A Bayesian Approach to Estimation of Dynamic Models with Small and Large Number of Heterogeneous Players and Latent Serially Correlated States *

A. Ronald Gallant
Penn State University

Han Hong
Stanford University

Ahmed Khwaja
University of Cambridge

April 2017

Abstract

We propose a Bayesian approach to estimating dynamic models that can have state variables that are latent, serially correlated, and heterogeneous. Our approach employs sequential importance sampling and is based on deriving an unbiased estimate of the likelihood within a Metropolis chain. Under fairly weak regularity conditions unbiasedness guarantees that the stationary density of the chain is the exact posterior, not an approximation. Results are verified by Monte Carlo simulation using two examples. The first is a dynamic game of entry involving a small number of firms whose heterogeneity is based on their current costs due to feedback through capacity constraints arising from past entry. The second is an Ericson-Pakes (1995) style game with a large number of firms whose heterogeneity is based on the quality of their products with firms competing through investment in product quality that affects their market share and profitability. Our approach facilitates estimation of dynamic games with either small or large number of players whose heterogeneity is determined by latent state variables, discrete or continuous, that are subject to endogenous feedback from past actions.

Corresponding author: A. Ronald Gallant, Penn State University, Department of Economics, 511 Kern Graduate Building, University Park PA 16802, aronaldg@gmail.com.

Keywords: Dynamic Games, Partially Observed State, Heterogeneous Agents, Endogenous State, Serially Correlated State, Particle Filter.

JEL Classification: E00, G12, C51, C52

*Supported by the National Science Foundation (SES-1024504, 0721015). We thank Neil Shephard, Michael Pitts, and Thomas Flury for helpful discussions.

1 Introduction

In many economic applications a researcher is interested in estimating a dynamic model where some of the agent specific states may be subject to feedback from past actions (e.g., Rust (1987)), partially observed and heterogeneous (e.g., Keane and Wolpin (1997)) or serially correlated (e.g., Miller (1984), Pakes (1986)).[1] In this paper we propose a Bayesian approach to estimating dynamic models with a partially observed state that has a Markovian representation of the dynamics and an algorithm to solve the model including those with serially correlated, endogenous, heterogeneous, state variables. The method uses sequential importance sampling (particle filter) to compute an unbiased estimate of the likelihood within a Metropolis MCMC chain. Here, unbiasedness means that expectation is with respect to the uniform draws that are used to compute the particle filter with all else held fixed. Unbiasedness guarantees that the stationary density of the chain is the exact posterior, not an approximation. The regularity conditions are weak and should be satisfied by most stationary, parametric models of interest.

Our approach contributes to the literature in several ways. We derive the unbiasedness of the estimator of the likelihood that leads to an exact posterior. This allows for tractable computation and feasible estimation of a dynamic model. In addition the latent state variables can be either discrete or continuous. Moreover, it permits endogenous feedback of past actions on the latent state variables that leads to heterogeneity among the players. We illustrate our method with two examples. The first example in Section 6.1 is based on a dynamic model of entry developed in Gallant, Hong, and Khwaja (2017). This is a dynamic discrete game in which a firm's cost of production is a continuous state variable that is serially persistent, unobserved (to the researcher), and endogenous to past actions. The endogenous feedback arises because past entry in markets for other products creates a capacity constraint that affects the costs of entering a market in the current period. This endogenous evolution of costs induces heterogeneity among firms over time in the model. Our second example, in Section 6.2, is an Ericson and Pakes (1995) style model from Weintraub, Benkard, and Roy (2010) with a large number of firms whose heterogeneity is based on the

[1] For more elaborate discussion on the motivations and rationale for estimating such dynamic models see e.g., Rust (1994), Keane (2010), Wolpin (2013).

quality of their products. The firms compete by investing in product quality that in turn affects their market share, revenues and profits.

Empirical models of static and dynamic games are differentiated by their information structures. A unified approach for dynamic models of incomplete information has been developed by Hu and Shum (2012). Blevins (2016) also uses sequential importance sampling for estimating games of incomplete information, while Blevins, Khwaja, and Yang (2017) allow for two-step estimation of such games with endogenous feedback. In our paper we focus on dynamic, complete information games because of the paucity of literature on estimation of such games. These games typically require the use of a combinatorial algorithm to search for an equilibrium instead of the continuous fixed point mapping used in incomplete information models. Unlike games of incomplete information, the complete information assumption requires that no player has any private information. However, it allows substantial unobserved heterogeneity at the level of the firms because the researcher does not observe all the information that the players have. In contrast, games of incomplete information require there to be no difference between what is observed by the players and the researcher. While static games of complete information have been estimated by, e.g., Bresnahan and Reiss (1991), Berry (1992), Ciliberto and Tamer (2009) and Bajari, Hong, and Ryan (2010), to our knowledge, we are the first to study dynamic games of complete information.

We prove the unbiasedness of an estimator of a likelihood obtained via particle filtering under regularity conditions that allow for endogenous feedback from the observed measurements to the dynamic state variables. Endogenous feedback is the feature that distinguishes this paper from the bulk of the particle filter literature. We establish our results by means of a recursive setup and an inductive argument that avoids the complexity of ancestor tracing during the resampling steps. This process allows elegant, compact proofs.

The rest of paper is organized as follows. We discuss related literature in Section 2. Section 3 describes the games to which our results apply. An algorithm for unbiased estimation of a likelihood is proposed and unbiasedness is proved in Section 4. The MCMC estimation algorithm is presented in Section 5. Two examples are described in Section 6: the first has a small number of players, the second a large number of players. Simulation results for the two examples are presented in Section 7. Section 8 concludes.

2 Related Literature

There is a growing literature on the estimation of games. Some of this literature focuses on games of incomplete information, either static (e.g., Haile, Hortaçsu, and Kosenok (2008), Ho (2009)) or dynamic (e.g., Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2007), Pakes, Ostrovsky, and Berry (2007)). The literature on estimating games of incomplete information is largely based on the two-step conditional choice probability estimator of Hotz and Miller (1993).[2] Arcidiacono and Miller (2011) have extended the literature on two step estimation of dynamic models of discrete choice to allow for unobserved heterogeneity in discrete types of latent states using the EM algorithm. Bayesian approaches that use MCMC for integrating out the unobserved state variables that are serially correlated over time have been developed by Imai, Jain, and Ching (2009) and Norets (2009). These papers focus on single agent dynamic discrete choice models with unobserved state variables. In contrast, we use sequential importance sampling to integrate out the unobserved state variables. Additionally, we are the first to apply this method to multi-agent dynamic games with strategic interaction, which are more computationally complex than single agent dynamic models.

The purely methodological papers most closely related to the econometric approach used here are Keane (1994), and more recently, Flury and Shephard (2010) and Pitt, Silva, Giordani, and Kohn (2012).[3] Fernandez-Villaverde and Rubio-Ramirez (2005) used sequential importance sampling to estimate dynamic stochastic general equilibrium models. Most of this literature, however, abstracts from the strategic interaction between agents. Ackerberg (2009) has developed a method for using importance sampling coupled with a change of variables technique to provide computational gains in estimating certain game theoretic and dynamic discrete choice models that admit a random coefficient representation.

[2]The two step estimation strategy makes the restrictive assumption that there is no market or firm level unobserved heterogeneity other than an IID random shock across both time and players. This assumption rules out any dynamics in the latent state variables. Moreover, it precludes any private information that an agent might possess about its rivals that is unavailable to the researcher.

[3]See Doucet, De Freitas, Gordon, et al. (2001) and Liu (2008) for an overview and examples of other applications.

3 The Game

A prominent example of a model to which our results apply is a dynamic game of complete information which we describe next. The game consists of I players, $i = 1, \ldots, I$, who can choose action a_{it} at each time period t. Let $a_t = (a_{1t}, a_{2t}, \ldots, a_{It})$. In an entry game as in our first Monte Carlo example below in Section 6.1, $a_{it} = 1$ if firm i enters at time t, if not, $a_{it} = 0$. Another example of a_{it} could be the level of investment which affects the quality of a firm's product as in an Ericson and Pakes (1995) style model, that is found in our second Monte Carlo example below in Section 6.2. Time runs in discrete increments from $t = 0, \ldots, \infty$. The state vector is $x_t = (x_{1t}, x_{2t})$, where in turn $x_{1t} = (x_{1t,i}, x_{1t,-i})$ and $x_{2t} = (x_{2t,i}, x_{2t,-i})$. The actions of all players a_t and the state vector $x_t = (x_{1t}, x_{2t})$ is observable by all players. We (the researchers) only observe the actions of all players a_t and the state x_{2t} but not x_{1t}. The $x_{1t} = (x_{1t,i}, x_{1t,-i})$ is an agent specific latent (to the researcher) state that is allowed to be serially correlated. In our first Monte Carlo example below in Section 6.1, this is considered to be the latent firm specific cost of production. In the our second Monte Carlo example below in Section 6.2, this is the firm's product quality. The $x_{2t} = (x_{2t,i}, x_{2t,-i})$ is any observable market level or firm specific observable state, such as a market level demand or cost shifter or firm characteristic such as past market entry experience. The game is indexed by a parameter vector θ that is known to the players and which we seek to estimate.

To formalize the model further, we define the reduced form one-shot payoff function (e.g., Bresnahan and Reiss (1991), Berry (1992)) as,

$$\begin{aligned}&\Pi_i(a_{it}, a_{-it}, x_{1t,i}, x_{1t,-i}, x_{2t,i}, x_{2t,-i}, \theta) \\ &\quad = R(a_{it}, a_{-it}, x^R_{1t,i}, x^R_{1t,-i}, x^R_{2t,i}, x^R_{2t,-i}, \theta^R) - C(a_{it}, a_{-it}, x^C_{1t,i}, x^C_{1t,-i}, x^C_{2t,i}, x^C_{2t,-i}, \theta^C), \quad (1)\end{aligned}$$

where $R(\cdot)$ is a revenue function, and $C(\cdot)$ is a cost function. The state variables can affect either the revenue or cost functions, and are denoted as $x^R_{qt,j}$ and $x^C_{qt,j}$ respectively as appropriate, for $q = 1, 2$ and $j = i, -i$. Moreover, the latent firm specific stochastic state, $x_{1t,i}$ can in principle affect either the revenue or cost function or both. In our first Monte Carlo example it is modeled as firm specific cost of production. However, in the our second Monte Carlo example it affects the revenue function. The θ^R and θ^C are vectors of model

parameters associated with the revenue and cost functions respectively. It should be noted that the "structural error" or the primary source of randomness from the point of view of the researcher is provided by x_{1t}.

The dynamics in the model arise because given the state x_t, the players choose actions a_t according to the probability density function $p(a_t|x_t, \theta)$. Furthermore, the transition density $p(x_t|a_{t-1}, x_{t-1}, \theta)$ governs the evolution of the state vector. This transition density allows for serial persistence and endogenous feedback from the past actions a_{t-1} on the latent states x_{1t}.

Given an initial state x_t at time t, the firm's expected present discounted profit is

$$\mathcal{E}\left[\sum_{\tau=t}^{\infty}\beta^{\tau-t}\Pi_i(a_{i\tau}, a_{-i\tau}, x_\tau)\,\middle|\, x_t\right], \tag{2}$$

where β is the discount factor, $0 \leq \beta < 1$. Each firm i seeks to maximize the present discounted value of its stream of profits at each time period t conditional on its rivals' action profiles being best response equilibrium strategies. The expectation operator here $\mathcal{E}$ is with respect to the future evolution of state variables and actions of all firms.

In specifying a dynamic game of complete information we focus on Markov Perfect Equilibria (MPE) in pure strategies. In doing this we build on the literature on estimating static games of complete information (e.g., Bresnahan and Reiss (1991), Berry (1992), Ciliberto and Tamer (2009)) and dynamic games of incomplete information (e.g., Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2007), Pakes, Ostrovsky, and Berry (2007)). We use the following notation in order to define the MPE strategies for the game. As stated above x_t is the payoff-relevant state for each firm i at time t. Let $\mathcal{X}$ represent the state space containing all feasible values of x_t for each firm i. Similarly, let $\mathcal{A}_i$ represent firm i's choice set (i.e., $a_{it} \in \mathcal{A}_i$). Then we define a Markov strategy for firm i as a function $\sigma_i : \mathcal{X} \to \mathcal{A}_i$ mapping payoff-relevant state variables to the set of all possible actions. Furthermore, $\sigma = (\sigma_1, \ldots, \sigma_I)$ denotes a profile of Markov strategies for all firms.

Dropping the time indices, and denoting the Markov Perfect Equilibrium strategy profile (to be defined below in the next paragraph) $(\tilde{\sigma}_i, \tilde{\sigma}_{-i})$, when the the realized actions and strategies in the current period are, $a_i = \sigma_i(x)$ and $a_{-i} = \sigma_{-i}(x)$, the choice specific value function is given by,

$$V_i(x;\sigma_i,\sigma_{-i}) = \Pi_i(\sigma_i(x),\sigma_{-i}(x),x_i) + \beta\mathcal{E}\left[V_i(x';\tilde{\sigma}_i,\tilde{\sigma}_{-i}) \mid x, a_i = \sigma_i(x), a_{-i} = \sigma_{-i}(x)\right] \quad (3)$$

In the above equation, the first component is the current payoff, Π_i. The first argument of the payoff function Π_i is $a_i = \sigma_i(x)$, which is the implied action by firm i under strategy σ_i when the state is x. Similarly, the second argument is $a_{-i} = \sigma_{-i}(x)$, which is a profile of rival actions given the state x. The second component in the above equation is the present discounted value of the stream of profits conditional on the current actions of all firms and the current state, when the firms play MPE strategies in the future. The expectation is with respect to the future states and actions of all players.

The Markov Perfect Equilibrium strategy profile $\tilde{\sigma}$ is defined such that no firm i unilaterally deviates from its strategy $\tilde{\sigma}_i$ when $\tilde{\sigma}_{-i}$ is the strategy profile adopted by its rivals. Hence, there is no alternative Markov strategy σ_i for any firm i that provides a higher level of present discounted value of expected profits (in terms of $V_i(\cdot)$) than $\tilde{\sigma}_i$ when the firm's rivals are using their equilibrium strategies $\tilde{\sigma}_{-i}$. Stated more formally, a Markov Perfect Equilibrium (MPE) strategy profile $\tilde{\sigma}$ is one such that for all firms i, all states x, and for all alternative Markov strategies σ_i the following condition holds:

$$V_i(x;\tilde{\sigma}_i,\tilde{\sigma}_{-i}) \geq V_i(x;\sigma_i,\tilde{\sigma}_{-i}). \quad (4)$$

From Equation (4), it should be noted that $V_i(x) \equiv V_i(x;\tilde{\sigma}_i,\tilde{\sigma}_{-i})$ is defined in the literature to be the ex-ante value function. It gives the expected present discounted value of profits obtained by firm i when players use MPE strategies $\tilde{\sigma}$ and the state is x. Implicit in what follows is an assumption that the game is stationary, an algorithm to solve the game is available and that an equilibrium to the game exists.

One limitation of relying on the Markov Perfect Equilibrium solution concept is that it is computationally intractable when a game consists of a large number of players. Thus, we show below in Section 6.2 that our approach can also apply in the context of the Oblivious Equilibrium solution concept developed by Weintraub, Benkard, and Roy (2008) to deal specifically with games with a large number of players. In particular we do this because Weintraub, Benkard, and Roy (2008) show that this approach can be quite accurate in approximating the MPE solution for games with a large number of players.

4 Evaluating the Likelihood

In this section we derive an unbiased particle filter estimate of the likelihood for a Markov process with partially observed state and endogenous feedback that is general enough to accommodate the game described in Subsection 3. Because of endogenous feedback, no previous proof of unbiasedness, e.g., Pitt, Silva, Giordani, and Kohn (2012), applies to our knowledge. In fact, we think that the remarkable simplicity of our proofs makes them of interest regardless. In Theorem 1 we establish a recursion using Bayes theorem. The idea is straightforward and is expressed as one four line equation. The remainder of the proof is algebra to reduce the basic expression to model primitives. In Corollary 1 we establish unbiasedness via a simple two line telescoping expression. In Theorem 2 we show that resampling is a mere footnote requiring only three sentences to dismiss.

Because we only require unbiasedness, our regularity conditions are quite weak – much weaker than is standard in the particle filter literature.[4] While the result does not require that the number of particles tend to infinity, the number of particles does affect the rejection rate of the MCMC chain so that the number of particles, like the scale of the proposal density, becomes a tuning parameter of the chain that has to be adjusted. Below in Section 5 we use MCMC to compute the posterior. If one has an unbiased estimator of the likelihood the posterior is the stationary distribution of the MCMC chain.[5]

What distinguishes our approach from, e.g., simulation methods, which would be applicable, is that an inference is exact within the Bayesian paradigm.[6] All other approaches known to us only approximate the likelihood to within an accuracy that is often not known. Therefore, inferences using other methods are approximate, not exact.

The last sentence of the preceding paragraph has to be taken with a grain of salt. To quote a referee:

> This is somewhat "snarky"; non-Bayesians are probably just as happy with their

[4] See, e.g. Andrieu, Douced, and Holenstein (2010) and the references therein.

[5] See e.g., Flury and Shephard (2010) and Pitt, Silva, Giordani, and Kohn (2012).

[6] A recent exception is the simulation estimator proposed by Gallant and Tauchen (2017). It is exact as to theory but does require that the distributional assumption regarding a semi-pivotal that implies exactness be verified by simulation in practice. Depending on one's attitude toward verification by simulation, one may or may not regard the method as exact.

asymptotic or approximate inference as Bayesians are with their "exact" inference (and mutually suspicious of each others approaches!)

In the case one has a conditional density of the form $f(y|x,\theta)$ and a marginal $f(x|\theta)$, where y, x, are vectors of the same length with y observed and x not, the property of a particle filter that is usually of interest is its ability to generate a sample from the density $f(x|y,\theta)$ for given θ. In the abstract, a particle filter algorithm can be viewed as a transformation of a vector u of independent uniform draws to a set of draws $\left\{x^{(k)}\right\}_{k=1}^{N}$ from $f(x|y,\theta)$. We are not concerned with that property of the algorithm here. Rather, we are concerned with a side effect of the algorithm which is that at an intermediate stage of the computations it generates an estimate $\hat{\ell}(u)$ of $\ell = \int f(y|x,\theta)f(x|\theta)\,dx$ that is unbiased in the sense that $\int \hat{\ell}(u)f(u)\,du = \ell$, where $f(u)$ is the density of u. In Subsection 4.2 we provide a bare bones description of a particle filter that is adequate to allow one to follow the theoretical discussion. The theoretical discussion is followed by a detailed description of the algorithm that we propose in Subsection 4.6.

4.1 Requirements

The essentials of the game of Section 3 relative to the requirements of filtering are as follows. The state vector is

$$x_t = (x_{1t}, x_{2t}), \tag{5}$$

where x_{1t} is not observed and x_{2t} is observed. The observation (or measurement) density is

$$p(a_t \,|\, x_t, \theta). \tag{6}$$

The transition density is denoted by

$$p(x_t \,|\, a_{t-1}, x_{t-1}, \theta). \tag{7}$$

The marginal for x_{1t} is

$$p(x_{1t}|a_{t-1}, x_{t-1}, \theta). \tag{8}$$

The marginal stationary density is denoted by

$$p(x_{1t} \,|\, \theta). \tag{9}$$

ASSUMPTION 1 We assume that we can draw from (8) and (9). As to the latter, one way to draw a sample of size N from (9) is to simulate the game and set $x_1^{(k)} = x_{1,\tau+M*k}$ for $k = 1, \ldots, N$ for some τ past the point where transients have died off and some M. Large M has the advantage that the $x_1^{(k)}$ are nearly serially uncorrelated. We can draw from (8) by drawing from (7) and discarding x_{2t}. We assume that there is either an analytic expression or algorithm to compute (6) and (7). We assume the same for (8) but if this is difficult some other importance sampler can be substituted as discussed in Subsection 4.5.

The full information likelihood is

$$\ell = p(a_0, x_{2,0}) \prod_{t=1}^{T} p(a_t, x_{2,t} \,|\, a_{t-1}, x_{2,t-1}, \ldots, a_0, x_{2,0}, \theta) = p(a_0, x_{2,0}) \prod_{t=1}^{T} C_t. \tag{10}$$

We are unwilling to impose the additional structure on the game necessary to be able to estimate $p(a_0, x_{2,0})$ unbiasedly. Therefore, as is routinely done in time series analysis, we discard the information in the stationary density for $(a_0, x_{2,0})$. With this convention, the object of interest is the partial information likelihood

$$\ell' = \prod_{t=1}^{T} C_t. \tag{11}$$

In the event that one can evaluate the density $p(a_t, x_{2t} \,|\, x_{1,t}, \theta)$ or has some other means to estimate $p(a_0, x_{2,0})$ unbiasedly, the modifications to what we propose are trivial and are discussed later.

4.2 A Particle Filter

A particle for the latent variable $x_{1,t}$ is a sequence of the form

$$x_{1,0:t}^{(k)} = \left(x_{1,0}^{(k)}, x_{1,1}^{(k)}, \ldots, x_{1,t}^{(k)} \right),$$

where $k = 1, 2, \ldots, N$ indexes the particles. They are i.i.d. draws from the conditional density

$$p(x_{1,0:t} \,|\, a_{1,0:t-1}, x_{2,0:t-1}, \theta).$$

An algorithm using particles to estimate the partial information likelihood ℓ' unbiasedly is as follows:

- Initialize by putting $\hat{C}_0 = 1$ and letting $\left\{x_{1,0:0}^{(k)}\right\}_{k=1}^{N}$ be draws $x_{1,0}^{(k)}$ from the marginal stationary density (9).

If one has particles $\left\{x_{1,0:t-1}^{(k)}\right\}_{k=1}^{N}$ in hand, one advances as follows:

- Draw $\tilde{x}_{1t}^{(k)}$ from the transition density

$$p(x_{1t} \,|\, a_{t-1}, x_{1,t-1}^{(k)}, x_{2,t-1}, \theta).$$

- Compute

$$\begin{aligned}
\bar{v}_t^{(k)} &= \frac{p\Big(a_t \,|\, \tilde{x}_{1,t}^{(k)}, x_{2,t}, \theta\Big)\, p\Big(\tilde{x}_{1,t}^{(k)}, x_{2,t} \,|\, a_{t-1}, x_{1,t-1}^{(k)}, x_{2,t-1}, \theta\Big)}{p\Big(\tilde{x}_{1,t}^{(k)} \,|\, a_{t-1}, x_{1,t-1}^{(k)}, x_{2,t-1}, \theta\Big)} \\
\hat{C}_t &= \frac{1}{N}\sum_{k=1}^{N} \bar{v}_t^{(k)} \\
\tilde{x}_{1,0:t}^{(k)} &= \Big(x_{1,0:t-1}^{(k)}, \tilde{x}_{1,t}^{(k)}\Big) \\
\hat{w}_t^{(k)} &= \frac{\bar{v}_t^{(k)}}{\sum_{k=1}^{N} \bar{v}_t^{(k)}}
\end{aligned}$$

- Draw $\left\{x_{1,0:t}^{(k)}\right\}_{k=1}^{N}$ i.i.d. from the discrete distribution $P\Big(x_{1,0:t} = \tilde{x}_{1,0:t}^{(k)}\Big) = \hat{w}_t^{(k)}$.

Repeat until $t = T$.

- An unbiased estimate of the partial information likelihood is $\hat{\ell}' = \prod_{t=0}^{T} \hat{C}_t$.

We next verify the algorithm.

4.3 Verification

In the Bayesian paradigm, the parameter vector θ is manipulated as if it were distributed jointly with the data even though one might actually regard θ as fixed. Thus, formally, we may let $\left\{\{a_t, x_t\}_{t=-T_0}^{\infty}\}, \theta\right\}$ be defined on a common probability space $\mathcal{P}$ over the support of a_t, x_t, and θ. Let $\mathcal{F}_t = \sigma\left\{\{a_s, x_{2s}\}_{s=-T_0}^{t}, \theta\right\}$. Let $P(\cdot|\mathcal{F}_t)$ denote the conditional probability measure given $\mathcal{F}_t$ over $\mathcal{P}$. Let $\mathcal{E}(\cdot|\mathcal{F}_t)$ denote conditional expectation given $\mathcal{F}_t$.

The elements of a_t and x_t may have either atomless or discrete distributions. No generality is lost by presuming that the discrete elements are positive integers. Let z denote a generic

vector some of whose coordinates are real numbers and the others positive integers. Let $\lambda(z)$ denote a product measure whose marginals are either counting measure on the positive integers or Lebesgue ordered as is appropriate to define an integral of the form $\int g(z)\, d\lambda(z)$. The densities defined in Subsection 4.1 are with respect to a measure of the form $\lambda(z)$. We will often write, e.g., $\int g(x_t)\, dP(x_t|a_{t-1}, x_{t-1}, \theta)$ to mean integration of $g(x_t)p(x_t|a_{t-1}, x_{t-1}, \theta)$ with respect to the product measure $\lambda(x_t)$.

Particle filters are advanced by drawing independent uniform random variables $u_{t+1}^{(1)}, \ldots, u_{t+1}^{(N)}$ and then evaluating a function of the form $X_{1,t+1}^{(k)}(u_{t+1}^{(k)}, a_{0:t}, \tilde{x}_{1,0:t}^{(k)}, x_{2,0:t})$ and putting $\tilde{x}_{1,t+1}^{(k)} = X_{1,t+1}^{(k)}(u_{t+1}^{(k)}, a_{0:t}, \tilde{x}_{1,0:t}^{(k)}, x_{2,0:t})$ for $k = 1, \ldots, N$. Denote expectation with respect to $u_{t+1}^{(1)}, \ldots, u_{t+1}^{(N)}$ by $\tilde{\mathcal{E}}_{1,t+1}$. Due to the recursive nature of a particle filter, $\tilde{x}_{1,0:t}^{(k)}$ will itself depend on uniform draws. Denote expectation with respect to the uniform draws involved in $\tilde{x}_{1,0:t}^{(k)}$ by $\tilde{\mathcal{E}}_{1,0:t}$. For now we are excluding resampling. In Subsection 4.4 we show that the extension of the results of this subsection to the case where resampling is employed amounts to simply increasing the scope of $\tilde{\mathcal{E}}_{1,0:t}$ to include not only the uniform draws that advance the filter but also the uniform draws involved in resampling.

Under these conventions, note that what is random in an expression such as $\mathcal{E}\left[g(\tilde{x}_{1,0:t}^{(k)}) \,|\, \mathcal{F}_t\right]$ are the uniform draws involved in $\tilde{x}_{1,0:t}^{(k)}$. All else is fixed due to conditioning on $\mathcal{F}_t$.

In this subsection we show that estimate of the likelihood given in Subsection 4.2 is unbiased for either the full information likelihood or a partial information likelihood, depending on how initial conditions are handled.

THEOREM 1 Under Assumption 1, if the particles $\tilde{x}_{1,0:t}^{(k)}$ and weights $\tilde{w}_t^{(k)}$, $k = 1, \ldots, N$, satisfy

$$\int g(x_{1,0:t})\, dP(x_{1,0:t}|\mathcal{F}_t) = \tilde{\mathcal{E}}_{1,0:t}\left\{\mathcal{E}\left[\sum_{k=1}^{N} \tilde{w}_t^{(k)} g(\tilde{x}_{1,0:t}^{(k)}) \,|\, \mathcal{F}_t\right]\right\} \tag{12}$$

for every integrable function $g(x_{1,0:t})$, then draws $\tilde{x}_{1,t+1}^{(k)}$ from $p(x_{1,t+1}|\tilde{x}_{1,0:t}^{(k)}, \mathcal{F}_t)$ and weights

$$\tilde{w}_{t+1}^{(k)} = \frac{\bar{v}_{t+1}^{(k)}}{C_{t+1}}\, \tilde{w}_t^{(k)} \tag{13}$$

satisfy

$$\int g(x_{1,0:t}, x_{1,t+1})\, dP(x_{1,0:t}, x_{1,t+1}|\mathcal{F}_{t+1})$$

$$= \tilde{\mathcal{E}}_{1,t+1}\tilde{\mathcal{E}}_{1,0:t}\left\{\mathcal{E}\left[\sum_{k=1}^{N}\tilde{w}_{t+1}^{(k)}\, g(\tilde{x}_{1,0:t}^{(k)}, \tilde{x}_{1,t+1}^{(k)})\,|\,\mathcal{F}_{t+1}\right]\right\}, \tag{14}$$

where

$$\bar{v}_{t+1}^{(k)} = \frac{p\Big(a_{t+1}\,|\,\tilde{x}_{1,t+1}^{(k)}, x_{2,t+1}, \theta\Big)\, p\Big(\tilde{x}_{1,t+1}^{(k)}, x_{2,t+1}\,|\,a_t, \tilde{x}_{1,t}^{(k)}, x_{2,t}, \theta\Big)}{p\Big(\tilde{x}_{1,t+1}^{(k)}\,|\,a_t, \tilde{x}_{1,t}^{(k)}, x_{2,t}, \theta\Big)} \tag{15}$$

and

$$C_{t+1} = p(a_{t+1}, x_{2,t+1}|\mathcal{F}_t). \tag{16}$$

Proof We first show the result for the weights

$$\tilde{w}_{t+1}^{(k)} = \frac{p(a_{t+1}, x_{2,t+1}|\tilde{x}_{1,0:t}^{(k)}, \tilde{x}_{1,t+1}^{(k)}, \mathcal{F}_t)}{p(a_{t+1}, x_{2,t+1}|\mathcal{F}_t)}\, \tilde{w}_t^{(k)}, \tag{17}$$

which we write as $w_{t+1}^{(k)}$ below when evaluated at $x_{1,t+1}$ instead of $\tilde{x}_{1,t+1}^{(k)}$, and then show that (17) and (13) are equivalent expressions for $\tilde{w}_{t+1}^{(k)}$.

Bayes theorem states that

$$p(x_{1,0:t}, x_{1,t+1}|a_{t+1}, x_{2,t+1}, \mathcal{F}_t) = \frac{p(a_{t+1}, x_{2,t+1}, x_{1,0:t}, x_{1,t+1}|\mathcal{F}_t)}{p(a_{t+1}, x_{2,t+1}|\mathcal{F}_t)}. \tag{18}$$

Note that

$$p(x_{1,0:t}, x_{1,t+1}|a_{t+1}, x_{2,t+1}, \mathcal{F}_t) = p(x_{1,0:t}, x_{1,t+1}|\mathcal{F}_{t+1}) \tag{19}$$

and that

$$p(a_{t+1}, x_{2,t+1}, x_{1,0:t}, x_{1,t+1}|\mathcal{F}_t)$$

$$= p(a_{t+1}, x_{2,t+1}|x_{1,0:t}, x_{1,t+1}, \mathcal{F}_t)p(x_{1,t+1}|x_{1,0:t}, \mathcal{F}_t)p(x_{1,0:t}|\mathcal{F}_t). \tag{20}$$

Then

$$\int g(x_{1,0:t}, x_{1,t+1})\, dP(x_{1,0:t}, x_{1,t+1}|\mathcal{F}_{t+1})$$

$$= \iint g(x_{1,0:t}, x_{1,t+1})\frac{p(a_{t+1}, x_{2,t+1}|x_{1,0:t}, x_{1,t+1}, \mathcal{F}_t)}{p(a_{t+1}, x_{2,t+1}|\mathcal{F}_t)}p(x_{1,t+1}|x_{1,0:t}, \mathcal{F}_t)$$

$$\times d\lambda(x_{1,t+1}) dP(x_{1,0:t}|\mathcal{F}_t) \tag{21}$$

$$= \tilde{\mathcal{E}}_{1,0:t} \int \mathcal{E}\left[\sum_{k=1}^{N} g(\tilde{x}_{1,0:t}^{(k)}, x_{1,t+1})\, w_{t+1}^{(k)}\, p(x_{1,t+1}|\tilde{x}_{1,0:t}^{(k)}, \mathcal{F}_t) \mid \mathcal{F}_t\right] d\lambda(x_{1,t+1}) \tag{22}$$

$$= \tilde{\mathcal{E}}_{1,t+1}\tilde{\mathcal{E}}_{1,0:t}\, \mathcal{E}\left[\sum_{k=1}^{N} g(\tilde{x}_{1,0:t}^{(k)}, \tilde{x}_{1,t+1}^{(k)})\, \tilde{w}_{t+1}^{(k)} \mid \mathcal{F}_{t+1}\right] \tag{23}$$

where (21) is due to (18) after substituting (19) and (20), (22) is due to (12) and (17), and (23) is due to the fact that $\tilde{x}_{1,t+1}^{(k)}$ is a draw from $p(x_{1,t+1}|\tilde{x}_{1,0:t}^{(k)}, \mathcal{F}_t)$. This proves the result for the weights (17).

We show that (17) and (13) are equivalent, by reexpressing the numerator of (17) in terms of the primitives (6), (7), and (8) as follows.

$$\begin{aligned}
& p(a_{t+1}, x_{2,t+1}|x_{1,0:t}, x_{1,t+1}, \mathcal{F}_t) \\
&= p(a_{t+1}, x_{2,t+1}|x_{1,0:t}, x_{1,t+1}, x_{2,0:t}, a_{0:t}, \theta) \\
&= \frac{p(a_{t+1}, x_{t+1}, a_{0:t}, x_{0:t}, \theta)}{\int p(a_{t+1}, x_{t+1}, a_{0:t}, x_{0:t}, \theta)\, d\lambda(a_{t+1}, x_{2,t+1})} \\
&= \frac{p(a_{t+1}|x_{t+1}, a_{0:t}, x_{0:t}, \theta) p(x_{t+1}|a_{0:t}, x_{0:t}, \theta) p(a_{0:t}, x_{0:t}, \theta)}{\int p(a_{t+1}|x_{t+1}, a_{0:t}, x_{0:t}, \theta) p(x_{t+1}|a_{0:t}, x_{0:t}, \theta) d\lambda(a_{t+1}, x_{2,t+1})\, p(a_{0:t}, x_{0:t}, \theta)} \\
&= \frac{p(a_{t+1}|x_{t+1}, \theta)\, p(x_{t+1}|a_t, x_t, \theta)}{\int p(a_{t+1}|x_{t+1}, \theta)\, p(x_{t+1}|a_t, x_t, \theta)\, d\lambda(a_{t+1}, x_{2,t+1})} \\
&= \frac{p(a_{t+1}|x_{t+1}, \theta)\, p(x_{t+1}|a_t, x_t, \theta)}{\int p(x_{t+1}|a_t, x_t, \theta)\, d\lambda(x_{2,t+1})} \\
&= \frac{p(a_{t+1}|x_{t+1}, \theta)\, p(x_{t+1}|a_t, x_t, \theta)}{p(x_{1,t+1}|a_t, x_t, \theta)}
\end{aligned} \tag{24}$$

□

REMARK 1 An implication of (14) is that N affects the second moment of

$$\mathcal{E}\left[\sum_{k=1}^{N} \tilde{w}_{t+1}^{(k)}\, g(\tilde{x}_{1,0:t}^{(k)}, \tilde{x}_{1,t+1}^{(k)}) \mid \mathcal{F}_{t+1}\right]$$

but not the first moment. □

REMARK 2 We may assume without loss of generality that all N of the weights $\tilde{w}_{t+1}^{(k)}$ given by (13) are positive because one can, e.g., discard all particles with zero weight then,

as often as necessary to get N particles, replicate the particle with the largest weight and divide that weight evenly between that particle and its replicates. □

COROLLARY 1 Under Assumption 1, if one starts the recursion of Theorem 1 with draws from the marginal stationary density (9) and weights $\tilde{w}_0^{(k)} = 1/N$, then

$$\hat{\ell}' = \left(\sum_{k=1}^{N} \bar{v}_T^{(k)} \frac{\tilde{w}_{T-1}^{(k)}}{\sum_{k=1}^{N} \tilde{w}_{T-1}^{(k)}}\right)\left(\sum_{k=1}^{N} \bar{v}_{T-1}^{(k)} \frac{\tilde{w}_{T-2}^{(k)}}{\sum_{k=1}^{N} \tilde{w}_{T-2}^{(k)}}\right)\cdots\left(\sum_{k=1}^{N} \bar{v}_1^{(k)} \frac{\tilde{w}_0^{(k)}}{\sum_{k=1}^{N} \tilde{w}_0^{(k)}}\right)\left(\sum_{k=1}^{N} \tilde{w}_0^{(k)}\right) \quad (25)$$

is an unbiased estimator of ℓ'.

Proof Set $g(x_{1,0:t}, u) \equiv 1$ in Theorem 1 whence $1 = \tilde{\mathcal{E}}_{1,0:T}\left\{\mathcal{E}\left[\sum_{k=1}^{N} \tilde{w}_t^{(k)} \,|\, \mathcal{F}_T\right]\right\}$. Write

$$\begin{aligned}\sum_{k=1}^{N} \tilde{w}_T^{(k)} &= \frac{1}{C_T}\left(\frac{\sum_{k=1}^{N} \bar{v}_T^{(k)} \tilde{w}_{T-1}^{(k)}}{\sum_{k=1}^{N} \bar{v}_{T-1}^{(k)} \tilde{w}_{T-2}^{(k)}}\right)\left(\frac{\sum_{k=1}^{N} \bar{v}_{T-1}^{(k)} \tilde{w}_{T-2}^{(k)}}{\sum_{k=1}^{N} \bar{v}_{T-2}^{(k)} \tilde{w}_{T-3}^{(k)}}\right)\cdots\left(\frac{\sum_{k=1}^{N} \bar{v}_1^{(k)} \tilde{w}_0^{(k)}}{\sum_{k=1}^{N} \tilde{w}_0^{(k)}}\right)\left(\sum_{k=1}^{N} \tilde{w}_0^{(k)}\right) \\ &= \frac{1}{\ell'}\left(\sum_{k=1}^{N} \bar{v}_T^{(k)} \frac{\tilde{w}_{T-1}^{(k)}}{\sum_{k=1}^{N} \tilde{w}_{T-1}^{(k)}}\right)\left(\sum_{k=1}^{N} \bar{v}_{T-1}^{(k)} \frac{\tilde{w}_{T-2}^{(k)}}{\sum_{k=1}^{N} \tilde{w}_{T-2}^{(k)}}\right)\cdots\left(\sum_{k=1}^{N} \bar{v}_1^{(k)} \frac{\tilde{w}_0^{(k)}}{\sum_{k=1}^{N} \tilde{w}_0^{(k)}}\right)\left(\sum_{k=1}^{N} \tilde{w}_0^{(k)}\right)\end{aligned}$$

The result follows. □

REMARK 3 Note that the terms $\frac{\tilde{w}_t^{(k)}}{\sum_{k=1}^{N} \tilde{w}_t^{(k)}}$ in (25) are invariant to C_t so that $\hat{\ell}'$ can be computed without knowledge of the C_t. Thus, one can arbitrarily set $C_t \equiv 1$ in (13) for weights that are to be used in (25) as is done in the algorithms of Subsections 4.2 and 4.6. □

REMARK 4 If one has an expression for $p(a_0, x_{2,0} \,|\, x_{1,0}, \theta)$, then

$$\hat{\ell} = \hat{\ell}' \frac{1}{N} \sum_{k=1}^{N} p(a_0, x_{2,0} \,|\, \tilde{x}_{1,0}^{(k)}, \theta)$$

is an unbiased estimator of ℓ. □

4.4 Resampling

One often resamples particles such as $\left\{\tilde{x}_{1,0:t}^{(k)}\right\}_{k=1}^{N}$ in order to prevent the variance of the weights $\tilde{w}_t^{(k)}$ from increasing with t. Most common is multinomial resampling where one samples $\left\{\tilde{x}_{1,0:t}^{(k)}\right\}_{k=1}^{N}$ with replacement with probability $\frac{\tilde{w}_t^{(k)}}{\sum_{k=1}^{N} \tilde{w}_t^{(k)}}$. Some particles will get copied and some particles will not survive. The new particles $\left\{x_{1,0:t}^{(k)}\right\}_{k=1}^{N}$ each have weight $\frac{1}{N}$.

Multinomial resampling is usually performed by drawing uniform random numbers $\{u_i\}_{i=1}^N$ on the interval $(0,1)$ and evaluating the quantile function $q(u)$ of the discrete density $p(k) = \frac{\tilde{w}_t^{(k)}}{\sum_{k=1}^N \tilde{w}_t^{(k)}}$ at the u_i to get the resampled particles $x_{1,0:t}^{(i)} = \tilde{x}_{1,0:t}^{(q(u_i))}$. Other resampling schemes seek to improve performance by having one uniform random number in each interval $[(i-1)/N, i/N]$ for $i = 1, \cdots, N$. One approach is stratified resampling where one uniform u is drawn inside each interval. Another is systematic resampling where the same uniform u is placed inside each interval. As with multinomial resampling, the weights for stratified and systematic sampling are $\frac{1}{N}$. In a comparison of stratified and systematic resampling, Douc, Cappé, and Moulines (2005) find that their performance is similar.

THEOREM 2 Theorem 1 and Corollary 1 remain valid if resampling is applied between recursive steps as in the algorithm of Subsection 4.2.

Proof By construction, if a set of particles and weights satisfy condition (12) then so will the particles and weights generated from them by resampling. Because a set of particles and weights satisfy condition (12) at the end of an iterate, the set of particles and weights generated from them by resampling will satisfy (12) when used at the beginning of an iterate. The only formal change to the development of Subsection 4.3 required is that $\tilde{\mathcal{E}}_{1,0:t}$ becomes expectation both with respect to the uniform draws that advance the filter and to the uniform draws involved in resampling. □

REMARK 5 For any resampling scheme that produces equal weights, the conclusion of Corollary 1 becomes

$$\hat{\ell}' = \left(\frac{1}{N}\sum_{k=1}^N \bar{v}_T^{(k)}\right)\left(\frac{1}{N}\sum_{k=1}^N \bar{v}_{T-1}^{(k)}\right)\cdots\left(\frac{1}{N}\sum_{k=1}^N \bar{v}_1^{(k)}\right)$$

is an unbiased estimator of ℓ'. □

4.5 An Alternative Importance Sampler

If computing $p(x_{1,t+1}|a_t, x_t, \theta)$ is costly or drawing from it troublesome, one can substitute an alternative importance sampler. The idea is that one can advance a filter from $(\tilde{\tilde{x}}_t^{(k)}, \tilde{\tilde{w}}_t^{(k)})$ that satisfies (12) to $(\tilde{\tilde{x}}_{t+1}^{(k)}, \tilde{\tilde{w}}_{t+1}^{(k)})$ that satisfies (14) by drawing $\tilde{\tilde{x}}_{t+1}^{(k)}$ from

$$f(x_{1,t+1}|x_{1t}, \mathcal{F}_t) = f\big(x_{1,t+1} \,|\, a_t, \tilde{\tilde{x}}_{1t}^{(k)}, x_{2t}, \theta\big) \tag{26}$$

for $k = 1, \ldots, N$, and setting

$$\tilde{\tilde{w}}_{t+1}^{(k)} = \frac{p\Big(a_{t+1} \,|\tilde{\tilde{x}}_{1,t+1}^{(k)}, x_{2,t+1}, \theta\Big)\, p\Big(\tilde{\tilde{x}}_{1,t+1}^{(k)}, x_{2,t+1} \,|\, a_t, \tilde{\tilde{x}}_{1t}^{(k)}, x_{2t}, \theta\Big)}{C_{t+1}\, f\Big(\tilde{\tilde{x}}_{1,t+1}^{(k)} \,|\, a_t, \tilde{\tilde{x}}_{1t}^{(k)}, x_{2t}, \theta\Big)} \tilde{\tilde{w}}_t^{(k)} \tag{27}$$

as is seen by noting that (21), (22), and (23) can be rewritten as

$$\begin{aligned} \iint g(x_{1,0:t}, x_{1,t+1}) & \frac{p(a_{t+1}, x_{2,t+1}|x_{1,0:t}, x_{1,t+1}, \mathcal{F}_t)}{p(a_{t+1}, x_{2,t+1}|\mathcal{F}_t)} \, \frac{p(x_{1,t+1}|x_{1,0:t}, \mathcal{F}_t)}{f(x_{1,t+1}|x_{1,t}, \mathcal{F}_t)} \, f(x_{1,t+1}|x_{1t}, \mathcal{F}_t) \\ & \times d\lambda(x_{t+1})\, dP(x_{1,0:t}|\mathcal{F}_t) \\ = \; & \tilde{\mathcal{E}}_{1,0:t} \int \mathcal{E}\left[\sum_{k=1}^{N} g(\tilde{\tilde{x}}_{1,0:t}^{(k)}, x_{1,t+1}) \tilde{\tilde{w}}_{t+1}^{(k)} \, f(x_{1,t+1}|\tilde{\tilde{x}}_t^{(k)}, \mathcal{F}_t) \,|\, \mathcal{F}_t\right] d\lambda(x_{1,t+1}) \\ = \; & \tilde{\mathcal{E}}_{1,t+1} \tilde{\mathcal{E}}_{1,0:t}\, \mathcal{E}\left[\sum_{k=1}^{N} g(\tilde{\tilde{x}}_{1,0:t}^{(k)}, \tilde{\tilde{x}}_{t+1}^{(k)})\, \tilde{\tilde{w}}_{t+1}^{(k)} \,|\, \mathcal{F}_{t+1}\right] \end{aligned} \tag{28}$$

due to the cancellation $p(x_{1,t+1}|x_{1,0:t}, \mathcal{F}_t)/p(x_{1,t+1}|a_t, x_t, \theta) = 1$ that occurs after the expression for $p(a_{t+1}, x_{2,t+1}|x_{1,0:t}, x_{1,t+1}, \mathcal{F}_t)$ given by (24) is substituted in (28).

The equations that replace (13) and (15) when an alternative importance sampler is used are

$$\tilde{\tilde{w}}_{t+1}^{(k)} = \frac{\tilde{\tilde{v}}_{t+1}^{(k)}}{C_{t+1}} \, \tilde{\tilde{w}}_t^{(k)} \tag{13$'$}$$

and

$$\tilde{\tilde{v}}_{t+1}^{(k)} = \frac{p\Big(a_{t+1} \,|\tilde{\tilde{x}}_{1,t+1}^{(k)}, x_{2,t+1}, \theta\Big)\, p\Big(\tilde{\tilde{x}}_{1,t+1}^{(k)}, x_{2,t+1} \,|\, a_t, \tilde{\tilde{x}}_{1,t}^{(k)}, x_{2,t}, \theta\Big)}{f\Big(\tilde{\tilde{x}}_{1,t+1}^{(k)} \,|\, a_t, \tilde{\tilde{x}}_{1,t}^{(k)}, x_{2,t}, \theta\Big)}. \tag{15$'$}$$

The requisite regularity condition is the following:

ASSUMPTION 2

$$g(x_{1,0:t}, x_{1,t+1}) \, \frac{p\left(a_{t+1} \,|\, x_{1,t+1}, x_{2,t+1}, \theta\right)\, p\left(x_{1,t+1}, x_{2,t+1} \,|\, a_t, x_{1,t}, x_{2,t}, \theta\right)}{f(x_{1,t+1} \,|\, a_t, x_{1,t}, x_{2,t}, \theta)}$$

is integrable with respect to $f(x_{1,t+1} \,|\, a_t, x_{1,t}, x_{2,t}, \theta)$, the support of which contains the support of $p\left(x_{1,t+1} \,|\, a_t, x_{1,t}, x_{2,t}, \theta\right)$.

Another reason to consider an alternative importance sampler is to improve efficiency. Pitt and Shephard (1999) suggest some adaptive importance samplers that one might consider. In addition to Pitt and Shephard's (1999) suggestions, one can use the notion

of reprojection (Gallant and Tauchen (1998)) to construct an adaptive density for (26) as follows. The model can be simulated. Therefore, for given θ^* a large simulation of $(a_t, x_{1t}, x_{2t}, a_{t+1}, x_{1,t+1}, x_{2,t+1})$ can be generated. Using multivariate regression one can determine the location $\mu(v)$ of $x_{1,t+1}$ as a linear function of

$$v = (a_t, x_{1t}, x_{2t}, a_{t+1}, x_{2,t+1}) \tag{29}$$

and the conditional variance Σ. The simulation can be taken so large that $\mu(v)$ and Σ can be regarded as population quantities. We put

$$h(x_{1,t+1}|x_{1t}, \mathcal{F}_{t+1}) = n(x_{1,t+1}|\mu(v), \Sigma), \tag{30}$$

where $n(\cdot|\mu, \Sigma)$ denotes the multivariate normal density and use (30) in place of (26), which is a slight abuse of notation because the argument lists are different. Substituting the multivariate Student-t density on five degrees of freedom with the same location and scale had little effect on results other than increase run times.

The mean or the mode of an informative prior is a reasonable choice of θ^* for the simulation that determines $\mu(v)$ and Σ. If the prior is flat, one can start with a guess, run a preliminary chain, and use the mean of the preliminary chain for θ^*.

4.6 Computing the Likelihood

In this subsection we refine the algorithm of Subsection 4.2 to incorporate the developments of the foregoing subsections and to to discuss how seeds must be handled in order for the MCMC chain that uses an unbiased estimate of the likelihood to target the intended posterior. Handling the seed correctly is of critical importance; on this see Flury and Shephard (2010) and Pitt, Silva, Giordani, and Kohn (2012).

Actually, the reasoning is straightforward: The seed s is a random variable distributed over a finite set of integers $\mathcal{S}$; its density is $u(s)$. Let $\mathcal{L}(\theta)$ be the likelihood of a game and $\mathcal{L}(\theta, s)$ its unbiased approximation in the sense that $\mathcal{L}(\theta) = \sum_{s\in\mathcal{S}} \mathcal{L}(\theta, s)u(s)$. If one draws (θ, s) from the joint density $\mathcal{L}(\theta, s)u(s)p(\theta)$, where $p(\theta)$ is the prior, and discards s then θ that remains is a draw from the marginal $\mathcal{L}(\theta)p(\theta)$. The crucial observation is that the ratio

$$\frac{\mathcal{L}(\theta^*, s^*)u(s^*)p(\theta^*)\lambda(\theta^*, \theta^\#)u(s^\#)}{\mathcal{L}(\theta^\#, s^\#)u(s^\#)p(\theta^\#)\lambda(\theta^\#, \theta^*)u(s^*)}$$

that determines acceptance in an MCMC algorithm becomes

$$\frac{\mathcal{L}(\theta^*, s^*)p(\theta^*)\lambda(\theta^*, \theta^\#)}{\mathcal{L}(\theta^\#, s^\#)p(\theta^\#)\lambda(\theta^\#, \theta^*)}$$

as in the algorithm described in Section 5, where $(\theta^\#, s^\#)$ is the end of the chain, (θ^*, s^*) is the proposed extension of the chain, and $\lambda(\theta^\#, \theta^*)u(s^*)$ is the proposal density. One can see that if $s^\#$ and s^* were not independent draws from $u(s)$ then the MCMC algorithm would not target $\mathcal{L}(\theta, s)u(s)p(\theta)$. One would not have to devote such attention to the seed were it not that one is usually advised to use the same seed every time to compute $\mathcal{L}(\theta, s)$ in order to reduce Monte Carlo jitter. Here the advice is exactly the opposite.

Random number algorithms have different designs varying from those that generate an array of random numbers intended to supply all that are ever needed with one call, through intermediate designs, to a simple functional recursion that may be conceptualized as follows:

> A draw from a density $f(v)$ is obtained by drawing a seed s from a uniform density $u(s)$ defined over a finite set of integers and executing an algorithm that evaluates a function $V(s)$ and returns $v' = V(s)$ and s' such that v' has density $f(v)$, s' has density $u(s)$, and s' is independent of s. The next draw from the same or a different density uses s' to return a draw v'' from that density and another new seed s'', and so on.

We may, without loss of generality, assume that whatever algorithm is used, a wrapper has been coded so that becomes a simple functional recursion.

The algorithm that we describe next may have a sequence of draws within it but viewed as a whole it is a functional recursion: One specifies θ and provides a random draw s from $u(s)$. The algorithm evaluates a function $\mathcal{L}(\theta, s)$ and returns $\hat{\ell}' = \mathcal{L}(\theta, s)$ and a draw s' from $u(s)$ that is independent of s.

Given seed s and parameter θ, the algorithm for evaluating $\mathcal{L}(\theta, s)$ follows. All draws use the seed returned by the previous draw; there are no fixed seeds anywhere within the algorithm.

1. For $t = 0$

   (a) Start N particles by drawing $\tilde{x}_{1,0}^{(k)}$ from $p(x_{1,0} \,|\, \theta)$ using s as the initial seed.

(b) If $p(a_t, x_{2t} \,|\, x_{1,t}, \theta)$ is available, compute $\hat{C}_0 = \frac{1}{N}\sum_{k=1}^{N} p(a_0, x_{2,0} \,|\, \tilde{x}_{1,0}^{(k)}, \theta)$ otherwise put $\hat{C}_0 = 1$.

(c) Set $x_{1,0:0}^{(k)} = \tilde{x}_{1,0}^{(k)}$ and $x_{1,0}^{(k)} = \tilde{x}_{1,0}^{(k)}$

2. For $t = 1, \ldots, T$

(a) For each particle, draw $\tilde{x}_{1t}^{(k)}$ from the transition density

$$p(x_{1t} \,|\, a_{t-1}, x_{1,t-1}^{(k)}, x_{2,t-1}, \theta). \tag{31}$$

(b) Compute

$$\begin{aligned} \bar{v}_t^{(k)} &= \frac{p\Big(a_t \,|\, \tilde{x}_{1,t}^{(k)}, x_{2,t}, \theta\Big)\, p\Big(\tilde{x}_{1,t}^{(k)}, x_{2,t} \,|\, a_{t-1}, x_{1,t-1}^{(k)}, x_{2,t-1}, \theta\Big)}{p\Big(\tilde{x}_{1,t}^{(k)} \,|\, a_{t-1}, x_{1,t-1}^{(k)}, x_{2,t-1}, \theta\Big)} \\ \hat{C}_t &= \frac{1}{N}\sum_{k=1}^{N} \bar{v}_t^{(k)} \end{aligned} \tag{32}$$

(Note that the draw pair is $(x_{1,t-1}^{(k)}, \tilde{x}_{1,t}^{(k)})$ and the weight is $\bar{v}_t^{(k)} \frac{1}{N}$.)

(c) Set

$$\tilde{x}_{1,0:t}^{(k)} = \Big(x_{1,0:t-1}^{(k)}, \tilde{x}_{1,t}^{(k)}\Big).$$

(d) Compute the normalized weights

$$\hat{w}_t^{(k)} = \frac{\bar{v}_t^{(k)}}{\sum_{k=1}^{N} \bar{v}_t^{(k)}}$$

(e) For $k = 1, \ldots, N$ draw $x_{1,0:t}^{(k)}$ by sampling with replacement from the set $\{\tilde{x}_{1,0:t}^{(k)}\}$ according to the weights $\{\hat{w}_t^{(k)}\}$.

(Note the convention: Particles with unequal weights are denoted by $\{\tilde{x}_{0:t}^{(k)}\}$. After resampling the particles are denoted by $\{x_{0:t}^{(k)}\}$.)

(f) For use in (31) and (32) at the next iterate, set $x_t^{(k)}$ to the last element of $x_{1,0:t}^{(k)}$.

3. Done

(a) An unbiased estimate of the likelihood is

$$\hat{\ell}' = \prod_{t=0}^{T} \hat{C}_t \tag{33}$$

(b) s' is the last seed returned in Step 2e.

Systematic or stratified sampling can be used at step 2e instead of multinomial resampling. To use the alternative importance sampler of Section 4.5, replace (31) with (26) or (30) and replace (32) with

$$\bar{\bar{v}}_t^{(k)} = \frac{p\Big(a_t \,|\, \tilde{x}_{1,t}^{(k)}, x_{2,t}, \theta\Big)\, p\Big(\tilde{x}_{1,t}^{(k)}, x_{2,t} \,|\, a_{t-1}, x_{1,t-1}^{(k)}, x_{2,t-1}, \theta\Big)}{f\Big(x_{1,t}^{(k)} \,|\, a_{t-1}, x_{1,t-1}^{(k)}, x_{2,t-1}, \theta\Big)}. \tag{34}$$

5 Computing the Posterior

In this section we let ℓ mean either ℓ of (10) or ℓ' of (11) as appropriate.

Metropolis algorithm is an iterative scheme that generates a Markov chain whose stationary distribution is the posterior of θ. To implement it, we require the particle filter algorithm for drawing (ℓ, s) described in Section 4.6, a prior $\gamma(\theta)$, and a transition density in θ called the proposal density. For a given $\theta^{\#}$, a proposal density $\lambda(\theta^{\#}, \theta^{*})$ defines a distribution of potential new values θ^{*}. We use the move-one-at-a-time, random-walk, proposal density that is built in to the public domain software that we use: `http://www.aronaldg.org/webfiles/emm/`.The algorithm for the Markov chain is as follows.

Start the chain at a reasonable value for θ and write to memory a draw $s^{\#\#}$ from the uniform density on a finite set of integers. Given a current $\theta^{\#}$ we obtain the next $\theta^{\#\#}$ as follows:

1. Draw θ^{*} according to $\lambda(\theta^{\#}, \theta^{*})$.
2. Set s^{*} to $s^{\#\#}$ retrieved from memory.
3. Compute ℓ^{*} corresponding to (θ^{*}, s^{*}) using the particle filter in Section 4.6 and write to memory the $s^{\#\#}$ returned by the particle filter.
4. Compute $\alpha = \min\left(1, \frac{\ell^{*}\,\gamma(\theta^{*})\,\lambda(\theta^{*}, \theta^{\#})}{\ell^{\#}\,\gamma(\theta^{\#})\,\lambda(\theta^{\#}, \theta^{*})}\right)$.
5. With probability α, set $\theta^{\#\#} = \theta^{*}$, otherwise set $\theta^{\#\#} = \theta^{\#}$.
6. Return to 1.

The choice for the parameter N of the particle filter in Section 4.6 influences the rejection rate of the MCMC chain. If N is too small then $\mathcal{L}(\theta, s) = \ell^{\#}$ given by (33) will be a jittery estimator of $\mathcal{L}(\theta)$ which will increase the chance that the chain gets stuck. Pitt, Silva, Giordani, and Kohn (2012) show that what is relevant is the variance

$$\text{Var}\left\{\log \mathcal{L}(\theta, s)\right\} = \int \left[\log \mathcal{L}(\theta, s) - \int \log \mathcal{L}(\theta, s)\, ds\right]^2 ds, \tag{35}$$

which can be computed from draws of $\ell^{\#}$ obtained by putting the filter in a loop. It is interesting that for an entry game such as in Subsection 6.1, the classification error rate can be so small that one is almost matching 0's and 1's and using the particle filter to solve backwards for $\{x_{1t}\}$ that will allow the match. The consequence is that N can be quite small. For our example, Pitt et. al.'s charts suggest that $N = 300$ will suffice. We actually use $N = 512$. In practice charts are unnecessary because one can easily determine N empirically by increasing it until the chain is no longer sticky.

One could set forth regularity conditions such that $\lim_{N\to\infty} \sup_\theta \text{Var}\left\{\log \mathcal{L}(\theta, s)\right\} = 0$. They will be stringent: see Andrieu, Douced, and Holenstein (2010). One can argue that there is no point to verifying that variance declines with N because one must still determine the requisite N empirically. If an acceptable N is found, it does not matter if variance declines with N or not. If an affordable N cannot be found, a proof that variance declines with N does not help except to provide support for a request for more computing resources.

6 Examples

6.1 A Game with a Small Number of Players

In this section we describe an dynamic oligopolistic entry game. It is a modest variation on the game presented in Gallant, Hong, and Khwaja (2017), and the description in this section is based on and borrows from that paper. We present it here for completeness of exposition. However, for additional details regarding the motivation and rationale for the specification, computation of the model and its estimation, the reader is referred to that paper. Having described the game, we then test our proposed methodology with several simulation experiments.

At each market entry opportunity t the actions available to profit maximizing firm i are whether to choose to enter the market, $\tilde{E}_{it} = 1$, or decide to not enter $\tilde{E}_{it} = 0$. In what follows, a time period uniquely identifies a market opening therefore we interchangeably use t to denote a time period or a market entry opportunity associated with it. Firms can not always achieve entry decision profiles $\tilde{E}_t$. Rather, firms are aware that the realized entry profiles E_t follow a conditional distribution $p(E_t|\tilde{E}_t)$ given $\tilde{E}_t$. We use the following specification for $p(E_t|\tilde{E}_t)$

$$p(E_t\,|\tilde{E}_t, \theta) = \prod_{i=1}^{I} (p_e)^{\delta(E_{it}=\tilde{E}_{it})}(1-p_e)^{1-\delta(E_{it}=\tilde{E}_{it})}, \tag{36}$$

where $0 < p_e < 1$ and $\delta(a = b) = 1$ if $a = b$ and 0 if not. The intended outcome $\tilde{E}_{it}$ is not observed by us. Instead E_{it} is observed, which is a Bernoulli random variable taking value $\tilde{E}_{it}$ with probability p_e and value $1 - \tilde{E}_{it}$ with probability $q_e = 1 - p_e$. The number of entrants in market t is

$$Q_t = \sum_{i=1}^{I} E_{it} \tag{37}$$

To illustrate, consider the generic pharmaceutical drug market. The entry decision of firm i is $\tilde{E}_{it} = 1$ if the firm submits an application to the Federal Drug Administration (FDA) and $E_{it} = 1$ if approved. The FDA reveals application approvals E_{it}; submissions $\tilde{E}_{it}$ are not revealed. Each application carries a small probability of being rejected by the FDA. Firms collectively decide on the equilibrium $\tilde{E}_t$. We observe the ex post realization of E_t.

The dynamics in the model come mainly through the evolution of costs based on past firm actions. The evolution of current total cost Z_{it} for firm i is determined by past entry decisions and random shocks. In the model estimated in Gallant, Hong, and Khwaja (2017), current entry can decrease the cost of an entry next period by, e.g., through learning by doing. However, in contrast in the simulations presented later in Section 6, the evolution of total costs will be configured to represent a situation where past entry constrains capacity and raises current total costs. The cost state variable evolves according to the actual outcome E_t rather than the intended outcome $\tilde{E}_t$. All firms know each others' costs and hence this is a game of complete information. We follow the convention that a lower case quantity denotes the logarithm of an upper case quantity, e.g., $Z_{it} = \log(z_{it})$, with the exception that

for the outcome both E and e denote variables that take the value zero or one. Log cost is the sum of two components

$$z_{i,t} = z_{u,i,t} + z_{k,i,t}. \tag{38}$$

We assume that $z_{u,i,t}$ cannot be observed by us and that $z_{k,i,t}$ can be observed. The first component follows a stationary autoregressive process of order one; the second accumulates the consequences of past entry outcomes:

$$z_{u,i,t} = \mu_z + \rho_z \left(z_{u,i,t-1} - \mu_z \right) + \sigma_z \epsilon_{it} \tag{39}$$

$$\begin{aligned} z_{k,i,t} &= \rho_e \, c_{k,i,t-1} + \kappa_e E_{i,t-1} \\ &= \sum_{j=0}^{\infty} \rho_e^j \kappa_e E_{i,t-j-1}. \end{aligned} \tag{40}$$

In the above, ϵ_{it} is a normally distributed shock with mean zero and unit variance, σ_z is a scale parameter, κ_e is the effect of entry in market $t-1$ on cost at market t, μ_z is the unconditional mean of the unobservable portion of log cost; ρ_z and ρ_e are autoregressive parameters that determine persistence. All firms are ex ante identical,[7] with heterogeneity arising endogenously in the model depending on the effects of past actions on the total costs of the firms.

The total size of the market t or the lump sum total revenue in a market is denoted by S_t. It evolves according to the following process,

$$s_t = \mu_s + \sigma_s \epsilon_{I+1,t} \, , \tag{41}$$

where $s_t = \log(S_t)$, and the $\epsilon_{I+1,t}$ are normally and independently distributed with mean zero and unit variance. In (41), μ_s represents the long run average total revenue for all firms across all market opportunities, while σ_s is a scale parameter. In this example we assume revenue to be exogenous, however, in Section 7 we provide an example where revenue may be endogenously determined.

At time t a firm's present discounted value of profits is given by

$$\sum_{j=0}^{\infty} \beta^j E_{i,t+j} \left(S_{t+j}/Q_{t+j} - Z_{i,t+j} \right), \tag{42}$$

[7] One could allow the initial condition z_{i0} to vary by firm if desired.

where the discount factor is β. Each firm maximizes the present discounted value of its profit in each time period t conditioning on the equilibrium strategy profiles of its rivals.

For firm i's dynamic problem at time t, the Bellman equation for the choice specific value function $V_i(\tilde{e}_{i,t}, \tilde{e}_{-i,t}, Z_{i,t}, Z_{-i,t}, S_t)$ is,

$$
\begin{aligned}
&V_i(\tilde{e}_{i,t}, \tilde{e}_{-i,t}, Z_{i,t}, Z_{-i,t}, S_t) \\
&= \sum_{m_1=0}^{1} p_E^{\delta(m_1=\tilde{e}_{1t})}(q_E)^{1-\delta(m_1=\tilde{e}_{1t})} \cdots \sum_{m_I=0}^{1} p_E^{\delta(m_I=\tilde{e}_{It})}(q_E)^{1-\delta(m_I=\tilde{e}_{It})} \Bigg\{ m_i \left(\frac{S}{\sum_{j=1}^{I} m_j} - Z_{it} \right) \\
&\quad + \beta \, \mathcal{E}\Big[V_i(\tilde{E}_{i,t+1}, \tilde{E}_{-i,t+1}, Z_{i,t+1}, Z_{-i,t+1}, S_{t+1}) \,|\, M_{i,t}, M_{-i,t}, Z_{i,t}, Z_{-i,t}, S_t \Big] \Bigg\},
\end{aligned}
\tag{43}
$$

where $M_{i,t} = m_i$ and $M_{-i,t}$ is $M_t = (m_1, \ldots, m_I)$ with m_i deleted.

For firm i at stage t of the game, $V_i(\tilde{e}_{i,t}, \tilde{e}_{-i,t}, Z_{i,t}, Z_{-i,t}, S_t)$ is the expected payoff if it chooses $\tilde{e}_{i,t}$ and its rivals choose $\tilde{e}_{-i,t}$. A a best response strategy profile $(\tilde{E}_{i,t},\ \tilde{E}_{-i,t})$ for a stationary, pure strategy, Markov perfect equilibrium of the dynamic game is one that satisfies the familiar (Nash) condition described above in Equation (4) in Section 3.

Since this is a game of complete information it implies that the equilibrium is known if the state that consists of the current total costs of all firms $(Z_{i,t}, Z_{-i,t})$ and total revenue S_t is known.

Using the above defined choice specific value function it is possible to compute the ex ante value function as follows,

$$
V_i(Z_{i,t}, Z_{-i,t}, S_t) = V_i(\tilde{E}_{i,t}, \tilde{E}_{-i,t}, Z_{i,t}, Z_{-i,t}, S_t). \tag{44}
$$

The ex ante value function satisfies the following Bellman equation

$$
\begin{aligned}
&V_i(Z_{i,t}, Z_{-i,t}, S_t) \\
&= \sum_{m_1=0}^{1} p_E^{\delta(m_1=\tilde{E}_{1t})}(q_E)^{1-\delta(l_1=\tilde{E}_{1t})} \cdots \sum_{m_I=0}^{1} p_E^{\delta(m_I=\tilde{E}_{It})}(q_E)^{1-\delta(m_I=\tilde{E}_{It})} \Bigg\{ m_i \left(\frac{S}{\sum_{j=1}^{I} m_j} - Z_{it} \right) \\
&\quad + \beta \, \mathcal{E}\Big[V_i(Z_{i,t+1}, Z_{-i,t+1}, S_{t+1}) \,|\, M_{i,t}, M_{-i,t}, Z_{i,t}, Z_{-i,t}, S_t \Big] \Bigg\}.
\end{aligned}
\tag{45}
$$

A comprehensive discussion of results for existence of equilibria in Markovian games is provided by Dutta and Sundaram (1998). Additional discussion may be found in Reny (1999). When the state space can only take on a finite set of values, Theorem 3.1 of Dutta and Sundaram (1998) implies that the entry game has a stationary Markov perfect equilibrium in mixed strategies. Parthasarathy (1973) showed that this support condition can be relaxed to include a state space with countable values. The regularity conditions of Theorem 5.1 of Dutta and Sundaram (1998) are closer to our model, i.e., that the revenue and cost do not have to be discrete but they do need to be bounded.

In our computations we rely on Theorem 5.1 of Dutta and Sundaram (1998) because we are able to compute pure strategy equilibria for our model that has a continuous state space. The equilibrium profiles guaranteed by Theorem 5.1 depend on period t of the state vector and might depend on period $t-1$ as well. We find that we can always compute pure strategy equilibria that depend only on period t of the state vector, and hence automatically satisfy the regularity conditions of Theorem 5.1. However, our model could also be transformed to satisfy the conditions of Theorem 3.1 of Dutta and Sundaram (1998) that require the state space to be finite and countable. The proof of Theorem 3.1 of Dutta and Sundaram (1998) is significant for us because it relies on a dynamic programming approach that motivates our computational strategy.[8]

In the numerical computation we have always been able to find an equilibrium for each trial value of the parameters. Thus, computationally a bigger problem we encounter is multiplicity of equilibria. The assumption we make to resolve the issue of multiple equilibria is to adopt an explicit equilibrium selection rule. When we encounter multiple equilibria we adopt an explicit equilibrium selection rule as follows. We pick the equilibrium with the lowest aggregate costs (i.e., entrants are such that the most profitable production takes place). This idea is similar to that used by Berry (1992). That is, the strategy profiles $\tilde{E}_t$ are ranked by aggregate cost in ascending order, $Z = \sum_{i=1}^{I} \tilde{E}_{it} Z_{it}$, and the first $\tilde{E}_t$ that satisfies the equilibrium condition (4) is accepted as the solution. Note that our distributional

[8]More details about computing equilibria for this example are provided in the web version of the paper, `http://www.aronaldg.org/papers/socc_web.pdf`. However, the basic idea in computing $\tilde{E}_t$ is that one uses (43) after substituting (44) to find the $\tilde{E}_{i,t}$ that satisfy the Nash condition (4). Then the Bellman equation (45) is used as a recursion to update $V_i(Z_{i,t}, Z_{-i,t}, S_t)$. One repeats until the terms on both sides of (45) are equal within a tolerance.

assumptions guarantee that no two Z can be equal so that this ordering of the $\tilde{E}_t$ is unique. Moreover, none of the Z_{it} can equal one another. When this condition is met we have never had trouble in being able to compute a pure strategy equilibrium for any parameter value as soon as we compute a sufficient number of value function iterations.

Additional details about the computation of the model and our procedure to deal with multiple equilibria may be found in the web version of the paper. Like much of the literature on estimation and computation of dynamic games, overall, our method works conditional on the existence of an equilibrium being guaranteed. In the event of multiple equilibria it also works if there is a procedure available or if the econometrician is willing to make an assumption to deal with such multiple equilibria. As an example, in our case we use an explicit equilibrium selection rule based on economic criteria to account for multiple equilibria.

The parameters of the entry game model are given by

$$\theta = (\mu_z, \rho_z, \sigma_z, \mu_s, \sigma_s, \rho_e, \kappa_e, \beta, p_e). \tag{46}$$

The function that provides the decision profile is

$$\tilde{E}_t = H(z_{u,t}, z_{k,t}, s_t, \theta), \tag{47}$$

where $z_{u,t} = (z_{1,u,t}, \ldots, z_{I,u,t})$; similarly $z_{k,t}$. The notation of the entry game maps to the notation of the generic game as follows

$$\begin{aligned}
x_{1t} &= z_{ut} \\
x_{2t} &= (z_{kt}, s_t) \\
a_t &= E_t \\
p(x_t \,|\, a_{t-1}, x_{t-1}, \theta) &= n[z_{ut} \,|\, \mu_z \mathbf{1} + \rho_z(z_{u,t-1} - \mu_z \mathbf{1}), \sigma_z^2 I] \\
&\quad \times \delta(z_{k,t} = \rho_e z_{k,t-1} + \kappa_e a_{t-1}) \\
&\quad \times n(s_t \,|\, \mu_s, \sigma_s^2) \qquad (48) \\
p(x_{1,t} \,|\, a_{t-1}, x_{t-1}, \theta) &= n[z_{ut} \,|\, \mu_z \mathbf{1} + \rho_z(z_{u,t-1} - \mu_z \mathbf{1}), \sigma_z^2 I] \qquad (49) \\
p(a_t \,|\, x_t, \theta) &= p[E_t \,|\, H(z_{u,t}, z_{k,t}, s_t, \theta), \theta] \qquad (50) \\
p(x_{1,t} \,|\, \theta) &= n[z_{ut} \,|\, \mu_z, \frac{\sigma_z^2}{(1-\rho_z^2)} I] \qquad (51)
\end{aligned}$$

In equations (50)

$$p[E_t \mid H(z_{u,t}, z_{k,t}, s_t, \theta), \theta] = p(E_t|\tilde{E}_t, \theta) = \prod_{i=1}^{I} p_E^{\delta(E_{i,t}=\tilde{E}_{i,t})}(1-p_E)^{1-\delta(E_{i,t}=\tilde{E}_{i,t})}.$$

Hidden within $H(z_{u,t}, z_{k,t}, s_t, \theta)$ are equations (4), (43), and (44), which describe the computation of the value function and the search algorithm for the Nash equilibrium used to compute the decision profiles $\tilde{E}_t$ from each $z_{u,t}$, $z_{k,t}$, and s_t.

The specialization of the formulas for the algorithm described in Subsection 4.6 to the game with a small number of players are as follows. Substituting (50) and (48) into the numerator of (32) and (49) into the denominator,

$$\bar{v}_t^{(k)} = p[E_t \mid H^f(z_{u,t}^{(k)}, z_{k,t}, s_t, \theta), \theta]\, n(s_t \mid \mu_s, \sigma_s^2)\, \delta[z_{k,t} = \rho_z z_{k,t-1} + \kappa_z H^f(z_{u,t-1}^{(k)}, z_{k,t-1}, s_{t-1}, \theta)]$$

For the alternative importance sampler,

$$\bar{\bar{v}}_t^{(k)} = \bar{v}_t^{(k)} \frac{n[z_{ut}^{(k)} \mid \mu_z \mathbf{1} + \rho_z(z_{u,t-1}^{(k)} - \mu_z \mathbf{1}), \sigma_z^2 I]}{f\left(z_{ut}^{(k)} \mid z_{u,t-1}^{(k)}, z_{k,t-1}, s_{t-1}, \theta\right)}.$$

The expressions needed at Step 1 to compute $\hat{C}_0$ are given by (51) and (51).

6.2 A Game with a Large Number of Players

In the last section we verified our results using Monte Carlo simulations from a dynamic game of complete information with the Markov Perfect Equilibrium solution concept. However, the MPE solution concept is intractable when a game consists of a large number of players as the time to compute the solution to a game can increase exponentially with the number of players. In this section we show how our method applies to a game with a large number of players by following Weintraub, Benkard, and Roy (2008) who develop the solution concept of Oblivious Equilibrium to deal with such games. Weintraub, Benkard, and Roy (2008) show that this approach can be quite accurate in approximating the MPE solution for Ericson and Pakes (1995) type games with five or more players. To illustrate, we apply our method to one of the examples taken from Weintraub, Benkard, and Roy (2010), which has been used in applications that they cite.[9]

[9]We used their Matlab® code, for which we thank the authors, translated verbatim to C++. See `http://www.columbia.edu/~gyw2105/GYW/GabrielWeintraub_files/programsMPE120726.zip`.

The example is as follows. The set up is based on Weintraub, Benkard, and Roy (2010) and we only present the details that are relevant for our Monte Carlo analysis and refer the reader to that paper for additional details. The industry has differentiated products. Firm $i, i = 1, \ldots, I$, produces at quality level x_{it} at time $t, t = 1, \ldots, T$, where x_{it} is integer valued. For our simulation, $I = 20$. In period t consumer $j, j = 1, \ldots, m$, receives utility

$$u_{ijt} = \theta_1 \ln\left(\frac{x_{it}}{\psi} + 1\right) + \theta_2 \ln\left(Y - p_{it}\right) + v_{ijt},$$

where Y is income, p_{it} is price, $(\theta_1, \theta_2, \psi)$ are the utility function parameters, and v_{ijt} are distributed i.i.d. Gumbel. Each consumer buys one product, choosing the one that maximizes utility. For our simulation, $m = 50$. This is a logit model for which there is a unique Nash equilibrium $\{p^*_{it}\}$ in pure strategies that yields profit $\pi(x_{it}, s_{-i,t})$ to firm i, where $s_{-i,t}$ is a list of the states of its competitors. Each firm has an investment strategy $\iota_{it} = \iota(x_{it}, s_{-i,t})$, which is successful with probability $\frac{a\iota}{1+a\iota}$, in which case the quality of its product increases by one level. Quality depreciates by one level with probability δ. Our simulation concerns the parameters of the utility function and the transition dynamics, namely $\theta = (\theta_1, \theta_2, \psi, a, \delta)$, set as shown in Table 3 for our simulation, which are the same as in the Matlab code on the authors' website. There are a number of subsidiary parameters, mostly respecting $\iota(x, s)$, that have also been set to the values in the distributed Matlab code: discount factor $\beta = 0.95$, marginal investment cost $d = 0.5$, sunk entry cost $\kappa = 35$, entry state $x^e = 10$, average income $Y = 1$, marginal cost of production $c = 0.5$, and the utility of the outside good $u_0 = 0$. The oblivious equilibrium is computed by replacing $s_{-i,t}$ by its expectation under the stationary transition density for states. Details and the solution algorithm are in Weintraub, Benkard, and Roy (2010).

This game is mapped to our notation as follows. The state vector (5) at time t is comprised of the quality levels $\{x_{it}\}_{i=1}^{I}$. The state vector is known to all firms but not observed by us. It follows the Markov process described above which has a transition density represented as a transition matrix evolving according to the investment function and a stationary density represented as a vector, the elements for both of which are supplied by the solution algorithm. The transition matrix defines (8) and the vector of stationary probabilities defines (9). No portion of the state is observed by us so that (7) is the same as (8). The

measurement vector a_t is comprised of the number of customers that each firm attracts at time t. It follows a multinomial distribution whose parameters are a vector of probabilities, the elements of which are supplied by the solution algorithm, and the total number of customers m. This multinomial distribution is the density (6).

7 Simulation Experiments

To assess the efficacy of the approach proposed here that directly contradicts current practice in that the seed is random and the number of particles small, we conduct a simulation exercise.

7.1 Simulations for a Game with a Small Number of Players

We simulate the entry game game described in Subsection 6, that was previously estimated in Gallant, Hong, and Khwaja (2017), configured to represent the manufacture of a single object where entry constrains capacity. There are three firms. The time increment is one year. We set parameters according to the following considerations. A hurdle rate of 20% is a standard assumption in business which leads to a discount factor of $\beta = 0.83333$. Setting $p_e = 0.95$ seems intuitively reasonable and is in line with the estimates of Gallant, Hong, and Khwaja (2017) who estimate a similar model from pharmaceutical data except that entry has the effect of reducing rather than increasing costs. We set $\rho_e = 0.5$, which gives the entry effect a half-life of six-months. Costs are usually persistent so $\rho_z = 0.9$ seems reasonable. The remaining parameters scale with μ_s. The parameter μ_s can be chosen arbitrarily because it is the log of the nominal price of the product. We chose $\mu_s = 10$. A gross margin of 30% puts $\mu_z = 9.7$. With $\kappa_e = 0.2$ the immediate impact of entry is to reduce the gross margin to 10%. The two scale parameters σ_z and σ_s are determined by the foregoing because, if one wants a sample that mimics competition to some extent, there is far less freedom in their choice than one might imagine. One can easily produce samples where one firm is dominant for long periods or a monopoly develops. By trial and error, we found $\sigma_z = 0.1$ and $\sigma_s = 2$ to be satisfactory. In general, σ_s must be fairly large, as it is here, to prevent a monopoly.

Gallant, Hong, and Khwaja (2017) reported that p_e was estimated precisely and varying it within reason had little effect on estimates. Because the parameter was of no intrinsic

interest, they fixed it to reduce computational cost. We estimated with p_e both fixed and free to see if that held true here.

The firm's discount rate β is extremely difficult to estimate in studies of this sort (see e.g., Magnac and Thesmar (2002)). On the other hand it is not difficult to form priors for β. As mentioned above, a common rule of thumb in business is not to undertake a project whose internal rate of return is less than 20%. Theoretically, a firm should not undertake a project whose rate of return is less than its cost of capital. The historical risk premia for various industries are available (e.g., Gebhardt, Lee, and Swaminathan (2001)) to which one can add a nominal borrowing rate of 5% to arrive at a value for β. We estimated with β both fixed and free to assess the value of prior information regarding β.

The model is recursive due to (40). The customary way of dealing with this situation in time series analysis (e.g. GARCH models) is to run the recursion over initial lags prior to estimation. We set the number of initial lags to a large value $T_0 = 160$ to reduce effect of the choice of T_0 in our results. The choice of large T_0 was also motivated by the Gallant, Hong, and Khwaja (2017) study where a structural break – a bribery scandal – gave rise to 160 initial lags that could be used to run the recursion (40) but could not be used for estimation. As in Gallant, Hong, and Khwaja (2017), we also pass (41) through the recursion as part of the likelihood which is equivalent to determining a loose prior for μ_s and σ_s from the initial lags. We simulated one long data set and from it constructed three data sets, small, medium, and large, with $T = 40, 120$, and 360, respectively. The initial observations in a larger data set are the same as those of its smaller predecessor.

We considered two cases keeping the data the same

1. The entry game model is fit to the data using a blind proposal and multinomial resampling. Estimates are in Table 1. Histograms of the marginals of the posterior density for the medium sample size are in Figure 1 with β constrained.

2. The entry game model is fit to the data using an adaptive proposal and systematic resampling. Estimates are in Table 2.

The key parameter in the study of games of this sort is κ_e so we focus on it although our remarks generally apply to the other parameters as well. Our results are not surprising.

- A large sample size is better. In Tables 1 and 2 the estimates shown in the columns labeled "lg" would not give misleading results in an application.
- Constraining β is beneficial as it reduces the bimodality of the marginal posterior distribution of σ_s and pushes all histograms closer to unimodality (Figure 1).[10] In consequence, the descriptive statistics in the columns labeled "β" and "β & p_e" of Tables 1 and 2 represent a posterior distribution better than those in the columns labeled "Unconstrained."
- Constraining p_e is irrelevant except for a small savings in computational cost: compare columns "β" and "β & p_e" in Tables 1 and 2.
- Improvements to the particle filter are helpful. In particular, an adaptive proposal is better than a blind proposal; compare Tables 1 and 2. In addition, the mean squared error in tracking the true unobserved cost by the average over the particles improves by 9%.[11] Systematic resampling is better than multinomial resampling.[12]

Tables 1 and 2 about here

Figure 1 about here

7.2 Simulations for a Game with a Large Number of Players

The structure of the model is such that for each θ proposed in the MCMC chain, the oblivious equilibrium only needs to be computed once to provide the information to define the transition and observation densities. This allows one to have a large number of particles at minimal computational cost. It is as well because we found that $N = 8174$ was required to get the rejection rate down to a reasonable value. The particle filtering does not become a significant component of the computational cost up to $N = 32696$, which is the value we used for the results reported in Table 3. As seen from Table 3, the method we propose here is viable for this example thereby providing estimates of precision that the calibration methods that are often used with this model cannot.

[10] Histograms for β unconstrained, which show more extreme bimodality, are available in the web version of the paper at `http://www.aronaldg.org/papers/socc_web.pdf`.

[11] Results available in the web version of the paper.

[12] The results for an adaptive proposal with multinomial sampling are available in the web version of the paper.

Table 3 about here

8 Conclusion

We propose a Bayesian approach based on sequential importance sampling (particle filtering) to estimate the parameters of a dynamic model that can have state variables that are partially observed, serially correlated, endogenous, and heterogeneous. We illustrated by application to two examples. The first is a dynamic game of entry involving a small number of firms whose heterogeneity is based on their current costs due to feedback through capacity constraints arising from past entry. The second is an Ericson-Pakes (1995) style game with a large number of firms whose heterogeneity is based on the quality of their products with firms competing through investment in product quality that affects their market share and profitability.

Our paper makes several contributions to the literature. Our approach is based on deriving an unbiased estimate of the likelihood that is used within a Metropolis chain to conduct Bayesian inference. Unbiasedness guarantees that the stationary density of the chain is the exact posterior, not an approximation. This allows for tractable computation and feasible estimation of a dynamic model. In addition the latent state variables can be either discrete or continuous. Moreover, the approach permits endogenous feedback of past actions on the latent state variables that allows for heterogeneity among the players. Thus, our approach facilitates computation and estimation of dynamic games with either small or large number of players whose heterogeneity is affected by latent state variables, discrete or continuous, that are subject to endogenous feedback.

References

Ackerberg, D. (2009): "A new use of importance sampling to reduce computational burden in simulation estimation," *Quantitative Marketing and Economics*, 7(4), 343–376.

Aguirregabiria, V., and P. Mira (2007): "Sequential estimation of dynamic discrete games," *Econometrica*, 75(1), 1–53.

Andrieu, C., A. Douced, and R. Holenstein (2010): "Particle Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society, Series B*, 72, 269–342.

Arcidiacono, P., and R. Miller (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, 7, 1823–1868.

Bajari, P., L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75, 1331–1370.

Bajari, P., H. Hong, and S. Ryan (2010): "Identification and Estimation of Discrete Games of Complete Information," *Econometrica*, 78(5), 1529–1568.

Berry, S. (1992): "Estimation of a model of entry in the airline industry," *Econometrica*, 60(4), 889–917.

Blevins, J. (2016): "Sequential Monte Carlo Methods for Estimating Dynamic Microeconomic Models," *Journal of Applied Econometrics*, 31(5), 773–804.

Blevins, J., A. Khwaja, and N. Yang (2017): "Firm Expansion, Size Spillovers and Market Dominance in Retail Chain Dynamics," *Management Science*, forthcoming.

Bresnahan, T., and P. Reiss (1991): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48, 57—81.

Ciliberto, F., and E. Tamer (2009): "Market structure and multiple equilibria in airline markets," *Econometrica*, 77(6), 1791–1828.

Douc, R., O. Cappé, and E. Moulines (2005): "Comparison of Resampling Schemes for Particle Filtering," in *In 4th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 64–69.

Doucet, A., N. De Freitas, N. Gordon, et al. (2001): *Sequential Monte Carlo methods in practice*, vol. 1. Springer New York.

Dutta, P. K., and R. K. Sundaram (1998): "The Equilibrium Existence Problem in General Markovian Games," in *Organizations with Incomplete Information*, ed. by M. Majumdar, pp. 159–207. Cambridge University Press.

Ericson, R., and A. Pakes (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *The Review of Economic Studies*, 62, 53–82.

Fernandez-Villaverde, J., and J. F. Rubio-Ramirez (2005): "Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood," *Journal of Applied Econometrics*, 20, 891–910.

Flury, T., and N. Shephard (2010): "Bayesian Inference Based Only on Simulated Likelihood: Particle Filter Analysis of Dynamic Economic Models," *Econometric Theory*, 27, 933–956.

Gallant, A. R., H. Hong, and A. Khwaja (2017): "The Dynamic Spillovers of Entry: An Application to the Generic Drug Industry," *Management Science*, forthcoming.

Gallant, A. R., and G. Tauchen (1998): "Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions," *Journal of the American Statistical Association*, 93, 10–24.

——— (2017): "Exact Bayesian Moment Based Inference for the Distribution of the Small-Time Movements of an Ito Semimartingale," Working paper, Penn State University, http://www.aronaldg.org/papers/hfemm.pdf.

Gebhardt, W. R., C. M. C. Lee, and B. Swaminathan (2001): "Toward an Implied Cost of Capital," *Journal of Accounting Research*, 39, 135–176.

Haile, P., A. Hortaçsu, and G. Kosenok (2008): "On the empirical content of quantal response equilibrium," *American Economic Review*, 98(1), 180–200.

Ho, K. (2009): "Insurer-provider networks in the medical care market," *American Economic Review*, 99(1), 393–430.

Hotz, J., and R. Miller (1993): "Conditional Choice Probabilties and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529.

Hu, Y., and M. Shum (2012): "Nonparametric identification of dynamic models with unobserved state variables," *Journal of Econometrics*, 171(1), 32–44.

Imai, S., N. Jain, and A. Ching (2009): "Bayesian Estimation of Dynamic Discrete Choice Models," *Econometrica*, 77(6), 1865–1899.

Keane, M., and K. Wolpin (1997): "The Career Decisions of Young Men," *The Journal of Political Economy*, 105(3), 473–522.

Keane, M. P. (1994): "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 95–116.

Keane, M. P. (2010): "Structural vs. atheoretic approaches to econometrics," *Journal of Econometrics*, 156(1), 3–20.

Liu, J. S. (2008): *Monte Carlo Strategies in Scientific Computing*. Springer.

Magnac, T., and D. Thesmar (2002): "Identifying dynamic discrete decision processes," *Econometrica*, 70(2), 801–816.

Miller, R. A. (1984): "Job Matching and Occupational Choice," *Journal of Political Economy*, 92, 1086–1120.

Norets, A. (2009): "Inference in dynamic discrete choice models with serially correlated unobserved state variables," *Econometrica*, 77(5), 1665–1682.

Pakes, A. (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54(4), pp. 755–784.

Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)," *RAND Journal of Economics*, pp. 373–399.

Parthasarathy, T. (1973): "Discounted, Positive, and Noncooperative Stochastic Games," *International Journal of Game Theory*, 2(1), 25–37.

Pitt, M. K., and N. Shephard (1999): "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94, 590–599.

Pitt, M. K., R. S. Silva, P. Giordani, and R. Kohn (2012): "On Some Properties of Markov Chain Monte Carlo Methods Based on the Particle Filter," *Journal of Econometrics*, 171(2), 134–151.

Reny, P. J. (1999): "On the Existence of Pure and Mixed Strategy Nash Equilibria in Discountinuous Games," *Econometrica*, 67, 1029–1056.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55(5), pp. 999–1033.

Rust, J. (1994): "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics, Vol. 4*, ed. by R. Engle, and D. McFadden, pp. 3081–3143. North Holland.

Weintraub, G. Y., C. L. Benkard, and B. V. Roy (2008): "Markov Perfect Industry Dynamics with Many Firms," *Econometrica*, 76, 1375–1411.

——— (2010): "Computational Methods for Oblivious Equilibrium," *Operations Research*, 58, 1247–1265.

Wolpin, K. I. (2013): *The Limits of Inference without Theory*. Cambridge, Mass. and London: MIT Press.

Table 1. Parameter Estimates for the Small Game
Blind Proposal, Multinomial Resampling

Parameter		Unconstrained			Constrained: β			Constrained: β & p_a		
	value	sm	md	lg	sm	md	lg	sm	md	lg
μ_z	9.70	10.10	9.72	9.68	9.94	9.67	9.68	9.86	9.72	9.68
		(0.15)	(0.12)	(0.06)	(0.19)	(0.11)	(0.06)	(0.18)	(0.12)	(0.06)
ρ_z	0.90	0.58	0.86	0.92	0.69	0.92	0.91	0.69	0.85	0.91
		(0.25)	(0.09)	(0.03)	(0.26)	(0.05)	(0.03)	(0.25)	(0.11)	(0.03)
σ_z	0.10	0.16	0.09	0.09	0.17	0.08	0.10	0.15	0.09	0.10
		(0.05)	(0.03)	(0.01)	(0.06)	(0.03)	(0.01)	(0.07)	(0.03)	(0.01)
μ_s	10.00	9.87	9.98	9.96	9.88	9.99	9.98	9.84	9.99	9.99
		(0.10)	(0.03)	(0.02)	(0.10)	(0.03)	(0.02)	(0.13)	(0.06)	(0.02)
σ_s	2.00	1.95	1.97	1.98	2.02	2.00	2.02	2.04	2.00	2.03
		(0.09)	(0.05)	(0.01)	(0.08)	(0.02)	(0.02)	(0.10)	(0.03)	(0.01)
ρ_e	0.50	0.76	0.56	0.58	0.59	0.57	0.56	0.76	0.57	0.52
		(0.09)	(0.07)	(0.06)	(0.22)	(0.09)	(0.05)	(0.10)	(0.07)	(0.04)
κ_e	0.20	0.04	0.24	0.19	0.15	0.26	0.20	0.14	0.22	0.22
		(0.05)	(0.05)	(0.02)	(0.07)	(0.05)	(0.03)	(0.06)	(0.06)	(0.03)
β	0.83	0.90	0.95	0.87	0.83	0.83	0.83	0.83	0.83	0.83
		(0.07)	(0.04)	(0.04)						
p_e	0.95	0.97	0.94	0.95	0.96	0.94	0.95	0.95	0.95	0.95
		(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)			

The data were generated according to the game with a small number of players with parameters set as shown in the column labeled "value". For all data sets $T_0 = -160$. For the small data set $T = 40$; for the medium $T = 120$; and for the large $T = 360$. The estimate is the mean of the posterior distribution. The values below each estimate in parentheses are the standard deviation of the posterior. The prior is uninformative except for the following support conditions $|\rho_z| < 1$, $|\rho_e| < 1$, $0 < \beta < 1$, and $0 < p_e < 1$. The likelihood for μ_s and σ_s includes the observations from T_0 to 0. In the columns labeled constrained, the parameters β and p_e are constrained to equal their true values as shown in the table. The number of MCMC repetitions is 240,000 with every 25th retained for use in estimation.

Table 2. Parameter Estimates for the Small Game
Adaptive Proposal, Systematic Resampling

Parameter		Unconstrained			Constrained					
					β			β & p_a		
	value	sm	md	lg	sm	md	lg	sm	md	lg
μ_z	9.70	9.87	9.82	9.72	9.81	9.78	9.68	9.78	9.76	9.65
		(0.24)	(0.07)	(0.05)	(0.12)	(0.07)	(0.06)	(0.15)	(0.09)	(0.05)
ρ_z	0.90	0.77	0.82	0.91	0.93	0.94	0.94	0.86	0.92	0.94
		(0.03)	(0.07)	(0.05)	(0.08)	(0.04)	(0.03)	(0.09)	(0.04)	(0.02)
σ_z	0.10	0.14	0.10	0.09	0.14	0.08	0.08	0.11	0.08	0.08
		(0.02)	(0.02)	(0.01)	(0.04)	(0.02)	(0.01)	(0.04)	(0.03)	(0.01)
μ_s	10.00	10.05	10.00	9.97	9.95	9.96	9.94	9.78	9.95	9.96
		(0.06)	(0.02)	(0.01)	(0.05)	(0.02)	(0.02)	(0.07)	(0.03)	(0.03)
σ_s	2.00	1.94	1.99	1.99	1.93	1.97	2.01	2.07	1.98	1.97
		(0.10)	(0.02)	(0.02)	(0.09)	(0.01)	(0.01)	(0.09)	(0.02)	(0.02)
ρ_e	0.50	0.61	0.53	0.56	0.41	0.36	0.61	0.71	0.58	0.64
		(0.21)	(0.09)	(0.06)	(0.17)	(0.06)	(0.06)	(0.20)	(0.07)	(0.05)
κ_e	0.20	0.21	0.22	0.18	0.20	0.18	0.18	0.17	0.19	0.18
		(0.02)	(0.03)	(0.02)	(0.06)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
β	0.83	0.93	0.96	0.90	0.83	0.83	0.83	0.83	0.83	0.83
		(0.10)	(0.03)	(0.04)						
p_e	0.95	0.96	0.94	0.95	0.95	0.93	0.95	0.95	0.95	0.95
		(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)			

The data were generated according to the game with a small number of players with parameters set as shown in the column labeled "value". For all data sets $T_0 = -160$. For the small data set $T = 40$; for the medium $T = 120$; and for the large $T = 360$. The estimate is the mean of the posterior distribution. The values below each estimate in parentheses are the standard deviation of the posterior. The prior is uninformative except for the following support conditions $|\rho_z| < 1$, $|\rho_e| < 1$, $0 < \beta < 1$, and $0 < p_e < 1$. The likelihood for μ_s and σ_s includes the observations from T_0 to 0. In the columns labeled constrained, the parameters β and p_e are constrained to equal their true values as shown in the table. The number of MCMC repetitions is 80,000 with every 25th retained for use in estimation.

Table 3. Parameter Estimates for the Large Game, Blind Importance Sampler, Stratified Resampling

		Posterior	
Parameter	Value	Mean	Std. Dev.
θ_1	1.00000	0.97581	0.04799
θ_2	0.50000	0.53576	0.07317
ψ	1.00000	1.01426	0.07070
a	3.00000	2.96310	0.06846
δ	0.70000	0.64416	0.05814

The data were generated according to the game with a large number of players with parameters for the consumer's utility function and firm's transition function set as shown in the column labeled "Value" and all others set to the values specified in Section 6.2. The number of firms is 20 and the number of consumers is 50. $T = 5$. The prior is uninformative except for a support condition that all values be positive. The number of MCMC repetitions is 109,000 and the number of particles per repetition is 32696.

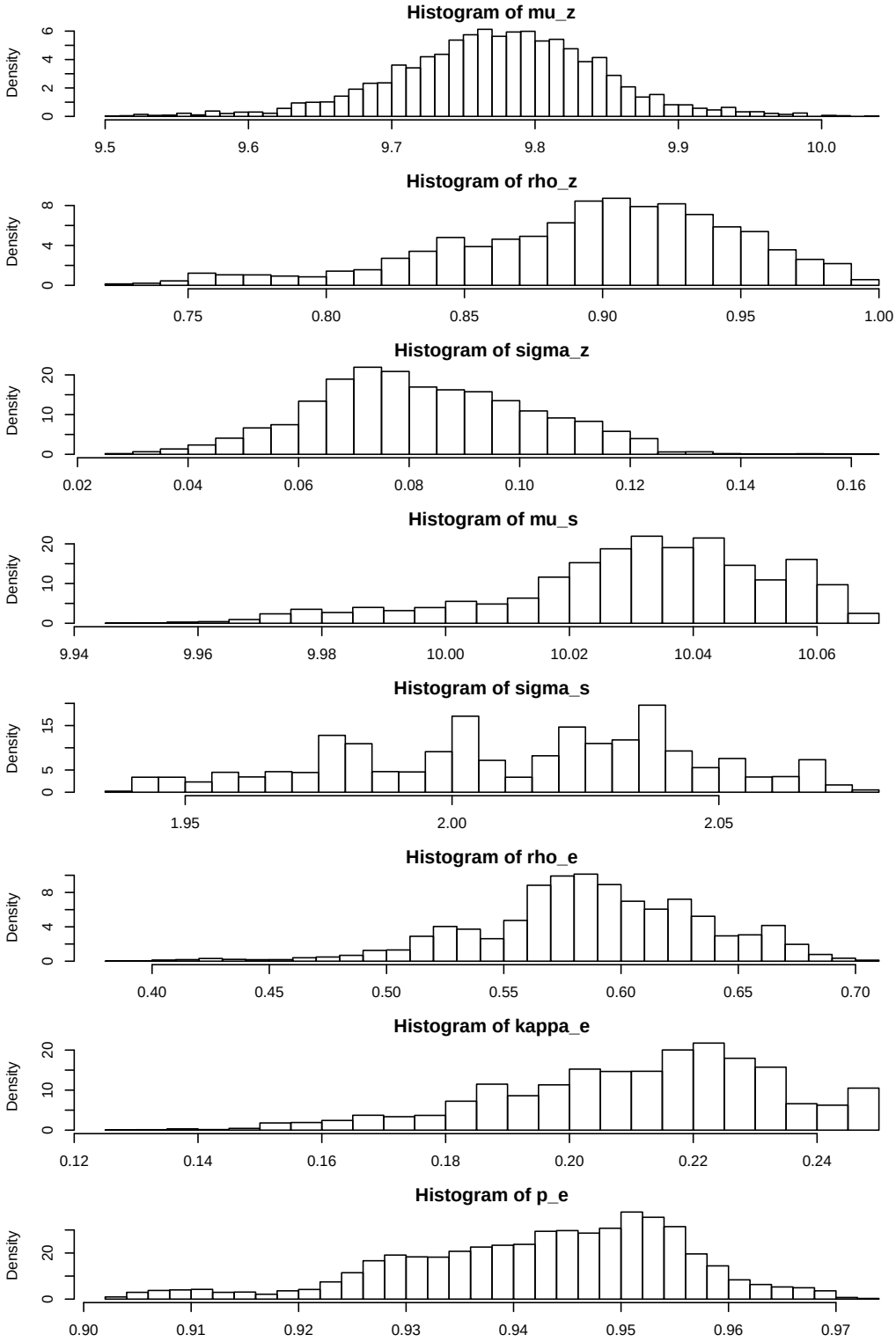

Figure 1. Small Game, β Constrained, Blind Proposal. Shown are histograms constructed from the MCMC repetitions for the column labeled "Constrained," "β," "md" in Table 1.