Identification and Estimation of a Discrete Game of Complete Information

Patrick Bajari, Han Hong, and Stephen P. Ryan[*]

University of Minnesota and NBER

Stanford University

MIT and NBER

August 7, 2010

Abstract

We discuss the identification and estimation of discrete games of complete information. Following Bresnahan and Reiss (1990, 1991), a discrete game is a generalization of a standard discrete choice model where utility depends on the actions of other players. Using recent algorithms to compute all of the Nash equilibria to a game, we propose simulation-based estimators for static, discrete games. We demonstrate that the model is identified under weak functional form assumptions using exclusion restrictions and an identification at infinity approach. Monte Carlo evidence demonstrates that the estimator can perform well in moderately-sized samples. As an application, we study entry decisions by construction contractors to bid on highway projects in California. We find that an equilibrium is more likely to be observed if it maximizes joint profits, has a higher Nash product, uses mixed strategies, and is not Pareto dominated by another equilibrium.

1 INTRODUCTION

In this paper, we study the identification and estimation of static, discrete games of complete information. These are the canonical normal form games of basic microeconomic theory, with a history dating back to the seminal work of Nash (1951). Econometrically, a discrete game is a generalization of a standard discrete choice model, such as the conditional logit or multinomial probit, that allows an agent's utility to depend on the actions of all other agents. The utilities of all agents are common knowledge, and we assume that observed outcomes are generated by a Nash equilibrium. Discrete game models been applied to diverse topics such as labor force participation (Bjorn and Vuong (1984), Soetevent and Kooreman (2007)), entry (Bresnahan and Reiss (1990, 1991), Berry (1992), and Jia (2008)), product differentiation (Seim (2001), Mazzeo (2002)), technology choice (Ackerberg and Gowrisankaran (2006) and Ryan and Tucker (2009), Manuszak and Cohen (2004)), advertising (Sweeting (2008)), long term care and family bargaining (Stern and Heideman(1999), Stern and Engers (2002)), analyst stock recommendations (Bajari, Hong, Nekipelov and Krainer (2004)) and production with discrete units (Davis (2006)).

A generic feature of normal form games is that, for a given set of payoffs, there are often multiple Nash equilibria. Therefore, the model does not satisfy the standard coherency condition of a one-to-one mapping between the model primitives and outcomes. This is problematic for identification and estimation. The literature has taken three approaches to deal with multiple Nash equilibria. The first approach is to introduce an equilibrium selection mechanism that specifies which equilibrium is picked as part of the econometric model. Examples include random equilibrium selection in Bjorn and Vuong (1984) and the selection of an extremal equilibrium, as in Jia (2008). The second approach is to restrict attention to a particular class of games, such as entry games, and search for an estimator which allows for identification of payoff parameters even if there are multiple equilibria. For example, Bresnahan and Reiss (1990, 1991) and Berry (1992) study models in which the number of firms is unique even though there may be multiple Nash equilibria. They propose estimators in which the number of firms, rather than the entry decisions of individual agents, is treated as the dependent variable. A third method, proposed by Tamer (2002), uses bounds to estimate an entry model. The bounds are derived from the necessary conditions for pure strategy Nash equilibria, which say that the entry decision of one agent must be a best response to the entry decisions of other agents. Bounds estimation has also been used by Ciliberto and Tamer (2009), Pakes, Porter, Ho, and Ishii (2005), and Andrews, Berry, Jia (2005). Berry and Tamer (2006) and Berry and Reiss (2007) survey the econometric analysis of discrete games.

In this paper, we study identification and estimation of discrete complete information games, explicitly allowing for both multiple and mixed strategy equilibria. We propose a simulation-based estimator for these games. The model primitives include player utilities and an equilibrium selection mechanism which determines the probability that a particular equilibrium of the game is played. Using these primitives, we define a Method of Simulated Moments (MSM) estimator. We exploit recent algorithms that compute all of the equilibria for general discrete games (see McKelvey and McLennan (1996)). Finding the entire set of Nash equilibria is computationally expensive in all except the most simple games. For example, in a game of five players with two actions each, we have found that it may take up to 20 minutes of CPU time on a 3.0 GHz single processor workstation to compute all Nash equilibria. Therefore, we construct a smooth simulator for our model using an approach related to work on importance sampling by Ackerberg (2009) and Keane and Wolpin (1997, 2001). As we shall demonstrate in Section 3, this algorithm significantly reduces the computational burden of estimation and can be easily implemented as a parallel process. In a Monte Carlo study, we find it takes less than a day of CPU time to construct estimates and standard errors for our model. We provide Monte Carlo evidence that our estimator works well even with moderately size samples. Finally, we apply our framework to study entry in an asymmetric first-price auction model. We study the strategic decision of contractors to bid on highway repair projects in California, estimating both the payoffs to the entry game and an equilibrium selection mechanism.

Our approach makes several contributions to the literature on estimating static discrete games. First, our approach can be applied to any normal form game of complete information. Several of the previous approaches in the literature restrict attention to specific classes of games, such as entry games or games with strategic complementarities. Also, our estimator allows agents to play mixed strategies. In Section 2, we demonstrate that unless strong restrictions are made on the underlying payoffs or on the support of the error terms, a discrete game model predicts mixed strategy equilibrium with strictly positive probability. Some research argues that mixed strategy equilibria are more likely to occur than pure strategies in some games. For example, in their study of penalty kicks, Chiapporri, Groseclose, and Levitt (2002) find evidence in favor of mixed strategies. Levin and Smith (2001) conduct an experimental study of entry in auctions and find evidence in favor of the mixed strategy entry equilibrium compared to the pure strategy entry equilibrium. In experimental studies, El-Gamal and Grether (1995) and Shachat and Walker (2004) both found that mixed strategy equilibria can be consistent with an unobserved mixture of Bayesian learning by players. While mixed strategy equilibria can be accounted for in a number of current estimation algorithms for the mean utility parameters that work off Nash equilibrium of a game, including

Pakes, Porter, Ho and Ishii (2005), our estimator is more efficient and also accounts for the equilibrium selection mechanism.

Second, we explicitly model and estimate the equilibrium selection mechanism. McKelvey and McLennan (1997, 1997) have established that normal form games generically have large numbers of Nash equilibria that increase at an exponential rate as the number of players and/or actions grow. Estimating the selection mechanism allows the researcher to simulate the model, which is central to performing counterfactuals. This contrasts with the earlier literature on discrete games, which propose estimators that do not specify which equilibrium to select, making it impossible to simulate the model.

Understanding how equilibria are selected in actual plays of a game is also a topic of independent interest. There is a large and influential literature on refinements of the Nash equilibrium solution concept, such as trembling hand perfection or stability. However, there may be a large number of Nash equilibria which satisfy even the strongest refinements. Currently, there is no generally accepted method in economic theory for selecting between alternative equilibria to a normal form game. As a result, in some applications, the usefulness of game theory may be limited because the economist is forced to either make simplifying assumptions which guarantee a unique outcome or propose an *ad hoc* rule for selecting between multiple equilibria. We contribute to the literature by taking an empirical approach to the problem of equilibrium selection. We believe that an empirical approach may be useful given the lack of theory for selecting between alternative equilibria in many applications.

Our third contribution is to propose conditions for the semiparametric identification of both the structural parameters underlying the payoff functions and the parameters of the equilibrium selection mechanism. We propose two separate sets of conditions. The first identification strategy is based on an identification at infinity argument. Here we suppose that the structural utility parameters can be defined as a linear index, and that the covariates have a sufficiently rich support. We demonstrate that it is possible to identify the structural parameters of our model by examining choice behavior for sufficiently large values of the covariates. The second strategy is based on finding an appropriate exclusion restriction. For example, if there are covariates that shift the utility of one player, but can be excluded from the utility of another player, then we demonstrate that both the payoffs and the equilibrium selection mechanism are locally identified. For example, in an entry game, we would search for a covariate that shifts the profitability of one firm but that can be excluded from the profits of all other firms. In oligopoly models, profits typically depend on a firm's costs, actions, and the actions of other firms in the market. The costs of competing firms are typically excluded from profits. Therefore, if a researcher can find firm specific cost shifters, our approach demonstrates that the model is identified. Firm specific cost shifters are

commonly used in empirical work. For example, Jia (2008) and Holmes (2008) demonstrate that distance from firm headquarters or distribution centers is a cost shifter for big box retailers such as Walmart.

Both exclusion restrictions and index restrictions have been previously used to identify econometric models of discrete games. Bresnahan and Reiss (1991) and Tamer (2002) use these restrictions to identify latent utility parameters in two by two games. Bresnahan and Reiss (1991) show without any restrictions, all outcomes are observationally equivalent in games other than two by two games. To the best of our knowledge, we are the first to use these restrictions to identify both payoffs and the equilibrium selection mechanism in general normal form games.

Finally, we consider an application to entry in auctions, where researchers have not formally treated the possibility of multiple equilibria to the auction game (see Bajari and Hortacsu (2003) and Athey, Levin, and Seira (2008)). We construct a data set of bidder entry into procurement auctions for highway paving projects in California. This application fits our modeling assumptions well. First, contractors' entry decisions can reasonably be modeled as a simultaneous move game. Contractors are prohibited by antitrust law from communicating before submitting their bids, which is enforced by the threat of both civil and criminal penalties. Second, the dependent variable in our model is the decision to bid for a single, precisely specified construction project with a fixed duration. In our application, we find that backlog and other dynamic factors are fairly minor in explaining bidding behavior. Thus, we argue that our entry decision can be reasonably modeled as static, isolated instances of the entry game. In other applications, entry decisions will involve competing in a market for an indeterminate period of time, which suggests that allowing for a dynamic model is important.

The focus of our application is the estimation of the equilibrium selection mechanism. We allow the probability that a particular equilibrium is observed to depend upon whether the equilibrium is in pure strategies, maximizes joint profits, has the highest Nash product among pure strategies, and whether it is dominated. To the best of our knowledge, this is the first empirical test of alternative criteria for equilibrium selection in a normal form game.

2   The Model

The model is a simultaneous move game of complete information, commonly referred to as a normal form game. There are $i = 1, \ldots, N$ players, each with a finite set of actions A_i. Define $A = \times_i A_i$ and let $a = (a_1, \ldots, a_N)$ denote a generic element of A. Player i's von Neumann-Morgenstern (vNM) utility is a map $u_i : A \to R$, where R is the real line. Let π_i

denote a mixed strategy over A_i. A Nash equilibrium is a vector $\pi = (\pi_1, \ldots, \pi_N)$ such that each agent's mixed strategy is a best response.

Following Bresnahan and Reiss (1990, 1991), assume that the vNM utility of player i when the action profile of all players is a can be written as:

$$u_i(a, x, \theta_1, \epsilon_i) = f_i(x, a; \theta_1) + \epsilon_i(a). \tag{1}$$

We will sometimes abuse notation and write $u_i(a)$ instead of $u_i(a, x, \theta_1, \epsilon_i)$. In Equation (1), i's vNM utility from action a, $u_i(a)$, is the sum of two terms. The first term is a function $f_i(a, x; \theta_1)$, which depends on a, the vector of actions taken by all of the players, covariates x, and parameters θ_1. The second term is $\epsilon_i(a)$, a random preference shock. The term $\epsilon_i(a)$ reflects information about utility that is common knowledge to the players, but not observed by the econometrician. Note that the preference shocks depend on the entire vector of actions a, not just the actions taken by player i. In our model, the $\epsilon_i(a)$ are assumed to be i.i.d. with a density $g_i(\epsilon_i(a)|\theta_2)$ and joint distribution $g(\epsilon|\theta_2) = \prod_i \prod_{a \in A} g_i(\epsilon_i(a)|\theta_2)$. In much of the literature, a different assumption is used, where stochastic shocks are only a function of player i's own actions. We could easily modify our estimator to allow the $\epsilon_i(a)$ to only depend on the actions of i or to drop the independence assumption, for example by including random effects to account for unobserved heterogeneity. We discuss the stochastic assumption of the error terms in more detail in our section on identification.

Let $u_i = (u_i(a))_{a \in A}$ denote the vector of utilities for player i, and let $u = (u_1, \ldots, u_N)$. Given that there may be more than one equilibrium for a particular u, let $\mathcal{E}(u)$ denote the set of Nash equilibria. We now introduce a mechanism for how a particular equilibrium is selected in the data. We let $\lambda(\pi; \mathcal{E}(u), \beta)$ denote the probability that equilibrium $\pi \in \mathcal{E}(u)$ is selected, where β is a vector of parameters. In order for λ to generate a well-defined distribution it must be the case that, for all u and β:

$$\sum_{\pi \in \mathcal{E}(u)} \lambda(\pi; \mathcal{E}(u), \beta) = 1.$$

Economic theory and the specifics of a particular application may suggest a potential model for λ. For example, the researcher may hypothesize that the probability that an equilibrium is played may depend on whether it is in pure strategies, Pareto dominated or maximizes the sum of the utilities of players in the game. Given $\pi \in \mathcal{E}(u)$, define the vector

$y(\pi, u)$ as follows:

$$y_1(\pi, u) = \begin{cases} 1 & \text{if } \pi \text{ is a pure strategy equilibrium,} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

$$y_2(\pi, u) = \begin{cases} 1 & \text{if } \pi \text{ is Pareto dominated,} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

$$y_3(\pi, u) = \begin{cases} 1 & \text{if } \left(\sum_i \sum_a \pi(a) u_i(a)\right) - \hat{u} = 0, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where $\pi(a) = \prod_i \pi_i(a_i)$ and $\hat{u} = \max_{\pi' \in \mathcal{E}(u)} \{\sum_i \sum_a \pi'(a) u_i(a)\}$.

A parsimonious, parametric model of λ is then:

$$\lambda(\pi; \mathcal{E}(u), \beta) = \frac{\exp(\beta \cdot y(\pi, u))}{\sum_{\pi' \in \mathcal{E}(u)} \exp(\beta \cdot y(\pi', u))}. \tag{5}$$

Note that in the denominator in (5) the sum is taken over all $\pi' \in \mathcal{E}(u)$. If β_1 is greater than zero, this means that a pure strategy equilibrium is more likely to be selected, all else held constant. This assumption is implicitly made in the bounds estimation literature on games which assumes that only pure strategy equilibrium are observed. If $\beta_2 < 0$, then Pareto dominated equilibrium are less likely to be observed. In economic theory, researchers frequently rule out such as equilibrium as a prior implausible. Finally, if $\beta_3 > 0$, then the equilibrium which maximizes joint payoffs is more likely to be observed and equilibrium which results in much lower total utility are less likely. This assumption is sometimes used in the theoretical literature on collusion or the empirical literature on entry.

A unique aspect of including λ in our model is that we explicitly estimate how equilibrium is selected rather than impose an ad hoc rule for selecting an equilibrium. While there is a large literature on refinements to normal form games, even the strongest refinements do not rule out enough equilibria to make most discrete games give a unique prediction. Our framework allows the researcher to treat equilibrium selection as an empirical problem. By estimating β, we can find which equilibrium best matches the outcomes observed in the data.

Combining the elements of the model together, we obtain the following expression for the probability of observing a specific action profile in a play of the game:

$$P(a|x, f, \lambda) = \int \left\{ \sum_{\pi \in \mathcal{E}(u(f,\epsilon), x)} \lambda\left(\pi; \mathcal{E}\left(u\left(f, \epsilon\right), x\right)\right) \left(\prod_{i=1}^{N} \pi(a_i)\right) \right\} g(\epsilon) d\epsilon. \tag{6}$$

Holding a draw of the error terms fixed, the utility of the agents can be written as

$$u(x, \theta, \epsilon) = f_i(a, x; \theta_1) + \epsilon_i(a)$$

This defines the payoffs in the normal form game. We can then compute the equilibrium set $\mathcal{E}(u(f, \epsilon), x)$, which appears in the index of the summation. Next, we sum over all of the $\pi \in \mathcal{E}(u(f, x, \epsilon, \theta_1))$ which are elements of the equilibrium set. For a particular equilibrium, $\prod_{i=1}^{N} \pi(a_i)$ is the probability that a is observed. We weight these terms by $\lambda(\pi; \mathcal{E}(u(f, x, \epsilon, \theta_1)), \beta)$, the probability that a particular equilibrium is observed.

To evaluate Equation (6) it is necessary to compute the set $\mathcal{E}(u)$. McKelvey and McLennan (1996) survey the available algorithms in detail. The free, publicly available software package, Gambit, has routines that can be used to compute the set $\mathcal{E}(u)$ using these methods.[1] Finding all of the equilibria to a game is not a polynomial time computable problem. However, the available algorithms are fairly efficient at computing $\mathcal{E}(u)$ for games of moderate size. Readers interested in the details of the algorithms are referred to McKelvey and McLennan (1996). In the next sections, we shall take the ability to compute $\mathcal{E}(u)$ as given. In reality, the computation burden for finding $\mathcal{E}(u)$ remains large for moderate to large number of players and actions, and increases exponentially with the number of players and actions. For example, Turocy (2008) tabulates average runtimes to compute all equilibria for various combinations of players and actions. The Gambit routine that we use computes both pure strategy and mixed strategy Nash equilibria by solving systems of polynomial equations and inequalities. Inequality solutions to the polynomial system are searched by enumeration. Equality solutions to the polynomial system, which correspond to mixed strategy equilibria, are searched through the use of a homotopy method.

2.1 Discussion

2.1.1 Mixed Strategies

Allowing for mixed strategies in our framework is necessary if the error term has large enough support. As a result, our estimator would be ill-defined if we restricted attention to pure strategy equilibria. Consider the well-known game of matching pennies, illustrated in the figure below:

Matching Pennies

---

[1] Gambit can be downloaded on the web from http://econweb.tamu.edu/gambit/.

	H	T
H	(1,-1)	(-1,1)
T	(-1,1)	(1,-1)

In matching pennies, each player simultaneously chooses heads (H) or tails (T). If the choice of strategies match, then player one receives utility of one and player two receives a utility of negative one. If the strategies differ, the payoffs are reversed. The only equilibrium of this game is in mixed strategies with each player placing probability 1/2 on H and 1/2 on T. Consider games that have payoffs in a neighborhood of matching pennies by perturbing the payoffs as follows:

<div align="center">Perturbed Game</div>

	H	T
H	$(1+\varepsilon_1(H,H),-1+\varepsilon_2(H,H))$	$(-1+\varepsilon_1(H,T),1+\varepsilon_2(H,T))$
T	$(-1+\varepsilon_1(T,H),1+\varepsilon_2(T,H))$	$(1+\varepsilon_1(T,T),-1+\varepsilon_2(T,T))$

For sufficiently small, but still non-zero, values of ϵ it can easily be verified that there is no pure strategy equilibrium to this game. For example, (H,H) cannot be a pure strategy equilibrium since player 2 would have an incentive to deviate and play T. Thus, there is an open set of payoffs for which discrete games only has an equilibrium in mixed strategies. As a result, our model must accomodate the possibility of equilirium in mixed strategies. If we only allowed for pure strategies, the model would have no equilibrium with probability greater than zero and would not be well defined. It is straightforward to show that this result can be generalized to games with more players and more strategies.

Previous research on complete information games generally limits attention to entry games (see Bresnahan and Reiss (1990, 1991), Berry (1992) and Tamer (2002)). These papers carefully restrict payoffs to guarantee the existence of a pure strategy equilibrium. Thus, the estimators proposed in these papers, which restrict attention to pure strategies, do not need to accommodate mixed strategies. However, since we are interested in a more general specification of payoffs, we must allow for mixed strategies.

2.1.2 Equilibrium Selection

A unique aspect of our framework is that we include the equilibrium selection mechanism, λ, in our econometric model. The inclusion of λ is useful for two reasons. First, there are frequently multiple Nash equilibria to a normal form game. Equilibrium selection is an extremely important question in theoretical economics, however there is very little empirical work in this area outside of experiments. Using our modeling framework, we are able to

empirically investigate equilibrium selection, which is important given that economic theory may provide little guidance about which equilibrium to select.

Second, including λ specifies the probability of each equilibrium and therefore allows us to simulate the model. This is necessary for both the construction of our estimator in the next section and for counterfactual analysis. The bounds approach to games has the advantage of remaining agnostic about λ, other than assuming that all equilibrium are in pure strategies. However, in the bounds approach, the researcher only estimates $f_i(a, x; \theta_1)$. This approach typically does not allow the researher to estimate $g(\epsilon; \theta_2)/$ and λ. As a result, the researcher lacks the infomation to compute $P(a|x, f, \lambda, \theta, \beta)$. Therefore the researcher cannot simulate the model, which greatly limits the scope of applications that can be considered.

Consider the pure coordination game below, where player one chooses {T,B}, top or bottom, and player 2 chooses {L,R}, left or right.

Coordination Game

	L	R
T	(1,1)	(0,0)
B	(0,0)	(1,1)

This game has three equilibria (T,L), (B,R), and a mixed strategy equilibrium where each player plays each strategy with probability 1/2. Economic theory provides little guidance as to which equilibrium is most likely in this game. It does not seem possible to use theory to predict whether the (T,L) or (B,R) equilibrium is most plausible, since equilibrium generate the same payoffs and only differ in the names assigned to the strategies. The inability of economic theory to select a unique equilibrium is not specific to this example. Many games generate multiple equilibria that satisfy the best known refinements in the theoretical literature.

Our approach allows for an empirical approach to equilibrium selection. Suppose that the payoff matrix is known and that the economist has access to data on repeated plays of this game. With a sufficiently large number of observations, the economist will be able to precisely estimate the probability of observing the strategy pairs (T,L), (T,R), (B,L), and (B,R). In this example, knowing λ requires the economist to specify the probability with which each of the three equilibria is played. Since the economist has knowledge of four probabilities, three of which are linearly independent, it follows that λ is identified. Therefore, while economic theory cannot be used to determine equilibrium selection, our simple example suggests that an empirical approach to this problem may be more fruitful.

In our identification section, we investigate conditions under which both the equilibrium selection mechanism and the payoff matrix can be simultaneously identified. We demonstrate, similar to Bresnahan and Reiss (1991) that in general our problem is underidentified.

10

However, we also describe two sets of sufficient conditions for identification that may be useful in some applications.

2.1.3 Comparison with Incomplete Information Games

An alternative approach used in the applied literature is to assume that the error terms only depend on player i's own actions and are private information. Incomplete information games are attractive for empirical work since it is often possible to estimate these models using a simple two-step procedure.[2] However, discrete games with incomplete information have a very different equilibrium structure than games with complete information. In a coordination game, Bajari, Hong, Krainer, and Nekipelov (2006) use numerical methods to show that the number of equilibria decreases as the number of players in the game increases. In a complete information game, by comparison, the average number of Nash equilibria will increase as players are added to the game (see McKelvey and McLennan (1996)). Thus, the assumption of incomplete information refines the equilibrium set. The properties of this refinement are not completely understood and is an active area of research, see Brock and Durlauf (2001), McKelvey and Palfrey (1995), and Sannikov (2007). We believe that evaluating the merits of both games with complete and incomplete information is an important topic for future research.[3]

2.2 Examples of Discrete Games

The model that we propose is quite general and could be applied to many discrete games considered in the literature. We discuss three examples: entry, technology adoption with network effects, and peer effects. The first example is static entry into a market (see Bresnahan and Reiss (1990, 1991), Berry (1992), Tamer (2002), Ciliberto and Tamer (2009), and Manuszak and Cohen (2004)). In applications of entry games, the economist observes a cross section of markets and the players correspond to a set of potential entrants. The potential entrants simultaneously choose whether to enter: $a_i = 1$ denotes a decision by i to enter the market and $a_i = 0$ not to enter the market. In empirical work, the profit function, f_i, typically takes the form:

$$f_i = \begin{cases} \theta_1 \cdot x + \delta \sum_{j \neq i} 1\{a_j = 1\} & \text{if } a_i = 1, \\ 0 & \text{if } a_i = 0. \end{cases} \quad (7)$$

---

[2]See Pesendorfer and Schmidt-Dengler (2003), Aguirregabiria and Mira (2007), Brock and Durlauf (2001), Sweeting (2008), Bajari, Benkard and Levin (2007) and Bajari, Hong, Krainer and Nekipelov (2006).

[3]In addition, it may be easier to compute the set of all equilibria in games of complete information. See Bajari, Hong, Nekipelov and Krainer (2004) for a discussion.

In Equation (7) the mean utility from not entering is set equal to zero.[4] The covariates x are variables which influence the profitability of entering a market, such as the number of consumers in the market, average income, and market-specific cost variables. The parameter δ measures the influence entry by other firms on firm i's profits. If profits decrease from having another firm enter the market, then $\delta < 0$. The error $\epsilon_i(a)$ captures shocks to the profitability of entry that are commonly observed by all firms in the market, but are unobserved to the econometrician. In applied work, it might be desirable to include a market-specific random effect in $\epsilon_i(a)$ in order to account for common shocks to profitability.

A second example is technology adoption in the presence of network effects, as in Ackerberg and Gowrisankaran (2006) who model the decision by banks in spatially separated markets to adopt the Automated Clearing House (ACH) payment system. The players in the game are the existing banks in some market. Let $a_i = 1$ denote a decision to adopt ACH and $a_i = 0$ denote non-adoption. A priori, network effects are likely since the customers of bank i are able to transfer funds to customers of bank j if both banks adopt ACH. An empirical model of network effects could take the form:

$$
f_i = \begin{cases} \theta_1 \cdot x_i + \delta \sum_{j \neq i} 1\{a_j = 1\} \cdot c_j \cdot c_i & \text{if } a_i = 1, \\ 0 & \text{if } a_i = 0. \end{cases}
\tag{8}
$$

In Equation (8), x_i denotes some factors which influence the costs and benefits to adoption by firm i, such as the number of customers of bank i and their characteristics (e.g. large corporate or government agencies commonly use ACH to make automatic payroll deposits). The term c_i is the current number of customers of bank i. The term $\delta \sum_{j \neq i} 1\{a_j = 1\} \cdot c_j \cdot c_i$ captures the network effect. The marginal benefit of i's adoption depends on $c_i \cdot c_j$.

A third example is peer effects, as in Manski (1993) and Brock and Durlauf (2001, 2006). A peer effect connotes a situation where there is a benefit from conforming to the average or norm behavior. For example, consider the decision by a high school senior to take calculus. The players in the game are all of the students who could potentially take the class. Let $a_i = 1$ if student i decides to take calculus and $a_i = 0$ otherwise. The utility of student i is:

$$
f_i = \begin{cases} \theta_1 \cdot x_i + \delta \sum_{j \neq i} 1\{a_j = 1\} \cdot s_j & \text{if } a_i = 1, \\ 0 & \text{if } a_i = 0. \end{cases}
\tag{9}
$$

In Equation (9), the covariates x_i could include terms that shift a student's incentives to take calculus, such as the educational status of her parents. The term s_i denotes the score

---

[4]We formally discuss this normalization in our section on identification.

of student i on a standardized achievement test and is commonly used to proxy for ability. If $\delta > 0$, the term $\delta \sum_{j \neq i} 1\{a_j = 1\} \cdot s_j$ captures a positive peer effect, i.e. the utility to student i from taking calculus in an increasing function of the number of other students who take calculus, interacted with the test scores of student i's peers.

The modeling framework we propose could be applied beyond these three examples. In principal, the framework above could be used to model any discrete choice where 1.) the payoffs of agents are interdependent, 2.) decisions are made simultaneously, and 3.) there is complete information. If the number of players or actions is very large, our estimator may not be computationally feasible due to the computational cost of solving for the entire equilibrium set. However, in the next section we describe an estimator which reduces the computational burden of estimation through the use of a parallel algorithm.

3   SIMULATION

Next, we propose a computationally efficient Method of Simulated Moments (MSM) estimator for θ and β, the parameters governing agents' payoffs and the equilibrium selection mechanism, respectively. As in section 2, let $P(a|x, \theta, \beta)$ denote the probability that a vector of strategies, $a = (a_1, \ldots, a_N)$, is observed conditional on x, θ, and β. MSM estimation requires an accurate and computationally efficient method for simulating $P(a|x, \theta, \beta)$, which can be written as:

$$P(a|x, \theta, \beta) = \int \left\{ \sum_{\pi \in \mathcal{E}(u(x, \theta, \epsilon))} \lambda\left(\pi; \mathcal{E}(u(x, \theta_1, \epsilon)), \beta\right) \left(\prod_{i=1}^{N} \pi_i(a_i)\right) \right\} g(\epsilon|\theta_2) d\epsilon. \qquad (10)$$

In principal, this integral could be simulated using a straightforward Monte Carlo procedure. First, pseudo random values of the random preference shocks $\epsilon = (\epsilon_1, \ldots, \epsilon_N)$ are drawn from the distribution $g(\epsilon|\theta_2)$. Second, for each pseudo random error draw $\epsilon = (\epsilon_1, \ldots, \epsilon_N)$, utilities are computed using Equation (1); we denote the utilities as $u(x, \theta, \epsilon)$ to emphasize their dependence on the parameters, covariates and preference shocks. Third, the equilibrium set $\mathcal{E}(u(x, \theta, \epsilon))$ is computed. And finally, the probability that an event is observed is computed by summing over the equilibria $\pi \in \mathcal{E}$ and computing 1.) $\lambda(\pi)$, the probability that the equilibrium π is selected, and 2.) $\prod_{i=1}^{N} \pi_i(a_i)$, the probability that a is observed given π. By averaging over a large number of draws of ϵ, the economist could precisely simulate $P(a|x, \theta, \beta)$.

Unfortunately, this straightforward approach is not practical for applied work in all but the simplest games. The reason is that druing estimation, $P(a|x, \theta, \beta)$ must be simulated for each candidate parameter value θ, β and each vector x that appears in the data. The

equilibrium set $\mathcal{E}(u(x,\theta,\epsilon))$ therefore must be computed a large number of times. We have found that it may take up to 20 minutes to compute $\mathcal{E}(u(x,\theta,\epsilon))$ for a 5 player game with two strategies. As a result, the computational costs of this straightforward approach will often be prohibitive in applied work.

In order to lower the computational burden of simulating $P(a|x,\theta,\beta)$, we borrow from Keane and Wolpin (1997), Keane and Wolpin (2001) and Ackerberg (2009). First, we change the variable of integration in Equation (10) from ϵ to u. Let $h(u|\theta,x)$ denote the density u, conditional on θ and x. In many models, this density is trivial to compute and simulate. For instance, suppose that the preference shocks $\epsilon_i(a)$ are i.i.d. standard normal with density $\phi(\cdot)$. Then, the density $h(u|\theta,x)$ is:

$$h(u|\theta,x) = \prod_i \prod_{a \in A} \phi(u_i(a) - f_i(x,a;\theta_1)|0,\sigma)$$

which can be computed easily using standard programming packages.

If we change the variable of integration from ϵ to u, then Equation (10) becomes:

$$P(a|x,\theta,\beta) = \int \left\{ \sum_{\pi \in \mathcal{E}(u)} \lambda\left(\pi; \mathcal{E}(u), \beta\right) \left(\prod_{i=1}^{N} \pi_i(a_i)\right) \right\} h(u|\theta,x)du. \tag{11}$$

Our simulator uses importance sampling; therefore, we rewrite Equation (11) as:

$$P(a|x,\theta,\beta) = \int \left\{ \sum_{\pi \in \mathcal{E}(u)} \lambda\left(\pi; \mathcal{E}(u), \beta\right) \left(\prod_{i=1}^{N} \pi_i(a_i)\right) \right\} \frac{h(u|\theta,x)}{q(u|x)} q(u|x)du,$$

where $q(u|x)$ is the importance density. For a given value of x, we draw a pseudo random sequence $u^{(s)} = (u_1^{(s)}, \ldots, u_N^{(s)})$, $s = 1, \ldots, S$ of random utilities from the importance density $q(u|x)$. Each $u_i^{(s)}$ is a vector of simulated utility indexes for all the possible action profiles for player i. We then compute the equilibrium sets $\mathcal{E}(u^{(s)})$, a step which can be performed in parallel across several CPU's.

We can then simulate $P(a|x,\theta,\beta)$ as follows:

$$\widehat{P}(a|x,\theta,\beta) = \frac{1}{S} \sum_{s=1}^{S} \left\{ \sum_{\pi \in \mathcal{E}(u^{(s)})} \lambda(\pi; \mathcal{E}(u^{(s)}), \beta) \left(\prod_{i=1}^{N} \pi_i(a_i)\right) \right\} \frac{h(u^{(s)}|\theta,x)}{q(u^{(s)}|x)} \tag{12}$$

This simulator has three practical advantages for applied work. First, $\widehat{P}(a|x,\theta,\beta)$ will be an unbiased estimator of $P(a|x,\theta,\beta)$. Second, given the simulation draws $u^{(s)}$, the parameters

θ and β do not enter into the expression for the equilibrium set $\mathcal{E}(u^{(s)})$. Therefore, when we change the parameter values, it is not necessary to recompute the equilibrium set $\mathcal{E}(u^{(s)})$. Exploiting this property will drastically reduce the time required to compute our estimator. Third, this simulator is a smooth function of the underlying parameters. As a result, the minimization of our MSM objective function will be numerically well behaved.[5]

The theory of importance sampling proves that $\widehat{P}(a|x,\theta,\beta)$ is a smooth and unbiased simulator for any choice of the importance density $q(u^{(s)}|x)$ that has sufficiently large support. However, as a practical matter, it is important that the ratio $\frac{h(u^{(s)}|\theta,x)}{q(u^{(s)}|x)}$ does not become too large. In order to ensure this, we need to make sure that the tails of the importance density $q(u^{(s)}|x)$ are not too thin in a neighborhood of the parameter that minimizes our MSM estimator. In our applied work, we have often constructed the importance density $q(u|x)$ by first estimating a version of the model in which the error terms $\epsilon = (\epsilon_1, \ldots, \epsilon_N)$ are private information instead of common knowledge. We then use the method proposed in Bajari, Hong, Nekipelov and Krainer (2004) to estimate the parameters of the private information version of the model. This is an extremely simple estimation problem and can be quickly programmed using a standard statistical package such as STATA. The importance density $q(u|x)$ is then set equal to the distribution of utilities conditional on x in the private information version of the game.[6]

3.1   The Estimator

The economist observes a sequence (a_t, x_t) of actions and covariates, $t = 1, \ldots, T$. Equation (12) can be used to form a maximum simulated likelihood estimator (MSL) for these observations. As is well known, MSL is biased for any fixed number of simulations. The number of simulation has to increase to infinity as the sample size increases in order for MSL to be consistent. In order to obtain $\sqrt{T}$ consistent estimates with an asymptotic distribution centered at zero, one needs to increase the number of draws S so that $\frac{S}{\sqrt{T}} \to \infty$. If $\frac{S}{\sqrt{T}} \to c$ for a constant c that is bounded away from 0, the MLS estimator is still $\sqrt{T}$ consistent but

---

[5] We note that while we follow Keane and Wolpin (1997), (2001) and Ackerberg (2009) in constructing the importance sampler, its use in normal form game estimation is new. In addition, there is also a subtle difference between our use of the importance sampling and its use by previous authors. The dynamic discrete choice model in these earlier papers requires a complete random coefficient specification to allow the importance sampler to reduce the computation burden. The complete information normal form game has the interesting feature that it does not require a random coefficient specification for the importance sampler to save on the computation burden of the estimator.

[6]This estimator can be performed in two stages. In the first stage, the economist flexibly estimates the choice probabilities $P(a|x)$ using standard methods. In the second stage, the economist assumes that these estimated choice probabilities represent the agent's equilibrium expectations. These choice probabilities are then substituted into the utility function.

the asymptotic distribution has a bias that is different from zero.[7]

Alternatively, one can estimate the parameters using MSM. An advantage of MSM is that it generates an unbiased and consistent estimator for a fixed value of S. To form the MSM estimator, enumerate the elements of A from $k = \{1, \ldots, \#A\}$. Note that, because the probabilities of all of the elements of $a \in A$ must sum to one, one of these probabilities will be linearly dependent on the others, so there are effectively $\#A - 1$ conditional moments. Let $w_k(x)$ be a vector of weight functions, with dimension larger than the number of parameters, for each k and let $1(a_t = k)$ denote the indicator function that the tth vector of actions is equal to k. The function $P(k|x, \theta, \beta)$ denotes the probability that the observed vector of actions is k given x and the parameters θ and β. This probability is defined in Equation (10). At the true parameters of the data-generating process the predicted probability of each action equals its empirical probability for each action k:

$$E\left[1(a_t = k) - P(k|x, \theta, \beta)\right] w_k(x) = 0.$$

Using the sample counterpart of the above expectation, we form a vector of $\#A - 1$ moments, where the k-th element is defined by:

$$m_{k,T}(\theta, \beta) = \frac{1}{T} \sum_{t=1}^{T} \left[1(a_t = k) - P(k|x_t, \theta, \beta)\right] w_k(x_t).$$

In practice, $P(k|x_t, \theta, \beta)$ is evaluated by simulation using the importance sampler in Equation (12). For each x_t, we draw a vector of S simulations $u_t^{(s)}$, where $s = \{1, \ldots, S\}$, from the importance density $q(u|x)$. We assume that the simulation draws $u_t^{(s)}$ are independent over both t and s, and are independent of the data. The k-th moment condition is then replaced by its simulation analog:

$$\hat{m}_{k,T}(\theta, \beta) = \frac{1}{T} \sum_{t=1}^{T} \left[1(a_t = k) - \widehat{P}(k|x_t, \theta, \beta)\right] w_k(x_t).$$

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

$$\left(\widehat{\theta}, \widehat{\beta}\right) = \arg\min_{(\theta, \beta)} \hat{m}_T(\theta, \beta)' \times W_T \times \hat{m}_T(\theta, \beta). \tag{13}$$

The asymptotic theory for estimating discrete choice models using MSL and MSM is well developed. See McFadden (1989), Pakes and Pollard (1989), or (Hajivassilou and Ruud

---

[7]In practice we have found that MSL can be useful for finding starting values for MSM. In our experience, the likelihood function is more concave around the maximum than in the MSM estimator.

1994) for detailed discussions.

4 IDENTIFICATION

Next, we develop two approaches to identification. To be clear, we are interested in conditions under which it is possible to recover the unknown primitives of the model in Section 2: $f(a, x)$ and $\lambda(x)$. We shall consider the identification of $f(a, x)$ and $\lambda(x)$ as general functions of x. Therefore, we shall drop their dependence on θ and β. In the first approach, we provide sufficient conditions to identify payoffs and the selection mechanism as the support of the covariates grows large. The second approach considers identification based on agent-specific payoff shifters. We first discuss two necessary restrictions on the data-generating process that are familiar from the discrete choice literature. We also present some negative results on identification before discussing the details of our approaches.

The results in section 4.3 show that, in general, the maximum number of equilibria can be much larger than the number of moment conditions $\#A - 1$, which means that the equilibrium selection mechanism cannot be identified for a certain range of the distribution of the random utility $u_i(x, \epsilon)$. This section also justifies more restrictive forms, parametric or semiparametric, of the equilibrium selection mechanism. Appendix B.4 also provides more details.

4.1 Scale and Location Normalizations

Assumption 1. *The payoffs of one action for each agent are fixed at a known constant.*

This restriction is similar to normalizing the mean utility from the outside good to a constant, usually zero, in a standard discrete choice model. It is clear from the definition of a Nash equilibrium that adding a constant to all deterministic payoffs does not perturb the set of equilibria, so a location normalization is necessary. A scale normalization is also necessary, as multiplying all deterministic payoffs by a positive constant does not alter the set of Nash equilibria either. This restriction is subsumed in the following assumption about the distribution of the error terms.

Assumption 2. *The joint distribution of $\epsilon = (\epsilon_i(a))$ is independent and known to all agents and the econometrician.*

Assumption 2 allows $\epsilon_i(a)$ to be any known joint parametric distribution. For expositional clarity, we shall assume that it has a standard normal distribution. Even in the simplest discrete choice models, it is not possible to identify both $f_i(a, x)$ and the joint distribution of

the $\epsilon_i(a)$ nonparametrically. Consider a standard binary choice model where the dependent variable is 1 if the index $u(x) + \epsilon$ is greater than zero, i.e.

$$y = 1(u(x) + \epsilon > 0) \tag{14}$$

All the population information about this model is contained in the conditional probability $P(y = 1|x)$, the probability that the dependent variable is equal to one given the covariates x. If the CDF of ϵ is G, then Equation (14) implies that:

$$P(y = 1|x) = G(u(x)), \tag{15}$$

Obviously, only the composition of $G(u(x))$ can be identified. It is therefore necessary to make parametric assumptions on one part (e.g. G or u) in order to identify the other part. For instance, if G is the standard normal CDF, we can perfectly rationalize the observed moments in Equation (15) by setting $u(x)$ to the inverse CDF evaluated at $P(y = 1|x)$. Therefore, we will assume that the error terms are independently and normally distributed.

4.2 Difficulty of Nonparametric Identification

A model is said to be identified if the model primitives can be recovered given the probability distributions the economist can observe. In a normal form game, the available population probabilities are $P(a|x)$ for $a \in A$. Again, the primitives we wish to identify are $f(a, x)$ and $\lambda(x)$.

We can generalize Equation (10) by writing $P(a|x)$ in a way that does not hinge on the specific parametric forms implicitly assumed in Section 2.

$$P(a|x, f, \lambda) = \int \left\{ \sum_{\pi \in \mathcal{E}(u(f,\epsilon),x)} \lambda\left(\pi; \mathcal{E}\left(u\left(f, \epsilon\right), x\right)\right) \left(\prod_{i=1}^{N} \pi(a_i)\right) \right\} g(\epsilon) d\epsilon \tag{16}$$

In Equation (16), we write the vNM utilities as $u(f, \epsilon)$ to remind ourselves that they are a sum of the mean utilities $f(a, x)$ and the shocks ϵ. Holding x fixed we can view Equation (16) as a finite number of equations that depend on the finite number of parameters, $f(a, x)$ and $\lambda(x)$, where $\lambda(x)$ implicily defines the vector of equilibrium selection probabilities for all π in $\mathcal{E}$. Denote this system as $P(a|x) = H(f(a, x), \lambda(x))$ where H is the map implicitly defined by Equation (16) across all the action profiles a. We drop one choice probability for each player when writing H; not doing so introduces a linear dependence between the rows of this system, since choice probabilities sum to one. In what follows, we shall invoke the following assumption:

Assumption 3. *The map H is continuously differentiable. The Jacobian formed by differentiating H with respect to the parameter vectors $f(a,x)$ and $\lambda(x)$ is denoted by $DH_{f,\lambda}(x)$.*

Given the probabilities $P(a|x)$, suppose that $f^0(a,x)$ and $\lambda^0(x)$ satisfy Equation (16). If no other pairs of $f(a,x)$ and $\lambda(x)$ also satisfy (16), $\left(f^0(a,x),\lambda^0(x)\right)$ is said to be globally identified. On the other hand, $\left(f^0(a,x),\lambda^0(x)\right)$ are said to be locally identified if there exists an open neighborhood N_x of $\left(f^0(a,x),\lambda^0(x)\right)$ such that there is no other vector $\left(\widetilde{f}(a,x),\widetilde{\lambda}(x)\right) \in N_x$, $\left(\widetilde{f}(a,x),\widetilde{\lambda}(x)\right) \neq \left(f^0(a,x),\lambda^0(x)\right)$, that also satisfies Equation (16).

$P(a|x)$ is analogous to the reduced form parameters and $(f(a,x),\lambda(x))$ to the structural parameters in the terminologies of Rothenberg (1971) and Gale and Nikaido (1965). The vector $(f(a,x),\lambda(x))$ is called a regular point of $DH_{f,\lambda}(x)$ if the rank of $DH_{f,\lambda}(x)$ is constant in a neighborhood of $(f(a,x),\lambda(x))$. However, H is usually highly nonlinear. Theorem 6 in Rothenberg (1971) states that a sufficient condition for $\left(f^0(a,x),\lambda^0(x)\right)$ to be locally identified is that the rank of $DH_{f,\lambda}(x)$ is equal to the total number of parameters in $(f(a,x),\lambda(x))$ at $\left(f^0(a,x),\lambda^0(x)\right)$. Sufficient conditions for global identification are more difficult. Gale and Nikaido (1965) require the existence of a square submatrix W of $DH_{f,\lambda}(x)$ with dimension of $(f(a,x),\lambda(x))$, such that W has a positive determinant and that $W+W'$ is positive definitive throughout the parameter space. While we can possibly check the rank identification condition of $DH_{f,\lambda}(x)$ for low dimension two by two games, it is difficult to do so for general games. At a minimum, this would require us to characterize the set of all equilibria that can be reached. This can be difficult in games with multiple players and strategies.

A necessary condition for the full rank of $DH_{f,\lambda}(x)$ is the order condition, which requires that the number of $P(a|x)$ in forming $DH_{f,\lambda}(x)$ is larger than the number of parameters in $(f(a,x),\lambda(x))$. The order condition is easier to investigate in general. To fix ideas in what follows, consider the simple game in table 1. Note that for each player, we have normalized the payoff of one action to zero.

Table 1: Example of two by two game

	L	R
T	$(0,0)$	$(0, f_2(TR,x) + \epsilon_2(TR))$
B	$(f_1(BL,x) + \epsilon_1(BL),0)$	$(f_1(BR,x) + \epsilon_1(BR), f_2(BR,x) + \epsilon_2(BR))$

We first note that, even if the selection mechanism λ is known, a two by two game has more utility parameters that need to be identified than the number of moment conditions that can be observed in the data. Holding a given realization of x fixed, the econometrician observes four conditional moments: $P(TL|x), P(TR|x), P(BL|x), P(BR|x)$. However, because the probability of the actions must sum to one, there are effectively three moments that the econometrician can use. Meanwhile, we have four utility parameters,

$f_1(BL, x)$, $f_1(BR, x)$, $f_2(TR, x)$, $f_2(BR, x)$ that need to be identified. Since there are more free parameters than moments, the model is unidentified. Note that variation in x does not help to reduce the total number of parameters that need to be identified because we place no restrictions on how mean utility varies with x.

The result above can easily be generalized to generic games. Consider a game with N players and $\#A_i$ strategies for player i. Holding x fixed, the total number of mean utility parameters $f_i(a, x)$ is equal to $N \cdot \prod_i \#A_i - \sum_i \prod_{j \neq i} \#A_j$. This is the cardinality of the number of strategies, times the number of players, minus the required normalizations. The number of moments that the economist can observe, conditional on x, is only $\prod_i \#A_i - 1$. If each player has at least two strategies, then for each given x the difference between the number of utility parameters, f_i, to estimate and the number of available moment conditions is bounded from below by

$$\left((N-1) - \frac{N}{2} \right) \prod_i \#A_i + 1 \geq 0.$$

Therefore, the model is underidentified.

4.3 Identification at Infinity

The first approach we propose is based on a strategy of identification at infinity. Suppose the covariates have full support and the mean utilities are defined by a linear index of the covariates with suitable sign restrictions on the coefficients. That is, $f_i(a, x) = x_i^a \beta_i^a$ for all a_i for which $f_i(a, x)$ is not normalized to zero. Identification at infinify strategies are often needed for linear index models of discrete outcomes, e.g. Manski (1988), Heckman (1990), and Tamer (2002). We demonstrate this approach can also be applied to discrete games.

The identification strategy involves two steps. In the first step, using arguments similar to Tamer (2002), we identify the mean utilities by focusing on a path of the covariate values that gives a unique equilibrium with probability close to 1. We then perturb the covariates locally to identify the utility parameters, all of which enter f_i linearly. In the second step, under an invariance assumption on the equilibrium selection mechanism, we identify the equilibrium selection probabilities from the observable choice probabilities.

As a simple example, consider the two by two game in Table 1 in subsection 4.2, where $\epsilon_1(TL) = \epsilon_1(TR) = \epsilon_2(TL) = \epsilon_2(BL) = 0$. Assume that $X = \{x_1, x_2\}$ has nondegenerate full support on $\mathbb{R}^2$. Since payoffs are defined as a linear index, it is possible to find a sufficiently large value of x_2 such that player two will play L with probability approaching one. For x_2 sufficiently large, the probability that player one chooses B is: $P(B|x) = P(f_1(BL, x_1) + \epsilon_1(BL) > 0)$. Note that this is a single-agent decision problem: x_2 is such

that player two is going to play L regardless of player one's decision. Therefore, player one will choose B if and only if the threshold condition above is satisfied. As long as $\epsilon_1(BL)$ is drawn from a known distribution with a strictly increasing cumulative distribution function, e.g. a standard normal distribution $\Phi(\cdot)$, the value of $f_1(BL, x_1)$ can be recovered by inverting the empirical analog of $P(B|x)$:

$$P(B|x) = P(\epsilon_1(BL) > -f_1(BL, x_1)) = \Phi(f_1(BL, x_1)),$$

The uniqueness of $f_1(BL, x_1)$ is guaranteed by the monotonicity of $\Phi(\cdot)$.

An analogous argument can be made to identify all of the unknown payoff parameters by a suitable choice of either x_1 or x_2. Once we have recovered all the payoff parameters, the only unknowns are those governing the equilibrium selection mechanism. We restrict attention to a region where the influence of the ϵ_i on payoffs is small. In this region, the utility function is known by the argument in the previous paragraph. The unknown part of the model is λ. Since there are fewer equilibrium than non-colinear moments, $P(a|x)$, the parameters of λ are identified. Our invariance assumption guarantees that identification in this region insures identification globally. The rest of this section we present formal results for general N-player games. The appendix contains the technical proofs and gives more details of how the results apply to two by two games.

Assumption 4. *For any $i = 1, \ldots, N$ and action profile $k_{-i} \in A_{-i}$, there exists a set $\mathcal{T}_{-i}^{k_{-i}}$ of covariates x such that $\lim_{||x|| \to \infty, x \in \mathcal{T}_{-i}^{k_{-i}}} P\left(a_{-i} = k_{-i}|x\right) = 1$.*

This assumption requires that for each player i, the covariates x can be shifted along a dimension such that each element in k_{-i} is a dominant strategy for each player in $-i$. This assumption allows for identifying $f_i\left(a_i, a_{-i}, x\right) = x_i^a \beta_i^a$ as a single agent discrete choice problem holding a_{-i} fixed at these values of the covariates x.

The next assumption, requires that utilities recovered from this path can be extended to the entire range of covariates using the linearity assumption on the deterministic payoff functions,

Assumption 5. *For all i and all $a \in A$ such that the mean utility $f_i\left(a, x\right)$ is not normalized, there exists some $L_0 > 0$ such that*

$$\inf_{L \geq L_0} \min \mathrm{eig} E\left[x_i^a x_i^a | x \in \mathcal{T}_{-i}^{a_{-i}}, ||x|| \geq L\right] > 0.$$

Theorem 1. *Under assumptions 1, 2, 4 and 5, β_i^a and hence $f_i(a, x)$ is identified up to the normalization in Assumption 1, for all i and all a for each x.*

21

The next assumption is an invariance property that is required to identify the equilibrium selection probabilities.

Assumption 6. *The equilibrium selection probabilities depend only on the latent utility indices:* $\rho(x, \epsilon) = \rho(u(a, x, \epsilon))$, *and are scale invariant with respect to the latent utility indexes, i.e. for all* $\alpha > 0$, $\rho(\alpha u(a, x, \epsilon)) = \rho(u(a, x, \epsilon))$.

The last assumption requires that the total utility indexes can be approximated arbitrarily well by the observable components.

Assumption 7. *There exists a set* $\mathcal{T}$ *such that for all* $\delta > 0$:

$$\lim_{|x| \to \infty, x \in \mathcal{T}} \min_{i,a} P\left[\left|\frac{f_i(a, x)}{u_i(a, x, \epsilon)} - 1\right| < \delta\right] = 1. \tag{17}$$

Theorem 2. *In addition to the conditions in theorem 1, under Assumptions 6 to 7, the equilibrium selection probabilities* $\rho(u(x, \epsilon))$ *are all identified from the observed choice probabilities whenever the cardinality of* $\mathcal{E}(u(x, \epsilon))$ *is less than or equal to* $\#A - 1$.

Remark: Note that the conditions in this theorem depend on the numbers of players and strategies, and generally also on the particular realization of $u(x, \epsilon)$. When the maximum number of equilibria for a game is less than or equal to $\#A - 1$, the condition in the above theorem holds uniformly for all realizations of $u(x, \epsilon)$. Such is the case, for example, for two by two games and for games with two players each equipped with four strategies.

4.4   Exclusion Restrictions

The results of the previous section are not surprising in light of Bresnahan and Reiss (1991) and Pesendorfer and Schmidt-Dengler (2003), who demonstrate failures in identification of discrete games. As we noted in the introduction, the structure of our models is not unlike treatment effect and sample selection models, with latent utilities f seem analogous to the treatment equation and λ to the selection equation. It is well known that these simpler models cannot be identified without exclusion restrictions. That is, we must search for variables that influence one equation, but not the other. In what follows, we demonstrate that a similar approach is possible in games.

The exclusion restrictions that we consider are covariates that shift the utility of agent i but do not enter as arguments into $u_j (j \neq i)$ or the equilibrium selection mechanism λ. In many applications, such covariates are not difficult to find.

Assumption 8. *For each agent i, there exists some covariate, x_i that enters the utility of agent i, but not the utility of other agents. That is, i's utility can be written as $f_i(a, \bar{x}, x_i)$. Furthermore, in addition to assumption 6, $\rho(u(\alpha, x, \epsilon))$ depends on $u(\alpha, x, \epsilon)$ only through a set of sufficient statistics of dimension $M \times (N-1)$ where M is a constant that does not depend on the number of players N.*

The first part of Assumption 8 implies that there are agent i specific utility shifters. While this Assumption is unlikely to be perfectly satisfied, to a first approximation it does seem reasonable in many applications. The second part of Assumption 8 is a weak assumption that will be satisfied, for example, if the equilibrium selection probabilities depend only on the total utilities of all players in each equilibrium. This assumption does impose some restrictions on λ. The function λ cannot depend freely on the utility indexes. Also, it cannot grow with the number of covariates, otherwise variation in the covariates will increase the number of parameters required to characterize the equilibrium selection mechanism.

Theorem 3. *Suppose that Assumptions 1, 2, 11, 6 and 8 hold. If $\#x_i$ are sufficiently large, the necessary order condition is satisfied.*

The intuition behind this theorem is quite simple. When K, the number of support points of x_i, increases, the number of conditional choice probabilities increases at the rate of K^N, which is larger than K^{N-1}, the order of the number of utility and equilibrium selection probability parameters. Our results demonstrate that identification is possible if we have covariates that are indexed by the agent's identity i. Such exclusion results are imposed in most existing applications of discrete games. For example, consider empirical studies of strategic entry. In the case of an airline deciding whether to serve a particular city-pair, one such shifter could be the number of connecting routes that airline has at both endpoints, or whether one or both of the cities is a hub for that airline. These covariates are typically excluded from the payoffs of an airline's competitors. Holmes (2008) and Jia (2008) study entry decisions by large retailers such as Walmart and Kmart. Their analysis suggests that a payoff shifter is the distance from the closest regional distribution centers or company headquarters.

As a second example, consider technology adoption in the presence of network effects, as in Ryan and Tucker (2009). Here employees within a firm decide whether to adopt a video-conferencing technology on their personal computers. The benefit to any given employee of adopting depends on the adoption decisions of other employees. Furthermore, the benefit of using this technology varies with an employee's rank in the firm, their geographic locale, and their job function. All of these characteristics shift the benefits of adoption on an individual

basis. For example, senior managing directors in equities are likely to have different payoffs from using the network than a junior administrator in human resources.

Finally, we note that the maintained exogeneity assumptions used in our identification results are quite strong. We assume that the only form of unobserved heterogeneity is an iid shock to payoffs. However, it is quite straightforward to include random effects in our econometric model by modifying the importance sampler to permit correlation between the error terms. For example, in a study of entry, it will be natural to include market specific random effect. In other applications, it will be natural to include an estimated markup equation in f_i. For example, in a differentiated product market, we may estimate markups using Berry, Levinsohn, and Pakes (1995). This allows the economist to control for unobserved cost and demand shocks.

There is a tradeoff between restricting the individual utility functions and restricting the equilibrium selection mechanism. In two by two games, as seen in section 4.2, even when the equilibrium selection mechanism is completely known, the mean utility functions are not nonparametrically identified without restrictions. However, if the mean utility parameters are all known, the three conditional choice probabilities will identify the (maximally) two equilibrium selection probability parameters without imposing additional restrictions. For general games with multiple players and multiple strategies, tables 10 and 11 in the appendix show that the maximum number of equilibria typically exceeds the maximum number of moments available in the data. Therefore even when the mean utility parameters are completely known, the equilibrium selection probabilities are not nonparametrically identified without imposing additional restrictions.

5 APPLICATION

As an application of our estimator, we model strategic entry by bidders into highway procurement auctions conducted by the California Department of Transportation (CalTrans) between 1999 and 2000. Econometric modeling of entry has been of considerable interest in empirical industrial organization; see Bresnahan and Reiss (1990, 1991), Berry (1992) , Mazzeo (2002), Tamer (2002), and Ciliberto and Tamer (2009). Bajari and Hortacsu (2003), Li and Zheng (2009), Athey, Levin and Seira (2008) and Krasnokutskaya and Seim (2005) have studied entry in bidding markets.

Bidder entry in highway procurements is an attractive application for our estimator for three reasons. First, CalTrans awards its contracts using an open competitive bidding system. For each highway contract, there is a fixed and publicly announced deadline for submitting bids. Any communication between bidders about entry or other bidding decisions

would be considered collusion and could lead to civil and criminal penalties. Therefore, the assumption of a simultaneous move game is plausibly satisfied in our application.

Second, there is a well developed empirical literature for estimating structural models of bidding for highway procurement contracts, see Porter and Zona (1999), Bajari and Ye (2003), Pesendorfer and Jofre-Bonet (2003), Krasnokutskaya and Seim (2005) and Li and Zheng (2009). The flexible econometric methods proposed by Guerre, Perrigne and Vuong (2000) are commonly used in this literature, and in the empirical auctions literature more generally, to estimate bidder markups. In a first step, we use these methods to precisely estimate the expected payoffs to each player for all possible configurations of entry. In a second step, we will use the methods from the previous sections to estimate the fixed costs of entry and the parameters of our equilibrium selection mechanism.

Finally, in our data set, the dependent variable is a decision by a contractor to submit a bid to complete a single and indivisible construction project. We focus on paving contracts, instead of all contracts awarded by CalTrans, as in Pesendorfer and Jofre-Bonet (2003), in order to reduce the importance of dynamics in our application. Most of the existing entry literature considers the decision by a firm to enter a spatially separated retail or service market and compete for an indefinite length of time. We believe that a static model is more plausible in our application than in much of the previous work on entry.

Our model of entry in auctions is similar to Athey, Levin, and Serra (2008). In the first stage, contractors simultaneously choose whether to incur a fixed cost in order to participate. In the second stage, participating contractors submit sealed bids in a first-price auction and the contract is awarded to the low bidder. Our model of entry often has multiple equilibria, and there is no clear criterion from economic theory that selects a unique equilibrium of our game. Previous empirical research on entry in auctions abstracts from the multiplicity problem by imposing assumptions that guarantee a unique equilibrium.

We contribute to the literature on entry in auctions by estimating λ, the probability of selecting a particular equilibrium. We parameterize λ to allow four criteria to influence equilibrium selection: that the equilibrium is in pure strategies, the equilibrium maximizes joint profits, the equilibrium is Pareto dominated, and the equilibrium has the highest Nash product among pure strategy equilibria. To the best of our knowledge, this is the first empirical analysis of equilibrium selection in a normal form game.

5.0.1 The Bidding Game

In the model, there are $i = 1, \ldots, N$ potential bidders who bid on $t = 1, \ldots, T$ highway paving contracts. Following previous researchers, we model bidding in this industry as an asymmetric first-price auction with independent private values (see Porter and Zona (1999),

Bajari and Ye (2003), Pesendorfer and Jofre-Bonet (2003), Krasnokutskaya and Seim (2005) and Li and Zheng (2009)). Let $N(t) \subseteq \{1, \ldots, N\}$ denote the set of contractors who submit bids on project t. We assume that the set of bidders is common knowledge at the time bids are submitted.[8]

Before submitting a bid, bidder i will prepare a cost estimate $c_{i,t}$. The cost estimate of bidder i is private information which has a distribution $F_i(x_{i,t})$ where $x_{i,t}$ are publicly observable covariates which influence bidder i's cost distribution. We follow previous research and include in $x_{i,t}$ an engineering cost estimate, the distance of contractor i to project t, a measure of i's backlog, contractor fixed effects and project fixed effects. We assume that the cost distribution has a common support for all bidders and satisfies the regularity conditions discussed in LeBrun (1996) and Maskin and Riley (2000) so that an equilibrium exists, is unique, and is strictly increasing in a bidder's private information.

Let $\mathbf{b}_{i,t}(c_{i,t})$ be the bidding strategy used by bidder i in auction t, and let $\phi_{i,t}(b_{i,t})$ denote the inverse bid function. Bidders are assumed to be risk neutral. The expected profit to bidder i from bidding $b_{i,t}$ is:

$$(b_{i,t} - c_{i,t}) \prod_{j \in N(t), j \neq i} \left(1 - F_j\left(\phi_{j,t}(b_{i,t}) | x_{i,t}\right)\right)$$

Expected profit is the product of two terms. The first term is a markup, $(b_{i,t} - c_{i,t})$, which reflects bidder i's profits conditional on winning the job. Note that since the bid functions are strictly increasing, the term $1 - F_j\left(\phi_{j,t}(b_{i,t}) | x_{i,t}\right)$ is the probability that firm j's bid is greater than i's bid $b_{i,t}$. As a result, the second term $\prod_{j \in N(t), j \neq i} \left(1 - F_j\left(\phi_{i,t}(b_{i,t}) | x_{i,t}\right)\right)$ is the probability that bidder i wins the contract with a bid of $b_{i,t}$. Thus, expected profits is a markup times the probability that firm i wins the contract.

Following Guerre, Perrigne, and Vuong (2000), we rewrite bidder i's profit maximization

---

[8]In principal, it is possible to consider a model where bidders are uncertain about which firms will participate. Changing our estimator to allow for this possibility would be straightforward. However, existence and uniqueness of equilibrium bidding functions in the first price asymmetric auction with random entry has not yet been established to the best of our knowledge.

Also, we believe that allowing the set of bidders to be common knowledge corresponds most closely to what happens in this industry. Bidders that we have spoken with feel like they are quite knowledgeable about which other contractors will submit bids. Typically, the closest firms and firms with the lowest backlogs of outstanding work are most likely to bid. Also, CalTrans provides a list of plan holders for the project shortly before bids are due which allows the contractors to learn about which competing firms are interested in the project. A similar modeling assumption is made in Athey, Levin and Serra (2008) and Kransnokutskaya and Seim (2005).

problem as:

$$\max_{b_{i,t}} (b_{i,t} - c_{i,t}) \prod_{j \in N(t), j \neq i} (1 - G_j(b_{i,t}|x_t)).$$

We let $G_j(b_{i,t}|x_t)$ denote the equilibrium distribution of bids submitted by firm j conditional on the publicly observed information $x_t = (x_{i,t})_{i \in N(t)}$. The first order conditions for profit maximization imply that:

$$c_{i,t} = b_{i,t} - \left[\sum_{j \in N(t), j \neq i} \frac{g_j(b_{i,t}|x_t)}{(1 - G_j(b_{i,t}|x_t))} \right]^{-1} \tag{18}$$

Note that the right hand side of the above equation is a function of $b_{i,t}$ and the distribution of bids, which can be estimated by pooling bidding data across contracts $t = 1, ..., T$. The left hand side is the structural parameter $c_{i,t}$ which is unobserved to the econometrician. Again, following Guerre, Perrigne and Vuong (2000), we will evaluate the empirical analogue of the right hand side of the above expression in order to recover the structural cost parameter $c_{i,t}$.[9]

5.0.2 The Entry Game

In the first stage of our model, bidders simultaneously and independently decide whether to bid for contract t. Submitting a bid is a costly decision. Based on extensive industry experience, Park and Chapin (1992) report that the costs of preparing a bid is typically one percent of $b_{i,t}$. Publicly traded firms in the construction industry typically report profit margins of one to five percent. This implies that the fixed costs of bidding are nontrivial compared to a firm's profit margins, hence bidders should selectively submit bids on projects they are most likely to win. Let θ_i denote the cost to firm i of submitting a bid.

We will allow the costs of bidding to vary across firms in order to rationalize differences in participation rates. Indeed, as we shall discuss in the next section, the size distribution of firms in our data is quite skewed. While there are 271 firms submiting bids, a small number of these firms account for the majority of total output.

In our application, we shall focus on entry decisions of the four largest firms, each having a market share of at least five percent, as measured by winning bids. We shall denote these firms as $i = 1, 2, 3, 4$. We shall take the entry decisions of the other bidders $N(t) \backslash \{1, 2, 3, 4\}$

---

[9]In Bajari, Houghton and Tadelis (2006), we argue that bidders' payoffs are somewhat more complicated than in the above model because of change orders and cost overruns. However, we find that the method of Guerre, Perrigne and Vuong (2000) estimates bidder profits quite well. As we report below, our estimates seem sensible given what is known about bidder markups and other structural parameters.

as predetermined. It would obviously be preferable to endogenize the entry decisions of all bidders. However, repeatedly solving for all Nash equilibria with approximately three hundred players is not computationally tractable. Furthermore, we believe that it innocuous to take the entry decisions of small, fringe firms as exogenous. Such firms rarely win large CalTrans contracts, due to lack of capital and managerial expertise to complete large projects at a competitive price. Fringe firms typically win much smaller jobs in the public sector, such as resurfacing streets for a mid-sized California city, or smaller private sector jobs, such as resurfacing parking lots for small businesses. In our CalTrans data, fringe firms have little influence on the winning bid and hence on profits at the margin. We believe that it is much more important to carefully model the largest firms' entry decisions and this is where we focus our attention.

Let $a_{i,t} = 1$ if firm i decides to submit a bid on project t and $a_{i,t} = 0$ otherwise. Given $a_{i,t}$, $i = 1, \ldots, 4$ for the largest bidders, the set of bidders who participate will be denoted as $N(t|a)$. This set includes all the fringe firms observed to participate in the data and those firms $i = 1, \ldots, 4$ for which $a_{i,t} = 1$. If one of our four largest firms i enters, then conditional on a and x_t, i's profit will be:

$$u_i(a; x_t, \theta_i) = \int (\mathbf{b}_{i,t}(c_{i,t}; x_t, N(t|a)) - c_{i,t}) \prod_{j \neq i} (1 - G_j(b_{i,t}|x_t, N(t|a))) dF(c_{i,t}|x_{i,t}) - \theta_i \quad (19)$$

In the above, $\mathbf{b}_{i,t}(c_{i,t}; x_t, N(t|a))$ denotes firm i's bid function and $G_j(b_{i,t}|x_t, N(t|a))$ firm j's bid distribution when the set of entrants is $N(t|a)$ and the publicly observed project characteristics are x_t.

The above expression implicitly assumes the following timeline for the game: First, all large firms simultaneously decide whether to enter. The project characteristics, x_t, and the entry decisions of the fringe firms are common knowledge. Second, after entering, each of the four largest firms observes which other large firms have entered the market. Third, all participating bidders independently make their cost draws $c_{i,t}$. Finally, firms submit sealed bids and the lowest bidder wins. In the above equation, $u_i(a; x_t, \theta_i)$ is i's profits conditional on the entry decisions of the other large firms, the publicly observed data x_t, and the parameter θ_i. Given $u_i(a; x_t, \theta_i)$, we can specify a normal form game in the framework of Section 2.

5.1 Estimation

Our estimation procedure consists of two steps. In the first step, we form an estimate of the term $\int (\mathbf{b}_{i,t}(c_{i,t}; x_t, N(t|a)) - c_{i,t}) \prod_{j \neq i} (1 - G_j(b_{i,t}|x_t, N(t|a))) dF(c_{i,t}|x_{i,t})$ in Equation 19 by adapting the approach proposed by Guerre, Perrigne, and Vuong ((2000)). In the second

step, we take the estimates from the first stage and estimate θ_i, the fixed cost of preparing a bid, and λ, the selection of equilibrium, using the methods from Sections 2 and 3.

5.1.1 Markup Estimation

The idea behind Guerre, Perrigne, and Vuong's estimator is quite simple. The left hand side of Equation (18) is the bidder's private information, $c_{i,t}$, which is unobserved to the econometrician. The right hand side is a function of the bid, $b_{i,t}$, the density of bids, $g_j(b_{i,t}|x_t)$, and the CDF of bids, $G_j(b_{i,t}|x_t)$. By pooling observations from contracts $t = 1, \ldots, T$, we construct an estimate $\widehat{g}_j(b_{i,t}|x_t)$ and $\widehat{G}_j(b_{i,t}|x_t)$ using standard nonparametric techniques. We then construct an estimate of firm i's private information $\widehat{c}_{i,t}$ by evaluating the empirical analogue of the right hand side of Equation (18). Once we have recovered the distribution of a firm's private information, we can then compute the ex-post entry profits in Equation (19).

5.1.2 Equilibrium Selection

In the second step, we estimate the fixed costs of bidding, θ_i, and the probability that a particular equilibrium is selected, taking the expected entry profits in Equation 19 as given. We use a conditional logit as a parsimonious specification of λ. Following the previous literature on entry games, we have found four criteria proposed for equilibrium selection on entry games. First, in empirical papers such as Tamer (2002), Ciliberto and Tamer (2009) and Andrews, Berry, and Jia (2005), and Jia (2008), it is usually assumed that only pure strategies are used in the entry game. The authors argue that mixed strategy equilibria are a priori implausible in these markets. However, in a related experimental literature, Levin and Smith (2001) argue that mixed strategy equilibrium seems the most reasonable in auction entry experiments. To acknowledge this possibility, we construct a dummy variable $MIXED(\pi)$, which equals one if the equilibrium π involves mixed strategies.

Second, we allow λ to depend on whether the equilibrium is efficient, in the sense that it maximizes joint payoffs. Economic theory and Ciliberto and Tamer (2009) have both proposed this criteria for equilibrium selection mechanism. Since the firms in our data interact repeatedly, they obviously have incentives to tacitly collude on an equilibrium that maximizes industry surplus.

Third, we include a dummy variable that equals one if an equilibrium is Pareto dominated. It is commonly assumed that Pareto dominated equilibria are less plausible and, therefore, less likely to be observed in the data.

Finally, we include the Nash product of a player's utilities for pure strategy equilibria.

Harsanyi and Selten (1988) argue that risk dominant equilibria are more plausible. An equilibrium having a large Nash product implies that deviating from the observed equilibrium behavior is especially costly, hence the equilibrium is more likely to be self-reinforcing.

5.2 The Data

We have constructed a unique data set of bidding by highway contractors in the State of California from 1999-2000. We observe 414 contracts awarded by the California Department of Transportation (CalTrans) during this time period.[10] The total value of winning bids in this data is \$1.326 billion. There are a total of 1,938 bids and 271 bidders in our sample. Highway improvement projects are awarded using open competitive bidding, which means that any qualified contractor can submit a bid and contracts are awarded to the lowest qualified bidder.[11] This data set is described in detail in Bajari, Houghton, and Tadelis (2006). We will describe some of the highlights of the data and the industry in this section.

For each contract t, we observe a detailed list of covariates including $b_{i,t}$, the bid of contractor i on project t, EST_t, the engineer's cost estimate for project t, $DIST_{i,t}$ the distance in miles of firm i to project t, $CAP_{i,t}$, the capacity utilization of firm i at the time of bidding for project t, and $FRINGE_{i,t}$, a dummy variable equal to one if firm i is a fringe firm, defined as firms with market shares of less than one percent. The data set includes the bids for all contractors, not just the winning bids. The engineer's estimate, EST_t, is constructed by CalTrans as a fair market value for completing the work. The project plans and specifications contain an exhaustive list of work items; the estimate is then formed using blue book prices for specific work items and local material prices.

Table 2 summarizes the market shares of the 10 largest firms in the industry, where share is defined using the winning bids. The market shares in this industry are quite skewed. The largest firm, Granite Construction Company, has a share of 27.2 percent, compared to a share of 1.9 percent for the 10th largest firm, Sully Miller Contracting. This skewed distribution suggests that productivity varies across firms and hence it is important to include firm fixed effects in our estimates of g_j and G_j.

---

[10]The data contain contracts for paving and excludes other contracts such as bridge repair. We look at only the subset of contracts where asphalt costs accounted for less than 1/3 of the winning bid. We focus on paving contracts since capacity constraints and the dynamics emphasized in Jofre-Bonet and Pesendorfer (2003) are less important for this set of contracts. In Bajari, Houghton, and Tadelis (2006) we produce closely related structural estimates. Here we adjust our estimates to allow for dynamics through non-trivial capacity constraints. We find that such capacity constraints have little effect on estimated markups. In order to simplify the presentation, we focus on a static model of profits, although it would be quite straightforward to extend the analysis to allow for capacity to influence profit margins and markups.

[11]In about 5 percent of the projects in our sample, CalTrans rejects all bids and awards the contracts again at a later date. We do not include these contracts in our sample.

Table 2: Bidder Identities and Summary Statistics

Company	Share	No. Wins	No. Bids Entered	Participation Rate	Total Bids for Contracts Awarded
Granite Construction Company	27.2%	76	244	58.9%	343,987,526
E. L. Yeager Construction Co Inc	10.4%	13	31	7.5%	132,790,460
Kiewit Pacific Co	6.6%	5	30	7.2%	112,057,627
M. C. M. Construction Inc	6.5%	2	6	1.4%	89,344,972
J. F. Shea Co Inc	3.3%	9	40	9.7%	43,030,861
Teichert Construction	3.3%	16	43	10.4%	40,177,076
W. Jaxon Baker Inc	2.9%	13	65	15.7%	37,702,631
All American Asphalt	2.2%	14	33	8.0%	30,764,962
Tullis And Heller Inc	2.1%	10	16	3.9%	27,809,535
Sully Miller Contracting Co	1.9%	17	49	11.8%	27,889,186

Table 2 shows that the largest firms tend to bid more often as measured by their participation rate. However, we note that the second largest firm only submits bids for 7.5 percent of the jobs compared to Granite Construction Company, which submits bids for 58.9 percent of the jobs. Hence it is important to account for firm specific differences in the costs of bidding, θ_i.

Table 3 provides summary statistics about the bids. In our data, the average winning bid is \$3.2 million dollars, which is about 6 percent below the engineer's estimate. Meanwhile, comparing the winning bid to the second highest bid, the average money left on the table is about 6 percent of the estimate. This suggests that there is asymmetric information in this market. If the low bidder knew the cost of the second lowest bidder, then in a Nash equilibrium we would expect these two bids to be much closer. Leaving money on the table does not increase the probability of winning and only decreases the profit of the low bidder.

Table 4 demonstrates that the the ranking of the bids corresponds closely to the ranking of the contractors' distances from the project. For instance, $DIST1$, the distance of the lowest bidder, is smaller than $DIST2$, the distance of the second lowest bidder. The closest contractor has a lower cost of hauling asphalt to the project site and is therefore more likely to win the project.

In Table 5, we regress the bids on the various cost controls. In the first column, we regress bids on the engineer's estimate. This has an R^2 of 0.987 with a coefficient of 1.02, suggesting that the engineer's estimate is a very powerful explanatory variable. Starting with second column, we change the dependent variable to $b_{i,t}/EST_t$ since the variance of the errors in the bid regressions are likely to be proportional to EST_t. The next set of regressions demonstrates that distance, the fringe firm dummy, project fixed effects, and firm fixed effects for the largest four firms are all economically and statistically significant regressors. Their resultant signs are as anticipated, for example, the positive distance coefficient reflects

Table 3: Bidding Summary Statistics

	Obs	Mean	Std. Dev.	Min	Max
Winning Bid	414	3,203,130	7,384,337	70,723	86,396,096
Markup: (Winning Bid-Estimate)/Estimate	414	-0.0617	0.1763	-0.6166	0.7851
Normalized Bid: Winning Bid/Estimate	414	0.9383	0.1763	0.3834	1.7851
Second Lowest Bid	414	3,394,646	7,793,310	84,572	92,395,000
Money on the Table: Second Bid-First Bid	414	191,516	477,578	68	5,998,904
Normalized Money on the Table: (Second Bid-First Bid)/Estimate	414	0.0679	0.0596	0.0002	0.3476
Number of Bidders	414	4.68	2.30	2	19
Distance of the Winning Bidder	414	47.47	60.19	0.27	413.18
Travel Time of the Winning Bidder	414	56.95	64.28	1.00	411.00
Utilization Rate of the Winning Bidder	414	0.1206	0.1951	0.0000	0.9457
Distance of the Second Lowest Bidder	414	73.55	100.38	0.19	679.14
Travel Time of the Second Lowest Bidder	414	82.51	97.51	1.00	614.00
Utilization Rate of the Second Lowest Bidder	414	0.1401	0.2337	0.000	0.9959

Table 4: Distance to Job Site

	Mean	Std. Dev.	Min	Max
DIST1	47.47	60.19	0.27	413.18
DIST2	73.55	100.38	70.19	679.14
DIST3	75.47	95.56	0.13	594.16
DIST4	84.38	89.87	1.45	494.08
DIST5	76.12	86.33	1.25	513.31

Table 5: Bid Function Regressions

Variable	$b_{i,t}$	$b_{i,t}/EST_t$	$b_{i,t}/EST_t$	$b_{i,t}/EST_t$
EST_t	1.025 (56.26)			
$DIST_{i,t}$		.000246 (5.66)	.000223 (5.01)	
$UTIL_{i,t}$			0.02539 (0.93)	
$FRINGE_{i,t}$				0.4288 (4.65)
Constant	-25686 (0.56)	1.19 (94.9)	1.001 (79.98)	
Fixed Effects	No	Project	Project	Project/Firm
R^2	0.989	0.5245	0.5292	0.5321

Number of observations = 1938; t-statistics are reported in parentheses. Fringe is a dummy variable that equals one for a fringe firm. Util denotes utilization rates.

Table 6: Logit Model of Entry

	I	II	III
Constant	-.9067	-1.6811	
	(7.91)	(7.53)	
$DIST_{i,t}$	-0.00218	-0.00322	-0.00854
	(5.42)	(5.66)	(4.85)
Granite		2.889	4.4537
		(13.28)	(7.31)
E. L. Yeager		-	-
Kiewit Pacific		-0.1527	1.1969
		(0.57)	(2.1)
M. C. M.		-1.786	-.70779
		(3.94)	(1.12)
Fixed Effects	No	No	Project
Observations	1656	1656	1068
Number of Groups			261
Log-Likelihood	-784.20	-511.86	-101.0728

The dependent variable is whether one of the four largest firms in the industry decides to submit a bid in a particular procurement; t-statistics are reported in parentheses.

bidders' large transportation costs. Indeed, the average distance of a firm from the project is 72 miles with a standard deviation of 92 miles. Our results imply that increasing the distance by a standard deviation will raise $b_{i,t}/EST_t$, relative to the engineer's estimate, by about 2.3 percent. This is substantial because firms are believed to have profit margins of less than five percent. Granite Construction, the largest firm in our sample, reports a profit margin of only 3.31 percent and an operating margin of 5.15 percent. Finally, $b_{i,t}/EST_t$ for fringe firms is 4.2 percent lower. This evidence supports using our simplifying assumption that fringe firm entry is exogenous. Fringe firms bid much higher and are unlikely to win many projects as a result.

In Table 6, we estimate a logit model of entry for the four largest firms in the industry. For each of these firms, we calculate their distance to each project t even the firm does not submit a bid. We find that participation is a decreasing function of the firm's distance to the project. Also, there is heterogeneity across the firms in terms of their participation decisions, suggesting that inclusion of firm level effects θ_i is important in modeling entry.

5.2.1 Estimates of Profits

We estimate bidder markups using the approach by Guerre, Perrigne and Vuong (2000). Given the number of covariates in our application, it is not feasible to nonparametrically estimate the distribution of bids g_j and G_j. Instead, we use a semiparametric approach. We first run a regression, as in Table 5:

$$\frac{b_{i,t}}{EST_t} = x_{i,t}'\alpha + u^{(t)} + \epsilon_{i,t},$$

where the dependent variable is normalized by dividing through by the engineer's estimate. In addition, we include an auction specific fixed effect, $u^{(t)}$. Let $\widehat{\alpha}$ denote the estimated value of α and let $\widehat{\epsilon}_{i,t}$ denote the fitted residual. We will assume that the residuals of this regression are iid. Let $\widehat{H}$ denote the Kaplan-Meier estimate of the CDF of the fitted residuals.[12]

Under these assumptions, the estimated bid distributions satisfy:

$$\begin{aligned}
\widehat{G}_i(b|z_{j,t}, N(t)) \qquad &= \Pr\left(\frac{b_{i,t}}{EST_t} \leq \frac{b}{EST_t}\right) \\
&= \Pr\left(x_{i,t}'\widehat{\alpha} + \widehat{u}^{(t)} + \widehat{\epsilon}_{i,t} \leq \frac{b}{EST_t}\right) \\
&= \widehat{H}\left(\widehat{\epsilon}_{i,t} \leq \frac{b}{EST_t} - x_{i,t}'\widehat{\alpha} - \widehat{u}^{(t)}\right).
\end{aligned}$$

That is, the distribution of the fitted residuals, $\widehat{\epsilon}_{i,t}$, can be used to infer the distribution of the bids. As the estimates in Table 5 suggest, variation in the estimated bid distribution will be driven by three factors. The first is the auction fixed effects, $\widehat{u}^{(t)}$. The second is the distance of each firm from the project; the further a particular firm is away from the project, the higher its bid will be. The third is the firm fixed effects. The largest four firms will bid more aggressively than the smaller fringe firms. [13]

Recall that earlier we demonstrated that firm specific profit shifters are sufficient to identify our model under fairly mild parametric assumptions. In our analysis, distance and firm fixed effects will be the primary shifters of individual firms' profits. Each firm has a unique distance to a particular contract t. The variation in transportation cost across

---

[12]We estimate the density h of the fitted residuals using kernel density estimation with an estimated optimal bandwidth. Since there are 1938 fitted residuals, the estimates of H and h are quite precise given θ. Ideally, our estimates would take into account the first stage estimation error in θ. However, the computational burden of performing a resampling procedure such as the bootstrap is considerable and beyond the scope of this research.

[13]We note that we must estimate the distribution of each firm i's bid even if it does not participate in a particular auction. We have therefore computed the distance of each of the four largest firms from all $t = 1, \ldots, 414$ projects even if they did not submit a bid. We use Equation 18 to infer the distribution of firm i's bid in this case. The estimates of Table 6 suggest that bidder i will have a low chance of winning a particular procurement t if it is a long distance from the project site.

Table 7: Margin Estimates

Variable	Num. Obs.	Mean	Std. Dev.	Median	25th Percentile	75th Percentile
Profit Margin	1938	0.0644	0.1379	0.0271	0.0151	0.0520

The markup is defined as 1 minus the ratio of the estimated cost, which is private information, to the actual bid.

projects generates shifts in the payoffs of individual firms and therefore allows us to identify our model.

In Table 7, we summarize the distribution of estimated markups on the 1,938 bids in our data set. The average markup is about 6 percent. Note that the distribution of markups is skewed: the median markup is 2.71 percent and the 75th percentile is 5.2 percent. These markups are comparable to the reported margins of Granite Construction and, more broadly, Census data on markups in the construction industry.

5.2.2 Equilibrium Selection Parameters and Bid Costs

As mentioned in Section B.2, the choice of the importance density is vital for our estimation procedure. To choose an importance density, we first estimate a private-information version of the entry game to obtain starting values for bid preparation costs and the profit scale parameter. While these parameters will not generally be consistent estimates for the complete information game, they should be in the correct neighborhood, which greatly aids convergence of our importance sampling MSM procedure. All equilibrium selection parameters are initially set at zero. We perform several sequential estimations, using 409 importance games initialized with the previous iteration's final value. When the parameter values are consistent after several iterations, we run the Laplace-type estimator of Chernozhukov and Hong (2003) to generate standard errors and to ensure that the estimated coefficients are robust to the optimization method used in the initial steps. The results are reported in Table 8.

The first parameter that needs interpretation is the coefficient on profit scale. The expected entry profits for a firm are expressed in tens of thousands of dollars. These expected profits are then multiplied by the profit scale parameter, equal to 0.0965. Therefore we should interpret each unit of the fixed costs of bidding as about $96,500. For example, Kiewit Pacific faces an average cost of approximately $161,650 to prepare a bid, which is roughly five percent of the average winning bid. This amount varies across the four firms, from $22,590 for Granite Construction to $235,460 for M. C. M. Construction. Given that winning bids are drawn from the left tail of the bid distribution, these numbers are roughly consistent

Table 8: Games Estimation Results

Variable	Mean	Median	Std.Dev.	95% Confidence Interval	
Equilibrium Selection Parameters (λ)					
Pure Strategy	-1.3524	-1.5345	0.7979	-2.4903	0.1954
Joint Profit Maximizing	6.4365	6.4226	0.5321	5.6151	7.5149
Dominated	-5.3841	-5.3316	0.7002	-6.7164	-4.0986
Nash Product	4.4143	4.2025	1.1017	2.9651	6.4836
Profit Scale					
Profit Scale	0.0965	0.0954	0.0015	0.0954	0.0984
Bid Preparation Costs (θ_i)					
Granite Construction	0.2341	0.2393	0.0977	0.0679	0.4271
E. L. Yeager	1.4583	1.4757	0.0941	1.2563	1.6227
Kiewit Pacific	1.6751	1.6720	0.0511	1.5775	1.7789
M. C. M. Construction	2.4490	2.4360	0.1144	2.2547	2.6966

Estimation and inference was performed using the LTE method of Chernozhukov and Hong (2003). A Markov chain was generated with 500 draws for each parameter. 409 importance games were used in the importance sampler for the 409 observations.

with Park and Chapin (1992), who argue that bid preparation costs are approximately equal to one percent of the total bid in magnitude.

A second check of validity is that bid costs should be monotonically and inversely related to participation rates. Indeed, the two firms with similar participation rates, E. L. Yeager and Kiewit Pacific, have almost identical bid preparation costs, while the bid costs for M. C. M. Construction, with a low entry rate of 1.4 percent, are much higher.

Turning to the parameters of the equilibrium selection mechanism, we have several interesting results. Firstly, mixed strategy equilibria are more likely than pure strategy equilibria. This is consistent with the results in Levin and Smith (2001), who find support for mixed strategy equilibria in auction entry experiments. They argue that a pure strategy equilibrium requires too much coordination of behavior on the part of agents. In a pure strategy equilibrium, agents must coordinate on which subset of the potential bidders will submit bids in an auction. The number of subsets of $\{1, ..., N\}$ is large even for moderately sized N. In our application, coordinating entry decisions through explicit communication would be collusion, making coordination even more difficult.

We find that efficiency has a strong effect on the probability of an equilibrium being chosen. Given that there are typically many pure and mixed strategies in a given game, this shifter is by far the most influential in deciding which equilibrium is played. This suggests firms are tacitly colluding, since they are more likely to choose the Nash equilibrium which

maximizes joint profits.

We find that the coefficient on dominated equilibria is strongly negative. This is also consistent with tacit collusion by bidders in the auction. Finally, the coefficient on the equilibrium with the highest Nash product is strongly positive. This indicates that among pure strategy equilibria, the one with the highest Nash product has a strong tendency to be played in the data.

6 Conclusion

Estimating models that are consistent with Nash equilibrium behavior is an important empirical problem. In this paper, we have proposed algorithms which estimate both the utilities and the equilibrium selection parameters for static, discrete games. Our algorithms can be applied to general normal form games, umlike those of previous research that frequently apply to specific examples such as entry games. The algorithms use computationally efficient methods and our Monte Carlo work demonstrates that they work well even with a moderate number of observations.

We also study the nonparametric identification of these games. We propose two strategies: identification at infinity and exclusion restrictions. If payoffs are restricted to be a linear index and covariates have full support, we can find a region in which players $-i$ will play a given strategy a_{-i}. This allows us to treat each agent i's choice of strategy as a single agent model. As a result, we can identify payoff parameters. Knowing payoffs, we can restrict attention to regions where ϵ has a small impact on total utility. This means that the payoff matrix is known, up to a small error. In turn, this allows us to identify λ, the equilibrium selection mechanism.

We also show that exclusion restrictions allow us to identify our model. We search for covariates x_i which shift the utility of i, but do not enter the utility of $-i$. The presence of these covariates allow us to shift the payoff matrix one cell at a time. Generating this variation allows us to identify the model. In both approaches, λ must not be too complicated. If there are more equilibrium than moments in the data, we cannot identify a fully general model of λ.

As an application of our methods, we study the decision of four large construction firms' entry into procurement auctions in California. We recover fixed bid preparation costs for each of the four firms which rationalize their entry rates into these auctions. The application also highlights one strength of our approach: the ability to estimate an equilibrium selection mechanism. Our estimates indicate that mixed strategy equilibria are selected with a greater probability than pure strategy equilibria. We also find that the equilibrium mechanism

favors joint profit maximizing and non-Pareto dominated equilibria. Among pure strategy equilibria, the one with the largest Nash product is selected with higher probability. The estimation method we propose is, to our knowledge, the most efficient approach capable of accommodating both multiplicity and mixed strategy equilibria.

References

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

Ackerberg, D. and G. Gowrisankaran (2006). Quantifying equilibrium network externalities in the ACH banking industry. *RAND Journal of Economics 37*(3), 738–761.

Aguiregabiria, V. and P. Mira (2007). Sequential Estimation of Dynamic Games. *Econometrica 75*, 1–53.

Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press, Cambridge, MA, U.S.A.

Andrews, D. W. K., S. Berry, and P. Jia (2005). Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location. Yale University Working Paper.

Athey, S., J. Levin, and E. Seira (2008). Comparing open and sealed bid auctions: Theory and evidence from timber auctions. Stanford University Working Paper.

Bajari, P., C. L. Benkard, and J. Levin (2007). Estimating dynamic models of imperfect competition. *Econometrica 75*, 1331–1370.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2006). Estimating static models of strategic interactions. working paper, forthcoming, Journal of Business and Economic Statistics.

Bajari, P. and A. Hortacsu (2003). The winner's curse, reserve prices, and endogenous entry: Empirical insights from ebay auctions. *RAND Journal of Economics 34*(2), 329–55.

Bajari, P., S. Houghton, and S. Tadelis (2006). Bidding for incomplete contracts: An empirical analysis. NBER Working Papers 12051, National Bureau of Economic Research, Inc.

Bajari, P. and J. Krainer (2004). An empirical model of analysts' stock recommendations: Market fundamentals, conflicts of interest, and peer effects. Duke University Working Paper.

Bajari, P. and L. Ye (2003). Deciding between competition and collusion. *The Review of Economics and Statistics 85*(4), 971–989.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in equilibrium. *Econometrica 63*(4), 841–90.

Berry, S. and P. C. Reiss (2007). *Empirical Models of Entry and Market Structure*, Volume III of *Handbook of Econometrics.* Elsevier, Amsterdam, Holland.

Berry, S. and E. Tamer (2006). Identification in models of oligopoly entry. In Blundell, Newey, and Persson (Eds.), *Advances in Economics and Econometrics, Vol 2, Ninth World Congress, 46-85.* Cambridge, England: Cambridge University Press.

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

Bjorn, P. and Q. Vuong (1984). Simultaneous equations models for dummy endogenous variables: A game theoretic formulation with application to labor force participation. SSWP No. 537, California Institute of Technology.

Bresnahan, T. F. and P. C. Reiss (1990). Entry in monopoly markets. *Review of Economic Studies 57*, 531–553.

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

Brock, W. and S. Durlauf (2001). Discrete choice with social interactions. *Review of Economic Studies 62*(2), 235–260.

Brock, W. and S. Durlauf (2006). Multinomial choice with social interactions. In L. Blume and S. Durlauf (Eds.), *The Economy as an Evolving Complex System III.* Oxford, England: Oxford University Press.

Chernozhukov, V. and H. Hong (2003). A MCMC approach to classical estimation. *Journal of Econometrics 115*(2), 293–346.

Chiappori, P.-A., S. Levitt, and T. Groseclose (2002). Testing mixed strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review 92*(4), 1138–1151.

Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in airline markets. Forthcoming, Econometrica.

Davis, P. (2006). Estimation of quantity games in the presence of indivisibilities and heterogeneous firms. *Journal of Econometrics 134*(1), 187–214.

El-Gamal, M. A. and D. M. Grether (1995). Are people Bayesian? Uncovering behavioral strategies. *Journal of the American Statistical Association 90*(432), 1137–1145.

Fox, J. (2007). Semiparametric estimation of multinomial discrete-choice models using a subset of choices. *The RAND Journal of Economics 38*(4), 1002–1019.

Gale, D. and H. Nikaido (1965). The Jacobian matrix and global univalence of mappings. *Mathematische Annalen 159*(2), 81–93.

Guerre, E., I. Perrigne, and Q. Vuong (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica 68*(3), 525–574.

Hajivassilou, V. and P. Ruud (1994). Classical estimation methods for ldv models using simulation. In *Handbook of Econometrics, Engle, R.F. and D. McFadden (eds)*, Volume IV, pp. 2384–2438.

Harsanyi, J. and R. Selten (1988). *A General Theory of Equilibrium Selection in Games.* MIT Press, Cambridge, MA.

Heckman, J. (1990). Varieties of selection bias. *American Economic Review 80*(2), 313–318.

Holmes, T. (2008). The diffusion of wal-mart and economies of density. *NBER Working Paper, forthcoming Econometrica.*

Jia, P. (2008). What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica 76*(6), 1263–1316.

Jofre-Bonet, M. and M. Pesendorfer (2003). Estimation of a dynamic auction game. *Econometrica 71*(5), 1443–1489.

Keane, M. and K. Wolpin (1997). The career decisions of young men. *The Journal of Political Economy 105*(3), 473–522.

Keane, M. and K. Wolpin (2001). The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review 42*(4), 1051–1103.

Krasnokutskaya, E. and K. Seim (2005). Bid preference programs and participation in highway procurement auctions. University of Pennsylvania Working Paper.

Lebrun, B. (1996). Existence of an equilibrium in first-price auctions. *Economic Theory 7*, 421–423.

Li, T. and X. Zheng (2009). Entry and competition effects in first-price auctions: theory and evidence from procurement auctions. *Review of Economic Studies 76*(4), 1397–1429.

Manski, C. (1988). Identification of binary response models. *Journal of the American Statistical Association 83*(403), 729–38.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies 60*(3), 531–42.

Manuszak, M. and A. Cohen (2004). Endogenous market structure with discrete product differentiation and multiple equilibria: An empirical analysis of the competition between banks and thrifts. Carnegie Mellon Working Paper.

Maskin, E. and J. Riley (2000). Equilibrium in sealed high bid auctions. *Review of Economic Studies 67*, 439–454.

Mazzeo, M. (2002). Product choice and oligopoly market structure. *RAND Journal of Economics 33*(2), 221–242.

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

McKelvey, R. and A. McLennan (1996). The computation of equilibrium in finite games. In H. Amman, D. A. Kendrick, and J. Rust (Eds.), *The Handbook of Computational Economics*, pp. 87–142. Elsevier, Amsterdam, Holland.

McKelvey, R. and A. McLennan (1997). The maximal number of regular totally mixed Nash equilibria. *Journal of Economic Theory 72*, 411–425.

McKelvey, R. and T. Palfrey (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior 10*, 6–38.

McLennan, A. (1997). The maximal generic number of pure Nash equilibria. *Journal of Economic Theory 72*, 408–410.

Nash, J. (1951). Non-cooperative games. *The Annals of Mathematics 54* (2), 286–295.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57* (5), 1027–1057.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2005). Moment inequalities and their application. working paper, Harvard University.

Park, W. R. and W. B. Chapin (1992). *Construction Bidding: Strategic Pricing for Profit* (2nd ed.). Hoboken, New Jersey: Wiley.

Pesendorfer, M. and P. Schmidt-Dengler (2003). Identification and estimation of dynamic games. LSE Working Paper.

Porter, R. H. and J. D. Zona (1999). Ohio school milk markets: An analysis of bidding. *RAND Journal of Economics 30* (2), 263–288.

Rothenberg, T. (1971). Identification in parametric models. *Econometrica 39* (3), 577–591.

Ryan, S. P. and C. E. Tucker (2009). Diversification and the dynamics of technology adoption. MIT Working Paper.

Sannikov, Y. (2007). Games with Imperfectly Observable Actions in Continuous Time. *Econometrica 75* (5), 1285–1329.

Seim, K. (2001). Spatial differentiation and market structure: the video retail industry. Yale University Ph.D. Dissertation.

Shachat, J. and M. Walker (2004). Unobserved heterogeneity and equilibrium: an experimental study of Bayesian and adaptive learning in normal form games. *Journal of Economic Theory 114*, 280–309.

Smith, J. L. and D. Levin (2001). Entry coordination in auctions and social welfare: An experimental investigation. *International Journal of Game Theory 30*, 321–350.

Soetevent, A. and P. Kooreman (2007). A discrete-choice model with social interactions: with an application to high school teen behavior. *Journal of Applied Econometrics 22* (3), 599–624.

Stern, S. and M. Engers (2002). Long-term care and family bargaining. *International Economic Review 43*(1), 73–114.

Stern, S. and B. Heidemann (1999). Strategic play among family members when making long-term care decisions. *Journal of Economic Behavior and Organization 40*(1), 29–57.

Sweeting, A. (2008). The strategic timing of radio commercials: An empirical analysis using multiple equilibria. Forthcoming, RAND Journal of Economics.

Tamer, E. (2002). Incomplete bivariate discrete response model with multiple equilibria. *Review of Economic Studies 70*, 147–167.

Turocy, T. (2008). *TOWARDS A BLACK-BOX SOLVER FOR FINITE GAMES: COMPUTING ALL EQUILIBRIA WITH GAMBIT AND PHCPACK*, Volume 148. Berlin, Germany: Springer.

A   Proofs of Theorems

Proof of theorem 1: For each i, and $a = (a_i, k_{-i})$, the observational data identifies the conditional probabilities $P(\epsilon_i(a) + x_i^a \beta_i^a > \epsilon_i(a_i', k_{-i}) + x_i^{a_i', k_{-i}} \beta_i^{a_i', k_{-i}} | a_{-i} = k_{-i}, x)$. Because of assumption 4, in the limit this converges to a single agent decision problem for player i:

$$\lim_{|x| \to \infty, x \in \mathcal{T}_{-i}^{k_{-i}}} P(\epsilon_i(a) + x_i^a \beta_i^a > \epsilon_i(a_i', k_{-i}) + x_i^{a_i', k_{-i}} \beta_i^{a_i', k_{-i}} | x).$$

This implies that for each k_{-i}, the ordering of $\epsilon_i(a) + x_i^a \beta_i^a$ is identified for $a = (a_i, k_{-i})$ across a_i, along the path $|x| \to \infty$ and $x \in \mathcal{T}_{-i}^{k_{-i}}$. Hence the linear utility indexes $x_i^a \beta_i^a$ are also identified along this path. (cf. Amemiya (1985) and Fox (2007)) Assumption 5 further identifies the coefficient parameters β_i^a. According to assumption 5, for every $\beta_i^a \neq \beta_{i0}^a$, there exist a set of x with positive probability such that $x_i^a \beta_i^a \neq x_i^a \beta_{i0}^a$, which implies identification of β_i^a. $\square$

Proof of theorem 2: Using assumption 7, we can recover the mixing probabilities with arbitrary precision using larger and larger values of the covariates x. By assumption 6, the equilibrium selection probabilities with smaller values of the latent utility indexes are obtained by extrapolation along the remote sections of a ray that emanates from the origin and goes through the latent utility indexes. $\square$

Proof of theorem 3: The proof follows similarly to that of the previous section. Hold x fixed. Consider a large but finite number of values of x_i equal to K for each agent. Consider all the K^N distinct vectors of the form $x = (x_1, \ldots, x_N)$ that can be formed.

Consider the moments generated by these K^N distinct covariates. The number of moments is equal to $K^N \cdot \left(\prod_i \#A_i - 1 \right)$. The number of mean utility parameters is equal to $\sum_i K (\#A_i - 1) \prod_{j \neq i} \#A_j$ plus the number of parameters required to characterize λ. The maximum number of parameters required to characterize λ depends on the total number of players and the number of strategies for each player, but does not depend directly on x_i. Thus, the second part of assumption 8 implies that the number of equilibrium selection probability parameters grows at most at the rate of K^{N-1}. The number of utility parameters depends linearly on K and the number of equilibrium selection probabilities grows at the rate of K^{N-1}, but the number of moments grows exponentially at the rate of K^N. Therefore, by choosing sufficiently large values for K, the order condition is satisfied. $\square$

B.1   Identification at Infinity in Two-by-Two Games

For the purpose of illustration, in this appendix we specialize the arguments for identification to the two by two game.

Identifying the mean payoff functions   Given a linear specification of the mean utility functions, let $f_i(\tau, x) = x_i^\tau \beta_i^\tau$ for $\tau = (2-i) Bk + (i-1) jR$. The following assumption requires a rich support of the covariates.

Assumption 9. *For each $i = 1, 2$, $j = T, B$, and $k = L, R$, there exists a set $\mathcal{T}_i^{j(2-i)+k(i-1)}$ of covariates x such that $\lim_{||x|| \to \infty, x \in \mathcal{T}_i^{j(2-i)+k(i-1)}} P\left[a_i = j^{2-i} k^{i-1} | x\right] = 1$.*

Assumption 9 requires that for each player i and for each of player i's strategies, we can shift the covariates x along a dimension such that action a_i is a dominant strategy for player i with probability arbitrarily close to 1. For example, for $i = 2$ and $k = L$, Assumption 9 requires that along a path of $||x|| \to \infty$, $x \in \mathcal{T}_2^L$, $P\left[a_2 = L | x\right] \to 1$, or

$$P\left[x_2^{TR}\beta_2^{TR} + \epsilon_2\left(TR\right) < 0, x_2^{BR}\beta_2^{BR} + \epsilon_2 < 0\left(BR\right)\right] \longrightarrow 1.$$

Assumptions 1, 2, and 9 allow us to identify the mean utilities along these paths. The next assumption requires that we can extrapolate knowledge of the deterministic utilities along this path to other values of x on its support.

Assumption 10. *For each $i = 1, 2$, $j = T, B$, $k = L, R$, $x \in \mathcal{T}_i^{j(2-i)+k(i-1)}$, there exists some $L_0 > 0$ such that*

$$\inf_{L \geq L_0} \min \text{eig} E\left[x_i^\tau x_i^{\tau\prime} | x \in \mathcal{T}_{3-i}^{j(i-1)+k(2-i)}, ||x|| \geq L\right] > 0.$$

Assumption 10 requires that the linear deterministic payoff functions $x_i^\tau \beta_i^\tau$ can be extrapolated from the path $||x|| \to \infty$, $x \in \mathcal{T}_{3-i}^{j(i-1)+k(2-i)}$ to the full support of x.

Theorem 4. *Under Assumptions 1, 2, 9, and 10, $f_i\left((2-i) Bk + (i-1) jR, x\right)$ is identified for all $i = 1, 2$, $j = T, B$ and $k = L, R$.*

Proof. For $i = 1, 2$, $k = L, R$, $j = T, B$, denote by $\bar{P}\left((2-i)Bk + (i-1)jR | x\right)$ the unconditional probabilities $P(\epsilon_i(\tau) + x_i^\tau \beta_i^\tau \geq 0 | x)$. The data does not directly identify this probability but only identifies the conditional probabilities:

$$P\left(a_i = (2-i) B + (i-1) R \middle| a_{3-i} = (2-i) k + (i-1)j, x\right).$$

However, because of Assumption 9,

$$\lim_{||x||\to\infty, x\in\mathcal{T}_{3-i}^{(i-1)j+(2-i)k}}\left[P\left(a_i = (2-i)B + (i-1)R\bigg|a_{3-i} = (2-i)k + (i-1)j, x\right)\right.$$
$$\left. -\bar{P}\left((2-i)Bk + (i-1)jR, x\right)\right] = 0.$$

This implies that $\bar{P}\left((2-i)Bk + (i-1)jR, x\right)$ and hence $f_i\left((2-i)Bk + (i-1)jR, x\right)$ can be identified along the path of $||x|| \to \infty, x \in \mathcal{T}_{3-i}^{(i-1)j+(2-i)k}$. Because the cumulative distribution function of $\epsilon_i(\tau)$ is strictly increasing, the linear index $x_i^\tau \beta_i^\tau$ is identified along this path. Assumption 10 further identifies the coefficient parameters β_i^τ. According to assumption 10, for every $\beta_i^\tau \neq \beta_{i0}^\tau$, there exists a set of x_i^τ with positive probabilitiess such that $x_i^\tau \beta_i^\tau \neq x_i^\tau \beta_{i0}^\tau$, which implies identification of β_i^τ. ∎

A special case of Assumption 9 is when $\epsilon = (\epsilon_i(jk)), i = 1, 2, j = T, B, k = L, R$ has finite support but the support for x_i^τ for $i = 1, 2$ and all τ is either larger or infinite. Denote by $\bar{U}$ an upper bound of the absolute value of the support of $\epsilon_i(jk)$ for all i, k, and j. Then a sufficient condition for Assumption 9 to hold is that for all i and τ, $P\left[x_i^{\tau'}\beta_i^\tau > 2\bar{U}\right] > 0$ and $P\left[x_i^{\tau'}\beta_i^\tau < -2\bar{U}\right] > 0$. Then we do not need the requirement that $||x|| \to \infty$. The sets $\mathcal{T}_i^{j(2-i)+k(i-1)}$ can be defined as $\mathcal{T}_i^{B(2-i)+R(i-1)} = \left\{x : x_i^\tau \beta_i^\tau > 2\bar{U}, \forall\tau\right\}$ and $\mathcal{T}_i^{T(2-i)+L(i-1)} = \left\{x : x_i^\tau \beta_i^\tau < -2\bar{U}, \forall\tau\right\}$. In this case, a sufficient condition for Assumption 10 to hold is that for all $i = 1$, $j = \{T, B\}$, and $i = 2$, $k = \{L, R\}$, the matrices $E\left[xx'|x \in \mathcal{T}_{3-i}^{(i-1)j+(2-i)k}\right]$ are positive definite and finite.

Identifying the Equilibrium Selection Mechanism Given that the deterministic utility components are identified in Theorem 4, the next goal is to identify the equilibrium selection mechanism. The equilibrium selection probabilities are only needed when there are three equilibria, which can be either (TL, BR, mix) or (BL, TR, mix). The mixing probabilities for these two cases are:

$$P_m(R; x, \epsilon) = \frac{f_1(BL, x) + \epsilon_1(BL)}{f_1(BL, x) - f_1(BR, x) + \epsilon_1(BL) - \epsilon_1(BR)}, \quad P_m(L; x, \epsilon) = 1 - P_m(R; x, \epsilon)$$

and

$$P_m(B; x, \epsilon) = \frac{f_2(TR, x) + \epsilon_2(TR)}{f_2(TR, x) - f_2(BR, x) + \epsilon_2(TR) - \epsilon_2(BR)}, \quad P_m(T; x, \epsilon) = 1 - P_m(B; x, \epsilon).$$

In the ideal case where there are no error terms: $\epsilon_1(BL) = \epsilon_1(BR) = \epsilon_2(TR) = \epsilon_2(BR) = 0$, all of $P_m(R), P_m(L), P_m(T)$, and $P_m(B)$ are functions of the known deterministic payoffs. Define the observed equilibrium selection probabilities as: $\rho(TL,x), \rho(BR,x), 1 - \rho(TL,x) - \rho(BR,x)$ in the case of (TL, BR, mix), and as $\rho(BL,x), \rho(TR,x), 1 - \rho(BL,x) - \rho(TR,x)$ in the case of (BL, TR, mix), where the dependence on covariates x is made explicit. Then for those values of x where (TL, BR, mix) is realized,

$$
\begin{aligned}
P(TL|x) &= \rho(TL,x) + (1 - \rho(TL,x) - \rho(BR,x)) P_m(T) P_m(L) \\
P(TR|x) &= (1 - \rho(TL,x) - \rho(BR,x)) P_m(T) P_m(R) \\
P(BL|x) &= (1 - \rho(TL,x) - \rho(BR,x)) P_m(B) P_m(L).
\end{aligned}
$$

These are three equations that identify the two unknown variables $\rho(TL,x)$ and $\rho(BR,x)$. Similarly, for values of x such that (BL, TR, mix) is realized,

$$
\begin{aligned}
P(BL|x) &= \rho(BL,x) + (1 - \rho(BL,x) - \rho(TR,x)) P_m(B) P_m(L) \\
P(BR|x) &= (1 - \rho(BL,x) - \rho(TR,x)) P_m(B) P_m(R) \\
P(TL|x) &= (1 - \rho(BL,x) - \rho(TR,x)) P_m(T) P_m(L),
\end{aligned}
$$

are the three equations that overidentify the two unknown variables $\rho(BL,x)$ and $\rho(TR,x)$.

In the presence of the unobservable error terms ϵ's, additional identification assumptions need to be imposed to isolate the effects of the error terms.

Assumption 11. *The equilibrium selection probabilities depend only on the utility indices:*

$$
\rho(x, \epsilon) = \rho\left(u_i\left((2 - i) Bk + (i - 1) jR, x\right) \forall i, j, k.\right),
$$

where $\rho(x, \epsilon) = [\rho(TL; x, \epsilon), \rho(BR; x, \epsilon), \rho(BL; x, \epsilon), \rho(TR; x, \epsilon)]$. In addition, the equilibrium selection probabilities are scale invariant with respect to the utility indexes. For all $\alpha > 0$, $\rho\left(\alpha u_i\left((2 - i) Bk + (i - 1) jR, x\right), \forall i, j, k\right) = \rho\left(u_i\left((2 - i) Bk + (i - 1) jR, x\right), \forall i, j, k\right).$

This assumption rules out the possibility that $\rho(x, \epsilon)$ might depend on x and ϵ nonseparably, independent of the latent utility indices. It also requires that the equilibrum selection probabilities only depend on the relative but not absolute scales of the latent utilities.

The scale invariance assumption, supplemented by the next support condition on the observables and unobservables, allows us to identify the equilibrium selection probabilities from the variations in the covariates x. In particular, Assumption 11 implies that the determinants for the equilibrium selection probabilities are the same as the determinants for the mixing probabilities. It allows for a rich class of equilibrium selection mechanisms but does

exclude some important ones. For example, it allows for the Pareto efficient equilibrium to be selected with a larger probability and for this probability to depend on the relative efficiency level. This restriction follows from the intuition that if all payoffs were scaled by a constant, we would not expect the distribution over outcomes to change. However, it does not allow this probability to depend on how much more efficient the efficient equilibrium is compared to the inefficient ones in absolute terms. It also rules out equilibrium selection probabilities that depend independently on some of the observed covariates but not on other observed covariates or the error terms. This potentially limits the antitrust implications of the model, because firms concerned with avoiding the suspicions of antitrust investigators might want to choose selection rules which depend on some variables that are easier to communicate but not others.

Assumption 12. *There exists a set* $\mathcal{T}$ *such that for all* $\epsilon > 0$:

$$\lim_{|x| \to \infty, x \in \mathcal{T}} P\left(\frac{f_i\left((2-i)\,Bk + (i-1)\,jR, x\right)}{u_i\left((2-i)\,Bk + (i-1)\,jR, x, \epsilon\right)} > 1 - \eta \right) = 1, \tag{20}$$

for all $i = 1, 2, j = T, B, k = L, R$ *and that for all* $\Lambda = R, B, T, L,$

$$\lim_{|x| \to \infty, x \in \mathcal{T}} P\left(\frac{P_m\left(\Lambda; x, \epsilon\right)}{P_m\left(\Lambda; x, 0\right)} > 1 - \epsilon \right) = 1.$$

This assumption is satisfied if ϵ has finite support but x has infinite support.

Theorem 5. *Under Assumptions 1 to 12, the equilibrium selection probabilities*

$$\rho\left(u_i\left((2-i)\,Bk + (i-1)\,jR, x\right), \forall i, j, k\right)$$

are all identified from the observed choice probabilities.

Proof. Assumptions 1 to 10 identify the payoff functions $f_i\left((2-i)\,Bk + (i-1)\,jR, x\right)$ for all i, j, k. Using Assumption 12, we can approximate the mixing probabilities with arbitrary precision by using larger and larger values of the covariates x. This allows us to recover the equilibrium selection probabilities with arbitrary precision at very large values of the covariates x. By assumption 11, the equilibrium selection probabilities with smaller values of the latent utility indexes are obtained by extrapolation along the remote sections of a ray that emanates from the origin and goes through the latent utility indexes. ∎

4

B.2 Monte Carlo

To demonstrate the performance of our estimator in small samples, we conduct a Monte Carlo experiment using a simple entry game with two players. Each player has the following profit function:

$$\pi_i(a) = 1(a_i = 1)\left\{\beta_1 x_{i1} + \beta_2 x_{2i} + \epsilon_i(a)\right\},$$

with the observable covariates defined by $x_{i1} \sim N(1,1)$ and $x_{2i} = n(a)$, where $N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2, and $n(a)$ is the number of competitors a firm faces given action profile a. The idiosyncratic error term, which is different for each player for each action profile a, is drawn independently from the standard normal distribution. The choice of unit variance in the random shock satisfies the need for a scale normalization, and assigning payoffs of zero to not entering the market satisfies the location normalization. x_{1i} represents variability in profits to firm i from entering that market, and x_{2i} captures the effects of having a competitor. The true payoff parameters are $\beta_1 = 2$ and $\beta_2 = -10$.

The distributions of the covariates were chosen such that when payoffs are evaluated at their means, it is optimal for only one of the two firms to enter the market. Under these circumstances the set of equilibria in this game, denoted by $\mathcal{E}$, has three elements: two pure strategies characterized by one firm or the other entering the market, and one mixed strategy where firms enter with some probability. We specify that the probability of equilibrium $\pi_i \in \mathcal{E}$ being played as:

$$Pr(\pi_i) = \frac{\exp(\theta_1 MIXED_i)}{\sum_{\pi_j \in \mathcal{E}} \exp(\theta_1 MIXED_j)},$$

where $MIXED_i$ is an indicator variable equal to one if equilibrium π_i is in mixed strategies. When $\theta_1 = 0$ one of the three equilibria is picked with equal chance. As that parameter tends to either negative or positive infinity, the mixed strategy is played with probability approaching zero or one, respectively. The true selection parameter is $\theta_1 = 1$.

Our game has three unknown parameters: β_1, β_2, and θ_1. The game generates moments from the probabilities of observing the four possible combinations of entry choices. Only three of these moments are linearly independent, as the probabilities must sum to one, implying that our model is exactly identified. We generate 500 samples of size $n = 25$, 50, 100, 200, and 400 to assess the finite sample properties of our estimator. We set the number of importance games per observation to be equal to the sample size, and generated new importance games for each observation and each replication. Asymptotic errors for each run

Table 9: Monte Carlo Results

	Mean	Median	StdDev	Mean Bias	Median Bias	Mean AD	Median AD	Mean ASE	Median ASE
N = 25									
β_1:	1.982	2.017	0.491	-0.018	0.017	0.346	0.218	0.624	0.595
β_2:	-9.952	-10	0.916	0.048	0	0.643	0.399	2.22	1.249
θ_1:	-1.743	-0.977	3.959	-2.743	-1.977	3.421	2.398	2.644	2.346
N = 50									
β_1:	2.048	2.039	0.401	0.048	0.039	0.265	0.113	0.588	0.574
β_2:	-9.962	-10	0.787	0.038	0	0.534	0.28	1.232	1.148
θ_1:	-0.62	0.526	3.395	-1.62	-0.474	2.04	0.588	2.545	2.124
N = 100									
β_1:	2.057	2.036	0.327	0.057	0.036	0.221	0.095	0.544	0.536
β_2:	-10.016	-10.012	0.663	-0.016	-0.012	0.442	0.213	1.086	1.071
θ_1:	0.14	0.611	2.297	-0.86	-0.389	1.122	0.498	2.218	1.882
N = 200									
β_1:	2.049	2.037	0.245	0.049	0.037	0.166	0.063	0.452	0.434
β_2:	-9.981	-10.003	0.543	0.019	-0.003	0.347	0.114	0.915	0.887
θ_1:	0.664	0.779	0.598	-0.336	-0.221	0.461	0.303	1.548	1.187
N = 400									
β_1:	2.044	2.041	0.224	0.044	0.041	0.166	0.105	0.274	0.261
β_2:	-10.013	-10.026	0.496	-0.013	-0.026	0.362	0.231	0.579	0.566
θ_1:	0.732	0.756	0.377	-0.268	-0.244	0.356	0.281	0.275	0.257

The true parameter vector is $\beta_1 = 2$, $\beta_2 = -10$, and $\theta_1 = 1$. Each sample size was evaluated 500 times. AD = Absolute Deviations. ASE = Asymptotic Standard Errors.

are calculated using the optimal weighting matrix from a two-step GMM procedure. The Laplace-type estimator of Chernozhukov and Hong (2003) is used to recover the parameters. The results of our Monte Carlo are reported in Table 9.

The results are encouraging even in the smallest sample sizes. The payoff parameters are tightly estimated near their true values, while the mixed strategy shifter is estimated with considerably lower precision. There is a distinct downward bias in the estimates of the equilibrium selection parameter that shrinks as the sample size grows. The median bias in all parameters is much better than the mean bias, implying that the mean bias is largely driven by occasional extreme outliers.

The standard deviation of the estimates of all three parameters shrinks as the sample

size increases, as do the mean and median absolute deviations. Significantly, the decrease in the standard deviation for the payoff parameters is close to $\sqrt{n}$, as theory implies. The rate of convergence of the equilibrium selection parameter is much more dramatic as the sample size increases, largely because this parameter is not precisely estimated at smaller sample sizes.

We include the means and medians of the asymptotic standard errors as well as the standard deviations of the Monte Carlo results. The asymptotic errors are usually comparable to the Monte Carlo standard errors, which are calculated by looking at the variance of the parameter estimates across the 500 replications for each sample size. The asymptotic errors tend to overstate the variance of the payoff parameters and understate the variance of the selection parameter relative to the Monte Carlo errors. Their magnitude decreases at an increasing rate as the sample size grows, also similar to the Monte Carlo errors. Overall, our results suggest that the asymptotic errors are good small-sample approximations to their Monte Carlo counterparts.

The precision of the estimated payoff coefficients relative to the equilibrium selection parameter follows from the intuition that the payoff-relevant covariates define the thresholds at which firms are willing to enter a market, and thus enter the likelihood of every observation directly. On the other hand, the equilibrium selection parameter enters the estimating moments in a more subtle manner. This parameter is identified using coordination failures between firms due to mixed strategy equilibrium.

To illustrate, suppose that all payoffs, including idiosyncratic shocks, are observed by the econometrician. For some realizations of the covariates, the model will predict two pure strategies, with one or the other of the firms entering the market, and a single mixed strategy. If the mixed strategy equilibrium is played, there is a chance of either no firms entering the market or both firms entering the market. It is only when these mistakes are observed is the econometrician certain that the mixed strategy is played. Behind this is a subtle and complex relationship between variables, as the probability of observing a mistake is a function of both θ_1, which controls how often a mixed strategy occurs, and the payoffs of the game, which determine the probability of observing a mistake conditional on playing a mixed strategy.

This interplay illustrates a more general point, which is that although the parameters are identified, in small samples the estimation of some parameters may depend on a relatively small subset of outcomes. Note well that this is true even in the extreme case when the payoff functions, including the idiosyncratic shocks, are known with certainty, since the model itself generates probabilistic outcomes through both the equilibrium selection mechanism and the random nature of mixed strategies. In light of this, the results here are very positive, as we

are able to recover estimates of the true parameters with acceptable precision in moderate sample sizes.

There is one caveat to our procedure that researchers have to address in practice. In each Monte Carlo, we know the true parameters of the game, and we are able to generate importance games using these. With real data, of course, these parameters are initially unknown. The importance sampler can generate imprecise parameter estimates with poor initial guesses, so it is necessary to derive starting parameters from a separate source. Below we use a related game of private information to generate initial starting values. Parametric identification is another difficult empirical issue. In the monte carlo example, we are able to calculate the rank condition explicitly at the true parameter values and find it to be nonsingular. This is an overly strong condition for local identification but not sufficient for global identification, which is difficult to obtain.

B.3   Rank Conditions in the Monte Carlo Example

As an illustration, we explicitly analyze the rank condition for identification, which requires that the Jacobian matrix is invertible everywhere, in the context of the Monte Carlo simulation example. The Jacobian matrix is formed by taking derivatives of the outcome probabilities with respect to the parameters:

$$
A = \begin{bmatrix} \frac{\partial P(1,1|x,\beta_1,\beta_1,\theta_1)}{\partial \beta_1} & \frac{\partial P(1,1|x,\beta_1,\beta_1,\theta_1)}{\partial \beta_2} & \frac{\partial P(1,1|x,\beta_1,\beta_1,\theta_1)}{\partial \theta_1} \\ \frac{\partial P(1,0|x,\beta_1,\beta_1,\theta_1)}{\partial \beta_1} & \frac{\partial P(1,0|x,\beta_1,\beta_1,\theta_1)}{\partial \beta_2} & \frac{\partial P(1,0|x,\beta_1,\beta_1,\theta_1)}{\partial \theta_1} \\ \frac{\partial P(0,0|x,\beta_1,\beta_1,\theta_1)}{\partial \beta_1} & \frac{\partial P(0,1|x,\beta_1,\beta_1,\theta_1)}{\partial \beta_2} & \frac{\partial P(1,1|x,\beta_1,\beta_1,\theta_1)}{\partial \theta_1} \end{bmatrix}.
$$

Despite the simple structure of the Monte Carlo setup, the observed outcome probabilities,

$$
P(1,1|x,\beta_1,\beta_1,\theta_1), \ P(1,0|x,\beta_1,\beta_1,\theta_1), \ \text{and} \ P(0,0|x,\beta_1,\beta_1,\theta_1),
$$

are highly complex functions of the model parameters, and the calculation of A is non-trivial.[14] We investigate the nonsingularity of the Jacobian matrix A. Because this is a conditional model, for each vector of parameters $(\beta_1, \beta_2, \theta_1)$ we look for a covariate x that gives a nonsingular Jacobian matrix A. Due to the computation cost of numerically evaluating an integral in the Jacobian matrix, it is prohibitively time consuming to verify nonsingularity for an exhaustive search over the parameter space. Therefore we take a grid

---

[14]We derive these relationships in supplementary material which is available upon request. Calculating the numerical values of A also requires integrating a one-dimensional integral numerically which we evaluate by simulation. Since the derivatives are expressed analytically, we do not need to choose a step size to compute the numerical derivatives.

of size 6 with interval length 0.3 symmetrically around each of the true parameter values. For each combination of the parameter values in the grid, we evaluate the smallest absolute value of the eigenvalues of the Jacobian A at 25 values of the covariates x. The maximum over the covariates x of the smallest absolute eigenvalue of the Jacobian matrix A is recorded and plotted in Figure 1. The histogram of the minimum absolute eigenvalues suggests that they are bounded away from zero and, as a result, that the rank condition is satisfied. While this is not definitive analytical proof that the Jacobian is invertible everywhere (this requires demonstrating strong conditions on the Jacobian, e.g. that it is a P-matrix, which is infeasible given the complexity of our system), as formally required for identification, it does suggest that our problem is identified in a neighborhood of the true parameters.

Figure 1: Distribution of absolute eigenvalues in the Monte Carlo example
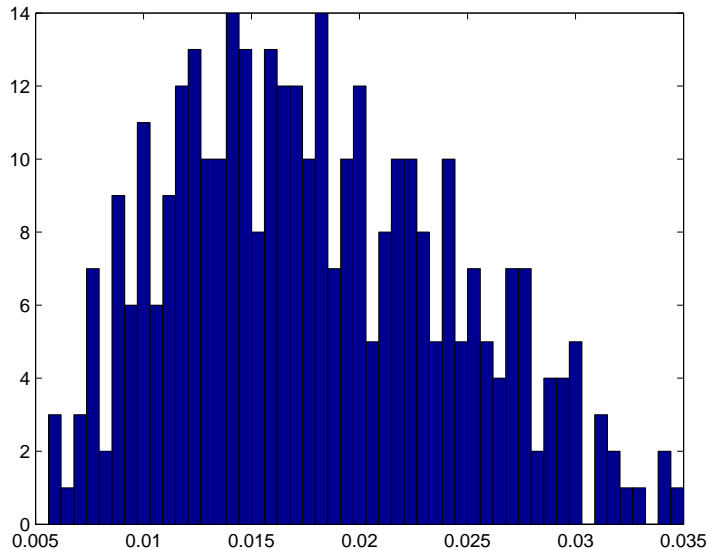
B.4   Maximal Number of Nash Equilibria

When considering the nonparametric identification of the equilibrium selection mechanism, knowledge of generic properties of the set of Nash equilibria proves useful. Here we briefly review results in the literature on the maximum number of equilibria to normal form games of the class considered here.

Solutions to normal form games can be characterized using polynomial equalities and inequalities. Therefore, before considering games, we review some important results on the solutions of a system polynomials. Let $F = \{f_i(x)\}_{i=1}^{n}$ be the system of n polynomials of n

variables, and we are looking for the set of all common roots of this system. A polynomial
$f_i(x) = \sum_{j=1}^{J} a_j \, x_1^{e_1^{ij}} \, x_2^{e_2^{ij}} \ldots x_n^{e_n^{ij}} = \sum_{j=1}^{J} a_j \prod_{k=1}^{n} x_k^{e_k^{ij}}$. The powers e_k^{ij} are integers in general: index
i refers to the equation number, index j refers to the number of monomial in polynomial i,
and index k refers to the specific variable x_k. The points $e^{ij} = \left(e_1^{ij}, \ldots, e_n^{ij}\right)$ form the finite
sets $E_i = (e^{ij}, j = 1, \ldots, J)$ and indicate which monomial terms appear in f_i. For instance,
in the polynomial $f_i(x_1, x_2) = x_1^2 \, x_2^3 + 2x_1^2$, the support set is $E_i = \{(2, 3), \, (2, 0)\}$.

The collection of sets $E = (E_1, \, E_2, \ldots, E_n)$ is called the support of the system of poly-
nomials. The convex hulls $Conv\,(E_i)$ are called Newton polytopes of f_i. For example, the
Newton polytope of the polynomial $f(x_1, x_2) = x_1 \, x_2 + x_1 + x_2 + 1$ is the unit square with
vertices in $(1, 1)$, $(1, 0)$, $(0, 1)$, and $(0, 0)$.

The degree of the polynomial i is $d_i = \max_j \sum_{k=1}^{n} e_k^{ij}$. One of the most important theorems
describing the behavior of zeros of F in the complex space $\mathbb{C}^n$ is Bézout's theorem, which
says that the total number of common complex roots of F is at most $\prod_{i=1}^{n} d_i$. Bézout's
theorem provides an upper bound to the number of common roots in the system, giving little
information on the polynomials that are sparse. In fact, for sparse systems, the number of
common roots of the polynomial system can be significantly less than the bound set by this
theorem. A universal and powerful tool for root counting in case of sparse polynomial is
Bernstein's theorem.

Let P_i be Newton polytopes of equations $f_i(x)$ in the system F defined previously. The
mixed volume of the system of polytopes is defined as:

$$\mathcal{M}\,(P_1, \ldots, P_n) = \sum_{S \subseteq \{1, \ldots, n\}} (-1)^{|S|} \,\mathrm{Vol}\left(\sum_{i \in S} P_i\right), \tag{21}$$

where S are all subsets of $\{1, 2, \ldots, n\}$, $|S|$ is the cardinality (number of elements) of a
particular subset, and $\mathrm{Vol}\,(\cdot)$ is the conventional geometric volume. The sum of the polytopes
is defined for two polytopes P and Q as $P + Q = \{p + q \,|\, p \in P, \, q \in Q\}$.

Theorem 6 (Bernstein). *The number of common roots in the system F is equal to the mixed
volume of the n Newton polytopes of this system.*

This is an extremely powerful result because the mixed volume is easy to compute. A
general problem with complex roots though is that they are not invariant with respect to the
group of polynomial transformations of F. For example, if the polynomial $f(x)$ has degree
d and thus has d distinct complex roots, then the polynomial $f(x)^2$ can have $2d$ distinct
complex roots. This is not the case with the real roots of a system of polynomials and thus
power transformations have no effect on the number of distinct real roots. This is captured

by Khovanskii's theorem, which sets the upper bound on the number of common real roots of the polynomial system which does not depend on the degrees of the equations in the system.

Theorem 7 (Khovanskii). *If m is the number of all monomials in F (or equivalently $m = |E| = \sum_{i=1}^{n} J - i$ - cardinality of the support) and there are n polynomials in F, then the maximum number of real solutions of the system is $2^{\left(\binom{m}{2}\right)} (n+1)^m$.*

In many cases, however, the so-called Kouchnirenko's conjecture holds: if the number of terms in f_i is at most m_i, then the number of isolated real roots is at most $\prod_{i=1}^{n}(m_i - 1)$. This conjecture is violated for some generic (although quite complex) counterexamples.

Consider an arbitrary N-person game where the player i has n_i strategies. Using Lagrangian multiplier techniques, it can be reduced to a system of $n + \sum_i n_i$ polynomial equations with $n + \sum_i n_i$ unknowns. Let $x_k^{(i)}$ denote strategy k of player i, $\xi_{j_1, j_2, \ldots, j_{i-1}, k, j_{i+1}, \ldots, j_N}^{(i)}$ be the payoff function, representing the payoff of player i when she plays the pure strategy k and the other players are playing $j_1, \ldots, j_N$, and $\pi^{(i)}$ be the expected payoff of player i. Let $\lambda_{k0}^{(i)}$ be the Lagrange multiplier for the constraint $x_k^{(i)} \geq 0$, and $\widetilde{\lambda}^{(i)}$ be the Lagrange multiplier for the constraint $\sum_{k=1}^{n_i} x_k^{(i)} = 1$. The Lagrangian for bidder i can be written as:

$$\mathcal{L}^{(i)} = \sum_{k=1}^{n_i} x_k^{(i)} \sum_{j_{-i}} \xi_{j_1, j_2, \ldots, j_{i-1}, k, j_{i+1}, \ldots, j_N}^{(i)} x_{j_1}^{(1)} \ldots x_{j_{i-1}}^{(i-1)} x_{j_{i+1}}^{(i+1)} \ldots x_{j_N}^{(N)}$$
$$- \sum_{k=1}^{n_i} \lambda_{k0}^{(i)} x_k^{(i)} + \widetilde{\lambda}^{(i)} \left(1 - \sum_{k=1}^{n_i} x_k^{(i)} \right).$$

The first order condition for the Lagrangian and the complementary slackness conditions for the multipliers $\lambda_{k0}^{(i)}$ are:

$$\frac{\partial \mathcal{L}^{(i)}}{\partial x_k^{(i)}} = \sum_{j_{-i}} \xi_{j_1, j_2, \ldots, j_{i-1}, k, j_{i+1}, \ldots, j_N}^{(i)} x_{j_1}^{(1)} \ldots x_{j_{i-1}}^{(i-1)} x_{j_{i+1}}^{(i+1)} \ldots x_{j_N}^{(N)} - \lambda_{k0}^{(i)} - \lambda^{(i)} = 0, \qquad (22)$$

$$1 - \sum_{k=1}^{n_i} x_k^{(i)} = 0, \qquad (23)$$

$$x_k^{(i)} \lambda_{k0}^{(i)} = 0, \quad k = 1, \ldots, n_i. \qquad (24)$$

If we multiply the first equation by x_k^i, it reduces to:

$$x_k^{(i)} \sum_{j_1,j_2,\ldots,j_{i-1},j_{i+1},\ldots,j_N} \xi_{j_1,j_2,\ldots,j_{i-1},k,j_{i+1},\ldots,j_N}^{(i)} x_{j_1}^{(1)} \ldots x_{j_{i-1}}^{(i-1)} x_{j_{i+1}}^{(i+1)} \ldots x_{j_N}^{(N)} - x_k^{(i)} \widetilde{\lambda}^{(i)} = 0, \quad (25)$$

$$1 - \sum_{k=1}^{n_i} x_k^{(i)} = 0, \quad (26)$$

$$x_k^{(i)} \lambda_{k0}^{(i)} = 0, \quad k = 1, \ldots, n_i. \quad (27)$$

Summing the first equation over all k, we obtain $\pi^{(i)} = \widetilde{\lambda}^{(i)}$. Then each bidder is characterized by the system of equations:

$$x_k^{(i)} \left(\pi^i - \sum_{j_1,j_2,\ldots,j_{i-1},j_{i+1},\ldots,j_N} \xi_{j_1,j_2,\ldots,j_{i-1},k,j_{i+1},\ldots,j_N}^{(i)} x_{j_1}^{(1)} \ldots x_{j_{i-1}}^{(i-1)} x_{j_{i+1}}^{(i+1)} \ldots x_{j_N}^{(N)} \right) = 0, \quad (28)$$

$$\sum_{k=1}^{n_i} x_k^{(i)} - 1 = 0, \quad k = 1, \ldots, n_i, \quad i = 1, \ldots, N. \quad (29)$$

Thus, for each player we have $n_i + 1$ equations and $n_i + 1$ unknown parameters (n_i mixed strategies and the expected payoff). The individual equation has $\prod_{j \neq i} n_j + 1$ terms (the number of strategies of the other players when the strategy of player i is fixed, plus the expected payoff of player i). In addition, the linear equations limiting the mixed strategies to the simplex have $n_i + 1$ terms each. So in total there are $\sum_i n_i + N$ equations and unknowns. The total number of terms is $\sum_i \prod_{j \neq i} n_j + \sum_i n_i + 2N$. If we consider purely mixed strategies, then $x_k^{(i)} > 0$, and thus the system can be rewritten as:

$$\sum_{j_{-i}} \left(\xi_{j_1,j_2,\ldots,j_{i-1},k,j_{i+1},\ldots,j_N}^{(i)} - \xi_{j_1,j_2,\ldots,j_{i-1},n_i,j_{i+1},\ldots,j_N}^{(i)} \right) x_{j_1}^{(1)} \ldots x_{j_{i-1}}^{(i-1)} x_{j_{i+1}}^{(i+1)} \ldots x_{j_N}^{(N)} = 0, \quad (30)$$

$$k = 1, \ldots, n_i - 1, \quad i = 1, \ldots, N. \quad (31)$$

This system has $n_i - 1$ unknowns for player i and $\sum_i n_i - N$ unknowns in total. The number of terms for each equation is $\prod_{j \neq i} n_j$, according to the number of strategies of the rival players when the strategy of the given player is fixed. The total number of terms is then given by the sum $\sum_i \left(\prod_{j \neq i} n_j \right) (n_i - 1)$.

McKelvey and McLennan (1996) directly apply Bernstein's theorem to the given system of equations and express the number of solutions in terms of the mixed volume of Newton polytopes for the case of totally mixed solutions (the case with possible pure strategies needs specific consideration for each payoff structure). We can also apply Khovanskii's result to this system as follows. First, by Kouchnirenko's conjecture, the maximum number of solutions

to this system is

$$\prod_{i=1}^{N} \left(\prod_{j \neq i} n_j - 1 \right)^{n_i - 1},\tag{32}$$

which gives an approximate upper bound on the number of solutions. An exact application of Khovanskii's formula with $m = \sum_i \left(\prod_{j \neq i} n_j \right) (n_i - 1)$ gives the maximum number of solutions as $2^{\frac{m!}{2(m-2)!}} \left(\sum_i n_i - N + 1 \right)^m$.

For a particular case when k equals the number of strategies, Kouchnirenko's formula gives $\left(k^{N-1} - 1 \right)^{N(k-1)}$ for the number of equilibria, while Khovanskii's bound is

$$2^{\binom{N k^{(N-1)}(k-1)}{2}} \left[N(k-1) + 1 \right]^{N k^{(N-1)}(k-1)}.$$

The number of moments with N players when each player has k strategies is $k^N - 1$. The corresponding numbers of moments are tabulated in Table 11. They are significantly smaller than Kouchnirenko's bounds.

Table 10: Tabulation of Kouchnirenko's formula

	Number of Strategies (k)				
N	2	3	4	5	6
2	3	16	729	65536	9765625
3	27	262144	38443359375	3.65203E+16	1.44884E+23
4	2401	2.08827E+11	3.90919E+21	3.12426E+33	4.45419E+46
5	759375	1.07374E+19	1.25344E+36	8.01109E+55	6.40832E+77
6	887503681	4.03445E+28	1.50578E+54	7.4656E+83	5.2603E+116

Table 11: Tabulation of the Number of Available Moments

	Number of Strategies (k)				
N	2	3	4	5	6
2	3	8	15	24	35
3	7	26	63	124	215
4	15	80	255	624	1295
5	31	242	1023	3124	7775
6	63	728	4095	15624	46655