TECHNICAL WORKING PAPER SERIES

IDENTIFICATION AND ESTIMATION OF DISCRETE GAMES OF COMPLETE INFORMATION

Patrick Bajari
Han Hong
Stephen Ryan

Technical Working Paper 301
http://www.nber.org/papers/T0301

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2004

We would like to thank seminar participants and Victor Aguirregabiria, Lanier Benkard, Jeremy Fox, Stephanie Houghton, John Krainer, Andy McLennan, Andrew Sweeting and Chris Timmins for helpful discussions. This research was funded in part by National Science Foundation grants to Bajari and Hong. The views expressed in this paper are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

© 2004 by Patrick Bajari, Han Hong, and Stephen Ryan. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Identification and Estimation of Discrete Games of Complete Information
Patrick Bajari, Han Hong, and Stephen Ryan
NBER Technical Working Paper No. 301
September 2004
JEL No. L0, L2, C1, C7

ABSTRACT

We discuss the identification and estimation of discrete games of complete information. Following Bresnahan and Reiss (1990, 1991), a discrete game is a generalization of a standard discrete choice model where utility depends on the actions of other players. Using recent algorithms to compute all of the Nash equilibria to a game, we propose simulation-based estimators for static, discrete games. With appropriate exclusion restrictions about how covariates enter into payoffs and influence equilibrium selection, the model is identified with only weak parametric assumptions. Monte Carlo evidence demonstrates that the estimator can perform well in moderately-sized samples. As an application, we study the strategic decision of firms in spatially-separated markets to establish a presence on the Internet.

Patrick Bajari
Department of Economics
Duke University
305 Social Science Building
Box 90097
Durham, NC 27708-0097
and NBER
bajari@econ.duke.edu

Han Hong
Duke University
han.hong@duke.edu

Stephen Ryan
Duke University
stephen.ryan@duke.edu

1 Introduction.

In this paper, we study the identification and estimation of static, discrete games of complete information. These games generalize standard discrete choice models to allow utility to also depend on the actions of other players. Examples of discrete games studied in the literature include Vuong and Bjorn (1984), Bresnahan and Reiss (1990,1991), Berry (1992), Seim (2001), Ackerberg and Gowrisankaran (2002), Mazzeo (2002), Tamer (2002), Ciliberto and Tamer (2003), Aguirregabiria and Mira (2002), Berry, Ostrovsky and Pakes (2003), Pesendorfer and Schmidt-Dengler (2003), Manuszak and Cohen (2004), Sweeting (2004) and Bajari and Krainer (2004).

The literature has considered both discrete games of complete and incomplete information. In incomplete information games the random preference shocks are private information, while in complete information games they are common knowledge. Games of incomplete information can often be estimated using a straightforward, two-step approach. (See Aguirregabiria and Mira (2003), Berry, Ostrovsky and Pakes (2003), Pesendorfer and Schmidt-Dengler (2003) and Bajari, Benkard and Levin (2003).) Games of complete information have proved much more difficult to estimate. To the best of our knowledge, there is no existing method that can be applied to discrete games with an arbitrary number of players and strategies. Also, little is known about conditions under which discrete games are identified beyond the case of certain entry games.[2]

[2] There have been three approaches for solving these problems in practice. First, Tamer (2002) and Ciliberto and Tamer (2003) propose bounds estimators for certain entry games. A second approach is to exploit the structure of a specific class of games, such as

In this paper, we study identification and estimation of discrete games allowing for both multiple and mixed strategy equilibria. We propose a simulation-based estimator for discrete games of complete information. The model primitives include the latent utilities and an equilibrium selection mechanism which determines the probability that a particular equilibrium to the game is played. Using these primitives, we define a Method of Simulated Moments (MSM) estimator. We exploit recent algorithms that compute all of the equilibria to discrete games which are included in the publicly available software package Gambit (see McKelvy and McLennan (1996)). Our estimators are computationally efficient and can be programmed easily using Gambit and a standard optimization toolbox. We also show that the models we study are identified under fairly weak exclusion restrictions on the latent utilities and the equilibrium selection mechanism. We provide some Monte Carlo evidence that our estimator works well even with moderately size samples. Finally, we provide an application of our estimator. We study the strategic decision of small businesses to go online by establishing a web page.

This paper makes several contributions to the literature. First, our estimator can be applied to games with a flexibly specified structure of players and strategies. Previous methods exploited the structure of specific types of games (e.g. entry games) in order to make estimation feasible. Arbitrarily large games may not be computationally feasible using our approach, due to the burden of computing all of the equilibria. However, our approach allows us to estimate a more general set of games than previous methods. The algorithms that we propose use new numerical techniques that generate two computational savings. First, they are parallel and the most computationally intensive steps can be run on separate processors. Second,

---

entry games, in order to make point estimation feasible. See, for example, Bresnahan and Reiss (1990,1991) and Berry (1992). Finally, some researchers directly parameterize the equilibrium selection process. Imrohoroglu (1993) estimates a money demand model following Sargent and Wallace (1987), who show, under a number of restrictions, how to uniquely index the continuum of equilibria with a finite vector of parameters. This strategy is also used in related contexts by Moro (2003), Ackerberg and Gowrisankaran (2002), and Sweeting (2004).

it is not necessary to recompute the equilibria to the games within the estimation procedure.

The second contribution is that we allow for both pure and mixed strategy equilibria. To the best of our knowledge, all previous studies ignore the mixed strategy equilibria to discrete games. Considering mixed strategy equilibria is important for two reasons. First, mixed strategy equilibria are a generic feature of discrete games, and some games have no equilibria in pure strategies. If an estimator does not allow for mixed strategies, the econometric model will be ill-defined since it will not yield *any* prediction for this set of games. Second, if mixed strategies are empirically important, which our application suggests they are, then any estimator which does not admit their existence will be biased.

A third contribution is that we formally model the equilibrium selection process. Most previous papers estimate the utility parameters but not the equilibrium selection mechanism. (See Bresnahan and Reiss (1990,1991), Seim (2001), Mazzeo (2002), Tamer (2002), Ciliberto and Tamer (2003), Aguirregabiria and Mira (2002), Berry, Ostrovsky and Pakes (2003), Pesendorfer and Schmidt-Dengler (2003), Manuszak and Cohen (2004), and Bajari and Krainer (2004).) Games typically generate multiple equilibria and theory provides little guidance on which equilibrium is most plausible. It is not possible to simulate a game without specifying, implicitly or explicitly, a model of equilibrium selection. Our paper therefore allows us to consider counterfactual simulations that are not possible using other approaches.

Fourth, we derive new, sufficient conditions for identification. Previous work, such as Bresnahan and Reiss (1990,1991), establishes that without restrictions on the game, it is not possible to identify utilities from the observed actions. We present two sets of sufficient conditions for identification. First, if some variable influences equilibrium selection but can be excluded from the utilities, then identification is possible. Second, the model is identified if there are covariates that 1) shift the utility of an individual player, but are excluded from the utilities of other players and 2) these covariates are excluded from the equilibrium selection mechanism. We discuss some examples of exclusion restrictions that might be found in applications. Our identification results are established under weak parametric assumptions. While our estimation approach is parametric, these results suggest that the identification does not hinge on ad hoc functional form restrictions. In moderate sample sizes, however, a parametric method provides a more feasible estimation approach.

Our identification strategy is closely related to approaches found in treatment effect and sample selection models. The probability that a particular equilibrium is played is analogous to the "selection equation" and the equation that determines utility to the "treatment equation". In sample selection models, it is well known that identification under weak functional form assumptions often requires an exclusion restriction (see Heckman (1990)). Exclusion restrictions are required to identify models that are much simpler than ours. It necessarily follows that equally, or even more, stringent assumptions will be required in our more complicated models.

The identification results discussed in this paper can be extended to other games. For instance, Bajari and Krainer (2004) use similar identification arguments in a discrete game with incomplete information. Also, the methods we propose could be useful for identifying peer effects. A "peer effect" implies that agents have an incentive to act in accordance with norm behavior. If we assume that people are rational economic actors, then the presence of a peer effect implies that a game is being played. We believe that our identification results shed some new light on these problems.

As an application of our estimator, we study the decision by small businesses to create a presence on the Internet by posting a web page. The specific industry we study is golf courses in the Carolinas. Many golfers learn about the characteristics of a course (e.g. price, difficulty and location) by browsing the web. Since there are only a handful of golf courses operating in most markets, the decision to create a web page is naturally modeled as a game. Golf courses are representative of many small businesses. They are primarily owner-operated small businesses that function in geographically separated markets. This industry is an attractive case study of an important segment of the Internet.

Our application allows us to learn whether the adoption decisions of other firms are strategic substitutes (lower the payoffs from adoption) or strategic complements (raise the payoffs from adoption). Also, our estimation allows us to infer what types of equilibria are most likely to be played. In particular, our results suggest that mixed strategy and efficient equilibria are more likely in the data.

2 The Model.

The model is a simultaneous move game of complete information (normal form game). There are $i = 1, ..., N$ players, each with a finite set of actions A_i. Define $A = \times_i A_i$ and let $a = (a_1, ..., a_N)$ denote a generic element of A. Player i's von Neumann-Morgenstern (vNM) utility is a map $u_i : A \to R$, where R is the real line. Let π_i denote a mixed strategy over A_i. A Nash equilibrium is a vector $\pi = (\pi_1, ..., \pi_N)$ such that each agent's mixed strategy is a best response.

Following Bresnahan and Reiss (1990,1991), assume that the vNM utility of player i can be written as:

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

We will sometimes abuse notation and write $u_i(a)$ instead of $u_i(a, x, \theta_1, \varepsilon_i)$. In equation (1), i's vNM utility from action a, $u_i(a)$, is the sum of two terms. The first is a function $f_i(x, a; \theta_1)$ which depends on a, the vector of actions taken by all of the players, the covariates x, and parameters θ_1. The second is $\varepsilon_i(a)$, a random preference shock. Note that the preference shocks depend on the entire vector of actions a, not just the actions taken by player i. In much of the literature, it is assumed that the stochastic shocks are only a function of player i's own actions. The framework that we propose allows for more general preference shocks. The $\varepsilon_i(a)$ are assumed to be i.i.d. with a density $g(\varepsilon_i(a)|\theta_2)$ across i, a and observations in the data. In principal, the i.i.d. assumption can be weakened. We could easily modify our estimator to allow the $\varepsilon_i(a)$ to only depend on the actions of i or to drop the independence assumption.[3] We discuss the independence assumption in more detail in our section on identification.

The term $\varepsilon_i(a)$ reflects information about utility that is common knowledge to the players, but not observed by the econometrician. In games where there are a small number of players who know each other well, they will observe important information about each other that is not observed to the econometrician. The $\varepsilon_i(a)$ captures this information. Under these circumstances, the assumption of perfect information would be more reasonable approximation than private information.

Aguirregabiria and Mira (2003), Berry, Ostrovsky and Pakes (2003), Pesendorfer and Schmidt-Dengler (2003), and Sweeting (2004) assume that $\varepsilon_i(a)$ is private information. The private information assumption is appropriate if what player i knows about player j's payoffs can be completely captured by f_j. We note, however, that private information games also can generate multiple equilibria (see McKelvy and Palfrey (1995), Brock and Durlauf (2001,2003), and Sweeting (2004)). If one wishes to use private information games to predict counterfactual outcomes, one will also need to account for the multiplicity of equilibria. We are currently exploring this extension.[4]

[3] If $\varepsilon_i(a)$ has full support, then the probability that any game u is drawn will be positive. Therefore, the support of the likelihood function is A for all parameter values and covariates. If $\varepsilon_i(a)$ only depends on i's actions, a_i, the likelihood may not have full support, A. This may lead to a severe specification problem since the model could predict that some events may have zero probability.

[4] An intermediate but unexplored case is when there are both commonly observed and private shocks to payoffs. For instance,

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

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

Theorists have suggested that an equilibrium may be more likely to be played if:

1. The equilibrium satisfies a particular refinement (e.g. trembling-hand perfection).
2. The equilibrium is in pure strategies.
3. The equilibrium is risk dominant.

The specifics of a particular application might also suggest factors which favor some equilibria. For instance, Berry (1992) and Ciliberto and Tamer (2003) suggest an equilibrium could be more likely if it maximizes industry profits or the profits of the largest incumbent firm. Given u and $\mathcal{E}(u)$ we could create dummy variables for whether a given equilibrium, $\pi \in \mathcal{E}(u)$ satisfies any of these criteria. Let $y(\pi, u)$ denote a vector of variables that we generate in this fashion. For instance,

$\varepsilon_i(a) = \eta_i(a) + \zeta_i(a)$ where $\eta_i(a)$ is commonly observed by the players, but not the econometrician, and $\zeta_i(a)$ is private information. This would generate a private information game with unoberved heterogeneity.

$$
\begin{aligned}
y_1(\pi, u) &= \begin{cases} 1 \text{ if } \pi \text{ is trembling-hand perfect} \\ 0 \text{ otherwise} \end{cases} \\
y_2(\pi, u) &= \begin{cases} 1 \text{ if } \pi \text{ is a pure strategy equilibrium} \\ 0 \text{ otherwise} \end{cases} \\
y_3(\pi, u) &= \left(\sum_i \sum_a \pi(a) u_i(a) \right) - \widehat{u} \\
\text{where } \widehat{u} &= \max \left\{ \sum_i \sum_a \pi'(a) u_i(a) | \pi' \in \mathcal{E}(u) \right\}.
\end{aligned} \tag{2}
$$

A parsimonious, parametric model of λ is:

$$
\lambda(\pi; \mathcal{E}(u), \beta) = \frac{\exp(\beta \cdot y(\pi, u))}{\sum_{\pi' \in \mathcal{E}(u)} \exp(\beta \cdot y(\pi', u))} \tag{3}
$$

Note that in (3) the sum runs over the distinct elements of the equilibrium set $\pi' \in \mathcal{E}(u)$. For each π', we calculate the vector $y(\pi, u) = (y_1(\pi, u), y_2(\pi, u), y_3(\pi, u))$ as in (2). Then we evaluate the standard logit formula where β weights the probability that a particular type of equilibrium is selected. The example above is meant to be a simple illustration of what a selection mechanism might look like in practice. It is easy to generalize λ to allow for a less restrictive functional form and a richer set of variables $y(\pi, u)$.

Computing the set $\mathcal{E}(u)$, all of the equilibrium to a normal form game, is a well studied problem. McKelvy and McLennan (1996) survey the available algorithms in detail. The free, publicly available software package, Gambit, has routines that can be used to compute the set $\mathcal{E}(u)$ using these methods.[5] Finding all of the equilibria to a game is not a polynomial time computable problem. However, the available algorithms are fairly efficient at computing $\mathcal{E}(u)$ for games of moderate size. Readers interested in the details of the

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

algorithms are referred to McKelvy and McLennan (1996). In the next sections, we shall take the ability to compute $\mathcal{E}(u)$ as given.

2.1 Examples of Discrete Games.

The model that we propose is quite general and could be applied to many discrete games considered in the literature. We discuss three examples. The first example is static entry games (see Bresnahan and Reiss (1990,1991), Berry (1992), Tamer (2002), Ciliberto and Tamer (2003), and Manuszak and Cohen (2004)). In applications of entry games, the economist observes a cross section of markets. The players in the game are a finite set of potential entrants. In each market, the potential entrants simultaneously choose whether to enter. Let $a_i = 1$ denote the decision to enter the market and $a_i = 0$ denote the decision not to enter the market. In applications, the function f_i takes a form such as:

$$f_i = \begin{cases} \theta_1 \cdot x + \delta \sum_{j \neq i} 1\{a_j = 1\} \text{ if } a_i = 1 \\ 0 \text{ if } a_i = 0 \end{cases} \tag{4}$$

In equation (4), the mean utility from not entering is set equal to zero.[6] The covariates x are variables which influence the profitability of entering a market. These might include the number of consumers in the market, average income and market specific cost indicators. The term δ measures the influence of j's choice on i's entry decision. If profits decrease from having another firm enter the market then $\delta < 0$. The $\varepsilon_i(a)$ capture shocks to the profitability of entry that are commonly observed by all firms in the market, but which are unobserved to the econometrician.

A second example is network effects, as in Ackerberg and Gowrisankaran (2002). They consider the decision by banks in spatially separated markets to adopt the Automated Clearing House (ACH) payment

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

system. The players are the existing banks in some market. Let $a_i = 1$ denote a decision to adopt ACH and $a_i = 0$ denote non-adoption. A priori, network effects are likely since if one bank in the market adopts ACH, this should raise the benefits to another bank of also adopting the system. A simple model of network effects might take the form:

$$f_i = \begin{cases} \theta_1 \cdot x_i + \delta \sum_{j \neq i} 1\{a_j = 1\} \cdot c_j \cdot c_i \text{ if } a_i = 1 \\ 0 \text{ if } a_i = 0 \end{cases} \tag{5}$$

In equation (5), x_i denotes some factors which influence the costs and benefits to adoption by firm i. For instance, i might have large corporate or government agencies as customers who require access to ACH. The variables in x_i therefore might include characteristics of the customer base. The term c_i is the current number of clients of bank i. If $\delta > 0$, the term $\delta \sum_{j \neq i} 1\{a_j = 1\} \cdot c_j \cdot c_i$ indicates that the marginal benefits to adopting are larger 1) as the number of clients of bank i, c_i increases and 2) as the number of clients of each adopting bank j increases. Since profits depend on the number of other banks that have adopted, we say that a network effect is present. The term $\varepsilon_i(a)$ captures benefits to adoption observed by the banks but unobserved by the economist.

A third example is peer effects (see Manski (1993) and Brock and Durlauf (2001,2003)). A peer effect connotes a situation in which there is a desire to conform to the norm. Consider the decision by a high school student of whether to take calculus during his senior year. The players in the game are students in a particular grade. Let $a_i = 1$ denote a decision to take calculus and $a_i = 0$ to take an easier math course (e.g., algebra). The utility of student i is:

$$f_i = \begin{cases} \theta_1 \cdot x_i + \delta \sum_{j \neq i} 1\{a_j = 1\} \cdot t_j \text{ if } a_i = 1 \\ 0 \text{ if } a_i = 0 \end{cases} \tag{6}$$

In equation (6), the covariates x_i could include terms that shift the incentives to take calculus. For instance, if i's parents have a college degree, they might encourage calculus more strongly. The term t_i denotes the score of student i on a standardized achievement test. The term $\delta \sum_{j \neq i} 1\{a_j = 1\} \cdot t_j$ models the peer effect. Here, i's decision depends on the average test score of other students who take calculus.[7]

The modeling framework we propose could be applied beyond these three examples. In principal, the framework above could be used to model any discrete choice where 1) the decisions of agents are interdependent, 2) decisions are made simultaneously and 3) there is complete information. If the number of players or actions is very large, our estimator may not be computationally feasible. However, computationally lighter estimators can be formed by assuming that the $\varepsilon_i(a)$ are private information. Our identification results can be extended to these models in a straightforward way. (See Bajari and Krainer (2004) for a discussion.)

3 Estimation.

Next, we propose a computationally efficient MSM estimator for θ and β. Let $P(a|x, \theta, \beta)$ denote the probability that a vector of strategies, a, is observed conditional on x, θ and β. Using the definitions from the previous section:

$$P(a|x, \theta, \beta) = \int \left\{ \sum_{\pi \in \mathcal{E}(u(a,x,\theta,\varepsilon))} \lambda\left(\pi; \mathcal{E}(u(a, x, \theta_1, \varepsilon)), \beta\right) \left(\prod_{i=1}^{N} \pi(a_i) \right) \right\} g(\varepsilon|\theta_2) d\varepsilon \qquad (7)$$

In equation (7), we compute $P(a|x, \theta, \beta)$ as follows. Given a realization of random preference shocks, $\varepsilon = (\varepsilon_1, ..., \varepsilon_N)$, we use equation (1) to compute the utilities, which we denote as $u(a, x, \theta, \varepsilon)$ to emphasize the dependence of the vNM utilities on the parameters, covariates and preference shocks. This determines

[7] Games with very large numbers of peers might be computationally infeasible. However, our analysis offers insight into the identification of peer effects. Computationally lighter estimators can be formed by assuming that the $\varepsilon_i(a)$ are private information. See Brock and Durlauf (2001,2003) for example.

the set of equilibria $\mathcal{E}(u(a, x, \theta, \varepsilon))$. Since this is generically a finite set, we sum over the equilibrium $\pi \in \mathcal{E}$ and compute 1) $\lambda(\pi)$, the probability that the equilibrium π is selected and 2) $\prod_{i=1}^{N} \pi(a_i)$, the probability that a is observed given π.

Following a suggestion by Ackerberg (2003), we change the variable of integration to u in order to reduce the computational burden of simulating (7). Let $h(u|\theta, x)$ denote the density for the utility vector u, conditional on θ and x. In many models, this is trivial to compute. For instance, suppose that the preference shocks $\varepsilon_i(a)$ are i.i.d. normal with density $\phi(\cdot|\mu, \sigma)$, with mean $\mu = 0$ and standard deviation σ. Then, the density $h(u|\theta, x)$ is:

$$h(u|\theta, x) = \prod_{i} \prod_{a \in A} \phi(u_i(a) - f_i(x, a; \theta_1)|0, \sigma) \tag{8}$$

If we change the variable of integration from ε to u, then (7) becomes:

$$P(a|x, \theta, \beta) = \int \left\{ \sum_{\pi \in \mathcal{E}(u)} \lambda\left(\pi; \mathcal{E}(u), \beta\right) \left(\prod_{i=1}^{N} \pi(a_i) \right) \right\} h(u|\theta, x) du \tag{9}$$

In the MSM estimator, we simulate the integral $P(a|x, \theta, \beta)$ using importance sampling. We rewrite the integral (9) as

$$P(a|x, \theta, \beta) = \int \left\{ \sum_{\pi \in \mathcal{E}(u)} \lambda\left(\pi; \mathcal{E}(u), \beta\right) \left(\prod_{i=1}^{N} \pi(a_i) \right) \right\} \frac{h(u|\theta, x)}{q(u|x)} q(u|x) du \tag{10}$$

In (10), we have modified the integral by adding the "importance density" $q(u|x)$. In importance sampling, for a given x, we draw a sequence $u^{(s)} = (u_1^{(s)}, ..., u_N^{(s)})$, $s = 1, ..., S$ of random utilities from $q(u|x)$. We can then simulate $P(a|x, \theta, \beta)$ as follows:

$$\widehat{P}(a|x,\theta,\beta) = \frac{1}{S}\sum_{s=1}^{S}\left\{\sum_{\pi\in\mathcal{E}(u)} \lambda(\pi;\mathcal{E}(u^{(s)}),\beta)\left(\prod_{i=1}^{N}\pi(a_i)\right)\right\}\frac{h(u^{(s)}|\theta,x)}{q(u^{(s)}|x)} \tag{11}$$

Under standard regularity conditions, $\widehat{P}(a|x,\theta,\beta)$ will be a unbiased estimator of $P(a|x,\theta,\beta)$. An importance density q that matches h closely will reduce the variance of $\widehat{P}$. Choosing an importance density carefully is critical in practice. In our Monte Carlo and application sections, we discuss this problem in detail.

3.1 The Estimator

The econometrician observes a sequence (a_t, x_t) of actions and covariates, $t = 1, ..., T$. Equation (11) can be used to form a maximum simulated likelihood estimator (MSL) for these observations. The simulated log likelihood function $\widehat{L}(\theta,\beta)$ is:

$$\widehat{L}(\theta,\beta) = \sum_t \log\left(\widehat{P}(a_t|x_t,\theta,\beta)\right)$$

where $\widehat{P}(a|x,\theta,\beta)$ is defined in equation (11).

As is well known, MSL is biased for any fixed number of simulations. In order to obtain $\sqrt{T}$ consistent estimates, one needs increase the number of draws S so that $\frac{S}{\sqrt{T}} \to \infty$.[8]

Alternatively, one can estimate the parameters using MSM. An advantage of MSM is it generates an unbiased and consistent estimator for a fixed value of S. To form the MSM estimator, enumerate the elements of A from $k = 1, ..., \#A$. Note that, because the probabilities of all of the elements of $a \in A$ must sum to one, one of these probabilities will be linearly dependent on the others, so there are effectively

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

$\#A-1$ conditional moments. Let $w_k(x)$ be a vector of weight functions, with dimension larger than the number of parameters, for each k and let $1(a_t = k)$ denote the indicator function that the tth vector of actions is equal to k. The function $P(k|x,\theta,\beta)$ denotes the probability that the observed vector of actions is k given x and the parameters θ and β. This is probability is defined in equation (7). For each k,

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

A set of moment conditions can be formed by taking the sample analog

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

In practice, $P(k|x_t,\theta,\beta)$ is evaluated by simulation using the importance sampler (11). For each x_t, we draw a vector of S simulations $u_t^{(s)}$, $s = 1, ..., S$ from the importance density $q(u|x)$. We assume that the simulation draws $u_t^{(s)}$ are independent over both t and s, and are independent of all x_t's. The moment conditions are then replaced by the simulation analog:

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

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

$$\left(\widehat{\theta},\widehat{\beta}\right) = \arg\min_{(\theta,\beta)} m_T(\theta,\beta)' \times W_T \times m_T(\theta,\beta). \tag{12}$$

The asymptotic theory for estimating discrete choice models using MSL and MSM are well developed. See McFadden (1989), Pakes and Pollard (1989) or Hajivassiliou and Ruud (1994) for a detailed discussion.

3.2 Discussion.

An important insight in Ackerberg (2003) is that the problem of simulating (11) can be broken into three steps.

1. Draw a large set of random games from the importance density $q(u^{(s)}|x)$.
2. Compute the set of equilibria, $\mathcal{E}(u^{(s)})$, from all games using Gambit or by using the algorithms described in McKelvy and McLennan (1996).
3. Evaluate the sum (11).

In an MSM estimator, it is necessary to evaluate expressions such as (11) a large number of times. However, once the equilibrium sets $\mathcal{E}(u^{(s)})$ have been precomputed, *it is not necessary to repeat steps 1 and 2 when evaluating (*11*)*. To see this, note first that $u^{(s)}$ does not depend on θ and β. Therefore the sets are independent of the model parameters. The parameter β only enters in the function $\lambda(\pi; \mathcal{E}(u^{(s)}), \beta)$. As in (3), changing the value of β only involves changing the parameters that weight the characteristics of the equilibrium. The parameters θ only enter into $h(u|\theta, x)$. Changing the values of θ will typically only involve evaluating a density such as (8) at new values.

In practice, we "precompute" the equilibrium sets $\mathcal{E}(u^{(s)})$ and the minimize the simple, parametric function (12). The step of precomputing the equilibrium sets, $\mathcal{E}(u^{(s)})$, is naturally parallel. If the economist has access to separate processors, he can send a subset of the games to each processor. In practice, computing the equilibrium is the most burdensome step of the computation. The computational time required to estimate the model is therefore roughly proportional to the inverse of the number of processors that the economist can exploit.

An appropriate choice of an importance density can improve the performance of the MSM estimator considerably. In our application, we found our importance density by first estimating the game assuming that the preference shocks in (1) are private information.[9] For each value of x_t the economist then simulates

[9] This estimator can be performed in two stages. In the first stage, the economist flexibly estimates the choice probabilities $P(a|x)$ using standard methods. In the second stage, the economist assumes that these estimated choice probabilities represent the agent's equilibrium expectations. These choice probabilities are then substituted into the utility function. The logic of this estimator is similar to Aguirregabiria and Mira (2002), Berry, Ostrovsky and Pakes (2003), Pesendorfer and Schmidt-Dengler (2003) and Bajari, Benkard and Levin (2004). Bajari and Krainer (2004) present a simple, static example of how to use these estimators.

the private information version of the model S times at the estimated parameter values. This generates $T \cdot S$ pseudorandom values of the latent utilities.

4 Monte Carlo.

The following Monte Carlo illustrates the use of the estimator and demonstrates that, at least in our example, it performs well in small samples. Consider a game of three players who must decide whether or not to enter a single market. The firms have two actions, $a_i = 0$ if they do not enter and $a_i = 1$ if they do. The payoffs from entering the market are given by:

$$u_i(a_i = 1) = \theta_1 - \theta_2(\sum_{j \neq i} 1(a_j = 1)) + \theta_3 x_1 + \theta_4 x_2 + \varepsilon_i(a), \tag{13}$$

where $1(\cdot)$ is the indicator function. The payoffs consist of a constant benefit of entry, θ_1, a payoff associated with the number of other firms in the market[10], θ_2, two market covariate shifters, and an action-specific error term. We interpret the two market-level covariates as population and average income, respectively. It is important to note that the error term is different for every profile of actions, as this is necessary to generate the full range of strategies in equilibrium.[11]

The payoff for staying out of the market is simply:

$$u_i(a_i = 0) = \varepsilon_i(a). \tag{14}$$

The $\varepsilon_i(a)$ are drawn from a standard normal distribution. The vector of market covariates are drawn independently from a uniform $U[0, 10]$ distribution.

The payoff structure generates games with many equilibria, so it is necessary to specify a process of

[10] This parameter could be generically positive or negative, say for a network externality or competition penalty, respectively.

[11] Restricting the error term to be firm-specific instead of action-specific generates dependence across the firm's payoffs. In this case many strategies become strictly dominated for any draw of ϵ, ruling out some types of equilibria.

equilibrium selection. We use a simple logit probability of selecting a given equilibrium:

$$\lambda(\pi_i; E(u), \beta) = \frac{\exp(\beta_1 \text{MIXED}_i)}{\sum_{\pi' \in E(u)} \exp(\beta_1 \text{MIXED}_i)}, \tag{15}$$

where MIXED_i is a dummy variable indicating if π_i is a mixed strategy equilibrium. This allows us to model whether or not mixed strategy equilibria are more likely to be played than pure strategy equilibria. Within the sets of pure and mixed strategies each strategy has an equal probability of being selected.

To perform the Monte Carlo, we created a large pool of games using the payoff structure described above, solved for their sets of equilibria in Gambit, and simulated an outcome for each game. For each run of the estimator, we drew a new data set from the pool of games and outcomes before running the estimation. The importance sampler games were also drawn from a pool of games generated with the same payoff structure as the observations. Following Ackerberg, we set the initial guess in the denominator of the importance sampler at the truth. We use an equal number of importance games and observations. The true parameters $(\theta_1, \theta_2, \theta_3, \theta_4, \beta_1)$ are (5, 1.5, 1.0, -1.0, 1.0). With these parameters the game can be thought of as the decision of three firms of whether or not to enter a market with a low-quality good. Firms prefer to have fewer competitors. Demand increases with population but decreases with average income. We set the equilibrium shifter to be positive to emphasize the ability of the estimator to deal outcomes generated by a mixed strategy. We ran the estimator 100 times on data sets of 25, 50, and 100 observations. The results are shown in Table 1.

The results are illuminating. The estimator performs well for even very small samples ($N = 25$). The decrease in the size of the confidence intervals is rapid as the number of observations increases for all parameters, especially for the equilibrium shifter. Interestingly, the equilibrium shifter is estimated as precisely as the utility parameters, despite the extra layer of uncertainty in its specification. Part of the intuition to explain this that the set of games generated by the true parameter vector are rich in multiple equilibria and mixed strategies. Each observation has additional structural requirements embedded in it that each of the parameters must satisfy. So even though the effect of the equilibrium selection shifter is indirect, it must satisfy a number of restrictions imposed by the game theory. This reinforces the idea that the restrictions of even a basic game like this one are substantial enough to allow the precise estimation of indirect parameters in the selection mechanism.

There is one caveat to our procedure that researchers have to address in practice. In the Monte Carlos, we knew the true parameters of the game, and we able to generate importance games using these. With real data, of course, these parameters are initially unknown. The importance sampler can generate biased parameter estimates if given poor initial guesses, so it is necessary to derive starting parameters from another source. We demonstrate this in our application by using a related game of private information.

5 Identification.

The estimation strategy we proposed above is parametric which is appropriate when sample sizes are small or there are many covariates.[12] Even if parametric methods are used, an estimation approach is more appealing if identification does not hinge on functional form assumptions. Therefore, in this section, we consider the nonparametric identification of our model.

5.1 Identifying Assumptions from Discrete Choice.

To begin with, we impose some common identifying assumptions used in the discrete choice literature. The model presented in Section 2 is a generalization of standard random utility models. Therefore, identifying assumptions required in these models will also be required here.

[12] We are currently pursuing nonparametric extensions to our estimation strategy.

The first identifying assumption we make is:

A1. For every i and $a_{-i} \in A_{-i}$, we let $f_i(\underline{a_i}, a_{-i}, x) = 0$ for some chosen $\underline{a_i} \in A_i$ and for all $a_{-i} \in A_{-i}$.

The rationale for A1 is similar to the argument that we can normalize the mean utility from the outside good equal to zero in a standard discrete choice model.[13]

A second assumption that we will make is:

A2. For every i and for every a, $\varepsilon_i(a)$ are distributed i.i.d. standard normal.

We could change A2 to allow $\varepsilon_i(a)$ to be any known joint parametric distribution. However, for expositional clarity, we shall assume that it has a standard normal distribution. Even in the simplest discrete choice models, it is not possible to identify both $f_i(a, x)$ and the joint distribution of the $\varepsilon_i(a)$ nonparametrically. Consider a standard binary choice model where the dependent variable is 1 if the index $u(x) + \varepsilon$ is greater than zero, i.e.

$$y = 1(u(x) + \varepsilon > 0) \tag{16}$$

All the population information about this model is contained in the conditional probability $P(y = 1|x)$, the probability that the dependent variable is equal to one given the covariates x. If the cdf of ε is G, then (16) implies that:

$$P(y = 1|x) = G(u(x)), \tag{17}$$

[13] Let π and π' be two arbitrary mixed strategies. In order to show that this assumption is without loss of generality, we must verify that it does not change i's ranking of π and π'.

Fix pure strategies $a^* = (a_1^*, ..., a_N^*)$ and $a' = (a_1', ..., a_N')$. Suppose that $u_i(a^*) \geq u_i(a')$. Then

$$f_i(a^*, x) + \varepsilon_i(a) \geq f_i(a_i, a_{-i}^*, x) + \varepsilon_i(a_i, a_{-i}^*)$$

This inequality will not be affected by subtracting $f_i(\underline{a_i}, a_{-i}, x)$ from both sides for all i. That is:

$$f_i(a^*, x) - f\left(\underline{a_i}, a_{-i}^*, x\right) + \varepsilon_i(a) \geq f_i(a_i, a_{-i}^*, x) - f\left(\underline{a_i}, a_{-i}^*, x\right) + \varepsilon_i(a_i, a_{-i}^*)$$

Hence assumption A1 does not change the ranking of pure strategies. Analogous arguments demonstrate that this normalization does not change the preference ordering over mixed strategies.

Obviously, only the composition of $G(u(x))$ can be identified, and it is necessary to make parametric assumptions on one part (e.g. G or u) in order to identify the other part. For instance, if G is the standard normal cdf, we could perfectly rationalize the observed moments (17) by setting $u(x)$ to the inverse cdf evaluated at $P(y = 1|x)$. Therefore, we will assume that the error terms are normally distributed.

We are also making the assumption that the $\varepsilon(a)$'s are independently distributed. This assumption is also required for the identification of single agent models if the function $f_i(a, x)$ is nonparametrically specified. For example, consider a simple single agent multinomial choice model with three options. Denote the possible choices as $a \in \{1, 2, 3\}$. For $a = 1, 2$ let

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

Also, the mean utility for the third option is normalized identically equal to 0:

$$u_i(3, x, \varepsilon_i) \equiv \varepsilon_i(3).$$

In the population, only two conditional probability functions are available to identify the model:

$$P\left(a = 1|x\right) \quad \text{and} \quad P\left(a = 3|x\right).$$

The last observable probability

$$P\left(a = 2|x\right),$$

is linearly dependent on the other two probabilities and does not help identification.

Holding x fixed, since there are only two moments, the unknowns $f_i\left(1, x\right)$ and $f_i\left(2, x\right)$ already exhaust the degrees of freedom available in the population. We can only hope to identify $f_i\left(1, x\right)$ and $f_i\left(2, x\right)$ by assuming that the *joint* distribution of the error terms is known. There are no additional degrees of freedom to identify the correlation structure between the error terms.

A multinomial probit model or a nested logit model allows one to estimate the correlation coefficients between $\varepsilon_i(1)$, $\varepsilon_i(2)$ and $\varepsilon_i(3)$. However, this comes at the cost of assuming a parametric functional form for the deterministic utility component $f_i(a, x)$.

5.2 Two by two games.

In what follows, it will be useful to consider simple examples in order to understand the main concepts. Therefore, we begin by analyzing the structure of equilibrium in a two player, two strategy game. To simplify notation, let the action sets of players 1 and 2 be denoted as

$$A_1 = \{T, B\}$$

and

$$A_2 = \{L, R\}.$$

We make the normalization A1 by setting $\underline{a_1} = T$ and $\underline{a_2} = L$. The resulting payoff matrix will take the form:

Payoff Matrix for Two-Player Game

	L	R
T	$(\varepsilon_1(TL), \varepsilon_2(TL))$	$(\varepsilon_1(TR), f_2(TR, x) + \varepsilon_2(TR))$
B	$(f_1(BL, x) + \varepsilon_1(BL), \varepsilon_2(BL))$	$(f_1(BR, x) + \varepsilon_1(BR), f_2(BR, x) + \varepsilon_2(BR))$

To analyze the structure of this game, we begin by observing that the set of equilibria $\mathcal{E}(u)$ can be characterized as follows:

Lemma 1. (2 by 2 Equilibrium). With probability one, the set of equilibrium is either unique or has three elements. If it has three elements (i) One equilibrium is in mixed strategies and (ii) In the two pure strategy equilibrium, no player plays the same strategy in both equilibria.

Proof: See Appendix.

The proof to the lemma above establishes that there are exactly two cases of possible multiplicity. If there are three equilibria, the possible sets of equilibria, $\mathcal{E}(u)$ are of the form:

1. $\{(T, L), (B, R)$, a mixed strategy equilibrium$\}$
2. $\{(T, R), (B, L)$, a mixed strategy equilibrium$\}$.

Next, we illustrate our approach for flexibly modeling λ, the equilibrium selection mechanism. We begin by making the following assumption about λ which we will maintain for the rest of the identification section:

A3. λ does not depend on the stochastic preference shocks ε.

This assumption is analogous to "selection based on observables" assumption in treatment effect and sample selection models. In these models, it is commonly assumed that the error terms in the treatment equation is independent of the error term in the selection equation. The implication of this assumption is that the treatment status is exogenous conditional on the observables and that the outcomes of the treated and untreated group can be compared conditional on observing the x's.

We will also require λ to have the following type of invariance. Consider two distinct sets of equilibrium $\mathcal{E}(u)$ and $\mathcal{E}(u')$. Suppose that the supports of the equilibria in $\mathcal{E}(u)$ and $\mathcal{E}(u')$ coincide in the following sense:

(i) For every $\pi \in \mathcal{E}(u)$ there is a $\pi' \in \mathcal{E}(u')$ with the same support.

(ii) For every $\pi' \in \mathcal{E}(u')$ there is a $\pi \in \mathcal{E}(u)$ with the same support.

Then, we will assume that $\lambda(\pi; x, \mathcal{E}(u)) = \lambda(\pi'; x, \mathcal{E}(u'))$ if π and π' have the same support. In words, this means that the measure λ only depends on the support of the elements in $\mathcal{E}(u)$, not the magnitude of the mixing probabilities. In the context of the two-player game, this would mean that if the equilibrium set was $\{(T, L), (B, R)$, a mixed strategy equilibrium$\}$, λ would always give the same probability to (T, L), (B, R) and the mixed strategy equilibrium conditional on x. If the mixing probabilities changed, this would not change the probability assigned to the mixed strategies. We will maintain these assumptions for the rest of the identification section and formalize this condition in the assumption below.

A4. Given x, λ only depends only on the support of the elements in $\mathcal{E}(u)$.

If we make assumption A3 and A4, then exactly four probabilities are required to parameterize the equilibrium selection process. We label these functions as

$$\lambda_1(x), ..., \lambda_4(x).$$

where:

1. If the equilibrium set is $\{(T,L),(B,R)$, a mixed strategy equilibrium$\}$, select (T,L) with probability $\lambda_1(x)$, (B,R) with probability $\lambda_2(x)$ and the mixed strategy equilibrium with probability $1-\lambda_1(x)-\lambda_2(x)$.
2. If the equilibrium set is $\{(T,R),(B,L)$, a mixed strategy equilibrium$\}$, select (T,R) with probability $\lambda_3(x)$, (B,L) with probability $\lambda_4(x)$ and the mixed strategy equilibrium with probability $1-\lambda_3(x)-\lambda_4(x)$.

Note that since there are two possible sets of equilibria (when there is multiplicity), we do not use $\lambda_3(x)$ and $\lambda_4(x)$ when the equilibrium set is $\{(T,L),(B,R)$, a mixed strategy equilibrium$\}$. Similarly, we do not use $\lambda_1(x)$ and $\lambda_2(x)$ if the equilibrium set is $\{(T,R),(B,L)$, a mixed strategy equilibrium$\}$. When the equilibrium is unique, the selection mechanism has no "bite" since only one outcome is possible.

Assumptions A3 and A4 allows us to characterize λ as a finite dimensional vector of parameters conditional on x. In our framework, the number of equilibria is finite with probability one because the ε's are drawn from an atomless distribution. Therefore, there are finitely many possibilities for which strategies can have positive support in equilibrium. In our two by two game, this allows us to characterize the selection mechanism with 4 parameters conditional on x. In more general games, by similar logic, one can always characterize the equilibrium selection mechanism using a finite number of parameters conditional on x. As we will show below, this simplifies our identification problem.

Lemma 2. Given A1-A4, λ can be characterized by a finite dimensional vector of parameters holding x fixed.

Given the results of Lemma 2, for the rest of this section we shall associate $\lambda(x)$ with a finite vector of real numbers.

5.3 Identification: Definitions and Preliminaries.

A model is said to be identified if the model primitives can be recovered given the probability distributions the economist can observe. In a normal form game, the available population probabilities are $P(a|x)$ for $a \in A$, the probability distribution of the observed actions conditional on the covariates x. The primitives we wish to identify are $f(a, x)$ and $\lambda(x)$. That is, we wish to learn the vector of mean utilities $f(a, x) = (f_i(a, x))$ and $\lambda(x)$ without making parametric assumptions about these objects.

Our approach can be criticized as not being completely "nonparametric". Assumption A2 implies that we have made normality and independence assumptions about the error terms. We note, however, it is not possible to identify simpler single agent discrete choice models without this assumption. Assumptions A3 and A4 are also restrictions on $\lambda(x)$. However, it does permit considerably flexibility. To the best of our knowledge, we establish the only results in the literature that allow for identification of the selection mechanism under weak functional form assumptions. At a minimum, we hope that this will be a useful starting point for further work.

We can generalize equation (7) by writing $P(a|x)$ in a way that does not hinge on the specific parametric forms implicitly assumed in Section 2.

$$P(a|x, f, \lambda) = \int \left\{ \sum_{\pi \in \mathcal{E}(u(f,\varepsilon))} \lambda\left(\pi; E\left(u\left(f, \varepsilon\right), x\right)\right) \left(\prod_{i=1}^{N} \pi(a_i)\right) \right\} g(\varepsilon) d\varepsilon \tag{18}$$

In equation (18), we write the vNM utilities as $u(f, \varepsilon)$ to remind ourselves that they are a sum of the mean utilities $f(a, x)$ and the shocks ε. By Lemma 2, holding x fixed we can view (18) as a finite number of equations that depend on the finite number of parameters, $f(a, x)$ and $\lambda(x)$. Denote this system as $P(a|x) = H(f(x), \lambda(x))$ where H is the map implicitly defined by (18). When writing H, assume that we drop one choice probability for each player. Since choice probabilities add up to one, this introduces a linear dependence between the rows of this system. We will let $DH_{f,\lambda}(x)$ denote the Jacobian formed by differentiating H with respect to the parameter vectors $f(x)$ and $\lambda(x)$.

Definition. Given the probabilities $P(a|x)$, suppose that $f^0(a, x)$ and $\lambda^0(x)$ satisfy (18). We will say that $\left(f^0(a, x), \lambda^0(x)\right)$ are *locally identified* if there exists an open neighborhood N_x of $\left(f^0(a, x), \lambda^0(x)\right)$ such there is no other vector $\left(\widetilde{f}(a, x), \widetilde{\lambda}(x)\right) \in N_x$, $\left(\widetilde{f}(a, x), \widetilde{\lambda}(x)\right) \neq \left(f^0(a, x), \lambda^0(x)\right)$, that also satisfies (18).

In what follows, we shall often invoke the following assumption:

A5. The map H is continuously differentiable. Also suppose that the n_1 by n_2 Jacobian matrix $DH_{f,\lambda}$ has rank n_2.

Assumption A5 implies that for our system of implicit equations (18) we can check local identification by comparing the number of moments, $P(a|x)$, to the number of free parameters $(f(x), \lambda(x))$. If the number of moments is greater than the number of parameters, then the implicit function theorem implies that the parameters are locally identified. While we can directly verify assumption A4 for certain games (e.g. a 2 by 2 game), we cannot do so for general games. At a minimum, this would require us to characterize the different sets of all equilibrium that can be reached, analogously to lemma 1. This is not feasible in games with many players and strategies.

5.4 Identification: Negative Results.

The first result we establish is that even if the selection mechanism λ is known, in a two by two game, the four deterministic utility parameters cannot be nonparametrically identified.

Theorem 1 *In a game with two players and two strategies, if we make assumptions A1-A5, the deterministic utility components*

$$f_1(BL, x), f_1(BR, x), f_2(TR, x), f_2(BR, x)$$

are not identified from the distribution of $P(a|x)$ *even if the selection mechanism*

$$\lambda_1(x), \ldots, \lambda_4(x)$$

is known.

Proof: To begin with, consider the identification problem holding a given realization of x fixed. Since there are two players with two strategies, the econometrician observes four conditional moments,

$$P(TL|x),\ P(TR|x),\ P(BL|x), \text{ and } P(BR|x),$$

Since the probability of the actions must sum to one, there are effectively three moments that the econometrician observes. This leaves us with 4 utility parameters,

$$f_1(BL, x), f_1(BR, x), f_2(TR, x), f_2(BR, x)$$

to identify. Clearly, for a given realization of x we are not identified. Q.E.D.

The result above can easily be generalized to generic games. Conditional on x, the number of mean utility parameters is greater than the number of moments available to the econometrician.

Theorem 2 *In a game with more than two players and at least two strategies per player, if we make assumptions A1-A5, the deterministic utility parameters* $f_i(a, x)$ *are not identified from the distribution of* $P(a|x)$*, even if the selection mechanism* $\lambda(\cdot)$ *is known.*

Proof: Consider a game with N players and $\#A_i$ strategies for player i. Holding x fixed, the total number of mean utility parameters $f_i(a, x)$ is equal to

$$N \cdot \prod_i \#A_i - \sum_i \prod_{j \neq i} \#A_j.$$

This is equal to the cardinality of the number of strategies, times the number of players, minus the normalizations allowed by assumption A1. The number of moments that the economist can observe, conditional on x, is only equal to

$$\prod_i \# A_i - 1.$$

If each player has at least two strategies and if there are at least 2 players in the game, then for each given x the difference between the number of utility parameters, f_i, to estimate and the number of available moment conditions is bounded from below by

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

Clearly, we will be unable to identify the utility parameters of the model. Q.E.D.

6 Exclusion Restrictions.

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

6.1 Identification Exclusion restrictions in equilibrium selection

We begin by supposing that there are some variables z that influence λ but which can be excluded from f. That is,

A6. λ is a function of x and z, where z can be excluded from f_i for all i.

Before we formally consider the problem of identification, it is useful to ask what variables might be plausible z's in some applications. One potential source of such variables is regulation which is a commonly exploited source of exogeneity in treatment effect models. Consider the following simple example. Choosing whether to drive on the left or the right side of the road can be modeled as a coordination game. There are naturally 3 equilibrium to this game. Driving on the left, driving on the right, and the mixed strategy equilibrium. In practice, the side of the road that we drive on is strictly regulated; however, it differs across geography. For instance, one drives on the left in England and the right in the United States. Regulation influences the choice of equilibrium in many problems. Frequently, government or industry groups will encourage market participants to coordinate on a particular equilibrium, such as a particular standard for a technology with multiple standards.

A more specific example can be found in Bajari and Krainer (2004). The authors model the determination of analyst recommendations (e.g. strong buy, buy, hold, sell, etc...) for high technology stocks. Their choice of an average recommendation is fundamentally indeterminate as is the assignment of any type of grade. A grade of any sort is usually judged by making a comparison to a relevant set of peers. Just as college professors benchmark their grading practices against colleagues, research analysts have an incentive to benchmark their recommendations against the practices of their peers.

Since this is a coordination game, there can be multiple equilibria. Beginning in June of 2001, the

State Attorney General of New York, Elliot Spitzer, began to question business practices in this industry. Spitzer criticized investment banks for issuing a large fraction of strong buy and buy recommendations, but few hold or sell recommendations. The authors argue that this intervention by the regulator encouraged the industry to focus on an equilibrium where more “sell” and “hold” recommendations are issued. They divide time into pre and post “Spitzer” eras and view this time dummy as an instrument z. Also, they interact this dummy with the number of analysts engaged in investment banking to construct additional instruments. The logic is that stocks covered by a disproportionate share of analysts from investment banks have incentives to coordinate on an equilibrium with lower grades.

A second set of variables that could enter z include previous plays of the game or behavior in surrounding markets. In many economic problems, the selection of equilibrium could be history dependent. For instance, college professors, when determining how to grade, will base their expectations about which equilibrium to play based on grades submitted in previous years. Also, professors in the economics department may benchmark their grades against more general grading practices in the university.

In Bajari and Krainer (2004), the authors argue that there could be some history dependence in industry practices so that lagged behavior might influence the equilibrium that is selected. Therefore, the authors use regulation and lagged behavior as excluded variables that influence the equilibrium selection process but not the current payoffs. Constructing measures of lagged behavior and behavior in adjacent markets can be done in many applications. Whether or not these variables are valid exclusion restrictions, however, will be application specific.

In addition to A6, we also need to assume that:

A7. The covariates, x, can be partitioned into

$$x = (x_\lambda, x_u)$$

such that $\lambda(x,z)$ depends only on x_λ and not x_u:

$$\lambda(x,z) = \lambda(x_\lambda, z).$$

That is, there must be some payoff relevant variables x_u that are also excluded from the equilibrium selection mechanism. The plausibility of A7 must be judged on a case by case basis. In some cases, such as the example of submitting stock recommendations, there seemed to be industry wide norms for grading stocks that transcended individual stocks. For instance, by definition, a recommendation of "market underperform" is issued for a fairly small fraction of stocks. This is despite the fact that, by definition, 1/2 of the stocks must underperform the market mode! One interpretation of the hesitance to issue low recommendations is that this represents a norm among stock analysts. If the norm transcends the game played for an individual stock, then A7 might be a plausible approximation to what occurs in the data. This would imply that some payoff relevant covariates, x_u, can be excluded from equilibrium selection.

Theorem 3 *In the two by two game, suppose that A1-A7 are satisfied. Also suppose that (x_λ, x_u, z) takes on a discrete number of values and $\#x_u > 3$ and $\#z > 3$. Then the mean utilities $f_i(a,x)$ and the selection parameters $\lambda(x_\lambda, z)$ are locally identified.*

Proof: The number of moments generated by observable population conditional outcome probabilities, $P(a|x_\lambda, x_u, z)$ is

$$3 \times (\#x_\lambda) \times (\#x_u) \times (\#z), \tag{19}$$

Note that, in this equation, we multiply by 3 because the probabilities of the various actions must sum to one. The total number of parameters needed to characterize both the utility functions and the equilibrium selection probabilities is

$$4 \times (\#x_\lambda) \times (\#x_u) + 4 \times (\#x_\lambda) \times (\#z). \tag{20}$$

Alternatively, we can think for each each given x_λ, there are

$$3 \times (\#x_u) \times (\#z)\,, \tag{21}$$

conditional outcome probabilities and there are

$$4 \times (\#x_u) + 4 \times (\#z)\,, \tag{22}$$

parameters to estimate. It is clear that as long as $\#x_u > 3$ and $\#z > 3$,

$$3 \times (\#x_u) \times (\#z) > 4 \times (\#x_u) + 4 \times (\#z) \tag{23}$$

and the model is locally identified by the implicit function theorem. Q.E.D.

These results will extend beyond the 2 by 2 game. The key to the proof is that the number of moments depends on the product of $(\#x_u)\times(\#z)$ in equation (21) while the number of parameters is a linear function of $\#x_u$ and $\#z$ in (22).

Theorem 4 *In a general N player game, suppose that A1-A7 are satisfied. Also suppose that (x_λ, x_u, z) takes on a discrete number of values. If $\#x_u$ and $\#z$ are sufficiently large, the model is locally identified.*

Proof. In a more general game, the number of parameters required to characterize the selection mechanism will be specific to the number of players and the number of actions. However, holding x_λ and z fixed, by Lemma 2, it must be possible to characterize λ with a finite dimensional parameter vector. Since this vector depends on the supports of the elements in $\mathcal{E}(u)$, it is possible to create a bound on the size of this vector that is independent of x_λ and z. Let $\#\mathcal{E}$ denote this number. Holding x_u fixed, the number of vNM utilities is equal to $N \cdot \prod_i \#A_i - \sum_i \prod_{j\neq i} \#A_j$.

As in the previous proof, hold x_λ fixed. Then the number of parameters is bounded by above by:

$$\#\mathcal{E}\left(\#z\right)+\left(N\cdot\prod_{i}\#A_{i}-\sum_{i}\prod_{j\neq i}\#A_{j}\right)\left(\#x_{u}\right)$$

The number of moments is proportional to

$$\left(\#A-1\right)\times\left(\#x_{u}\right)\times\left(\#z\right).$$

The number of moments grows at a rate involving the product of $(\#x_u) \times (\#z)$ while the number of parameters is a linear combination of these terms. For sufficiently large $\#x_u$ and $\#z$, the number of moments is greater than the number of parameters. Hence, by the implicit function theorem and A5, the model is identified. Q.E.D.

In Theorem 4, we required the variables to live in a discrete set. However, if the variables are all continuous, identification follows from the arguments above trivially. While the requirements of this theorem are quite stringent, this is a very powerful result. If the economist is willing to assume that the exclusion restrictions in A6-A7 are valid, then all of the primitive parameters can be recovered under fairly weak parametric assumptions.

6.2 Identification using exclusions in payoffs.

A second approach to establishing identification is to search for covariates that shift the utility of agent i, but which do not enter as arguments into u_j $(j \neq i)$ or λ. In some applications, such covariates might be easier to find than variables z which shift the entire equilibrium as in the previous section.

A8. For each agent i, there exists some covariate, x_i that enters the utility of agent i, but not the utility of other agents. That is, i's utility can be written as $f_i(a, x, x_i)$. Furthermore, x_i can be excluded from λ.

Assumption A8 implies that there are agent i specific utility shifters. In the Bajari and Krainer (2004) example, this could include the amount of investment banking business done by the firm or other brokerage-specific covariates. Thus, A8 would imply that the investment banking work done by Merrill Lynch does not directly influence the utility of Goldman Sachs for issuing a particular recommendation. While this assumption is unlikely to be perfectly satisfied, to a first approximation it does seem reasonable. In the information released by prosecutors, firms were accused of biasing reports to favor their own investment banking clients. They were not alleged to bias their reports in favor of the clients of other firms.

Theorem 5 *Suppose that A1-A5 and A8 hold. If $\#x_i$ are sufficiently large, the model is nonparametrically (locally) identified.*

Proof: The proof follows similarly to the previous section. Hold x fixed. Consider a large, but finite number of values of x_i equal to K for each agent. Consider the all the distinct vectors of the form $x = (x_1, ..., x_N)$ that can be formed. The number will be equal to K^N. Consider the moments generated by these K^N distinct covariates. The number of moments is equal to $K^N \cdot \sum_i (\#A_i - 1)$. The number of mean utility parameters is equal to $\sum_i K(\#A_i - 1) \prod_{j \neq i} \#A$ plus the number of parameters required to characterize λ (which is independent of the x_i). Thus, the number of moments depends linearly on K but the but the number of moments grows exponentially with K. By choosing sufficiently large values for K the model is identified. Q.E.D.

The intuition behind the theorem above is quite simple. By shifting the individuals' utilities one at a time, it is possible to increase the number of moments at a faster rate than the number of parameters. Thus, identification of the model is possible. As in the model of the previous section, the theorem generalizes trivially to the case of continuous covariates.

7 Application.

As an application of our estimator, we consider the decision by small businesses to create a web page on the Internet. We follow recent work in empirical industrial organization by limiting the focus of our application to a single industry. The particular industry we study is golf courses in the Carolinas. Golf courses are typically owner-operated enterprises that to a first approximation compete in spatially separated markets. Also, many golfers will search for a course web page to learn about the price, course characteristics (slope, length, etc.) and the course location. The largest businesses in the U.S. were early and widespread adopters of the Internet. We would expect small businesses to be the marginal adopters. Golf courses, therefore, are a case study of an interesting segment of the Web.

7.1 Background.

Robert Metcalfe, who founded 3Com and helped develop the Ethernet protocol, stated what has become known as Metcalfe's Law: the usefulness of a network increases in the square of the number of users. The Law applies particularly well to information technologies, such as telephones, fax machines, and most recently, the Internet. Network dynamics are characterized by initially low adoption rates that increase dramatically once the network achieves a "critical mass" in the number of users.

The speed of this adoption varies by technology and cost, with progressively more complex technologies taking longer. The evolution of major communication media reflects this: radio was quickest, followed by television, with the Internet taking the longest. The foundation of the Internet has been around for decades, but it was only during the mid-1990's that it became a cultural and economic phenomenon. For instance, the value of retail goods sold online in 1990 was virtually zero. By 1998, it had reached \$4.9 million; four years later it had increased by an order of magnitude to \$44.2 million. This trend also holds true for manufacturing, wholesale goods, and service markets.

Like all products, network good markets are two-sided: there must be consumers of network goods, but there also must be providers. Clearly the consumer side of the Internet has reached critical mass and continues to grow and mature. The response to the Internet over the same time horizon by firms has been varied, both across and within industries. From the consumer's perspective, the Internet augments at least two aspects of a firm's operation: advertising and the delivery or procurement of goods. Almost every large firm in the world now has an informational web site, while most retailers also have a full-fledged electronic storefront.

Adoption of Internet technologies to smaller firms and fringe markets has been slower and more varied. Vacationing markets are a good example. There are literally thousands of possible destinations for consumers to choose from. While every town with a beach could choose to advertise itself separately from other towns, most do not do so. Instead, there are a number of web sites that aggregate information, utilizing their local expertise to help potential customers select among the large number of choices. However, the choice of whether to adopt an Internet web site or not has important strategic considerations. The number of competitors that have adopted a web site can be important for several reasons. First, choosing not to adopt while your competitors do can have negative demand consequences if consumers use web sites to help decide where they want to vacation. Second, firms with no competitors on the web can also face a negative supply-side network effect—consumers cannot find enough information about beaches in the area and will turn to other information outlets to help them decide.

For these reasons, it is important to model the decision of whether to adopt an Internet technology as the outcome of a strategic game. Our estimator provides a natural way of estimating games of this type, where there are a relatively low number of firms in regional markets that are deciding on a discrete action. We focus on the decision of golf courses in North Carolina and South Carolina to create and maintain their own web sites. Golf is a popular recreational activity, and the Carolinas are a popular destination for people looking for a golf-oriented vacation.

Some vacationers are going to decide on their vacation itinerary after consulting a number of golfing sources, including print guides, magazines, and the Internet. There are a large number of information aggregators in the online golf market, since choosing among the hundreds of golf courses is daunting. Every golf course in the Carolinas is covered in some form by one of these general golfing web sites. Each course has a generic listing with location, course rating, par, yardage, slope, and price. In addition, a number of golf courses have decided to create and maintain their own web sites. The purpose of these web sites is largely informational since most sites do not allow you to buy time at the course. However, the nature of the information is more specific and detailed than the information on the general golf web sites. Typically the web site has information about the course's history, its designer and layout, tee time policies and other regulations, and attempts to make the golf course seem as attractive as possible. This text is usually accompanied by flattering photographs of the fairways and greens under optimal weather conditions and lighting.

7.2 Data.

We construct a data set on the golf industry in North Carolina and South Carolina. We obtain the location of every golf course in the two states. Where possible, we collect information on a number of relevant covariates: course type (public, private, semi-private, resort, military), location, whether they have a web site or not, number of holes, par, yardage, course rating, slope, weekday and weekend prices, local population, median rent, median house value, and median household income. The data were collected using Internet golf guides to get course characteristics. Housing prices and income are from Census sources.

Good market definitions were critical given we model the decision to adopt a web site as the outcome of a strategic game among local competitors. Yahoo! Yellow Pages has a useful feature which gives distances between firms in the same category. We use this feature to define markets by chaining all courses that are within 10 miles of any other course in the market. As a result, some courses within the market will be further than 10 miles apart while others may be next to each other, but this definition produces a range of market sizes which track geographic features and population centers fairly well. For tractability in the estimation, we only consider markets with five or fewer golf courses.

Table 2 gives some summary statistics of the data set. The total number of golf courses is 261 and the number of markets is 124. The typical golf course in our data set has a par of 68, is about 6,000 yards, and has a weekend price of $27.26. The Carolinas offer a considerable variety to the golfer as can be seen in the variation in course characteristics. Table 3 displays the rates of web site adoption broken down by market size. In all of our market sizes, less than half of the firms choose to go online. While larger markets tend to have a higher adoption rate, the pattern is far from linear.

7.3 Results.

In the model, firms are assumed to play a simultaneous move game of whether to create a web page with other firms in the same local market. The decision to create a web page (or not) is denoted as $a_i = 1$ ($a_i = 0$). The utility of firm i for adopting the web page is assumed to be a linear function of the weekend price, the number of competing firms in the market, the population of the market, the average home price, and the average income. The mean utility for not adopting is normalized to zero. The weekend price of golf course i will be excluded from the adoption decision of other firms in the market. Theorem 5 demonstrates that if we also exclude this variable from the selection mechanism, the model would be identified under weak parametric assumptions.

We also parameterize the equilibrium selection rule. The equilibrium selection rule takes the form:

$$\lambda(e;\mathcal{E}(u)) = \frac{\exp\left(\beta_1 MIXED(e) + \beta_2\left(\overline{e}_{eff} - \overline{e}\right)\right)}{\sum_{e'\in\mathcal{E}(u)} \exp\left(\beta_1 MIXED(e') + \beta_2\left(\overline{e}_{eff} - \overline{e'}\right)\right)}$$

The term $MIXED(e)$ is a dummy variable equal to one if the equilibrium is in mixed strategies. The term $\overline{e}_{eff} - \overline{e}$ is the difference between the efficient equilibrium (in the sense of maximizing the sum of the players' utilities) less the sum of utilities in equilibrium e. Our model of equilibrium selection lets us ask whether more efficient equilibria to the game and pure strategy equilibria are more likely.

Good starting values are critical to the estimation procedure. To obtain these we use a two-step procedure. Consider a related game where the error term is private information. In this setting, each firm has expectations about the number of competitors it will face if it decides to enter. We model these expectations by regressing the number of actual adopters in a market on that market's characteristics. In the second step, we use this predicted number of competitors, along with firm-specific characteristics, in a probit model of adoption choice. We then used the regression coefficients from this model as our starting values. We found that the constant, number of competitors, and price of a round of golf on the weekend were the only statistically significant parameters. See Bajari and Krainer (2004) for the details of this estimator.

Starting with these values, we estimated the full model on the model using markets with two firms. This involved seeding the games from the importance density with the starting values from the first stage private information game, solving for their equilibria in Gambit, and then using these in the estimation procedure. The initial guess in the importance density, $q(u|x)$, is very important for maintaining numerical stability in the optimization procedure. If the current values of the parameters are far from the initial guess the optimization routine can behave pathologically. Also, the estimates are biased towards the starting values if the truth is far from the initial guess. Therefore after each full estimation, we re-seeded the importance density with the new parameters and re-solved the games in Gambit. After running the estimation procedure again, we checked to see if the new estimates were substantially different than the starting values. We iterated the re-seeding procedure until convergence was achieved; typically, this took only two full iterations of the estimation procedure. Once convergence was achieved for the small markets, we increased the sample size to include markets with up to five firms.

Once we had solved for the parameter estimates, we added market-level covariates and the equilibrium selection parameters to the basic model. We added population, median house value, and median house income to the utility specification for each firm. Since the firm makes the adoption decisions based on the opportunity cost of doing so, these covariates should be interpreted as marginal effects on the payoff of adopting a web site.

Table 4 shows the results of the estimation. The constant is negative, meaning there is a disincentive to adopt a web site absent all other effects. This captures the fact that adopting a web site is costly to the firm. Adoption is driven in large measure by the difference in prices across firms. High prices are associated with much higher rates of web site adoption. The price is almost certainly a proxy for the golf course's quality and a measure of how desirable the course is. Therefore it is intuitive that these firms may have the largest benefits from advertising their courses on a web site, as consumers are willing to travel further to play nicer courses. Higher quality courses also have the largest incentives to differentiate themselves from average or poor quality courses.

There are significant penalties associated with adopting a web site when your competition has also done so. This indicates that the potential network benefits from adopting with other producers are swamped by the diminished market exposure associated with having to share the online market. The market level demographic estimates indicate that they are not significant determinants in the decision to adopt a web site.[14] It appears that price soaks up most of the information about quality that would be relevant to the decision of whether to adopt or not.

The equilibrium selection parameters are particularly interesting. The coefficient on the MIXED dummy in the selection equation shifts the probability of selecting the set of mixed strategy equilibria over the set of pure strategy equilibria. Mixed strategy equilibria are more likely to be played than pure strategy equilibria, although the marginal effect is small.

The literature typically chooses to ignore the mixed strategy equilibria in empirical applications. Our results suggest that this assumption may be problematic. Not only are the mixed strategy equilibria slightly more likely, but there can also be a large number of these equilibria. For instance, in the five firm case, we draw several games from the importance sampler with over 30 equilibria, the majority of which are either partially or fully mixed. The unconditional probability of a mixed strategy using our estimated parameters is 0.701 with a standard deviation of 0.2197. Also, the importance sampler did draw examples of games that only equilibria in which at least one player mixed. An estimation method that did not allow for mixed strategies would be undefined in this case.[15]

---

[14] The market-level covariates were normalized to lie in between 0 and 100.

[15] For example, consider an entry game with 3 firms where payoffs are defined as follows:{ "In, In, In" -5, -5, -5 }{ "In, In, Out" 1, 1, 0 }{ "In, Out, In" 1, 0, 1 }{ "In, Out, Out" 7, 0, 2 }{ "Out, In, In" 0, 1, 1 }{ "Out, In, Out" 2, 1, 0 }{ "Out, Out, In" 0, 2, 1 }{ "Out, Out, Out" 0, 0, 0 }. In this game, there is a unique mixed strategy equilibrium. The probabilities of playing In for the three players are: 0.213495, 0.670985, and 0.350388. We have verified computationally that this equilibrium is generic in the

The results also suggest that an efficient equilibrium is more likely to be played than an inefficient equilibrium. We view this finding as particularly interesting. Unlike stylized models of perfect competition, it is not true in general that decentralized actions should lead to efficiency when there is strategic behavior. Nevertheless, our estimates suggest that firms tend to realize more efficient configurations of web adoption patterns.

We acknowledge that our model does not allow for dynamics which may be important for understanding the incentives to adopt this technology. However, compared to previous applications of discrete games, the assumption of static behavior is probably more plausible in our application. In entry games, the model is static despite the fact that the data exhibits decade(s) old incumbents. In our application, most web pages were established in the last 3 to 4 years. While dynamics may be present, they seem less problematic than in entry games.

7.4 Model Simulations.

A powerful feature of our estimator is the ability to handle multiple equilibria, particularly those in mixed strategies. Previous estimation techniques have used special features of the econometric problem to avoid having to deal with the multiplicity of equilibria. For example, Berry, Ostrovsky, and Pakes emphasize this point: "It is important to note that though our assumptions do not guarantee a unique equilibrium, they do insure that *there is only equilibrium that is consistent with a given data generating process*. As a result, we will be able to use the data itself to 'pick out' the equilibrium that is played, and at least for large enough samples, we will pick out the correct one." One drawback of this type of approach is the inability to researchers to simulate the model forward and perform counterfactual analysis, since the estimator provides

sense that the mixed strategy equilibrium persists for nearby payoffs. An estimator that did not accomodate mixed strategies would be undefined for this vector of utilities.

no guidance about which equilibrium will be played under different economic conditions. On the other hand, since we explicitly account for the equilibrium selection mechanism, this is something that we can calculate with confidence.

For example, it is interesting to gauge how close the expected utility of an observed outcome is to the joint utility-maximizing equilibrium. To compute this distance, we form games with payoffs determined by the observed covariates and the parameters of the estimation. We use Gambit to solve for all of the equilibria of these games, and compute the expected payoff of each equilibrium. The expected surplus is simply the difference between the expected payoff of the observed action and the most efficient equilibrium.[16] This type of calculation is simply not possible without a equilibrium selection mechanism, since it requires knowing the expected payoffs under all possible equilibria. Figure 1 graphs the distribution of observed utility surplus, broken down by the size of the market. Table 1 shows summary statistics of the surplus distribution.

An interesting feature of the data is that, for all market sizes, the distribution of surplus has a mass of probability centered at zero. The distributions are negatively skewed. This is to be expected as we compare the expected utility of one outcome against the maximum expected payoff of all possible equilibria. The opposite is also possible, as sometimes the observed outcome is a fortuitous element of the most efficient equilibrium. The mass to the right of zero is clearly going to be less than that to the left, since at best the most favorable market outcome can be one part of the efficient equilibrium while poor outcomes can be part of dominated equilibria. In fact, for the markets with 2 and 4 firms, no expected outcome does better than the expected outcome of the most efficient equilibrium.

Table 5 shows the distribution of the percentage difference between the expected utility of the observed

[16] The expectations are relevant here since the firms' payoffs are subject to unobserved shocks.

outcome and of the maximum expected utility. There are generally two modes near 0 and -100, as outcomes are close to the efficient outcome or are far away from it. When the value is -100 it corresponds the case where no firms decided to adopt, where the efficient outcome has a positive payoff. Interestingly our results suggest that in the three smaller markets there are a number of lost opportunities to adopting a website. Only in the largest market do the observed outcomes tend to be close to efficient on average. These results are also summarized in figures 1 and 2.

8 Conclusion.

Estimating models that are consistent with Nash equilibrium behavior is an important empirical problem. In this paper, we have developed algorithms that can be used to estimate both the utilities and the equilibrium selection parameters for static, discrete games. Our algorithms, unlike previous research, can be applied to general normal form games, not just specific examples such as entry games. The algorithms use computationally efficient methods and our Monte Carlo work suggests that they may work well even with a moderate number of observations.

We also study the nonparametric identification of these games. Previous researchers have proved negative results in the class of models we study. In general, these games are not identified. However, we demonstrate that identification is possible through the use of exclusion restrictions. Two types of exclusion restrictions are sufficient for identification. First, if there are variables that influence equilibrium selection but do not directly enter into payoffs. Second, if there are variables that: a) shift i's utility but which do not enter into the utility of other players and b) can be excluded from the equilibrium selection mechanism.

As an application of our methods, we studied the decision by firms to go online by creating a web page. Our data set was a cross section of many markets and there is no particular reason to assume that the same equilibrium was played in each market. The results suggested that web page adoption has a negative effect on the decision to create a web page. Also, the data displays a tendency toward selecting efficient equilibrium. The application highlights the strengths of our approach. Five player games had upwards of 30 equilibria and in some cases there were no pure strategy equilibrium. The estimation methods we propose to the best of our knowledge are the only current approach capable of accommodating both multiplicity and mixed strategy equilibria.

9 References.

[1] Ackerberg, D. (2003): "A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation," University of Arizona Working Paper.

[2] Ackerberg, D., and Gowrisankaran, G. (2003): "Quantifying Equilibrium Network Externalities in the ACH Banking Industry," University of Arizona and Ohlin School of Business, Washington University at St. Louis Working Paper.

[3] Aguirregabiria, V., and Mira, P. (2003): "Sequential Simulated-Based Estimation of Dynamic Discrete Games," Boston University Working Paper.

[4] Bajari, P., Benkard, C.L., and Levin, J. (2003): "Estimating Dynamic Models of Imperfect Competition," Stanford and Duke University Working Paper.

[5] Bajari, P., and Krainer, J. (2004): "An Empirical Model of Analysts' Stock Recommendations: Market Fundamentals, Conflicts of Interest and Peer Effects," mimeo, Duke University.

[6] Berg, J., and McLennan, A. (2002): "The Asymptotic Expected Number of Nash Equilibria of Two Player Normal Form Games", University of Minnesota Working Paper.

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

[8] Berry, S., Ovstrovsky, M. and Pakes, A. (2003): "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry / Exit Examples)," Harvard University Working Paper.

[9] Bresnahan, T., and P. Reiss (1990): "Entry in Monopoly Markets," *Review of Economic Studies*, 57, 531–553.

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

[11] Brock, W. and Dulauf, S. (2001): "Discrete Choice with Social Interactions," *Review of Economic Studies*, 62(2), 235-260.

[12] Brock, W. and Durlauf, S. (2003): "Multinomial Choice with Social Interactions," University of Wisconsin Working Paper.

[13] Chernozhukov, V., and Hong , H. (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115(2), 293-346.

[14] Ciliberto, F., and Tamer, E. (2003): "Market Structure and Multiple Equilibria in Airline Markets,"

Princeton and North Carolina State University Working Paper.

[15] Conklin, J., Judd, K., and Yeltekin, S. (2003): “Computing Supergame Equilibria,” *Econometrica*, 71(4), 1239-54.

[16] Hajivassiliou, V., and Ruud, P. (1994): “Classical Estimation Methods for LDV Models Using Simulation,” in *The Handbook of Econometrics, Volume 4*, ed. by R. Engle and D. McFadden, Elsevier, 2383-2441.

[17] Heckman, J. (1990): “Varieties of Selection Bias,” *American Economic Review*, 80(2), 313-318.

[18] Hotz, V., and Miller, R. (1993): “Conditional Choice Probabilities and the Estimation of Dynamic Models,” *The Review of Economic Studies* 60, 497-531.

[19] Imrohoroglu, S. (1993): “Testing for Sunspot Equilibria in the German Hyperinflation,” *Journal of Economic Dynamics and Control*, 17(1-2), 289-317.

[20] Manski, C. (1993): “Identification of Endogenous Social Effects: The Reflection Problem,” *Review of Economic Studies*, 60(3), 531-42.

[21] Manuszak, M. and Cohen, A. (2004): “Endogenous Market Structure with Discrete Product Differentiation and Multiple Equilibria: An Empirical Analysis of Competition Between Banks and Thrifts,” Carnegie Mellon Working Paper.

[22] Mazzeo, M. (2002): “Product Choice and Oligopoly Market Structure,” *Rand Journal of Economics*, Summer 2002, 221-242.

[23] McFadden, D. (1989): “A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration,” *Econometrica*, 57(5), 995-1026.

[24] McKelvy, R., and McLennan, A. (1996): “The Computation of Equilibrium in Finite Games”, in *The Handbook of Comptutational Economics*, *Vol I.*, ed. by H. Amman, D. A. Kendrick, and J. Rust, Elsevier, 1996, 87-142.

[25] McLennan, A. (2002): “The Expected Number of Nash Equilibria of a Normal Form Game”, forthcoming in *Econometrica.*

[26] Moro, A. (2003): “The Effect of Statistical Discrimination on Black-White Wage Inequality: Estimating a Model with Multiple Equilibria,” *International Economic Review*, 44(2), 467-500.

[27] Pakes, A. and Pollard, D. “Simulation and the Asymptotics of Optimization Estimators,” *Econometrica*, 57, 1027-1057.

[28] Pesendorfer, M., and Schmidt-Dengler, P. (2003): “Identification and Estimation of Dynamic Games,” London School of Economics Working Paper.

[29] Sargent, T., and Wallace, N. (1987): ““Inflation and the Government Budget Constraint,” in *Economic Policy in Theory and Practice*, ed. by Assaf Razin and Efraim Sadka, MacMillan Press, London, 1987.

[30] Seim, K. (2001): “Spatial Differentiation and Market Structure: The Video Retail Industry,” Yale University Ph.D. Dissertation.

[31] Sweeting, A. (2004): “Coordination Games, Multiple Equilibria and the Timing of Radio Commercials”, MIT Working Paper.

[32] Tamer, E. T. (2002): “Incomplete Bivariate Discrete Response Model with Multiple Equilibria,” *Review of Economic Studies*, 70, 147–167.

[33] Vuong, Q. and Bjorn, P.A. (1984): “Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation,” SSWP No. 537, California Institute of Technology.

10 Appendix.

10.1 Proof of Lemma 1.

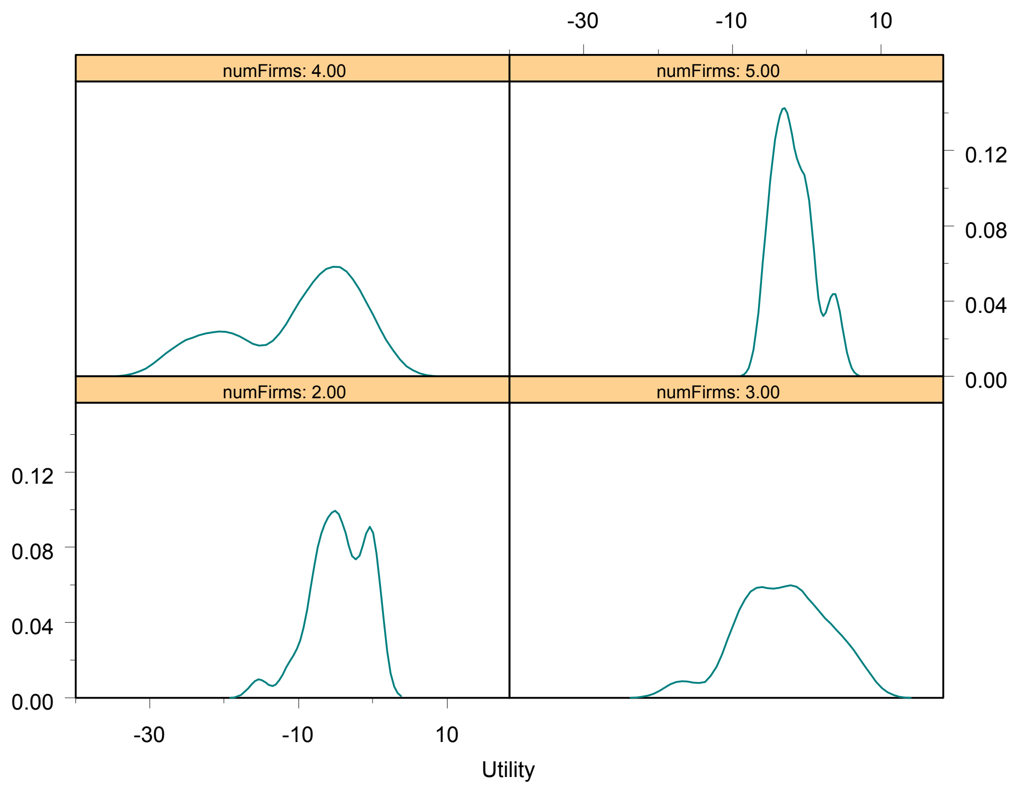
There are at most 4 pure strategy equilibrium. We note that, generically, there is at most one mixed strategy equilibrium. This equilibrium can be represented as the solution to a system of two equations and two unknowns (representing the mixing probabilities) and the solution to such a system is unique with probability one. Therefore, an upper bound on the number of equilibrium is 5 for generic u.

Next, we demonstrate that the number of pure strategy equilibrium must generically be less than two, given u. Suppose not. Note that for one player, some strategy has to be an element of two or more equilibrium. Without loss of generality, let (T, L), (T, R) and (B, R) be three pure strategy equilibrium. By the definition of equilibrium:

$$\begin{aligned} u_2(T, L) &\geq u_2(T, R) \\ u_2(T, R) &\geq u_2(T, L) \end{aligned}$$

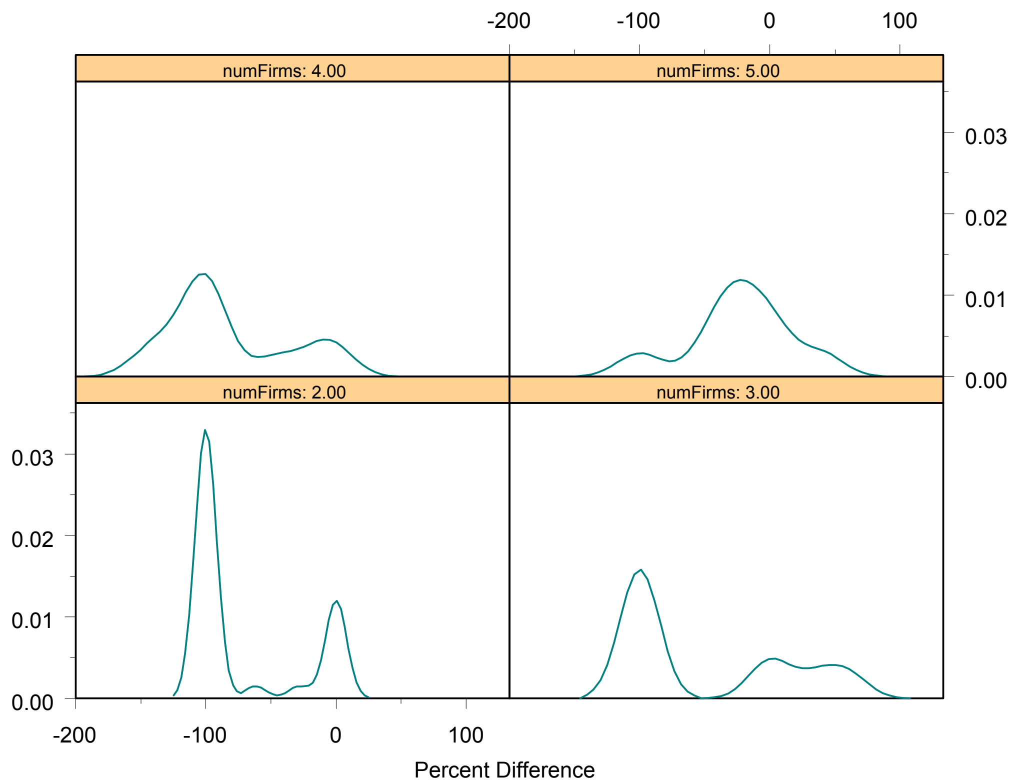
The first inequality follows because (T, L) is an equilibrium and the second because (T, R) is an equilibrium. Therefore, $u_2(T, L) = u_2(T, R)$. This event is not generic. Therefore, there are at most two pure strategy equilibrium and generically the number of equilibrium is three. If the number is three, one of them must be mixed, which establishes (i). Arguments analogous demonstrate (ii). Q.E.D.

Figure 1: Distribution of Surplus by Size of Market



This figure shows the distribution of expected surplus conditional on the size of the market. The expected surplus is calculated as the difference between the expected utility of the observed outcome and the expected utility of the most efficient equilibrium.

Figure 2: Distribution of Percentage of Expected Maximum Surplus by Size of Market



This figure shows the distribution of the percentage of expected maximum surplus conditional on the size of the market. The distribution is calculated as the percent difference between the expected utility of the observed outcome and the expected utility of the most efficient equilibrium.

Table 1: Monte Carlo Results.

Parameter	*Mean*	*Median*	*IQR*	*Confidence Interval*	*Std. Dev*
T=25					
θ_1	5.0342	5.0404	0.21418	[4.5684, 5.5686]	0.26308
θ_2	1.4877	1.4675	0.23879	[1.0771, 1.9592]	0.26361
θ_3	1.1283	1.0394	0.19277	[0.72062, 1.9593]	0.36912
θ_4	-1.0711	-1.0244	0.04973	[-1.2794, -0.97999]	0.19687
β	1.0330	0.99922	0.01356	[0.92224, 1.1125]	0.33465
T=50					
θ_1	5.0218	5.0288	0.08552	[4.6863, 5.3162]	0.20485
θ_2	1.4794	1.4752	0.16211	[1.1823, 1.8771]	0.21398
θ_3	1.0322	1.0176	0.06815	[0.82246, 1.3440]	0.19895
θ_4	-1.0293	-1.0116	0.02993	[1.0837, 0.98737]	0.08357
β	1.0172	1.0000	0.00448	[0.95352, 1.0168]	0.23988
T=100					
θ_1	5.0323	5.0224	0.05724	[4.8738, 5.2040]	0.12690
θ_2	1.4938	1.4766	0.08945	[1.3246, 1.7082]	0.14590
θ_3	1.0253	1.0165	0.03221	[0.92628, 1.2529]	0.17947
θ_4	-1.0221	-1.0112	0.01746	[-1.0594, -0.99170]	0.058417
β	0.99637	1.0001	0.00100	[0.96680, 1.0083]	0.017842

Monte Carlo results of the estimator on simulated data sets of N=25, 50, and 100. The estimation was repeated 100 times for each data set size. The true parameters (θ_1,θ_2,θ_3,θ_4,β) are (5, 1.5, 1.0, -1.0, 1.0).

Table 2: Summary Statistics for Golf Courses.

Variable	*Mean*	*Median*	*Standard Deviation*
Par	67.81	72.00	11.39
Yardage	6091.60	6,456.00	1,208.81
Course Rating	69.28	70.90	7.77
Slope	121.94	121.50	7.61
Weekday Price	22.93	20.00	14.58
Weekend Price	27.26	25.00	15.07
Population	14,282.78	7544.50	18,509.33
Median Rent	324.18	311.04	82.20
Median House Value	85,717.35	78,400.00	32,582.27
Median Income	29,172.32	28,686.00	6,259.55

Table 3: Website Adoption Rates for Golf Courses in Carolinas.

Number of Firms in Market	*Adoption Rate*
1	15%
2	19%
3	16%
4	45%
5	27%

Table 4: Parameter Estimates.

Parameter	*Mean*	*Median*	*Confidence Interval*
Constant	-1.1830	-1.1511	[-1.5017,-1.1218]
Competition Penalty	-6.1076	-6.1226	[-6.0331,-6.4116]
Weekend Price	0.3365	0.3337	[0.3321, 0.3447]
Population	0.0006	-0.0012	[-0.0207,0.0161]
Median House Value	-0.0264	-0.0243	[-0.0766, 0.0065]
Median House Income	-0.0878	-0.0525	[-0.2520,0.0128]
Mixed Strategy Selection	0.1984	0.1841	[0.1841, 0.2459]
Most Efficient Strategy Selection	0.5807	0.6239	[0.6407,0.3242]

Table 5: Simulation Results.

Number of Firms in Market	*2*	*3*	*4*	*5*
Minimum	-15.33	-16.89	-26.02	-5.38
1st Quartile	-6.76	-6.79	-15.73	-3.7
Mean	-4.56	-3.22	-10.23	-1.74
Median	-4.58	-2.94	-7.84	-2.26
3rd Quartile	-1.06	0.46	-4.74	0
Maximum	0	7.2	0	3.68
Number Observations	32	20	10	8
Standard Deviation	3.83	5.98	8.67	2.89