AUSTRALIAN SCHOOL OF BUSINESS™
THE UNIVERSITY OF NEW SOUTH WALES

The University of New South Wales
Australian School of Business

School of Economics Discussion Paper: 2010/22

Estimation of Finite Sequential Games

Shiko Maruyama

School of Economics
Australian School of Business
UNSW Sydney NSW 2052 Australia
http://www.economics.unsw.edu.au

ISSN 1837-1035
ISBN 978-0-7334-2966-8

Estimation of Finite Sequential Games

Shiko Maruyama*

October 18, 2010

Abstract

I study the estimation of finite sequential games with perfect information. The major challenge in estimation is computation of high-dimensional truncated integration whose domain is complicated by strategic interaction. I show that this complication resolves when unobserved off-the-equilibrium-path strategies are controlled for. Separately evaluating the likelihood contribution of each subgame perfect strategy profile that rationalizes the observed outcome allows the use of the GHK simulator, the most widely used importance-sampling probit simulator. Monte Carlo experiments demonstrate the performance and robustness of the proposed method, and confirm that misspecification of the decision order leads to underestimation of strategic effect.

KEYWORD: Inference in discrete games, sequential games, Monte Carlo integration, GHK simulator, subgame perfection, perfect information

*Comments from Victor Aguirregabiria, Han Hong, Susumu Imai, and Robert Porter significantly improved the paper at various stages. Earlier versions of the paper were circulated under the title "Estimating Sequential-Move Games by a Recursive Conditioning Simulator." School of Economics, University of New South Wales. E-mail: s.maruyama@unsw.edu.au

1 INTRODUCTION

Sequential games are a standard tool to investigate sequential strategic interactions such as first-mover (dis-)advantages and strategic precommitment. The predetermined order of moves provides players an opportunity to strategically commit their decision that is not revertible once the decision is made. Preemptive behavior to deter a rival's entry is a classical textbook example. A clear sequence and strategic interactions are also observed in heavily regulated industries, decisions among siblings, organizational decision making, judicial cases, labor disputes, drafts in sports leagues, parlor and TV show games, and so on.

While countless theoretical studies on sequential games exist, there has been little empirical work devoted to quantifying the relevance and implications of sequential interaction. This presumably reflects not a lack of interest in this important topic, but rather the considerable computational challenges. Existing empirical studies that consider sequential games, whether exclusively or in addition to simultaneous games, range over the entry of firms (Bresnahan and Reiss (1991), Berry (1992), Mazzeo (2002), Maruyama(2011)), technology adoption (Schmidt-Dengler (2006)), the labor participation of couples (Kooreman (1994) and Hiedeman (1998)), the retirement behavior of elderly couples (Jia (2005)), the location choice of siblings (Konrad, Kunemund, Lommerud, and Robledo (2002)), political science and international relations (Bas, Signorino, and Walker (2008) and Signorino and Tarar (2006)), tax competition (Redoano (2007)), and the validity and limit of subgame perfection in experimental economics (Andreoni and Blanchard (2006)). All of the existing literature has so far focused on simple cases where: the number of players is very small (two in most cases); the game structure is very simple; a certain degree of symmetry is assumed among players; or emphasis is not on the structural estimation of strategic effect.[1]

In this paper, I propose a practical estimation method for discrete-choice pure-strategy sequential games with perfect information, in which each player makes a decision in publicly known exogenous decision order. The econometrician knows the decision order either from institutional knowledge, by assumption, or from observation of data, and imposes the sequential structure onto an econometric model to draw inferences on payoff function and the nature of strategic interaction. Table 1 shows an example of the typical data structure for which the method proposed in this paper is intended to be used. This data set on young adult siblings records their tertiary education decision at the age of 18. Within each family, the decision is first made by the firstborn sibling, followed by the secondborn and so on. Among siblings, strategic interaction may be at work. Following an older sibling's path may facilitate the decision and positively affect motivation and future learning, generating strategic complementarity. At the same time, limited parental resources may create strategic substitution effects among their decisions. The sequential structure may generate first-mover (dis-)advantages.

In the proposed framework, I assume a payoff function with random components that follow normal distribution so that the game almost surely has a unique subgame perfect equilibrium. By assuming a parametric model of payoffs and random components, estimation relies on the maximum likelihood principle. Using micro data on heterogeneous players and

[1] An exception is Maruyama (2011), which is based on the approach outlined in this paper.

Table 1: DATA EXAMPLE: DECISION ON TERTIARY EDUCATION AMONG SIBLINGS

Family ID	Age	Sex	Parental income	Tertiary education decision at 18
1	25	M	High	University - Law
1	22	M	High	University - Law
2	21	M	Low	Work
3	26	F	Middle	University - Arts
3	24	F	Middle	University - Overseas
3	22	M	Middle	University - Arts
4	23	M	Low	University - Engineering
4	20	F	Low	Stay home
5	25	F	Middle	Work
⋮	⋮	⋮	⋮	⋮

observed decisions of players, the econometrician aims to: make a statistical inference on the payoff function of players by utilizing the identification power provided by the sequential structure; examine the nature of strategic interaction; evaluate its implications for resource allocation; and conduct counterfactual simulations.

The major challenge in estimation is the computation of high-dimensional truncated integration whose domain is complicated by sequential strategic interaction. If the game is a binary choice game played by two players, the dimension of random components in a market is two after normalization and its estimation is straightforward. The likelihood for a particular observed game outcome is analytically solved by using backward induction and bivariate normal distribution function. As the number of players and the size of the choice set grows, however, the dimension exceeds three and the likelihood function in general no longer has an analytical solution, thus requiring simulation techniques that approximate high-dimensional truncated integrals. For high-dimensional integration in standard probit models, the most popular solution is the Geweke-Hajivassiliou-Keane (GHK) simulator. This importance-sampling simulator recursively truncates the multivariate normal probability density function, by decomposing the multivariate normal distribution into a set of univariate normal distribution using Cholesky triangularization. However, sequential strategic interaction causes interdependence of truncation thresholds, which undermines the ground of the recursive conditioning approach.

I propose the use of the GHK simulator for each subgame perfect strategy profile that rationalizes the observed equilibrium outcome. I show that the interdependence of truncation thresholds in the integration domain stems from changes in off-the-equilibrium-path strategies, which are counterfactuals and, from the econometrician's viewpoint, are the source of the indeterminacy of the strategy profile that yields the observed game outcome. Thus, the separate evaluation of likelihood contribution for each subgame perfect strategy profile allows us to control for unobserved off-the-equilibrium-path strategies so that the domain of Monte Carlo integration becomes (hyper-)rectangle and the recursive conditioning of the GHK simulator can be used.

To demonstrate the performance and robustness of the proposed estimation method, I conduct Monte Carlo experiments using artificial data generated for a simple airline industry entry game inspired by Berry (1992). The data consists of 3,000 city-pair markets. The number of firms varies across markets from one to six and the error component has strong within-market correlation. Overall, when the estimated model is correctly specified, the simulation bias inherent in the method of simulated likelihood tends to be small and the model parameters are estimated fairly quickly and precisely with a small number of simulation draws such as twenty. I also conduct a number of misspecification experiments. The size of the bias due to misspecification of the decision order depends on the extent to which the econometrician imposes the correct decision order. While misspecifying the decision order by up to 10 percent of observations does not lead to significant bias, misspecification of the decision order in general leads to significant downward bias of the estimate of strategic effect. In addition, the misspecification also leads to bias of the coefficient estimates of variables that are correlated to the true decision order. Imposing independent univariate normal distribution for each error term allows researchers to avoid high-dimensional integration, but the result shows that ignoring the existent covariance structure is another potential source of significant bias.

The structural estimation of non-cooperative discrete games has rapidly developed since the seminal works by Bjorn and Vuong (1984) and Bresnahan and Reiss (1991).[2] Recent development has centered around the estimation of dynamic games (Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2007), and Su and Judd (2008)). However, the recent development of dynamic game estimation mostly focuses on Markov perfect equilibrium in infinite repetition of simultaneous move games. By contrast, despite its voluminous theoretical counterpart, the empirical analysis of sequential games has so far attracted limited attention and its existing literature is confined to a simple framework or reduced-form analysis. No study has so far investigated the estimation of a general class of asymmetric sequential

[2] For example, see Bajari, Hong, Krainer, and Nekipelov (2007) for incomplete information games, and Ciliberto and Tamer, (2009), Bajari, Hong, and Ryan (2008), Soetevent and Kooreman (2007) for complete information games.

games.[3]

After formally presenting the setup in the next section, I explain in Section 3 how the GHK simulator can aid high-dimensional integration under subgame perfection. Monte Carlo experiments are conducted in Section 4. Section 5 provides discussions and extensions. Section 6 concludes.

2 MODEL

2.1 The Sequential Game

The model is a finite sequential game with perfect information. There are $i = 1, ..., N$ players, each makes a decision in publicly known exogenous order. The game can be set up so that players take multiple turns alternately. Each player chooses an "action" a_i from a finite set of actions A_i, e.g. ("left", "right") and ("enter", "not enter").[4] Define $A \equiv \times_i A_i$ and let $a \equiv (a_1, ..., a_N)$ denote a generic element of A. Player i's payoff, such as utility or profit, from action a_i depends on a_{-i}, the vector of actions taken by the other players. Thus the payoff function of player i is a map $\pi_i : A \to \mathbb{R}$. The payoff of player i given a is

$$\pi_i (a, x, \varepsilon_i; \theta_1) = \overline{\pi}_i (a, x; \theta_1) + \varepsilon_i (a_i), \tag{1}$$

[3] This paper also builds upon the literature on sequential discrete choice models. Widely used nested logit models and sequential multinomial models incorporate the sequence of decision making by one agent. The method I propose here extends these models by introducing sequential decision making by multiple agents.

[4] In many potential empirical applications, every player faces the same choice set A_i. regardless of another player's decision. However, allowing the choice set to vary across decision nodes is possible by defining A_i as a union of available alternatives at each decision node.

where vector x contains exogenous characteristics that describe players and the environment in which the game is played and θ_1 is a vector of parameters. The first term, $\overline{\pi}_i(a, x; \theta_1)$, is an assumed parametric function of mean payoffs. The second term, $\varepsilon_i(a_i) \in \mathbb{R}$, is a random preference shock with continuous parametric density function, $g_i(\varepsilon_i(a_i); \theta_2)$, where θ_2 is a vector of parameters.[5] Define $\varepsilon_i \equiv \{\varepsilon_i(a_i)\}_{a_i \in A_i}$ and $\varepsilon \equiv (\varepsilon_1, ..., \varepsilon_N)$. Both x and ε are common knowledge to the players, but the econometrician observes only x, not ε.

All the game theoretical concepts used in this paper are textbook standard, except for "action profile", a, defined above, which records decisions made on the equilibrium path and corresponds to what the econometrician observes as a game outcome in data, whether the game is sequential or simultaneous. An extensive form game is a *perfect information* game if every information set is a singleton decision node. With perfect information, every decision made earlier is observable for the following players. Player i's (pure) *strategy*, $s_i \in S_i$, specifies her decision at each decision node.[6] Define $S \equiv \times_i S_i$ and let $s \equiv (s_1, ..., s_N) \in S$ denote a *strategy profile.* Since s uniquely determines a game outcome, define $a(s) : S \to A$ and $a_i(s) : S \to A_i$. A *subgame* of an extensive form game with perfect information is a subset of the game that begins with a single decision node, contains all the decision nodes that are successors of this node, and contains only these nodes. A *subgame perfect equilibrium*, s^e, is a strategy profile in which each player's strategy is the best response to the strategies of the other players in *every* subgame. It is a well-known fact that every finite game with

[5] I assume the additive separability of the random shock term following much of the existing literature, such as Bresnahan and Reiss (1991). In the following discussion, this assumption is not essential as long as the identification of parameter estimates is established.

[6] Incorporating mixed strategies in the present framework is computationally impractical and beyond the scope of this paper.

perfect information has a pure strategy subgame perfect equilibrium (Zermelo's theorem). In addition, in the present setup, the game almost surely has a unique equilibrium. Denote this subgame perfect equilibrium, $s^e(x, \varepsilon; \theta_1)$ and its i'th component, $s_i^e(x, \varepsilon; \theta_1)$. An *equilibrium outcome function* is also defined as $a^e(x, \varepsilon; \theta_1) \equiv a(s^e(x, \varepsilon; \theta_1))$, with its i'th component, $a_i^e(x, \varepsilon; \theta_1)$. Given $(x, \varepsilon, \theta_1)$, the game can be solved to obtain s^e by backward induction.

2.2 Data

The econometrician observes T independent realizations of the game, $(\Gamma_1, ..., \Gamma_T)$, e.g., T different markets, T different families, or T periods of time. I index each realization of the game by t. The structure and environment of the game may vary across t in terms of the number and identity of players, the choice set of each player, the decision order, and covariates x. The parametric forms of $\pi_i(a_t, x_t, \varepsilon_{it}; \theta_1)$ and $g_i(\varepsilon_i(a_{it}); \theta_2)$ and parameters, $\theta \equiv (\theta_1, \theta_2)$, are assumed to be invariant across t to draw statistical inferences. In each t, the econometrician observes equilibrium outcome a_t^o and covariate vector x_t. Equilibrium strategy s_{it} is not observed as it contains counterfactuals. The econometrician knows the structure of game Γ_t, such as the number of players and the decision order. In the following I drop the subscript for each game, t, when no ambiguity arises.

To utilize a probit simulator below, I assume a normal distribution for ε_t as

$$\varepsilon_t \sim N(0, \Omega). \tag{2}$$

Covariance matrix Ω has dimension of $\Pi_{i=1}^N$ [the number of action alternatives for i] and is parameterized by θ_2. For the parameterization of Ω, the usual identification conditions of probit models apply. In particular, the fact that payoff π_{it} is an unobserved latent construct means that what the econometrician can infer from observed decisions concerns only the relative comparison of payoffs among alternatives and, consequently, two types of normalization for ε are required. First, the random shock of an alternative is normalized to zero so that the interpretation of ε_t is the relative difference in random shocks between the normalized alternative and other alternatives. Second, the variance of ε is also not identified. Following the convention, it is normalized to one.[7] Below I abuse notation and use ε and Ω to denote the error structure after normalization.

2.3 Estimation and the High-Dimensional Integration

The task of the econometrician is to make statistical inferences on θ based on the structure of game Γ_t and the assumed parametric forms of $\pi_i\left(a, x, \varepsilon_i; \theta_1\right)$ and $g_i\left(\varepsilon_i\left(a_i\right); \theta_2\right)$. Since the distribution of ε is specified fully parametrically, the estimation procedure relies on the maximum likelihood principle. Game Γ_t is the unit for which individual likelihood is defined. The individual likelihood is defined as

$$l\left(\theta; x_t, a_t^o\right) = \Pr\left[a_t^o = a_t^e(x_t, \varepsilon_t; \theta_1) | \theta_2\right]. \tag{3}$$

[7] In some applications, information on the level of payoffs is available and aids identification, typically making the normalization of the variance of error terms unnecessary. For example, in the analysis of entry decisions of health insurance plans, Maruyama (2011) uses equilibrium variable profits that are recovered from the demand estimation and the level of fixed costs is identified.

This leads to the following maximum likelihood problem:

$$\widehat{\theta}_{ML} = \arg\max_{\theta} \left[\frac{1}{T} \sum_{t}^{T} \ln l \left(\theta; x_t, a_t^o \right) \right]. \tag{4}$$

The challenge in this maximum likelihood framework is that the probability term in the likelihood involves high-dimensional integrals and generally does not have an analytical solution. The dimension depends on the number of players and the number of alternatives each player has.[8] There are several cases where this likelihood function is easily computed. First is the two dimensional case (Stackelberg games), which arises, for example, if the number of players is two and the number of alternatives is two. The econometrician can then solve the two threshold values for $(\varepsilon_{1t}, \varepsilon_{2t})$ in accordance with the observed equilibrium outcome, a_t^o. The bivariate normal distribution function then produces an analytical solution for the probability term. The vast majority of the existing literature on sequential games focuses on the two player case. If the dimension of integration increases to three, an analytical solution is generally not available, but the quadrature method enables numerical approximation. Another special case is when each stochastic component in ε_t follows an independent univariate normal distribution. In most applications, this is a strong assumption. It implies no game specific error (e.g. market specific random component). It also implies a quite restrictive substitution pattern among alternatives when the choice set is larger than the binary case.

For high-dimensional integration, the literature has developed the maximum simulated

[8]The dimension also depends on the number of turns each player has, if multiple decisions are assumed.

likelihood (MSL) method, which utilizes Monte Carlo integration.[9] The most straightforward simulator for MSL is the crude frequency simulator, first proposed by Lerman and Manski (1981). The simulator for the current setup is given by

$$\begin{aligned} \widehat{\theta}_{CF} &= \arg\max_{\theta} \left\{ \frac{1}{T} \sum_{t=1}^{T} \ln \widehat{l}_R^{CF} (\theta; x_t, a_t^o) \right\} && (5) \\ &\equiv \arg\max_{\theta} \left\{ \frac{1}{T} \sum_{t=1}^{T} \ln \frac{1}{R} \sum_{r=1}^{R} I \left[a_t^o = a_t^e(x_t, \widetilde{\varepsilon}_t^r; \theta_1) | \theta_2 \right] \right\}, && (6) \end{aligned}$$

where $I[\;]$ denotes an indicator function. The simulation procedure takes R sets of random draws from the assumed distribution. For each random draw $\widetilde{\varepsilon}_t^r$, an equilibrium outcome a_t^e is solved by backward induction. The probability simulator is based on, out of R times repetition of simulation draws, how many times the predicted equilibrium outcome coincides with the observed equilibrium outcome. Although this simulator provides estimates that are consistent with R and T, the simulated probability is a discontinuous function of the parameters and is not bounded away from 0 and 1. The use of the indicator function makes its variance quite large. Due to these problems, Lerman and Manski find that their estimator requires a very large number of simulations for satisfactory performance.[10] Since a likelihood evaluation of relatively large asymmetric extensive form games tends to be quite expensive, the frequency simulator is practically infeasible.

[9] The method of simulated moments (MSM) and the method of simulated scores (MSS) are alternative options. These may improve the finite sample property of estimators by removing the simulation bias that results from the logarithm in the log likelihood function (Hajivassiliou and McFadden (1998)), though Geweke, Keane, and Runkle (1994) does not find such an advantage of MSM over MSL.

[10] Moreover, the discontinuity of the likelihood function requires an optimization method that does not require differentiability of the optimand, such as the nonlinear simplex algorithm of Nelder and Mead (1964).

2.4 The GHK Simulator

For high-dimensional integration over a region of the multivariate normal, the most popular simulator is the GHK simulator, due to Geweke (1992), Hajivassiliou and McFadden (1994), and Keane (1994). The GHK simulator recursively truncates the multivariate normal probability density function. Its algorithm draws recursively from truncated univariate normal distributions, and relies on Cholesky triangularization to decompose the multivariate normal distribution into a set of univariate normal distributions. The combination of the recursive conditioning approach and the smooth univariate truncated variate generation algorithm produces an unbiased and smooth importance sampling simulator. Compared to the frequency simulator, it requires many fewer draws for alternatives with low probability of being chosen and is unlikely to have boundary problems. A number of studies have confirmed its usefulness and relative accuracy, especially when considering the low computational effort required (Hajivassiliou, McFadden, and Ruud (1996), Hajivassiliou and McFadden (1998), Geweke, Keane, and Runkle (1994), Börsch-Supan and Hajivassiliou (1993)).

The complication in using the GHK simulator for sequential games arises from the recursive conditioning approach. The GHK algorithm repeats recursive simulation draws from truncated univariate normal distributions so that the resulting random shocks, $\widetilde{\varepsilon}^r$, generate equilibrium outcome a^o, which is observed by the econometrician. The requirement for this recursive conditioning is that, in the ε space, the truncation threshold for each simulation draw is independent of other simulation draws and the truncation thresholds are orthogonal to each other. However, because of sequential strategic interaction, the truncation threshold for a draw may depend on other simulation draws. Due to this dependence of truncation thresholds on random shocks of other players, recursive conditioning simulation breaks down.[11]

3 USING THE GHK SIMULATOR

The problem of interdependent truncation thresholds arises as a result of changes in unobserved off-the-equilibrium-path strategies. This point is best illustrated by an example entry game that is played by two players, firm 1 and firm 2.[12] Firm 1 is the Stackelberg leader. Firm 2 makes its entry decision having observed firm 1's entry decision. Firms 1 and 2 incur random shocks ε_1 and ε_2 respectively in their profit functions. The rival's entry reduces payoff. Each firm enters the market when it expects nonnegative profits from entry. If not enter, a firm earns zero profit. Hence four possible market configurations exist, and given the assumed payoff functions, the realized values of ε_1 and ε_2 determine which market outcome occurs (Figure 1). A firm with a larger random shock is more likely to enter the market. However, the effects of ε_1 and ε_2 are not symmetric and the decisions of the two firms are not independent of each other, due to the sequential nature of the game. The center part of Figure 1 shows the asymmetry; when neither ε_1 nor ε_2 has dominant influence, only firm 1 enters.

The goal of this paper is to establish a computationally practical Monte Carlo integration

[11] Chernew, Gowrisankaran, and Fendrick (2002) use the GHK simulator in their entry model of hospitals, but strategic interactions are not explicitly modeled in their empirical specification.

[12] This entry game is only for explanation purposes, as its likelihood function can easily be solved analytically.

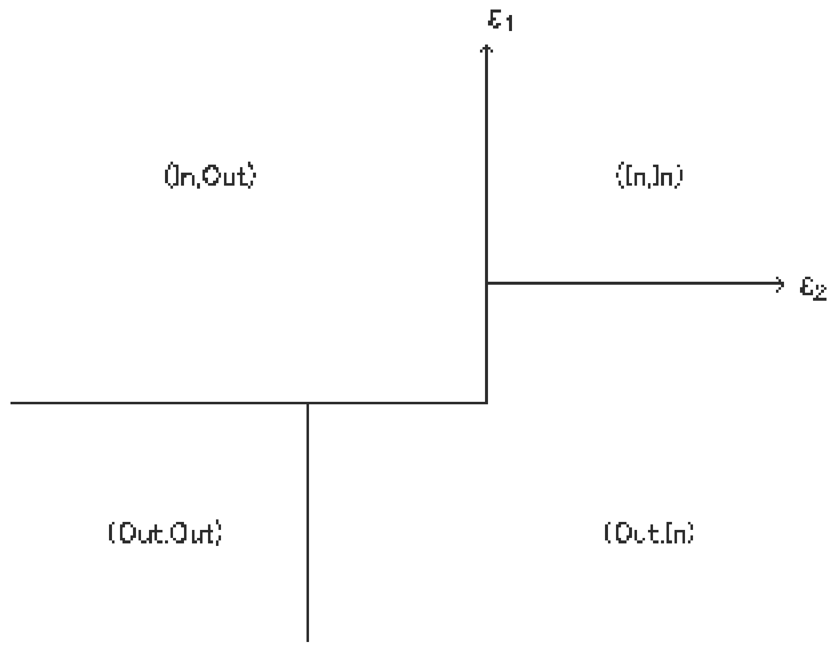

Figure 1: TWO PLAYER STACKELBERG ENTRY GAME

method to evaluate the probability for each market outcome in the likelihood function. Figure 2 illustrates this task by superimposing the probability density function of ε_1 and ε_2. In this example, market configuration (Out,In) does not allow the use of the standard GHK simulator, because the domain of integration is not a rectangle, and thus drawing ε_1 cannot be conditional on ε_2 and vice-versa.

The notion of subgame perfection solves this dependency. Indeed, this non-rectangular shaped domain of integration stems from a behavioral change in an off-the-equilibrium path. The strategic interaction in this sequential game is illustrated by its extensive form (Figure 3). With perfect information, firm 2 has two singleton decision nodes, and the choice set of firm 2 consists of four strategies: "never enter", "imitate", "preempted", and "always enter". Since firm 1 has two alternatives, "In" and "Out", there are eight combinations as a whole.

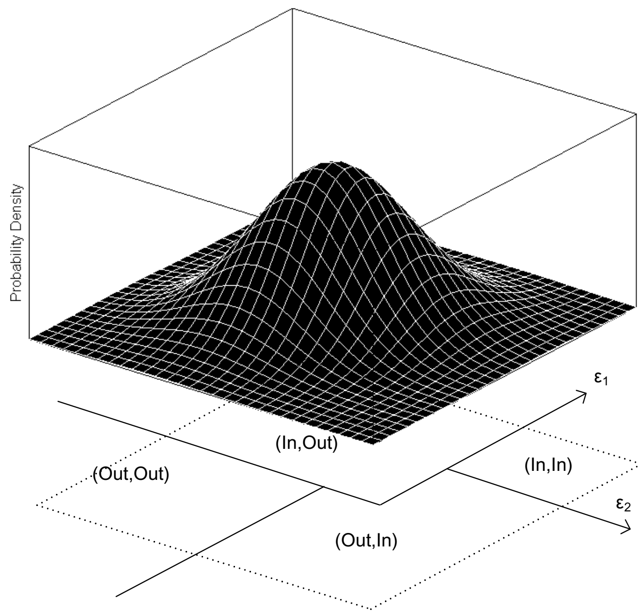

Figure 2: INTEGRATION WITH NORMAL DENSITY FUNCTION

Figure 3 assumes "Out" for firm 1 and shows four equilibrium profiles. The extensive form highlights several important facts. First, subgame perfection implies that firm 2 chooses the best option based on its random shock, ε_2, *irrespective of* ε_1. Facing a large negative shock, firm 2 chooses "never enter". For a large positive shock, firm 2 chooses "always enter". For a medium value of ε_2, firm 2 chooses "preempted", i.e. it enters the market only if firm 1 does not. Due to the assumed negative impact of a rival's entry, firm 2 never chooses the "imitate" strategy. Secondly, different strategy profiles may generate game outcomes that are observationally equivalent to the econometrician. In Figure 3, strategy profiles (3) and (4) both result in (Out,In). However, the two strategies of firm 2 under (3) and (4) have different implications for firm 1's decision. When preemption is possible, the entry threshold for firm 1 is lower and the integration domain of ε_1 is larger.

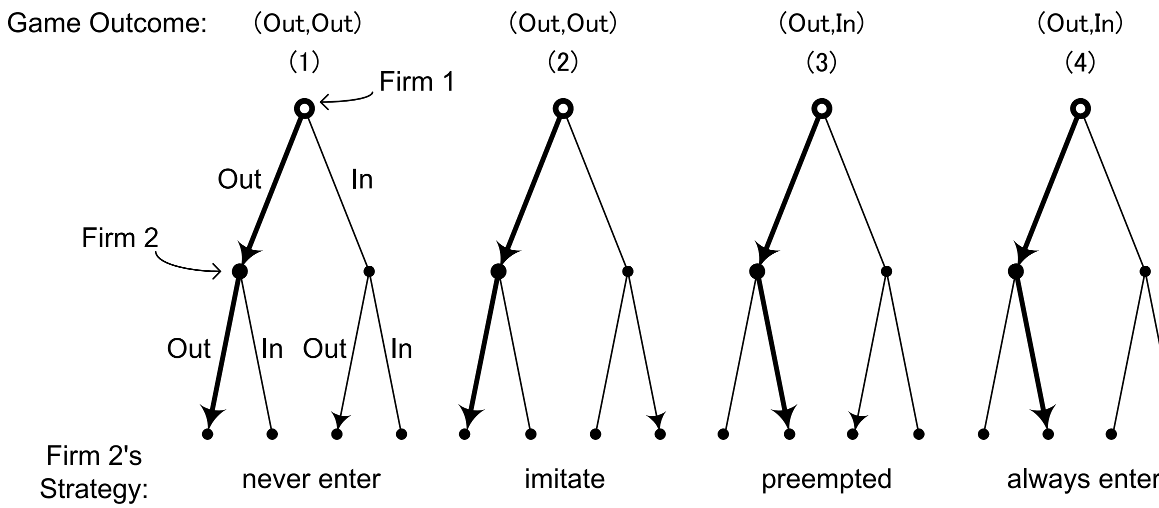

Figure 3: STRATEGIES AND OUTCOMES IN THE EXTENSIVE FORM WHEN FIRM 1 CHOOSES "OUT"

Figure 4 incorporates these considerations into the $(\varepsilon_1, \varepsilon_2)$ space. Now the (Out,In) area is divided into two rectangles, each representing different strategy profiles, i.e. (3) "preempted" and (4) "always enter" as named in Figure 3. The standard GHK procedure works as long as the domain of integration is rectangular, or hyperrectangular in a general n-dimensional space, and therefore, we can simulate the likelihood function by evaluating each subgame perfect equilibrium separately.

To formalize the discussion so far in the general n-dimensional case, let s_{-i} denotes the subvector of strategy profile s that excludes component i, and let $s_i^{BR}(x, \varepsilon_i, s_{-i}; \theta_1)$ denotes the function that determines the best response strategy of player i given x, ε_i, and s_{-i}. Then, the following result holds.

Proposition 1 *For any strategy profile $s^* \in S$,*

$$\{\varepsilon | s^e(x, \varepsilon; \theta_1) = s^*\} = \times_i \left\{\varepsilon_i | s_i^{BR}\left(x, \varepsilon_i, s_{-i}^*; \theta_1\right) = s_i^*\right\}.$$

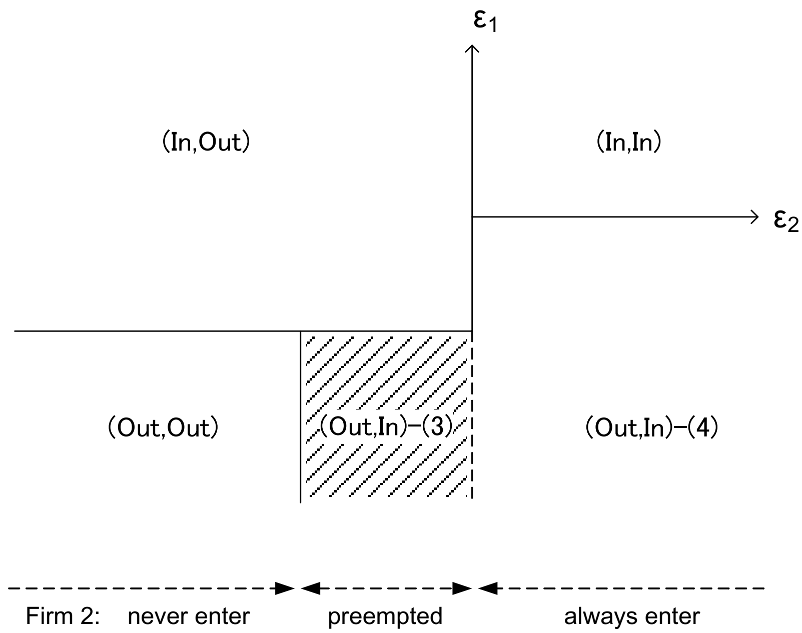

Figure 4: DIVIDING AN OBSERVED MARKET OUTCOME INTO STRATEGY PROFILES

Proof. Player i's best response strategy is uniquely determined by s_{-i}, x, and ε_i. Thus, given s^*_{-i} and x, the set of ε_i under which s^*_i is the best response strategy to s^*_{-i} does not depend on another player's component of ε. Therefore, the set of ε under which s^* solves the game as a subgame perfect equilibrium can be written as a Cartesian product of each player's set of ε_i under which s^*_i is the best response strategy to s^*_{-i}. ■

The logic underlying this proposition comes directly from the Nash equilibrium concept, not specifically from subgame perfection. What is important here is its implication in empirical sequential games. When the econometrician ignores subgame perfection and only considers observed actions, a^o, the realized value of ε_j may change player j's off-the-equilibrium-path decisions, which in turn affects the set of ε_i under which player i chooses a^o_i on her equilibrium path. In contrast, the proposition clarifies that this interdependency across players does not occur at the level of subgame perfect strategy profiles.

The main virtue of the proposition is that for any observed market outcome, a^o, by dividing the integration problem into the subgame perfect equilibria that rationalize a^o, the interdependency of integral intervals across players resolves and the standard GHK procedure can be used. Specifically, to obtain $\widehat{\theta}_{ML}$ using Monte Carlo integration, the estimation procedure evaluates the GHK simulator for every strategy profile that rationalizes observed outcome a_t^o. Let $S^o(a) \equiv \{s \in S | a(s) = a\}$. Rewrite the individual likelihood in the original maximum likelihood problem, (3), as

$$
\begin{aligned}
l(\theta; x, a^o) &= \Pr\left[a^o = a^e(x, \varepsilon; \theta_1) | \theta_2\right] \\
&= \sum_{s \in S^o(a^o)} \Pr\left[s = s^e(x, \varepsilon; \theta_1) | \theta_2\right].
\end{aligned}
$$

The second equality holds owing to the fact that any ε leads to a unique subgame perfect equilibrium. The GHK simulator is used to evaluate $\Pr\left[s = s^e(x, \varepsilon; \theta_1) | \theta_2\right]$ for each $s \in S^o(a^o)$, following the standard procedure. The rest of this section sets out the procedure.

The probability that the event, $s = s^e(x, \varepsilon; \theta_1)$, occurs can be rewritten using an integral. Let $n(\varepsilon, \Omega)$ denote the probability density function of the multivariate normal variates, ε, with zero mean and covariance matrix Ω. Then

$$
\begin{aligned}
\Pr\left[s = s^e(x, \varepsilon; \theta_1) | \theta_2\right] &= \int I\left[s = s^e(x, \varepsilon; \theta_1)\right] n(\varepsilon, \Omega(\theta_2))\, d\varepsilon \\
&= \int \prod_i I\left[s_i = s_i^{BR}(x, \varepsilon_i, s_{-i}; \theta_1)\right] n(\varepsilon, \Omega(\theta_2))\, d\varepsilon.
\end{aligned}
$$

The last equality holds from the proposition. Covariance matrix $\Omega(\theta_2)$ takes a parametric form of θ_2 that allows identification. Defining a set $\Delta_i(x, s; \theta_1) \equiv \left\{\varepsilon_i | s_i^{BR}(x, \varepsilon_i, s_{-i}; \theta_1) = s_i\right\}$,

$$\Pr[s = s^e(x, \varepsilon; \theta_1) | \theta_2] = \int \prod_i I[\varepsilon_i \in \Delta_i(x, s; \theta_1)] \, n(\varepsilon, \Omega(\theta_2)) \, d\varepsilon.$$

The set $\Delta_i(x, s; \theta_1)$ represents the conditions that random shocks ε_i needs to satisfy for s_i to be player i's best response given s_{-1}. The derivation of $\Delta_i(x, s; \theta_1)$ is based on finding thresholds of ε_i by comparing payoffs across available strategies given s_{-1}. There may be a strategy that is dominated by another strategy regardless of the value of ε_i. For such a dominated strategy s_i, $\Delta_i(x, s_i, s_{-i}; \theta_1) = \emptyset$, and strategy profile s that contains s_i occurs with probability zero. Define $\overline{S}^o(a^o, \theta_1) \subset S^o(a^o)$ as the set of strategy profiles each element of which leads to market outcome a^o and occurs with positive probability. Then the likelihood function becomes

$$\begin{aligned} l(\theta; x, a^o) &= \sum_{s \in S^o(a^o)} \Pr[s = s^e(x, \varepsilon; \theta_1) | \theta_2] \\ &= \sum_{s \in \overline{S}^o(a^o, \theta_1)} \Pr[s = s^e(x, \varepsilon; \theta_1) | \theta_2]. \end{aligned}$$

In the following I focus on $\overline{S}^o(a^o, \theta_1)$ so that $\Delta_i(x, s; \theta_1)$ is not the empty set.

Before applying the GHK simulator, I introduce Cholesky decomposition. For the simplicity of explanation, I assume the choice set of every player is binary. Then, after normalization, $\varepsilon \in \mathbb{R}^N$ and $\Omega(\theta_2)$ is a $N \times N$ matrix. Allowing more than two alternatives is straightforward under the GHK procedure. Denote the lower-triangular Cholesky factor of

$\Omega(\theta_2)$ as L so that $LL' = \Omega(\theta_2)$. Denote $\eta = (\eta_1, ..., \eta_N)$ an N-dimensional multivariate standard normal vector; $\eta \sim N(0, I_N)$. Hence we can write $\varepsilon = L\eta \sim N(0, \Omega(\theta_2))$. I introduce some notation to simplify the following presentation. For a vector of indexes $(1, ..., N)$, the notation "$< i$" denotes the subvector $(1, ..., i-1)$ and "$\leq i$" denotes the subvector $(1, ..., i)$. Thus, for a vector ε, $\varepsilon_{<i}$ is the subvector of the first $i-1$ components, and ε_{-i} is the subvector excluding component i. For a matrix L, L_{ii} is the i-th diagonal elements of L, and $L_{i,<i}$ and $L_{i,\leq i}$ denote vectors containing the first $i-1$ and i elements of row i, respectively. Using this notation, $\varepsilon_i = L_{i,\leq i}\eta_{\leq i}$.

Then the probability expression becomes

$$\begin{aligned} \Pr\left[s = s^e(x, \varepsilon; \theta_1)|\theta_2\right] &= \int_{\Re^N} \prod_i I\left[\varepsilon_i \in \Delta_i(x, s; \theta_1)\right] n\left(\varepsilon, \Omega(\theta_2)\right) d\varepsilon \\ &= \int_{\Re^N} \left[\prod_i I\left[L_{i,\leq i}\eta_{\leq i} \in \Delta_i(x, s; \theta_1)\right]\right] \cdot \left[\prod_i \phi(\eta_i)\right] d\eta \qquad (7) \\ &= \int_{\Re^N} \prod_i \left[I\left(L_{i,\leq i}\eta_{\leq i} \in \Delta_i(x, s; \theta_1)\right) \cdot \phi(\eta_i)\right] d\eta, \end{aligned}$$

where $\phi()$ is the probability density function of the univariate standard normal distribution.

The simulated likelihood with the GHK simulator is constructed as follows. For each simulation, $r = (1, ..., R)$, prepare an N-dimensional vector of independent uniform $(0, 1)$ random variables, $\widetilde{u}^r = (\widetilde{u}_1^r, ..., \widetilde{u}_N^r)$. For $u \in (0, 1)$ and a non-empty set $\Delta \subset \mathbb{R}$, define a function $q(u, \Delta)$ which is a mapping that takes u into a truncated standard normal distribution which ranges over Δ. For example, if $\Delta = (-\infty, a]$, then $q(\cdot)$ is a mapping into a standard normal random variate that is right-hand truncated at a, i.e. $q(u, (-\infty, a]) = \Phi^{-1}(\Phi(a) \cdot u)$, where

$\Phi(a)$ is the standard normal distribution function. For given x, s, θ_1, L, and $\widetilde{u}^r$, recursively define a sequence of simulated $\widetilde{\eta}_i^r$ so as to satisfy $s_i = s_i^{BR}(x, \varepsilon_i, s_{-i}; \theta_1)$ for $i = 1, ..., N$ as

$$
\begin{aligned}
\widetilde{\eta}_1^r &\equiv q\left(\widetilde{u}_1^r, \{\eta_1 | L_{1,1}\eta_1 \in \Delta_1(x, s; \theta_1)\}\right) \\
\widetilde{\eta}_2^r &\equiv q\left(\widetilde{u}_2^r, \{\eta_2 | L_{2,1}\widetilde{\eta}_1^r + L_{2,2}\eta_2 \in \Delta_2(x, s; \theta_1)\}\right) \\
&\quad ... \\
\widetilde{\eta}_N^r &\equiv q\left(\widetilde{u}_N^r, \{\eta_N | L_{N,<N}\widetilde{\eta}_{<N}^r + L_{N,N}\eta_N \in \Delta_N(x, s; \theta_1)\}\right).
\end{aligned}
$$

After obtaining simulated $\widetilde{\eta}^r$, the probability for ε_i to satisfy $s_i = s_i^{BR}(x, \varepsilon_i, s_{-i}; \theta_1)$, which I denote Q_i^s, is recursively calculated. For $\Delta \subset \mathbb{R}$, define $\Psi(\Delta) \equiv \int_\Delta \phi(\eta)\, d\eta$. For example, if $\Delta = (-\infty, a]$, then $\Psi(\Delta) = \Phi(a)$. Then

$$
\begin{aligned}
Q_1^s &\equiv \Psi\left(\{\eta_1 | L_{1,1}\eta_1 \in \Delta_1(x, s; \theta_1)\}\right) \\
Q_2^s\left(\widetilde{\eta}_{<2}^r\right) &\equiv \Psi\left(\{\eta_2 | L_{2,1}\widetilde{\eta}_1^r + L_{2,2}\eta_2 \in \Delta_2(x, s; \theta_1)\}\right) \\
&\quad ... \\
Q_N^s\left(\widetilde{\eta}_{<N}^r\right) &\equiv \Psi\left(\{\eta_N | L_{N,<N}\widetilde{\eta}_{<N}^r + L_{N,N}\eta_N \in \Delta_N(x, s; \theta_1)\}\right).
\end{aligned}
$$

Repeat this simulation R times for each element of $\overline{S}^o(a^o, \theta_1)$ and define the likelihood simulator as

$$
\widehat{l}_R^{GHK}(\theta; x, a^o) \equiv \sum_{s \in \overline{S}^o(a^o, \theta_1)} \frac{1}{R} \sum_{r=1}^{R} \left[Q_1^s \cdot \prod_{i=2}^{N} Q_i^s(\widetilde{\eta}_{<i}^r) \right].
$$

Using this simulator, the estimation procedure solves the following maximum simulated likelihood problem,

$$\widehat{\theta}_{MSL-GHK} = \arg\max_{\theta} \left\{ \frac{1}{T} \sum_{t}^{T} \ln \widehat{l}_R^{GHK}(\theta; x_t, a_t^o) \right\}.$$

This maximum likelihood problem is solved using numerical derivatives. In searching $\widehat{\theta}$, each iteration should use the same simulation draws $\left(\widetilde{u}^1, ..., \widetilde{u}^R\right)$ to minimize standard errors.

4 MONTE CARLO EXPERIMENTS

4.1 Experimental Design

In this section I conduct Monte Carlo experiments and demonstrate the performance and robustness of the estimation method presented in this paper. I pay particular attention to potential simulation bias in the Monte Carlo integration and robustness with respect to the decision order. The latter is especially important, as the precise decision order may not be available in many empirical applications. Inspired by Berry (1992), I employ a simple binary-choice entry game in the passenger airline industry, in which at most 6 heterogeneous airline firms compete to serve different markets.

A market, defined as a city pair route that connects major U.S. cities, constitutes the unit of observation. The six largest national carriers of differing sizes (as defined by the number of existing served routes) non-cooperatively play a sequential entry game independently in each market, based on predicted profitability in the market. The number of players in each market varies from 1 to 6. Following the early literature on static entry games, I assume a one-shot game and make no distinction between entry by new entrants and "entry" by incumbent firms. The econometrician has a cross section data set in which she observes which firms chose to enter into each market in the following year, in addition to the list of "potential entrants". Also available are variables in the base period that explain the potential profitability from entry. These variables are either at the market level, firm level, or market-firm level. In the base model, potential entrants are assumed to make their decisions in order of size, possibly reflecting advantages of access to the regulation bureaucracy and airport infrastructure.

Ten artificial data sets are generated using pseudo-random numbers. Each data set consists of 3,000 market observations and around 8,300 market-firm observations on average and contains information on the list of potential entrants, covariates, and generated random shocks in each market. Throughout all the experiments conducted in this study, I use the same ten data sets for better compatibility of the simulation results. The experiments I conduct vary in three aspects. First, I investigate the effects of changes in strategic effect and decision order. These changes in the data generating process alter market outcomes in the data, i.e. the entry decision of each firm in each market, which is generated by solving the game. Second, to check the computational performance, I examine the effects of changing the simulation framework, such as the number of simulation draws. Third, to study the effect of misspecification, I impose restrictions on estimated models.

4.2 Model and Data Generating Process

In market t, firms $(1, ..., N_t)$ play the entry game. Firm i in market t chooses to enter if it expects a non-negative profit. The expected profit from entry, π_{it}, is

$$\pi_{it}(n) = x_{it}^{'}\beta - \delta \ln(n) + \varepsilon_{it}$$

where x_{it} is a vector of covariates that are specific to either market t, firm i, or firm-market pair (i, t), ε_{it} is the firm-market specific random component, and n is the number of firms that choose to enter market t. The key parameter, δ, captures the strategic effect. For simplicity, the strategic effect is assumed to depend only on the number of competitors, not their identity. The random term ε_{it} is not observed to the econometrician but is known to every firm, and follows a multivariate normal distribution: $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{N_t,t})' \sim N(0, \Omega_t)$. The payoff when a firm does not enter is normalized to 0. The econometrician desires to learn about β, δ, and Ω based on observed entry decisions and x_{it}.

The covariate vector contains the following variables: two market-specific continuous variables, population (pop) and distance ($dist$); a firm-market specific continuous variable, past profitability in neighboring markets ($pastp$); a firm-market level dummy variable that indicates the firm's presence at both airports of the route in the previous period, $city2$; and $nroute$, a firm-specific variable for the number of existing routes in the country (in 100's) that indicates the size of each firm and determines the decision order.

Data on the pool of entrants and covariates are generated using pseudo-random numbers. For each of 3,000 markets, I first generate market population, *pop*, the number of potential entrants, $NCity1$, and the number of potential entrants with a presence at two airports, $NCity2$ based on trivariate normal distribution. These three variables are assumed to be positively correlated with covariance matrix $\begin{bmatrix} 1.0 & 0.3 & 0.3 \\ 0.3 & 1.0 & 0.6 \\ 0.3 & 0.6 & 1.0 \end{bmatrix}$. For *pop*, generated normal variable values are transformed to a log-normal variable with mean 4.0 and standard deviation 1.0. To constrain the number of players in each market between 1 and 6, the two generated normal variables are transformed into truncated normal distributions. For $NCity1$, the generated normal variable is transformed to a truncated normal variable with mean 3.0 and standard deviation 1.5 and with the truncation points at 1.0 and 7.0. Likewise, for $NCity2$, the third generated normal variable is transformed to a truncated normal variable with mean 1.5 and standard deviation 1.0 with truncation points at 0.0 and 7.0. Both variables are then rounded down to integers. To guarantee $NCity2 \leq NCity1$, $NCity2$ is replaced with the value of $NCity1$ where $NCity2 > NCity1$. The numbers of existing routes, *nroute*, are set as $(2.8, 2.5, 2.0, 1.7, 1.1, 0.75)$ for the six airlines. In each market, potential entrants are randomly chosen up to the number of $NCity1$ with probabilities proportional to *nroute*. This determines the list of players in each market. Potential entrants with a presence at both airports of the market are also randomly chosen up to the number of $NCity2$ (each firm with same probability). This generates the dummy variable, $city2$. The two remaining variables, *dist* and *pastp* are independently generated from the standard normal distribution.

The error component ε_{it} is also generated for the ten data sets and is kept fixed throughout all experiments. The covariance matrix of the error component, Ω_t, is assumed to be a $N_t \times N_t$ matrix with diagonal elements, 1.0, and off-diagonal elements, ρ^2. In other words, ε_{it} consists of two independent standard normal errors, (ν_{it}, ν_t), as

$$\varepsilon_{it} = \sqrt{(1 - \rho^2)}\nu_{it} + \rho\nu_t$$

where ρ is a correlation among the error terms within a market and ν_t measures a market-specific factor that makes entry more attractive for all firms in the market. The correlation, ρ, is set to be 0.7, which implies ν_{it} and ν_t have about the same weights in the error term.

The coefficients on (constant, $pop, dist, pastp, city2, nroute$) are set to be (−5.0, 1.2, 0.0, 0.4, 1.5, 0.0). To highlight the misspecification bias, the coefficient on firm size, $nroute$, is set to zero so that the firm size affects profits not directly, only via the decision order. Once I specify these parameter values, the value of strategic effect parameter, δ, and the decision order, I can solve the game by backward induction and obtain data on market outcome. The default specification is $\delta = 2.0$ and assumes that firms make decisions in order of $nroute$. I also conduct experiments with $\delta = 1.0$ to study the effect of the degree of strategic effect and experiments with randomized decision order to study the robustness of the proposed method with respect to decision order.

Tables 2 and 3 report descriptive numbers from one of the 10 artificial data sets as an example. Similar patterns are observed in the other data sets. The equilibrium number of entrants presented in the tables is generated with two different values of δ, 1.0 and 2.0. The majority of the 3,000 markets have two or three potential entrants. It is most likely that markets end up with one entrant, with no entrant being the second likely outcome. The higher value of δ magnifies the competitive effect and leads to fewer entrants.

Table 2: EXAMPLE OF DATA SET: DISTRIBUTION OF MARKETS BY NUMBER OF ENTRANTS

Outcome number of entrants		Number of potential entrants 1	2	3	4	5	6	Total
Total		556	794	846	528	215	61	3,000
(a) $\delta = 1.0$	0	317	244	176	59	14	6	766
	1	239	368	296	161	64	8	1,201
	2	0	182	230	140	55	9	612
	3	0	0	144	92	40	11	272
	4	0	0	0	76	24	8	104
	5	0	0	0	0	18	8	34
	6	0	0	0	0	0	11	11
(b) $\delta = 2.0$	0	317	244	176	59	14	6	816
	1	239	442	429	250	102	15	1,477
	2	0	108	187	149	66	21	531
	3	0	0	54	52	20	10	136
	4	0	0	0	18	12	4	34
	5	0	0	0	0	1	3	4
	6	0	0	0	0	0	2	2

Since the pool of potential entrants is constructed randomly but with probability proportional to firm size, firm 1 appears in the data set most frequently and firm 6 least frequently. When $\delta = 1.0$, the early-mover advantages are smaller, so the entry propensity does not vary much across firms, whereas, when $\delta = 2.0$, the larger early-mover advantages reduce the entry propensity of followers.

Table 3: EXAMPLE OF DATA SET: NUMBER OF OBSERVATIONS AND ENTRY PROFITABILITY BY AIRLINES

Airline ID	Number of observations	Entry frequency			
		$\delta = 1.0$		$\delta = 2.0$	
1	2,091	1,005	48.1%	930	44.5%
2	1,931	934	48.4%	779	40.3%
3	1,589	724	45.6%	565	35.6%
4	1,379	628	45.5%	448	32.5%
5	803	361	45.0%	249	31.0%
6	442	205	46.4%	144	32.6%
Total	8,235	3,857	46.8%	3,115	37.8%

4.3 Results of the Experiments

The first set of Monte Carlo experiments is based on the correct model specification and concerns about the size of potential simulation bias inherent in the method of simulated likelihood for a small number of simulation draws. A debate exists in the literature on the choice between the method of simulated likelihood and the method of simulated moments. While the method of simulated likelihood may suffer from simulation bias given a fixed number of simulation draws, it is simple to implement, numerically stable, and potentially efficient under the correct specification. Geweke, Keane, and Runkle (1997) and McFadden and Ruud (1994) provide evidence of the instability of the method of simulated moment estimator. Nevertheless, the number of simulation draws that will lead to a sufficiently small bias is an empirical question specific to each application, and in particular depends on the complexity of the covariance structure of error terms. Table 4 compares the estimates of four different simulation draw settings. The data generating process assumes $\delta = 2.0$. The first experiment makes 20 independent simulation draws, while the second experiment uses antithetic sampling to make 20 simulation draws, i.e. 10 symmetric replications of 10 independent pseudo-random draws to make simulation draws more systematic. The results show that, first, even with only 20 independent simulation draws, the comparison of the true parameter values and estimated values indicates overall accuracy given the estimated standard errors. Second, however, the use of antithetic sampling considerably improves the model fit in terms of the average log likelihood value. Third, increasing the number of draws to 40 and 300 shows a further improvement in the fit, though the improvement is rather small. This pattern is consistently observed in simulations with different values of parameters and different seeds of pseudo-random number generator. Since the covariance structure in the present model is rather simple, the result shows accuracy even with a very small number of simulation draws, albeit small simulation bias is observed. Though not shown here, for a smaller value of ρ, i.e. a smaller market level random effect, the number of simulation draws required to generate the same level of accuracy is even smaller, since the distribution of each random error is closer to the univariate standard normal distribution.

Table 4: POTENTIAL SIMULATION BIAS: $\delta = 2.0$

		20 draws no antithetics			20 draws			40 draws			300 draws		
θ	DGP	$\bar{\bar{\theta}}$	$\overline{ASE}$	MSE	$\bar{\bar{\theta}}$	$\overline{ASE}$	MSE	$\bar{\bar{\theta}}$	$\overline{ASE}$	MSE	$\bar{\bar{\theta}}$	$\overline{ASE}$	MSE
cons	−5.0	−4.990	0.145	0.140	−4.982	0.146	0.142	−4.978	0.146	0.144	−4.977	0.147	0.145
pop	1.2	1.193	0.032	0.024	1.197	0.033	0.023	1.197	0.033	0.022	1.198	0.033	0.023
dist	0.0	0.001	0.011	0.010	0.001	0.011	0.010	0.001	0.011	0.010	0.001	0.011	0.010
pastp	0.4	0.389	0.018	0.023	0.388	0.018	0.023	0.388	0.018	0.023	0.388	0.018	0.023
city2	1.5	1.502	0.041	0.033	1.496	0.041	0.034	1.495	0.041	0.034	1.494	0.041	0.034
nroute	0.0	0.004	0.032	0.035	−0.002	0.032	0.036	−0.003	0.032	0.036	−0.004	0.032	0.036
δ	2.0	1.992	0.074	0.068	2.009	0.075	0.067	2.012	0.075	0.070	2.016	0.075	0.069
ρ	0.7	0.691	0.028	0.031	0.701	0.027	0.030	0.703	0.027	0.030	0.705	0.027	0.031
$\overline{LogL}$		−3145.26			−3141.20			−3140.96			−3140.57		

Note: $\theta \equiv$ parameter, DGP $\equiv$ data generating value, $\bar{\bar{\theta}} \equiv$ average parameter estimate, $\overline{ASE} \equiv$ average asymptotic standard error, MSE $\equiv$ root mean square error, $\overline{LogL} \equiv$ average log likelihood value.

The next series of experiments examines the effect of misspecification by imposing restrictions on the correctly specified model (Table 5). The data generating process assumes $\delta = 2.0$ and each estimation makes 40 simulation draws using antithetic sampling. The first restricted model assumes that the econometrician has no correct knowledge about the decision order so estimates the model imposing a completely random decision order. The lack of decision order information reduces the model fit and leads to significant bias of most estimates. The serious underestimation of δ and ρ and the overestimation of *nroute* are particularly notable. In the data generating process, early movers enjoy their advantages, but without correct information on the decision order, these advantages are not captured as a strategic effect in δ and instead are captured in the positive coefficient of *nroute*, which determines the decision order but has no direct effect on payoff in the true data generating process. Inability to well explain the entry decision of each firm results in higher weights on individual random components, which leads to the underestimation of ρ. The two variables that have no correlation with the decision order, *dist* and *pastp*, are nevertheless precisely estimated, which is the case for all the experiments conducted below. The next restricted model assumes the correct specification of the decision order but imposes zero market level random effect, $\rho = 0$. Since this restriction removes the correlation between multivariate normal variates, high-dimensional integration is no longer necessary and the estimation procedure is significantly simplified. This misspecification, however, leads to considerable reduction in the model fit and significant bias of estimates. The strategic effect, δ, is underestimated because ignoring market random errors that generate correlation between entry decisions of firms blurs the true harshness of strategic interaction. The last restricted model assumes no market error and no interaction effect ($\delta = 0$ and $\rho = 0$). These restrictions degenerate the model to a probit model. The model fit is the worst in this table. Ignoring early mover advantages again leads to a spurious positive estimate of the size effect.

Table 5: RESTRICTED MODELS: $\delta = 2.0$, 40 SIMULATION DRAWS

		Full Model			No Order Info			No Market Error			Probit		
θ	DGP	$\bar{\hat{\theta}}$	$\overline{ASE}$	MSE	$\bar{\hat{\theta}}$	$\overline{ASE}$	MSE	$\bar{\hat{\theta}}$	$\overline{ASE}$	MSE	$\bar{\hat{\theta}}$	$\overline{ASE}$	MSE
cons	−5.0	−4.978	0.146	0.144	−5.338	0.140	0.359	−5.167	0.128	0.217	−4.425	0.125	0.588
pop	1.2	1.197	0.033	0.022	1.070	0.031	0.132	1.060	0.026	0.142	0.623	0.021	0.577
dist	0.0	0.001	0.011	0.010	0.001	0.010	0.009	0.000	0.008	0.010	0.001	0.010	0.007
pastp	0.4	0.388	0.018	0.023	0.402	0.019	0.016	0.416	0.019	0.025	0.385	0.018	0.023
city2	1.5	1.495	0.041	0.034	1.565	0.038	0.072	1.619	0.038	0.124	1.560	0.036	0.066
nroute	0.0	−0.003	0.032	0.036	0.269	0.028	0.270	0.173	0.031	0.175	0.371	0.027	0.372
δ	2.0	2.012	0.075	0.070	1.472	0.070	0.531	1.394	0.052	0.609			
ρ	0.7	0.703	0.027	0.030	0.488	0.039	0.214						
$\overline{LogL}$ / $\overline{BIC}$		−3140.96 / 6350.32			−3226.62 / 6521.64			−3211.20 / 6483.21			−3598.46 / 7250.13		

Note: $\theta \equiv$ parameter, DGP $\equiv$ data generating value, $\bar{\hat{\theta}} \equiv$ avg parameter estimate, $\overline{ASE} \equiv$ avg asymptotic standard error, MSE $\equiv$ root mean square error, $\overline{LogL} \equiv$ avg log likelihood value, $\overline{BIC} \equiv$ avg Bayesian information criterion.

Table 6 reports the results of the same comparison for $\delta = 1.0$, which reflects a weaker strategic effect. Overall the results are consistent with the previous table. One notable difference is that misspecifying and ignoring the sequential interaction leads to much less reduction in the model fit.

Table 6: RESTRICTED MODELS: $\delta = 1.0$, 40 SIMULATION DRAWS

		Full Model			No Order Info			No Market Error			Probit		
θ	DGP	$\bar{\hat{\theta}}$	$\overline{ASE}$	MSE	$\bar{\hat{\theta}}$	$\overline{ASE}$	MSE	$\bar{\hat{\theta}}$	$\overline{ASE}$	MSE	$\bar{\hat{\theta}}$	$\overline{ASE}$	MSE
cons	−5.0	−4.863	0.133	0.163	−5.003	0.131	0.096	−4.876	0.119	0.161	−4.739	0.126	0.288
pop	1.2	1.171	0.030	0.037	1.143	0.030	0.060	1.026	0.024	0.175	0.893	0.022	0.308
dist	0.0	0.001	0.010	0.011	0.001	0.010	0.011	0.001	0.008	0.011	0.001	0.008	0.010
pastp	0.4	0.391	0.018	0.021	0.398	0.018	0.018	0.417	0.018	0.025	0.421	0.019	0.028
city2	1.5	1.472	0.041	0.044	1.510	0.039	0.033	1.597	0.037	0.102	1.638	0.038	0.144
nroute	0.0	−0.005	0.028	0.026	0.073	0.026	0.077	0.098	0.029	0.102	0.151	0.028	0.154
δ	1.0	1.016	0.064	0.039	0.888	0.059	0.118	0.462	0.036	0.539			
ρ	0.7	0.699	0.027	0.020	0.650	0.028	0.056						
$\overline{LogL}$ / $\overline{BIC}$		−3368.05 / 6804.50			−3376.77 / 6821.95			−3447.04 / 6954.89			−3432.56 / 6918.33		

Note: $\theta \equiv$ parameter, DGP $\equiv$ data generating value, $\bar{\hat{\theta}} \equiv$ avg parameter estimate, $\overline{ASE} \equiv$ avg asymptotic standard error, MSE $\equiv$ root mean square error, $\overline{LogL} \equiv$ avg log likelihood value, $\overline{BIC} \equiv$ avg Bayesian information criterion.

The next set of experiments introduces various degrees of randomness in the decision order. In many potential applications the econometrician may have a priori information that reflects the true decision order only approximately. This limited knowledge about the true decision order motivates this experiment. Specifically, while the estimated models still assume that firms make decisions in order of *nroute*, I modify the data generating process in such a way that the true decision order is determined by a weighted sum of *nroute* and a random variable that follows a uniform distribution with the same mean and variance as *nroute*. Thus, the weight of this uniform random variable captures the level of imprecision of the decision order information used in the estimation. Table 7 reports the results for different degrees of randomness. The results show that when the econometrician correctly specifies more than 90 percent of the decision order, the differences between the estimated coefficients and their population values tend to be smaller than the estimated standard error. The fact that the model fit reduces with randomness in the decision order suggests the potential use of non-nested model selection tests. Estimating and comparing models with different decision order assumptions allows us to examine which decision order best fits the data.

5 DISCUSSION AND EXTENSIONS

5.1 The Perfect Information Assumption

I use the perfect information assumption to guarantee a unique subgame perfect equilibrium. The uniqueness is necessary to specify the domain of integration in the ε space for

Table 7: EFFECT OF RANDOMNESS IN DECISION ORDER

Randomness in sequence	misspecified order (%)	δ ($\delta_0 = 2.0$) $\bar{\bar{\delta}}$	$\overline{ASE}$	MSE	ρ ($\rho_0 = 0.7$) $\bar{\bar{\rho}}$	$\overline{ASE}$	MSE	$\overline{LogL}$
0%	0.0%	2.012	0.075	0.070	0.703	0.027	0.030	−3140.96
10%	0.0%	2.012	0.075	0.070	0.703	0.027	0.030	−3140.96
20%	1.9%	2.006	0.075	0.065	0.699	0.028	0.028	−3141.24
30%	13.4%	1.939	0.075	0.098	0.667	0.030	0.047	−3162.22
40%	26.0%	1.876	0.076	0.140	0.640	0.032	0.069	−3188.38
50%	37.1%	1.823	0.076	0.192	0.623	0.032	0.086	−3209.99
60%	45.7%	1.786	0.077	0.220	0.608	0.033	0.097	−3228.39
70%	52.0%	1.737	0.077	0.266	0.590	0.034	0.115	−3245.10
80%	56.9%	1.701	0.077	0.304	0.571	0.036	0.133	−3264.87
90%	60.7%	1.685	0.077	0.321	0.568	0.036	0.137	−3272.85
100%	63.7%	1.682	0.077	0.323	0.570	0.036	0.135	−3275.09

Note: misspecified order indicates how many observations are assigned with different decision order. $\bar{\bar{\delta}}, \bar{\bar{\rho}} \equiv$ avg parameter estimate, $\overline{ASE} \equiv$ avg asymptotic standard error, MSE $\equiv$ root mean square error, $\overline{LogL} \equiv$ avg log likelihood value, 40 simulation draws using antithetic sampling.

each strategy profile that rationalizes the observed game outcome, without making a strong (often ad hoc) assumption on the equilibrium selection mechanism. Admittedly the perfect information assumption is strong in many applications. Even though the main point of introducing sequentiality in empirical studies is to study implications of publicly known decision order and publicly known decision history, results might be affected by possibilities that some players may move simultaneously, there may be some private information, and "nature" may bring in uncertainty. Relaxing the perfect information assumption is possible as long as the uniqueness of an equilibrium is guaranteed for any possible values of random shocks, ε. In general the following approaches potentially help relaxing the perfect information assumption. First, we can specify the game and payoff function in such a way that a unique subgame perfect equilibrium is guaranteed. Second, focusing on a set of equilibria might provide uniqueness. An example is an entry game in which the identity of the entering firms is not uniquely determined but the number of entrants is uniquely determined (Berry (1992)). Third, an equilibrium concept that is stronger than subgame perfection may help to avoid the multiplicity of equilibria. For example, sequential equilibrium (Kreps and Wilson (1982)) may reduce the set of subgame perfect equilibrium strategy profiles when decision nodes that are never reached exist (Litan and Pimienta (2008)). Fourth, some equilibrium selection mechanism can be assumed. The use of the notions of Pareto and risk-dominance may provide a reasonable option if it leads to a unique equilibrium.

5.2 Decision Order

The entry game example in the previous section assumes that each firm makes a one-shot decision sequentially. In general, the proposed estimation framework allows players to take multiple turns alternately. In simulating the likelihood function, all turns of player i must be simulated at once, as the strategy of each player consists of a decision at every decision node.

A more fundamental issue on decision order is the empirical analogue of decision order. The proposed estimation framework utilizes a publicly known exogenous decision order. In some applications, even if sequential interaction appears likely, such decision order may not be available or may be endogenously determined. The above Monte Carlo experiments illustrate that misspecifying the true decision order may lead to significant bias of the estimate of strategic effect. If the game is correctly specified except for decision order, we can draw an inference about not only structural parameters but also decision order. Specifically, the econometrician can estimate different models, each with a different imposed decision order, then conduct a model selection test for non-nested specifications. Advancing this idea further, estimation of the population decision order by selecting the decision order that maximizes the likelihood function may be a possibility. The statistical properties of an estimated decision order and how to deal with the discontinuity that arises from maximization over decision orders are left for future research.[13]

5.3 Computation

For applications with relatively simple games such as the entry game example in this paper, the computation burden of the proposed estimation procedure is fairly manageable. This is due to the high performance of the GHK simulator, and also because, while solving a sequential game requires a large number of calculations, it does not require much multiplication and division. For example, conducting all the Monte Carlo experiments shown in the tables of this paper only requires a half day or so with a standard stand-alone desktop computer.[14]

However, increasing the number of players, the number of turns, or the number of alternatives increases the dimension of integration, which may quickly make computation infeasible. Although applications with many players, many alternatives, and many turns generally entail less value in the structural estimation of sequential interaction, middle-sized games with a

[13] Endogenizing the order of decision is another possible extension. This class of games is called a leadership game or a commitment game (Hamilton and Slutsky (1990)) and has attracted some theoretical applications (e.g. Kempf and Rota-Graziosi (2010)). These games endogenize the order by introducing a pre-play stage that determines the order of decision. Consequently, these games are no longer perfect information games, but as long as a unique outcome is secured, estimation may be possible as discussed in the previous subsection. However, the empirical analogue of leadership games seems to be rather unclear.

[14] Most of the time is spent on the experiments with 300 simulation draws.

complex covariance structure may considerably benefit from the following computation techniques that reduce computational burden. First, structures of payoff function and strategic interaction implied by assumed economic theory can be utilized to skip the unnecessary part of the calculation in the backward induction algorithm. In the above entry game example, the assumed negative effect of a rival's entry excludes one strategy ("imitate" in Figure 3) from the simulation procedure. In Maruyama (2011) I exploit the non-increasing property of the profit function in the number of entering rival firms and reduce the computation time by more than 95 percent. As a result, in the estimation of sequential games with at most 16 heterogeneous firms, the computational burden is not found to be a significant problem.

Second, variance reduction techniques will enhance the performance of the simulator. The Monte Carlo experiments above show the gain from antithetic sampling. Instead of using pseudo-random numbers, systematic simulation draws by quasi-Monte Carlo sampling, such as Halton sequences, and sampling methods based on orthogonal arrays will produce better performance (Train (2003), Sándor and András (2004)). Lastly another potential avenue is the use of a more efficient importance-sampling algorithm to enhance the GHK simulator (Liesenfeld and Richard (2010)).

6 CONCLUSION

In this paper I study the estimation of finite sequential games with perfect information and propose a computationally practical estimation method that overcomes high-dimensional truncated integration with complication due to sequential strategic interaction. I show that separate evaluation of each subgame perfect strategy profile that rationalizes the observed equilibrium outcome allows us to use the GHK simulator, the most widely used importance sampling probit simulator, for Monte Carlo integration, by controlling for unobserved off-the-equilibrium-path strategies. The method allows researchers to empirically study strategic interactions in a large asymmetric game. Specifically, researchers can draw inferences on strategic complementarity and perform counterfactual simulations that take sequential strategic interactions into account. Monte Carlo experiments for a simple entry game example demonstrate the performance and robustness of the proposed method and the potential bias resulting from misspecification.

References

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

[2] Andreoni, J., and E. Blanchard (2006): "Testing Subgame Perfection Apart from Fairness in Ultimatum Games," *Experimental Economics*, 9(4), 307-321.

[3] Bajari, P., C. L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75(5), 1331-1370.

[4] Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2007): "Estimating Static Models of Strategic Interactions," mimeo.

[5] Bajari, P., H. Hong, and S. P. Ryan (2008): "Identification and Estimation of a Discrete

Game of Complete Information," mimeo.

[6] Bas, M. A., C. S. Signorino, and R. W. Walker (2008): "Statistical Backwards Induction: A Simple Method for Estimating Recursive Strategic Models," *Political Analysis*, 16(1), 21-40.

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

[8] Bjorn, P. A., and Q. H. Vuong (1984): "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation," California Institute of Technology Social Science Working Paper 537.

[9] Börsch-Supan, A., and V. Hajivassiliou (1993): "Smooth Unbiased Multivariate Probability Simulation for Maximum Likelihood Estimation of Limited Dependent Variable Models," *Journal of Econometrics*, 58, 347-368.

[10] Bresnahan, T. F., and P. C. Reiss (1991): "Empirical-Models of Discrete Games," *Journal of Econometrics*, 48(1-2), 57-81.

[11] Chernew, M., G. Gowrisankaran, and A. M. Fendrick (2002): "Payer Type and the Returns to Bypass Surgery: Evidence from Hospital Entry Behavior," *Journal of Health Economics*, 21, 451-474.

[12] Ciliberto, F., and E. Tamer (2007): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, forthcoming.

[13] Geweke, J. (1992): "Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments (with Discussion)," in *Bayesian Statistics*, J. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), Volume 4, 169-193, Oxford, Oxford University Press.

[14] Geweke, J., M. Keane, and D. Runkle (1994): "Alternative Computational Approaches to Inference in the Multinomial Probit Model," *Review of Economics and Statistics*, 76, 609-632.

[15] Geweke, J., M. Keane, and D. Runkle (1997): "Comparing Simulation Estimators for the Multinomial Probit Model," *Journal of Econometrics*, 80, 125-166.

[16] Hajivassiliou, V., and D. McFadden (1994): "A Simulation Estimation Analysis of the External Debt Crises of Developing Countries," *Journal of Applied Econometrics*, 9, 109-131.

[17] Hajivassiliou, V., and D. McFadden (1998): "The Method of Simulated Scores for the Estimation of LDV Models," *Econometrica*, 66, 863-896.

[18] Hajivassiliou, V., D. McFadden, and P. Ruud (1996): "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results," *Journal of Econometrics*, 72, 85-134.

[19] Hamilton, J. H., and S. M. Slutsky (1990): "Endogenous timing in duopoly games: Stackelberg or Cournot equilibria," *Games and Economic Behavior*, 2(1), 29-46.

[20] Hiedeman, B. (1998): "A Stackelberg Model of Social Security Acceptance Decisions

in Dual-Career Households," *Journal of Economic Behavior & Organization*, 34, 263-78.

[21] Jia, Z. (2005): "Labor Supply of Retiring Couples and Heterogeneity in Household Decision-Making Structure," Discussion Papers No. 404, Research Department, Statistics Norway.

[22] Keane, M. (1994): "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62(1), 95-116.

[23] Konrad, K. A., H. Künemund, K. E. Lommerud, and J. R. Robledo (2002): "Geography of the Family," *American Economic Review*, 92(4), 981-998.

[24] Kempf, H., and G. Rota-Graziosi (2010): "Endogenizing leadership in tax competition," *Journal of Public Economics*, 94, 768-776.

[25] Kooreman, P. (1994): "Estimation of Econometric Models of Some Discrete Games," *Journal of Applied Econometrics*, 9, 255-268.

[26] Kreps, D., and R. Wilson (1982): "Sequential Equilibria," *Econometrica*, 50, 863-894.

[27] Lerman, S., and C. Manski (1981): "On the Use of Simulated Frequencies to Approximate Choice Probabilities," *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. Manski and D. McFadden, Cambridge, MA: MIT Press, 305-319.

[28] Maruyama, S. (2011): "Socially Optimal Subsidies for Entry: The Case of Medicare Payments to HMOs," *International Economic Review*, forthcoming.

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

[30] McFadden, D., and P. A. Ruud (1994): "Estimation by Simulation," *Review of Economics and Statistics*, 76, 591-608.

[31] Nelder, J. A., and R. Mead (1964): "A Simplex Method for Function Minimization," *Computer Journal*, 7(4), 308-313.

[32] Liesenfeld, R., and J. F. Richard (2010): "Efficient Estimation of Probit Models with Correlated Errors," *Journal of Econometrics*, 156, 367-376.

[33] Litan, C., and C. Pimienta (2008): "Conditions for Equivalence between Sequentiality and Subgame Perfection," *Economic Theory*, 35(3), 539-553.

[34] Redoano, M. (2007): "Fiscal Interactions among European Countries. Does the EU Matter?" CESifo Working Paper No. 1952.

[35] Sándor, Z., and P. András (2004): "Alternative Sampling Methods for Estimating Multivariate Normal Probabilities," *Journal of Econometrics*, 120, 207-234.

[36] Schmidt-Dengler, P. (2006): "The Timing of New Technology Adoption: The Case of MRI," mimeo.

[37] Signorino, C. S., and A. R. Tarar (2006): "A Unified Theory and Test of Extended Immediate Deterrence," *American Journal of Political Science*, 50, 586-605.

[38] Soetevent, A. R., and P. Kooreman (2007): "A Discrete-Choice Model with Social Interactions: With an Application to High School Teen Behavior," *Journal of Applied*

Econometrics, 22, 599-624.

[39] Su, C., and K. Judd (2008): "Constrained Optimization Approaches to Estimation of Structural Models," mimeo.

[40] Train, K. (2003): *Discrete Choice Methods with Simulation, Cambridge*, UK, Cambridge University Press.