

Estimation of Finite Sequential Games

Shiko Maruyama*

Economics Discipline Group, University of Technology Sydney

Abstract

I propose a new estimation method for finite sequential games that is efficient, computationally attractive, and applicable to a fairly general class of finite sequential games that is beyond the scope of existing studies. The major challenge is computation of high-dimensional truncated integration whose domain is complicated by strategic interaction. This complication resolves when unobserved off-the-equilibrium-path strategies are controlled for. Separately evaluating the likelihood contribution of each subgame-perfect equilibrium that generates the observed outcome allows the use of the GHK simulator, a widely used importance-sampling probit simulator. Monte Carlo experiments demonstrate the performance and robustness of the proposed method.

JEL Codes: C35, C72, C63, C31

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

*Please address correspondence to: Shiko Maruyama, Economics Discipline Group, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia. E-mail: shiko.maruyama@uts.edu.au.

1 INTRODUCTION

In this paper, I study the structural estimation of finite sequential games and propose a new estimation method that is efficient, computationally attractive, and applicable to a fairly general class of finite sequential games that is beyond the scope of existing studies. Existing empirical studies that consider sequential games (at least as an 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; Hiedeman, 1998), the retirement behavior of elderly couples (Jia, 2005), the location choice of siblings (Konrad et al., 2002; Maruyama and Johar, 2013), political science and international relations (Signorino and Tarar, 2006; Bas et al., 2008), tax competition (Kempf and Rota-Graziosi, 2010), and the validity of subgame perfection in experimental economics (Andreoni and Blanchard, 2006). All of the existing literature on sequential games 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 (e.g. a binary choice symmetric game); or emphasis is not on the structural estimation of strategic effect.¹

The class of games I study in this paper is finite sequential games, i.e., finite-horizon pure-strategy discrete-choice sequential games with perfect information, in which each player makes a decision in publicly known exogenous decision order. The econometrician knows the decision order and uses data on players and their decisions to estimate a parametric

¹Exceptions are Maruyama (2011) and Maruyama and Johar (2013), which are based on the approach outlined in this paper.

model of payoffs and random components. The random components serve as structural errors that are observed by players, but not by the econometrician. Conceptually, solving such sequential games is straightforward by backward induction. When the random errors follow continuous distribution, such as multivariate normal distribution, the game becomes even simpler to solve, because ties occur with probability measure zero and there always exists a unique subgame-perfect equilibrium. Once the relationship from realized values of the random errors to a unique equilibrium outcome is established, the remaining task is simply to seek parameter values that minimize a certain distance metric between the predicted and observed game outcomes. Computationally, however, except for extremely simple games, estimating sequential games is challenging. Even for a fairly simple game in which four players sequentially make binary decisions, the standard maximum likelihood method is not feasible because the likelihood function does not have an analytical solution due to high-dimensional integration. Maximum likelihood based on simulation techniques is an alternative, but its computation is a daunting task; the game needs to be solved for each observation of game plays for each simulation draw for each set of candidate parameter values.

The proposed method in this paper relies on two ideas. First, I propose the use of the Geweke-Hajivassiliou-Keane (GHK) simulator, the most popular solution for approximating high-dimensional truncated integrals in standard probit models. 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.

Sequential strategic interaction, however, complicates high-dimensional truncated integration in the probit framework, causing *interdependence of truncation thresholds*, which undermines the ground of the GHK's recursive conditioning approach. As the second building block of the proposed method, I propose the use of the GHK simulator not for the observed equilibrium outcome per se, but separately for each of all the subgame-perfect strategy profiles that rationalize the observed equilibrium outcome. In the sequential game framework, the observed equilibrium outcome arises according to the underlying subgame-perfect equilibrium, but the econometrician does not observe the underlying equilibrium, because an equilibrium strategy consists of a complete contingent plan, which includes off-the-equilibrium-path strategies as unobserved counterfactuals. Even if a unique subgame-perfect equilibrium is guaranteed, from the econometrician's viewpoint, there may exist different realizations of unobservables that lead to different subgame-perfect equilibria that generate an observationally identical game outcome.

The use of subgame perfection allows us to uniquely determine the corresponding subgame-perfect equilibrium for each realization of random components. I show that the separate evaluation of likelihood contribution for each subgame-perfect strategy profile allows us to control for the unobserved off-the-equilibrium-path strategies so that the recursive conditioning of the GHK works by making the domain of Monte Carlo integration (hyper-)rectangular. The econometrician then obtains the probability of the observed outcome by summing the probabilities of each equilibrium that rationalizes the observed outcome, and the use of

maximum likelihood follows.

Section 2 positions the proposed method in the empirical game literature and discusses the potential usefulness of the method. After formally presenting the setup in Section 3, I explain in Section 4 how the GHK simulator can aid high-dimensional integration under subgame perfection. In Section 5, to demonstrate the performance and robustness of the proposed estimation method, I conduct Monte Carlo experiments. Section 6 discusses potential extension and computation issues.

2 RELATION TO THE LITERATURE AND APPLICABILITY

This paper builds on a line of research on the estimation of non-cooperative discrete games, initiated by Bjorn and Vuong (1984) and Bresnahan and Reiss (1991). Recent development in this literature has mostly centered around two issues: the identification problem due to multiple equilibria² and the computation problem. This paper contributes to the latter by providing a direction different from recent developments. There is a very active literature on the estimation of dynamic discrete games. Recent work by Pesendorfer and Schmidt-Dengler (2003), Pakes et al. (2007), Aguirregabiria and Mira (2007), and Bajari et al. (2007) is based on a computationally convenient two-step approach, developed by Hotz and Miller (1993), which exploits the mapping in discrete-choice problems between conditional

²For example, see Chernozhukov et al. (2007), Ciliberto and Tamer (2009), and Pakes et al. (2011).

choice probabilities and choice-specific value functions. The computational advantage of this type of method comes from the fact that it only uses necessary conditions of equilibrium and does not explicitly compute equilibrium. Two general drawbacks to these estimators are the information loss that may lead to substantial finite sample bias and the difficulty of conducting counterfactual simulations. The approach proposed in this paper, although its computational advantage depends on each application, does not have these drawbacks as it is based on the explicit calculation of equilibrium.

The recent work by Jia (2008) on the location choice of discount chains has some similarities to this paper in that she studies a discrete simultaneous complete-information game with a large choice set. Her innovative approach to the dimensionality problem relies on the lattice theory. For her approach to work, however, the model has to satisfy several strong restrictions.³ Similarly, the literature on incomplete-information static games computes equilibrium by using fixed point theorem (e.g. Seim, 2007). The fixed point algorithm works well as long as the underlying assumptions are satisfied. My approach, on the other hand, relies on the backward induction algorithm to find equilibrium, a conceptually much simpler approach, which works in a fairly general class of finite sequential games.

Whether sequentiality is a reasonable assumption to make depends on each application. In entry games, for example, there may not be an explicit sequence in the first place. It may be natural that the recent empirical game literature has centered around the identification issue under the possibility of multiple equilibria. The sequential game assumption allows

³In Jia's setup, externality across markets must be positive and the number of players cannot exceed two.

this paper to circumvent the issue of multiple equilibria but the validity of the assumption needs to be warranted in each application.

More importantly, however, the sequential game framework is not merely a technical assumption to avoid multiple equilibria but a tool to investigate sequential strategic interaction, such as the first-mover (dis-)advantage and preemptive behavior to deter a rival's action. Sequential strategic interaction is observed in a wide range of real world phenomena: heavily regulated industries, organizational decision making, labor disputes, judicial cases, decisions among siblings, drafts in sports leagues, parlor and TV show games, and so on. Innumerable theoretical studies on sequential games exist, but there has been little empirical work devoted to quantifying the relevance and implications of sequential interaction.

It is worthwhile to point out that the proposed method does not fully resolve the high dimensionality problem. The GHK significantly facilitates high-dimensional integration, but as the game size increases, the number of possible strategy profiles increases exponentially. Although there are ways to further improve computational efficiency, as discussed in the last section, computational practicality remains a challenge when a game is very large.⁴

The benefit of the proposed method will be fully exploited when an application focuses on sequential interaction in a middle-sized game, which is not overly large but if larger than the two-player binary-choice game. On one hand, the two-player Stackelberg game, which has been widely studied in the theoretical literature, has limited use in empirical research. On the other hand, an application with a game played by a large number of players may

⁴The maximum size of the game a researcher can practically estimate depends on various factors, such as imposed game structures, the sample size, and the availability of high-performance computing.

entail less value in inference on sequential interaction; a simultaneous game framework may be more appropriate for such a game.

The paper by Maruyama and Johar (2013) offers an example of the intended use of the proposed method. The paper concerns the location choice of adult siblings. In our setup, adult siblings make location decisions in their birth order—the order they finish their schooling—while the well-being of their elderly parents is their shared concern. This setting creates a public good problem and sequential strategic interaction. While statistical inferences are based on a quite large data set, the game is not large: we consider families with up to four siblings and the decision is binary—whether to live far away from the parent or not. The model, instead, features very rich heterogeneity. The error term has a complex covariance structure in which correlation among siblings depends on their characteristics, such as age and gender differences. We do not impose a priori assumption on strategic complementarity, allowing families to play different types of games. The proposed algorithm performs very well, leading us to find economically insignificant sequential interaction but a significant public good problem.

In the rest of the paper I repeatedly use entry game examples. Readers should note that this is primarily for illustration purposes. I employ the entry game examples because of their simple form as a discrete game that is a well-understood classic in the empirical game literature (Bresnahan and Reiss, 1991; Berry, 1992). The applicability and fruitfulness of the finite sequential game framework depends on each application.

3 MODEL

3.1 The Sequential Game

The model is a finite sequential game with perfect information. There are $i = 1, \dots, N$ 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").⁵ Define $A \equiv \times_i A_i$ and let $a \equiv (a_1, \dots, a_N)$ 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. The payoff function of player i , $\pi_i : A \rightarrow \mathbb{R}$, is

$$\pi_i(a, x, \varepsilon_i; \theta_1) = \bar{\pi}_i(a, x; \theta_1) + \varepsilon_i^{a_i}, \quad (1)$$

where θ_1 is a vector of parameters and vector x contains exogenous characteristics that describe the players and the environment in which the game is played. The first term, $\bar{\pi}_i(a, x; \theta_1)$, is an assumed parametric function of mean payoffs. The second term, $\varepsilon_i^{a_i} \in \mathbb{R}$, is a random preference shock player i incurs when a_i is chosen. Define a vector, $\varepsilon_i \equiv \{\varepsilon_i^{a_i}\}_{a_i \in A_i}$ and $\varepsilon \equiv (\varepsilon_1, \dots, \varepsilon_N)$. ε follows continuous parametric density function, $g(\varepsilon; \theta_2)$, where θ_2 is a vector of parameters.^{6,7} Both x and ε are common knowledge to the players, but the

⁵Allowing the choice set to vary across decision nodes is a straightforward extension.

⁶ $g(\varepsilon; \theta_2)$ may depend on x as well.

⁷I 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.

econometrician observes only x .

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 (but not off-the-equilibrium-path decisions) 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.⁸ Define $S \equiv \times_i S_i$ and let $s \equiv (s_1, \dots, s_N) \in S$ denote a *strategy profile*. Since s uniquely determines a game outcome, define $a(s) : S \rightarrow A$ and $a_i(s) : S \rightarrow A_i$. In the example of a two-player sequential entry game, if $s_{\text{leader}} = (\text{"In"})$ and $s_{\text{follower}} = (\text{"In"} \text{ if leader stays out, "Out"} \text{ if leader enters})$, then $a(s) = (\text{In, Out})$.

Given the primitives defined above, each player chooses a strategy s_i that maximizes the payoff taking rivals' strategies as given. The solution concept of the game in this paper is subgame perfection, which is a refinement of Nash equilibrium to exclude certain strategies such as noncredible threat. 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

⁸Incorporating mixed strategies in the present framework is computationally impractical and beyond the scope of this paper.

finite game with perfect information has a pure-strategy subgame-perfect equilibrium (Zermelo's theorem). Furthermore, in the current setup, the game almost surely has a unique equilibrium, because ties occur with probability measure zero. 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. In other words, given (x, θ_1) , a realization of ε results in a unique subgame-perfect equilibrium.

3.2 Data

The econometrician observes T independent realizations of the game, $(\Gamma_1, \dots, \Gamma_T)$, e.g., T different markets, T different families, and T periods of time. Each realization of the game is indexed by $t = 1, \dots, T$. 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 $g_i(\varepsilon_i^{ait}; \theta_2)$ 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}^e 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 either from institutional knowledge, by assumption, or from observation of data. 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

$$\varepsilon_t \sim N(0, \Omega). \quad (2)$$

Covariance matrix Ω has a dimension of $\prod_{i=1}^N$ [the number of alternatives for i] and is parameterized by θ_2 . For the parameterization of Ω , the standard 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 requires normalization of the mean and variance of ε .⁹ For example, in the setup of standard binary-choice games, the dimension of Ω is the number of players. Below, I abuse notation and use ε and Ω to denote the error structure after normalization.

3.3 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 $\pi_i(a, x, \varepsilon_i; \theta_1)$ and $g_i(\varepsilon_i^{a_i}; \theta_2)$. Since the distribution of ε is specified fully parametrically, the estimation procedure relies on maximum likelihood. Game Γ_t is the unit for which individual likelihood is defined as follows:

$$l(\theta; x_t, a_t^o) = \Pr[a_t^o = a_t^e(x_t, \varepsilon_t; \theta_1) | \theta_2]. \quad (3)$$

⁹In applications with more model structures, information on the level of payoffs may be available and aid identification, making the normalization of the variance of error terms unnecessary.

This leads to the following maximum likelihood problem:

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

The challenge in this maximum likelihood framework is that the probability term in (3) 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. There are several cases where this likelihood function is easily computed. First is the two-dimensional case (Stackelberg games), which arises if the number of players is two and the decision to be made is binary. 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. 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 this case, though the game still needs to be solved for an equilibrium, once it is solved, obtaining an analytical solution is trivial. In most applications, however, the independent normal assumption is restrictive. It implies no game specific error (e.g. market specific random component), and when the choice set is larger than the binary case, it also implies a quite restrictive substitution pattern among alternatives.

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

likelihood (MSL) method, which utilizes Monte Carlo integration.¹⁰ The most straightforward simulator for MSL is the crude frequency simulator, first proposed by Lerman and Manski (1981). This simulation procedure takes R sets of random draws from the assumed distribution. For each random draw $\tilde{\varepsilon}_t^r$, an equilibrium outcome a_t^e is solved by backward induction. The probability simulator is based on how many times the predicted equilibrium outcome coincides with the observed outcome out of R times repetition of simulation draws. Although this simulator provides estimates that are consistent with R and T , it has two major limitations. First, the simulated probability is a discontinuous function of the parameters and is not bounded away from 0 and 1. Second, the use of the indicator function makes the variance of this simulator quite large, especially in high-dimensional cases. As a result of these problems, Lerman and Manski (1981) find that their estimator requires a very large number of simulations for satisfactory performance. McFadden (1989) develops smoothed simulators that solve the discontinuity problem. Smoothed simulators simplify the iterative computation of the estimator, allowing researchers to use an optimization method that relies on the differentiability of the optimand. McFadden's (1989) smoothed simulators, however, do not address the large variance problem in high-dimensional cases. Since a likelihood evaluation of relatively large asymmetric extensive form games tends to be particularly expensive, these simulators are practically infeasible.

¹⁰The 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 et al. (1994) do not find such an advantage of MSM over MSL.

3.4 The GHK Simulator

For high-dimensional integration over a region of the multivariate normal, the most popular simulator is the GHK simulator (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 algorithm to generate a smooth univariate truncated variate produces an unbiased and smooth importance-sampling simulator. Importance sampling aims to achieve higher efficiency by adjusting the weight or "importance" of different points in the sample space. Compared with the frequency simulator, the GHK simulator requires remarkably fewer draws for alternatives with low probability of being chosen. A number of studies have confirmed its usefulness and relative accuracy, especially when considering the low computational effort required (Börsch-Supan and Hajivassiliou, 1993; Geweke et al., 1994; Hajivassiliou et al., 1996; Hajivassiliou and McFadden, 1998).

The complication in using the GHK simulator for empirical games stems from the recursive conditioning approach. The GHK algorithm repeats recursive simulation draws from truncated univariate normal distributions so that the resulting random shocks, $\tilde{\varepsilon}^r$, generate the observed equilibrium outcome, a^o . 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 hence, 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, and recursive conditioning simulation breaks down.

4 USING THE GHK SIMULATOR

The problem of interdependent truncation thresholds arises as a result of changes in unobserved off-the-equilibrium-path strategies. Before formally presenting the general case results, I illustrate this point by a simple two-player entry game. Note that this two-player entry game is only for explanation purposes, as the game is two-dimensional and its likelihood function can easily be solved analytically.

The entry game is played by two players, firm 1 and firm 2. Firm 1 is the Stackelberg leader. Having observed firm 1's entry decision, firm 2 makes its entry decision. Firms 1 and 2 incur random shocks ε_1 and ε_2 respectively in their profit functions. For illustration purposes, assume that the rival's entry reduces payoff (this is not essential for the proposed method). Each firm enters the market when it expects nonnegative profits from entry. If it does not enter, a firm earns zero profit. 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 dominating impact, only firm 1, the leader with the first-mover advantage, enters. In this example, the probability of market configuration (Out,In) cannot

be computed by the GHK simulator directly, because the domain of integration is not a rectangle, and thus drawing ε_1 cannot be conditional on ε_2 and vice versa.

[Insert Figure 1]

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 2). 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". Assuming "Out" for firm 1, Figure 2 shows four possible equilibria. The extensive form highlights several important facts. First, subgame perfection implies that firm 2 chooses the best strategy 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.¹¹ Thus, ε_1 does not affect the thresholds of ε_2 that determine the choice of firm 2. Second, different strategy profiles may generate game outcomes that are observationally equivalent to the econometrician. In Figure 2, strategy profiles (3) and (4) both result in (Out,In). Third, firm 1's decision does depend on the strategy of firm 2, and hence, it does depend on ε_2 . When preemption is possible, the entry threshold for firm 1 is lower and the integration domain of ε_1 is larger. However, if the strategy of firm 2 is given, the threshold of ε_1 does not depend on the value of ε_2 .

¹¹Firm 2 never chooses the "imitate" strategy, due to the assumed negative impact of a rival's entry.

[Insert Figure 2]

Figure 3 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 2. 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.

[Insert Figure 3]

To formalize the discussion so far in the general n -dimensional case, let s_{-i} denote the subvector of strategy profile s that excludes component i , and let $s_i^{BR}(x, \varepsilon_i, s_{-i}; \theta_1)$ denote the function that determines the best response strategy of player i given x, ε_i , and s_{-i} . Given $(x, \varepsilon_i, s_{-i})$, the best response strategy of player i is uniquely determined almost surely by comparing payoffs at each decision node. Then, the following result holds.

Proposition 1 *For any strategy profile $s^* \in S$, if there exists a set of $\{\varepsilon\}$ that rationalizes s^* as a subgame-perfect strategy profile given x and θ_1 , then*

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

In words, 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}^* .*

Proof. In a finite sequential game with continuous random unobservables, player i 's best response strategy is uniquely determined by s_{-i} , x , and ε_i almost surely. 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 . Then the proposition follows trivially. ■

The logic underlying this proposition comes directly from the Nash equilibrium concept, not specifically from subgame perfection. However, for this result to hold, the best response needs to be uniquely determined. The subgame perfection (and hence the assumption of a sequential game) plays the key role in avoiding indeterminacy from the ε space to each player's best response.¹²

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. 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_i^o on her equilibrium path. The proposition clarifies that this interdependency across players does not occur as far as each subgame-perfect equilibrium is concerned.

To obtain $\hat{\theta}_{ML}$ 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

¹²Without subgame perfection, indeterminacy arises in off-the-equilibrium paths and makes Monte Carlo integration impossible.

problem, (3), as

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

The second equality holds owing to the fact that any ε leads to a unique subgame-perfect equilibrium. Given the discussion so far, using the GHK simulator for each subgame-perfect equilibrium is trivial. The rest of this section sets out this standard procedure to evaluate $\Pr[s = s^e(x, \varepsilon; \theta_1) | \theta_2]$ for each $s \in S^o(a^o)$. Readers familiar with the standard GHK procedure can turn directly to the Monte Carlo experiments.

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

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

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 $\Delta_i(x, s; \theta_1) \equiv \{\varepsilon_i | s_i^{BR}(x, \varepsilon_i, s_{-i}; \theta_1) = s_i\}$,

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

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_{-i} . 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_{-i} . 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 $\bar{S}^o(a^o, \theta_1) \subset S^o(a^o)$ 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

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

In the following I focus on $\bar{S}^o(a^o, \theta_1)$ 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 exposition, assume the choice set of player $i = 1, \dots, N$ 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 $\eta = (\eta_1, \dots, \eta_N)$ 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 notation to simplify the following presentation. For a vector of indexes $(1, \dots, N)$, the notation " $< i$ " denotes the subvector $(1, \dots, i - 1)$ and " $\leq i$ " denotes the subvector $(1, \dots, i)$.

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

$$\begin{aligned}
\Pr [s = s^e(x, \varepsilon; \theta_1) | \theta_2] &= \int_{\mathbb{R}^N} \prod_i I[\varepsilon_i \in \Delta_i(x, s; \theta_1)] n(\varepsilon, \Omega(\theta_2)) d\varepsilon \\
&= \int_{\mathbb{R}^N} \left[\prod_i I[L_{i,\leq i}\eta_{\leq i} \in \Delta_i(x, s; \theta_1)] \right] \cdot \left[\prod_i \phi(\eta_i) \right] d\eta \quad (5) \\
&= \int_{\mathbb{R}^N} \prod_i [I(L_{i,\leq i}\eta_{\leq i} \in \Delta_i(x, s; \theta_1)) \cdot \phi(\eta_i)] d\eta,
\end{aligned}$$

where $\phi(\cdot)$ 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, $r = (1, \dots, R)$, prepare an N -dimensional vector of independent uniform $(0, 1)$ random variables, $\tilde{u}^r = (\tilde{u}_1^r, \dots, \tilde{u}_N^r)$. For $u \in (0, 1)$ and a non-empty set $\Delta \subset \mathbb{R}$, define an inverse distribution 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 \tilde{u}^r , recursively define a sequence of simulated $\tilde{\eta}_i^r$ so as to satisfy

$s_i = s_i^{BR}(x, \varepsilon_i, s_{-i}; \theta_1)$ for $i = 1, \dots, N$ as

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

After obtaining simulated $\tilde{\eta}^r$, 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

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

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

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

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

likelihood problem,

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

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

5 MONTE CARLO EXPERIMENTS

5.1 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 (1) potential simulation bias in the Monte Carlo integration and (2) robustness with respect to the misspecification of 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 six 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 one to six. By construction, there is no distinction between entry by new

entrants and "entry" by incumbent firms. The econometrician observes the list of "potential entrants" and which firms choose to enter each market in the following year. 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 make their decisions in order of size.

Twenty 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 and contains information on the list of potential entrants, covariates, and generated random shocks in each market. Throughout all the experiments conducted below, I use the same twenty data sets for better compatibility of the simulation results. I conduct three sets of experiments. First, I examine the effects of changing the simulation setting, such as the number of simulation draws, to check the size of potential simulation bias. Second, to study the effect of misspecification, I impose restrictions on underlying models to be estimated. Third, I introduce various degrees of randomness in the decision order to address the possibility that the econometrician has only imprecise information about the true decision order.

5.2 Model and Data Generating Process

In market t , N_t firms play the entry game, where $N_t \in \{1, \dots, 6\}$. 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 by the econometrician but is known to every firm, and follows a multivariate normal distribution: $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{N_t,t})' \sim N(0, \Omega_t)$. The payoff when a firm does not enter is normalized to zero. 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 three thousand 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 one and six, 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 generated for the twenty 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 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 the strategic effect parameter, δ , and the decision order, I can solve the game by backward induction and obtain the data on market outcomes. The default specification is $\delta = 2.0$ and assumes that firms make decisions in order of *nroute*.

Tables 1 and 2 report descriptive numbers from one of the twenty 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 three thousand markets have two or three potential entrants. A monopoly is the most frequent outcome, with no entrant being the second likely outcome. The higher value of δ magnifies the competitive effect and leads to fewer entrants.

[Insert Table 1]

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 (Table 2). 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.

[Insert Table 2]

5.3 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 et al. (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 3 compares the estimates of four different simulation draw settings. The data generating process assumes $\delta = 2.0$. The first experiment makes twenty independent simulation draws, while the second experiment uses antithetic sampling to make twenty simulation draws, i.e. ten symmetric replications of ten independent pseudo-random draws to make simulation draws more systematic. The results show that, first, even with only twenty independent simulation draws, the comparison of the true parameter values and estimated values indicates overall accuracy given the estimated standard errors. Second, the use of antithetic sampling improves the model fit in terms of the average log likelihood value and reduces bias in terms of mean squared error. Third, increasing the number of draws to forty and one hundred shows a further improvement in the fit, though the improvement is small. This pattern is consistently observed in simulations with different values of parameters and different seeds of pseudo-random number generator. The results show accuracy even with a very small number of simulation draws, because the covariance structure in the model is simple and the correctly specified specification is used in these experiments. 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.

[Insert Table 3]

The next series of experiments examines the effect of misspecification by imposing restrictions on the correctly specified model (Table 4). The data generating process assumes $\delta = 2.0$ and each estimation makes forty simulation draws using antithetic sampling. The

first restricted model assumes that the econometrician has no correct knowledge about the decision order thus estimates the model by 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 binary 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.

[Insert Table 4]

Table 5 reports the results of the same comparison for the case of weak strategic effect: $\delta = 1.0$. Overall the results are consistent with the previous table. While misspecification still leads to substantial bias, the impact of misspecification is smaller when the strategic effect is smaller.

[Insert Table 5]

The last set of experiments introduces various degrees of randomness in the decision order. This is motivated by the fact that in many potential applications, the econometrician may have limited information that reflects the true decision order only approximately or with measurement error. Specifically, while the estimated models still assume that the 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 6 reports the results for different degrees of randomness. The results indicate that when the econometrician correctly specifies more than about 85 percent of the decision order, the differences between the estimated coefficients and their population values are smaller than the estimated standard error.

[Insert Table 6]

6 DISCUSSION AND EXTENSIONS

6.1 The Perfect Information Assumption

The perfect information assumption plays a key role in guaranteeing a unique subgame-perfect equilibrium. The uniqueness is necessary to specify the domain of integration in the ε space for each strategy profile that rationalizes the observed game outcome, without making a strong assumption on the equilibrium selection mechanism. The perfect information assumption, however, may be too strong in many applications. The assumption does not hold when some players have private information, when players move simultaneously, and when "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 to relax 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 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.

6.2 Decision Order

The entry game example in the previous section assumes that each firm makes a decision sequentially. In general, the proposed method 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 method 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 a significantly biased estimate of strategic effect. At the same time, 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

¹³Endogenizing 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

6.3 Computational Feasibility

For applications with relatively simple games, the computation burden of the proposed estimation procedure is fairly manageable. This is due to the high performance of the GHK simulator. Conducting all the Monte Carlo experiments shown in this paper requires less than a half day with a standard high-end DELL desktop computer purchased in 2007.

However, as the number of players, the number of turns, or the number of alternatives increases, the size of the game tree increases exponentially and computation quickly becomes infeasible. Though this exponential computational burden is inherent in the nature of sequential games, the following computation techniques may significantly 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 2) from the simulation procedure. In Maruyama (2011), I exploit the non-increasing property of the profit function in the number of entering rival firms; imposing this structure dramatically reduces the computation time.

Second, given the assumed independence across each game play, parallel computing is a promising way to reduce computational burden; the parallelization of the maximum likelihood evaluation loop is straightforward. Third, variance reduction techniques will enhance

(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.

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).

Acknowledgements Comments from Victor Aguirregabiria, Han Hong, Susumu Imai, Andy McLennan, Robert Porter, William Schworm, and Junichi Suzuki 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."

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] 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.
- [5] Berry, S. T. (1992) Estimation of a Model of Entry in the Airline Industry. *Econometrica*, 60(4), 889-917.
- [6] 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.
- [7] 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.
- [8] Bresnahan, T. F., and P. C. Reiss (1991) Empirical-Models of Discrete Games. *Journal of Econometrics*, 48(1-2), 57-81.
- [9] Chernozhukov, V., H. Hong, and E. Tamer (2007) Estimation and Confidence Regions for Parameter Sets in Econometric Models. *Econometrica*, 75, 1243-1284.
- [10] Ciliberto, F., and E. Tamer (2009) Market Structure and Multiple Equilibria in Airline Markets. *Econometrica*, 77(6), 1791-1828.
- [11] 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.

- [12] 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.
- [13] Geweke, J., M. Keane, and D. Runkle (1997) Comparing Simulation Estimators for the Multinomial Probit Model. *Journal of Econometrics*, 80, 125-166.
- [14] 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.
- [15] Hajivassiliou, V., and D. McFadden (1998) The Method of Simulated Scores for the Estimation of LDV Models. *Econometrica*, 66, 863-896.
- [16] 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.
- [17] 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.
- [18] Hiedeman, B. (1998) A Stackelberg Model of Social Security Acceptance Decisions in Dual-Career Households. *Journal of Economic Behavior & Organization*, 34, 263-78.
- [19] Hotz, V.J., and R.A. Miller (1993) Conditional Choice Probabilities and the Estimation of Dynamic Models. *Review of Economic Studies*, 60, 497-529.
- [20] Jia, P. (2008) What Happens When Wal-Mart Comes to Town: An Empirical Analysis

- of the Discount Retailing Industry. *Econometrica*, 76(6), 1263-1316.
- [21] Jia, Z. (2005) Labor Supply of Retiring Couples and Heterogeneity in Household Decision-Making Structure. *Review of Economics of the Household*, 3(2), 215-233.
- [22] Keane, M. (1994) A Computationally Practical Simulation Estimator for Panel Data. *Econometrica*, 62(1), 95-116.
- [23] Kempf, H., and G. Rota-Graziosi (2010) Endogenizing Leadership in Tax Competition. *Journal of Public Economics*, 94, 768-776.
- [24] Konrad, K. A., H. Künemund, K. E. Lommerud, and J. R. Robledo (2002) Geography of the Family. *American Economic Review*, 92(4), 981-998.
- [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*, C. Manski and D. McFadden (Eds.), Cambridge, MA: MIT Press, 305-319.
- [28] Liesenfeld, R., and J. F. Richard (2010) Efficient Estimation of Probit Models with Correlated Errors. *Journal of Econometrics*, 156, 367-376.
- [29] Litan, C., and C. Pimienta (2008) Conditions for Equivalence between Sequentiality and Subgame Perfection. *Economic Theory*, 35(3), 539-553.

- [30] Maruyama, S. (2011) Socially Optimal Subsidies for Entry: The Case of Medicare Payments to HMOs. *International Economic Review*, 52(1), 105-129.
- [31] Maruyama, S., and M. Johar (2013) Do Siblings Free-Ride in ‘Being There’ for Parents. UNSW Australian School of Business Research Paper, No. 2013-ECON06, UNSW.
- [32] Mazzeo, M. J. (2002) Product Choice and Oligopoly Market Structure. *RAND Journal of Economics*, 33(2), 221-242.
- [33] McFadden, D. (1989) A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration. *Econometrica*, 57(5), 995-1026.
- [34] McFadden, D., and P. A. Ruud (1994) Estimation by Simulation. *Review of Economics and Statistics*, 76, 591-608.
- [35] Pakes, A., M. Ostrovsky, and S. Berry (2007) Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples). *RAND Journal of Economics*, 38(2), 373-399.
- [36] Pakes, A., J. Porter, K. Ho, and J. Ishii (2011) Moment Inequalities and Their Application. Unpublished Manuscript, Harvard University.
- [37] Pesendorfer, M., and P. Schmidt-Dengler (2003) Identification and Estimation of Dynamic Games. Working Paper 9726, NBER.
- [38] Sándor, Z., and P. András (2004) Alternative Sampling Methods for Estimating Multivariate Normal Probabilities. *Journal of Econometrics*, 120, 207-234.
- [39] Schmidt-Dengler, P. (2006) The Timing of New Technology Adoption: The Case of

MRI. mimeo.

- [40] Seim, K. (2007) An Empirical Model of Firm Entry with Endogenous Product-Type Choices. *RAND Journal of Economics*, 37, 619-640.
- [41] 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.
- [42] Train, K. (2003) *Discrete Choice Methods with Simulation*. Cambridge, Cambridge University Press.

Table 1: EXAMPLE OF DATA SET: DISTRIBUTION OF MARKETS BY NUMBER OF ENTRANTS

Number of actual entrants		Number of potential entrants						Total
		1	2	3	4	5	6	
(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
Total		556	794	846	528	215	61	3,000

Table 2: 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%

Table 3: POTENTIAL SIMULATION BIAS: $\delta = 2.0$

DGP θ	20 draws no antithetics			20 draws			40 draws			100 draws		
	$\bar{\hat{\theta}}$	\overline{ASE}	MSE	$\bar{\hat{\theta}}$	\overline{ASE}	MSE	$\bar{\hat{\theta}}$	\overline{ASE}	MSE	$\bar{\hat{\theta}}$	\overline{ASE}	MSE
cons -5.0	-5.086	0.146	0.133	-5.081	0.147	0.130	-5.078	0.147	0.127	-5.077	0.147	0.128
pop 1.2	1.204	0.032	0.027	1.207	0.032	0.027	1.207	0.032	0.027	1.208	0.032	0.027
dist 0.0	0.008	0.022	0.019	0.008	0.022	0.019	0.008	0.022	0.019	0.008	0.022	0.019
pastp 0.4	0.402	0.018	0.018	0.402	0.018	0.017	0.402	0.018	0.017	0.401	0.018	0.017
city2 1.5	1.521	0.041	0.040	1.517	0.041	0.037	1.516	0.041	0.036	1.516	0.041	0.036
nroute 0.0	0.016	0.033	0.027	0.012	0.033	0.024	0.011	0.033	0.023	0.010	0.033	0.023
δ 2.0	1.986	0.074	0.083	1.997	0.074	0.077	2.001	0.074	0.076	2.003	0.074	0.076
ρ 0.7	0.677	0.029	0.033	0.685	0.029	0.026	0.688	0.029	0.027	0.689	0.029	0.025
\overline{LogL}	-3136.85			-3134.57			-3133.95			-3133.84		

Note: Monte Carlo experiments for 20 independent data sets with 3,000 markets. DGP $\theta \equiv$ parameter values used to generate data, $\bar{\hat{\theta}} \equiv$ average parameter estimate, $\overline{ASE} \equiv$ average asymptotic standard error, MSE \equiv root mean squared error, $\overline{LogL} \equiv$ average log likelihood value.

Table 4: RESTRICTED MODELS: $\delta = 2.0$

DGP θ	Full Model			No Order Info			No Market Error			Probit		
	$\bar{\theta}$	\overline{ASE}	MSE	$\bar{\theta}$	\overline{ASE}	MSE	$\bar{\theta}$	\overline{ASE}	MSE	$\bar{\theta}$	\overline{ASE}	MSE
cons -5.0	-5.078	0.147	0.127	-5.417	0.141	0.427	-5.241	0.129	0.260	-4.474	0.126	0.534
pop 1.2	1.207	0.032	0.027	1.077	0.031	0.124	1.072	0.026	0.130	0.625	0.021	0.575
dist 0.0	0.008	0.022	0.019	0.007	0.020	0.017	0.007	0.017	0.017	0.006	0.020	0.014
pastp 0.4	0.402	0.018	0.017	0.413	0.019	0.022	0.423	0.019	0.029	0.391	0.018	0.019
city2 1.5	1.516	0.041	0.036	1.576	0.039	0.083	1.627	0.039	0.131	1.561	0.036	0.072
nroute 0.0	0.011	0.033	0.023	0.281	0.028	0.282	0.180	0.031	0.182	0.382	0.027	0.383
δ 2.0	2.001	0.074	0.076	1.460	0.070	0.543	1.416	0.051	0.588			
ρ 0.7	0.688	0.029	0.027	0.466	0.041	0.237						
$\overline{LogL} / \overline{BIC}$	-3133.95 / 6339.95			-3218.43 / 6508.92			-3200.29 / 6472.63			-3609.46 / 7290.99		

Note: 40 simulation draws with antithetic sampling for 20 data sets with 3,000 markets. DGP $\theta \equiv$ parameter values to generate data, $\bar{\theta} \equiv$ average parameter estimate, $\overline{ASE} \equiv$ average asymptotic standard error, MSE \equiv root mean squared error, $\overline{LogL} \equiv$ average log likelihood value, $\overline{BIC} \equiv$ average Bayesian information criterion.

Table 5: RESTRICTED MODELS: $\delta = 1.0$

DGP θ	Full Model			No Order Info			No Market Error			Probit		
	$\bar{\theta}$	\overline{ASE}	MSE	$\bar{\theta}$	\overline{ASE}	MSE	$\bar{\theta}$	\overline{ASE}	MSE	$\bar{\theta}$	\overline{ASE}	MSE
cons -5.0	-5.046	0.140	0.130	-5.183	0.139	0.217	-4.986	0.121	0.122	-4.786	0.126	0.242
pop 1.2	1.204	0.032	0.031	1.172	0.032	0.041	1.048	0.024	0.155	0.893	0.022	0.308
dist 0.0	0.005	0.021	0.019	0.005	0.020	0.019	0.006	0.016	0.019	0.006	0.017	0.018
pastp 0.4	0.401	0.018	0.017	0.409	0.018	0.019	0.416	0.018	0.024	0.418	0.018	0.024
city2 1.5	1.506	0.042	0.034	1.541	0.041	0.051	1.602	0.038	0.105	1.633	0.038	0.138
nroute 0.0	0.007	0.028	0.022	0.087	0.026	0.089	0.108	0.029	0.111	0.167	0.028	0.169
δ 1.0	0.993	0.064	0.075	0.859	0.060	0.155	0.489	0.035	0.513			
ρ 0.7	0.690	0.028	0.029	0.637	0.029	0.070						
$\overline{LogL} / \overline{BIC}$	-3329.17 / 6730.40			-3341.70 / 6755.45			-3431.66 / 6935.39			-3464.88 / 7001.83		

Note: 40 simulation draws with antithetic sampling for 20 data sets with 3,000 markets. DGP $\theta \equiv$ parameter values to generate data, $\bar{\theta} \equiv$ average parameter estimate, $\overline{ASE} \equiv$ average asymptotic standard error, MSE \equiv root mean squared error, $\overline{LogL} \equiv$ average log likelihood value, $\overline{BIC} \equiv$ average Bayesian information criterion.

Table 6: EFFECT OF RANDOMNESS IN DECISION ORDER

Randomness in sequence	misspecified order (%)	δ ($\delta_0 = 2.0$)			ρ ($\rho_0 = 0.7$)			\overline{LogL}
		$\widehat{\delta}$	\overline{ASE}	MSE	$\widehat{\rho}$	\overline{ASE}	MSE	
0%	0.0%	2.001	0.074	0.076	0.688	0.029	0.027	-3133.95
10%	0.0%	2.001	0.074	0.076	0.688	0.029	0.027	-3133.95
20%	1.8%	1.992	0.075	0.080	0.683	0.029	0.029	-3135.86
30%	13.0%	1.928	0.075	0.111	0.654	0.031	0.052	-3154.22
40%	25.7%	1.857	0.076	0.161	0.623	0.033	0.081	-3176.04
50%	37.1%	1.801	0.077	0.209	0.596	0.034	0.107	-3199.64
60%	45.9%	1.754	0.077	0.254	0.573	0.036	0.130	-3218.71
70%	52.4%	1.717	0.077	0.291	0.560	0.037	0.145	-3233.78
80%	57.2%	1.696	0.077	0.311	0.554	0.037	0.151	-3243.32
90%	60.9%	1.679	0.077	0.327	0.547	0.038	0.157	-3250.80
100%	63.8%	1.669	0.077	0.338	0.544	0.038	0.161	-3251.66

Note: misspecified order indicates how many observations are assigned with different decision order. $\widehat{\delta}, \widehat{\rho} \equiv$ average parameter estimate, $\overline{ASE} \equiv$ average asymptotic standard error, MSE \equiv root mean squared error, $\overline{LogL} \equiv$ average log likelihood value. Experiments based on 40 simulation draws using antithetic sampling for 20 independent data sets with 3,000 market.

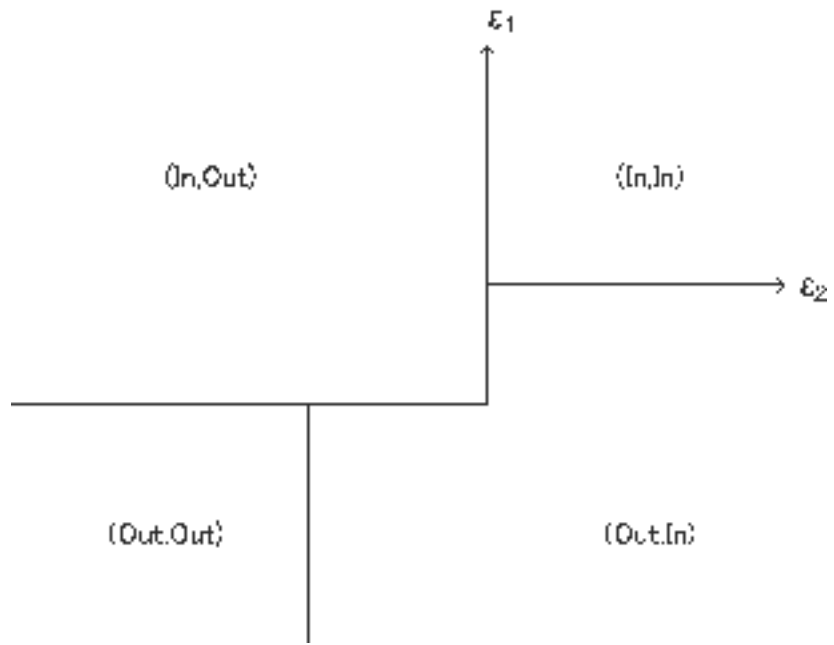


Figure 1: TWO PLAYER STACKELBERG ENTRY GAME

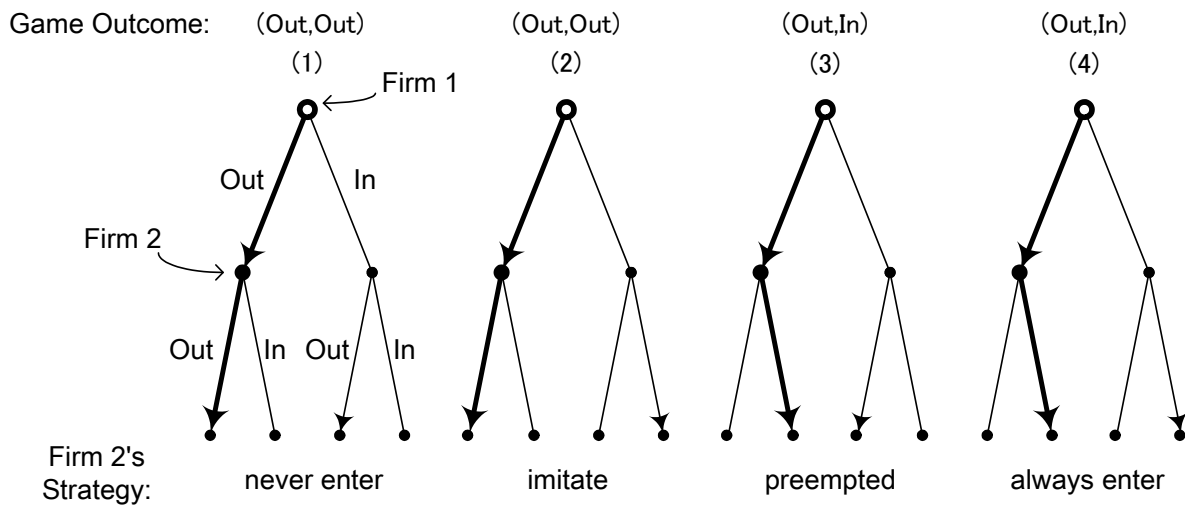


Figure 2: STRATEGIES AND OUTCOMES IN THE EXTENSIVE FORM WHEN FIRM 1 CHOOSES "OUT"

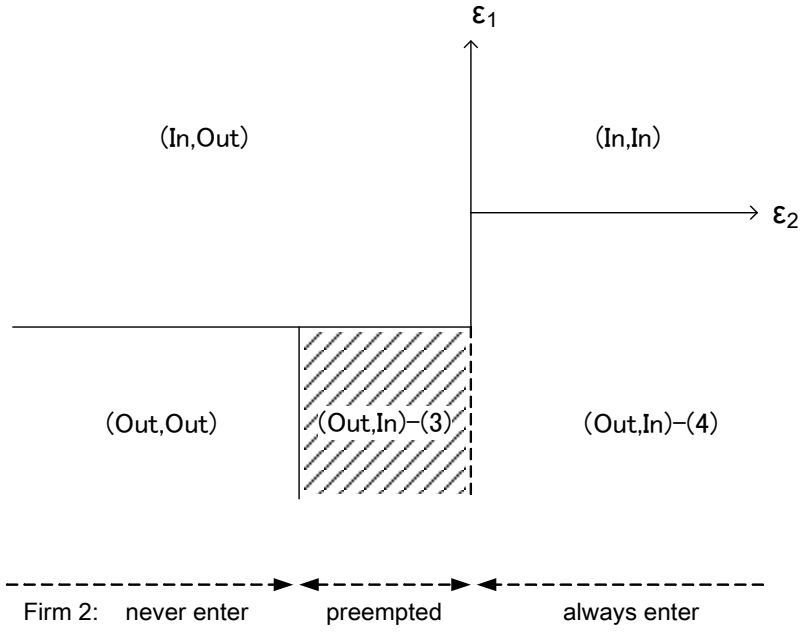


Figure 3: DIVIDING AN OBSERVED MARKET OUTCOME INTO STRATEGY PROFILES