Correlated Equilibrium and the Estimation of Discrete Games of Complete Information*

Zhou Yang[†]

March 2009

Abstract

In order to understand strategic interactions among firms, we often need to estimate the parameters of static discrete games with complete information. This class of games is difficult to estimate because the possibility of multiple equilibria invalidates the use of methods such as MLE and GMM. We propose a two-step estimator to get around the issue of multiple equilibria by exploiting the fact that all of the Nash equilibria are contained in the set of correlated equilibria. In the first step, we estimate the conditional choice probabilities by which each possible outcome is realized. In the second step, we obtain the bounds on estimates of the parameters by minimizing the average distance between the set of correlated equilibria and the probability distribution obtained in the first step. Compared to previous approaches through which the issue of multiple equilibria has been tackled, our method has two important advantages. First, it explicitly takes into account the possible existence of mixed strategy equilibria. Second, it is computationally easy to implement: due to the inherent linearity of correlated equilibria, we can obtain the bounds estimates by solving a series of linear programming problems.

KEYWORDS: Correlated Equilibrium, Discrete Games, Two-step Estimator

[†]Department of Economics, Virginia Tech. Email: zyang74@vt.edu

1 Introduction

The structural estimation of games has recently gained much attention in the empirical Industrial Organization literature, because many markets involve strategic interactions among firms. In order to make inferences about the fundamental parameters that govern these strategic interactions, researchers often need to econometrically model such interactions as games and then structurally estimate the payoff functions of these games. A simple class of games that is very useful for econometrically modeling and estimating firms' strategic interactions is the group of static discrete games with complete information. Examples of such games include those that involve entry decisions (Bresnahan and Reiss, 1990, 1991; Ciliberto and Tamer, 2007), adoption of technology standard (Augereau, Greenstein and Rysman, 2006), and network effects (Ackerberg and Gowrisankaran, 2006).

In this paper, we provide a method for making inferences about the payoff functions of simultaneous games with complete information and a discrete strategy space. Although they are conceptually straightforward and can be used to model many types of strategic interactions among firms, such games pose nontrivial challenges for identification and estimation. One challenge is that such games often have multiple equilibria, which means that the likelihood function of the observed data is not well defined, rendering the maximum likelihood approach invalid. Furthermore, unlike the complete-information static games with a *continuous* strategy space, where all the equilibria must satisfy some first order conditions from which one can form moment conditions and apply the GMM approach, there are *no* analogous first order conditions that incorporate all the equilibria for a static *discrete* game having multiple equilibria. The other challenge is on the computational front. It is well known that finding all the equilibria of a game having nontrivial scale—which is often the case in empirical applications—is computationally infeasible. Consequently, any inferential approach that involves explicitly solving for the equilibria of a game is computationally burdensome. The method we propose here is an

attempt to make inferences about the payoff structures while explicitly addressing the above difficulties.

The basic idea behind the approach is to exploit the relationship between the Nash equilibria and correlated equilibria of static discrete games with complete information. Such a game can be expanded by allowing the players to communicate with each other. The set of *correlated equilibria* of the original game is then defined as the set of all Nash equilibria of the expanded game. Since strategic independence is a special form of communication among players, the set of Nash equilibria of a static game with complete information is contained within the set of correlated equilibria of this game. Thus, all the Nash equilibria of a game must satisfy whatever restrictions are satisfied by the correlated equilibria. This insight forms the basis for our estimator of static discrete games with complete information.

Our estimator is composed of two steps. In the first step we estimate the conditional choice probabilities by which each possible outcome is realized by parametrically or semiparametrically regressing the observed outcomes on the covariates. In the second step we exploit the relationship between the Nash and correlated equilibria to obtain the confidence regions for the parameters in the payoff functions. Heuristically, this second step goes as follows. We know that the set of correlated equilibria is defined by a system of linear inequalities. Therefore, the conditional choice probabilities that we obtain in the first step must also satisfy this system of linear inequalities. We then obtain the parameter estimates by minimizing the average distance (with a metric to be described more precisely in the section on estimation) between the set of correlated equilibria and the probability distribution that we have obtained in the first step. Depending on whether the model is completely or partially identified, our second stage yields point or set estimates, respectively. For completely identified models, our estimator is consistent and asymptotically normal. For partially identified models, we first obtain the set estimates following a similar procedure as in Manski and Tamer (2002), and then apply the method developed in Chernozhukov, Hong and Tamer (2007) to form the confidence

regions for the parameters of interest.

This paper belongs to a growing literature on estimating static discrete games with complete information. Bresnahan and Reiss (1990, 1991), who first explicitly addressed the issue of multiple equilibria, sought to find the common features among all the equilibria and formed the likelihood function based on the observed common features. Their approach, however, did not allow for mixed strategy equilibria and required the computation of multi-dimensional integrals, the dimensions of which grow quickly as the number of players and/or actions increases, thus making them difficult to implement numerically. In a recent paper, Bajari, Hong and Ryan (2007) explicitly modeled the equilibrium selection mechanism and used MLE to estimate both the parameters in the payoff function and those in the equilibrium selection equation. Their estimator is efficient, but requires one to make assumptions regarding the underlying equilibrium selection mechanism. Ciliberto and Tamer (2007) is the closest to the present paper. They also used a two-step estimator with the first step being the same as ours. In the second step, they formed the lower and upper bounds for the conditional choice probabilities based on the predictions of the model. They then minimized the distance between the choice probabilities obtained in the first step and the set of predicted choice probabilities determined by the lower and upper bounds. Their method, however, does not consider the cross-outcome restrictions on the conditional choice probabilities imposed by economic theory, and is best suitable for handling pure strategy equilibrium only.

Compared to previous approaches to solving the problem of multiple equilibria, our method has several desirable properties, both conceptually and computationally. First, the inequality restrictions on which our estimator is based have a rigorous theoretical foundation and are mathematically easy to characterize. These restrictions are robust to the "true" equilibrium-selection mechanism that underlies the data generating process, and more importantly, can accommodate the possibility that the observed outcomes are generated by different equilibria across different observation units. Second, our method allows both pure strategy and mixed strategy equilibria. Ruling out mixed

strategy equilibria can cause bias in interpreting observed outcomes. Third, our method is computationally easy to implement. The metric we use to measure the distance between the estimated choice probability and the set of predicted choice probabilities that satisfy the restrictions imposed by correlated equilibria is defined in a specific way so that the sample objective function is piecewise linear and convex. As a consequence, we can use linear programming to obtain the set estimates and confidence regions in the second step. This feature is especially desirable when the game has many players and/or actions. Fourth, given the number of parameters, our estimator can provide tighter bounds estimates for games with a larger scale without substantially increasing the computational burden. This is due to the fact that the number of inequality restrictions increases polynomially when the scale of the game becomes large. Hence, the bounds estimates are potentially tighter for games with larger scale, as long as we have enough data to estimate consistently the conditional choice probabilities in the first stage. Our method, however, also has a weakness. In some situations, the set of correlated equilibria of a game can be quite large. Hence, the incentive constraints imposed by correlated equilibria might not be restrictive enough. As a consequence, the bounds estimates and confidence regions might be uninformative. Nonetheless, despite this weakness, we believe that our method provides a new and tractable approach to estimating static discrete games with complete information.

The remainder of the paper is organized as follows. Section 2 sets up the econometric model for static discrete games of complete information. Section 3 introduces the main ideas of imposing restrictions based on correlated equilibria. Section 4 defines the distance metric used in the estimator. Section 5 presents in detail the estimation procedure. Section 6 gives Monte Carlo evidence on how the estimator performs. Section 7 provides extension and concluding remarks. The proofs can be found in the Appendix.

2 Econometric Modeling of Static Discrete Games with Complete Information

2.1 Model Setup

In this section we set up the econometric model of a static discrete game with complete information. Let I denote the set of players, each of whom has a vector of covariates $X_i \in \mathbb{R}^K$ representing their characteristics. We allow X_i to possibly overlap with X_j for $i \neq j$. For example, in the entry game that we will look at in the next subsection, the common market conditions are part of X_i for $\forall i \in I$. We denote $X \equiv \cup_{i \in I} X_i$ as the set of all the covariates for the players, which has distribution F_X and support S_X. Let A_i denote the set of actions available to player i and a_i stand for a generic action that could be taken by player i. Furthermore, $\epsilon_i = [\epsilon_i(a_i)]_{a_i \in A_i}$ denotes the vector of payoff shocks for player i. Finally, let $a \equiv (a_1, ..., a_{\|I\|})$ denote a generic element in the set of all possible action profiles, $A \equiv \times_{i \in I} A_i$. In the following sections of the paper, we will sometimes term an action profile a as an *outcome*[1] of the game. Since the number of action profiles is finite, *i.e.*, $\|A\| < \infty$, we can enumerate all of the possible action profiles as $1, 2, ..., l, ..., \|A\|$. We let $a^{(l)} = \left[a_i^{(l)} \right]_{i \in I}$ denote the lth profile of actions in set A.

Player $i's$ payoff function $u_i(\cdot)$ is given as:

$$u_i(a, X_i, \theta, \epsilon_i) = f_i(a, X_i, \theta) + \epsilon_i(a_i), \tag{1}$$

where $\theta \in \mathbb{R}^J$ is the vector of parameters to be estimated. In the above payoff specification, the deterministic part is a function of the strategy profile, a, $i's$ characteristics, X_i, and the parameter vector, θ; the preference shock is assumed to depend only on $i's$ own strategy, a_i. In most empirical applications (see, for example, Bresnahan and Reiss,

---

[1]The term *outcome* is sometimes used in a slightly different way to refer to the realized payoffs of the players.

1990, 1991; Ciliberto and Tamer, 2007), $f_i(a, X_i, \theta)$ is assumed to be linear in θ for tractability. Note that this specification can accommodate observed heterogeneity, since the term of characteristics, X_i, is player specific. Our specification is similar to that in Ciliberto and Tamer (2007) in this regard, and more general than that in Bresnahan and Reiss (1990, 1991) and Mazzeo (2002).

We assume that the entire structure of the game is common knowledge among the players. That is, each player's characteristics and preference shock vector, $\epsilon_i \equiv (\epsilon_i(1), ..., \epsilon_i(\|A_i\|))$, as well as the parameter vector, θ, the functional form, $f_i(\cdot)$, are common knowledge among all the players. Furthermore, we assume $\epsilon_i(a_i)$ is $i.i.d.$ across i and a_i, and has distribution function $G(\cdot)$ and support S_ϵ. As we will discuss below, the econometrician cannot observe the realization of ϵ_i despite his perfect knowledge of its distribution. Note here that we assume $\epsilon_i(a_i)$ depends only on player $i's$ own strategy, but not on other players' strategies. Hence, this assumption says that the random vector ϵ_i is an inherent attribute of player i and does not vary with respect to the strategies that might be taken by the other players. Many applications coincide with this assumption. For instance, in Seim (2006), the profitability shock depends solely upon a video store's own choice of location.

We denote the parameterized game we are considering as $\Gamma(X, \theta, \epsilon)$. Player $i's$ *strategy* in $\Gamma(X, \theta, \epsilon)$ is a probability distribution $\sigma_i = (\sigma_i(a_i))_{a_i \in A_i} \in \triangle(A_i)$, where $\sigma_i(a_i)$ is player $i's$ probability of choosing action a_i, and $\triangle(A_i)$ denotes the set of all probability distributions over A_i. Since each player's action set A_i is finite, σ_i is a probability vector, each element of which corresponds to the probability of choosing a certain action. Throughout the paper, we will assume that the players' behavior accords with the Nash equilibrium. Due to the complete information assumption, the equilibrium strategies not only depend on a player's own econometric error, but also on the errors of all the other players. This fact makes the estimation of θ a much more difficult problem than it would otherwise be if the $\epsilon_i's$ were private information among the players (Bajari, Hong and Ryan, 2007). In the next subsection, we present the simple 2×2 entry game as

an example to illustrate the econometric modeling of static discrete games of complete information.

2.2 Example: Simultaneous Entry Game

Two firms consider whether or not to enter a market. Each of the two firms $i \in \{1, 2\}$ has to choose an action $a_i \in \{1, 0\} \equiv A_i$, where $a_i = 1$ denotes that i chooses "enter" and $a_i = 0$ means that i chooses "not enter." We normalize each firm's deterministic payoff to be zero if it chooses not to enter. We include market and firm specific characteristics X_i in firm $i's$ profit functions in order to capture the impact of these variables on its profits. Furthermore, we include a spillover effect measuring the impact of a firm's entry decision on the other firms' profits. We can write firm $i's$ payoff function as follows:

$$u_i \left(a, X_i, \theta, \epsilon_i \right) = \left[X_i \cdot \beta - \delta \cdot 1 \left\{ a_{-i} = 1 \right\} + \epsilon_i \left(a_i \right) \right] \cdot 1 \left\{ a_i = 1 \right\},$$

where $1 \left\{ \cdot \right\}$ is an indicator function. The parameter vector is $\theta = \left(\beta, \delta \right)$, where β captures the effects of the covariates on the profits and δ measures the spillover effect of entry between the firms. We expect $\delta > 0$. Finally, we can represent this entry game using the following matrix form:

	Enter	Not Enter
Enter	$X_1 \cdot \beta - \delta + \epsilon_1 \left(1 \right), X_2 \cdot \beta - \delta + \epsilon_2 \left(1 \right)$	$X_1 \cdot \beta + \epsilon_1 \left(1 \right), 0$
Not Enter	$0, X_2 \cdot \beta + \epsilon_2 \left(1 \right)$	$0, 0$

Intuitively, corresponding to each realization of $X = \left(X_1, X_2 \right)$ and of $\epsilon = \left(\epsilon_1, \epsilon_2 \right)$ we have a payoff matrix that describes the strategic interactions between the firms. In the following section, we will also use this simple example to illustrate the incentive restrictions imposed by a correlated equilibrium.

2.3 Formulation of the Inferential Problem

To formally describe the inferential problem facing the econometrician, let $\mu : S_X \times \Theta \times S_\epsilon \to \triangle(A)$ be an *equilibrium function* such that for each $(x, \theta, \epsilon) \in S_X \times \Theta \times S_\epsilon$, $\mu(x, \theta, \epsilon) \in \triangle(A)$ is *an* equilibrium of $\Gamma(x, \theta, \epsilon)$. Furthermore, let $U : S_X \times \Theta \times S_\epsilon \Rightarrow \triangle(A)$ be an *equilibrium correspondence* such that for each $(x, \theta, \epsilon) \in S_X \times \Theta \times S_\epsilon$, $U(x, \theta, \epsilon) \subseteq \triangle(A)$ constitutes *all* the equilibria of $\Gamma(x, \theta, \epsilon)$. Finally, denote by $\mu \in U$ if $\mu(x, \theta, \epsilon) \in U(x, \theta, \epsilon)$, $\forall (x, \theta, \epsilon) \in S_X \times \Theta \times S_\epsilon$. That is, we say that an equilibrium function μ belongs to an an equilibrium correspondence U if and only if the graph of μ is contained in the graph of U.

For the inferential problem, the econometrician has access to cross-sectional data $\{x^n\}_{n=1}^N$, which are realizations of an *i.i.d.* sample of covariates $\{X^n\}_{n=1}^N$ with each X^n being drawn independently from the same distribution $F_X(\cdot)$. For each observation unit n, she also observes a^n, an action profile generated by *an* equilibrium of the game $\Gamma(x^n, \theta, \epsilon^n)$, whose payoff structure is specified as in (1). Furthermore, let $\{\lambda^n\}_{n=1}^N$ be the true equilibrium selection mechanism, where for each observation unit n, λ^n is a probability distribution over the correspondence U, such that an equilibrium function $\mu \in U$ is chosen with probability $\lambda^n(\mu)$. Suppose an equilibrium function $\mu^n \in U$ is selected for observation unit n, then an outcome a^n of the game $\Gamma(x^n, \theta, \epsilon^n)$ will be realized according to the equilibrium $\mu^n(x^n, \theta, \epsilon^n)$. Due to the multiplicity of equilibria, the probability distribution λ^n over U is non-degenerate. The objective of the econometrician is to use the data $\{x^n, a^n\}_{n=1}^N$ as well as the model to make inferences about the structural parameter θ.

If the econometrician knew the true equilibrium selection mechanism $\{\lambda^n\}_{n=1}^N$, the multiplicity of equilibria would not cause any difficulty for making inferences about θ, since the likelihood of the data would still be well defined and the maximum likelihood approach could still be used to estimate θ.[2] To get around the issue that the true equilibrium selection mechanism is in fact unknown to the econometrician, Bajari, Hong and Ryan

---

[2] To see this, note that under the equilibrium selection mechanism $\{\lambda^n\}_{n=1}^N$, the log likelihood of

(2007) assume λ^n takes a parametric form, which enables them to form the likelihood function of the observed data and apply MLE to estimate both the structural parameter θ and the parameters in the equilibrium selection mechanism. As will become clear, our estimation approach does not impose any parametric or nonparametric assumption on λ^n, except that the true equilibrium selection mechanism is assumed to be the same across different observations. That is, we assume $\lambda^n \equiv \lambda, \forall n$, but do not impose any structure on λ. To put it intuitively, we allow different equilibria to "bounce around" in the data but follow the same pattern across different observations.

3 Inference via Correlated Equilibrium: The Main Ideas

In this section we formally introduce our approach to estimating static discrete games with complete information. We provide the general framework here and discuss the implementation in the next two sections.

The basic idea of the estimator we propose is based on the following insight of Bresnahan and Reiss (1990, 1991) and Tamer (2003): in games with multiple equilibria, theory provides inequality restrictions on the probabilities that certain outcomes can occur. To understand this point, let us consider the discrete choice models (logit models, for example). In these models, theory provides an exact prediction of the probabilities of the outcomes, namely, the choice probabilities. This observation carries over to the

---

observing the data $\{a^n, x^n\}_{n=1}^N$ is:

$$\pounds = \sum_{n=1}^N \log \left[\Pr \left(a^n | x^n, \theta \right) \right] + \sum_{n=1}^N \log \left[f_X \left(x^n \right) \right],$$

where

$$\Pr \left(a^n | x^n, \theta \right) = \int_U \left\{ \int_{S_\epsilon} \left[\mu \left(x^n, \theta, \epsilon \right) \left(a^n \right) \right] P \left(d\epsilon \right) \right\} \lambda^n \left(d\mu \right).$$

games with unique equilibrium, where theory provides a unique prediction of the choice probabilities as well. However, in games with multiple equilibria, the uniqueness of the choice probabilities breaks down. Nevertheless, as Tamer (2003) argues, equilibrium theory can still provide the lower and upper bounds on the choice probabilities. He then uses these inequality restrictions on regressions to find the regions in which the parameters of interest lie.

Our approach is similar to Tamer (2003) and Ciliberto and Tamer (2007) in that we also seek to put restrictions on the choice probabilities using the theoretical predictions of the model. However, Tamer (2003) and Ciliberto and Tamer (2007) mainly considers pure strategy equilibria only. Furthermore, their method relies on simulation to find the lower and upper bounds, which is computationally intensive for games with many players and/or many strategies. The method we develop here exploits the relationship between Nash and correlated equilibrium to impose restrictions on the choice probabilities. Roughly, theory says that all the Nash equilibria of a game must be contained in the set of correlated equilibria of that game. Hence, even though a game might have multiple equilibria, all of them must satisfy whatever conditions a correlated equilibrium satisfies. Furthermore, the restrictions we impose apply to both pure and mixed strategy equilibria. Before we use this relationship to derive our theoretical restrictions, we give a brief review of the solution concept of correlated equilibria.

3.1 Correlated Equilibria of Static Discrete Games with Complete Information

In this subsection we give a brief presentation of the concept of correlated equilibria for static discrete games of complete information. A probability distribution $\mu = (\mu(a))_{a \in A}$ $\in \triangle(A)$, where $\mu(a)$ denotes the probability that a profile of actions a is chosen, is a

correlated equilibrium if and only if

$$\sum_{a_{-i} \in A_{-i}} \mu(a) \left[u_i(a_{-i}, a_i) - u_i(a_{-i}, d_i) \right] \geq 0, \quad \forall i \in I, \forall a_i \in A_i, \forall d_i \in A_i. \qquad (2)$$

First, note that there are a total of $\sum_{i \in I} \|A_i\| \cdot (\|A_i\| - 1)$ inequalities, all of which are *linear* in the payoffs. This fact will facilitate the application of linear programming in the estimation. Second, to understand this definition, imagine a mediator who randomly draws an action profile, a, from the set A with probability $\mu(a)$. The mediator then tells each player *privately* to play the action as specified in the action profile a. Based on the probability distribution μ and the instruction given to him, the player can calculate his expected payoff from obeying or disobeying the instruction. Hence, a correlated equilibrium is defined as a probability distribution such that no player has an incentive to deviate from the instruction given to him.

The following theorem characterizes the relationship between the correlated and Nash equilibria of a strategic-form game.

Theorem 1 *Given any finite strategic-form game Γ, suppose $\sigma = (\sigma_i)_{i \in I}$ is a Nash equilibrium of Γ, where $\sigma_i = (\sigma_i(a_i))_{a_i \in A_i} \in \triangle(A_i)$ denotes the probability distribution according to which player i chooses an action from A_i. Define the probability distribution $\mu = (\mu(a))_{a \in A} \in \triangle(A)$ where $\mu(a) = \prod_{i \in I} \sigma_i(a_i)$, then μ is a correlated equilibrium of Γ.*

This theorem suggests that a Nash equilibrium of the original game without communication remains a Nash equilibrium of the augmented game where players are allowed to communicate with each other. The proof and further explanation of this theorem can be found in any standard advanced game theory text such as Myerson (1991).

3.2 Inequality Restrictions on Conditional Choice Probabilities via Correlated Equilibria

In this subsection, we derive the inequality restrictions imposed by correlated equilibrium on the conditional choice probabilities. We first write down the incentive constraints implied by correlated equilibria for a game $\Gamma(X, \theta, \epsilon)$. Based on these incentive constraints, we then derive a set of inequality restrictions on conditional choice probabilities. These inequality restrictions do not involve the ϵ and will form the basis for our estimation.

Fix a game $\Gamma(X, \theta, \epsilon)$. Let player $i's$ Nash equilibrium strategy in this game be $\sigma_i(X, \theta, \epsilon) = (\sigma_i(X, \theta, \epsilon)(a_i))_{a_i \in A_i} \in \triangle(A_i)$. Hence, in our notation, $\sigma_i(X, \theta, \epsilon)$ is a vector of probabilities with the element corresponding to the action a_i being $\sigma_i(X, \theta, \epsilon)(a_i)$. From now on, if confusion does not arise, we will suppress the dependence of σ_i on (X, θ, ϵ) to make our notation less cumbersome. However, the readers should keep in mind the dependence of the equilibrium strategy on the specific game being played. Suppose $\sigma = (\sigma_i)_{i \in I}$ is an equilibrium of the game $\Gamma(X, \theta, \epsilon)$. Let $\mu = (\mu(a))_{a \in A} \in \triangle(A)$ denote the induced probability distribution over the set of action profiles A. It then follows that $\mu(a) = \prod_{i \in I} \sigma_i(a_i)$. Thus, each equilibrium σ uniquely determines a probability distribution over the set of action profiles.

Since the game $\Gamma(X, \theta, \epsilon)$ may have multiple equilibria, there are possibly multiple equilibrium probability distributions corresponding to the game. Let $CE(X, \theta, \epsilon)$ denote the set of correlated equilibria for the game $\Gamma(X, \theta, \epsilon)$. Suppose the data on the observed action profiles are generated by a specific equilibrium distribution $\mu(X, \theta, \epsilon) = (\mu(X, \theta, \epsilon)(a))_{a \in A} \in \triangle(A)$. It then follows that $\mu(X, \theta, \epsilon) \in CE(X, \theta, \epsilon)$ since $\mu(X, \theta, \epsilon)$ is a Nash equilibrium, and by theorem 1, all the Nash equilibria are contained in the set of correlated equilibria. In the following we will derive the population restrictions on (X, θ) from the relationship $\mu(X, \theta, \epsilon) \in CE(X, \theta, \epsilon)$.

Because it belongs to the set of correlated equilibria of $\Gamma(X, \theta, \epsilon)$, $\mu(X, \theta, \epsilon)$ satisfies the

incentive constraints in (2) required by the definition of correlated equilibria. Formally, for $\forall i \in N, \forall a_i \in A_i, \forall d_i \in A_i,$

$$\sum_{a_{-i} \in A_{-i}} [\mu(X, \theta, \epsilon)(a)] u_i(a_{-i}, a_i, X_i, \theta, \epsilon_i) \geq \sum_{a_{-i} \in A_{-i}} [\mu(X, \theta, \epsilon)(a)] u_i(a_{-i}, d_i, X_i, \theta, \epsilon_i),$$

where the left-hand side is player $i's$ expected payoff for following the instruction while the right hand-side is player $i's$ expected payoff from deviating to play d_i.

Note that the expectations on both sides are taken with respect to the probability distribution $\mu(X, \theta, \epsilon)$. By making implicit the dependence of the vector μ on (X, θ, ϵ), we can rewrite the above incentive constraints as

$$\sum_{a_{-i} \in A_{-i}} \mu(a) u_i(a_{-i}, a_i, X_i, \theta, \epsilon_i) \geq \sum_{a_{-i} \in A_{-i}} \mu(a) u_i(a_{-i}, d_i, X_i, \theta, \epsilon_i).$$

By substituting the specification of the payoff function (Equation 1) and rearranging terms, we have that for $\forall i \in I, \forall a_i \in A_i, \forall d_i \in A_i,$

$$\sum_{a_{-i} \in A_{-i}} \mu(a) [f_i(a_{-i}, a_i, X_i, \theta) - f_i(a_{-i}, d_i, X_i, \theta)] \geq \sum_{a_{-i} \in A_{-i}} \mu(a) [\epsilon_i(d_i) - \epsilon_i(a_i)]. \quad (3)$$

The left-hand side is player $i's$ expected *loss* in the deterministic part of the payoff by deviating from the instructed action a_i to another action d_i. The right-hand side is his expected *gain* in the random part of the payoff from such deviation. Thus the above incentive constraints stipulate that the expected loss in the deterministic part of the payoff must be greater than the expected gain in the random part of the payoff.

The expectation of the left-hand side of inequality (3) conditional on the covariates X is:

$$E \left\{ \sum_{a_{-i} \in A_{-i}} \mu(a) [f_i(a_{-i}, a_i, X_i, \theta) - f_i(a_{-i}, d_i, X_i, \theta)] | X \right\}$$
$$= \sum_{a_{-i} \in A_{-i}} E[\mu(a) | X] \cdot [f_i(a_{-i}, a_i, X_i, \theta) - f_i(a_{-i}, d_i, X_i, \theta)]$$
$$\equiv \sum_{a_{-i} \in A_{-i}} \Pr(a | X, \mu) \cdot [f_i(a_{-i}, a_i, X_i, \theta) - f_i(a_{-i}, d_i, X_i, \theta)],$$

where $\Pr(a|X,\mu)$ denotes the *conditional* choice probability that the action profile a is chosen under equilibrium distribution μ.

Similarly, the expectation of the right-hand side of inequality (3) conditional on the covariates X is:

$$E\left\{\sum_{a_{-i}\in A_{-i}} \mu(a)\left[\epsilon_i(d_i) - \epsilon_i(a_i)\right]|X\right\} = E\left\{\sigma_i(a_i)\left[\epsilon_i(d_i) - \epsilon_i(a_i)\right]|X\right\}, \qquad (4)$$

where the equality follows from the fact that $\mu(a) = \prod_{i\in I}\sigma_i(a_i)$. Since player $i's$ equilibrium strategy $\sigma_i(a_i)$ depends on its own vector of payoff shocks, ϵ_i, (4) in general is not equal to zero. We can, however, obtain its lower bound:

$$E\left\{\sigma_i(a_i)\left[\epsilon_i(d_i) - \epsilon_i(a_i)\right]|X\right\} \geq -E\left\{|\epsilon_i(d_i) - \epsilon_i(a_i)||X\right\} = -E\left\{|\epsilon_i(d_i) - \epsilon_i(a_i)|\right\},$$

$$(5)$$

where the weak inequality follows from the fact that $\sigma_i(a_i) \in [0,1]$, and the equality from the assumption that payoff shocks are independent of the covariates.

Since $\epsilon_i(a_i)$ is *i.i.d.* across i and a_i, and has a known distribution $G(\cdot)$, then $E_G\left\{|\varepsilon_i(d_i) - \varepsilon_i(a_i)|\right\}$ is also known. Let ℓ_G denote $E_G\left\{|\varepsilon_i(d_i) - \varepsilon_i(a_i)|\right\}$. Then (3) implies the following system of $\sum_{i\in I}\|A_i\|\cdot(\|A_i\| - 1)$ inequalities:

$$\sum_{a_{-i}\in A_{-i}}\Pr(a|X,\mu)\cdot\left[f_i(a_{-i}, a_i, X_i, \theta) - f_i(a_{-i}, d_i, X_i, \theta)\right] \geq -\ell_G,$$
$$\forall i \in I, \forall a_i \in A_i, \forall d_i \in A_i.$$

This system of inequalities imposes restrictions on $\Pr(a|X,\mu)$, the conditional choice probability under a certain equilibrium distribution μ. Suppose the true equilibrium selection mechanism (unobserved to the econometrician) is such that μ is chosen from the correspondence U with probability $\lambda(\mu)$. Since $\Pr(a|X) = \int_U \Pr(a|X,\mu)\lambda(d\mu)$, it immediately follows that the conditional choice probability $\Pr(a|X)$ satisfies the same system of inequalities, *i.e.*,

$$\sum_{a_{-i}\in A_{-i}}\Pr(a|X)\cdot\left[f_i(a_{-i}, a_i, X_i, \theta) - f_i(a_{-i}, d_i, X_i, \theta)\right] \geq -\ell_G,$$
$$\forall i \in I, \forall a_i \in A_i, \forall d_i \in A_i. \qquad (6)$$

The inequalities in (6) are linear in the payoffs due to the linearity of correlated equilibria. In many empirical applications, the payoff function $f_i(a_{-i}, a_i, X_i, \theta)$ is specified to be linear in parameter θ. Hence, the above system is linear in θ. As will become clear later when we present the estimation procedure, this feature proves to be very important in reducing the computational burden in our set estimates. We summarize the discussion above in the following proposition:

Proposition 1 *Consider a parameterized static discrete game of complete information* $\Gamma(X, \theta, \epsilon)$. *Suppose the payoff structure of this parameterized game is specified as in* (1) *and the econometric error* $\epsilon_i(a_i)$ *is i.i.d.* $\forall i \in I, \forall a_i \in A_i$, *and has a known distribution* $G(\cdot)$. *Furthermore, let* $\ell_G = \mathrm{E}_G\{|\epsilon_i(d_i) - \epsilon_i(a_i)|\}$. *Then the conditional choice probability* $\Pr(a|X)$ *satisfies the system of inequalities in* (6).

Remark 1. The restrictions on conditional choice probabilities imposed by (6) allow both pure and mixed strategies. ∎

Relationship (6) is a population restriction. Under the formulation of the inferential problem in Section 2.3, suppose the true equilibrium selection mechanism is the same across different observations, *i.e.*, $\lambda^n = \lambda$. If the econometrician has cross-sectional data of $\{(a^n, X^n)\}_{n=1}^N$, the sample restriction associated with (6) is that for each $n = 1, ..., N$,

$$\sum_{a_{-i} \in A_{-i}} \Pr(a^n|X^n) \cdot \left[f_i\left(a_{-i}^n, a_i^n, X_i^n, \theta\right) - f_i\left(a_{-i}^n, d_i^n, X_i^n, \theta\right) \right] \geq -\ell_G \quad , \qquad (7)$$
$$\forall i \in I, \forall a_i \in A_i, \forall d_i \in A_i$$

where $\Pr(a^n|X^n) = \int_U \Pr(a^n|X^n, \mu)\, \lambda(d\mu)$. This system of restrictions is a generalization of standard statistical models in the sense that the sample relationship here does not take the form of equalities. Also, one subtlety arises when we apply the sample relationship (7) to the inferential problem facing the econometrician. Since we do not impose any assumption on λ, the conditional choice probability $\Pr(a^n|X^n)$ is unknown. In a typical two-stage approach, researchers would use $\{(a^n, X^n)\}_{n=1}^N$ to estimate $\Pr(a^n|X^n)$.

This paper also uses a two-stage estimation approach, which is discussed in detail in the section on estimation.

Before moving on to the next section on identification and estimation, we look at the above population restrictions in the entry example from earlier. In this example, there are two players each of whom has two available actions. Hence, there are 4 inequalities. The true data generating process implies the following 4 conditional choice probabilities: $\Pr(E, E|X)$, $\Pr(E, N|X)$, $\Pr(N, E|X)$, and $\Pr(N, N|X)$, where $\Pr(E, E|X)$ means the conditional choice probability that both firms enter and $\Pr(E, N|X)$ is the conditional choice probability that firm 1 enters but firm 2 does not, etc. Finally, we assume $\epsilon_i(a_i) \sim N(0, 1), \forall i \in I, \forall a_i \in A_i$. Hence $\ell_G \approx 1.80$. The concept of correlated equilibrium requires that when a firm is instructed to play "enter" or "not enter," it does not have incentives to deviate. Therefore,

$$\Pr(E, E|X) \cdot (\beta \cdot X^1 - \delta) + \Pr(E, N|X) \cdot (\beta \cdot X^1) + 1.80 \geq 0$$

$$-\Pr(N, E|X) \cdot (\beta \cdot X^1 - \delta) - \Pr(N, N|X) \cdot (\beta \cdot X^1) + 1.80 \geq 0$$

$$\Pr(E, E|X) \cdot (\beta \cdot X^2 - \delta) + \Pr(N, E|X) \cdot (\beta \cdot X^2) + 1.80 \geq 0$$

$$-\Pr(E, N|X) \cdot (\beta \cdot X^2 - \delta) - \Pr(N, N|X) \cdot (\beta \cdot X^2) + 1.80 \geq 0.$$

4 Identification and Estimation

As we have argued above, economic theory imposes the restrictions described by (6). Let the vector $y = [y_1, ..., y_{\|A\|}]$ list all the possible outcomes. The vector

$$\Pr(y|X) = [\Pr(y_1|X), ..., \Pr(y_l|X), ..., \Pr(y_{\|A\|}|X)]$$

then provides the conditional choice probability for each possible outcome. Furthermore, for $\forall a_i \in A_i$, define a mapping $\Upsilon^{a_i} : \triangle(A) \to \triangle(A_{-i})$ as follows. Given any probability distribution $\nu \in \triangle(A)$, let $\Upsilon^{a_i}(\nu) = [\nu(a_i, a_{-i})]_{a_{-i} \in A_{-i}}$. That is, $\Upsilon^{a_i}(\nu)$

is a subvector of ν that consists of all the elements in ν that correspond to those action profiles having a_i as their *ith* component. As an example to illustrate the meaning of mapping $\Upsilon^{a_i}(\cdot)$, suppose we have a two-player game, where $A_1 = \{T, B\}$ and $A_2 = \{L, R\}$. Let vector $\nu = [\nu(T, L), \nu(T, R), \nu(B, L), \nu(B, R)]$. If we take $a_1 = T$, then $\Upsilon^T(v) = [\nu(T, L), \nu(T, R)]$. Similarly, $\Upsilon^B(v) = [\nu(B, L), \nu(B, R)]$, etc.

Let

$$F_i(a_i, X_i, \theta) \equiv [f_i(a_{-i}, a_i, X_i, \theta)]_{a_{-i} \in A_{-i}}.$$

We can then rewrite condition (6) in the following vector form:

$$\Upsilon^{a_i}[\Pr(y|X)] \cdot [F_i(a_i, X_i, \theta) - F_i(d_i, X_i, \theta)] + \ell_G \geq 0, \quad \forall i \in I, \forall a_i \in A_i, \forall d_i \in A_i. \quad (8)$$

Definition 1. Given covariates X and a parameter θ, define $CE(X, \theta)$ to be the set of conditional choice probabilities that satisfy the system of inequalities (8):

$$CE(X, \theta) = \{\nu \in \triangle(A) | \Upsilon^{a_i}[\nu] \cdot [F_i(a_i, X_i, \theta) - F_i(d_i, X_i, \theta)] + \ell_G \geq 0, \quad \forall i \in I, \forall a_i, d_i \in A_i\}$$

The identified features of the model are the parameter values that satisfy the following restrictions:

$$\Theta_I = \{\theta \in \Theta : \Pr(y|X) \in CE(X, \theta), a.s.\}.$$

Intuitively, the identified set of parameters Θ_I represents the set of parameter vectors that satisfy the restrictions imposed by theory for all X almost surely; in other words, the economic models corresponding to Θ_I are consistent with the data almost everywhere. Depending on whether Θ_I is a singleton or a set, our econometric model of the game is *just* or *under* identified.

In order to define the identified set Θ_I via minimizing a population objective function, we need to formally define the notion of distance between a probability vector and a set of probability vectors. The choice of distance metric depends on the nature of the problem at hand. Ciliberto and Tamer (2007) chose the minimum distance between a

17

point and the points in the set as the distance metric between a point and a set. In this paper, we define the distance metric in a way that facilitates the transformation of the optimization problem into a linear program. Specifically, we measure the distance between $\Pr(y|X)$ and $CE(X,\theta)$ in the following way:

$CE(X,\theta)$ is defined in terms of a total of $\sum_{i \in I} \|A_i\| \cdot (\|A_i\| - 1)$ linear inequalities, which can be indexed by a combination of player and action pair (i, a_i, d_i). Roughly, we check whether $\Pr(y|X)$ violates these inequalities, one by one. Whenever $\Pr(y|X)$ violates an inequality, we "penalize" it by the extent to which this inequality is violated. Formally, if we denote the distance between $\Pr(y|X)$ and $CE(X,\theta)$ as $d[\Pr(y|X), CE(X,\theta)]$, then

$$
\begin{aligned}
& d\left[\Pr(y|X), CE(X,\theta)\right] \\
= & \sum_{(i,a_i,d_i)} \left| \min\left\{0, \Upsilon^{a_i}\left[\Pr(y|X)\right] \cdot \left[F_i(a_i, X_i, \theta) - F_i(d_i, X_i, \theta)\right] + \ell_G\right\} \right|.
\end{aligned}
\tag{9}
$$

In expression (9), the sign of the second term in the minimum function indicates whether the associated restriction is violated. When it is positive, the restriction is not violated and the corresponding penalty is 0. When it is negative, the restriction is violated and the penalty is the distance of term from 0. It is straightforward to see that $d[\Pr(y|X), CE(X,\theta)] = 0$ if and only if $\Pr(y|X) \in CE(X,\theta)$. It is important to note that $d[\Pr(y|X), CE(X,\theta)]$ is piecewise linear and convex in θ if the deterministic part of the payoff function $f_i(a, X_i, \theta)$ is specified to be linear in θ. This piecewise linearity and convexity proves to be a crucial property that we will exploit in computing our estimates. Details appear in the next section.

Having defined the distance metric d, we are now ready to express the identified set Θ_I via a minimization problem. Define a function $D : \Theta \to \mathbb{R}^+$ such that

$$
D(\theta) = \int_{S_X} d\left[\Pr(y|X), CE(X,\theta)\right] dF_X
$$

18

where F_X is the distribution of the covariates X. Therefore, Θ_I can be rewritten as

$$\Theta_I = \arg\min_{\theta \in \Theta} D\left(\theta\right). \tag{10}$$

5   Estimation Approach

In this section, we provide detailed procedures for implementing the two-step estimator outlined in the last section. In the first stage, we consistently estimate the conditional choice probabilities $\Pr\left(y|X\right)$ using a multinomial logit model. We then recover the structural parameters in the second stage. For partially identified models, we seek the bounds estimates of the true parameters by applying recently developed techniques in the literature on set estimation of partially identified models. We provide both the set estimates and confidence regions for the identified set of parameters. As stated in proposition 3 below, our method provides a consistent (in the Hausdorff sense) set estimate for partially identified models. We then proceed to consider the case in which the model is completely identified. For these models, our two-step approach reduces to a sharp two-step minimum distance estimator. We also provide the asymptotic theory of our estimator when the model is completely identified: under suitable regularity conditions, our two-step approach gives a consistent and asymptotically normal estimate of the true parameter.

5.1   First Stage: Estimating the Conditional Choice Probabilities

Suppose the econometrician has access to cross-sectional data on $n = 1, ..., N$ repetitions of the game, (Y^n, X^n), where $Y^n \in A$ and X^n are the observed outcome and covariates, respectively. In the first stage, we form an estimate $\widehat{\Pr}\left(y|X\right) = \left[\widehat{\Pr}\left(y_1|X\right), ..., \widehat{\Pr}\left(y_l|X\right), ..., \widehat{\Pr}\left(y_{\|A\|}|X\right)\right]$ of $\Pr\left(y|X\right) = \left[\Pr\left(y_1|X\right), ..., \Pr\left(y_l|X\right), ..., \Pr\left(y_{\|A\|}|X\right)\right]$ using a multinomial logit model.

That is, we parameterize the conditional choice probabilities as follows:

$$\Pr(y_l|X) = \frac{\exp(X \cdot \beta_l)}{1 + \sum_{m=2}^{\|A\|} \exp(X \cdot \beta_m)}, \ l = 2, ..., \|A\|,$$

where we normalize β_1 to 0. We can then use MLE to obtain the estimate $\widehat{\beta}_l$ of β_l, $l = 2, ..., \|A\|$. Therefore, given the nth observation of the covariates, X^n, the predicted choice probability for the outcome y_l ($l = 1, ..., \|A\|$) is

$$\widehat{\Pr}(y_l|X^n) = \frac{\exp\left(X^n \cdot \widehat{\beta}_l\right)}{\sum_{m=1}^{\|A\|} \exp\left(X^n \cdot \widehat{\beta}m\right)},$$

with $\widehat{\beta}_1 \equiv 0$.

For this stage, we have assumed that the conditional choice probabilities $\Pr(y|X, \beta)$ are parameterized by a finite parameter vector β. In particular, we have assumed $\Pr(y|X, \beta)$ are multinomial logit as specified in the above. In principle, one could use other methods such as nonparametric or semiparametric approaches to estimate these conditional choice probabilities. However, in many applications, the number of covariates included in X is likely to be large, which would give rise to the "curse of dimensionality" for the nonparametric approach. As for the semiparametric approach, we can use the sieve estimator to obtain estimates of the equilibrium strategies, but at a lower convergence rate than root N (Ai and Chen, 2004). In the point-identified model below, the consistency of the second-stage estimator does not depend on the convergence rate of the first stage. The asymptotic normality does, however. Thus, if we care only about the consistency of the second stage, we can use the more flexible sieve estimator for our first-stage estimation.

5.2 Second Stage: Bounds Estimation

As mentioned earlier, the identified features of our model are the solutions to the minimization problem in (10) above. In general, one does not know whether the population

objective function $D(\theta)$ has a unique or multiple minimizers in the parameter space. Unless one imposes strict identifying assumptions, the model is not point identified. In light of the recent econometrics literature on partially identified models, we seek to make inferences directly on the set Θ_I without imposing identifying assumptions that ensure Θ_I is a singleton.

To obtain an estimate of the set Θ_I, we minimize a feasible sample analog of the population objective function $D(\theta)$ in (10). In so doing, we replace the conditional choice probability $\Pr(y|X)$ by its estimate $\widehat{\Pr}(y|X)$, which we obtained in the first stage. The feasible sample objective function $\widehat{D_N}(\theta)$ is then:

$$\widehat{D_N}(\theta) = \frac{1}{N} \sum_{n=1}^{N} d\left[\widehat{\Pr}(y|X^n), CE(X^n, \theta)\right]. \tag{11}$$

Following Manski and Tamer (2002), we define $\widehat{\Theta_I}$, the estimate of the set Θ_I as follows:

$$\widehat{\Theta_I} = \left\{\theta \in \Theta | \widehat{D_N}(\theta) \leq \min_{\theta \in \Theta} \widehat{D_N}(\theta) + \tau_N\right\} \tag{12}$$

for some $\tau_N > 0$, where $\tau_N \to 0$ as the sample size $N \to \infty$. That is, we define our estimate of the set Θ_I to include all the $\theta's$ that are within τ_N from minimizing the sample objective function $\widehat{D_N}(\theta)$. The proposition below characterizes the set estimate $\widehat{\Theta_I}$.

Proposition 2 $\widehat{\Theta_I}$ *is a convex polytope in* $\mathbb{R}^J$.

This property of $\widehat{\Theta_I}$ is convenient since it implies that the confidence interval of each individual parameter is a closed interval without any "holes" in it. Hence, in order to construct the confidence interval of an individual parameter, it is sufficient to find the minimum and maximum values of this parameter, given the constraint that the whole parameter vector has to be in the set $\widehat{\Theta_I}$.

To understand the large sample behavior of the set $\widehat{\Theta_I}$, we follow Manski and Tamer (2002) to define the distance between two sets $\Theta, \Theta' \subseteq \mathbb{R}^J$ as

$$\rho(\Theta, \Theta') \equiv \sup_{\theta \in \Theta} \inf_{\theta' \in \Theta'} |\theta - \theta'|.$$

Roughly, $\rho(\Theta, \Theta')$ measures the greatest distance between a point $\theta \in \Theta$ and the set Θ'. Note that $\rho(\Theta, \Theta')$ may not equal $\rho(\Theta', \Theta)$ under this definition of distance between two sets Θ and Θ'. We state the following proposition that characterizes the large sample behavior of $\widehat{\Theta_I}$, which is adapted from a similar proposition given in Ciliberto and Tamer (2007).

Proposition 3 *Suppose the estimated conditional choice probability distribution $\widehat{\Pr}(y|X)$ converges to the true conditional choice probability $\Pr(y|X)$ almost surely as the sample size $N \to \infty$. Furthermore, let the perturbation parameter τ_N be such that (i) $\tau_N > 0$; (ii) $\tau_N \to 0$ as $N \to \infty$; (iii) $\sup_\theta |\widehat{D_N}(\theta) - D(\theta)| = o_p(\tau_N)$. Then both $\rho\left(\Theta_I, \widehat{\Theta_I}\right)$ and $\rho\left(\widehat{\Theta_I}, \Theta_I\right)$ converge to 0 almost surely.*

This proposition says that if the first stage estimator is consistent, then under standard regularity conditions the distance between the identified set Θ_I and the estimated set $\widehat{\Theta_I}$ converges with probability 1 to zero. Hence, under the distance metric ρ defined above, $\widehat{\Theta_I}$ is a "consistent" estimate of Θ. Note that the hypothesis in Proposition 3 does not require the first stage to be $\sqrt{N}$−consistent. Thus, the first stage can be estimated more flexibly using sieve or other semiparametric approaches, and we still retain the consistency of the set estimates in the second stage. We do not have results on the convergence rate of the set estimate due to the fact that the speed of set convergence is a concept as yet not well defined in the literature.

To compute the set $\widehat{\Theta_I}$, we first minimize the sample objective function $\widehat{D_N}(\theta)$ as defined in (11), and then solve the system of inequalities $\widehat{D_N}(\theta) \leq \min_{\theta \in \Theta} \widehat{D_N}(\theta) + \tau_N$. In our context, these two tasks are computationally convenient because of the inherent linearity of our problem. To see this, note that in many applications, the deterministic part of the payoff functions f_i is linear in θ. Thus, by definition, the set $CE(X^n, \theta)$ is also linear in θ. It then follows that $\widehat{D_N}(\theta)$ is piecewise linear and convex in θ due to the specific way we defined the distance metric d. Therefore, we can transform the minimization problem $\min_{\theta \in \Theta} \widehat{D_N}(\theta)$ into a linear program, which can be solved very efficiently using

the interior point algorithm. After solving this minimization problem, we obtain $\widehat{\Theta}_I$ by finding all the $\theta's$ that satisfy the inequality $\widehat{D_N}(\theta) \leq \min_{\theta \in \Theta} \widehat{D_N}(\theta) + \tau_N$, which can also be achieved via linear programming because of the piecewise linearity and convexity of $\widehat{D_N}(\theta)$.

Finally, we can construct the confidence region for Θ_I. Following the recent literature on set inference, we say that a set $\widehat{\Theta}_\alpha$ is the confidence region of the identified set Θ_I at a significance level of α if $\widehat{\Theta}_\alpha$ covers Θ_I asymptotically with probability α:

$$\lim_{n \to \infty} \Pr\left(\Theta_I \subseteq \widehat{\Theta}_\alpha\right) = \alpha.$$

To construct such a confidence region, we adapt the nonparametric technique proposed by Chernozhukov, Hong and Tamer (2007, and henceforth CHT) to our current context. Denote the minimum value of the sample objective function $\min_{t \in \Theta} \widehat{D_N}(t)$ as q_n and the level set as

$$C_n(c) \equiv \left\{ \theta \in \Theta : n\left[\widehat{D_N}(\theta) - \min_{t \in \Theta} \widehat{D_N}(t)\right] \leq c \right\}.$$

CHT proved that if we can choose an appropriate cutoff value $c_{(\alpha)}$, the corresponding level set $C_n\left[c_{(\alpha)}\right]$ will cover the identified set Θ_I asymptotically with probability α, that is, $\widehat{\Theta}_\alpha \equiv C_n\left[c_{(\alpha)}\right]$ will be the confidence region we seek.

Therefore, our task boils down to choosing the appropriate cutoff value $c_{(\alpha)}$. In the context of our problem, the subsampling procedure (CHT, 2007) used to pick $c_{(\alpha)}$ consists of the following steps:

1. Choose the initial value of the cutoff, $c_{(0)}$. Following CHT's suggestion, we choose this initial cutoff value to be 25% above the minimum sample objective function, *i.e.*, we let $c_{(0)} = 1.25 q_n$.

2. Construct B_n subsamples of size $b \ll n$. For each subsample $b_n = 1, ..., B_n$, we compute $\widehat{c}_{b_n} = \sup_{\theta \in C_n[c_{(0)}]} b\left[\widehat{D_{b_n}}(\theta) - q_{b_n}\right]$, where $\widehat{D_{b_n}}(\theta)$ is the sample objective function corresponding to the subsample b_n, and q_{b_n} is the minimum of $\widehat{D_{b_n}}(\theta)$ (over the entire parameter space Θ).

3. Let $c_{(\alpha)}$ be the $\alpha-$quantile of $\{\widehat{c}_{b_n}\}_{b_n=1,...,B_n}$.

This subsampling procedure would usually be computationally demanding since it involves solving optimization problems many times along the way. In fact, CHT (2007) and subsequent applied work using this procedure has had to rely on simulated annealing to solve each of these optimization problems. Hence, the computational burden for applying this procedure is very high. In our context, however, the functions $\widehat{D_N}(\theta)$ as well as $\widehat{D_{b_n}}(\theta)$ $(b_n = 1, ..., B_n)$ are piecewise linear and convex in θ, which enables us to apply linear programming to solve the optimization problems. This feature dramatically reduces our computational burden.

5.3   Estimation of Point Identified Models

In the above bounds estimation, we did not assume the model to be identified, and our estimator provides a set estimate $\widehat{\Theta}_I$ and confidence region $\widehat{\Theta}_\alpha$ for the identified features Θ_I of the model. However, under some conditions, a game-theoretic model can be point identified. For example, if there is a set of covariates that affects the payoff functions of some but not all players (Tamer 2003, Bajari, Hong and Ryan, 2007; Bajari, Hong, Kraner and Nekipelov, 2007), then the model might be point identified. It is worth pointing out, however, that identified models may still have multiple equilibria, rendering the traditional direct approaches such as MLE or GMM invalid.

For point-identified models, the estimator described in the above two subsections reduces to the following sharp two-step minimum distance estimator. The first step is the same as in the case of partial identification discussed above; that is, we estimate the conditional choice probabilities using a multinomial logit model with parameter vector β. In the second step, we obtain a *point* estimate $\widehat{\theta}_I$ for the identified parameter θ_I by minimizing

the sample objective function $\widehat{D_N}\left(\theta,\widehat{\beta}\right)$[3]. That is,

$$\widehat{\theta}_I = \arg\min_{\theta\in\Theta} \widehat{D_N}\left(\theta,\widehat{\beta}\right). \tag{13}$$

Now we can recast our two-step estimator as a GMM estimator with the identity weighting matrix, namely, an MOM estimator. To see this, note that $\widehat{\theta}_I$ solves the following first-order condition for the minimization problem (13),

$$\frac{1}{N}\sum_{n=1}^{N} \nabla_\theta d\left(X^n,\theta,\widehat{\beta}\right) = 0, \tag{14}$$

except when $\nabla_\theta d\left(X^n,\theta,\widehat{\beta}\right)$ does not exist. Also, note that in (14), with a slight abuse of notation, we can let $d\left(X^n,\theta,\widehat{\beta}\right)$ denote $d\left[\widehat{\Pr}\left(y|X^n,\widehat{\beta}\right), CE\left(X^n,\theta\right)\right]$. By looking at the definition of the distance metric in (9), we can see that $\nabla_\theta d\left(X^n,\theta,\widehat{\beta}\right)$ does not exist at the kinked points of the function $d\left[\widehat{\Pr}\left(y|X^n,\widehat{\beta}\right), CE\left(X^n,\theta\right)\right]$. Hence, the set of realized values for X^n that make $\nabla_\theta d\left(X^n,\theta,\widehat{\beta}\right)$ nonexistent equal the event

$$\cup_{(i,a_i,d_i)} \left\{ X : \Upsilon^{a_i}\left[\widehat{\Pr}\left(y|X^n,\widehat{\beta}\right)\right] \cdot \left[F_i\left(a_i,X_i,\theta\right) - F_i\left(d_i,X_i,\theta\right)\right] + \ell_G = 0 \right\},$$

which has measure 0. Thus, with probability equal to 1, $\widehat{\theta}_I$ solves the first-order condition (14).

Furthermore, recall that $\widehat{\beta}$ is the maximizer of the log likelihood function

$$\frac{1}{N}\sum_{n=1}^{N} \ln\left[\Pr\left(y_l|X^n,\beta\right)\right] \equiv \frac{1}{N}\sum_{n=1}^{N} l\left(X^n,\beta\right),$$

and thus solves

$$\frac{1}{N}\sum_{n=1}^{N} \nabla_\beta l\left(X^n,\beta\right) = 0. \tag{15}$$

---

[3]Note that there is a slight change in the notation in the current subsection: we explicitly include the first-stage parameter β as an argument in the conditional choice probabilities $\Pr\left(y|X,\beta\right)$, the population objective function $D\left(\cdot,\cdot\right)$, and the sample objective function $\widehat{D_N}\left(\cdot,\cdot\right)$. The purpose of making β explicit will become clear below.

Hence, if we let $\widetilde{g}(X, \theta, \beta) = \left[\nabla_\theta d(X^n, \theta, \beta)', \nabla_\beta l(X^n, \beta)'\right]$, then (14) and (15) are just the two components of the joint moment equation $\frac{1}{N} \sum_{n=1}^N \widetilde{g}\left(X^n, \widehat{\theta}_I, \widehat{\beta}\right) = 0$. Therefore, our two-step estimator can be viewed as a GMM estimator. Finally, note that the first-stage parameter β does not depend on θ, thus, we do not have a recursion problem. The following lemma summarizes the above reformulation of our two-step estimator as a GMM estimator.

Lemma 1 *With probability* 1, $\left(\widehat{\beta}, \widehat{\theta}_I\right)$ *solves the system of equations in* (14) *and* (15). *Hence, the two-step estimator is a GMM estimator with moment functions defined by* (14) *and* (15).

In the following, we will state a proposition that establishes the asymptotic normality of our estimator when the model is point identified. The standard argument for obtaining the asymptotic normality for a GMM estimator is based on the Taylor expansion and requires the moment functions to be once differentiable. Although the moment function $\nabla_\theta d\left(X^n, \theta, \widehat{\beta}\right)$ associated with the second stage is not differentiable in our model, we can still obtain the asymptotic normality of $\widehat{\theta}_I$ by applying theorem 7.2 in Newey and McFadden (1994). This theorem provides sufficient conditions for a GMM estimator to be asymptotically normal when the sample moment is a nonsmooth function. In the following, we state the primitive conditions on our model that guarantee the sufficient conditions in theorem 7.2 of Newey and McFadden (1994):

A1. The conditional choice probabilities $\Pr(y|X, \beta)$ are parameterized by a finite vector β and take the form of multinomial logit. The true parameter β^0 is in the interior of parameter space B.

A2. The parameter space Θ is convex. θ_I is an interior point of Θ and uniquely minimizes $D(\theta, \beta^0)$.

A3. The deterministic part of payoff function $f_i(a_{-i}, a_i, X_i, \theta)$ is bounded in $S_X \times \Theta$, is twice differentiable in X_i and θ, and has bounded first and second derivatives.

A4. X^n are *i.i.d.* across n with $E\|X^n\|^2 < \infty$.

A1 is a convenient assumption that researchers often invoke in practice. However, the multinomial logit specification is not essential for our asymptotic result on $\widehat{\theta}_I$. One could alternatively specify $\Pr(y|X,\beta)$ to be probit. *A2* is an identification assumption. *A3* and *A4* provide certain dominance properties required to establish the stochastic equicontinuity of the sample moments, an essential condition for obtaining asymptotic normality. *A4* rules out some types of covariates that are fat tailed.

We are now ready to state the following proposition that establishes the asymptotic normality of $\widehat{\theta}_I$ when the model is point identified:

Proposition 4 *Suppose A1-A4 hold. Let $g^{(1)}(X,\beta)$ and $g^{(2)}(X,\beta,\theta)$ denote the first-and second-stage moment functions as in (15) and (14), respectively. Let $g_0^{(1)}(\beta^0) \equiv E\left[g^{(1)}(X,\beta^0)\right]$ and $g_0^{(2)}(\beta^0,\theta_I) \equiv E\left[g^{(2)}(X,\beta^0,\theta_I)\right]$ denote the expectations of these moment functions evaluated at the true parameter value (β^0,θ_I), respectively. Furthermore, let*

$$G_\beta^{(1)} \equiv \nabla_\beta g_0^{(1)}(\beta^0), \quad G_\beta^{(2)} \equiv \nabla_\beta g_0^{(2)}(\beta^0,\theta_I), \quad G_\theta^{(2)} \equiv \nabla_\theta g_0^{(2)}(\beta^0,\theta_I),$$

$$\Psi(X,\beta^0,\theta_I) \equiv -G_\beta^{(2)} \cdot \left[G_\beta^{(1)} \cdot g^{(1)}(X,\beta^0)\right],$$

$$\Lambda \equiv E\left\{\left[g^{(2)}(X,\beta^0,\theta_I) + \Psi(X,\beta^0,\theta_I)\right]\left[g^{(2)}(X,\beta^0,\theta_I) + \Psi(X,\beta^0,\theta_I)\right]'\right\}.$$

Then, as $N \to \infty$, $\sqrt{N}\left(\widehat{\theta}_I - \theta_I\right) \xrightarrow{d} N(0,V)$, where $V \equiv \left[G_\theta^{(2)}\right]^{-1} \Lambda \left[G_\theta^{(2)}\right]^{'-1}$.

As is typical of two-step estimators, the variance of $\widehat{\theta}_I$ is adjusted by taking into account the first-stage variances, as reflected in the term Λ. Also, V depends on the true parameters θ_I and β^0. In practice, we may replace θ_I and β^0 by their estimates $\widehat{\theta}_I$ and $\widehat{\beta}$, respectively. We can also use bootstrapping to estimate V. As will be detailed in the proof, our sample objective function $d\left[\widehat{\Pr}\left(y|X^n,\widehat{\beta}\right), CE(X^n,\theta)\right]$ can be written as the objective function in the quantile regression models by setting the appropriate quantile

to be 0. Heuristically, since quantile estimates are generally asymptotically normal, it is not surprising that our estimator is also asymptotically normal. Of course, the proofs differ in some important ways.

5.4 Revisiting the Equilibrium Selection Assumption

We have assumed that the true equilibrium selection mechanism $\{\lambda^n\}_{n=1}^N$ is the same across different observations; *i.e.*, $\lambda^n = \lambda$ for all n, even though the econometrician is assumed not to possess any knowledge about λ. The purpose of assuming that the equilibrium selection mechanism is the same across observations is to allow the econometrician to consistently estimate the conditional choice probability distribution $\Pr(y|X)$. If this assumption does not hold in the data, the conditional choice probability distribution $\Pr^n(y|X)$ in general would be observation-specific, since the functional form of $\Pr^n(y|X) = \int_U \Pr(y|X,\mu) \lambda^n(d\mu)$ depends upon n. As a consequence, the first stage would not offer an estimate of $\Pr(y|X)$, but rather an estimate of the mixing distribution among the different $\Pr^n(y|X)$.

To see more clearly the consequence of violating this assumption, suppose that for some observations, the conditional choice probability distribution equals $\Pr'(y|X)$, while for the other observations have $\Pr''(y|X)$. In this situation, the first-stage estimation will return an estimate of the mixing distribution between $\Pr'(y|X)$ and $\Pr''(y|X)$, denoted as $\Pr^{mix}(y|X)$. Consider, specifically, two observation units, l and k, which have conditional choice probability distributions $\Pr'(y|X)$ and $\Pr''(y|X)$, respectively. Then according to (7), we have:

$$\sum_{a_{-i} \in A_{-i}} \Pr'\left(y|X^l\right) \cdot \left[f_i\left(a^l_{-i}, a^l_i, X^l_i, \theta\right) - f_i\left(a^l_{-i}, d^l_i, X^l_i, \theta\right)\right] \geq -\ell_G,$$
$$\forall i \in I, \forall a_i \in A_i, \forall d_i \in A_i. \tag{16}$$

and

$$\sum_{a_{-i} \in A_{-i}} \Pr''\left(y|X^k\right) \cdot \left[f_i\left(a^k_{-i}, a^k_i, X^k_i, \theta\right) - f_i\left(a^k_{-i}, d^k_i, X^k_i, \theta\right)\right] \geq -\ell_G,$$
$$\forall i \in I, \forall a_i \in A_i, \forall d_i \in A_i. \tag{17}$$

Note that $\Pr^{mix}\left(y|X^l\right)$ does not necessarily satisfy condition (16), since $\Pr''\left(y|X^l\right)$ may violate it. Similarly, $\Pr^{mix}\left(y|X^k\right)$ does not necessarily satisfy (17). Hence, the mixing distribution $\Pr^{mix}\left(y|X\right)$ may not satisfy the sample restrictions as in (7). As a consequence, the two-step approach would not be valid, since the first-stage estimation gives an estimate of the mixing between these two conditional choice probability distributions, which no longer necessarily satisfy the sample restrictions.

6 Monte Carlo Evidence

In this section we present Monte Carlo evidence for the two-step estimation approach described in the last section. In order to illustrate the performance of the method for a general class of models, we do not impose point identification assumptions when designing the Monte Carlo experiments below. Therefore, we seek the *set* estimates and confidence intervals for the identified features of these models. Section 6.1 describes the design; Section 6.2 presents the results of the experiments.

6.1 Design of the Monte Carlo Experiments

In designing the experiments, we use a 2×2 entry game to generate the outcome data. In this model, each of the two firms decides whether or not to enter a single market. Hence $A_i = \{1, 0\}$, where $a_i = 1$ means player i enters the market and $a_i = 0$ means the opposite. For the purpose of our Monte Carlo experiments, we specify the profit function for each firm i as follows:

$$\pi_i\left(a, x, \beta, \delta, \epsilon_i\right) = \left[\beta^1 x^1 + \beta^2 x^2 - \delta \cdot 1\left\{a_{-i} = 1\right\} + \epsilon_i\right] \cdot 1\left\{a_i = 1\right\}. \qquad (18)$$

In this specification, x^1 and x^2 are the covariates representing the market-specific characteristics, which may be, for example, demand shifters. Thus, β^1 and β^2 capture the effects of x^1 and x^2 on each firm's profits, respectively. In addition, δ measures the

spillover effect of one firm's entry on the other's profits; ϵ_i represents firm $i's$ profit shock, which is common knowledge among the firms, but unobserved by the econometrician. Finally, each firm earns zero profit if it does not enter the market.

After setting up the payoff structures, we specify the values of the parameters and the distributions of the covariates and profit shocks. The parameter vector is set to be $(\beta^1, \beta^2, \delta) = (1.0, -1.0, 5.5)$. Furthermore, we let the profit shock $\epsilon_i's$ be $i.i.d.$ across i and distributed as standard normal $N(0,1)$, as well as be independent of the covariates x^1 and x^2, which are themselves assumed to be independent of each other. We specify x^1 and x^2 to be normally distributed with $x^1 \sim N(0,5)$ and $x^2 \sim N(1,7)$, respectively. It is important to observe that the supports of the normal distributions are *unbounded*. As Manski and Tamer (2002) have pointed out, for set-identified models, wide variations in the covariates help shrink the set estimates of the parameters. Hence, the specifications of the distributions for the covariates here tend to create a bias towards finding narrower intervals for the parameters. Thus, if we instead specified the covariates to be uniformly distributed, we may tend to find wider interval parameter estimates.

To generate the outcome data, we assume the firms play a Nash equilibrium. However, given the parameters specified above, for some realizations of the covariates and of the econometric errors, the models we consider might have multiple equilibria. We assume each equilibrium is picked with equal probability in such situations. Also, if a mixed-strategy equilibrium happens to be chosen, the outcome of the game will be determined randomly according to the probability distribution defined by this mixed-strategy equilibrium.

Finally, we specify the sample size for the experiment. We draw 600 samples of size $N = 500$, $N = 1000$, and $N = 2000$, using the distributions specified above. Corresponding to each specification of the sample size, we will report the averages of the set estimates and of the confidence intervals, where the average is taken over the 600 different samples. The next subsection presents these results.

30

6.2 Results from the Monte Carlo Experiments

Table 1 reports the results corresponding to the Monte Carlo experiment. Each entry in this table has two closed intervals associated with it. The top intervals provide the means of the interval parameter estimates across the 600 samples with corresponding sizes; the bottom intervals provide the means of the confidence intervals. We normalize β^1 to be equal to 1.0, since the players' choices remain unchanged if we multiply all the parameters by the same positive number. Hence, in the table, we list only the entries associated with β^2 and δ, with the understanding that β^1 equals 1.0.

Before looking closely at the results displayed in the table, let us explain briefly how we obtained these numbers using the corresponding procedures described in Section 4. For each sample, whether its size equals $N = 500$, $N = 1000$, or $N = 2000$, we first obtain the set of parameter vectors $\widehat{\Theta}_I$ that are within $\tau_N = \frac{1}{N}$ from minimizing the corresponding sample objective function. This set is described in (12). After solving the embedded minimization problem $\min_{\theta \in \Theta} \widehat{D_N}(\theta)$ using linear programming, we insert the corresponding minimum values back into (12) and can completely characterize $\widehat{\Theta}_I$ in terms of a system of *linear* inequalities. After obtaining $\widehat{\Theta}_I$, we then project it onto each axis of the linear space $\mathbb{R}^J$ and obtain—again, via linear programming—the lower and upper bounds of each interval parameter estimate, as displayed in the table.

Computing the confidence intervals is more complex. We follow the subsampling procedure presented in Section 5. We set the size of the subsample equal to 2% of the size of the original sample. Thus, corresponding to $N = 500$, 1000, and 2000, the sizes of the subsamples are $b = 10$, 20 and 40, respectively. We then let the number of subsamples $b_N = 20N$; *i.e.*, for the original sample with size equal to 500, we construct 10,000 subsamples, etc. The literature provides little guidance as to how to choose the appropriate size for the subsamples as well as how many subsamples should be constructed (Ciliberto and Tamer, 2007; Chernozhukov, Hong and Tamer, 2007). In the current context, the subsample size and the number of subsamples we are choosing are heuristic and based

largely on computational considerations.

From this table, we can immediately see that none of the parameters is point identified, to the extent that no interval parameter estimate is tightly centered around the true value of the parameter. Nevertheless, these interval parameter estimates are more or less informative of the true parameters. This observation is especially true for β^2 in the sense that most of the interval parameter estimates in the table contain the associated true parameter value. In comparison, the interval parameter estimates for δ are less successful, since these interval estimates are systematically biased downward. In addition, the increase in sample size helps make the bounds tighter. Relative to the set estimates, the confidence intervals are quite wide, which may be caused by the difficulty of obtaining the accurate cutoff value $c_{(\alpha)}$ in the subsampling procedure. Nonetheless, the overall picture is encouraging, in the sense that we have obtained quite sensible set estimates and confidence intervals at relatively low computational cost.

In principle, our estimator should produce tighter bounds in a game with more players and/or actions. More specifically, recall that the number of inequality restrictions imposed by correlated equilibria equals $\sum_{i \in I} \|A_i\| \cdot (\|A_i\| - 1)$, which is monotonically increasing in both the number of players and the number of actions each player has. As such, for games of larger scale, the parameters need to satisfy more inequality restrictions. Thus, the identified regions for the parameters in a larger game tend to be smaller. However, when we conduct the set inferences, we need to consistently estimate the conditional choice probabilities in the first stage, which requires more data points when the number of potential outcomes grows. Note that the number of outcomes for a game equals $\prod_{i \in I} \|A_i\|$. Thus, we need more data to consistently estimate the conditional choice probabilities in games with more than two players. Therefore, the relative performance of our estimator in large- vs. small-scale games depends on the sample size: when sample size is small, the estimate of the conditional choice probabilities is less accurate, which counteracts the advantage of the three-player game in having more inequality restrictions. When sample size is large enough, so that the first stage estimation for the

larger game is consistent, the greater number of inequality restrictions associated with the larger game enables our estimator to produce narrower bounds. The Monte-Carlo results obtained from games with three or more players are available upon request from the author.

7   Conclusion

This paper has proposed a two-step approach for estimating static discrete games with complete information. This approach exploits the relationship between correlated equilibria and Nash equilibria; namely, all the Nash equilibria of a static discrete game with complete information are contained in the set of correlated equilibria of that same game. Starting from this theoretic relationship, we derive a system of linear inequalities that the conditional choice probabilities satisfy. This system of inequality restrictions suggest the following two-step procedure to estimate the structural parameters in the payoff functions. In the first step, we estimate the conditional choice probabilities that each possible outcome is realized by parametrically or semiparametrically regressing the observed outcomes on the covariates. In the second step, we recover the structural parameters by minimizing the sample average distance between the set defined by the system of linear inequalities and the conditional choice probability vector that we obtained in the first step.

Our estimation approach has both conceptual and computational advantages over some previous approaches for estimating static discrete games with complete information. At the conceptual level, first, the restrictions we exploit have a rigorous theoretical foundation; second, our estimator addresses both the issues of multiple equilibria and of the possible nonexistence of pure strategy equilibria. At the computational level, our estimator exploits the inherent linearity of correlated equilibria and applies the linear programming techniques to solve the optimization problems that we encounter.

33

Hence, compared to the previous approaches, our estimator substantially reduces the computational costs. This feature is especially important for games with a large number of players and/or a large strategy space.

Our estimator is robust to equilibrium behavior of the players. In this paper we have assumed that players behave according to a Nash equilibrium. However, our estimation approach does not rely on this assumption. For example, we can allow the players to communicate with each other and play correlated equilibrium rather than Nash. As long as the stipulated equilibrium which the players are assumed to play is contained in the set of correlated equilibria, our approach still applies. In this sense, our estimator is robust to the behavioral assumptions imposed on the players.

Our estimator is applicable to a wide range of game-theoretic models that are interesting to I.O. economists. Examples include entry games, technology-adoption games, and labor force participation games, among others. For instance, in a companion paper, we apply the method developed here to study the equilibrium behavior of Texas hotels in choosing chain-affiliation status (i.e., operating independently vs. becoming affiliated with a large chain). Since a hotel's franchising status may affect the market competition that ensues, we model hotels' choices regarding franchising status as a game. We specify the reduced form of the profit functions and apply our method to estimate the parameters therein. The estimation results suggest that the choice of franchising status is a strategic complement among Texas hotels.

A Mathematical appendix

A.1 Proof of Proposition 2

Note that $\widehat{D_N}(\theta)$ is piecewise linear and convex in θ. The right hand side of the inequality in the set defined by (12) does not depend on θ. Take $\forall \theta^1, \theta^2 \in \widehat{\Theta_I}$. By the convexity of $\widehat{D_N}(\theta)$,

$$\widehat{D_N}\left[\alpha \cdot \theta^1 + (1-\alpha) \cdot \theta^2\right] \leq \alpha \widehat{D_N}\left(\theta^1\right) + (1-\alpha)\widehat{D_N}\left(\theta^2\right),$$

for $\forall \alpha \in [0,1]$. It then follows that $\alpha \cdot \theta^1 + (1-\alpha) \cdot \theta^2 \in \widehat{\Theta}_I$, which implies that $\widehat{\Theta}_I$ is convex. Furthermore, note that $\widehat{\Theta}_I$ is enclosed by hyperplanes due to the piecewise linearity of $\widehat{D_N}(\theta)$. The conclusion then immediately follows. ∎

A.2 Proof of Proposition 3

It follows from theorem $5b$ in Manski and Tamer (2002). ∎

A.3 Proof of Proposition 4

The proof of proposition 4 consists of two parts. In the first part, we establish consistency. We then prove the asymptotic normality in the second part.

Part I. Consistency

Here, we want to prove that $\widehat{\beta} \overset{p}{\to} \beta^0$ (consistency in the first stage) and $\widehat{\theta}_I \overset{p}{\to} \theta_I$ (consistency in the second stage). The first stage consistency follows immediately from the fact that we obtained $\widehat{\beta}$ using Maximum Likelihood Estimation. Therefore, we only need to prove the consistency for the second stage. Recall that

$$\widehat{\theta}_I = \arg\min_{\theta \in \Theta} D_N\left(\widehat{\beta}_N, \theta\right) \equiv \arg\min_{\theta \in \Theta} \frac{1}{N} \sum_{n=1}^{N} d\left[\Pr\left(y|X^n, \widehat{\beta}_N\right), CE\left(X^n, \theta\right)\right],$$

where we write $\widehat{\beta}$ as $\widehat{\beta}_N$ in order to make it explicit that $\widehat{\beta}_N$ is a sample statistic. Given assumptions $A1-A4$, by theorem 2.7 in Newey and McFadden (1994), it suffices to show the following: $1. D_N\left(\widehat{\beta}_N, \theta\right)$ is convex in θ; 2. There exists a function $D_0(\theta)$ such that $D_0(\theta)$ is uniquely minimized at θ_I and $D_N\left(\widehat{\beta}_N, \theta\right) \overset{p}{\to} D_0(\theta)$ as $N \to \infty$ for all $\theta \in \Theta$. The convexity of $D_N\left(\widehat{\beta}_N, \theta\right)$ follows from the fact that $d\left[\Pr\left(y|X^n, \widehat{\beta}_N\right), CE\left(X^n, \theta\right)\right]$ is convex in θ due to the specific way we constructed the distance metric d. Let

$$D_0(\theta) = E\left\{d\left[\Pr\left(y|X, \beta^0\right), CE\left(X, \theta\right)\right]\right\} \equiv D\left(\beta^0, \theta\right),$$

then θ_I uniquely minimizes $D\left(\beta^0, \theta\right)$ by the identification assumption $A2$. Therefore, it only remains to show that $D_N\left(\widehat{\beta}_N, \theta\right) \overset{p}{\to} D\left(\beta^0, \theta\right)$ for all $\theta \in \Theta$.

Note that

$$D_N\left(\widehat{\beta}_N,\theta\right) - D_0\left(\theta\right) \equiv D_N\left(\widehat{\beta}_N,\theta\right) - D\left(\beta^0,\theta\right)$$
$$= \left[D_N\left(\widehat{\beta}_N,\theta\right) - D_N\left(\beta^0,\theta\right)\right] + \left[D_N\left(\beta^0,\theta\right) - D\left(\beta^0,\theta\right)\right].$$

Hence, it suffices to show that both these bracketed terms are $o_p\left(1\right)$. Obviously, $D_N\left(\beta^0,\theta\right) \overset{p}{\to} D\left(\beta^0,\theta\right)$ for all θ by the weak law of large numbers. It needs a bit more work to show $D_N\left(\widehat{\beta}_N,\theta\right) - D_N\left(\beta^0,\theta\right) = o_p\left(1\right)$ for all θ. Since the function $D_N\left(\widehat{\beta}_N,\theta\right)$ is not continuously differentiable, we cannot write out $D_N\left(\widehat{\beta}_N,\theta\right) - D_N\left(\beta^0,\theta\right)$ using Taylor expansion. However, as shown below, there exist two sequences of *i.i.d.* random variables $\{K^n\}_{n=1}^N$ and $\{J^n\}_{n=1}^N$, both of which have bounded first moments, such that for all θ,

$$\left|D_N\left(\widehat{\beta}_N,\theta\right) - D_N\left(\beta^0,\theta\right)\right| \leq \left|\widehat{\beta}_N - \beta^0\right| \cdot \left[\frac{1}{N}\sum_{n=1}^N K^n\right] + o_p\left(\left|\widehat{\beta}_N - \beta^0\right|\right) \cdot \left[\frac{1}{N}\sum_{n=1}^N J^n\right] \tag{19}$$

Then by the consistency in the first stage, the left hand side of the above inequality is $o_p\left(1\right)$. Our task now has been reduced to construct such sequences of *i.i.d.* random variables, $\{K^n\}_{n=1}^N$ and $\{J^n\}_{n=1}^N$.

Recall that

$$D_N\left(\beta,\theta\right) = \frac{1}{N}\sum_{n=1}^N \sum_{i,a_i,d_i} \min\left\{0, \Upsilon^{a_i}\left[\Pr\left(y|X^n,\beta\right)\right] \cdot \triangle F_i\left(a_i,d_i,X_i^n,\theta\right) + \ell_G\right\},$$

where $\triangle F_i\left(a_i,d_i,X_i^n,\theta\right)$ denotes the change in player $i's$ deterministic payoff vector caused by the deviation from a_i to d_i. That is,

$$\triangle F_i\left(a_i,d_i,X_i^n,\theta\right) = F_i\left(a_i,X_i^n,\theta\right) - F_i\left(d_i,X_i^n,\theta\right).$$

Hence,

$$
\left| D_N\left(\widehat{\beta}_N, \theta\right) - D_N\left(\beta^0, \theta\right) \right|
$$

$$
= \left| \frac{1}{N} \sum_{n=1}^{N} \sum_{i, a_i, d_i} \left[\begin{array}{c} \min\left\{0, \Upsilon^{a_i}\left[\Pr\left(y|X^n, \widehat{\beta}_N\right)\right] \cdot \triangle F_i\left(a_i, d_i, X_i^n, \theta\right) + \ell_G\right\} \\ -\min\left\{0, \Upsilon^{a_i}\left[\Pr\left(y|X^n, \beta^0\right)\right] \cdot \triangle F_i\left(a_i, d_i, X_i^n, \theta\right) + \ell_G\right\} \end{array} \right] \right|
$$

$$
\leq \frac{1}{N} \sum_{n=1}^{N} \sum_{i, a_i, d_i} \left| \left[\begin{array}{c} \min\left\{0, \Upsilon^{a_i}\left[\Pr\left(y|X^n, \widehat{\beta}_N\right)\right] \cdot \triangle F_i\left(a_i, d_i, X_i^n, \theta\right) + \ell_G\right\} \\ -\min\left\{0, \Upsilon^{a_i}\left[\Pr\left(y|X^n, \beta^0\right)\right] \cdot \triangle F_i\left(a_i, d_i, X_i^n, \theta\right) + \ell_G\right\} \end{array} \right] \right|
$$

$$
\leq \frac{1}{N} \sum_{n=1}^{N} \sum_{i, a_i, d_i} \left| \left\{ \Upsilon^{a_i}\left[\Pr\left(y|X^n, \widehat{\beta}_N\right)\right] - \Upsilon^{a_i}\left[\Pr\left(y|X^n, \beta^0\right)\right] \right\} \cdot \triangle F_i\left(a_i, d_i, X_i^n, \theta\right) \right|,
$$

where the first inequality follows by the triangular inequality and the second one follows from the fact that $|\min\{0, a\} - \min\{0, b\}| \leq |a - b|$ for $\forall a, b \in \mathbb{R}$. Since $\Pr\left(y|X, \beta\right)$ is continuously differentiable in β,

$$
\Upsilon^{a_i}\left[\Pr\left(y|X^n, \widehat{\beta}_N\right)\right] - \Upsilon^{a_i}\left[\Pr\left(y|X^n, \beta^0\right)\right] = \left|\widehat{\beta}_N - \beta^0\right| \cdot \nabla_\beta \Upsilon^{a_i}\left[\Pr\left(y|X^n, \beta^0\right)\right] + o_p\left(\left|\widehat{\beta}_N - \beta^0\right|\right) \tag{20}
$$

Therefore, by substituting (21) into the second inequality in the above and rearranging terms, we have

$$
\left| D_N\left(\widehat{\beta}_N, \theta\right) - D_N\left(\beta^0, \theta\right) \right|
$$

$$
\leq \left|\widehat{\beta}_N - \beta^0\right| \cdot \frac{1}{N} \sum_{n=1}^{N} \left\{ \begin{array}{c} \sum_{i, a_i, d_i} \left|\nabla_\beta \Upsilon^{a_i}\left[\Pr\left(y|X^n, \beta^0\right)\right] \cdot \triangle F_i\left(a_i, d_i, X_i^n, \theta\right)\right| \\ +o_p\left(1\right) \cdot \sum_{i, a_i, d_i} \left|\triangle F_i\left(a_i, d_i, X_i^n, \theta\right)\right| \end{array} \right\}.
$$

Let

$$
K^n = \sum_{i, a_i, d_i} \left| \nabla_\beta \Upsilon^{a_i}\left[\Pr\left(y|X^n, \beta^0\right)\right] \cdot \triangle F_i\left(a_i, d_i, X_i^n, \theta\right) \right|
$$

and

$$
J^n = \sum_{i, a_i, d_i} \left| \triangle F_i\left(a_i, d_i, X_i^n, \theta\right) \right|,
$$

By the weak law of large numbers,

$$
\frac{1}{N} \sum_{n=1}^{N} K^n \xrightarrow{p} E\left(K\right),
$$

37

and

$$\frac{1}{N}\sum_{n=1}^{N} J^n \xrightarrow{p} E(J).$$

Thus, the right hand side of (20) is $o_p(1)$, which implies that $D_N\left(\widehat{\beta}_N, \theta\right) - D_N\left(\beta^0, \theta\right) = o_p(1)$. Therefore, the consistency in the second stage has been proved. It is worth pointing out that $E(K) > 0$, which implies that the first-stage consistency is necessary in order for us to establish the second-stage consistency. To see this, let S_X denote the support of X. Take a compact set $Q \subseteq S_X$ such that Q has a positive probability measure. Since K is continuous for each possible realization of X, it follows that K has minimum value $\underline{k} > 0$ over the set Q. Hence,

$$E(K) = \Pr(Q) E(K|Q) + [1 - \Pr(Q)] E(K|S_X \backslash Q) \geq \underline{k} \Pr(Q) > 0,$$

where the equality follows from conditioning and the weak inequality follows from the fact that $K \geq 0$ for $\forall X$.

Part II. Asymptotic Normality

To prove the asymptotic normality of $\widehat{\theta}_I$, we apply theorem 7.2 in Newey and McFadden (1994). This theorem provides sufficient conditions that guarantee the asymptotic normality of a GMM estimator whose moment functions are nonsmooth. As we have seen in Lemma 1, our two-step estimate $\left(\widehat{\beta}, \widehat{\theta}_I\right)$ solves the following moment conditions

$$\frac{1}{N}\sum_{n=1}^{N} g(X^n, \beta, \theta) = 0,$$

where

$$g(X, \beta, \theta) \equiv \left[g^{(1)}(X, \beta, \theta)', g^{(2)}(X, \beta, \theta)'\right]' = \left[\nabla_\beta l(X, \beta)', \nabla_\theta d(X, \theta, \beta)'\right]'.$$

It is easy to see that $\nabla_\theta d(X, \theta, \beta)$ is not differentiable. Let

$$g_0(\beta, \theta) \equiv \left[g_0^{(1)}(\beta, \theta)', g_0^{(2)}(\beta, \theta)'\right]' = \left[E[\nabla_\beta l(X, \beta)]', E[\nabla_\theta d(X, \theta, \beta)]'\right]'.$$

Also, we let $\widehat{g}_N(\beta, \theta) \equiv \frac{1}{N}\sum_{n=1}^{N} g(X^n, \beta, \theta)$ denote the sample average of the moment functions. We now verify all the hypotheses in theorem 7.2 following the steps below:

Condition 1: $g_0\left(\beta^0, \theta_I\right) = 0$ *where β^0 and θ_I are the true parameters for β and θ, respectively.*

By the assumption that the model is point identified, we have

$$\theta_I = \arg\min_{\theta} E\left\{d\left[\Pr\left(y|X, \beta^0\right)\right], CE\left(X, \theta\right)\right\}.$$

By the first order condition of this minimization problem,

$$\nabla_\theta E\left\{d\left[\Pr\left(y|X, \beta^0\right), CE\left(X, \theta_I\right)\right]\right\} = 0.$$

Changing the order of integration and differentiation, we have

$$E\left\{\nabla_\theta d\left[\Pr\left(y|X, \beta^0\right), CE\left(X, \theta_I\right)\right]\right\} = 0.$$

That is,

$$g_0^{(2)}\left(\beta^0, \theta_I\right) = 0. \tag{21}$$

By applying a similar argument to the first-stage estimation, we have

$$g_0^{(1)}\left(\beta^0, \theta_I\right) = 0. \tag{22}$$

Hence, equations (22) and (23) imply that $g_0\left(\beta^0, \theta_I\right) = 0$.

Condition 2: $g_0\left(\beta, \theta\right)$ *is differentiable at $\left(\beta^0, \theta_I\right)$.*

We will verify this condition together with the stochastic equicontinuity condition (condition 5) by applying theorem 7.3 in Newey and McFadden (1994) after checking conditions 3 and 4.

Condition 3: $\left(\beta^0, \theta_I\right)$ *is an interior point of the whole parameter space.*

This condition is guaranteed by assumptions $A1$ and $A2$.

Condition 4: $\widehat{g}_N\left(\beta, \theta\right)$ *is asymptotically normal.*

By assumptions $A3$ and $A4$,

$$\Sigma \equiv E\left[g\left(X, \beta^0, \theta_I\right) \cdot g\left(X, \beta^0, \theta_I\right)'\right] < \infty.$$

Then by the Central Limit Theorem and *condition 1* (which we have verified),

$$\sqrt{N}\left[\widehat{g}_N\left(\beta,\theta\right)\right] \xrightarrow{d} N\left(0,\Sigma\right).$$

Condition 5: Stochastic Equicontinuity: for any $\delta_N \to 0$,

$$\sup_{\|\eta-\eta^0\|\leq\delta_N} \sqrt{N}\|\widehat{g}_N\left(\eta\right) - \widehat{g}_N\left(\eta^0\right) - g_0\left(\eta\right)\| \Big/ \left[1 + \sqrt{N}\|\eta-\eta^0\|\right] \xrightarrow{p} 0,$$

where we denote $\eta \equiv \left(\beta,\theta\right)$, $\eta^0 \equiv \left(\beta^0,\theta_I\right)$ and $\widehat{\eta} \equiv \left(\widehat{\beta},\widehat{\theta}_I\right)$. In the following, we apply theorem 7.3 of Newey and McFadden (1994) to verify condition 5 together with condition 2 above. This theorem provides a sufficient condition for 2 and 5 to hold. This sufficient condition says that if there exists a matrix $\triangle X$ and a number $\varepsilon > 0$ such that:

(i) With probability one,

$$r\left(X,\eta\right) \equiv \|g\left(X,\eta\right) - g\left(X,\eta^0\right) - \triangle X \cdot \left(\eta - \eta^0\right)\| \Big/ \|\eta - \eta^0\| \to 0,$$

as $\eta \to \eta^0$

(ii)

$$E\left[\sup_{\|\eta-\eta^0\|<\varepsilon} r\left(X,\eta\right)\right] < \infty$$

(iii)

$$\frac{1}{N}\sum_{n=1}^{N}\triangle\left(X^n\right) \xrightarrow{p} E\left[\triangle X\right]$$

then conditions 2 and 5 stated above are satisfied. We now check $(i) - (iii)$. As will be clear below, the key in proving the satisfaction of $(i) - (iii)$ is to construct the matrix $\triangle X$.

Recall that the moment functions corresponding to the first and second stage are $g^{(1)}\left(X,\beta,\theta\right) = \nabla_\beta l\left(X,\beta\right)$ and $g^{(2)}\left(X,\beta,\theta\right) = \nabla_\theta d\left(X,\theta,\beta\right)$, respectively. Let

$$L\left(a_i,d_i,X,\beta,\theta\right) \equiv \Upsilon^{a_i}\left[\Pr\left(y|X,\beta\right)\right] \cdot \left[F_i\left(a_i,X_i,\theta\right) - F_i\left(d_i,X_i,\theta\right)\right] + \ell_G.$$

Then,

$$g^{(2)}\left(X,\beta,\theta\right) = \sum_{(i,a_i,d_i)} \left[-1\left\{L\left(a_i,d_i,X,\beta,\theta\right)<0\right\}\right] \cdot \nabla_\theta L\left(a_i,d_i,X,\beta,\theta\right).$$

Note that $g^{(2)}\left(X,\beta,\theta\right)$ is not differentiable. However, if we ignored the part containing the indicator function, the other part $\nabla_\theta L\left(a_i,d_i,X,\beta,\theta\right)$ is continuously differentiable in both β and θ. Based on this observation, we construct the matrix $\triangle\left(X\right)$ as follows:

$$\triangle\left(X\right) = \begin{bmatrix} \frac{\partial}{\partial\beta\partial\beta'}l\left(X,\beta^0\right) & 0 \\ \sum_{(i,a_i,d_i)}\left[-1\left\{L<0\right\}\right]\cdot\frac{\partial^2 L}{\partial\theta\partial\beta} & \sum_{(i,a_i,d_i)}\left[-1\left\{L<0\right\}\right]\cdot\frac{\partial^2 L}{\partial\theta^2} \end{bmatrix},$$

where we have suppressed the arguments in the function L and evaluate all the elements in this matrix at the true parameter value $\left(\beta^0,\theta_I\right)$.

Let us fix X and partition the matrix $\triangle X$ as

$$\triangle X = \begin{bmatrix} \triangle^{(1)}X \\ \triangle^{(2)}X \end{bmatrix},$$

where $\triangle^{(1)}X$ has the same row dimension as $\frac{\partial}{\partial\beta\partial\beta'}l\left(X,\beta^0\right)$. By substituting $\triangle\left(X\right)$ into the expression of $r\left(X,\eta\right)$ in the above, and use the triangular inequality, we have

$$\begin{aligned} r\left(X,\eta\right) &= \left\|g\left(X,\eta\right) - g\left(X,\eta^0\right) - \triangle X\cdot\left(\eta-\eta^0\right)\right\| \diagup \left\|\eta-\eta^0\right\| \\ &\leq \left\|g^{(1)}\left(X,\beta\right) - g^{(1)}\left(X,\beta^0\right) - \triangle^{(1)}X\cdot\left(\beta-\beta^0\right)\right\| \diagup \left\|\beta-\beta^0\right\| \\ &\quad + \left\|g^{(2)}\left(X,\eta\right) - g^{(2)}\left(X,\eta^0\right) - \triangle^{(2)}X\cdot\left(\eta-\eta^0\right)\right\| \diagup \left\|\eta-\eta^0\right\| \\ &\equiv r^{(1)}\left(X,\beta\right) + r^{(2)}\left(X,\eta\right). \end{aligned} \tag{23}$$

Note that $r^{(1)}\left(X,\beta\right)$ can be equivalently expressed as:

$$r^{(1)}\left(X,\beta\right) = \left\|\nabla_\beta l\left(X,\beta\right) - \nabla_\beta l\left(X,\beta^0\right) - \frac{\partial}{\partial\beta\partial\beta'}l\left(X,\beta^0\right)\cdot\left(\beta-\beta^0\right)\right\| \diagup \left\|\beta-\beta^0\right\| \tag{24}$$

Since $\nabla_\beta l\left(X,\beta\right)$ is continuously differentiable, by the definition of derivatives, as $\beta\to\beta^0$, the right hand side of equation (25) converges to 0. Hence, as $\beta\to\beta^0$,

$$r^{(1)}\left(X,\beta\right)\to 0.$$

41

Therefore, it only remains to show that

$$r^{(2)}(X, \eta) \to 0.$$

By substituting the expression for $\triangle^{(2)} X$, we have:

$$
\begin{aligned}
r^{(2)}(X, \eta) = \| \sum_{(i,a_i,d_i)} &\frac{\partial}{\partial \theta} L\left(a_i, d_i, X, \beta, \theta\right) \cdot \left[-1\left\{L\left(a_i, d_i, X, \beta, \theta\right) < 0\right\}\right] \\
- \sum_{(i,a_i,d_i)} &\frac{\partial}{\partial \theta} L\left(a_i, d_i, X, \beta^0, \theta_I\right) \cdot \left[-1\left\{L\left(a_i, d_i, X, \beta^0, \theta_I\right) < 0\right\}\right] \\
- \left\{ \sum_{(i,a_i,d_i)} \frac{\partial^2}{\partial \theta \partial \theta'} L\left(a_i, d_i, X, \beta^0, \theta_I\right) \cdot \left[-1\left\{L\left(a_i, d_i, X, \beta^0, \theta_I\right) < 0\right\}\right] \right\} \cdot \left(\theta - \theta^0\right) \\
- \left\{ \sum_{(i,a_i,d_i)} \frac{\partial^2}{\partial \theta \partial \beta'} L\left(a_i, d_i, X, \beta^0, \theta_I\right) \cdot \left[-1\left\{L\left(a_i, d_i, X, \beta^0, \theta_I\right) < 0\right\}\right] \right\} \cdot \left(\beta - \beta^0\right) \|.
\end{aligned}
$$

Here, we consider two possibilities:

1. $-1\left\{L\left(a_i, d_i, X, \beta^0, \theta_I\right) < 0\right\} = 0.$

2. $-1\left\{L\left(a_i, d_i, X, \beta^0, \theta_I\right) < 0\right\} = -1.$

In the first possibility, since the function $L\left(a_i, d_i, X, \beta, \theta\right)$ is continuous in (β, θ), as $(\beta, \theta) \to (\beta^0, \theta_I)$, $-1\left\{L\left(a_i, d_i, X, \beta, \theta\right) < 0\right\} \to 0$. Therefore, in some neighborhood around (β^0, θ_I),

$$r^{(2)}(X, \eta) = 0.$$

For the second possibility, by the continuity of the function L, $-1\left\{L\left(a_i, d_i, X, \beta, \theta\right) < 0\right\}$ is equal to -1 in a small neighborhood around (β^0, θ_I). Therefore, in this neighborhood,

$$
\begin{aligned}
r^{(2)}(X, \eta) = \| \sum_{(i,a_i,d_i)} &\frac{\partial}{\partial \theta} L\left(a_i, d_i, X, \beta, \theta\right) - \sum_{(i,a_i,d_i)} \frac{\partial}{\partial \theta} L\left(a_i, d_i, X, \beta^0, \theta_I\right) \\
- &\left[\sum_{(i,a_i,d_i)} \frac{\partial^2}{\partial \theta \partial \theta'} L\left(a_i, d_i, X, \beta^0, \theta_I\right) \right] \cdot \left(\theta - \theta^0\right) \\
- &\left[\sum_{(i,a_i,d_i)} \frac{\partial^2}{\partial \theta \partial \beta'} L\left(a_i, d_i, X, \beta^0, \theta_I\right) \right] \cdot \left(\beta - \beta^0\right) \|.
\end{aligned}
$$

Since $L\left(a_i, d_i, X, \beta, \theta\right)$ is continuously differentiable, it then follows that for each X,

$$r^{(2)}\left(X, \eta\right) = o\left(\left\|\eta - \eta^0\right\|\right).$$

Hence, condition (i) is satisfied for this second possibility as well. We now move on to show that conditions (ii) and (iii) are satisfied.

To verify condition (ii), observe that by the inequality in (24), it is sufficient to find $\varepsilon > 0$ such that

$$E\left[\sup_{\|\eta - \eta^0\| < \varepsilon} r^{(1)}\left(X, \beta\right)\right] < \infty \tag{25}$$

and

$$E\left[\sup_{\|\eta - \eta^0\| < \varepsilon} r^{(2)}\left(X, \eta\right)\right] < \infty \tag{26}$$

To prove (26), note that by (25) and triangle inequality it is sufficient to show

$$E\left[\sup_{\|\eta - \eta^0\| < \varepsilon} \|\frac{\partial}{\partial\beta\partial\beta'} l\left(X, \beta^0\right)\|\right] < \infty,$$

which follows from writing out the term $\frac{\partial}{\partial\beta\partial\beta'} l\left(X, \beta^0\right)$ and then applying assumption $A.4$. We obtain (27) using a similar argument as follows. First note that for $\forall X \in S_X$, we have the following inequality:

$$r^{(2)}\left(X, \eta\right) \leq \| \sum_{(i,a_i,d_i)} \frac{\partial}{\partial\theta} L\left(a_i, d_i, X, \beta, \theta\right) - \sum_{(i,a_i,d_i)} \frac{\partial}{\partial\theta} L\left(a_i, d_i, X, \beta^0, \theta_I\right)$$

$$- \left[\sum_{(i,a_i,d_i)} \frac{\partial^2}{\partial\theta\partial\theta'} L\left(a_i, d_i, X, \beta^0, \theta_I\right)\right] \cdot \left(\theta - \theta^0\right)$$

$$- \left[\sum_{(i,a_i,d_i)} \frac{\partial^2}{\partial\theta\partial\beta'} L\left(a_i, d_i, X, \beta^0, \theta_I\right)\right] \cdot \left(\beta - \beta^0\right) \|.$$

We then explicitly compute the derivatives on the right hand side of this inequality. (27) then follows if we apply the triangle inequality as well as assumptions $A3.$ and $A4.$

Hence, condition (ii) is satisfied. Finally, note that $\triangle X$ has bounded first moment. Thus, by the weak law of large numbers, condition (iii) is satisfied. So far, we have

verified that all the hypotheses in theorem 7.2 are satisfied. Let $G = \nabla_\eta g_0\left(\eta^0\right)$. By theorem 7.2,

$$\sqrt{N}\left(\widehat{\eta} - \eta^0\right) \xrightarrow{d} N\left(0, G^{-1}\Sigma G'^{-1}\right).$$

By partitioning the matrix $G^{-1}\Sigma G'^{-1}$ and selecting the block corresponding to the parameter θ, the conclusion of proposition 4 follows. ∎

References

Ackerberg, D., and G. Gowrisankaran (2006), "Quantifying Equilibrium Network Externalities in the Ach Banking Industry," Forthcoming in *Rand Journal of Economics.*

Andrews, D., 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.

Aradillas-López, A. (2007), "Semiparametric Estimation of a Simultaneous Game with Incomplete Information," Princeton University Working Paper.

Augereau, A., Greenstein, S., and M. Rysman (2006), "Coordination vs. Differentiation in a Standards War: 56K Modems," *Rand Journal of Economics*, 37, 887-909.

Aumann, R. (1974), "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics,* 1, 67-96.

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

Bajari, P., H. Hong, and S. Ryan (2007), "Identification and Estimation of a Discrete Game of Complete Information," University of Minnesota Working Paper.

Bajari, P., H. Hong, J. Kraner, and D. Nekipelov (2007), "Estimating Static Models of Strategic Interactions," University of Minnesota Working Paper.

Berry, S. (1992), "Estimation of a Model of Entry in the Airline Industry," *Econometrica,* 60, 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," Caltech Working Paper.

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

Bresnahan, T., and P. Reiss (1991), "Empirical Models of Discrete Games," *Journal of Econometrics,* 81, 57-81.

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Inference on Parameter Sets in Econometric Models," *Econometrica,* 75, 1243-1284.

Chwe, M. (2004), "Statistical Game Theory" Department of Political Science, U.C.L.A. Working Paper.

Ciliberto, F., and E. Tamer (2007), "Market Structure and Multiple Equilibria in Airline Markets," University of Virginia and Northwestern University Working Paper.

Hart, S., and D. Schmeidler (1989), "Existence of Correlated Equilibria," *Mathematics of Operations Research,* 14,18-25.

Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica,* 46, 931-959.

Manski, C., and E. Tamer (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica,* 62, 1349-1382.

Mazzeo, M. (2002), "Product Choice and Oligopoly Market Structure," *Rand Journal of Economics,* 32, 1-22.

McLennan, A., and R. McKelvey (1996), "The Computation of Equilibrium in Finite Games," *Handbook of Computational Economics*, Vol.1, 87-142, Elsevier.

Murty, K. (1976), *Linear and Combinatorial Programming,* John Wiley &Sons.

Myerson, R. (1991), *Game Theory,* Harvard University Press, Cambridge, MA.

Myerson, R. (1986), "Acceptable and Predominant Correlated Equilibria," *International Journal of Game Theory,* 15, 133-154.

Nau, R., S. Canovas, and P. Hansen (2003), "On the Geometry of Nash Equilibria and Correlated Equilibria," *International Journal of Game Theory,* 32, 443-453.

Newey, W., and D. McFadden (1994), "Large Sample Estimation," *Handbook of Econometrics,* Vol.4, 2111-2245, North-Holland, Amsterdam.

Pakes A., J. Porter, K. Ho, and J. Ishii (2006), "Moment Inequalities and Their Application," Harvard University Working Paper.

Politis, D., J. Romano, and M. Wolf (1999), *Subsampling,* Springer, New York.

Powell, J. (1986), "Censored Regression Quantiles," *Journal of Econometrics,* 32, 143-155.

Reiss, P. (1996), "Empirical Models of Discrete Strategic Choices," *American Economic Review,* 86, 421-426.

Seim, K. (2006), "Spatial Differentiation and Market Structure: the Video Rental Industry," *Rand Journal of Economics,* 37, 619-640.

Sposito, V. (1989), *Linear Programming with Statistical Applications*, Iowa State University Press, Iowa.

Tamer, E. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies,* 70, 147-165.

Vaart, A. (1999), *Asymptotic Statistics,* Cambridge University Press

Table 1: Monte Carlo Results

	Results
$N = 500$	
$\beta^2 = -1.0$	[-1.59, -0.23]
	[-2.20, 0.08]
$\delta = 5.5$	[4.98, 5.34]
	[4.21, 5.60]
$N = 1000$	
$\beta^2 = -1.0$	[-1.60, -0.88]
	[-2.10, -0.32]
$\delta = 5.5$	[4.31, 5.48]
	[5.06, 6.12]
$N = 2000$	
$\beta^2 = -1.0$	[-1.42, -0.93]
	[-1.63, 0.10]
$\delta = 5.5$	[5.15, 5.49]
	[5.28, 5.95]

Note: Since we have normalized β^1 to equal 1.0, we do not list its values in the table. The top interval corresponding to each parameter is its set estimate while the bottom interval is its 95% confidence interval.