of the latent states.

4.2.2 Joint State Filtering and Parameter Learning

Following the developments in earlier chapters, the LW algorithm can be adapted to the entry game as follows

1. For $i = 1, \cdots, N$, sample $x_0 \sim p(x_0)$, $\theta \sim p(\theta_0)$ and set $t = 0$.

2. For $i = 1, \cdots, N$, compute look-ahead particles for the state variables by $\mu_{t+1}^{(i)} = E(x_{t+1}|x_t^{(i)}, \theta^{(i)})$, and new kernel locations by $m_t^{(i)} = a\theta_t^{(i)} + (1-a)\bar{\theta}_t$.

3. Solve the game at market opening $t+1$ with the prediction particles $(\mu_{t+1}^{(i)}, m_t^{(i)})$ and resample $\left\{x_t^{(i)}, m_t^{(i)}\right\}_{i=1}^{N}$ with the now available predictive weights $g_{t+1}^{(i)} \propto w_t^{(i)} p(y_{t+1}|\mu_{t+1}^{(i)}, m_t^{(i)})$, the resampled particles will be indexed by k.

4. For $i = 1, \cdots, N$, sample new parameter particles $\theta_{t+1}^{(k)} \sim N(\cdot|m_t^{(k)}, (1-a^2)V_t)$, sample new particles for the latent state $x_{t+1}^{(k)} \sim p(x_{t+1}|x_t^{(k)})$.

5. Solve the game at market opening $t+1$ with the updated states and parameters $(x_{t+1}^{(k)}, \theta_{t+1}^{(k)})$. The posterior weights are given by $w_{t+1}^{(k)} \propto p(y_{t+1}|x_{t+1}^{(k)}, \theta_{t+1}^{(k)})/p(y_{t+1}|\mu_{t+1}^{(k)}, m_t^{(k)})$.

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

The proposed SMC method performs sequential state filtering and parameter learning in one single pass through the data, and provides posterior samples at each market opening that approximate the marginal posterior $p(x_{t+1}, \theta|y_{1:t+1})$. The method is generally applicable to the estimation of other games as parameters don't need to have close-form posterior densities.

4.2.3 Particle Smoothing

The earlier SMC filtering methods provide consistent posterior inference to the marginal filtering distributions $p(x_{t+1}|y_{1:t+1})$ and $p(x_{t+1}, \theta|y_{1:t+1})$ but will fail to produce reliable inference to joint filtering distributions $p(x_{1:t+1}|y_{1:t+1})$. While this is of no concern in the context of parameter estimation, it's sometimes of interest to study the latent state paths jointly after receiving all available information. Obtaining samples from $p(x_{1:T}|y_{1:T})$ is referred to as a smoothing problem in the SMC literature and is closely related to the earlier filtering problem. Smoothing algorithms usually employ filtering results in a forward-backward recursion where filtered results are sampled in reverse order to accommodate additional information obtained in the forward filtering process. As we already have filtering algorithms designed for the dynamic entry game, no special smoothing algorithm is needed. Once filtering results are available, the smoothing process doesn't require any additional games to be solved, thus providing an efficient way to make joint inference on $x_{1:T}$ conditional on $y_{1:T}$. The following is a smoothing algorithm that incorporates parameter posterior uncertainty from Carvalho et al. (2010).

1. Sample $(\tilde{\theta}, \tilde{x}_T) \sim p(x_{1:T}, \theta | y_{1:T})$ and set $t = T$.

2. For $i = 1, \cdots, N$, compute smoothing weights $w_{t|t+1}^{(i)} \propto p(\tilde{x}_{t+1} | x_t^{(i)}, \tilde{\theta})$.

3. Choose $\tilde{x}_t = x_t^{(i)}$ with probabilities proportional to $w_{t|t+1}^{(i)}$.

4. If $t \neq 1$, set $t = t - 1$ and go to step 2.

By fixing θ and using pure filtering results, this smoothing algorithm reduces to the pure state smoothing algorithm in Godsill et al. (2004). Running the above smoothing algorithm once gives one approximate realization from the full smoothing distribution $p(x_{1:T}, \theta | y_{1:T})$. A representative approximation of the full smoothing density can be obtained by repeating the process enough number of times.

4.2.4 Computational Issues

The above SMC resample-sample method offers substantial savings in computational cost when compared to MCMC algorithms. MCMC estimation of dynamic structural models involves getting parameter proposals from an outer loop (e.g., a Metropolis-Hastings proposal density) and then filtering through the latent state variables with the model parameter fixed at the proposal. As mentioned in section 2, the value iteration DP solver used to find game equilibria is a computational bottleneck in any Bayesian simulation-based estimation. The slow mixing of MCMC chains for this class of models coupled with the embedded state filter translates to the computation of a huge number of game solutions. Assume that the simplest bootstrap filter is used for state filtering, then for a data set of length T and a total of M particles used in the state filter, obtaining N MCMC posterior samples requires solving $\mathcal{O}(NMT)$ games. For each parameter proposal, the states have to be re-filtered through the data set, resulting in the computation of $\mathcal{O}(MT)$ game solutions. In comparison, an SMC method such as the one proposed in this paper requires only $\mathcal{O}(NT)$ games to

be solved to obtain N SMC posterior samples. It's clear to see that when $M = N$, the number of games to be solved grows quadratically in the number of posterior samples for MCMC estimation, and only linearly for SMC estimation. In practice, MCMC methods need a much larger N than SMC methods to adequately explore the posterior parameter space of dynamic discrete choice models due to the particularly slow Markov chain convergence for this class of models[5] , and consequently end up solving a lot more than NMT games where N is the actual number of MCMC posterior samples used for inference.

4.3 Results

Using the above SMC method, I estimate the dynamic entry game with three dominant firms using both boundedly rational and fully rational game solutions (see appendix B for differences between the two). I do not present estimation results for the four dominant firm case as i) the computational cost of solving repeated games with four players is prohibitively high (It takes almost one month to estimate a four firm entry model with the same data set using the proposed SMC method with 4000 particles), ii) the three dominant firm model captures almost as much market dynamics as the four firm model does[6], and iii) MCMC estimation of the four firm model from GHK did not produce posterior modes that vary greatly from the three firm model. In GHK, the nonlinear parameters (γ, β, p_a) are fixed for more efficient MCMC performance. SMC results that follow will include both the partial model uncertainty case (with γ, β and p_a fixed) and the full model uncertainty case (with those parameter estimated). As particle approximations of parameter posteriors can

[5] In GHK's MCMC estimation of the dynamic entry game, 3 million MCMC steps are performed with a one-step-at-a-time random-walk Metroplis Hastings kernel, 1 out of every 375 MCMC samples are used in actual posterior computation due to high sample autocorrelation.

[6] The top three firms account for 55% of all market entries in the dataset used, whereas the top four firms account for 60% of all market entries.

be multimodal, the posterior mode might not always be a characteristic measure for central tendency. I thus present both posterior modes and medians for comparison purposes.

One measure of estimation quality for dynamic discrete choice models is the classification error rate (CER), which is essentially an in-sample prediction error measure. At any time t during the sequential estimation procedure, one can average the computed firm equilibrium action profiles conditional on the newly sampled x_t, θ) particles at step 5 of the proposed resample-sample algorithm. With a predetermined predictive threshold, one can use the proportion of missed predictions as the CER. All subsequent CER results are presented with a predictive threshold of 0.5 (e.g. an observed entry with a corresponding particle equilibrium prediction of under 0.5 would be considered a missed prediction).

4.3.1 Prior Distribution $p(\theta)$

Table 4.3 describes the prior distributions used in the estimation, which are all uniform distributions with different supports. With the exception of parameter ρ, the prior supports for all other parameters are either chosen to be non-informative or meant to capture economically realistic and meaningful scenarios. For the unconditional mean parameters μ_c and μ_r, the proposed prior range more than covers all possible realistic values of log revenue and log cost. For the standard deviation parameters σ_c and σ_r, the given uniform support could be seen as appropriate a posteriori as the estimated posteriors are well inside the prior support. The support for κ_c is determined by model definition to be between 0 and 1. While κ_c could be identified on this prior support, a smaller interval from 0 to 0.2 can facilitate faster SMC sample convergence. Support for parameters γ, β, p_a) are selected to capture all possible values within an economically meaningful range, see GHK for the economic intuitions.

Table 4.3: Prior distribution $p(\theta)$ for the entry game.

θ	$p(\theta)$
μ_c	Uniform(0,50)
ρ_c	Uniform(0.75,1)
σ_c	Uniform(0,1)
κ_c	Uniform(0,0.2)
μ_r	Uniform(0,50)
σ_r	Uniform(0,5)
γ	Uniform(0.9,1)
β	Uniform(0.75,1)
p_a	Uniform(0.75,1)

The cost persistency parameter ρ_c can theoretically take any value between -1 and 1 to make the centered AR(1) cost process stationary. Applying the above SMC estimation with initial particles of ρ_c generated uniformly from (-1,1) yields unstable estimates of the parameter. One source of the weak identifiability is the fact that multiple parameter values can lead to the same game equilibrium solution. In MCMC algorithms, the resulting plateaus on the likelihood surface create difficulties in the adequate exploration of the parameter space. For SMC estimation, the evaluated observation equations in the form of (4.13) are used as important weights to resample the parameter particles. The non-unique mapping from particles $(x_t, \theta)^{(i)}$ to equilibrium particles $(A^E)^{(i)}$ means parameter particles that best rationalize the data will be under represented as parameter values around those particles will have similar importance weights. While parameter ρ_c appears to suffer the most from this source of weak identifiability, its effect on the other model parameters is less pronounced. A more specific cause for the low identifiability for ρ_c is the large estimates of σ_c. The marginal posterior distribution of σ_c from repeated SMC estimation in the partial model uncertainty case all center around values between 0.4 and 0.5. As the state equations are all in log scale, such a large σ_c can generate cost shocks big enough to make up for possible low cost persistence. To impose identifiability, I gen-

Table 4.4: Effect of ρ priors on CER for the three firm prediction error model (fixed γ, β and p_a) based on 1000 particle estimation.

prior on ρ	MYLAN	NOVOPHARM	LEMMON	all firms
Uniform (-1,1)	0.175	0.275	0.25	0.2333
Uniform (-0.5,1)	0.1	0.225	0.25	0.1916
Uniform (0,1)	0.125	0.225	0.2	0.1833
Uniform (0.75,1)	0.1	0.05	0.075	0.075

erate initial particles of ρ_c from uniform distributions defined on different supports and compare the resulting effect on classification errors, the results are summarized in Table 4.4. As consistent with MCMC results from GHK, a high cost persistency region of (0.75,1) give rise to the lowest CER. Thus I use the informed uniform interval of (0.75,1) as the prior for ρ_c. Note that this only applies to the initial particles of ρ_c and that subsequent posteriors of ρ_c still has the stationary support of (-1,1). The same applies to κ_c if a subset of the model-defined support (0,1) is used as prior support.

4.3.2 Boundedly Rational vs Fully Rational

The fully rational equilibrium concept should in theory provide better estimation results in terms of lower CERs as firms explicitly take into account the misclassification error in the observation equation (4.14) when making equilibrium predictions about entry. This means the equilibrium action profiles returned by a fully rational solver will have a higher chance of matching the observed actions than the boundedly rational solver. MCMC results from GHK does show slightly lower CERs from estimation with the full rationality model, although parameter estimates are nearly indistinguishable across the two solution concepts. To get a sense of how the solution concept will affect SMC estimation of the dynamic game, models with both solution concepts are estimated with 1000 particles. As seen from Table 4, estimation of the boundedly rational model yields substantially higher CERs than that of the fully

Table 4.5: Effect of game solution concept on SMC estimation with (γ, β, p_a) fixed at (0.9375,0.96875,0.9375). Presented here are posterior modes, medians and CERs under the two solution concepts.

θ	boundedly rational		fully rational	
	post. mode	post. mean	post. mode	post. mean
μ_c	9.5235	9.5371	10.5481	10.5895
ρ_c	0.5583	0.5570	0.5186	0.5571
σ_c	0.2828	0.2858	0.5892	0.4303
κ_c	0.0427	0.0478	0.0190	0.0234
μ_r	9.7290	9.7629	9.7498	9.8441
σ_r	1.7766	1.7829	1.6697	2.0724
CER Mylan	0.125		0.125	
CER Novopharm	0.225		0.075	
CER Lemmon	0.2		0.05	
CER all	0.1833		0.0833	

rational model and this remains to be the case when the number of particles are increased. Most parameter estimates are similar, the discrepancies between estimates for the cost shock standard deviation and the spill-over cost reduction could be due to the low number of particles used. In conclusion, SMC estimation appears to be more sensitive to the equilibrium concept used in solving the game than MCMC estimation. As SMC methods don't rely on the convergence of a single Markov chain, this increased sensitivity is somewhat expected. In a slow mixing Markov chain of parameter samples, any effect the game solution concept might have on the posterior distributions is essentially diluted by sample autocorrelation. Since imposing full rationality will noticeably increase the model's in-sample predictive power, all subsequent results are presented on the fully rational model.

4.3.3 Partial vs Full Model Uncertainty

Table 4.6 shows the estimation summary of the dynamic entry game with parameters (γ, β, p_a) fixed at the given values in the table. All posterior information presented are from the marginal posterior distribution at the last observation (i.e. $p(\theta, x_{40}|y_{1:40})$)

in the data set, which can be seen as equivalent to the marginal posterior from MCMC batch estimation of the same data set. In the partial model uncertainty case, the standard deviations for both the cost shock and the revenue shock are estimated to be around 30% higher than that of MCMC results. The immediate spill-over effect is estimated to be around 10%, which is slightly higher than the 7% from GHK. As mentioned earlier, the large cost shock standard deviation is a direct cause for the low identifiability of cost persistence ρ_c. This is demonstrated in the posterior trace plot of ρ_c in Figure 4.1. Those trace plots essentially plot out the 95% posterior quantiles at each observation in the data since the above SMC method generates posterior samples sequentially through observations. The rate at which those quantiles converge can be used as a diagnostic of how well identified the underlying parameter is. For well identified parameters such as μ_c and σ_r, their posterior quantiles quickly converge as the SMC algorithm picks up and zeroes in on regions of high model likelihood. For ρ_c however, the final 95% quantile is only slightly smaller than the initial prior support due to weak identifiability for reasons discussed earlier.

Estimation results of the model with full parameter uncertainty are summarized in Table 4.7. The revenue shock standard deviation is estimated to be around 45% higher than that of MCMC estimation while the cost shock standard deviation coincides with MCMC results. Posterior distribution for spill-over cost reduction is clearly multimodal under SMC estimation, with a lower mode around the MCMC mode from GHK for an immediate $5-6\%$ cost reduction from entry in the last market opening, and a higher mode around $14-15\%$ which MCMC estimation fails to pick up after being trapped in lower mode. Notice that such a high mode for the spill-over effect is countered by the relatively low estimate of the cost persistence ρ_c at 0.73, which is noticeably lower than the MCMC mode of 0.98 in GHK. With an average of eight market openings annually, and taking into account full model un-

Table 4.6: Three firm SMC estimation summary with partial model uncertainty, parameters (γ, β, p_a) are fixed at (0.9375,0.96875,0.9375) and SMC results are from the marginal posterior at the end of the data set $(t = 40)$.

θ	MCMC mode	SMC mode	SMC median	SMC 95% CI
μ_c	10.05	7.2927	9.227	(6.4012 13.1523)
ρ_c	0.9866	0.8575	0.8466	(0.7485 0.9185)
σ_c	0.3721	0.5051	0.4895	(0.3827 0.6445)
κ_c	0.06655	0.1109	0.0932	(0.0632 0.1405)
μ_r	9.906	10.3077	10.3451	(10.1068 10.6106)
σ_r	1.591	1.9378	2.0555	(1.8065 2.3421)
CER Mylan	0.09	0.075		
CER Novopharm	0.08	0.05		
CER Lemmon	0.1	0		
CER all	0.09	0.0417		
SMC particles		3000		
MCMC samples	8000			

certainty, the annual cumulative cost reduction for a dominant firm that entered all eight openings is 12% at the lower κ_c mode, 37% at the higher κ_c mode and 24% at the posterior median. For comparison purposes with the partial model uncertainty case, SMC estimation results gave an average cumulative cost reduction of around 41%, and MCMC results from GHK implies a 51% yearly average cost reduction. The trace plot in Figure 4.2 again demonstrates the weak identifiability of cost persistence in the full model uncertainty scenario. In contrast to ρ_c, the trace plots for all other parameters show evidence of successful parameter learning from the SMC algorithm, albeit at a slightly slower pace than the partial model uncertainty case.

All three nonlinear parameters (γ, β, p_a) are estimated to be lower than the fixed values used in GHK. SMC estimate of γ is slightly lower than the fixed calibration of 0.9375 in the partial model uncertainty scenario, which implies that non-dominant firms have a larger share of market profit (i.e. $R^{1-\gamma}$) when full model uncertainty is taken into account. The discount factor β is estimated to be 0.784, much lower

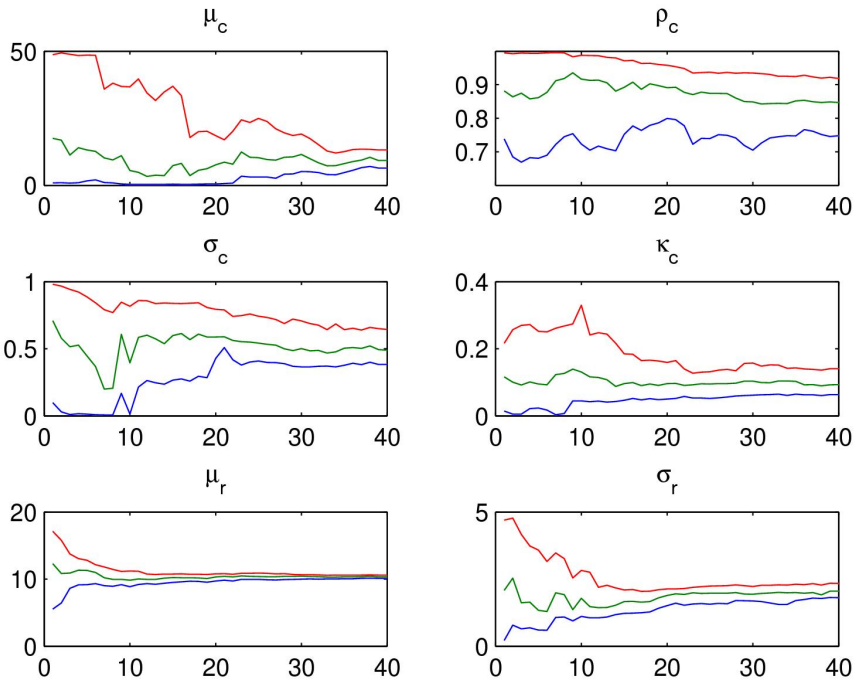

FIGURE 4.1: Trace plot from sequential estimation of three firm fully rational model (fixed γ, β and p_a). The three lines trace out the 95% posterior confidence interval at each point in the data set.

than the calibrated value of 0.96875. A discount factor this low would lead to an unrealistic annual internal rate of over 200%. Discount factors in dynamic models of the type considered are typically hard to estimate as other parameters in the Bellman equation (4.8) can counteract with it. For example, a low β can be countered by a high μ_r or a low μ_c. It would be perhaps wise to leave β fixed at a value consistent with empirical findings in similar models.

Regardless of whether the model is being estimated under partial uncertainty or full uncertainty, it appears that the mean market revenue and the revenue shock standard deviation are consistently over estimated when compared to the MCMC modes from GHK. To test whether the MCMC posterior modes for those two parameters are true local modes, I re-estimate the three firm model with partial uncertainty

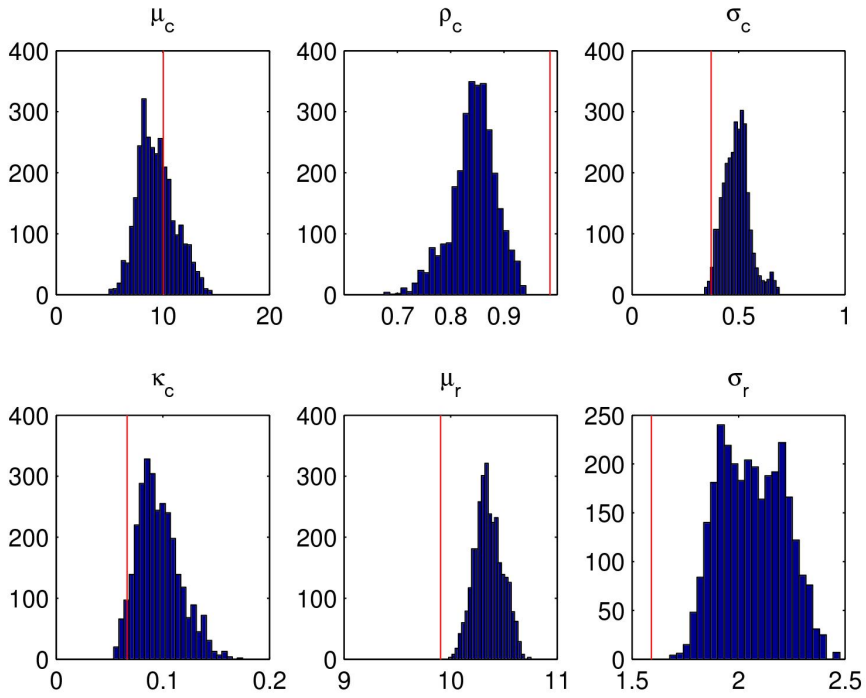

FIGURE 4.2: Marginal posteriors at last market opening from sequential estimation of three firm fully rational model (fixed γ, β and p_a). The red vertical lines mark the posterior modes from GHK's MCMC estimation.

with informative priors on all parameters. The informative priors used are normal densities centered around GHK's posterior modes with variances of 10 MCMC standard deviations, these priors represent a strong belief that the true parameter values are within close vicinity of the MCMC modes. If these MCMC modes are indeed likelihood maximizing locally (say within 10 posterior std. dev.), then the resulting SMC modes with informative priors should coincide with them. Figure 4.7 suggests that while the MCMC modes for $(\mu_c, \rho_c, \sigma_c, \kappa_c)$ are likely local posterior modes, the true local posterior modes for μ_r and σ_r are larger than those reported in GHK. Specifically, SMC mode for μ_r is around 10.4, which is 6 posterior std. dev. away from the MCMC mode, and around 1.75 for σ_r, which is 2.6 posterior std. dev. away from the MCMC mode. Despite the seemingly small differences compared to

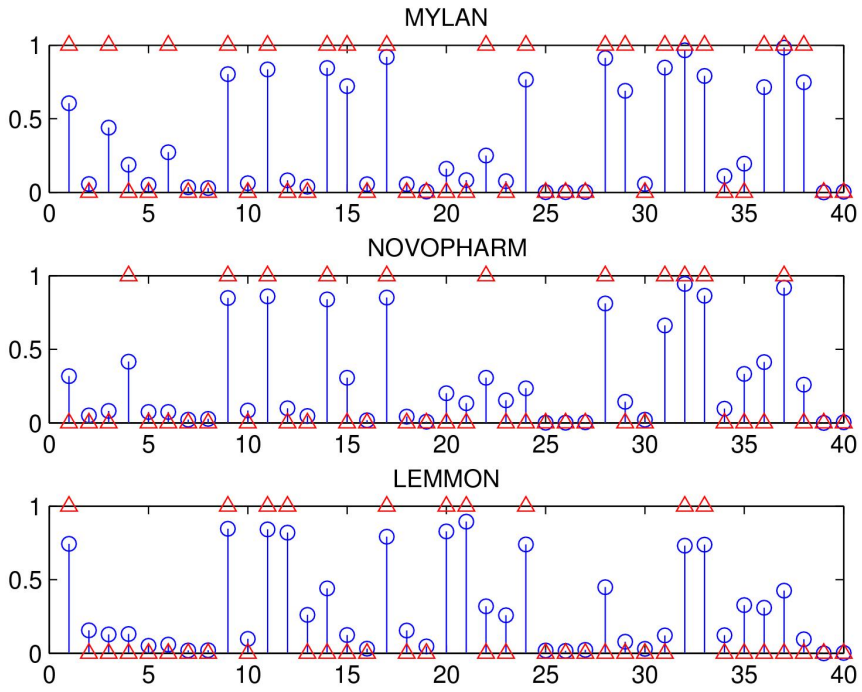

FIGURE 4.3: Particle prediction of firm actions from sequential estimation of three firm fully rational model (fixed γ, β and p_a). The triangles denote the observed entry decisions of the firms, the stems denote the computed particle average of equilibrium action profiles A_{it}^E conditional on $(r_t, x_t, y_{t-1}, \theta)$.

reported MCMC modes, these estimates translate to significant discrepancies in real dollar amounts. Take the mean market revenue for example, an estimate of 10.4 for log revenue is 65% larger than the MCMC mode of 9.9 in real dollars.

Figures 3 and 6 show that the particle average predictions of equilibrium firm actions match accurately with reality during the sequential estimation process. The overall classification errors are very low with a firm average CER of 0.04 for partial model uncertainty and 0.08 for full model uncertainty. C++ implementation of the proposed SMC procedure can estimate the three firm model with 4000 particles in roughly 5-6 days, which is a mere fraction of the time it takes to get the MCMC results in earlier tables.

Table 4.7: Three firm SMC estimation summary with full model uncertainty, results are from the marginal posterior at the end of the data set ($t = 40$). MCMC results for comparison is the same as that in Table 4.6, which has (γ, β, p_a) fixed at (0.9375,0.96875,0.9375).

θ	MCMC mode	SMC mode	SMC median	SMC 95% CI
μ_c	10.05	10.2659	10.1534	(9.4667 10.9154)
ρ_c	0.9866	0.7217	0.7333	(0.6428 0.8193)
σ_c	0.3721	0.4085	0.3642	(0.2689 0.5186)
κ_c	0.06655	0.0450	0.0949	(0.0378 0.2132)
μ_r	9.906	10.4627	10.5921	(10.2250 10.9653)
σ_r	1.591	2.3009	2.3285	(2.1016 2.6678)
γ	0.9375*	0.9058	0.9073	(0.9040 0.9197)
β	0.96875*	0.7847	0.7844	(0.7692 0.8213)
p_a	0.9375*	0.8865	0.8879	(0.8302 0.9124)
CER Mylan	0.09	0.1275		
CER Novopharm	0.08	0.075		
CER Lemmon	0.1	0.05		
CER all	0.09	0.0833		
SMC particles		4000		
MCMC samples	8000			

4.3.4 Extracting Latent Cost Paths

A benefit of SMC methods is that it provides a way to generate realizations from the full smoothing distribution, which in the context of the dynamic game, allows a researcher to extract the latent cost paths $\{c_{i,u,t}\}_{t=1}^{T}$ conditional on all the information contained in the data. The extracted cost paths can reveal important information regarding the firms entry decisions and are crucial for understanding the market share breakdown amongst the dominant firms in the industry. As mentioned earlier, smoothing can either be performed with parameter posterior uncertainty taken into account, or without in a traditional state smoothing setting. The former has the advantage that we could just re-use the posterior samples from the SMC estimation to generate realizations from the smoothing distribution, thus avoiding solving any games in the process. The downside to this is of course the fact that the cost paths

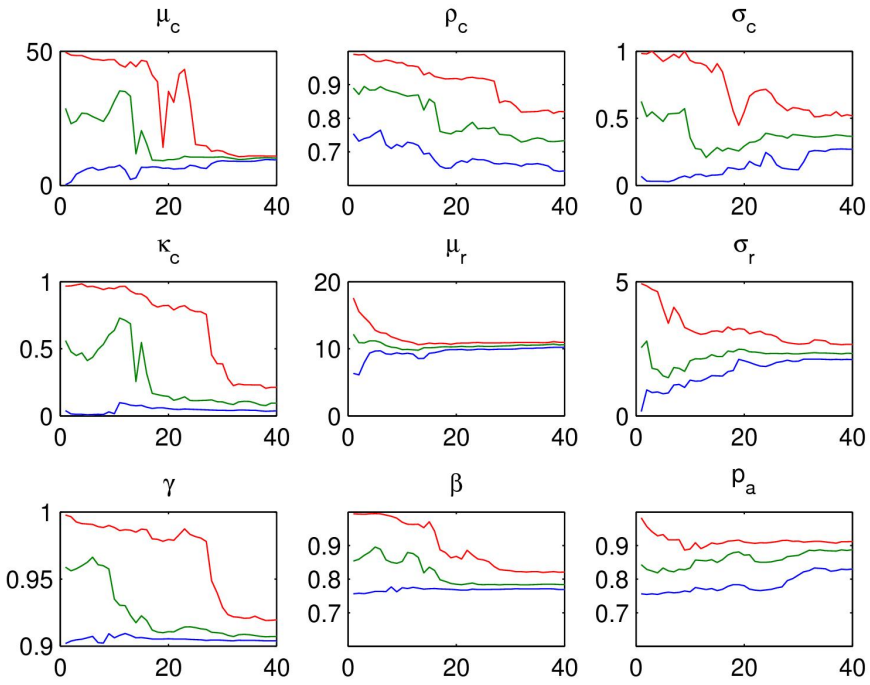

FIGURE 4.4: Trace plot from sequential estimation of three firm fully rational model with full model uncertainty. The three lines trace out the 95% posterior confidence interval at each point in the data set.

are unreliable during the earlier part of the sample period as the SMC algorithm hasn't seen enough data to pin down the parameter estimates. Figure 4.8 shows the average of 100 realizations of the total cost paths for the three dominant firms, the unobserved cost component is obtained by applying the smoothing algorithm with posterior parameter uncertainty. We can see that the cost paths become stable just prior to the market opening at $t = 15$. For the 26 market opening between $t = 15$ and the end of the sample, the leading firm Mylan is shown to have a distinctive cost advantage over the other two leading firms. Mylan's cumulative cost during these 26 markets is 7.5% lower than that of Novopharm and 9.3% lower than that of Lemmon. Mylan has a lower cost than both the other two dominant firms in 24 out of the 26 markets. The firm with the second highest market share, Novopharm, has

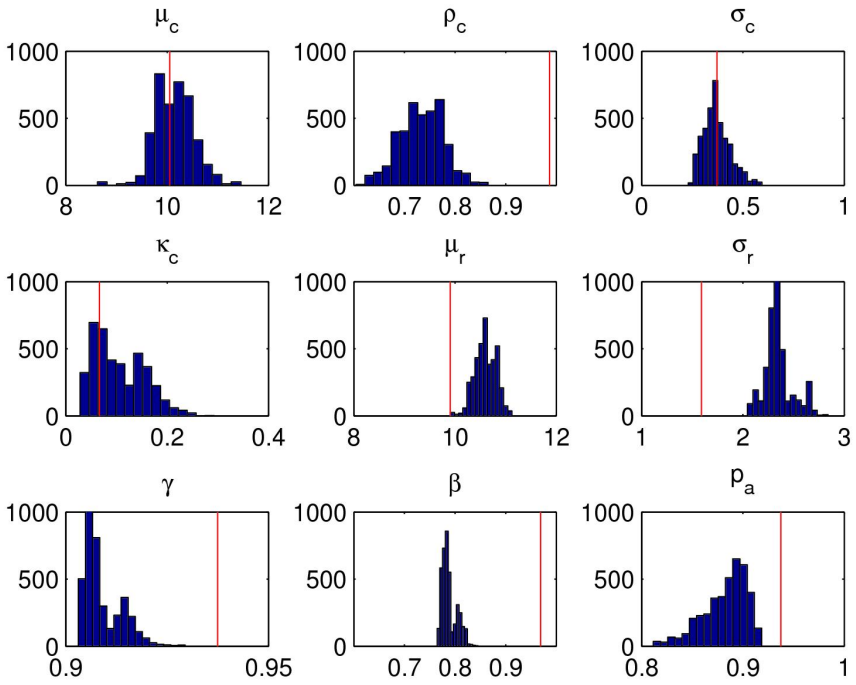

FIGURE 4.5: Marginal posteriors at last market opening from sequential estimation of three firm fully rational model with full model uncertainty. The red vertical lines mark the posterior modes from GHK's MCMC estimation.

a competitive cost advantage over Lemmon in 17 out of those 26 markets.

To get a more complete picture on the firms' cost structure at the cost of neglecting parameter uncertainty, we can first perform state filtering whilst fixing the model parameters at their posterior medians[7]. This process takes roughly the same amount of time as it takes to perform SMC parameter estimation as games have to be solved to evaluate the observation equation during state filtering. Once forward filtering is completed, we can use the smoothing algorithm without parameter uncertainty to sample the latent cost paths backwards through data. Figure 4.9 shows the average of 100 realizations from the full smoothing distribution as well as actual market entries by the three dominant firms. During the entire sample period, My-

[7] Fixing the parameters at the posterior modes or means produces near identical smoothing results.

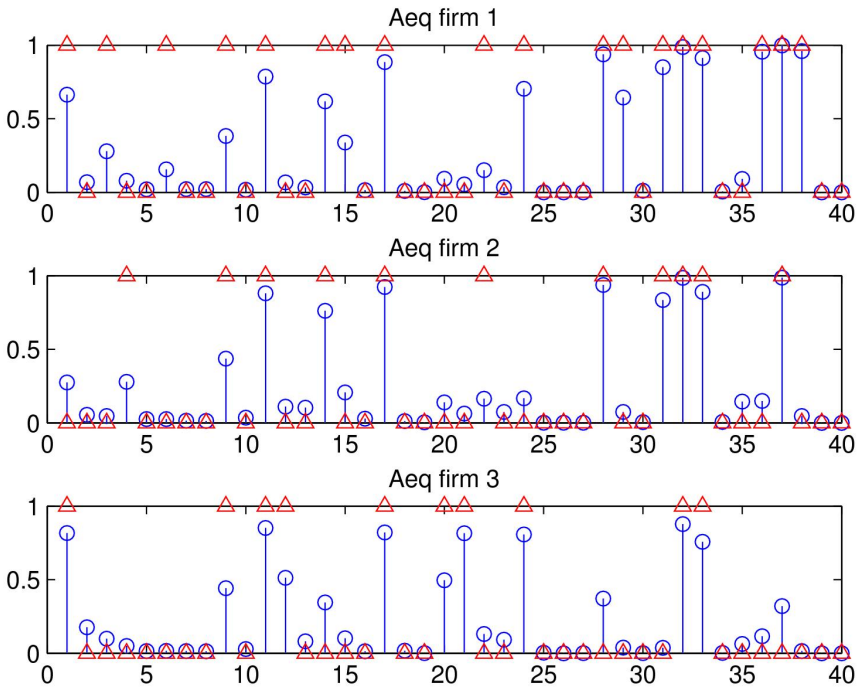

FIGURE 4.6: Particle prediction of firm actions from sequential estimation of three firm fully rational model with full model uncertainty. The triangles denote the observed entry decisions of the firms, the stems denote the computed particle average of equilibrium action profiles A_{it}^{E} conditional on $(r_t, x_t, y_{t-1}, \theta)$.

lan's cumulative cost is 15% lower than that of both Novopharm and Lemmon. Of the 40 market openings, Mylan has a competitive edge in cost over both Novopharm and Lemmon in 32 openings, and Novopharm has a cost advantage over Lemmon in 21 openings. These results solidify the earlier smoothing finding that Mylan enjoys a substantial cost advantage to give it the competitive edge in the dynamic entry game.

4.4 Discussion

This chapter presents an alternative SMC-based Bayesian method for the estimation of dynamic discrete games with serially correlated latent endogenous state variables.

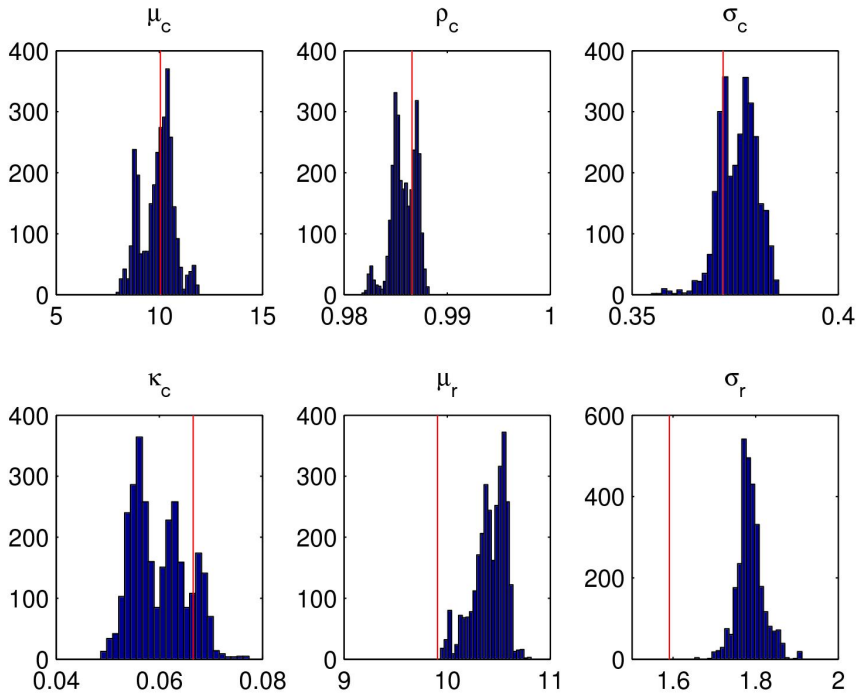

FIGURE 4.7: Marginal posteriors at last market opening from sequential estimation of three firm fully rational model (fixed γ, β and p_a) with fully informed priors. Initial particles for all parameters are sampled from $N(\theta_{GHK}, 10\sigma^2_{GHK})$ where θ_{GHK} and σ^2_{GHK} are the posterior modes and variances from GHK's MCMC estimation. The red verticle lines mark the posterior modes from GHK's MCMC estimation.

The main advantages of SMC over MCMC in the estimation of dynamic discrete choice models is three folds. First, SMC methods are more efficient. By combining state filtering and parameter learning, the resample-sample algorithm can identify most structural parameters from non-informative uniform priors whilst needing only a fraction of the computational resource required in MCMC estimation of similar models. Second, SMC methods can better explore parameter posterior spaces as they don't rely on Markov chain convergence. This is particularly useful in the context of game estimation as model parameters in dynamic discrete choice models in general are often times weakly identified. Lastly, particle smoothing can be used

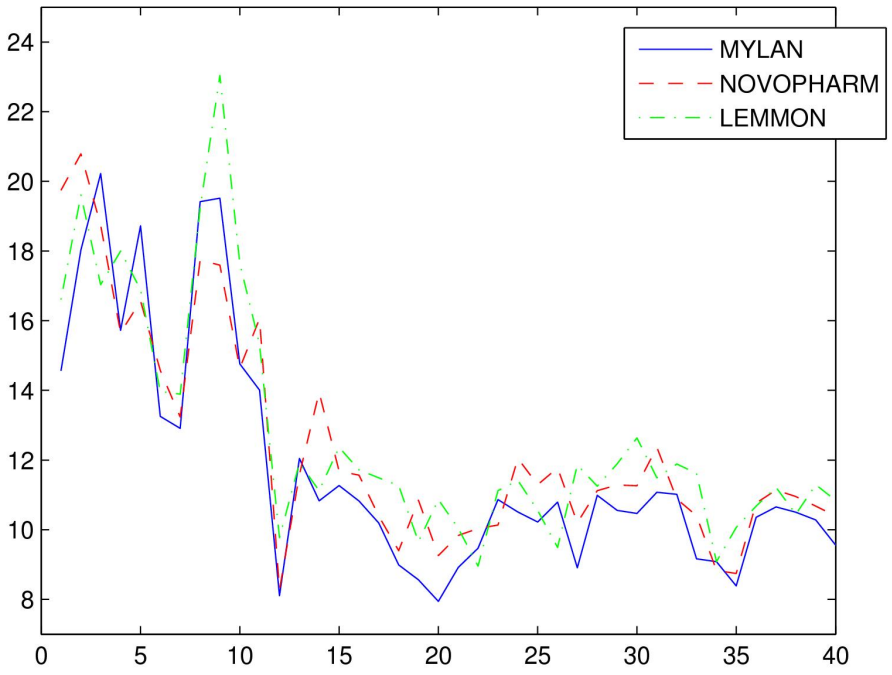

FIGURE 4.8: Smoothed total log costs from SMC estimation of three firm model with full rationality (fixed γ, β and p_a). The latent cost component is averaged over 100 approximate realizations from the full smoothing distribution $p(x_{1:T}, \theta | y_{1:T})$, which takes into account parameter posterior uncertainty.

to extract the unobserved state paths from their full smoothing distribution in the presence of model nonlinearity. In contrast, forward-filtering-backward-sampling for MCMC are typically done in close-form, thus requiring linearity and/or normality assumptions on the model. The proposed SMC method is flexible, and can in general be applied to a wide variety of similar models without much further modification.

There are several directions of future work in relation to this paper. First, despite the substantial computational improvement over MCMC estimation, SMC estimation of the entry game is still expensive to perform, and limits studies to only model a handful of game participants. The computational bottleneck comes from the full DP solution method used to evaluate the observation equation. It would thus be

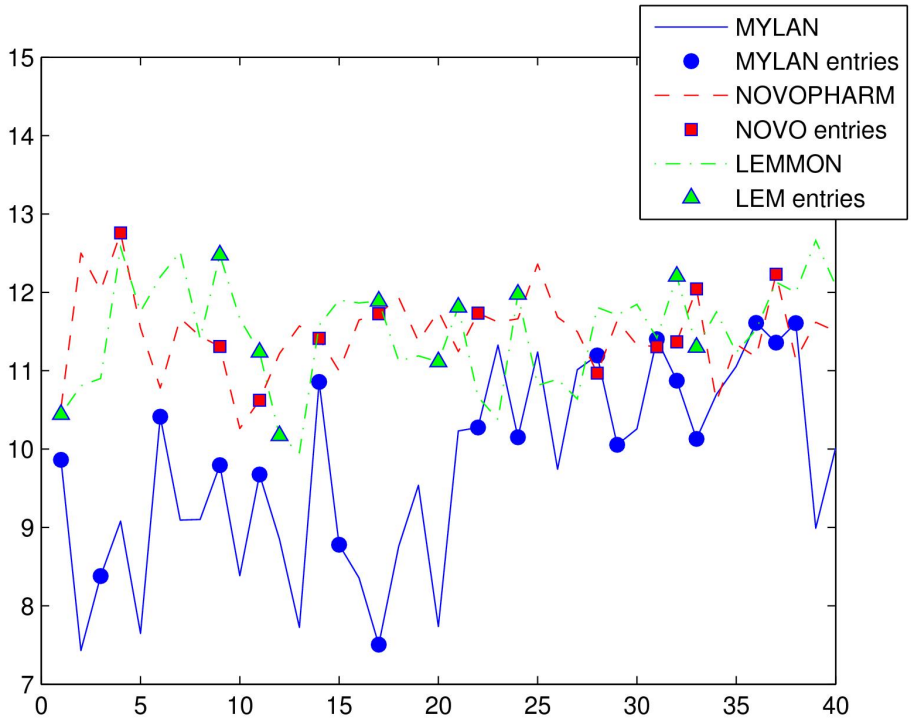

FIGURE 4.9: Smoothed total log costs from auxiliary filter run with all parameters fixed at SMC posterior medians of three firm model with full rationality (fixed γ, β and p_a). The latent cost component is averaged over 100 approximate realizations from $p(x_{1:T}|y_{1:T})$, the full smoothing distribution in a pure state smoothing context. The solid dots mark which markets the three firms entered in reality.

worthwhile to explore potential approximate DP methods to use in conjunction with Bayesian simulation methods. One potential candidate is artificial neural networks, which was successfully used in Norets (2008) to approximate DP solutions in the MCMC estimation of the bus engine replacement problem.

Another area of future work is to apply this method to other dynamic games such as dynamic games of incomplete information. This game theoretic setting permits richer hierarchical statistical modeling as players don't fully observe the moves of other players, which presents a unique challenge in the structural estimation context.

Appendix A

Linearized SW DSGE Model

Here we provide the complete description of the linearized DSGE model for the Euro area from Smets and Wouters (2003) used in solving the model in DYNARE, 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:

$$C_t = \frac{h}{1+h}C_{t-1} + \frac{1}{1+h}C_{t+1} - \frac{1-h}{(1+h)\sigma_c}(R_t - \pi_{t+1}) + \frac{1-h}{(1+h)\sigma_c}(\epsilon_t^b - \epsilon_{t+1}^b) \quad \text{(A.1)}$$

$$I_t = \frac{1}{1+\beta}I_{t-1} + \frac{\beta}{1+\beta}I_{t+1} + \frac{\varphi}{1+\beta}Q_t + \beta\epsilon_{t+1}^I - \epsilon_t^I \quad \text{(A.2)}$$

$$Q_t = -(R_t - \pi_{t+1}) + \frac{1-\tau}{1-\tau+\bar{r}^k}Q_{t+1} + \frac{\bar{r}^k}{1-\tau+\bar{r}^k}r_{t+1}^k + \eta_t^Q \quad \text{(A.3)}$$

$$K_t = (1-\tau)K_{t-1} + \tau I_{t-1} \quad \text{(A.4)}$$

$$\pi_t = \frac{\beta}{1+\beta\gamma_p}\pi_{t+1} + \frac{\gamma_p}{1+\beta\gamma_p}\pi_{t-1} + \frac{1}{1+\beta\gamma_p}\frac{(1-\beta\xi_p)(1-\xi_p)}{\xi_p}\left[\alpha r_t^k + (1-\alpha)w_t - \epsilon_t^a + \eta_t^p\right] \quad \text{(A.5)}$$

$$w_t = \frac{\beta}{1+\beta}w_{t+1} + \frac{1}{1+\beta}w_{t-1} + \frac{\beta}{1+\beta}\pi_{t+1} - \frac{1+\beta\gamma_w}{1+\beta}\pi_t + \frac{\gamma_w}{1+\beta}\pi_{t-1} \quad \text{(A.6)}$$

$$- \frac{1}{1+\beta}\frac{(1-\beta\xi_w)(1-\xi_w)}{\left(1+\frac{(1+\lambda_w)\sigma_L}{\lambda_w}\right)\xi_w}\left[w_t - \sigma_L L_t - \frac{\sigma_c}{1-h}(C_t - hC_{t-1}) - \epsilon_t^L - \eta_t^w\right] \quad \text{(A.7)}$$

$$L_t = -w_t + (1+\psi)r_t^k + K_{t-1} \quad \text{(A.8)}$$

$$Y_t = (1-\tau k_y - g_y)C_t + \tau k_y I_t + g_y\epsilon_t^G = \phi\epsilon_t^a + \phi\alpha K_{t-1} + \phi\alpha\psi r_t^k + \phi(1-\alpha)L_t \quad \text{(A.9)}$$

$$R_t = \rho R_{t-1} + (1-\rho)\left[\bar{\pi}_t + r_\pi(\pi_{t-1} - \bar{\pi}_t) + r_Y Y_t\right] + r_{\Delta\pi}(\pi_t - \pi_{t-1}) \quad \text{(A.10)}$$

$$+ r_{\Delta y}(Y_t - Y_{t-1}) - r_a\eta_t^a - r_L\eta_t^L + \eta_t^R, \quad \text{(A.11)}$$

where variables dated at $t+1$ refer to their rational expectations. 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) pro-

cesses, 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.

Appendix B

Solution Method for the Dynamic Entry Game

To evaluate the observation equation (4.13), the dynamic game must be solved for A_{it}^E. As the model is defined over an infinite horizon, we would be solving for a Markov perfect equilibrium which corresponds to solving the dynamic program characterized by Bellman equation (4.11). The solution method I use is the same value iteration method implemented in GHK, as such I will give a brief description of how the method works and interested readers are directed to GHK for details of the solver algorithm. As the term state variable is used in both the context of state-space models and that of dynamic programming, a clarification is needed before I move on to describe the DP solver. In the context of dynamic programming, the state variables at time t consist of the firm specific costs $(C_{1t}, \cdots, C_{It})$ and the market revenue R_t, these are the only information needed to solve the period-t subproblem of the original dynamic program as the solution concept is Markov perfect. This is slightly different from the state variables in the context of the state-space formulation of the game model, which excludes the observation R_t. For the following solution algorithm description, state variables refer to $S_t = (C_{1t}, \cdots, C_{It}, R_t)$. Everywhere

else in chapter 4, state variables refer to the firm specific costs only.

Given the log state variables $s_t = (c_{1t}, \cdots, c_{It}, r_t)$, the ex ante value functions are approximated by an affine function $V(s_t) = b + Bs_t$, where $V(s_t) = (V_1(s_t), \cdots, V_I(s_t))$. The goal of the solver is to find values of b and B such that the affine approximation of the ex ante value function satisfies equation (4.11). The solver works as follows given s_t:

1. Start with an initial guess of $V^{(0)}(s_t)$ by setting $(b^{(0)}, B^{(0)})$ to 0.

2. Obtain a set of unique state variables $\left\{s_t^j\right\}_{j=1}^J$ around s_t and compute for each of those s_t^j the choice specific value function in equation (4.8) at all possible action profiles A_t. In doing so, the continuation value $\beta\mathcal{E}\left[V_i^{(0)}(A_{i,t+1}^E, A_{-i,t+1}^E, C_{i,t+1}, C_{-i,t+1}, R_{t+1})|A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t\right]$ is computed using Gauss-Hermite quadrature and $(b^{(0)}, B^{(0)})$.

3. For each s_t^j, compute the corresponding equilibrium action profile A_j^E by checking the optimality condition (4.10). To deal with the possibility of multiple equilibria, the action profiles are ranked in the order of increasing total cost, and the action profile that satisfies (4.10) with the lowest total cost is chosen as the equilibrium profile.

4. Denote the choice specific functions evaluated at (s_t^j, A_j^E) as $V^{(1)}(A_j^E, s_t^j) = (V_1^{(1)}(s_t^j, A_j^E), \cdots, V_I^{(1)}(s_t^j, A_j^E))$. Using $\left\{V^{(1)}(A_j^E, s_t^j), s_t^j\right\}_{j=1}^J$ as data, update the coefficients in the affine approximation of the ex ante value function to $(b^{(1)}, B^{(1)})$ using multivariate regression.

5. Go back to step 3, and repeat the procedure until the affine coefficients converge to $(b^{(*)}, B^{(*)})$. With $V(s_t) = b^* + B^* s_t$, we can now compute all choice

specific value functions for state s_t and thus the equilibrium strategy profile that corresponds to s_t.

Equilibrium action profile A_t^E computed using the above algorithm is referred to as a boundedly rational equilibrium in GHK. Recall that in adopting the measurement error form for the likelihood for the observed entry decisions in (4.14), we assume that there is a small probability p_a a firm will implement an entry decision contrary to its original plan. The DP algorithm above does not take into account the fact that equilibrium profiles might not actually be played in reality. One could account for this misclassification explicitly in the DP algorithm by adopting the following choice specific value function instead of (4.8),

$$V_i^{FR}(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t) = \tag{B.1}$$

$$\prod_{i=1}^{I}(p_a)^{I(A_{it}^O = A_{it})}(1-p_a)^{I(A_{it}^O \neq A_{it})}V_i(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t) \tag{B.2}$$

where $A_t^O = (A_{1t}^O, \cdots, A_{It}^O)$ is the vector of observed entry actions at time t. Equilibrium profiles computed using (B.1) is referred to as a fully rational equilibrium in GHK. GHK demonstrated that MCMC estimation is not sensitive to the type of equilibrium concept used to construct the likelihood. With SMC estimation however, the posterior distributions vary considerably depending on which solution concept is used.

Regardless of solution type, this DP solver is the computational bottleneck during estimation as each new MCMC proposal or SMC particle calls for multiple games to be solved. For efficient estimation, the state space is mapped to a grid and states that fall into the same grid space share the same coefficients in the affine approximation of the ex ante value functions, see GHK for details regarding the grid increments. It's worthwhile to explore approximate DP solution methods to avoid solving repeated DPs and greatly speed up estimation of dynamic game models such as this.

Bibliography

Ackerberg, D., J. Geweke, and J. Hahn (2009), "Comments on convergence properties of the likelihood of computed dynamic models." *Econometrica*, 77, 2009–2017.

Aguirregabiria, V. and P. Mira (2002), "Swapping the nexted fixed point algorithm: A class of estimators for discrete markov decision models." *Econometrica*, 70, 1519–1543.

Aguirregabiria, V. and P. Mira (2004), "Dynamic discrete choice structural models: A survey." *Journal of Econometrics*, 156, 38–67.

Arcidiacono, P. and R. Miller (2008), "CCP estimation of dynamic discrete choice models with unobserved heterogeneity." Working paper, Duke University.

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

Bajari, P., H. Hong, and S. Ryan (2004), "Identification and estimation of discrete games of complete information." Working paper, Duke University.

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

Berry, S. (1992), "Estimation of a model of entry in the airline industry." *Econometrica*, 60, 889–917.

Bresnahan, T. and P. Reiss (1990), "Entry in monopoly markets." *Review of Economic Studies*, 57, 531–553.

Bresnahan, T. and P. Reiss (1991), "Empirical models of discrete games,." *Journal of Econometrics*, 48, 57–81.

Carvalho, C.M., M.S. Johannes, H.F. Lopes, and N.G. Polson (2010), "Particle learning and smoothing." *Statistical Science*, 25, 88–106.

Chen, H., H. F. Lopes, and F. Petralia (2010), "Smc estimation of dsge models." Working paper, Duke University and University of Chicago.

Chib, S. and M. Srikanth (2010), "Tailored randomized block mcmc methods with application to dsge models." *Journal of Econometrics*, 155, 19–38.

Ching, A. (2009), "A dynamic oligopoly structural model for the prescription drug market after patent expiration." *International Economic Review*. Forthcoming.

Chopin, N. (2004), "Central limit theorem for sequential monte carlo and its application to bayesian inference." *The Annals of Statistics*, 32, 2385–2411.

Ciliberto, F. and E. Tamer (2009), "Market structure and multiple equilibria in airline markets." *Econometrica*, 77, 1791–1828.

Crisan, D. and A. Doucet (2002), "A survey of convergence results on particle filtering methods for practitioners." *IEEE Transcations on Signal Processing*, 50, 736–746.

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

Doucet, A., N. de Freitas, and N. Gordon, eds. (2001), *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York.

Doucet, A. and A.M. Johansen (2008), "A tutorial on particle filtering and smoothing: Fifteen years later." Manuscript, University of British Columbia and University of Warwick.

Dutta, P. K. and R. K. Sundaram (1998), *The Equilibrium Existence Problem in General Markovian Games*, 159–207. Cambridge University Press.

Fearnhead, P. (2002), "Markov chain monte carlo, sufficient statistics, and particle filters." *Journal of Computational and Graphical Statistics*, 11, 848–862.

Fearnhead, P. and P. Clifford (2003), "On-line inference for hidden markov models via particle filters." *Journal of the Royal Statistical Society*, 65, 887–899.

Fernández-Villaverde, J. (2009), "The econometrics of dsge models." PIER Working Paper 09-008.

Fernández-Villaverde, J. and J.F. Rubio-Ramírez (2005), "Estimating dynamic equilibrium economies: Linear versus nonlinear likelihood." *Journal of Applied Econometrics*, 20, 891–910.

Fernández-Villaverde, J., J.F. Rubio-Ramírez, and M.S. Santos (2006), "Convergence properties of the likelihood of computed dynamic models." *Econometrica*, 74, 93–119.

Gallant, A.R., H. Hong, and A. Khwaja (2010), "Dynamic entry with cross product spillovers: An application to the generic drug industry." Working paper. Duke University and Stanfor University.

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

Giacomini, R. and B. Rossi (2007), "Model selection and forecast comparison in unstable environments." Working Paper, Duke University.

Gilks, W. and C. Berzuini (2001), "Following a moving targat: Monte carlo inference for dynamic bayesian models." *Journal of Royal Statistical Society*, 63, 127–146.

Godsill, S.J., A. Doucet, and M. West (2004), "Monte carlo smoothing for nonlinear time series." *Journal of the American Statistical Association*, 99, 156–168.

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

Hitsch, G. (2006), "An empirical model of optimal dynamic product launch and exit under demand uncertainty." *Marketing Science*, 25, 25–50.

Hotz, J. and R. Miller (1993), "Conditional choice probability and the estimation of dynamic models." *Review of Economic Studies*, 60, 497–529.

Hotz, J., R. Miller, S. Sanders, and J. Smith (1994), "A simulation estimator for dynamic models of discrete choice." *Review of Economic Studies*, 61, 265–289.

Imai, S., N. Jain, and A. Ching (2005), "Bayesian estimation of dynamic discrete choice models." Working paper, Concordia University, Northern Illinois University and University of Toronto.

Johannes, M. and N.G. Polson (2008), "Exact particle filtering and learing." Working paper.

Judd, K.L. and S.M. Guu (1993), *Perturbation solution methods for economic growth model*. New York: Springer.

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

Künsch, H.R. (2005), "Recursive monte carlo filters: Algorithms and theoretical analysis." *The Annals of Statistics*, 33, 1983–2021.

Liu, J. and M. West (2001), *Combined parameters and state estimation in simulation-based filtering*, chapter 10, 197–217. Springer-Verlag, New York.

Ljungqvist, L. and T.J. Sargent (2000), *Ljungqvist, L., and Sargent T.J.* MIT Press, Cambridge, Massachusetts.

Magill, J.P.M. (1977), "A local analysis of n-sector capital under uncertainty." *Journal of Economic Theory*, 15, 219–221.

Mazzeo, M. (2002), "Product choice and oligopoly market structure." *Rand Journal of Economics*, 33, 221–242.

Moral, P. Del and A. Guionnet (1999), "Central limit theorem for nonlinear filtering and interacting particle systems." *The Annals of Applied Probability*, 9, 275–297.

Norets, A. (2008), "Estimation of dynamic discrete choice models using artificial neural network approximations." Working paper, Princeton University.

Norets, A. (2009), "Inference in dynamic discrete choice models with serially correlated unobserved state variables." *Econometrica*, 77, 1665–1682.

Orhun, Y. (2006), "Spatial differentiation in the supermarket industry." Working paper, University of Chicago.

Papaspiliopoulos, O. and G. Roberts (2008), "Stability of the gibbs sampler for bayesian hierarchical models." *Annals of Statistics*, 36, 95–117.

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

Polson, N.G., J.R. Stroud, and P. Müller (2008), "Practical filtering with sequential parameter learning." *J. R. Statist. Soc*, 70, 413–428.

Rust, J. (1987), "Optimal replacement of gmc bus engines: An empirical model of harold zurcher." *Econometrica*, 55, 999–1033.

Rust, J. (2006), "Dynamic programming." Entry for consideration by the New Palgrave Dictionary of Economics.

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

Scott-Morton, F. (1999), "Entry decisions into generic pharmaceutical markets." *Rand Journal of Economics*, 30, 421–440.

Seim, K. (2006), "An empirical model of rm entry with endogenous product-type choices." *Rand Journal of Economics*, 37, 619–640.

Shen, Q. (2010), "A dynamic model of entry and exit in a growing industry." Working paper, University of Pennsylvania.

Shiga, T. and H. Tanaka (1985), "Central limit theorem for a system of markovian particles with mean field interaction." *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 69, 439–459.

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.

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

Tamer, E. (2003), "Incomplete simultaneous discrete response model with multiple equilibria." *Review of Economics Studies*, 70, 147–165.

West, M. (1993a), "Approximating posterior distributions by mixtures." *Journal of Royal Statistical Society*, 55, 409–422.

West, M. (1993b), *Mixture models, Monte Carlo, Bayesian updating and dynamic models*, 325–333. Interface Foundation of North America, Fairfax Station, Virginia.

Zhu, T. and V. Singh (2006), "Spatial competition with endogenous location choices: An application to discount retailing." *Quantative Marketing and Economics*. Forthcoming.

Biography

Hao Chen was born on February 23, 1980 in Shanghai, China. He received his bachelor degree in computer science in May 2002 from Duke University. He received an M.S. in applied mathematics from UMass Amherst in May 2004. He will receive a Ph.D. in Business Administration under the Decision Sciences department upon successful defense of this dissertation. Upon graduation, he will join SAS Institute as a senior research statistical developer.