Simulation Based Estimation of Discrete Sequential Move Games of Perfect Information

Wang, Yafeng and Graham, Brett

Wang Yanan Institute for Studies in Economics, Xiamen University

8 July 2010

Simulation Based Estimation of Discrete Sequential Move Games of Perfect Information*

Yafeng Wang and Brett Graham†

Wang Yanan Institute for Studies in Economics (WISE)

Xiamen University

June 20, 2010

Abstract

We propose simulation based estimation for discrete sequential move games of perfect information which relies on the simulated moments and importance sampling. We use importance sampling techniques not only to reduce computational burden and simulation error, but also to overcome non-smoothness problems. The model is identified with only weak scale and location normalizations, monte Carlo evidence demonstrates that the estimator can perform well in moderately-sized samples.

Keywords: Game-Theoretic Econometric Models; Sequential-Move Game; Method of Simulated Moments; Importance Sampling; Conditional Moment Restrictions.

JEL Classification Numbers: C01, C13, C35, C51, C72.

*Preliminary version, comments are welcome.

†We would like to thank participants at Chinese Economists Society 2010 Annual Conference for helpful discussions. Yafeng Wang: econwang@econwang.co.cc. Brett Graham: brett.d.graham@gmail.com

1 INTRODUCTION

Nash equilibrium is one of the cornerstones of modern economic theory, with substantive application in all major fields in economics, particularly industrial organization. It is the benchmark theoretical model for analyzing strategic interactions among a handful of players. Given the importance of gaming in economic theory, the empirical analysis of games has been the focus of a recent literature in econometrics and industrial organization, such as Tamer (2003), Berry & Tamer (2007), Aguirregabiria & Mira (2007), Aradillas-Lopez (2007, 2008), Ciliberto & Tamer (2009), Bajari, Hong, Krainer & Nekipelov (2010) and Bajari, Hong & Ryan (2010) (hereafter BHR).

Econometrically, a discrete game is a generalization of a standard discrete choice model, such as the conditional logit or multinomial probit. An agent's utility is often assumed to be a linear function of covariates and a random preference shock. However, unlike a discrete choice model, utility is also allowed to depend on the actions of other agents. Such modeling strategy was first suggested by the seminal work of Bresnahan & Reiss (1990, 1991). Although there are numerous studies on both methodology and empirical applications of game-theoretic models, the most widely studies is the class of incomplete information simultaneous-move games (normal form) and dynamic games, see Tamer (2003), Bajari, Hong, Krainer & Nekipelov (2010) and Aguirregabiria & Mira (2007). The complete information games received fewer studies due to its computational complexity, since it involves multidimensional integrals. More recently, Ciliberto & Tamer (2009) and BHR (2010) provide simulation-based estimators for static complete information discrete games. Furthermore, estimation of sequential-move (extensive form) games has been quite limited, especially on its' general form, Berry (1992), Mazzeo (2002) and Schmidt-Dengler (2006) estimate some simplified sequential-move games with special game structure. The estimation of the general class of sequential move games has suffered from its computational complications, Maruyama (2009) provides a simulation-based estimator for the general class of discrete-choice perfect information sequential move games with a modified version of the GHK simulator (Geweke (1989, 1991), Hajivassiliou & McFadden (1998) and Keane (1990, 1994)), which he called as "sequential GHK". The estimator provided by Maruyama (2009) essentially is a maximum simulated likelihood (MSL) estimator, As is well known, MSL is biased for any fixed number of simulations, in order to obtain $\sqrt{T}$ consistent estimators, one needs to increase the number of draws NS so that $\frac{NS}{\sqrt{T}} \to \infty$. Wang & Graham (2009) provides a generalized maximum entropy (GME) estimator for this class of games which avoids the usual multidimensional integrals by using the data constraints instead of the moment constraints, they reformulate the estimation problem as a mixed-integer nonlinear optimization problem since there are logical connections between endogenous variables among the equilibrium conditions, although the computational burden is acceptable for most applications, it is hard to construct large sample properties for this GME estimator, since essentially it is a nonsmooth estimation[1].

In this paper, we propose a simulation based estimator for discrete sequential move games of perfect information which relies on the simulated moments and importance sampling. As noted by Maruyama (2009), the estimation of sequential games has some distinctive features and advantages over simultaneous games, the most advantage is that perfect information sequential games can utilize the notion of subgame perfection, which guarantees the existence of unique equilibria, however, in simultaneous games of complete information, the existence of multiple equilibria is sometimes considered problematic or at least an issue to deal with (see for example, Ciliberto & Tamer, 2007; BHR, 2010).

The moment conditions implied by the model equilibrium conditions in discrete sequential move games of perfect information contain multidimensional integrals, in principle, one can use straightforward monte Carlo simulations to get unbiased estimators for

[1]One can use the bootstrap or other resampling methods to do the inference with this GME estimation, but little is known about the ability of such methods to provide asymptotic refinements or even the consistent approximation to the asymptotic distribution.

such multidimensional integrals, but there are several problems that can arise with estimators based on such simulations. First, there are discrete parts of the model, the objective function in the estimation procedure is typically discontinuous in the parameter vector, making it hard to minimize (maximize) correctly; Second, the straightforward monte carlo simulations need to solve the game numerous times, typically once for every draw, for every observation, for every parameter vector that is ever evaluated in an optimization procedure. If we have T observations, performs NS simulation draws, and optimization requires R function evaluations, estimation requires solving the model $NS * T * R$ times, this can be computationally time consuming since R can be quite large. In spirit of Ackerberg (2009) and BHR (2010), we make use of importance sampling to overcome both of the problems, by finding the right change of variables to do the importance sampling over, the simulated approximation of the multidimensional integral (expectation) will generally be continuous in the parameter vector, and also one reduce the times of solving the game from $NS * T * R$ to $NS * T$.[2] In order to make use of importance sampling, it is important to make sure that the tails of the importance density are not too thin in a neighborhood of the parameter that minimizes (maximizes) the objective function in the estimation procedure, the GME estimator proposed by Wang & Graham (2009) can be used to construct the importance density, or one can make use of the MSL estimator proposed by Maruyama (2009). Based on such simulated moments, we propose two estimators for the discrete sequential move games of perfect information, one is the method of simulated moments (MSM), which is same as the usual GMM estimation but use the simulated moments instead of the true moments. Given that the equilibrium conditions are conditional moment restrictions, same as the GMM estimation, MSM estimation may induce inconsistent estimates due to the number of arbitrarily chosen instruments is finite, we make use of the always consistent estimation procedure that is directly based on the definition

[2]One can even reduce this computation times to NS by using the same simulation draws for different observations, see Ackerberg (2009).

of the conditional moments proposed by Dominguez & Lobato (2004). Our monte Carlo experiments show that the always consistent estimator performs better than the MSM estimator, especially in the small sample size.

The paper is organized as follows. In section 2 we outline the general discrete sequential-move games to be estimated and formulate its equilibrium conditions, the assumptions for the identification and estimation also are presented. Section 3 formalizes our simulation and estimation approach. Monte Carlo simulations are conducted in section 4. Section 5 concludes, and provides limitations and future work.

2 THE MODEL

In the model, there are T independent repetitions of a sequential move game of perfect information (extensive form game). In each game there are $i = 1, ..., N_t$ players, each with the finite set of actions A_{it}. Define $A_t = \times_i A_{it}$ and let $a_t = (a_{1t}, ..., a_{it}, ...a_{Nt})$ denote a generic element of A_t. Without loss of generality, the order of subscripts for players $(1, ..., N_t)$ also represents the decision order of the sequential move game in each repetition, that means player 1 makes decision first and player N_t at the end. Player i's von Neumann-Morgenstern (vNM) utility is a map $u_{it} : A_t \to R$, where R is the real line. Since we study the sequential move game, the corresponding equilibrium concept is the subgame perfect equilibria (SPE), this can be achieved when every player expects no gain from individually deviating from its equilibrium strategy in its every subgame, the standard technique for solving the SPE is backward induction, furthermore, the finite sequential move game of perfect information where there is no player is indifference between any two outcomes has a unique SPE. We will sometimes drop the subscript t for simplicity when no ambiguity would arise.

Following Bresnahan & Reiss (1990, 1991), assume that the vNM utility of player i

can be written as:

$$u_i(a, x, \epsilon_i; \theta) = \Pi_i(x, a; \theta) + \epsilon_i(a) \tag{1}$$

In Equation (1), player i's vNM utility from action a is the sum of two terms. The first term $\Pi_i(x, a; \theta)$ is a function which depends on a, the vector of actions taken by all of the players, covariates x, the players' characteristics and some other variables which influence the utility, and parameters θ, covariates x are observed to the econometrician. The second term is $\epsilon_i(a)$, a random preference shock which reflects the information about utility that is common knowledge to the players but not observed by the econometrician. Unlike Maruyama (2009), here the preference shocks depend on the entire vector of actions a, not just the actions taken by player i. As argued by BHR (2010), this is a more general setting and seems straightforward within the game framework, think about a simple entry game, the unobserved information of one player to econometrician may be different not only among players but also action vector dependent[3]. $\epsilon_i(a)$ are assumed to be independent, let ϵ_i denote the vector of the individual $\epsilon_i(a)$ and ϵ_i denote the vector of all the shocks. we will discuss more about the structure of ϵ_i in the model assumptions.

As noted above, the equilibrium concept corresponding to the sequential move game of perfect information, SPE, is a equilibrium strategy profile which means that every player expects no gain from individually deviating from its equilibrium in every subgame. A strategy of player $i \in N$ is a function that assigns an action in A_i to each nonterminal history, a player's deviation form equilibrium holding other's decisions fixed does not mean that all the others make the same decision, it means the others follow the same strategy. But what can be observed is only the equilibrium actions (i.e., equilibrium outcome). Thus, for deriving the equilibrium conditions in our econometric model, we should make the others' action profile when one player deviating as endogenous variable. Formally, an

[3]One can find that this specification of the preference shock also facilitates the use of importance sampling, since the usual use of importance sampling in the discrete choice models requires the random coefficients specification.

SPE action profile, $a^{SPE} = (a_1^{SPE}, ...a_i^{SPE}, ...a_N^{SPE})$, is any solution for the decisions of the players that satisfies:

$$u_i(a_i^{SPE}, a_{-i}^{SPE}, x, \epsilon_i; \theta) - u_i(a_i, a_{<i}^{SPE}, a_{>i}^*(a_{<i}^{SPE}, a_i), x, \epsilon_i; \theta) \geq 0 \tag{2}$$
$$\textit{for all } i = 1, ..., N \textit{ and all } a_i \neq a_i^{SPE}.$$

where $a_{>i}^*(a_{<i}^{SPE}, a_i)$ is a SPE action profile for the subgame that starts from player $i+1$ given the decisions of the preceding players, $a_{\leq i}$. This equilibrium conditions are defined recursively and the solution can be easily calculated by the backward induction for any given parameters θ, observed covariates, x, and unobservable shocks ϵ. Kuhn's theorem ensures the existence of solutions of the inequality system (2) but makes no claim of uniqueness, thus we can conclude that every finite sequential move game of perfect information has a SPE. As noted by Berry & Tamer(2007), dealing with multiple equilibria complicate the identification problem, fortunately, a modified version of Kuhn's theorem ensures the uniqueness of equilibria of finite sequential move games of perfect information, which is presented in theorem 1.

Theorem 1 *Every finite sequential move game with perfect information in which no player is indifference between any two outcomes has a unique subgame perfect equilibrium.*

Proof. See Osborne & Rubinstein (1994). ■

Obviously, the indifference case can be ignored in our econometric model since we work with continuous latent payoffs ($\epsilon_i(a)$ has an atomless distribution). Given such structure of the discrete choice sequential move game, our task is to estimate and draw an inference about the parameters of payoff functions, θ, with the observation of action profile a^o, some covariates which have effect on the payoffs, x, and an exogenous decision order. Note that the actual payoff levels are unobserved, since in most case, we can not determine what they should be, i.e. they are the latent variables. Before presenting our estimation strategy, some assumptions about the model structure are introduced.

2.1 Assumptions

Assumption 1 (Exogenous Decision Order) *The decision order of agents in the sequential move game is exogenous.*

Although the exact decision order of agents is rarely observed, we can estimate sequential move games by imposing different decision order assumptions, this restriction only excludes the endogenous decision order which may alter the uniqueness of the game structure.

Assumption 2 (Scale and Location Normalizations) *The payoff of one action for each player are fixed at a known constant.*

As argued by BHR (2010), this restriction is similar to the argument that we can normalize the mean utility from the outside good equal to a constant, usually zero, in a standard discrete choice model. One clearly find that from the equilibrium condition (2) that adding a constant to all deterministic payoffs does not perturb the set of equilibria, so a location normalization is necessary. A scale normalization is also necessary, as multiplying all deterministic payoffs by a positive constant does not alter the SPE. This restriction is subsumed in the following assumption about the distribution of the error terms.

Assumption 3 (Regularity Conditions of Random Shocks) *The joint distribution of $\epsilon = (\epsilon_i(a))$, $G(\epsilon|\beta)$ is independent and known to all agents and the econometrician.*

This restriction allows G to be any known joint parametric distribution, identification in this game with unknown G is complicated, and since our estimation is based on the simulation which relies on the distribution of error terms, the case with unknown G will not be dealt with here.

3 ESTIMATION

Next, we propose computationally efficient simulation based estimators for θ and β, the parameters governing agents' deterministic payoffs and the error terms' distribution, given the observations of a sequence (a_t, x_t) of action profiles and covariates. To form the estimation framework, enumerate the elements of A from $k = \{1, ..., \#A\}$. Denote the observation at t^{th} repetition of the game with y_t and

$$y_t = \begin{bmatrix} I(a_t = 1) \\ . \\ I(a_t = k) \\ . \\ I(a_t = \#A) \end{bmatrix} = f(x_t, \epsilon_t, \theta_0) \tag{3}$$

where $I(\cdot)$ is the usual indicator function, $f(x_t, \epsilon_t, \theta)$ is an algorithm which solves the game for any given x_t, ϵ_t and θ, obviously, it is corresponding to the model equilibrium conditions (2). Denote the probability that a specific action profile k is played implied by the model as $P(k|x_t; \theta)$ and collect them into a vector $P(a|x_t; \theta)$, where

$$P(a|x_t; \theta, \beta) = E[f(x_t, \epsilon_t, \theta)|x_t] = \int f(x_t, \epsilon_t, \theta) dG(\epsilon, \beta) \tag{4}$$

At the true parameters of the data-generating process the predicted probability of each action equals its empirical probability of each action k:

$$E[(y_t - P(a|x_t; \theta, \beta))|x_t] = 0 \quad \text{at } \theta = \theta_0, \beta = \beta_0 \tag{5}$$

Note that, because the probability of all of the elements of must sum to one, one of these probabilities will be linearly dependent on the others, so there are effectively $\#A - 1$ conditional moment restrictions. Obviously, the expectation of any function $w(x_t)$ of the conditioning variables multiplied by the difference between y_t and the predicted probabilities is identically zero at the true parameters, i.e.

$$E[w(x_t) * (y_t - P(a|x_t; \theta, \beta))] = \mathbf{0} \text{ at } \theta = \theta_0, \beta = \beta_0 \tag{6}$$

In principle, the value of θ and β, say $\hat{\theta}$ and $\hat{\beta}$, that set the sample analog of this moment

$$G_T(\theta, \beta) = \frac{1}{T} \sum_t [w(x_t) * (y_t - P(a|x_t; \theta, \beta))]$$

equal to zero or as close as possible to zero is a consistent estimator of θ_0 and β_0. Under appropriate regularity conditions, one obtains asymptotic normality of the estimators (Hansen, 1982), and as the number of moments used increases, one can approach asymptotic efficiency by the right choice of instruments (i.e. the w function).

To make use of such GMM estimation, we should overcome some obstacles, the first obstacle is that the predicted probabilities $P(a|x_t; \theta, \beta)$ which defined by (4) is not easily computable, since it involves a multidimensional integral, thus simulation enters the picture. As can be found below, a straightforward Monte Carlo procedure is not practical due to the computational burden and discreteness in $f(x_t, \epsilon_t, \theta)$, we make use of importance sampling to overcome such problems.

3.1 Simulation

The straightforward way of simulating

$$P(a|x_t; \theta, \beta) = E[f(x_t, \epsilon_t, \theta)|x_t] = \int f(x_t, \epsilon_t, \theta) dG(\epsilon, \beta)$$

is by averaging $f(x_t, \epsilon_t, \theta)$ over a set of NS random draws $(\epsilon_1, ..., \epsilon_{NS})$ from the distribution of ϵ_t, $G(\epsilon|\beta)$, i.e.

$$\tilde{P}(a|x_t; \theta, \beta) = \frac{1}{NS} \sum_{ns} f(x_t, \epsilon_t, \theta) \tag{7}$$

$\tilde{P}(a|x_t; \theta, \beta)$ is trivially an unbiased simulator of the true expectation $P(a|x_t; \theta, \beta) = E[f(x_t, \epsilon_t, \theta)|x_t]$. McFadden (1989) and Pakes & Pollard (1989) prove statistical properties of the MSM estimator that set the simulated moment:

$$\begin{aligned} \tilde{G}_T(\theta, \beta) &= \frac{1}{T} \sum_t [w(x_t) * (y_t - \tilde{P}(a|x_t; \theta, \beta))] \\ &= \frac{1}{T} \sum_t [w(x_t) * (y_t - \frac{1}{NS} \sum_{ns} f(x_t, \epsilon_t, \theta))] \end{aligned} \tag{8}$$

as close as possible to zero. The most important of these statistical properties is the fact that these estimators are typically consistent for finite NS. The intuition behind this is that simulation error averages out over observations as $T \to \infty$. This consistency property gives the estimator an advantage over alternative estimation approaches such as maximum simulated likelihood (MSL), which typically is not consistent for a finite number of simulation draws. Another nice property of these estimators is that the extra variance imparted on the estimates due to the simulation is relatively small, asymptotically it is $1/NS$. As noted above, an important obstacle of making use of MSM estimation procedure in our sequential game estimation is that $f(x_t, \epsilon_t, \theta)$ typically is not continuous in θ, since the algorithm for solving the discrete sequential move game of perfect information essentially is a combination of several indicator functions, which is not continuous in θ. The discreteness in $f(x_t, \epsilon_t, \theta)$ will generate the discreteness in $\tilde{P}(a|x_t; \theta, \beta)$, as can be found via a simple entry game conducted in example 1. Thus the simulated moments, $\tilde{G}_T(\theta, \beta)$, will tend not to be continuous in θ, typically having both flats and jumps. This can be very problematic in the numeric minimization of $\tilde{G}_T(\theta, \beta)$, derivative based methods are useless.

Example 1 *To illustrate the discreteness problem, consider a simple two-firm sequential entry game, where firm 1 moves first. Each firm has the following profit function:*

$$u_i(x, a, \epsilon_i; \theta) = 1(a_i = 1)\{x_i\theta_1 + N(a)\theta_2 + \epsilon_i(a)\}$$

where $a_i \in \{0, 1\}$ *is firm* i*'s action,* $N(a)$ *is the number of entrants for a action profile* a*. Function* f *maps* (x, ϵ, θ) *into the market structure (outcome)* y*,*

$$y = \begin{bmatrix} I(0,0) \\ I(0,1) \\ I(1,0) \\ I(1,1) \end{bmatrix} = f(x, \epsilon, \theta)$$

For exposition we focus on the 2nd element of y*, we can write this out explicitly as:*

$$y_2 = I(0,1) = I\left(\begin{array}{c} [0 > x_1\theta_1 + \theta_2 + \epsilon_1(1,0) \quad \cap \quad 0 > x_2\theta_1 + 2\theta_2 + \epsilon_2(1,1)] \\ \cup \\ [0 > x_1\theta_1 + 2\theta_2 + \epsilon_1(1,1) \quad \cap \quad 0 \leq x_2\theta_1 + 2\theta_2 + \epsilon_2(1,1)] \\ \cap \\ x_2\theta_1 + \theta_2 + \epsilon_2(0,1) \geq 0 \end{array}\right)$$

Obviously, function f *is not continuos in* θ*. The straightforward simulator*

$$\tilde{P}((0,1)|x_t; \theta, \beta) = \frac{1}{NS}\sum_{ns} I\left(\begin{array}{c} [0 > x_1\theta_1 + \theta_2 + \epsilon_{1,ns}(1,0) \quad \cap \quad 0 > x_2\theta_1 + 2\theta_2 + \epsilon_{2,ns}(1,1)] \\ \cup \\ [0 > x_1\theta_1 + 2\theta_2 + \epsilon_{1,ns}(1,1) \quad \cap \quad 0 \leq x_2\theta_1 + 2\theta_2 + \epsilon_{2,ns}(1,1)] \\ \cap \\ x_2\theta_1 + \theta_2 + \epsilon_{2,ns}(0,1) \geq 0 \end{array}\right)$$

is also not continuos in θ.

In spirit of Ackerberg (2009) and BHR (2010), we make use of importance sampling to reduce the non-smoothness problem[4]. Importance sampling is most noted for its ability to reduce simulation error and computational burden, and was first used in game-theoretic models estimation by BHR (2010), who estimated norm form complete information games. First, we change the variable of integration in Equation (4) from ϵ to u. Let $h(u|x,\theta,\beta)$ denote the density of u, conditional on x, θ and β, and $g(\epsilon_i(a)|\beta)$ the density of $\epsilon_i(a)$. Then the density $h(u|x,\theta,\beta)$ is:

$$h(u|x,\theta,\beta) = \prod_i \prod_{a\in A} g(u_i(a,x,\epsilon_i;\theta) - \Pi_i(x,a;\theta)|\beta) \tag{9}$$

If we change the variable of integration in

$$\begin{aligned} P(a|x_t;\theta,\beta) = E[f(x_t,\epsilon_t,\theta)|x_t] &= \int f(x_t,\epsilon_t,\theta)dG(\epsilon,\beta) \\ &= \int f(x_t,\epsilon_t,\theta)g(\epsilon|\beta)d\epsilon \end{aligned}$$

from ϵ to u, then $P(a|x_t;\theta,\beta)$ becomes:

$$P(a|x_t;\theta,\beta) = \int f(u)h(u|x_t,\theta,\beta)du \tag{10}$$

In order to use importance sampling, introduce the importance density $q(u)$, rewrite Equation (10) as:

$$P(a|x_t;\theta,\beta) = \int f(u)\frac{h(u|x_t,\theta,\beta)}{q(u)}q(u)du \tag{11}$$

[4]McFadden (1989) noted the ability to use importance sampling to smooth simulations which is extended by Ackerberg (2009).

We can then simulate $P(a|x_t;\theta,\beta)$ by draw random variables $u_1, ...u_{NS}$ from $q(u)$ and construct

$$\hat{P}(a|x_t;\theta,\beta) = \frac{1}{NS}\sum_{ns=1}^{NS} f(u_{ns})\frac{h(u_{ns}|x_t,\theta,\beta)}{q(u_{ns})} \tag{12}$$

Note that

$$\begin{aligned}
E[\hat{P}(a|x_t;\theta,\beta)] &= E[f(u)\frac{h(u|x_t,\theta,\beta)}{q(u)}] \\
&= \int f(u)\frac{h(u|x_t,\theta,\beta)}{q(u)}q(u)du \\
&= E[f(x_t,\epsilon_t,\theta)|x_t] \\
&\equiv P(a|x_t;\theta,\beta)
\end{aligned}$$

So the importance sampling simulator $\hat{P}(a|x_t;\theta,\beta)$ is an unbiased simulator for the true expectation. The most important property of this simulator is that $\hat{P}(a|x_t;\theta,\beta)$ will generally be continuous in θ and β since it only depends on θ and β through $h(u|x_t,\theta,\beta)$ which is continuous in θ and β given that $g(\epsilon|\beta)$ is continuous, this can be revealed by using this simulator in the simple two-player entry game which conducted in Example 1.

Example 2 (Ex.1 Cont') *Consider the two-player entry game conducted in Example 1. For exposition we also only focus on the 2nd element of y:*

$$y_2 = I(0,1) = I\begin{pmatrix} [0 > x_1\theta_1 + \theta_2 + \epsilon_1(1,0) \quad \cap \quad 0 > x_2\theta_1 + 2\theta_2 + \epsilon_2(1,1)] \\ \cup \\ [0 > x_1\theta_1 + 2\theta_2 + \epsilon_1(1,1) \quad \cap \quad 0 \leq x_2\theta_1 + 2\theta_2 + \epsilon_2(1,1)] \\ \cap \\ x_2\theta_1 + \theta_2 + \epsilon_2(0,1) \geq 0 \end{pmatrix}$$

A change of variables from ϵ to u resulting in

$$\hat{P}((0,1)|x_t;\theta,\beta) = \frac{1}{NS}\sum_{ns} \frac{I\begin{pmatrix} [0 > u_{1,ns}(1,0) & \cap & 0 > u_{2,ns}(1,1)] \\ & \cup & \\ [0 > u_{1,ns}(1,1) & \cap & 0 \leq u_{2,ns}(1,1)] \\ & \cap & \\ & u_{2,ns}(0,1) \geq 0 & \end{pmatrix} h(u_{ns}|x_t,\theta,\beta)}{q(u_{ns})}$$

obviously, given that $g(\epsilon|\beta)$ is continuous, this simulator is smooth in the underlying parameters.

Although the theory of importance sampling proves that $\hat{P}(a|x_t;\theta,\beta)$ is a smooth and unbiased simulator for any choice of the importance density $q(u)$ which has sufficiently large support. However, as noted by BHR(2010), as a practical matter, it is important to make sure that the tails of the importance density $q(u)$ are not too thin in a neighborhood of the parameter that minimizes the objective function in our estimator. One natural choice of $q(u)$ is $h(u|x,\mathring{\theta},\mathring{\beta})$ where $\mathring{\theta}$ and $\mathring{\beta}$ are some guess or preliminary estimate of θ and β. To ensure that the importance density $q(u)$ are not too thin in a neighborhood of the estimated parameters, we found that the generalized maximum entropy (GME) estimator proposed by Wang & Graham (2009) is a good choice for $\mathring{\theta}$ and $\mathring{\beta}$, also we can set the importance density equals to the distribution of utilities conditional on x in the GME estimation, this means that for each value of x we simulate the GME estimation NS times. At the same time, since $\hat{P}(a|x_t;\theta,\beta)$ only depends on θ and β through $h(u|x_t,\theta,\beta)$ which is continuous in θ and β given that $g(\epsilon|\beta)$ is continuous, in computations, the $f(u_{ns})$ and $q(u_{ns})$ should be stored as they do not vary as the underlying parameters changes in the estimation procedure, then as the underlying parameters changes, one only need re-compute the density $h(u|x,\theta,\beta)$.

3.2 The Estimator

Given the importance simulator $\hat{P}(a|x_t;\theta,\beta)$, we can replace the moment conditions in Equation (6) by its simulation analog:

$$\hat{G}_T(\theta,\beta) = \frac{1}{T}\sum_t [w(x_t) * (y_t - \hat{P}(a|x_t;\theta,\beta))]$$

Then for a positive definite weighting matrix W_T, the MSM estimator is:

$$(\hat{\theta}_{MSM}, \hat{\beta}_{MSM}) = \arg\min_{\{\theta,\beta\}} \hat{G}_T(\theta,\beta)' W_T \hat{G}_T(\theta,\beta) \tag{13}$$

The asymptotic theory for estimating discrete choice models using MSM is well developed by McFadden (1989) and Pakes & Pollard (1989). Christian Gouriéoux & Alain Monfort (2002) has done a formal analysis of the MSM estimation in the GMM framework, involved the optimal choice of the weighting matrix W_T and instrumental matrix $w(x_t)$. However, this MSM estimator which relies on the conditional moment restrictions (5), just as the GMM, can render inconsistent estimates since the number of arbitrarily chosen instruments is finite. In fact, consistency of the GMM estimators relies on additional assumptions that imply unclear restrictions on the data generating process. To avoid such inconsistent case, we can make use of the consistent estimation of models defined by conditional moment restrictions proposed by Dominguez & Lobato (2004)[5], but use the simulation analog instead of the usual sample analog. The always consistent estimator can be defined as:

$$(\hat{\theta}_{AC}, \hat{\beta}_{AC}) = \arg\min_{\{\theta,\beta\}} \frac{1}{T^3}\sum_{l=1}^{T}\left[\left(\sum_{t=1}^{T}\hat{m}(y_t,x_t)I(x_t \le x_l)\right)' \left(\sum_{t=1}^{T}\hat{m}(y_t,x_t)I(x_t \le x_l)\right)\right] \tag{14}$$

[5]The main idea behind this estimation is that use the whole information about the parameters contained in the conditional moments $E[h(Y_t,\theta_0)|X_t] = 0$ by the fact: $E[h(Y_t,\theta_0)|X_t] = 0 \iff E[h(Y_t,\theta_0)I(X_t \le x)]$.

where

$$\hat{m}(y_t, x_t) = y_t - \hat{P}(a|x_t; \theta, \beta) \tag{15}$$

This estimator is always consistent but inefficient since it does not control the minimization of the covariance, Dominguez & Lobato (2004) briefly discussed that by carrying out a single Newton-Raphson step in the direction of the efficient GMM estimator, an asymptotically efficient estimator can be constructed. Another choice of the efficient estimation is Kitamura, Tripathi & Ahn (2004)'s local estimation, but it needs to introduce a bandwidth number, although this bandwidth number allows the estimator to be $root - n$ asymptotically normal and efficient, statistical inference with this estimator can be sensitive to the selection of the bandwidth number.

4 MONTE CARLO

To demonstrate the performance of our estimator in finite samples, we conducted a simple Monte Carlo experiment using the simple sequential entry game introduced in Example 1. There are two players and each player has the following profit function:

$$u_i(x, a, \epsilon_i; \theta) = 1(a_i = 1)\{\theta_1 x_{i1} + \theta_2 x_{i2} + \theta_3 x_{i3} + \epsilon_i(a)\} \tag{16}$$

where player 1 moves first. We assume that

$$x_{11} \sim N(20, 1)$$

$$x_{12} \sim N(11, 3)$$

$$x_{21} \sim N(26, 1)$$

$$x_{22} \sim N(11, 3)$$

$$x_{i3} = N(a)$$

where $N(a)$ is the number of entrants for a action profile a, and $\epsilon_{it}(a)$, the idiosyncratic error term, are drawn from standard normal distribution. As discussed previously, our model requires both scale and location normalizations, so we assume the variance of the error terms is one and the payoffs of not entering are zero. Thus our game has three unknown parameters: θ_1, θ_2 and θ_3. We generated 1000 samples of size $T = 25, 50, 100, 200$ and 400 to assess the finite sample properties of our estimator, first use importance simulator (12) get $\hat{P}(a|x_t; \theta, \beta)$ for each t then generate the simulated analog (15). The true parameter vector was chosen as

$$\theta_1 = 1, \theta_2 = -1, \theta_3 = -8$$

the random draws in the importance sampling, NS, is 1000.

In Table I we report the mean, median, standard deviation, mean bias, median bias and mean square error (MSE) for the MSM estimator defined in (13) for five sample sizes, $T = 25, 50, 100, 200$ and 400 and Table II for the AC estimator defined in (14), which show that both estimators can perform well in moderately-sized samples, the payoff parameters are estimated near their true values, and as the sample size increase, the estimates become more precisely. The comparison between the MSM estimator and AC estimator shows the superiority of AC estimator, especially in small samples. One may find that parameters are estimated much less precision when sample size is 400, this may due to the large scale non-linear algorithm we've chosen. Actually, since the objective function of our estimate is not globally convex, we choose the global optimization algorithm "LGO"[6] in GAMS, a more meticulous modification on the algorithm details should increase the performance of our estimation in large samples.

[6]The Lipschitz-Continuous Global Optimizer (LGO) developed by János D. Pintér.

Table I: Monte Carlo Results for MSM

Parameter	Mean	Median	Standard Deviation	Mean Bias	Median Bias	MSE
			$T = 25$			
θ_1	1.0769	1.0774	0.1332	0.0769	0.0774	0.0236
θ_2	−1.1254	−1.1039	0.2196	−0.1254	−0.1039	0.0639
θ_3	−8.4267	−8.4045	1.0817	−0.4267	−0.4045	1.2637
			$T = 50$			
θ_1	1.0721	1.0730	0.1311	0.0721	0.0730	0.0224
θ_2	−1.1140	−1.0953	0.2203	−0.1140	−0.0953	0.0615
θ_3	−8.3077	−8.2143	1.0196	−0.3077	−0.2143	1.2206
			$T = 100$			
θ_1	1.0239	1.0242	0.1111	0.0239	0.0242	0.0129
θ_2	−1.0201	−1.0133	0.1371	−0.0201	−0.0133	0.0192
θ_3	−8.1938	−8.1152	0.9040	−0.1938	−0.1152	0.8539
			$T = 200$			
θ_1	1.0089	1.0065	0.1070	0.0089	0.0065	0.0115
θ_2	−0.9960	−0.9937	0.1239	0.0040	0.0063	0.0154
θ_3	−8.0946	−8.0422	0.8521	0.0946	0.0422	0.7343
			$T = 400$			
θ_1	1.0071	1.0022	0.0886	0.0071	0.0022	0.0079
θ_2	−0.9979	−0.9944	0.1011	0.0021	0.0056	0.0102
θ_3	−8.0918	−8.0367	0.7070	−0.0918	−0.0367	0.5078

True value: $\theta_1 = 1$, $\theta_2 = -1, \theta_3 = -8$; Monte Carlo Times: 1000

Table II: Monte Carlo Results for AC						
Parameter	Mean	Median	Standard Deviation	Mean Bias	Median Bias	MSE
			$T = 25$			
θ_1	1.0409	1.0325	0.1296	0.0409	0.0325	0.0184
θ_2	-1.0618	-1.0385	0.2671	-0.0618	-0.0385	0.0751
θ_3	-8.1269	-8.7044	0.8386	-0.1269	-0.7044	0.7186
			$T = 50$			
θ_1	1.0328	1.0270	0.1299	0.0328	0.0270	0.0179
θ_2	-1.0496	-1.0220	0.2027	-0.0496	-0.0220	0.0435
θ_3	-8.1054	-8.1655	0.8017	-0.1054	-0.1655	0.6532
			$T = 100$			
θ_1	1.0109	1.0157	0.0663	0.0109	0.0157	0.0045
θ_2	-1.0050	-1.0027	0.0907	-0.0050	-0.0027	0.0082
θ_3	-8.1199	-8.1546	0.6529	-0.1199	-0.1546	0.4402
			$T = 200$			
θ_1	1.0038	1.0137	0.0666	0.0038	0.0137	0.0044
θ_2	-1.0010	-1.0027	0.0805	-0.0010	-0.0027	0.0065
θ_3	-8.0401	-8.0788	0.6129	-0.0401	-0.0788	0.3770
			$T = 400$			
θ_1	1.0103	1.0164	0.0655	0.0103	0.0164	0.0043
θ_2	-1.0096	-1.0089	0.0750	-0.0096	-0.0089	0.0057
θ_3	-8.0817	-8.1049	0.5677	-0.0817	-0.1049	0.3287

True value: $\theta_1 = 1$, $\theta_2 = -1, \theta_3 = -8$; Monte Carlo Times: 1000

5 CONCLUSION

In this paper, we developed the simulation based estimation for the discrete sequential move game of perfect information, which relies on the simulated moments and importance sampling. We use importance sampling techniques not only to reduce computational burden and simulation error, but also to overcome non-smoothness problems. Monte Carlo evidence demonstrates that the estimator can perform well in moderately-sized samples. The most limitation of our estimation is that it relies on the known distribution of random preference shocks which is rarely known to researchers, working with the unknown $G(\epsilon|\beta)$ is an important topic for future research. Another interesting issue concerns the efficient estimation of the simulated conditional moments, although in a full parametric model, we can make use of the first order condition of the likelihood function, the simulated score may not exist since the simulated choice probability can be zero in some random draws.

References

Aguirregabiria, V. & P. Mira (2007). "Sequential Estimation of Dynamic Discrete Games." *Econometrica* 75 : 1-53.

Ackerberg, D. A. (2009). "A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation." *Quantitative Marketing and Economics* 7 (4): 343–376.

Bajari, P. & H. Hong, et al. (2010). "Estimating Static Models of Strategic Interactions." forthcoming *Journal of Business and Economic Statistics*.

Bajari, P. Han Hong & Stephen Ryan (2010). "Identification and Estimation of a Discrete Game of Complete Information." forthcoming *Econometrica*.

Berry, Steven T. (1992). "Estimation of a Model of Entry in the Airline Industry." *Econometrica* 60 (4): 889-917.

Berry, Steven T. & Tamer, E. (2007). "Identification in Models of Oligopoly Entry." in *Advanced in Economics and Econometrics: Theory and Application*, vol. II, chap. 2, pp.46-85. Cambridge University Press, Ninth World Congress.

Bresnahan, T. F. & P. C. Reiss (1990). "Entry in Monopoly Markets." *Review of Economic Studies* 57 : 531-553.

Bresnahan, T. F. & P. C. Reiss (1991). "Empirical models of Discrete Games." *Journal of Econometrics* 48 : 57-81.

Brooke, A., D. Kendrick, A. Meeraus, & R. Raman (2008). GAMS - A User's Guide, Washington, DC: GAMS Development Corporation. Online at: http://www.gams.com/dd/docs/bigdocs/GAMSUsersGuide.pdf.

Ciliberto & Tamer, E. (2009). "Market Structure and Multiple Equilibria." *Econometrica*, 77 (6): 1791-1828.

Dominguez, M. A. & I. N. Lobato (2004). "Consistent Estimation of Models Defined by Conditional Moment Restrictions." *Econometrica* 72 (5): 1601-1615.

Christian Gouriéoux & Alain Monfort (2002). *Simulation-based econometric methods.* Oxford university press.

Haile, P., & A. Hortacsu, et al. (2003). "On the Empirical Content of Quantal Response Equilibrium." Working Paper.

Hansen, L. P. (1982). "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50 (4): 1029-54.

Aradillas-Lopez (2007). "Pairwise Difference Estimation of Incomplete Information Games." Working Paper.

Aradillas-Lopez (2008). "Semiparametric Estimation of a Simultaneous Game with Incomplete Information." Working Paper.

Maruyama, S. (2009). "Estimating Sequential-move Games by a Recursive Conditioning Simulator." Working Paper.

Mazzeo, Michael J. (2002). "Product Choice and Oligopoly Market Structure." *Rand Journal of Economics* 33 (2): 221-242.

Martin J. Osborne & Ariel Rubinstein. (1994). *A Coure in Game Theory.* The MIT Press.

McFadden, D. (1989). "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration." *Econometrica* 57 (5): 995-1026.

Kitamura, Y. & G. Tripathi, et al. (2004). "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models." *Econometrica* 72 : 1667 - 1714.

Klaus Ritzberger (2002). *Foundations of Non-Cooperative Game Theory.* Oxford University Press.

Pakes, A. & D. Pollard (1989). "Simulation and the Asymptotics of Optimization Estimators." *Econometrica* 57 (5): 1027-57.

Schmidt-Dengler, P. (2006). "The Timing of New Technology Adoption: The Case of MRI." Working Paper.

Seim,. K. (2005). "An Empirical Model of Firm Entry with Endogenous Product-Type Choices." *Rand Journal of Economics* 37 (3): 619-640.

Tamer, E. (2003). "Coordination Games, Multiple Equilibria and the Timing of Radio Commercials." Working Paper.

Wang, Yafeng & Graham, Brett (2009). "Generalized Maximum Entropy Estimation of Discrete Sequential Move Games of Perfect Information." Working Paper.