Structural Estimation of Sequential Games of Complete Information*

JASON R. BLEVINS

Department of Economics, The Ohio State University

November 1, 2014†

Forthcoming in *Economic Inquiry*

Abstract. In models of strategic interaction, there may be important order of entry effects if one player can credibly commit to an action (e.g., entry) before other players. If one estimates a simultaneous-move model, then the move-order effects will be confounded with the payoffs. This paper considers nonparametric identification and simulation-based estimation of sequential games of complete information. Relative to simultaneous-move games, these models avoid the problem of multiple equilibria and require fewer payoff normalizations. We apply the estimator in several Monte Carlo experiments and to study entry-order effects using data from the airline industry.

Keywords: static games, sequential games, identification, simulation-based estimation, airline industry.

JEL Classification: C57, C15, L93.

1. Introduction

There has been much recent work on identification and estimation of static models of strategic interaction. Static models can be classified according to the timing of players' moves, which can either be *simultaneous* or *sequential*, and the informational assumptions, where players have either *complete* or *incomplete* information about the payoffs of their rivals. These two dimensions of differentiation are shown in Table 1.

Most previous work involving structural econometric models of static games has focused on games where players move simultaneously (panels B and D of Table 1). Bresnahan and Reiss

*I am grateful to Han Hong, Paul Ellickson, and Arie Beresteanu for many helpful discussions and to the members of the Duke Applied Microeconometrics Reading Group for useful feedback. This work was supported in part by an allocation of computing time from the Ohio Supercomputer Center.

†This paper is a revision of a working paper whose first draft appeared in October 2006.

(1991), Berry (1992), Tamer (2003), Ciliberto and Tamer (2009), and Bajari, Hong, and Ryan (2010), among others, have studied simultaneous-move games of complete information in a static setting (panel B). In contrast, Bajari, Hong, Krainer, and Nekipelov (2010b) considered simultaneous-move games under the assumption of incomplete information (panel D). A notable exception is Einav (2010), who considered a static, incomplete-information game with sequential moves (panel C). In contrast, this paper focuses on sequential-move games of complete information (panel A), which have received no attention in the literature to date to the author's knowledge.

Information \ Timing of Moves		
	A. Sequential moves, Complete information	B. Simultaneous moves, Complete information
	C. Sequential moves, Incomplete information	D. Simultaneous moves, Incomplete information

TABLE 1. Classification of Empirical Models of Static Games

From the perspective of estimation, simultaneous-move games have the unfortunate property that they give rise to multiple Nash equilibria. This makes estimation of the structural parameters difficult and so the literature has focused on ways to overcome the issues that arise. An early solution proposed by Bresnahan and Reiss (1991) and Berry (1992) was to consider a reduced outcome space over which there is a unique equilibrium outcome. They estimated entry models using information on outcomes only in terms of the number of players that enter rather than the identities of the players that enter. This permitted their models to be estimated but entailed a loss of information.

Tamer (2003) later showed that in a simple 2×2 simultaneous-move game, the parameters can be point identified even when there are multiple equilibria provided that a player-specific variable with large support is available. Bajari et al. (2010) considered joint estimation of the payoffs and equilibrium selection mechanism in models with many players. Ciliberto and Tamer (2009) developed a partially identified approach which avoids specifying the equilibrium selection mechanism and estimates the identified of structural parameters.

In contrast, we show that empirical models based on sequential games of complete information have an almost-surely unique subgame perfect Nash equilibrium (SPNE) and do not suffer from the problems of multiplicity that plague estimation of simultaneous-move games. Furthermore, the equilibrium can be computed very easily, which is far from the case in simultaneous-move games (Bajari, Hong, Krainer, and Nekipelov, 2010a; Bajari, Chernozhukov, Hong, and Nekipelov, 2007). Finally, counterfactual experiments can be carried out using the model without concern for the complications of multiple equilibria.

In their classic paper, Bresnahan and Reiss (1991) develop an empirical model of static discrete games of complete information and consider the estimation of both simultaneous- and sequential-move games in their framework. They briefly consider a two-player sequential-move entry game and note that uniqueness of the equilibrium seems to favor using sequential-move games over simultaneous-move games but that the apparent need to specify the order of moves, with little guidance on how to do so, is a drawback. In this paper, we show that the researcher does not need to specify the order of moves, rather, it can be treated as stochastic with the distribution of possible orderings being an unknown of interest.

We note that specifying a sequential-move game is different than specifying a simultaneous-move game with a specific equilibrium selection rule. Berry (1992) took the latter approach in estimating a simultaneous-move model of airline entry. The existence of multiple equilibria, due in large part to firm-heterogeneity, substantially complicates estimation of the model. In order to select a particular equilibrium, Berry assumes that firms enter in order of profitability. Gayle and Luo (forthcoming) use non-nested model selection tests (Vuong, 1989) to assess various order-of-entry assumptions in an entry game between McDonald's and Burger King. Although similar in spirit, the difference is that the sequential-move counterpart has a unique equilibrium for any order of moves and one can estimate the distribution of the order of moves as opposed to specifying that a particular equilibrium (selected implicitly by an order assumption) is played.

In addition to the technical advantages mentioned so far, sequential games can also shed light on the nature of competition in industry studies by capturing phenomena such as market power and entry deterrence. If early-movers have advantages in a market but one estimates a simultaneous-move game, then the early move advantage could be mistakenly attributed to higher entry costs or overly large competitive effects. Thus, ignoring order of entry effects can result in biased structural parameter estimates.

In this paper, we consider the case where the order of moves is not observed by the econometrician (although is known by the players) and treat it as an object of interest along with the payoff functions. This is by far the leading case since players in the model can typically make actions that cannot easily be observed by the researcher. For example, in a model of entry it is easy to observe when a firm has entered but difficult to know precisely, in terms of timing, when a firm decided not to enter. In other words, even if we are able to observe the order of *entry* we may not be able to discern the order of *moves*, which is the full sequence of decisions about whether to enter the market or remain out.

In our identification analysis, we show that even in very simple entry models the observable distribution of outcomes provides information about the order of moves. We compare in detail the role of the sequential- or simultaneous-move structure of the game in determining which payoffs can be identified. Additionally, we examine the nonparametric identification of the model and the role of payoff and order-selection exclusion restrictions.

In previous work on sequential games, Einav (2010) estimated a model of movie release timing under the assumption of incomplete information. He showed that the incomplete information assumption and a parametric assumption about the distribution of unobservables yields an analytic form for the likelihood function conditional on the order of moves. He cited an intractable likelihood function as one reason to prefer the assumption of incomplete information over that of complete information. However, complete information models may be preferable in cases where players have accurate information about the payoffs of their rivals, for example, due to a high level of industry experience, transparency, or to long-term, repeated interactions with the same rivals. Furthermore, we show that despite the intractable likelihood function, the complete information model can be estimated using simulation methods. This issue is certainly not unique to sequential games: Berry (1992), Bajari et al. (2010), and others have proposed the use of simulation for integration in simultaneous-move games as well.

A second issue of interest is the distribution of the order of moves. When the order of moves is unobserved the model is incomplete unless an assumption is made about its distribution. Einav ultimately assumed that the order of moves is uniformly distributed (i.e., that each of the $N!$ possible permutations of the N players is assigned equal probability). However, this assumption may not be innocuous and one may be interested in estimating the distribution of the order of moves. Relative to Einav, we consider games with a different informational assumption, we formally analyze identification of the model, including the distribution of the order of moves, and we apply our model to study order of entry effects in a different industry. Therefore, the methods we propose fill a gap in the literature (represented by panel A of Table 1) and allow researchers to choose from the full menu of simultaneous or sequential games of complete or incomplete information as needed.

In Section 2, the basic stochastic sequential game framework is developed. Section 3 establishes a minimal set of normalizations that must be imposed and considers nonparametric identification of the remaining free mean payoff values. Section 4 introduces our proposed simulation-based estimator using maximum simulated likelihood (MSL) (Lerman and Manski, 1981; Gouriéroux and Monfort, 1991; Lee, 1992).[1] Section 5 discusses a set of Monte Carlo experiments based on a simple two-player entry model. Finally, Section 6 applies the proposed estimator to study entry into city-pair markets in the airline industry.

2. Econometric Models of Sequential Games

Sequential games are games of *perfect information*, meaning that at each stage, each player observes the moves of all preceding players but not the moves of subsequent players. In such

[1] The method of simulated moments (MSM) (McFadden, 1989) can also be used, as we discuss in Appendix A.

games, every information set is a singleton and players know precisely where they are in the game tree at each move. If the players have complete information, meaning that the payoffs (including the unobservables) are common knowledge to all players, then a unique SPNE can be found through backwards induction. In this paper we consider the properties of econometric models based on such games.

Sequential games provide additional flexibility relative to simultaneous games by allowing for order of move effects. Depending on the shape of the reaction functions, there may be advantages to moving first or moving last (Gal-Or, 1985). A well-known result of the classic Stackelberg-Cournot model is that the firm that chooses it's output first will earn higher profits. Yet, in the Stackelberg-Bertrand variation of the model in which firms choose prices instead of quantities, the follower can undercut the price of the leader to earn higher profits. Models with heterogeneous firms yield even more potential patterns of interaction and move-order effects.

Because of the possibility of either first- or last-mover advantage in sequential move games, simultaneous-move games may seem preferable since they do not give market power to any particular player. Yet, allowing some firms to have market power due to move-order effects (while also not imposing that any firm has such power) may be important in the industry of interest and therefore is a useful feature for empirical models to possess. Otherwise, if one does not allow for move order effects they become confounded with the direct competitive effects on profits in a complicated way, leading to biased estimates of the model parameters.

In this section, we propose a framework for modeling strategic interaction as a sequential game without making strict assumptions about the order of moves. One can estimate the distribution of the order of moves in an attempt to determine which firms are more or less likely to lead or follow and thus enjoy such market advantages (or fail to enjoy them, as the case may be for the model at hand). This is possible because the conditional outcome distribution provides information about the order of moves in the sense that certain outcomes are more likely under some orderings than others. For example, if there are two identical players in a Stackelberg model then we could identify the first mover with a high probability by choosing the player with the highest observed output.

2.1. The Basic Model

Suppose that the econometrician observes M independent instances of some economic interaction that can be modeled as a sequential game of complete information (e.g., entry decisions in geographically separate markets). Each observation $m = 1, \dots, M$ will be referred to as a market. Each market consists of N players indexed by $i = 1, \dots, N$ and each player can choose one action $a_i \in \mathscr{A}_i$, where $\mathscr{A}_i = \{0, 1, \ldots, J_i - 1\}$ is a finite set representing player i's choice set.[2] A profile of actions of all players $a = (a_1, \ldots, a_N)$ is called an *outcome* and the set of all outcomes is $\mathscr{A} = \mathscr{A}_1 \times \cdots \times \mathscr{A}_N$. Players receive payoffs $u_i(a)$ for each outcome $a \in \mathscr{A}$.

In many cases, for simplicity, we will consider the case where all players have the same number of choices (i.e., $J_i = J$ for all i). Furthermore, we assume that the identities of the players are the same across markets. Identity may be taken literally (e.g., a particular airline) or it may be thought of more broadly as a category (e.g., a low cost carrier, a chain store, etc.).

In sequential games, unlike simultaneous games, the order of moves must be specified to complete the model. Each possible ordering is a permutation of the numbers $1, \ldots, N$, of which there are $N!$ possible permutations. It is customary to denote the initial stage, corresponding to the root node of the game tree, as stage 0. Hence, the last player moves at stage $N - 1$ and the outcome of the game can be thought of as stage N. Let $o : \{0, \ldots, N - 1\} \to \{1, \ldots, N\}$ denote a particular order of moves, where $o(k)$ is the index of the player who moves at stage k. This ordering is a random variable O with support $\mathscr{O}$ where $\mathscr{O}$ is the set of all $N!$ permutations of N players. We also use the inverse permutation $o^{-1}(i)$ to denote the stage at which player i moves.

Formally speaking, the model is an extensive form game with N stages. At each stage, only one player moves and each player moves exactly once during the course of the game. Each information set consists of a single node. Let $h^k = (a_{o(0)}, a_{o(1)}, \ldots, a_{o(k-1)})$ denote the history of actions at the beginning of stage k and let $\mathscr{H}^k$ denote the set of possible stage-k histories, where $\mathscr{H}^0 \equiv \varnothing$.

It is important to distinguish between histories and outcomes. When speaking of a particular history, it is implicitly assumed that the order of moves is known. On the other hand, the order of moves is irrelevant in an outcome, which is simply a profile of actions listed in the order in which players were labeled. This labeling is arbitrary.

Players move at information sets. However, since there is a unique history of actions leading to each information set we can equivalently express a strategy for player i (moving at stage $k = o^{-1}(i)$) as a function $s_i : \mathscr{H}^k \to \mathscr{A}_i$ which maps $\mathscr{H}^k$, the set of possible stage-k histories, to $\mathscr{A}_i$, player i's choice set. Thus, a strategy must specify a unique action for each possible history leading up to player i's move. The set of possible strategies for player i is $\mathscr{S}_i = \{s_i : \mathscr{H}^k \to \mathscr{A}_i\}$. There is no history for the leading player to consider, so we define $\mathscr{H}^0 = \varnothing$. A strategy for this player is simply an action, so $\mathscr{S}_{o(0)} = \mathscr{A}_{o(0)}$. For each $h \in \mathscr{H}^k$, there are J possible actions and

[2] In general multistage games, players may move simultaneously at each stage and the choices available to each player depend on the history of previous actions of all players. In this situation player i's choice set at stage k is $\mathscr{A}_i(h^k)$ for some history h^k. However, in this framework, only one player moves at each stage (the other players have empty choice sets). Each player moves exactly once and $\mathscr{A}_i$ cannot depend on the history of moves or the order of play.

there are J^k histories in $\mathscr{H}^k$ and therefore $J \cdot J^k = J^{k+1}$ possible strategies for player $o(k)$.

Sequential games have well-known and attractive equilibrium properties. For our purposes, the most important feature is given by the following theorem. See Zermelo (1913), Kuhn (1953), or Fudenberg and Tirole (1991) for details.

Zermelo's Theorem (1913)**.** *Every finite game of perfect information has a pure-strategy Nash equilibrium.*

2.2. The Econometric Model

In each market $m = 1, \dots, M$, the econometrician observes a vector of covariates x which is the realization of some random vector X with support $\mathscr{X}$. These covariates may include variables which affect the payoffs of all players (market-specific variables) as well as variables that affect only the payoffs of particular players. In addition, an equilibrium outcome a is observed in each market. For a given outcome a, player i receives payoff $u_i(a, x, \varepsilon_i(a))$, where $\varepsilon_i(a)$ is a random variable which captures any factors that are unobserved by the econometrician but affect the payoff player i receives in the event that outcome a obtains. Note that the payoffs depend on the actions of all players, not just the action of player i. This is the fundamental difference between game theoretic models and single agent discrete choice models. In single agent models, the payoffs are independent of a_{-i} since there is no strategic interaction.

We now state our assumptions about the structure of the payoffs. Let $\varepsilon_i = (\varepsilon_i(a))_{a \in \mathscr{A}}$ and let ε denote the vector $(\varepsilon_1, \dots, \varepsilon_N)$. Identifying the distribution of ε is difficult even in simple single agent models (Matzkin, 1992, 1993; Rust, 1994; Magnac and Thesmar, 2002; Aguirregabiria, 2010; Blevins, 2014). As such, we take this distribution as given.

Assumption 1 (Distribution of Unobservables)**.** For all $x \in \mathscr{X}$, the conditional distribution of ε given $X = x$ is absolutely continuous with respect to Lebesgue measure, has support $\mathscr{E} \subset \mathbb{R}^{NJ^N}$, and has a known conditional cdf $G(\cdot \mid X = x)$.

Under this assumption and in light of Zermelo's Theorem, because the game in question is a game of complete information, each realization of ε induces an optimal strategy profile $s = (s_1, \dots, s_N)$, which is unique with probability one since ε is continuously-distributed.

Proposition 1. *Under Assumption 1, the model has an almost surely unique SPNE.*

Proof. The existence of a SPNE is well known (see, e.g., Fudenberg and Tirole (1991) for a proof). Now, suppose to the contrary that there are two distinct SPNE strategy profiles $s = (s_1, \dots, s_N)$ and $\tilde{s} = (\tilde{s}_1, \dots, \tilde{s}_N)$. Let i denote the player moving at the latest stage in the game for which $s_i \neq \tilde{s}_i$. Let $k = o^{-1}(i)$ denote that stage and let $\overline{\mathscr{H}}$ denote the set of stage-k histories for which $s_i \neq \tilde{s}_i$.

For both strategies s_i and $\tilde{s}_i$ to be elements of SPNE strategy profiles it must be the case that $f_i(s_i(h), s_{-i}(h), x) + \varepsilon_i(s_i(h), s_{-i}(h)) = f_i(\tilde{s}_i(h), s_{-i}(h), x) + \varepsilon_i(\tilde{s}_i(h), s_{-i}(h))$ for all $h \in \overline{\mathcal{H}}$. In other words, $\varepsilon_i(s_i(h), s_{-i}(h)) = f_i(\tilde{s}_i(h), s_{-i}(h), x) - f_i(s_i(h), s_{-i}(h), x) + \varepsilon_i(\tilde{s}_i(h), s_{-i}(h))$ for all $h \in \overline{\mathcal{H}}$. By Assumption 1, this event has probability zero and so the SPNE is unique with probability one. ■

Henceforth, we will drop the "almost surely unique" qualifier and simply refer to *the* SPNE. Let $\alpha(u, o)$ denote the SPNE outcome of the game given payoffs u and an order of moves o. In light of Proposition 1, α is a function (as opposed to a correspondence) with probability one. Now, for simplicity we focus on the case of additively separable payoffs.

Assumption 2 (Additive Separability)**.** Player i's payoff function can be written as

$$u_i(a, x, \varepsilon_i) = f_i(a, x) + \varepsilon_i(a),$$

where we refer to $f_i(a, x) \equiv \mathrm{E}[u_i(a, x, \varepsilon_i) \mid a, x]$ as the mean payoff and to $\varepsilon_i(a)$ as the unobserved component.

When the order of moves is unobserved, to complete the model the econometrician must consider the probability distribution over the possible permutations that can occur.[3] This is similar to the situation with static simultaneous move games, where completing the model involves incorporating a latent equilibrium selection variable (Tamer, 2003; Bajari et al., 2010). In sequential games with a small number of players, one might specify a fully nonparametric order selection mechanism which assigns a separate probability to each permutation in $\mathcal{O}$. However, this requires estimating $N! - 1$ parameters and so in practice it is probably only feasible for small values of N. At the other extreme, it is possible to simply assign probability one to a single permutation if there is some compelling reason to believe that it always occurs. For example, the econometrician may believe that more profitable firms enter before their less profitable rivals (Berry, 1992). In between these two extremes there are many possible parametric distributions over $\mathcal{O}$. For example, Einav (2010) estimates a sequential game of incomplete information in which each permutation is assumed to occur with equal probability. Other possible distributions can depend on the covariates x. Specifying a flexible parametric form allows a compromise between a fully non-parametric approach and imposing *ad hoc* assumptions about the probabilities. This approach also remains feasible for large values of N, where the number of nonparametric order selection primitives would increase factorially in N. As such, we make the following assumptions about the order selection mechanism.

[3] An incomplete model would not condition on O and would therefore admit an equilibrium *correspondence* containing all outcomes which are equilibrium outcomes for *some* ordering $o \in \mathcal{O}$.

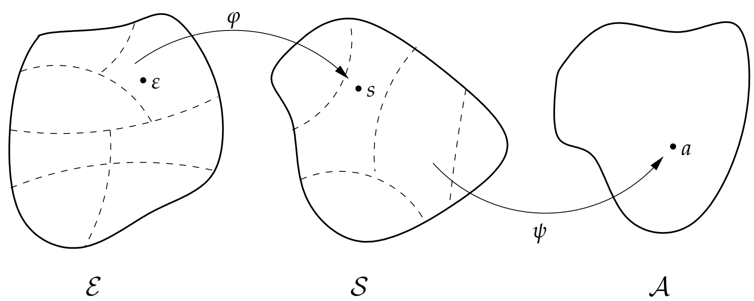


FIGURE 1. Unobservables, strategies, and outcomes (conditional on X and O)

Assumption 3 (Order Selection Mechanism)**.** The order of moves, O, is independent of ε conditional on X and is iid across markets. Let $\mu : \mathscr{O} \times \mathscr{X} \to [0,1]$ with $\mu(o,x) = \Pr(O = o \mid X = x)$ denote the associated conditional probability mass function.

Both the mean payoffs, f_i, and the order selection mechanism, μ, are left quite general for now so as to discuss nonparametric identification in Section 3. Later, when we turn to estimation, we will assume that these primitives are known up to a finite vector of parameters.

2.3. Stochastic Properties

Expressing the probability of an outcome in the sequential game framework is more complex than doing so in a simultaneous-move game because of the recursive nature of the model. In a normal form game, a strategy is simply a single action. In a sequential game a strategy is a mapping from the set of player i's information sets (or the set of histories) to the choice set $\mathscr{A}_i$. This is crucial to the construction of the choice probabilities, but it also makes it more difficult to express the outcome probabilities in a tractable way.

Thus, before formally looking at identification, it is useful to develop some notation and look more closely at the relevant probability spaces. Conditional on X and O, the randomness in the model derives from the payoff disturbances $\varepsilon_i(a)$. By Proposition 1, each $\varepsilon \in \mathscr{E}$ induces a unique SPNE strategy profile $s \in \mathscr{S}$ which in turn induces a unique outcome $a \in \mathscr{A}$. Let $\phi : \mathscr{E} \to \mathscr{S}$ denote the function which maps each ε to the corresponding unique SPNE strategy profile s and let $\psi : \mathscr{S} \to \mathscr{A}$ be the function that maps each strategy profile s to the resulting outcome in $\mathscr{A}$. The relationships between these spaces is illustrated in Figure 1 (conditional on X and O). Since ε is unobserved, inference is based on the induced probability measure over the set of observed outcomes $\mathscr{A}$.

The following proposition establishes that these functions are surjective (or onto) and that the inverse functions define disjoint sets. These properties are of interest for two reasons. First, we require the model to be saturated so that it assigns nonzero probability to any potentially observable outcome. Second, it allows us to characterize the inverse images under these functions, which are the regions of integration that define the outcome probabilities used for estimation.

Proposition 2. *If Assumptions 1 and 2 hold, then the mappings ϕ and ψ are surjective for all $X \in \mathscr{X}$ and $O \in \mathscr{O}$. Furthermore, the sets in each of the collections $\{\phi^{-1}(s)\}_{s \in \mathscr{S}}$ and $\{\psi^{-1}(a)\}_{a \in \mathscr{A}}$ are pairwise disjoint.*

Proof. Since $\mathscr{S}_i = \{s_i : \mathscr{H}^{o^{-1}(i)} \to \mathscr{A}_i\}$ and since $\varepsilon_i(a)$ has full support for all i and a, for any $s_i \in \mathscr{S}_i$ we can choose the vector $\varepsilon_i(a)$ so that s_i is a dominant strategy. Hence, given any strategy profile s, we can construct an ε for which s is the SPNE. Similarly, for each $a \in \mathscr{A}$, we can construct an s for which a is the equilibrium outcome. In particular, $s_i(h) = a_i$ for all $h \in \mathscr{H}^{o^{-1}(i)}$ is a valid strategy for each player i. The strategy profile $s = (s_1, \ldots, s_N)$ trivially induces the outcome a for any ordering o. Hence, the inverse images of these functions are nonempty.

The second conclusion follows from a fundamental property of functions. Suppose to the contrary that there exist elements $s, s' \in \mathscr{S}$ so that $\varphi^{-1}(s) \cap \varphi^{-1}(s') \neq \varnothing$ and let $\varepsilon \in \varphi^{-1}(s) \cap \varphi^{-1}(s')$. Then, $\varphi(\varepsilon) = s$ since $\varepsilon \in \varphi^{-1}(s)$. But this is a contradiction since $\varepsilon \in \varphi^{-1}(s')$ implies $\varphi(\varepsilon) = s' \neq s$. The proof is analogous for ψ. ■

Proposition 2 guarantees that for any outcome a, there is a corresponding *nonempty* set of strategies $\psi^{-1}(a)$ in $\mathscr{S}$ which induce a. Similarly, for any strategy profile s, there is a corresponding nonempty set of unobservables $\phi^{-1}(s)$ which induce s. Finally, combining these results, for any outcome a, there is a set of unobservables which induce a, given by $\phi^{-1} \circ \psi^{-1}(a)$. This is the region of integration in the support of ε corresponding to the probability of observing outcome a, conditional on X and O. Proposition 2 guarantees that this region is nonempty for each a.

The probability of observing any particular outcome is well-defined under Proposition 2 and can be constructed for a given set of primitives by using the inverse structural mappings to express the set of unobservables which induce the outcome in question. First of all, the probability of an outcome a is the probability that any strategy in $\psi^{-1}(a)$ is played, so

$$\Pr(a \mid x, o) = \Pr(\phi^{-1} \circ \psi^{-1}(a) \mid x, o).$$

Furthermore, for all a

$$\phi^{-1} \circ \psi^{-1}(a) = \bigcup_{s \in \psi^{-1}(a)} \phi^{-1}(s).$$

Since $\psi^{-1}(a) \subset \mathscr{S}$ is finite and since the sets $\phi^{-1}(s)$ are disjoint for all s by Proposition 2, we can use finite additivity to write the probability of outcome a as

$$\Pr(a \mid x, o) = \sum_{s \in \psi^{-1}(a)} \Pr(\phi^{-1}(s) \mid x, o). \tag{1}$$

This provides a concise representation of an otherwise intractable recursive expression.

Let $h_+^{k+1}(h^k, s, x)$ denote the moves of subsequent players following the stage-k history h^k, given that players follow the strategies specified by the strategy profile s. For each $s \in \psi^{-1}(a)$ we have

$$
\begin{aligned}
\Pr(\phi^{-1}(s) \mid x, o) = \Pr\Big(& s_{o(0)}(\varnothing) = \arg\max_j u_{o(0)}(j, h_+^2(j,s), x, \varepsilon), \\
& s_{o(1)}(h^1) = \arg\max_j u_{o(1)}(h^1, j, h_+^3(h^1, j, s), x, \varepsilon) \; \forall h^1 \in \mathscr{H}^1, \\
& \vdots \\
& s_{o(N-1)}(h^{N-1}) = \arg\max_j u_{o(N-1)}(h^{N-1}, j, x, \varepsilon) \; \forall h^{N-1} \in \mathscr{H}^{N-1} \Big| \, x, o \Big)
\end{aligned}
\tag{2}
$$

This expression is analytically intractable, but it is straightforward to approximate via simulation. We describe the details of this approach when we discuss estimation in Section 4. Before considering estimation, we turn to identification.

3. Identification

In this section, we determine which features of the model are nonparametrically identified and under what conditions. One benefit of establishing identification in this general sense is that it also lends credibility to the estimates of carefully specified parametric models. That is, even though data limitations might dictate a parametric specification in practice, one can be sure that identification is not achieved only through functional form assumptions if the model is shown to be nonparametrically identified.

The following sections present several definitions and assumptions that will be required to establish the identification results that follow. Several simple two-player examples will be used to emphasize the subtle differences between simultaneous- and sequential-move games. As with single-agent, static discrete choice models, several normalizations will need to be imposed. We will then use these normalizations as a benchmark to examine the conditions required to nonparametrically identify the model.

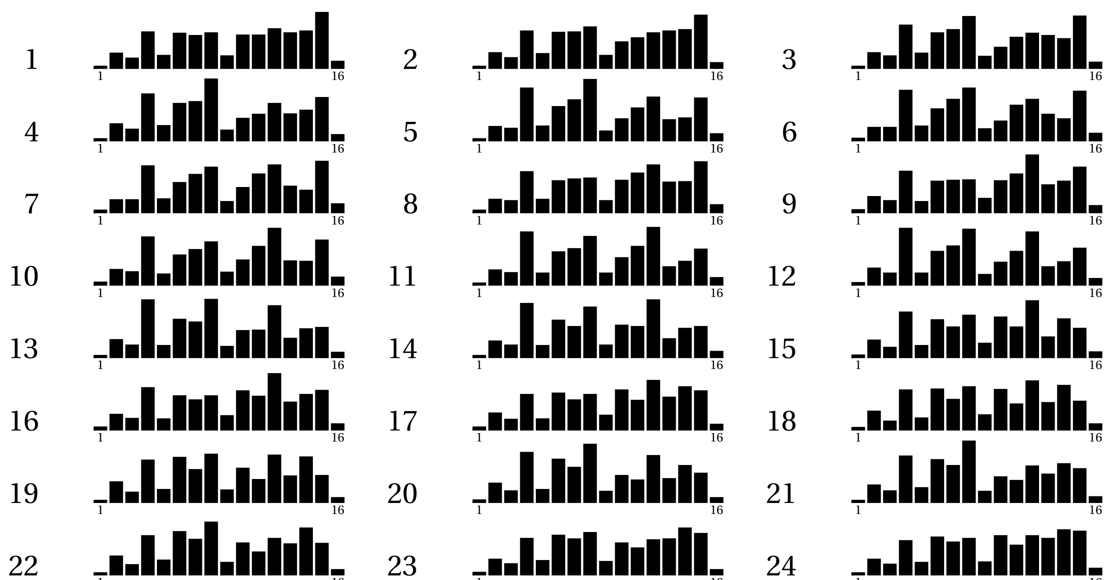
3.1. Intuition for Identification

To provide a brief example of the identification problem faced, consider a simple entry model with $N = 4$ players. We have simulated 1,000 markets (according to the model used below for the Monte Carlo experiments described in Section 5, holding X fixed) and plotted the observed reduced form distribution of outcomes and the unobserved structural distributions—the outcome distributions for each particular order of moves as well as the distribution of permutations.

Figure 2(a) depicts the simulated frequency of observed outcomes. Along the horizontal axis are the $2^4 = 16$ possible outcomes. This distribution corresponds to the reduced form of the model. Figure 2(b) depicts the simulated frequency of the order of moves, with the $4! = 24$ different permutations along the horizontal axis. Finally, Figure 2(c) depicts the different underlying distributions of the 16 outcomes conditional on each of the 24 permutations. Figures 2(b) and 2(c) correspond to the structural model. Essentially, the goal is to identify these structural distributions using only the observed distribution of outcomes, the observed covariates, and the properties of the model. Identification therefore requires there to be sufficient variation in the conditional distributions implied by the model.

(a) Distribution of Outcomes $A \mid X$ (observed)

(b) Order Selection Mechanism $O \mid X$ (unobserved)



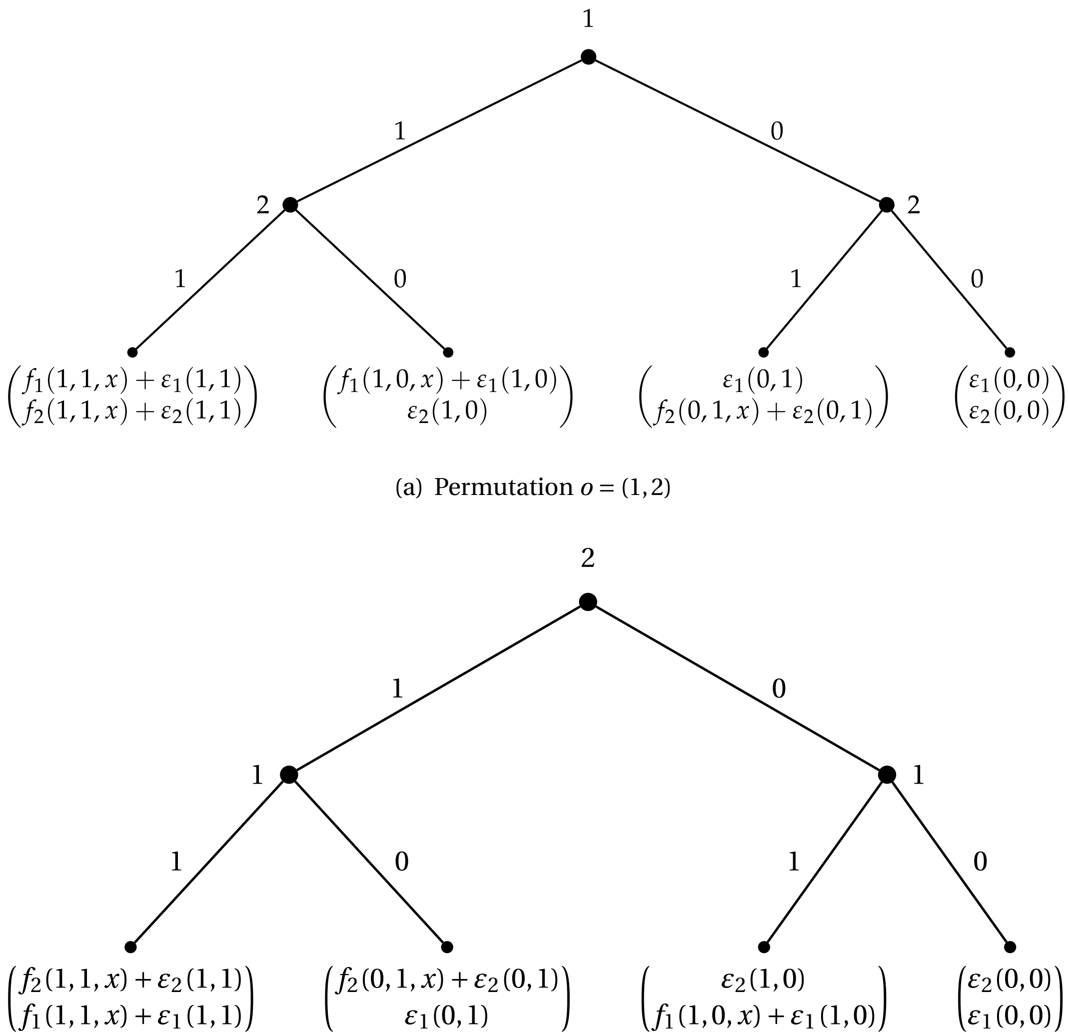
(c) Conditional Outcome Distributions $A \mid X, O = o, o \in \mathscr{O}$ (unobserved)

FIGURE 2. Simulated Reduced Form and Structural Distributions (Conditional on X)

Payoffs				Outcome given o		Different	Consistent
$f_1(1,1)$	$f_1(1,0)$	$f_2(1,1)$	$f_2(0,1)$	$o=(1,2)$	$o=(2,1)$	Outcomes	w/Theory
-	-	-	-	$(1,1)$	$(1,1)$	No	Yes
+	-	-	-	$(1,1)$	$(1,1)$	No	No
-	+	-	-	$(0,1)$	$(0,1)$	No	Yes
+	+	-	-	$(0,1)$	$(0,1)$	No	Yes
-	-	+	-	$(1,1)$	$(1,1)$	No	No
+	-	+	-	$(0,0)$	$(0,0)$	No	No
-	+	+	-	$(1,1)$	$(0,1)$	Yes	No
+	+	+	-	$(0,0)$	$(0,0)$	No	No
-	-	-	+	$(1,0)$	$(1,0)$	No	Yes
+	-	-	+	$(1,0)$	$(1,1)$	Yes	No
-	+	-	+	$(0,1)$	$(1,0)$	Yes	Yes
+	+	-	+	$(0,1)$	$(0,1)$	No	Yes
-	-	+	+	$(1,0)$	$(1,0)$	No	Yes
+	-	+	+	$(0,0)$	$(0,0)$	No	No
-	+	+	+	$(1,0)$	$(1,0)$	No	Yes
+	+	+	+	$(0,0)$	$(0,0)$	No	Yes

TABLE 2. Payoff Structures and Outcomes, $N=2$, $\#A=2$

There is little such variation in the entry model so a relatively large sample size is required to obtain accurate parameter estimates. To see why, consider the deterministic two-player entry model depicted in Figures 3(a) and 3(b). If the profit from not entering is zero, then there are only four nonzero payoffs remaining: the monopoly and duopoly profits for each player. Considering cases where each of these payoffs is either positive or negative, there are 16 possible games. These possibilities are listed in Table 2 along with the resulting outcomes in each case for both possible orders of play. There are only three cases which yield different outcomes. Furthermore, theory dictates that monopoly profit should dominate duopoly profit. This rules out two of the these cases. Therefore, there is only a single case (negative duopoly payoff, positive monopoly payoff) which is both theoretically consistent and yields observationally distinct outcomes depending on the order of moves. Thus we may need a large sample size in order to obtain good estimates of the order selection mechanism in this model. Fortunately, this case is a very simple two-player entry model. Models with more players, more actions, or with fewer *a priori* theoretical restrictions can yield more observationally distinct outcomes from which to identify the order selection mechanism.



(a) Permutation $o = (1,2)$

(b) Permutation $o = (2,1)$

FIGURE 3. A sequential two-player entry game

3.2. Reduced Form and Structure

A reduced-form analysis in this setting would attempt to explain the relationship between the equilibrium outcomes and the covariates based on a random sample $\{y_m, x_m\}_{m=1}^M$. The system through which players interact is viewed as a black box. The reduced form of the model can be summarized by $\lambda(x)$, the vector of conditional outcome probabilities $\lambda(a,x) = \Pr(a \mid x)$ for all $a \in \mathscr{A}$.

Now consider the structural model. The fundamental component of the structural model is the structure of the game itself, with an important part of that structure being the equilibrium function α. In light of Assumptions 1, 2, and 3, the primitives of the model are the mean payoffs $f_i(a,x)$ and the order selection probabilities $\mu(o,x)$ for each $i = 1,\dots,N$, $a \in \mathscr{A}$, $x \in \mathscr{X}$, and $o \in \mathscr{O}$.

For simplicity, since we have assumed the state space is finite, let f denote the vector of all mean payoff primitives and let μ denote the vector of all order selection probabilities. Let Γ denote the set of all structures, feasible (f,μ) combinations. We are interested in identifying as many of the components of (f,μ) as possible. Essentially, we want to be able to express each primitive as a function of observed population moments.

Every structure (f,μ) induces a reduced form $\lambda(x)$ through the mapping $\lambda(x) = \Lambda(f(x),\mu(x)) = (\Lambda(f(a,x),\mu(x)))_{a\in\mathscr{A}}$, where

$$\Lambda(f(a,x),\mu(x)) = \sum_{o\in\mathscr{O}} \left[\int 1\{\alpha(f(x)+\varepsilon,o) = a\}\, dG(\varepsilon \mid x)\right] \mu(o,x).$$

The goal is to uncover the true structure (f_0,μ_0) which satisfies

$$\Lambda(f_0(a,x),\mu_0(x)) = \sum_{o\in\mathscr{O}} \Pr(a \mid x,o)\mu(o,x) = \Pr(a \mid x) = \lambda_0(x).$$

3.3. Definitions and Assumptions

For a given conditional distribution of unobservables, G, through the structure of the model the primitives induce a particular distribution of observable equilibrium outcomes. This relationship is the mapping $\Lambda(f,\mu)$ described in Section 3.2. The inverse mapping Λ^{-1} partitions the set of primitives into equivalence classes in the sense that any primitives in the same equivalence class induce the same reduced form. This leads us to the following definitions of observational equivalence and identification (Hurwicz, 1950; Koopmans, 1949; Matzkin, 2007).

Definition. The primitives (f,μ) and $(\tilde{f},\tilde{\mu})$ are *observationally equivalent* if $\Lambda(f,\mu) = \Lambda(\tilde{f},\tilde{\mu})$.

If two distinct sets of primitives are consistent with the observed reduced form, the researcher cannot determine which of these primitives generated it. More formally, if for two distinct sets of primitives (f,μ) and $(\tilde{f},\tilde{\mu})$ in Γ we have $\Lambda(f,\mu) = \Lambda(\tilde{f},\tilde{\mu})$, then the model is not identified because (f,μ) and $(\tilde{f},\tilde{\mu})$ are observationally equivalent. Upon observing the reduced form $\Lambda(f,\mu)$, one can distinguish the model primitives (f,μ) from $(\tilde{f},\tilde{\mu})$. This leads to the following notion of identification.

Definition. The model is *identified* if $(f,\mu) \neq (\tilde{f},\tilde{\mu})$ implies $\Lambda(f,\mu) \neq \Lambda(\tilde{f},\tilde{\mu})$.

Thus, if each structure $(f,\mu) \in \Gamma$ induces a unique reduced form $\Lambda(f,\mu)$, then no two structures are observationally equivalent. Stated differently, if Λ is one-to-one, then the model is identified. First, we narrow the list of primitives which are identified by showing that without additional restrictions, some mean payoff values cannot be identified and must be normalized.

3.4. Mean Payoff Normalization: Simultaneous and Sequential Games Compared

Although the necessary mean payoff normalizations are well known for the cases of single-agent discrete choice models (McFadden, 1974; Maddala, 1983) and simultaneous-move discrete games (Bresnahan and Reiss, 1991; Tamer, 2003; Bajari et al., 2010), the nature of strategies and interactions is subtly different in sequential move discrete games. As we will show, this allows us to identify more outcome-specific mean payoffs, and thus requires fewer mean payoff normalizations.

Normal form games are straightforward generalizations of single-agent discrete choice problems. In a normal form game, for given values of the explanatory variables, x, and rival actions, a_{-i}, player i faces a simple choice between the elements of the choice set $\mathscr{A}_i$ which correspond to the set of available strategies. Thus, for each x and a_{-i}, one of player i's mean payoffs must be normalized because the level of payoffs cannot be identified (i.e., only differences in payoffs can be identified). This is analogous to the normalization required in single-agent discrete-choice problems.

Sequential games are also generalizations of single-agent discrete choice problems, however the relationship is not as straightforward due to the more complex nature of strategies in extended form games. For a given player i, holding the strategies of player i's rivals s_{-i} fixed, player i faces a discrete choice problem of choosing a strategy $s_i \in \mathscr{S}_i$. We show below that the resulting pattern of payoff comparisons is much richer than in simultaneous games, and as a result fewer of the mean payoffs need to be normalized.

To motivate the need to normalize the mean payoffs, consider the simultaneous two-player, two-action game in Figure 4(a) and the corresponding sequential game in Figure 4(b). In both cases, the choice sets are $\mathscr{A}_1 = \mathscr{A}_2 = \{0,1\}$. There are four possible outcomes $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$ and therefore eight mean payoff primitives. For the purposes of this example, suppose that the idiosyncratic outcome-specific shocks $\varepsilon_i(a)$ are iid across players and outcomes so that outcome probabilities are more tractable.

The standard normalization in the simultaneous game of Figure 4(a) requires normalizing four of the eight mean payoffs. That is, for each strategy of player 1, one of player 2's mean payoffs must be normalized and similarly for player 1's payoff. The reason for this normalization is due to the nature of comparisons that occur. Player 1 will choose 0 if, given $s_2 = 0$,

$$f_1(0,0,x) + \varepsilon_1(0,0) > f_1(1,0,x) + \varepsilon_2(1,0),$$

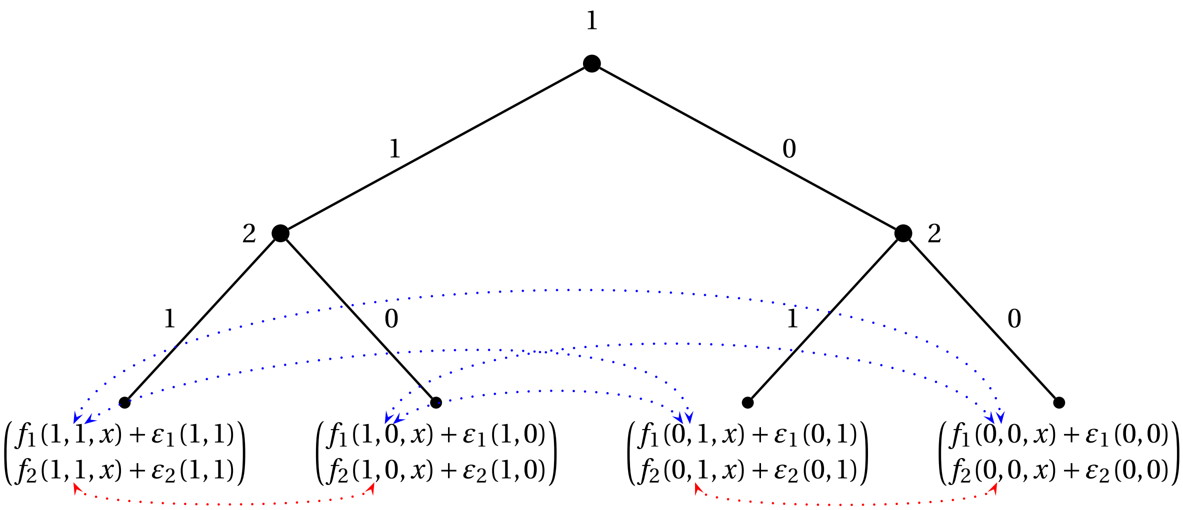
or if, given $s_2 = 1$,

$$f_1(0,1,x) + \varepsilon_1(0,1) > f_1(1,1,x) + \varepsilon_2(1,1).$$

Similar comparisons are made by player 2. Notice that for each player, there are only two

		Player 2	
		0	1
Player 1	0	$f_1(0,0,x)+\varepsilon_1(0,0)$; $f_2(0,0,x)+\varepsilon_2(0,0)$	$f_1(0,1,x)+\varepsilon_1(0,1)$; $f_2(0,1,x)+\varepsilon_2(0,1)$
	1	$f_1(1,0,x)+\varepsilon_1(1,0)$; $f_2(1,0,x)+\varepsilon_2(1,0)$	$f_1(1,1,x)+\varepsilon_1(1,1)$; $f_2(1,1,x)+\varepsilon_2(1,1)$

(a) Simultaneous-Move Game



(b) Sequential-Move Game, $o = (1,2)$

FIGURE 4. Payoff Comparisons in Two-Player, Two-Action Games

independent pairwise comparisons of payoffs. These comparisons are represented by the dotted lines in Figure 4(a). Both $f_1(0,0,x)$ and $f_1(1,0,x)$ cannot be jointly identified because, for example, an isomorphic game can obtained by adding some constant c_0 to both payoffs. Similarly, both $f_1(0,1,x)$ and $f_1(1,1,x)$ cannot be jointly identified because the outcome distribution is unchanged when a constant c_1 is added to both terms.

Now, consider the sequential two player entry game shown in Figure 4(b). Because there are two possible permutations, we first consider the permutation $o = (1,2)$. A strategy for player 1 is simply a single action s_1 and a strategy for player 2 consists of two history-contingent actions, $s_2 = (s_2(0), s_2(1))$. Although the payoffs are common knowledge and players can perfectly predict the equilibrium strategies, from the econometrician's point of view the probability that player 1 plays a particular strategy depends on all possible strategies of player 2. The conditional probability of an outcome $a = (a_1, a_2)$ is $\Pr(a \mid x, o) = \Pr(a_2 \mid a_1, x, o)\Pr(a_1 \mid x, o)$. Because we have assumed independence of the unobservables, we can immediately write down the probabilities $\Pr(a_2 \mid a_1, x, o)$ for each a_1:

$$\Pr(a_2 = 1 \mid a_1 = 0, x, o) = \Pr\left(f_2(0,1,x) + \varepsilon_2(0,1) > f_2(0,0,x) + \varepsilon_2(0,0) \mid x\right),$$
$$\Pr(a_2 = 1 \mid a_1 = 1, x, o) = \Pr\left(f_2(1,1,x) + \varepsilon_2(1,1) > f_2(1,0,x) + \varepsilon_2(1,0) \mid x\right).$$

Again using independence, the probability of a strategy is simply the sum of the probabilities of the individual actions at each information set. For example, the probability of $s_2 = (1,1)$ is

$$\begin{aligned}\Pr(s_2 = (1,1) \mid x, o) = \Pr&\left(f_2(0,1,x) + \varepsilon_2(0,1) > f_2(0,0,x) + \varepsilon_2(0,0) \mid x\right)\\ &+ \Pr\left(f_2(1,1,x) + \varepsilon_2(1,1) > f_2(1,0,x) + \varepsilon_2(1,0) \mid x\right).\end{aligned}$$

Here, there are only two independent pairwise comparisons between the four mean payoffs for player 2. So far, this is analogous to the simultaneous-move case.

Now, given s_2, player 1 compares $f_1(0, s_2(0), x) + \varepsilon_1(0, s_2(0))$ with $f_1(1, s_2(1), x) + \varepsilon_1(1, s_2(1))$. For example, the probability of $s_1 = 1$ is

$$\begin{aligned}\Pr(s_1 = 1 \mid x, o) &= \sum_{s_2 \in S_2} \Pr(s_1 = 1 \mid x, o, s_2) \Pr(s_2 \mid x)\\ &= \Pr(f_1(1,1,x) + \varepsilon_1(1,1) > f_1(0,1,x) + \varepsilon_2(0,1) \mid x) \Pr(s_2 = (1,1) \mid x)\\ &+ \Pr(f_1(1,0,x) + \varepsilon_1(1,0) > f_1(0,1,x) + \varepsilon_2(0,1) \mid x) \Pr(s_2 = (0,1) \mid x)\\ &+ \Pr(f_1(1,1,x) + \varepsilon_1(1,1) > f_1(0,0,x) + \varepsilon_2(0,0) \mid x) \Pr(s_2 = (1,0) \mid x)\\ &+ \Pr(f_1(1,0,x) + \varepsilon_1(1,0) > f_1(0,0,x) + \varepsilon_2(0,0) \mid x) \Pr(s_2 = (0,0) \mid x).\end{aligned}$$

For player 1, as shown in Figure 4(b), there are four distinct pairwise comparisons between the four payoffs (instead of the two comparisons in the simultaneous-move game). Therefore, it is only necessary to normalize one mean payoff here. To see this, note that adding a constant to all payoffs would indeed leave the strategy probability unchanged. In contrast, two normalizations are needed in the corresponding simultaneous-move game.

Thus, the system of payoff comparisons is richer than in the simultaneous-move game. Importantly, when we consider the other permutation, $o = (2,1)$, the roles of the players are switched. Each of player 2's pairwise payoff comparisons are relevant and so only one of player 2's mean payoffs must be normalized. In general, as long as the order selection probabilities are all nonzero, each player moves first with some positive probability and so all possible pairwise payoff comparisons are made. Thus, only one of the mean payoffs of each player needs to be normalized.

The following proposition generalizes the above result, that the mean payoffs can at most be identified up to a constant, beyond the simple two-player entry model.

Proposition 3. *For any $x \in \mathcal{X}$, at most $\prod_{i=1}^{N} J_i - 1$ of each player's mean payoffs are identified.*

Proof. Suppose that the vector of covariates $x \in \mathscr{X}$ is given. Let $a \in \mathscr{A}$ be an arbitrary outcome and let $\{f_i(a,x)\}_{i,a}$ be a collection of mean payoff primitives. The proof proceeds by constructing another set of mean payoffs $\{\tilde{f}_i(a,x)\}_{i,a}$ which differ from $\{f_i(a,x)\}$ by a fixed constant but which yield the same outcome probabilities. Note that for each player i, for any order of moves o with $o(0) = i$ (i.e., player i moves first) player i will make $\prod_{l=1}^{N} J_l$ distinct pairwise payoff comparisons. However, if one adds the same constant to each of player i's outcome-specific payoffs, then the outcome probabilities will remain unchanged. This also remains true for any order of moves, since any other ordering will result in fewer pairwise payoff comparisons. Therefore, at most $\prod_{l=1}^{N} J_l - 1$ of each player's mean payoff values can be identified. ■

3.5. Local Identification

Henceforth, in light of Proposition 3, we normalize the payoff for the outcome $a = (0,\dots,0)$ to zero for each player i and each x.

Assumption 4 (Mean Payoff Normalization)**.** For each $i = 1,\dots,N$ and $x \in \mathscr{X}$, $f_i(0,\dots,0,x) = 0$.

Let $f(x)$ denote the vector of the remaining free mean payoffs. Since the order selection probabilities must sum to one, let $\mu(x)$ denote the vector of probabilities for all but one permutation. Let Γ denote the set of all feasible values of $f(x)$ and $\mu(x)$. Finally, since the conditional outcome probabilities must sum to one, let $\Lambda(f(x),\mu(x))$ denote the induced outcome probabilities, with the outcome $(0,0,\dots,0)$ omitted. We will consider the nonparametric identification of the model in the following sense.

Definition (Local Identification)**.** Let $\Pr(a \mid x)$ be given and suppose that the primitives $(f,\mu) \in \Gamma$ satisfy $\Lambda(f(a,x),\mu(x)) = \Pr(a \mid x)$ for all $a \in \mathscr{A}$. Then the primitives (f,μ) are *locally identified* if there exists an open neighborhood $B(x) \subset \Gamma$ such that for each $(\tilde{f},\tilde{\mu}) \in B(x)$ with $(\tilde{f},\tilde{\mu}) \neq (f,\mu)$, we have $\Lambda(\tilde{f}(x),\tilde{\mu}(x)) \neq \Lambda(f(x),\mu(x))$.

Under the following conditions, $\Lambda(f(x),\mu(x))$ is locally invertible, which is a sufficient condition for local identification.

Assumption 5. $\Lambda(f,\mu)$ is continuously differentiable and for all $x \in \mathscr{X}$, $\mathrm{D}\,\Lambda(f(x),\mu(x))$ has full column rank, where $\mathrm{D}\,\Lambda$ denotes the Jacobian of Λ.

The rank assumption condition must be verified on a model-specific basis. Therefore, we focus here on finding general conditions under which the necessary order condition holds:

(3) $$\dim \Lambda(f(x),\mu(x)) \geq \dim \Gamma = \dim f(x) + \dim \mu(x).$$

This requires that there are at least as many equations as unknowns and requires counting the observable moments and unknown primitives and placing additional restrictions on the model when necessary.

3.6. Moments and Primitives

For a given x, there are a finite number of moment conditions and primitives. For the model to be identified, we require the number of moments to be at least as large as the number of primitives, the number of mean payoffs and order selection probabilities remaining after normalization. First we count the number of observed moment conditions and then the number of order selection and payoff primitives. For simplicity, in this section we focus on the case where all players i have $J_i = J$ choices.

The observed population moments are $\Pr(a \mid x)$ for all $a \in \mathscr{A}$. Since there are a finite number of outcomes, we know that the distribution of equilibrium outcomes must satisfy the adding-up condition $\sum_{a\in\mathscr{A}} \Pr(a \mid x) = 1$. One degree of freedom is lost and so there are only

$$J^N - 1 \tag{4}$$

linearly independent outcome probabilities.

Now, for a given value of x, the payoff structure of any single player can be described by $J^N - 1$ values, corresponding to the set of possible outcomes, with the payoff of one outcome normalized to zero. The total number of mean payoff primitives is thus

$$N(J^N - 1). \tag{5}$$

Finally, since the order selection mechanism is a probability mass function on a finite set $\mathscr{O}$, for a given x, it can be described by

$$N! - 1 \tag{6}$$

nonnegative real numbers, or rather a vector in the simplex $\Delta^{N!-1}$.

Proposition 4. *If Assumptions 1–5 are satisfied and $N \geq 2$, then the model is nonparametrically unidentified.*

Proof. A sufficient condition for showing the model is nonparametrically unidentified is that the number of primitives is larger than the number of moments. Thus, summing (5) and (6) gives $N(J^N - 1) + (N! - 1)$, but

$$N(J^N - 1) + (N! - 1) > N(J^N - 1) > J^N - 1$$

since $N \geq 2$. The right-hand side corresponds to (4), the number of potentially observable outcome probabilities, so the model is unidentified without further restrictions. ■

In light of Proposition 4, the model is nonparametrically unidentified without additional restrictions. To achieve identification, the additional restrictions must increase the number of observed moments faster than the number of primitives.

3.7. Exclusion Restrictions

We will discuss two types of exclusion restrictions: player-specific exclusions and order-selection exclusions. First, suppose that some covariates $z_i \in \mathcal{Z}_i$ are observed for player i which affect only the payoff of player i, given by

$$u_i(a, x, z_i, \varepsilon_i(a)) = f_i(a, x, z_i) + \varepsilon_i(a)$$

but not the payoffs of player i's rivals. Suppose further that there are covariates $z_\mu \in \mathcal{Z}_\mu$ which affect the order selection mechanism $\mu(o, x, z_\mu)$, but do not affect the payoff structure. For the purpose of discussing identification and counting moments, we restrict these variables to lie in finite sets.

Assumption 6. For all i, $\mathcal{Z}_i$ is finite with L elements and $\mathcal{Z}_\mu$ is finite with L_μ elements.

After taking these exclusion restrictions into account, after normalization there are $N(J^N - 1)L_\mu$ payoff primitives, $(N! - 1)\, L_\mu$ order selection probabilities, and $(J^N - 1)L^N L_\mu$ population moments. Nonparametric identification requires that the number of moments equals or exceeds the number of primitives:

$$(J^N - 1)L^N L_\mu \geq N(J^N - 1)L + (N! - 1)L_\mu. \tag{7}$$

Now, the number of moments is increasing exponentially in the number of exclusion restrictions L while the number of primitives is only linear in L. Thus, the model can always be identified if the player-specific covariates lie in a sufficiently rich set.

Proposition 5. *If Assumptions 1–6 hold, then:*

1. *For any values of N and J, the model is nonparametrically identified if L is sufficiently large.*
2. *For any value of N, if there is a binary payoff exclusion ($L = 2$) then the model is nonparametrically identified for a sufficiently large value of J.*

For example, in the airline entry application below with $N = 6$ airlines and a binary choice ($J = 2$), the order condition in (7) is satisfied when there are binary excluded variables z_i (i.e., $L \geq 2$), regardless of whether there is an order-selection exclusion (i.e., $L_\mu \geq 1$). In the application, we include continuous player-specific variables that satisfy the payoff exclusion restriction.

4. Estimation

4.1. Maximum Likelihood Estimation

Let θ denote the vector of parameters of interest, let $f_i(a, x, \theta)$ and $\mu(o, x; \theta)$ be parametric specifications for the mean payoff function and order selection mechanism, and let G be the conditional cdf of ε given X.

There are many possible parametric specifications for μ. For example, one can model the probability that a particular player i moves first as

$$\frac{p_i}{p_1 + \cdots + p_N}. \tag{8}$$

Conditional on player j moving first, the probability that player $i \neq j$ moves second is

$$\frac{p_i}{p_1 + \cdots + p_{j-1} + p_{j+1} + \cdots + p_N}$$

and so on. This order selection mechanism is described by $N-1$ parameters.

A similar specification, suggested by Einav (2010), depends on observed covariates in a logistic model for what we will refer to as the *first move propensity*. Let the probability that player i moves first be

$$\frac{e^{z_i^\top \rho}}{e^{z_1^\top \rho} + \cdots + e^{z_N^\top \rho}}. \tag{9}$$

Define the second and subsequent probabilities similarly.

Now, given a sample $\{y_m, x_m\}$ of outcomes and covariates in M markets, the log likelihood for the sample is

$$\ln L(\theta) = \sum_{m=1}^{M} \ln \Pr(y_m \mid x_m; \theta). \tag{10}$$

where the likelihood for a single observation is

$$\Pr(y_m \mid x_m; \theta) = \sum_{o \in \mathcal{O}} \Pr(y_m \mid x_m, o; \theta) \mu(o, x_m; \theta).$$

Given the parametric forms, the conditional outcome probability $\Pr(y_m \mid x_m, o; \theta)$ can be derived as a function of θ according to (1).

Unfortunately, estimating the model using maximum likelihood with (10) is likely to be infeasible. Deriving a closed form for (2) without an independence assumption on the shocks, $\varepsilon_i(a)$, requires taking the integral over a very irregular subset of $\mathcal{E}$. This may be possible through the use of truncated distributions or the GHK algorithm (cf. Hajivassiliou and Ruud, 1994) in the normal case, however there are no known closed forms for commonly used distributions such as the type I extreme value distribution. Thus, estimating the model using MLE may require numerical or Monte Carlo integration over a set that is difficult to quantify. For these reasons, we focus on simulation-based estimation methods.

4.2. Maximum Simulated Likelihood

For a given vector of parameters θ, the likelihood can be approximated using simulation by simulating outcomes in each market R times using R draws from the joint distribution of unobservables. Each draw is a vector of NJ^N outcome-player-specific payoff shocks. For each evaluation of the log-likelihood function, this requires a total of $MRNJ^N$ unobservables. For computational efficiency and to reduce chatter in the objective function, these draws should be stored at the beginning of the estimation routine and used for each evaluation of the simulated log-likelihood function.

Let $\varepsilon^{(r,m)}$ denote the r-th draw from the distribution of unobservables for market m and let $u^{(r,m)}(\theta) = f(x_m, \theta) + \varepsilon^{(r)}$ denote the resulting payoffs for the current value of θ. For each $u^{(r,m)}$ and each o, the simulated SPNE outcome $\alpha(u^{(r,m)}(\theta), o)$ is found using backwards induction. The simulated likelihood for market m is the simulated frequency of y_m, given by

$$\hat{P}_R(y_m \mid x_m; \theta) = \sum_{o \in \mathcal{O}} \left[\frac{1}{R} \sum_{r=1}^{R} 1\{\alpha(u^{(r,m)}(\theta), o) = y_m\} \right] \mu(o, x_m; \theta). \tag{11}$$

This is simply the proportion of the R simulated outcomes which are equal to the observed outcome y_m. The full simulated log likelihood is

$$\ln \hat{L}_M(\theta) = \sum_{m=1}^{M} \ln \hat{P}_R(y_m \mid x_m; \theta).$$

Simulating $\hat{P}_R(y_m \mid x_m; \theta)$ for each market can be done in parallel on multiple processors by simulating a particular subset of markets on each processor. R should be large relative to the total number of outcomes (and we compare several choices of R in the next section). In the appendix, we also discuss how importance sampling can be used to reduce the number of simulations required as in Ackerberg (2009).

As N increases, the number of terms in the summation in (11) increases as $N!$ since each permutation $o \in \mathcal{O}$ is considered. An alternative approach is to take a sample of draws from $\mu(o, x; \theta)$ and use them as representative draws to evaluate the sum. This is analogous to using Monte Carlo integration to approximate an integral or using a small subset of possible subsamples when using subsampling (Politis, Romano, and Wolf, 1999).

5. Monte Carlo Experiments

This section describes the results of a number of Monte Carlo experiments designed to shed light on both the small-sample and asymptotic properties of the estimators under various specifications. The experiments are carried out using a static entry model with $N = 4$ players which will also form the basis for the empirical application of Section 6.

The payoff function for firm i depends on common market characteristics $x \in \mathbb{R}$, firm attributes $z_i \in \mathbb{R}$, and outcome $a \in \mathscr{A}$ in the following way:

$$(12) \quad u_i(a, x, z_i, \varepsilon) = \begin{cases} \varepsilon_i(0, a_{-i}) & \text{if } a_i = 0, \\ \beta_0 + \beta_1 x + \gamma z_i - \delta \left(\sum_{j \neq i} a_j\right) + \varepsilon_i(1, a_{-i}) & \text{if } a_i = 1. \end{cases}$$

The decision to enter is denoted by $a_i = 1$ and the decision not to enter is denoted by $a_i = 0$. The coefficient δ denotes the competitive effect of the presence of other firms. The stochastic error term $\varepsilon_i(a)$ captures the unobserved factors that affect firm i's profits when the outcome a obtains. We choose the true parameters to be $\beta_0 = 0$, $\beta_1 = 1$, $\gamma = 0.5$, $\delta = 1$ and we assume that $\varepsilon_i(a)$ follows the standard normal distribution.

Data for each of M markets was generated by choosing values for each parameter and generating draws for each characteristic and unobservable term. Specifically, the market-wide covariates are $X_m \sim \chi^2(1)$, the firm-market-specific excluded variables are $Z_{i,m} \sim \mathrm{N}((3-i)/10, 2)$ for each player $i = 1, \ldots, N$.

The order of moves in each market was determined by a draw from the distribution of possible permutations. We considered three specifications for the order selection mechanism:

1. Uniform: $\mu(o) = 1/N!$ for all $o \in \mathscr{O}$.

2. Logistic: $\mu(o; p_1, \ldots, p_N)$ is calculated according to (8) for all $o \in \mathscr{O}$.

3. Index: $\mu(o, z_\mu; \rho)$ is calculated according to (9) for all $o \in \mathscr{O}$ and $z_\mu \in \mathscr{Z}_\mu$.

In the logistic case, we choose the true first-move propensities to be $p_1 = 0.25$, $p_2 = 0.05$, $p_3 = 0.20$, and $p_4 = 0.50$. In the index case, for each market m, $Z_{\mu,m} = (Z_{\mu,1,m}, \ldots, Z_{\mu,N,m})$ is a vector of length N, with one variable for each player which is drawn as $Z_{\mu,i,m} \sim \mathrm{N}((1-i)/10, 1)$ and we choose the true order selection mechanism parameter to be $\rho = 0.2$. The outcome in each market was found through backwards induction, using the true parameters, the randomly drawn market and firm characteristics and order of moves, and draws of the unobservables.

We carried out a series of Monte Carlo experiments for each of the three specifications described above (uniform, logistic, and index). For each experiment, we choose values of M, the number of markets, and R, the number of simulation draws used for estimation. For each specification and each choice of M and R, we simulated 25 datasets and obtained parameter estimates using MSL for each. We report the mean bias and the square root of the mean square bias (RMSE) for each parameter.

We seek to learn about the finite sample properties of the estimators and verify the asymptotic properties. As such, we perform experiments with $M = 200, 400, 800$ markets and with $R = 25, 50, 100, 200, 400$ simulation draws. MSL estimators are consistent as long as $R \to \infty$ and

		β_0		β_1		γ		δ	
M	R	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
	25	-0.063	0.320	-0.183	0.237	-0.408	0.464	0.267	0.389
	50	0.022	0.265	-0.059	0.168	-0.185	0.269	0.080	0.256
200	100	0.072	0.243	-0.023	0.121	-0.076	0.194	0.017	0.211
	200	-0.043	0.209	0.002	0.111	0.006	0.187	0.042	0.181
	400	-0.024	0.207	0.017	0.114	0.014	0.158	0.020	0.184
	25	0.002	0.247	-0.234	0.263	-0.522	0.542	0.267	0.341
	50	0.066	0.174	-0.135	0.191	-0.253	0.287	0.122	0.211
400	100	0.046	0.208	-0.037	0.119	-0.086	0.150	0.044	0.161
	200	0.037	0.137	-0.009	0.110	-0.039	0.141	0.005	0.113
	400	0.018	0.128	0.007	0.098	0.000	0.107	0.006	0.127
	25	0.047	0.170	-0.226	0.259	-0.493	0.508	0.214	0.262
	50	0.042	0.199	-0.148	0.197	-0.332	0.353	0.123	0.194
800	100	0.019	0.178	-0.066	0.111	-0.158	0.207	0.083	0.146
	200	0.043	0.099	-0.029	0.083	-0.079	0.116	0.019	0.087
	400	0.033	0.116	-0.012	0.076	-0.047	0.092	0.003	0.099

TABLE 3. Monte Carlo Results with Uniform O

$M \to \infty$ and they are $\sqrt{M}$-consistent and asymptotically normal when $R/\sqrt{M} \to \infty$ as $M \to \infty$. Therefore, we should expect the mean bias and RMSE to decrease as we increase both M and R.

Table 3 reports the results for the uniform order selection mechanism. In this case, each permutation arises with equal probability and so there are no parameters to estimate other than the payoff parameters. The top panel of the table shows how the mean bias and RMSE vary as we increase the number of simulation draws from $R = 25$ to $R = 400$, holding the number of markets fixed at $M = 200$. As expected, the bias and RMSE decrease for all parameters when increasing the number of simulation draws because the log likelihood approximation is more precise. Proceeding to the middle and lower panels, the number of markets doubles to $M = 400$ and then $M = 800$. Again, as expected increasing the number of markets also results in lower mean bias and RMSE values nearly uniformly.

Table 4 reports the results for the logistic order selection mechanism, which introduces three new parameters p_1, p_2, and p_3 (with $p_4 = 1 - p_1 - p_2 - p_3$). These first-move propensities are estimated fairly precisely even for as few as $R = 50$ simulation draws even though the order of moves is treated as being unobserved (i.e., the simulated permutations are not being used during estimation). Because we have introduced additional parameters, the remaining parameter estimates have slightly more variation but are still estimated rather precisely for even small values of R.

		β_0		β_1		γ		δ		p_1		p_2		p_3	
M	R	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
	25	0.022	0.275	-0.165	0.207	-0.066	0.132	0.167	0.267	0.030	0.082	0.132	0.154	0.041	0.103
	50	0.080	0.237	-0.023	0.157	-0.018	0.099	0.016	0.202	-0.016	0.080	0.051	0.065	0.099	0.174
200	100	0.088	0.217	0.013	0.114	0.017	0.104	-0.032	0.164	-0.003	0.078	0.056	0.064	0.076	0.153
	200	0.090	0.226	0.025	0.135	-0.004	0.069	-0.048	0.206	-0.024	0.061	0.043	0.053	0.078	0.153
	400	0.085	0.206	0.053	0.106	0.016	0.070	-0.078	0.163	-0.020	0.069	0.046	0.055	0.077	0.158
	25	0.032	0.202	-0.181	0.217	-0.081	0.139	0.191	0.280	0.019	0.070	0.110	0.117	0.062	0.087
	50	0.064	0.197	-0.052	0.111	-0.033	0.081	0.037	0.155	0.008	0.062	0.075	0.089	0.045	0.081
400	100	0.043	0.114	-0.003	0.082	-0.005	0.075	0.012	0.110	0.020	0.072	0.064	0.072	0.009	0.087
	200	0.064	0.196	0.026	0.087	0.010	0.067	-0.044	0.168	-0.010	0.059	0.048	0.058	0.035	0.098
	400	0.052	0.131	0.039	0.097	0.013	0.061	-0.048	0.141	-0.003	0.066	0.055	0.071	0.042	0.105
	25	0.015	0.126	-0.217	0.235	-0.112	0.147	0.216	0.244	0.014	0.043	0.135	0.139	0.047	0.062
	50	0.038	0.154	-0.093	0.122	-0.040	0.071	0.092	0.165	0.016	0.060	0.094	0.100	0.029	0.063
800	100	0.019	0.127	-0.051	0.087	-0.031	0.058	0.066	0.152	0.000	0.065	0.065	0.073	0.043	0.086
	200	0.035	0.092	0.025	0.067	-0.007	0.061	-0.018	0.088	-0.017	0.040	0.049	0.056	0.030	0.074
	400	0.043	0.099	0.011	0.048	0.000	0.043	-0.019	0.097	-0.024	0.046	0.042	0.047	0.031	0.076

TABLE 4. Monte Carlo Results with Normal ε and Logistic O

		β_0		β_1		γ		δ		ρ	
M	R	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
	25	0.014	0.233	-0.160	0.233	-0.041	0.125	0.172	0.279	-0.145	0.221
	50	0.046	0.226	-0.033	0.175	-0.003	0.132	0.036	0.217	-0.165	0.301
200	100	0.073	0.238	0.037	0.152	0.029	0.119	-0.051	0.204	-0.066	0.224
	200	0.067	0.222	0.035	0.140	0.019	0.082	-0.061	0.193	-0.009	0.326
	400	0.030	0.176	0.035	0.125	0.003	0.085	-0.028	0.127	-0.015	0.285
	25	0.059	0.176	-0.170	0.223	-0.078	0.113	0.169	0.260	-0.093	0.121
	50	-0.012	0.143	-0.049	0.145	0.005	0.076	0.086	0.196	-0.089	0.140
400	100	-0.019	0.108	0.012	0.106	0.005	0.071	0.019	0.126	0.026	0.219
	200	-0.034	0.137	0.020	0.115	0.018	0.074	0.016	0.128	0.006	0.179
	400	-0.015	0.109	0.036	0.097	0.013	0.076	-0.019	0.109	0.012	0.206
	25	0.051	0.132	-0.198	0.213	-0.103	0.135	0.180	0.208	-0.151	0.163
	50	0.026	0.118	-0.087	0.127	-0.051	0.088	0.085	0.167	-0.124	0.145
800	100	-0.007	0.134	-0.056	0.099	-0.035	0.056	0.071	0.148	-0.058	0.132
	200	0.052	0.109	0.004	0.064	0.004	0.054	-0.020	0.092	-0.055	0.112
	400	0.017	0.074	0.006	0.063	0.002	0.047	-0.006	0.074	-0.017	0.141

TABLE 5. Monte Carlo Results with Normal ε and Index O

Finally, Table 5 reports the results for the index order selection mechanism, which introduces a single new parameter ρ, which is the coefficient on the player-specific covariates in the order selection probabilities. The estimates behave similarly to the logistic case. Some parameter estimates appear to have more variation and other less than the logistic case, but this could be due to the specific parameterizations chosen. The situation is similar to that with the logistic order selection mechanism: we have more parameters than the uniform case but are still able to estimate both the payoff parameters and the order selection parameter accurately.

Overall, the results suggest that even with a small number of simulation draws and a modest number of markets, we can obtain useful estimates of both the payoff parameters and the order selection mechanism parameters. The same basic entry model and payoff function and the same index order selection mechanism used here will form the basis for the model used in the empirical application that follows.

6. An Application to Entry in the Airline Industry

6.1. Data

To investigate the importance of move-order effects in the airline industry, we use the data of Ciliberto and Tamer (2009), who in turn followed Borenstein (1989) and Berry (1992) in

TABLE 6. Summary Statistics for Market-Level Variables

Variable	Min.	Max.	Mean	S.D.	Obs.
Wright amendment	0	1	0.029	0.169	2742
Dallas airport	0	1	0.070	0.255	2742
Market size (million people)	0.310	15.236	2.259	1.846	2742
Per capita income ($10,000)	1.702	4.580	3.240	0.391	2742
Income growth rate (%)	-0.300	10.050	4.051	1.478	2742
Market distance (1,000 miles)	0.067	2.724	1.085	0.624	2742
Closest airport (100 miles)	0.102	1.505	0.346	0.205	2742
U.S. center distance (1,000 miles)	0.283	3.390	1.571	0.594	2742

constructing the dataset. The data was taken from the second quarter of the 2001 Airline Origin and Destination Survey (DB1B) provided by the U.S. Department of Transportation and which based on a 10% random sample of airline tickets sold within the quarter.

As is standard in this literature, a market is defined to be a trip between two cities without regard to the intermediate stops. The dataset contains $M = 2,742$ markets. The DB1B data be used to determine which airlines serve which markets (i.e., entry decisions). This information was supplemented with additional demographic and geographic data. Summary statistics for the variables we use are listed in Table 6 for the market-level variables and Table 7 for the carrier-level variables. For the entry decision, Southwest (WN) serves 25% of markets, United (A) serves 28%, American (AA) serves 43%, and Delta (DL) serves 55%. Other medium airlines (MA) serve 55% of the markets and other low-cost carriers (LC) serve 16% of the markets. We now briefly review the other variable definitions; see the above references for additional details.

Cost and Airport Presence The variable referred to as "cost" is the additional distance that a flight must travel if it makes a connection to the nearest hub rather than traveling non-stop, taken as a percentage of the total non-stop distance. This can be interpreted as an opportunity cost: the next best alternative to serving the market directly is for the carrier to serve it with a connecting flight via it's nearest hub. For the main carriers used in our analysis, average cost (in percentage terms) ranges from 0.736 for American to 0.303 for Southwest. As an example, if the non-stop distance is 1,000 miles and the connecting itinerary distance is 1,500 miles, then the cost is defined as $(1500 - 1000)/1000 = 0.5$ (i.e., the connecting flight distance is 50% higher than that of the direct flight).

Market power is measured by "airport presence" which is a function of the number of markets served out of the endpoint airports and was constructed by following Berry (1992). It is defined

TABLE 7. Summary Statistics for Carrier-Level Variables

Variable	Carrier	Min.	Max.	Mean	S.D.	Obs.
Entry (%)	AA	0	1	0.426	0.495	2742
	DL	0	1	0.551	0.497	2742
	UA	0	1	0.275	0.447	2742
	MA	0	1	0.548	0.498	2742
	LC	0	1	0.162	0.369	2742
	WN	0	1	0.247	0.431	2742
Cost (%)	AA	0	27.570	0.736	1.609	2742
	DL	0	27.570	0.420	1.322	2742
	UA	0	21.096	0.784	1.476	2742
	MA	0	11.620	0.229	0.615	2742
	LC	0	3.095	0.043	0.174	2742
	WN	0	16.180	0.303	0.860	2742
Airport presence (%)	AA	0	0.873	0.422	0.167	2742
	DL	0	0.987	0.540	0.181	2742
	UA	0	0.689	0.265	0.153	2742
	MA	0	0.850	0.376	0.135	2742
	LC	0	0.650	0.098	0.077	2742
	WN	0	0.863	0.242	0.176	2742

by taking the shares of markets served out of each endpoint airport (of the total served by at least one carrier) and averaging over both endpoints. Delta, American, and Southwest all have markets with over 85% presence. Southwest, followed closely by United, have the lowest average presence (24% and 27%, respectively). Delta has the highest average presence (54%). Airport presence influences profits, but it can also serve as a commitment device in the order selection mechanism.

Dallas and the Wright Amendment The Wright amendment was passed in 1979 to promote development of Dallas/Fort Worth airport. Traffic out of Dallas Love airport, the second major airport in the area, was restricted: only flights traveling within Texas or to neighboring states were permitted originally. Additional states were added in 1997 resulting in permitted flights between Dallas Love to destinations in Louisiana, Arkansas, Oklahoma, New Mexico, Alabama, Kansas, and Mississippi. The amendment was partially repealed in 2006 with some restrictions left intact until 2014. We control for the Wright Amendment using an indicator variable equaling

Parameter	Estimate	S.E.
Payoff function		
Airport presence	8.621	(0.063)
Cost	-0.439	(0.008)
Market size	0.117	(0.006)
Market distance	-0.133	(0.011)
Close airport	0.120	(0.014)
U.S. center distance	0.218	(0.009)
Per capita income	0.129	(0.011)
Income growth	0.049	(0.005)
Wright amendment	-1.813	(0.130)
Dallas	0.342	(0.070)
Competitive effect	-0.349	(0.007)
Constant	-5.299	(0.019)
First-move propensity		
Airport presence	-2.984	(0.895)
Cost	3.821	(0.622)
Log likelihood		-6063.548
Markets		2742

TABLE 8. Airline Entry Model Parameter Estimates ($R = 400$)

1 if traffic in the market is restricted by the amendment and 0 otherwise. Additionally, we include an indicator variable for markets with Dallas as an endpoint.

Demographic and Geographic Variables We use six additional control variables. The first three are demographic variables that measure the size and economic conditions of the market. These are the average per capita incomes and average income growth rates of the endpoint cities as well as "market size", which is is defined as the geometric mean of the populations of the endpoint cities. The remaining three are geographic properties of the market: "market distance" is defined as the non-stop distance (in thousands of miles) between the endpoints, "close airport" is the minimum of the distances (in hundreds of miles) from each endpoint airport to the nearest alternative airport, and "U.S. center distance" is the sum of the distances (in thousands of miles) from each endpoint to the geographical center of the United States.

6.2. Results

We estimate the payoff function in (12) using the data for all 2,742 markets. The market-specific variables in the vector x include a constant, the market size, distance, and income measures, and the Dallas and Wright amendment indicators. The carrier-specific variables included in the vectors z_i include airport presence and cost. Finally, we also include both airport presence and cost in the logit first-move probabilities for the index order selection mechanism in (9). This yields two additional parameters represented by the vector ρ.

The estimates for the empirical model are reported in Table 8. Standard errors for the maximum simulated likelihood estimates are calculated by inverting the outer product form of the information matrix using standard formulas and numerical gradients with stepsize $h = 10^{-6}$ (see, e.g., Cameron and Trivedi (2005)). We find that carriers are more likely to enter larger markets with higher per capita income and higher rates of income growth. Firms serving a market connecting to Dallas were more profitable than others, on average. Yet, the negative effects of the Wright amendment were over five times as large as the benefit from serving a Dallas market.

The firm-specific measures were among the most important determinants of profits, with airport presence having a large positive effect and cost having a large negative effect. Interestingly, both of these measures have the opposite signs in the first-move propensity and are both significantly different from zero at the 1% level. Firms with more airport presence are more likely to move last; firms with higher costs tend to move earlier. Because of firm-specific heterogeneity, it appears that firms with more presence in a market are able to enjoy last mover advantages. Due to their strong profit advantages, they can credibly commit entry despite entering after other, weaker firms. Therefore, our results indicate that allowing for move-order effects is important in the airline industry.

7. Conclusion

The methods introduced here expand the array of available empirical models of static games available to researchers. We have developed and analyzed an econometric model based on a sequential game of complete information with discrete choices. The model has interesting differences in terms of identification from its simultaneous-move counterpart and allows for additional flexibility in that it can capture the effects of the order of moves separately from payoff effects. We have demonstrated the simulation-based estimator proposed through a series of Monte Carlo experiments and used data on entry in airline markets to illustrate that move-order seems to be an important determinant of market structure in this industry.

A. Alternative Estimators and Methods

A.1. Method of Simulated Moments

The parametric model can also be estimated using the Method of Simulated Moments (MSM) of McFadden (1989). Let θ be a $l \times 1$ vector of parameters. For simplicity, we will refer to outcomes by a number $a \in \{1, \ldots, J^N\}$ instead of using action profiles. For each a,

$$\mathrm{E}[1\{y_m = a\} - \Pr(a \mid x_m; \theta) \mid x_m] = 0.$$

Let $w_a(x_m)$ be a $Q \times 1$ vector of instruments with $Q \geq l = \dim \theta$. Then, by the law of iterated expectations we have

$$\mathrm{E}\left[\left(1\{y_m = a\} - \Pr(a \mid x_m; \theta)\right) w_a(x_m)\right] = 0.$$

Method of Moments estimation uses the sample analog of these moment conditions:

$$\frac{1}{M} \sum_{m=1}^{M} \sum_{a=1}^{J^N - 1} \left[1\{y_m = a\} - \Pr(a \mid x_m; \theta)\right] w_a(x_m).$$

As before, due to the difficulty of calculating $\Pr(a \mid x_m; \theta)$ we can use the simulated probabilities $\widehat{\Pr}(a \mid x_m; \theta)$ calculated as in (11). Then, the simulation analog of the unconditional moment conditions is

$$q(\theta) = \frac{1}{M} \sum_{m=1}^{M} \sum_{a=1}^{J^N - 1} \left[1\{y_m = a\} - \widehat{\Pr}(a \mid x_m; \theta)\right] w_a(x_m),$$

and the MSM estimator is defined by

$$\hat{\theta} = \arg\min_{\theta} q(\theta)^\top W q(\theta),$$

where W is a positive definite $Q \times Q$ weighting matrix corresponding to some metric on $\mathbb{R}^Q$.

A.2. Importance Sampling

For estimation purposes, we are typically interested in the probability of a particular outcome a occurring conditional on the covariates x, denoted by $\Pr(a \mid x)$. This is the case for both MSL and MSM. For a given parametric model, this probability can be expressed in integral form as

$$\Pr(a \mid x; \theta) = \sum_{o \in \mathcal{O}} \int 1\{\alpha(u(x, \varepsilon, \theta), o) = a\} \, dG(\varepsilon \mid x) \mu(o, x; \theta),$$

where $\alpha(u(x, \varepsilon, \theta), o)$ denotes the subgame perfect Nash equilibrium outcome for a given set of covariates, payoffs, and an order of moves. Importantly, α does not depend on θ except through

u. For simplicity, let u denote the matrix of payoffs and let $h(u \mid x, \theta)$ denote the implied density of u. The outcome of the game can be determined by simply knowing o and u, but evaluating this integral requires solving the game for each value of u, or rather, for each ε and θ. We can change the variable of integration from ε to u and write

$$\Pr(a \mid x; \theta) = \sum_{o \in \mathcal{O}} \mu(o, x; \theta) \int 1\{\alpha(u, o) = a\}\, h(u \mid x; \theta)\, du$$

since given x and θ, the distribution of ε induces a distribution of u which can be easily found and the density h can be evaluated. Then, it is straightforward to apply another transformation which allows the integral to be simulated using importance sampling, following Ackerberg (2009). Simply multiplying and dividing by the value of some sampling density $q(\cdot)$ does not change the value of the integral, but facilitates easier simulation of the integral. This yields

$$\Pr(a \mid x; \theta) = \sum_{o \in \mathcal{O}} \mu(o, x; \theta) \int 1\{\alpha(u, o) = a\}\, \frac{h(u \mid x; \theta)}{q(u)} q(u)\, du.$$

Now, the integral can be simulated using values of u that are pre-drawn from $q(\cdot)$. The game has to be solved for each of these values only once, at the beginning of the estimation procedure instead of at each iteration for a new value of θ. Instead, the fraction $h(u \mid x; \theta)/q(u)$ provides an *importance weight* for each draw from $q(u)$.

Now, given a simulated sample $\{u^{(r)}\}_{r=1}^{R}$ with $u^{(r)} \sim q(\cdot)$, the likelihood can be approximated through simulation by the sum

$$\sum_{o \in \mathcal{O}} \mu(o, x; \theta) \frac{1}{R} \sum_{r=1}^{R} 1\left\{\alpha(u^{(r)}, o) = a\right\} \frac{h(u^{(r)} \mid x, \theta)}{q(u^{(r)})}$$

By the law of large numbers,

$$\frac{1}{R} \sum_{r=1}^{R} 1\left\{\alpha(u^{(r)}, o) = a\right\} \frac{h(u^{(r)} \mid x; \theta)}{q(u^{(r)})} \xrightarrow{\text{p}} \int 1\{\alpha(u, o) = a\}\, \frac{h(u \mid x; \theta)}{q(u)} q(u)\, du.$$

It only remains to specify the density q. This density could be chosen by first estimating the model in the incomplete information case (Einav, 2010). The estimate $\hat{\theta}$, the payoff function specification, and the distribution of ε then yield a candidate density q.

References

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

Aguirregabiria, V. (2010). Another look at the identification of dynamic discrete decision processes: An application to retirement behavior. *Journal of Business and Economic Statistics 28*, 201–218.

Bajari, P., V. Chernozhukov, H. Hong, and D. Nekipelov (2007). Semiparametric estimation of dynamic games of incomplete information. Working paper, University of Minnesota.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010a). Computing equilibria in static games of incomplete information using the all-solution homotopy. Working paper, University of Minnesota.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010b). Estimating static models of strategic interactions. *Journal of Business and Economic Statistics 28*, 469–482.

Bajari, P., H. Hong, and S. P. Ryan (2010). Identification and estimation of a discrete game of complete information. *Econometrica 78*, 1529–1568.

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

Blevins, J. R. (2014). Nonparametric identification of dynamic decision processes with discrete and continuous choices. *Quantitative Economics*. Forthcoming.

Borenstein, S. (1989). Hubs and high fares: Dominance and market power in the US airline industry. *The RAND Journal of Economics 20*(3), 344–365.

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

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

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

Einav, L. (2010). Not all rivals look alike: Estimating an equilibrium model of the release timing game. *Economic Inquiry 48*, 369–390.

Fudenberg, D. and J. Tirole (1991). *Game Theory*. Cambridge, MA: MIT Press.

Gal-Or, E. (1985). First mover and second mover advantages. *International Economic Review 26*, 649–653.

Gayle, P. and Z. Luo (forthcoming). Choosing between order-of-entry assumptions in empirical entry models: Evidence from competition between burger king and mcdonaldâĂŹs restaurant outlets. *Journal of Industrial Economics*.

Gouriéroux, C. and A. Monfort (1991). Simulation based inference in models with heterogeneity. *Annales d'Économie et de Statistique 20/21*, 69–107.

Hajivassiliou, V. A. and P. A. Ruud (1994). Classical estimation methods for LDV models using simulation. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 40. Amsterdam: North Holland.

Hurwicz, L. (1950). Generalization of the concept of identification. In T. C. Koopmans (Ed.), *Statistical Inference in Dynamic Models*. New York: John Wiley & Sons.

Koopmans, T. C. (1949). Identification problems in economic model construction. *Econometrica 17*, 125–144.

Kuhn, H. W. (1953). Extensive games and the problem of information. *Annals of Mathematics Studies 28*, 193–216.

Lee, L.-F. (1992). On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. *Econometric Theory 8*, 518–552.

Lerman, S. R. and C. F. Manski (1981). On the use of simulated frequencies to approximate choice probabilities. In C. F. Manski and D. L. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press.

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

Matzkin, R. L. (1992). Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica 60*, 239–270.

Matzkin, R. L. (1993). Nonparametric identification and estimation of polychotomous choice models. *Journal of Econometrics 58*, 137–168.

Matzkin, R. L. (2007). Nonparametric identification. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, Chapter 73. Amsterdam: North Holland.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica 57*, 995–1026.

McFadden, D. L. (1974). Conditional logit analysis of qualitative choice analysis. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. Academic Press.

Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling.* Springer.

Rust, J. (1994). Structural estimation of Markov decision processes. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Amsterdam. North Holland.

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

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*, 307–333.

Zermelo, E. (1913). Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels. In *Proceedings of the Fifth International Congress Mathematics*, Cambridge, pp. 501–504. Cambridge University Press.