Identification and Estimation of Normal Form Games.

Patrick Bajari, Han Hong and Stephen Ryan.[1]
Preliminary Version.

March 11, 2004.

Abstract

We discuss the identification and estimation of normal form games. Following Bresnahan and Reiss (1990, 1991) utilities are a function of strategies, observed covariates, and random preference shocks. We also include an equilibrium selection rule as part of our model. Using recent algorithms to compute all of the Nash equilibrium to a game, we propose simulation-based estimators for parametric games. We also study nonparametric identification of the model parameters. With appropriate exclusions restrictions about which variables can influence payoffs and the selection of equilibrium, we establish that the model is identified with minimal parametric assumptions.

[1] Affiliations: Bajari- Duke University Department of Economics and NBER, Hong and Ryan- Duke University Department of Economics. This research was funded in part by National Science Foundation grants to Bajari and Hong.

1 Introduction.

Nash equilibrium is one of the cornerstones of modern social science. It is the benchmark theoretical model for analyzing strategic interactions among a handful of players. This theory has been applied to diverse problems in almost all subfields of economics as well as other disciplines ranging from sociology to evolutionary biology.

The empirical literature on Nash equilibrium is considerably less developed than the theory. A sizeable number of studies have conducted hypothesis testing to judge the extent to which observed behavior corresponds to Nash equilibrium. However, comparatively few studies have attempted to econometrically model the relationship between independent and endogenously determined variables in a manner that is consistent with the logic of Nash equilibrium in a normal form game.

The existing literature on estimating normal form games can be classified into two broad categories. The first includes the papers of Bresnahan and Reiss (1990,1991), Berry (1992), Tamer (2002), and Ciliberto and Tamer (2003). These papers generalize standard discrete choice models to allow for strategic interactions between agents. Utility is a function of the vector of strategies, covariates, and latent, outcome specific preference shocks. The second set of papers includes the work of Seim (2001), Aguirregabiria and Mira (2002), Berry, Ovtrovsky and Pakes (2003), Pesendorfer and Schmidt-Dengler (2003) and Sweeting (2004). These models are similar, except that the random preference shocks are assumed to be private information. This second class of models can be estimated with considerable generality in a computationally simple manner. For instance, Aguirregabiria and Mira (2002), Berry, Ovtrovsky and Pakes (2003), Pesendorfer and Schmidt-Dengler (2003) and Bajari, Benkard and Levin (2003) propose methods for estimating rich dynamic models using computationally light two-step estimators.

Games in which the outcome specific preference shocks are common knowledge among the agents, but unobserved to the econometrician, are considerably more difficult to estimate. Because preference shocks are observed by the agents, but not the econometrician, the equilibrium will depend on a vector of unobserved shocks. Moreover, the multiplicity of equilibrium poses difficulties for both computation and the econometric theory.

There have been three approaches for solving these problems in practice. First, Tamer(2002) and Ciliberto and Tamer (2003) propose bounds estimators for the games. A second approach is to simplify the structure of the equilibrium to make point estimation feasible. For example, Bresnahan and Reiss (1990,1991) and Berry (1992) derive a well-defined likelihood function for the number of players in entry games. Finally, one could 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). However, to the best of our knowledge, the literature has not proposed a computationally feasible strategy for point estimation of general normal form games. Also, to the best of our knowledge, positive identification results under weak parametric assumptions are not available for this class of models.

In this paper, we propose a method that can be used to estimate a rich class of normal form games of the type first discussed by Bresnahan and Reiss (1990,1991). Our approach differs from previous research in several ways. First, we explicitly model the equilibrium selection process. Using previously proposed methods, it may be possible to estimate the Von Neumann-Morgenstern (vNM) utilities for players in certain special cases (e.g. entry games). However, without a model of which equilibrium is selected, we will be unable to simulate behavior, which is a necessary step for forecasting or counterfactual analysis. Additionally, characterizing the selection process is an important empirical problem. As McLennan (2002) and Berg and McLennan (2002) have demonstrated, the expected number of Nash equilibrium increases at an exponential rate in the number of players and the number of strategies in the game. Without an empirical model of equilibrium selection, game theory will have very little predictive power due to the large number of equilibria.

Second, we allow for both pure and mixed strategy equilibrium. To the best of our knowledge, all previous studies ignore the mixed strategy equilibria to the game. If mixing does in fact occur, ignoring these equilibria will generate inconsistent estimates.

Third, our model allows for a more general specification of how preference shocks influence utility. In entry games, such as Bresnahan and Reiss (1990,1991) and Tamer (2002), the dimensionality of the error terms is less than the number of strategies. While these restrictions may be reasonable in entry games, they may not be a priori reasonable in more general games. Finally, to the best of our knowledge, for this general class of games, we present the only positive results about identification of vNM utilities and the selection of equilibrium under weak parametric restrictions.

We estimate the structural vNM utilities and equilibrium selection process using two recent advances in computation. First, we use algorithms described by McKelvy and McLennan (1996) to compute all of the Nash equilibrium to a game, including the mixed strategy equilibria. These algorithms have been distributed in the publicly available software Gambit. Second, we modify an importance sampling method to reduce the computational burden of our simulation-based estimator (see Ackerberg (2003)). Normally in structural models, it is necessary to recompute the equilibrium many times in order to evaluate moment conditions at different parameter values. Ackerberg (2003) demonstrates that by making an appropriate change of variables, recomputing the equilibrium is not required. Applying this insight to our problem massively reduces the computational burden of estimation.

Finally, we turn to the problem of identification. While the estimation procedure that we propose is parametric, it is useful to known whether identification hinges on parametric restrictions. Therefore, we consider whether our model is nonparametrically identified. As we shall discuss in the conclusion to the paper, our nonparametric identification arguments naturally suggest nonparametric estimation strategies. We are currently pursuing this in related research. However, given the lack of algorithms that can be used to estimate normal form games in general, parametric models are a natural starting place.

Suppose that the economist observes the joint probability distribution of the players' actions conditional on the covariates. The model is identified if knowledge of this distribution is sufficient to determine the structural vNM utility and equilibrium selection parameters. We begin by establishing a negative result: without restrictions on the equilibrium selection process, it is not possible to identify the structural parameters without parametric restrictions. This result is similar to results previously established by Bresnahan and Reiss (1991) and Pesendorfer and Schmidt-Dengler (2003).

We then consider an approach to identification that exploits exclusion restrictions. The strategy is similar to approaches used in treatment effect and sample selection models. We look for some variable that influences the probability that a given equilibrium is played but which does not enter into players utilities. 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 requires an exclusion restriction (see Heckman (1990)). Exclusion restrictions generate variation in the selection process, so that individuals who are otherwise similar, differ in their likelihood of receiving the treatment. This variation allows one to distinguish between selection versus treatment. Our finding that similar conditions are required in game theoretic models should not be surprising in light of these results since formally, our model has a similar structure.

In some applications, plausible exclusion restrictions might be available. For instance, consider the game college professors play when determining grade distributions. In principal, a C is supposed to be assigned to an average student. However, in practice, as seasoned instructors know, grade inflation has occurred. If a particular instructor gave a C to an average student, while the average grade from all other instructors in the department was a B+, many complaints would be generated. The problem of determining a grade distribution can be viewed as a game of selecting standards. There are potentially multiple equilibria. A first possible equilibrium is where grading is "tough" and C's given to an average student. In a second, grades are inflated and the average student receives a B+. There may be a rich set of additional equilibria as well.

An exclusion restriction is a variable that influences the choice of equilibrium, but which does not enter into the vNM utilities. For instance, grading guidelines established by administrators can help departments coordinate on a particular equilibrium. A second possibility is history dependence. As any faculty member knows, expectations about the distribution of grades is strongly influenced by grades given in the previous quarter or at comparable institutions. Therefore, grading guidelines, lagged grades or grades a comparable institutions are potential exclusion restrictions since they may influence the particular equilibrium that is played even though they do not enter into current utility.

Beyond this simple example, exclusion restrictions may be available in many applications. Variation in laws and regulation, a common source of exogeneity in treatment effect and sample selection models, might be appropriate exclusion restrictions in our models. Pronouncements by industry associations or leading equilibrium figures can influence the choice of equilibrium while not directly entering payoffs. Finally, previously played equilibrium or equilibrium in comparable markets might influence beliefs about which equilibrium should be played. We note that such variables can easily be constructed in panel data sets.

If appropriate exclusion restrictions can be found, then under some conditions, our model can be identified with weak parametric restrictions. As in the models of previous authors, including Sargent and Wallace (1987), Imrohoroglu (1993), Moro (2003), Ackerberg and Gowrisankaran (2003) and Sweeting (2004), the presence of multiple equilibria increases the number of moments the model must match. Multiplicity can actually assist with identification.

One limitation to our analysis is that it will not work for very large normal form games with many players and many strategies. This is because there is no polynomial time algorithm to compute all of the equilibrium to a game (see McKelvy and McLennan (1996)). However, our method will work for games with a moderate number of players and strategies. Also, our algorithm is easily parallelizable, which can reduce the computational burden of the estimator if multiple processors are available to the econometrician.

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 $A = \prod_i A_i$ and let $a = (a_1, ..., a_N)$ denote a generic element of A. Player i's 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:

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

We will often abuse notation an write $u_i(a)$ instead of $u_i(a, x, \varepsilon, \theta_1)$. 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 vector of random preference shocks. Note that the preference shocks depend on the entire vector of actions a, not just the actions taken by player i. The assumption of complete information implies that both $f_i(x, a; \theta_1)$ and $\varepsilon_i(a)$ are common knowledge. We will assume that the $\varepsilon_i(a)$ are distributed i.i.d. according to a parametric distribution $g(\varepsilon|\theta_2)$. Standard discrete choice models, such as the multinomial logit or probit are special cases of (1). In these models, u_i only depends on a_i, not the actions of other agents. Therefore, the framework that we consider generalizes this important class of models to allow utility to depend on a, the actions of all players.

Specifying utility similarly to (1) is common in the literature on estimating games. For instance, in studies of strategic entry, A_i represents the discrete choice of whether or not to enter a market. The term $f_i(x, a; \theta_1)$ corresponds to a reduced form representation of profits conditional on i's entry decision. The vector of covariates x includes variables which influence profits such as the population of a city or observable cost information.

The term $\varepsilon_i(a)$ reflects information about utility that is common knowledge to the players, but not observed by the econometrician. Note that Aguirregabiria and Mira (2002), Berry, Ovtrovsky and Pakes (2003), Pesendorfer and Schmidt-Dengler (2003) and Sweeting (2004) assume that $\varepsilon_i(a)$ is private information. Assuming that $\varepsilon_i(a)$ is private information is an appropriate assumption if $f_i(x, a; \theta_1)$ completely summarizes i's information about player j's payoffs for $j \neq i$. If there is information about payoffs publicly observed by the players that is unobserved by the economist, then assuming the $\varepsilon_i(a)$ are common knowledge is perhaps more appropriate.[2]

Let $u = (u_1, ..., u_N)$ denote a vector of vNM utilities. Frequently, a game will have multiple equilibria. Let $E(u)$ denote the set of Nash equilibrium corresponding to the utilities u. Let $\lambda(\pi; E(u), \beta)$ denote the

[2] A more general model would allow for a vector of ε's that includes some preference shocks that are public information to all players and other preference shocks that are private information. We note that our identification results suggest that the model where all the ε's are public information can be rejected by the data. This suggests that the more general model might be identified under some conditions. This would be closely related to an incomplete information game with unobserved state variables.

probability that a particular mixed strategy equilibrium $\pi \in E(u)$ is selected given a vector of parameters β. Since the number of Nash equilibrium is finite with probability one, $\lambda(\pi; E(u), \beta)$ corresponds to a finite vector of probabilities.[3] In order for λ to generate a well defined distribution it must be the case that for all u and β that:

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

In an application, we could generate a parsimonious, parametric model of λ using standard tools from discrete choice. Theorists have suggested that an equilibrium may be more likely to be played if it:

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 $E(u)$ we could create dummy variables for whether a given equilibrium, $\pi \in E(u)$ satisfies criteria 1-3 above. Let $y(\pi, u)$ denote a vector of variables that we generate in this fashion. Then a parsimonious, parametric model of λ is:

$$\lambda(\pi; E(u), \beta) = \frac{\exp(\beta \cdot y(\pi, u))}{\sum_{\pi' \in E(u)} \exp(\beta \cdot y(\pi', u))} \tag{2}$$

Computing the set $E(u)$, 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 $E(u)$ using these methods.[4] Finding all of the equilibrium to a game is not a polynomial time computable problem. However, the

[3] It can easily be verified that the number of Nash equilibrium is finite with probability one. There are at most a finite number of equilibrium in pure strategies. A mixed strategy equilibrium can be represented as the solution to a system of linear equations where the number of equations and unknowns are both equal. It can easily be verified that the Jacobian to this equation will be singular at a Lebesgue measure zero set of vNM utilities. So long as the distribution of ε is absolutely continuous, this event will have probability zero.

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

available algorithms are fairly efficient at computing $E(u)$ for games of moderate size. Readers interested in the details of the algorithms are referred to McKelvy and McLennan (1996). In the next sections, we shall take the ability to compute $E(u)$ as given.

3 Estimation.

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

$$P(a|x,\theta,\beta) = \int \left\{ \sum_{\pi \in E(u(x,\theta,\varepsilon))} \lambda\left(\pi; E(u(x,\theta_1,\varepsilon)), \beta\right) \left(\prod_{i=1}^{N} \pi(a_i) \right) \right\} g(\varepsilon|\theta_2) d\varepsilon \tag{3}$$

In equation (3), we compute $P(a|x,\theta,\beta)$ as follows. Given a realization of random preference shocks ε, we use equation (1) to compute the utilities, which we denote as $u(x,\theta,\varepsilon)$ to emphasize the dependence of the vNM utilities on the parameters, covariates and preference shocks. This determines the set of equilibria $E(u(x,\theta,\varepsilon))$. Since this is generically a finite set, we sum over the equilibrium $\pi \in E$ 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.

In practice, we will simulate the integral (3) using importance sampling. Instead of integrating over ε, the vector of latent utility shocks, we will make a change of variables and integrate over u, the vector of latent utilities. Ackerberg (2003) noted that this change of variables reduced the computational burden in models similar to those that we consider. This change of variables will lead to significant computational savings.

Let $h(u|\theta,x)$ denote the pdf for u, conditional on θ and x. This can easily be derived through a straightforward change of variables. For instance, suppose that the preference shocks ε are iid normal. Let $\phi(\cdot|\mu,\sigma)$ denote the normal density with mean μ and standard deviation σ. Then, the density $h(u|\theta,x)$ is:

$$h(u|\theta, x) = \prod_i \prod_{a \in A} \phi(\varepsilon_i(a); f_i(\theta, x, \theta) + \mu, \sigma) \tag{4}$$

$$\text{where for all } i \text{ and all } a,\ \varepsilon_i(a) = u_i(a) - f_i(x, a; \theta_1) \tag{5}$$

If we change the variable of integration from ε to u, then (3) becomes:

$$P(a|x, \theta, \beta) = \int \left\{ \sum_{\pi \in E(u)} \lambda\left(\pi; E(u), \beta\right) \left(\prod\nolimits_{i=1}^{N} \pi(a_i) \right) \right\} h(u|\theta, x) du \tag{6}$$

In practice, we shall evaluate the integral (6) using importance sampling. Let $u^{(s)} = (u_1^{(s)}, ..., u_N^{(s)})$, $s = 1, ..., S$ denote a sequence of pseudorandom utilities drawn from an importance density $q(u)$. We can then simulate $P(a|x, \theta, \beta)$ as follows:

$$\widehat{P}(a|x, \theta, \beta) = \frac{1}{S} \sum_{s=1}^{S} \left\{ \sum_{\pi \in E(u)} \lambda(\pi; E(u^{(s)}), \beta) \left(\prod\nolimits_{i=1}^{N} \pi(a_i) \right) \right\} \frac{h(u^{(s)}|\theta, x)}{q(u^{(s)})} \tag{7}$$

As $S \to \infty$, under standard regularity conditions, $\widehat{P}(a|x, \theta, \beta)$ will be a consistent estimator of $P(a|x, \theta, \beta)$. Given $\widehat{P}(a|x, \theta, \beta)$ we can easily form a simulated likelihood or method of moments estimator as we will describe in the next section. Even for finite S, this $\widehat{P}$ is an unbiased estimate of P.

The problem of simulating (7) can be broken into three steps.

1. Draws a large set of random games from the importance density $q(u^{(s)})$.
2. Compute the set of equilibrium, $E(u^{(s)})$ from all games using Gambit or by using the algorithms described in McKelvy and McLennan (1996).
3. Evaluate the sum (7).

The key to the computational savings, as noticed by Ackerberg (2003), is once the equilibrium sets $E(u^{(s)})$ have been precomputed, *it is not necessary to repeat steps1 and 2 when evaluating (7)*. This is a consequence of changing the variable of integration from ε to u. Completing step 2, in particular, can be computationally burdensome for games with larger numbers of players or strategies. However, having performed steps 1 and 2 once, it will not be necessary to recompute equilibrium sets. As we shall show below, this change variables also guarantees that our simulated method of moments objective function will be continuously differentiable in the parameters.

Given precomputed equilibrium sets, $E(u^{(s)})$, $s = 1, ..., S$, evaluating (7) is computationally inexpensive. The term $h(u^{(s)}|\theta, x)$, in most specifications, will correspond to a standard density, such as the normal pdf and can evaluated quickly. The term $\lambda(\pi; E(u^{(s)}), \beta)$ will often correspond to a simple model such as the logit in equation (2).

The econometrician observes a sequence (a_t, x_t) of actions and covariates, $t = 1, ..., T$. One possible method that can be used to estimate the parameters θ and β is maximum simulated likelihood (MSL). As is well known, MSL is biased for any fixed number of simulations. In order to obtain $\sqrt{T}$ consistent estimates, one needs to draw S independent simulations per observation t and let $\frac{S}{\sqrt{T}} \to \infty$. Alternatively, one can estimate the parameters using the method of simulated moments (MSM). An advantage of MSM is that the moment functions can be simulated without bias which yields a consistent estimator for a fixed value of S. Appropriate methods and 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.

For completeness, we describe how the MSM estimator is formed. Enumerate the elements of A from $k = 1, ..., \#A - 1$. Note that, because the probabilities of all of the elements of $a \in A$ must sum to one, one of these probabilities will be linearly dependent on the others, so there are effectively $\#A - 1$ moments. Let $w_k(x)$ be a vector of weight functions 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 the equation (3). For each k,

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

Therefore, a set of moment conditions can be formed by taking the sample analog

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

In practice,$P(k|x_t, \theta)$ is evaluated by simulation using the importance sampler (7). The moment conditions are then replaced by the simulation analog:

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

We assume that the simulation draws $u^{(s)}$ are independent over both t and s, and are independent of all x_t's. Then for a positive definite weighting matrix W the MSM estimator is:

$$\left(\widehat{\theta},\widehat{\beta}\right)=\arg\min_{(\theta,\beta)} m(\theta,\beta)'\times W\times m(\theta,\beta).$$

4 Identification.

The estimation strategy we discussed above is parametric. While we shall discuss possible non-parametric extensions of our estimators, often parametric methods are more appropriate because the number of observations is not sufficiently large or the curse of dimensionality makes nonparametric approaches unreliable. 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 non-parametric identification of our model.

4.1 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 (which can be viewed as a game with one player). Therefore, we begin by imposing the identifying assumptions used in random utility models.

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. 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

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

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

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

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.

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 parametric distribution. However, for expositional clarity, we shall assume that it has a standard normal distribution. In standard discrete choice models, it is not possible to identify both $f_i\left(a, x\right)$ and $\varepsilon_i\left(a\right)$ nonparametrically. Distributional assumptions on the error term are necessary even in a binary choice model. Consider a standard binary choice model where the dependent variable is 1 if the index $u(x) + \varepsilon$ is greater than zero, i.e.

$$y = 1(u(x) + \varepsilon > 0) \tag{8}$$

Suppose that the economist observes $P(y = 1|x)$, the probability that the dependent variable is equal to one given the covariates x. If the cdf of ε is G, then (8) implies that:

$$P(y = 1|x) = G(u(x)),$$

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. Therefore, we will assume that the error terms are normally distributed.

Notice that we are also making the assumption that the $\varepsilon(a)$'s are independently distributed. This also has no loss of generality because the correlation structure of the error terms can not be identified if the deterministic part of the utility function $f_i(a,x)$ is nonparametrically specified. For example, consider a simple single agent multinomial choice model with three options, for $a=1,2$:

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

The utility for the third option is normalized identically equal to 0:

$$u_i(3,x) \equiv 0.$$

In the population, essentially two conditional probabilities functions are available for identification purposes:

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

The last observable probability

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

is linearly dependent on the other two probabilities and does not add to the source of identification. The two observable probabilities can be written as functions of the parameters of interest:

$$P\left(a=3|x\right) = P\left(\varepsilon_i(1) \leq -f_i\left(1,a\right), \varepsilon_i(2) \leq -f_i\left(2,a\right)\right),$$

and

$$P\left(a=1|x\right) = P\left(\varepsilon_i(1) \geq -f_i\left(1,x\right), \varepsilon_i(1) - \varepsilon_i(2) \geq f_i\left(2,x\right) - f_i\left(1,x\right)\right).$$

Given that there are only two equations, 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 $\varepsilon_i(1)$ and $\varepsilon_i(2)$ is known. There are no additional degrees of freedom to identify the correlation structure between $\varepsilon_i(1)$ and $\varepsilon_i(2)$. The insufficiency of the degrees of freedom holds regardless of the degree of variation of x in the population.

A multinomial probit model or a nested logit model allows one to estimate the correlation coefficients between $\varepsilon_i(1)$ and $\varepsilon_i(2)$. However, this comes at the cost of assuming a parametric functional form for the deterministic utility component $f_i(a,x)$. Since any choice of the error distribution $F(\cdot)$ involves no loss of generality, we will specify it to be the standard normal distribution.

4.2 The Structure of Equilibrium in 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

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

The payoff matrix in the two by two game can be normalized for example by taking $\underline{a_1} = T$ and $\underline{a_2} = L$. The resulting payoff matrix will take the form:

Payoff Matrix for Two Player Game.

	L	R
T	$(0, 0)$	$(0, f_2(TR, x) + \varepsilon_2(TR))$
B	$(f_1(BL, x) + \varepsilon_1(BL), 0)$	$(f_1(BR, x) + \varepsilon_1(BR), f_2(BR, x) + \varepsilon_2(BR))$

We begin by observing that the set of equilibrium $E(u)$ can be characterized as follows:

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

Proof: See Appendix.

The lemma above establishes that there are exactly two cases of possible multiplicity. If there are three equilibria, the possible sets of equilibrium, $E(u)$ are of the form:

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

Next, we turn to the problem of flexibly specifying the equilibrium selection probabilities. A completely flexible specification of these probabilities would require us to model λ as a function of both x and ε. However, this case is analytically more complicated and we are unable to establish positive results about identification. Therefore, we will assume that λ depends only on the variables x that are observed to the economist. This assumption is analogous to "selection based on observables" assumption in treatment effect and sample selection models. This implies that the error terms in the treatment equation do not enter the selection equation. The implication of this assumption is that the treatment status is exogenous conditional on the observables and that the outcome 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 $E(u)$ and $E(u\prime)$. Suppose that the supports of the equilibria in $E(u)$ and $E(u\prime)$ coincide in the following sense:

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

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

Then, we will assume that $\lambda(\pi; x, z, E(u)) = \lambda(\pi'; x, z, E(u'))$ 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 $E(u)$, 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. We will maintain these assumptions for the rest of the identification section and formalize this condition in the assumption below.

A3. Assume λ is a function of the observed covariates x, but not the unobserved preference shocks ε. Also, given x, λ only depends on the support of the elements in $E(u)$.

If we make assumption A3, 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 $1 - \lambda_3(x) - \lambda_4(x)$.

4.3 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 observed. In a normal form game, the probabilities the economist can observe 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 $\lambda(x)$. We begin by demonstrating that conditional on x, both f and λ correspond to a finite vector of real numbers.

Lemma 1 *Suppose that A1-A3 hold. Conditional on x, for any normal form game, $f(a,x)$ and $\lambda(x)$ correspond to a finite vector of real numbers.*

Proof: Since A is finite, $f(a,x)$ corresponds to a finite vector of real numbers. The equilibrium set is finite with probability one. This is because a mixed strategy equilibrium correspond to the solution of a linear system of equations in u, where the elements of π are the unknowns. This system requires the exact indifference between all of the actions in the support of π_i for all i. This system has as many equations as unknowns and it can easily be verified that it has a unique solution so long as the appropriately defined Jacobian is not zero. This is true with probability one because ε has a normal distribution. By A3, λ only depends on the elements in the support of $E(u)$. There are only finitely many possibilities for the possible supports of the elements of $E(u)$. Therefore, conditional on x, λ corresponds to a finite vector of parameters. Q.E.D.

Conditional on x, the following equation relates $P(a|x)$ and the primitive parameters $f(a,x)$ and $\lambda(x)$:

$$P(a|x,f,\lambda) = \int \left\{ \sum_{\pi \in E(u(f,\varepsilon))} \lambda\left(\pi; E\left(u\left(f,\varepsilon\right),x\right)\right) \left(\prod_{i=1}^{N} \pi(a_i)\right) \right\} g(\varepsilon) d\varepsilon \tag{9}$$

In equation (9), 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. Holding x fixed, by our previous lemma, we can view (9) as a finite number of equations in a finite number of parameters, $f(a,x)$ and $\lambda(x)$.We will denote this system as $P(a|x) = H(f(x),\lambda(x))$ where H is the map implicitly defined by (9). We will let $DH_{f,\lambda}$ denote the Jacobian formed by differentiating H in the finite vector of parameters f and λ.

Definition. Suppose that $f^0(a,x)$ and $\lambda^0(x)$ satisfy (9). We will say that $\left(f^0(a,x), \lambda^0(x)\right)$ are *locally identified* if for all x, there exists an open neighborhood N_x of the finite vector of parameters $\left(f^0(a,x), \lambda^0(x)\right)$ such there is no vector $(f(a,x), \lambda(x)) \in N_x$, $(f(a,x), \lambda(x)) \neq \left(f^0(a,x), \lambda^0(x)\right)$ that also satisfies (9).

Note that this definition of identification is nonparametric in the sense that it does not require directly parameterization of the mean utilities f or selection mechanism λ. In what follows, we shall often invoke the following assumption:

A4. The map H is continuously differentiable. Also suppose that the n_1 by n_2 Jacobian matrix $DH_{f,\lambda}$ has rank $min\{n_1 - 1, n_2\}$.

Assumption A4 implies that our system of implicit equations (9) will satisfy the conditions of the implicit function theorem so that we may confirm local identification by, holding x fixed, comparing the number of moments, $P(a|x)$ to the number of free parameters $(f(x), \lambda(x))$. While we can directly verify assumption A4 for certain games (e.g. a 2 by 2 game), we cannot do so for general games.[5]

The continuous differentiability assumption is fairly easy to verify for general games. Note that holding ε fixed, the elements of $E(u(f,\varepsilon))$ will change continuously except on a set of measure zero. This is because a mixed strategy equilibrium corresponds to the linear system in the utilities u with the number of equations are exactly equal to the number of unknowns. The π which solve this system will therefore, with probability one, change continuously in the f.

The rank condition in A4 is typically difficult to directly verify. There is a linear dependence between the rows of the Jacobian matrix x because of the requirement that the probability of all the elements $a \in A$ sum to one. The rank condition requires that, after we account for this dependence, the rank of the Jacobian matrix can be computed directly by comparing the number of rows and columns. Finding the rank of the Jacobian requires a complete enumeration of the equilibrium set as in Section 4.2. While this is possible in small games, it can be difficult to enumerate these equilibrium sets for larger games.

4.4 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

[5] Our results can potentially be restated as global identification results by applying the Gale-Nikaido theorem.

four deterministic utility parameters cannot be nonparametrically identified.

Theorem 2 *In a game with two players and two strategies, if we make assumptions A1-A4, the deterministic utility components*

$$f_1\left(BL,x\right), f_1\left(BR,x\right), f_2\left(TR,x\right), f_2\left(BR,x\right)$$

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

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

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,

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

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,

$$f_1\left(BL,x\right), f_1\left(BR,x\right), f_2\left(TR,x\right), f_2\left(BR,x\right)$$

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

The result above can easily be generalized beyond 2 by 2 games for similar reasons. Conditional on x, the number of mean utility parameters is greater than the number of moments available to the econometrician.

Theorem 3 *In a game with more than two players and at least two strategies per player, if we make assumptions A1-A2, the deterministic utility parameters $f_i(a,x)$ are not identified from the distribution of $p(a|x)$, even if the selection mechanism $\lambda\left(\cdot\right)$ is known.*

Proof: Consider a game with $i = 1, ..., N$ players and $\#A_i$ strategies for player i. Holding x fixed, the total number of mean utility parameters $f_i\left(a,x\right)$ 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 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 more than 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.

These results are not surprising in light of the analysis of Bresnahan and Reiss (1991) and Pesendorfer and Schmidt-Dengler (2003) who demonstrate in similar models that identification of mean utilities $f_i(a, x)$ is not possible. Given that the mean utilities $f_i(a, x)$ are not identified even with known equilibrium selection probabilities, it is then obvious that there is no hope of jointly identifying $f_i(a, x)$ and $\lambda(x)$ nonparametrically.

5 Exclusion Restrictions.

The previous section demonstrates that restricting the dependence of λ on x does not allow for nonparametric identification. This result is not surprisingly in light of results in the literature on treatment effects and sample selection models. It is well known that identification in these models hinges on parametric assumptions without exclusion restrictions. Therefore, we proceed analogously to the treatment effects literature and search for variables z that influence λ, but not f_i.

Before we formally consider the problem of identification, it is useful to ask what variables might plausibly generate exclusion restrictions in real applications. One potential source of such variables is regulation which is a commonly exploited source of exogeneity in treatment effect models. For instance, 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. There choice of an average recommendation is fundamentally indeterminate as in 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 had an incentive to benchmark their recommendations against the practices of their peers.

Since this is a coordination game, there are obviously 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. 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.

5.1 Negative Results with Exclusion Restrictions.

If there is an exclusion restriction, we write the selection process as $\lambda(x, z)$ where z are the variables that influence selection, but are excluded from $f_i(a, x)$.

A5. The function λ is a function of x and z, where z can be excluded from f_i.

We begin by showing that, without additional assumptions on $\lambda(x, z)$, nonparametric identification is still not possible. This negative result can be seen in the case of a two by two game, where there are four elements in the $\lambda(x, z)$ vector:

$$\lambda(x, z) = (\lambda_1(x, z), \ldots, \lambda_4(x, z)).$$

Suppose that the economist observes P pairs of the vectors (x, z). Given every x, z, there are three conditional outcome probabilities that are observed by the econometrician. But there are four parameters in $\lambda(x, z)$ and another four utility parameters. While the utility parameters do not depend on z, the total number of parameters to estimate for a total of P pairs of (x, z) is

$$4 \times P + 4 \times (\#x),$$

which is clearly more than the $3 \times P$ number of available conditional outcome probabilities. Hence, identification is not possible with a discrete number of pairs P. It follows trivially that identification is not possible with continuous (x, z), since if we knew the structural parameters everywhere, we would know them at any P discrete pairs. This logic can be generalized to a general normal form game.

Theorem 4 *Suppose A1-A5 hold so that there are variables z that enter $\lambda(x, z)$ but not $f_i(a, x)$. Without additional assumptions, identification is still not possible in an N player game.*

Proof: In the general case, the number of available conditional outcome probabilities in the data is

$$\left(\prod_{i} \#A_i - 1\right) \times P \tag{10}$$

We demonstrate that this is less than a lower bound on the number of parameters so that identification is not possible.

First consider the special case when $N = 2$. Given (x, z), suppose the realizations of the error terms in the utility functions of the two players are such that all except two strategies are strictly dominated. There are a total of

$$\binom{\#A_1}{2}\binom{\#A_2}{2} \tag{11}$$

possible pairs of such strategies (a_1, a_2). For each pair of strategies, there are four parameters corresponding to $\lambda(x, z)$ are needed to calculate the equilibrium selection probabilities. Therefore the number of parameters in the equilibrium selection mechanism in a general two player game is bounded from below by

$$4 \times \binom{\#A_1}{2}\binom{\#A_2}{2} \times P. \tag{12}$$

Adding in the number of utility parameters that must be estimated, the total number of parameters in a general two player game is therefore bounded from below by

$$4 \times \binom{\#A_1}{2}\binom{\#A_2}{2} \times P + (2 \times \#A_1 \cdot \#A_2 - (\#A_1 + \#A_2)) \times (\#x) \tag{13}$$

This is more than the number of moments available in (10).

With more than $N > 2$ players, given each pair (x, z), suppose the realizations of the error terms in the utility functions of $N - 2$ players are such that it is a dominant strategy for them to play one of the strategies. For each such case the number of equilibrium selection parameters for the remaining two player game is bounded from below by (12). Therefore the total number of equilibrium selection parameters in the general case is bounded from below by

$$4 \sum_{i \neq j} \prod_{k \neq i,j} \#A_k \binom{\#A_i}{2}\binom{\#A_j}{2} \times P, \tag{14}$$

and the total number of parameters is therefore bounded from below by

$$4 \sum_{i \neq j} \prod_{k \neq i,j} (\#A_k) \binom{\#A_i}{2} \binom{\#A_j}{2} \times P + \left(N \cdot \prod_i \#A_i - \sum_i \prod_{j \neq i} \#A_j \right) \times (\#X) . \tag{15}$$

This is clearly larger than the number of observable conditional outcome probabilities in the population. For example, in the symmetric case where $\#A_i = \#A$ for all i, the lower bound on the total number of parameters is

$$\binom{N}{2} (\#A)^N \frac{(\#A - 1)^2}{2} \times P + \left(N (\#A)^N - N (\#A)^{N-1} \right) \times \#X, \tag{16}$$

and the total number of conditional outcome probabilities is

$$\left((\#A)^N - 1 \right) \times P \tag{17}$$

and it is clearly smaller. Q.E.D.

These results suggest that in addition to exclusion restrictions, we will need to make additional assumptions to identify the model parameters. We consider three possibilities. First we make restrictions similar to identification at infinity arguments in the sample selection literature. Second, we propose variables from x that do not enter λ. Finally, we allow λ to be semi-parametric instead of fully nonparametric.

5.2 Identification at Infinity.

We begin by considering the case when $\lambda(x, z)$ is known for some configurations of (x, z). We will call this a case of "identification at infinity" because it is related to sample selection models in which the intercept of the outcome equation can be identified by choosing the regressor such that the conditional probability of observing the outcome is close enough to 1. We begin considering a two player game and by making an assumption, A6, that strongly restricts the equilibrium selection mechanism.

A6. There are values $z = z_{(1)}, z_{(2)}$ such that for all values of x:

$$\lambda_1 \left(x, z_{(1)} \right) = 1 \qquad \lambda_2 \left(x, z_{(2)} \right) = 1,$$

and

$$\lambda_3 \left(x, z_{(1)} \right) = 1 \qquad \lambda_4 \left(x, z_{(2)} \right) = 1.$$

We can interpret this assumption in words as follows. When $z = z_{(1)}$ and there are multiple equilibrium, the pure strategy equilibrium with player one selecting T is played and when $z = z_{(2)}$, the equilibrium in which B is played is selected. Note that while this is a strong assumption, there are some applications where this can be argued to hold. For instance, when a law or industry group require a particular standard in some, but not all markets in the data.

Theorem 5 *Suppose that A1-A6 are satisfied . Then in a game with two players and two strategies, the $f_i(a, x)$ and $\lambda(x, z)$ are (locally) identified.*

Proof: See Appendix.

The strategy behind the proof is quite straightforward. Holding x fixed, the vNM utilities are identical in the case when $z = z_{(1)}$ and $z = z_{(2)}$ since z is excluded from f_i. Since we know the selection mechanisms in both of these cases, we double the number of moments that the $f_i(a, x)$ must satisfy. Similar to the intuition of Sweeting (2004), the multiplicity of equilibrium generate additional moment conditions that actually improve our ability to identify the structural parameters.

The strategy of this theorem can be generalized to games with more players and more strategies in a straightforward fashion. The basic idea is that we must find an instrument which forces players to choose a particular pure strategy equilibrium and rule out the mixed strategy equilibria when the realization of ε makes multiple equilibria possible. The number of exclusion restrictions will depend on the number of latent mean vNM utilities, $f_i(a, x)$, holding x fixed. As in the above proof, we must count equations and unknowns to determine if there is a sufficient number of cases where we know the selection mechanism to identify the $f_i(a, x)$. Of course, as the number of players and number of strategies increase, these assumptions are less likely to be satisfied in practice.

5.3 Exclusion restrictions in equilibrium selection

In this section we consider a different restriction on the $\lambda(x, z)$ function. This restriction is stricter than allowing $\lambda(x, z)$ to be completely nonparametric, but might be more flexible than assuming that $\lambda(x, z)$ is known for some values of z.

A7. We will assume that x can be partitioned into

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

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

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

Assumption A7 can be viewed as a weaker version of A6. We are now assuming that, instead of knowing the selection mechanism perfectly, we can exclude some covariates from λ. The appropriateness of such an assumption must be judged on a case by case basis. However, many applications of game theory, particularly in industrial organization, consider fairly narrowly defined markets. In principal, the economist can learn a lot about industry norms by reading industry sources or by directly questioning key market participants. Based on this industry knowledge, restrictions of the form A4, may be feasible. We demonstrate that, in the two by two game, A4 is sufficient for nonparametric identification.

Theorem 6 *In the two by two game, suppose that A1-A5 and 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

$$3 \times (\#x_\lambda) \times (\#x_u) \times (\#z), \tag{18}$$

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 in needed to characterize both the utility functions and the equilibrium selection probabilities is

$$4 \times (\#x_\lambda) \times (\#x_u) + 4 \times (\#x_\lambda) \times (\#z). \tag{19}$$

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

$$3 \times (\#x_u) \times (\#z), \tag{20}$$

conditional outcome probabilities and there are

$$4 \times (\#x_u) + 4 \times (\#z), \tag{21}$$

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

$$3 \times (\#x_u) \times (\#z) < 4 \times (\#x_u) + 4 \times (\#z) \tag{22}$$

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

These results will extend beyond the 2 by 2 game. The key to the proof was that the number of moments depends on the product of $(\#x_u) \times (\#z)$ in equation (20) while the number of parameters is a linear function of $\#x_u$ and $\#z$ in (21).

Theorem 7 *In a general N player game, suppose that A1-A3 and A4 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 nonparametrically (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 A3, it must be possible to characterize λ with a finite dimensional parameter vector. Since this vector on depends on the possible supports of the elements in $E(u)$, it is possible to create a bound on the size of this vector that is independent of x_λ and z. Let $\#E$ 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:

$$\#E\,(\#z) + \left(N \cdot \prod_i \#A_i - \sum_i \prod_{j \neq i} \#A_j \right) (\#x_u)$$

The number of moments is proportional equal to

$$(\#A - 1) \times (\#x_u) \times (\#z)\,.$$

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, the model is identified. Q.E.D.

In the theorem, we required the variables to live in a discrete set. However, if the variable 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 has access to the exclusion restrictions in A4, then all of the primitive parameters can be recovered without additional parametric assumptions.

5.4 Identification using semiparametric restrictions.

In this section, we consider a final strategy that allows for nonparametric identification of all of the model parameters if we make semiparametric restrictions on λ. We did this implicit in assumption A3 by forcing the selection process to depend on only the support of the equilibrium strategies. However, we could imagine, more generally, that the selection mechanism depends on the variables (x, z) and a finite dimensional vector of parameters.

We will assume that λ has the following form:

A8. There are vector valued functions $c(x)$ and $d(z)$ with range in R^d and a known, parametric function, e such that $\lambda(\cdot|x, z) = e(c(x), d(z); \beta)$ where β is a finite vector of parameters of dimension k.

The first assumption is a non-trivial restriction on the selection mechanism. The selection mechanism is only allowed to depend on covariates that are observable to the economist, not on the vector of latent preference shocks. The second assumption requires that the selection mechanism depend on a finite dimensional vector of parameters and some functions $c(x)$ and $d(z)$. The assumption restricts potential interactions between x and z. However, the vector of parameters can is only restricted to be finite and the functions c and d are arbitrary.

It is useful to consider a particular example. Consider the two player game with two strategies. Suppose that there are functions, $e_1, ..., e_4, c_1, ..., c_4$ and $d_1, ..., d_4$ that depend on the parameters β such that:

$$\lambda_1(x,z) = \frac{\exp(e_1(c_1(x),d_1(z)))}{1+\exp(e_1(c_1(x),d_1(z)))+\exp(e_2(c_2(x),d_2(z)))} \tag{23}$$

$$\lambda_2(x,z) = \frac{\exp(e_2(c_2(x),d_2(z)))}{1+\exp(e_1(c_1(x),d_1(z)))+\exp(e_2(c_2(x),d_2(z)))} \tag{24}$$

$$\lambda_3(x,z) = \frac{\exp(e_3(c_3(x),d_3(z)))}{1+\exp(e_3(c_3(x),d_3(z)))+\exp(e_4(c_4(x),d_4(z)))} \tag{25}$$

$$\lambda_4(x,z) = \frac{\exp(e_4(c_4(x),d_4(z)))}{1+\exp(e_3(c_3(x),d_3(z)))+\exp(e_4(c_4(x),d_4(z)))} \tag{26}$$

Note that the specification in equations (23)-(26), the selection mechanism is we defined in the sense that the probabilities associated with each equilibrium add up to one. (We suppress the dependence on β in order to simplify the notation). Also, it allows the probabilities to change in a flexible way as a function of β.

Next, we shall show that the above restrictions on the equilibrium selection mechanism are sufficient to identify the parameters of the model.

Theorem. Suppose that assumptions A1-A5 and A8 are satisfied. Then f_i and λ are (locally) identified.

Proof. Let $(x^{(1)}, z^{(1)}), ..., (x^{(P)}, z^{(P)})$ be P arbitrary pairs of x and z. If we considered all, non-redundant combinations the x's and z's. There would be P^2 such pairs. For each pair, the economist observes $\#A - 1$ moments corresponding to the probability that a particular combination of actions is observed in the data. In order to compute the probability of the joint distribution of actions for all $(x^{(r)}, z^{(s)})$, the following values are required:

- The k parameters β.
- The $P\left(N \cdot \prod_i \#A_i - \sum_i \prod_{j \neq i} \#A_j\right)$ values characterizing the mean utilities, $f(a, x^{(p)})$ for every $p = 1, ..., P$.
- The $2 \cdot P \cdot d$ values of $c(x^{(1)}), ..., c(x^{(P)})$ and $d(z^{(1)}), ..., d(z^{(P)})$.

Note that the number of parameters grows linearly in P while the number of moment conditions grows at a quadratic rate. Therefore, by choosing P sufficiently large, the number of moments is greater than the number of free parameters. Hence, by the implicit function theorem, the model is identified for an arbitrary value of x and z. Q.E.D.

6 Monte Carlos.

The following Monte Carlo illustrates the use of the estimator and provides some insights into its small sample properties. Consider a simultaneous game of entry between two firms in two separate markets. Each firm can enter one of two markets. The payoffs from entering the first market are given by:

$$u_i(\text{In}) = \beta X - \delta 1(a_j = \text{In}) + \epsilon_i(a),$$

where $1(\cdot)$ is the indicator function. The exogenous X is a measure of the first market's size. It is natural to assume that payoffs increase with market size and decrease with the presence of another competitor, so we set $\beta = 1.0$ and $\delta = 3.0$. The payoff to entering the second market is:

$$u_i(\text{Out}) = \gamma Y - \delta 1(a_j = \text{Out}) + \epsilon_i(a),$$

where Y is the market size of the second market. We set $\gamma = 2$. For both In and Out the error term depends on the action of the other agent. The errors for each player and each action are drawn independently from the standard normal distribution.

We generate data by finding the outcomes for many independently drawn games. For each game, we draw errors from the standard normal, X from $U[0, 5]$, and Y from $U[0, 6]$. We resolve multiple equilibria with a simple selection mechanism. Each equilibrium is assigned a weight, w_i, and the probability of choosing this equilibrium out of the set of possible equilibria is:

$$P(e_i|X, \epsilon, \theta) = \frac{\exp(w_i)}{\sum_{e \in E} \exp(w_j)}.$$

To identify these parameters, it is necessary that each equilibrium belong to a set of multiple equilibria. Otherwise the probability of selecting that equilibrium is equal to one independent of the weight. For the purposes of the Monte Carlo, we estimate the probability of choosing the mixed strategy over any pure strategy. All of the pure strategy weights are equal and set to log(10). We set the weight on playing a mixed

Table 1: Monte Carlo Results

Run One $n = 40$, $s = 40$.

Parameter	Median	Mean	IQR	Std Dev
beta	0.812	0.771	0.293	0.217
delta	2.852	2.842	0.378	0.286
weight	2.851	3.028	1.212	2.002
gamma	1.741	1.68	0.376	0.302

Run Two $n = 80$, $s = 80$.

Parameter	Median	Mean	IQR	Std Dev
beta	0.785	0.75	0.239	0.195
delta	2.848	2.816	0.337	0.31
weight	2.822	3.089	0.949	1.69
gamma	1.745	1.683	0.342	0.299

Run Three $n = 160$, $s = 160$.

Parameter	Median	Mean	IQR	Std Dev
beta	0.759	0.742	0.232	0.192
delta	2.838	2.827	0.277	0.21
weight	2.914	3.001	0.714	0.661
gamma	1.681	1.679	0.323	0.233

strategy equal to log(20) so that the mixed strategy is played half of the time when it is a choice in the equilibrium set.

Once we generated the data, we ran the estimator as described above. Following Ackerberg, the importance density was chosen to mimic the data generating mechanism. The number of data points and simulations were set equal, and we ran the estimator on 100 simulated data sets for each n. The method of maximum likelihood was used to recover the parameters for three different sample sizes, with the results shown in Table 1.

The results show a few general trends. First, overall the estimator performs well even for small samples. Second, the parameters on the utility function are estimated more precisely than the parameters of the equilibrium selection process. This is because there is still a degree of inherent uncertainty in the equilibrium selection process even when everything else about the model is known. Third, parameters are estimated more precisely as the sample size grows, and the interquartile range shrinks at rate $O(n^{1/2})$, which is a necessary and sufficient condition for pointwise consistency. In particular, the estimation of the weight benefits from an increase in the number of observations for small sample sizes. The reason for this is that the mixed strategy weight is only identified off of coordination failures. To see this, suppose we observe a game with two pure strategies, In/In and Out/Out. If we observe In/In, we cannot separate that outcome as the result of either the pure or mixed strategy. However, if we observe In/Out, then we know it has to be the result of the mixed strategy. For smaller samples, such as 40 observations, it is possible to generate a data set with very few coordination failures, which implies that the variance of these estimates is going to be very high. Increasing the sample size from 40 to 80 helps avoid the case where only a handful coordination failures identify the mixed strategy weight. Once the expected number of failures increases away from zero the benefit of a larger sample size in estimating the weight is more in line with the other parameters.

Following the methods of Chernozhukov and Hong (cite), we also examine the posterior distribution for the parameters using a MCMC sampler. This is a computationally simple and robust method for examining the behavior of the log likelihood function around the estimated parameters. Starting at the truth, a new point in the parameter space, y, is drawn from a multivariate normal distribution centered on the old point, x. The new point is accepted if the negative log likelihood function is lower than the previous point. If it is higher, then the probability of acceptance follows the Metropolis criterion:

$$\text{Pr(Accept)} = \exp(-(f(y) - f(x))/T). \tag{27}$$

The sampling parameter T controls this probability and was set to generate a roughly 50% chance of acceptance. This sequence generates a Markov chain which converges to the true posterior distribution of the parameters. We generated two 5000-length sample chains, each corresponding to $n = s = 40$ and $n = s = 80$. A burn-in period was not necessary as we started the chain at the true parameters.

The marginal distributions for the utility parameters β and δ are shown in Figures 1 and 2. For all four cases the posterior distributions are approximately normal, with larger sample sizes producing smoother densities and sharper maxima. We use a quantile-quantile plot (see Figures 3 and 4) to evaluate the normality of the distributions. The values from the chain are ordered and then plotted against the inverse cumulative

Table 2: Properties of the Posterior Marginal Distributions

Run One n = 40, s = 40		
Parameter	Median	IQR
beta	0.816853	1.538443
delta	2.992445	4.002998

Run Two n = 80, s = 80		
Parameter	Median	IQR
beta	0.703052	1.239434
delta	3.240878	3.412659

distribution for the normal. If the data are distributed normally $N(\mu, \sigma^2)$, then the points on the plot should lie on a straight line with intercept μ and slope σ. The steeper line on both graphs corresponds to the smaller sample size. From the graph, the distributions are consistent with a normal distribution, with most of the data points falling along the line except at the extreme tails. A formal Kolmogorov-Smirnov test fails to reject normality for all four cases. Table 2 gives the summary statistics for the posterior distributions. The empirical rate of decrease in the interquartile range, approximately 1.2, is just shy of the theoretical prediction of $\sqrt{2}$. This deviation is due in large part to the very high variance in estimating the weight for such low sample sizes. In our experiments it smooths out towards the correct rate for higher numbers of observations.

This Monte Carlo shows that the performance of the estimator is good even in small samples. The utility parameters are estimated very precisely, with the equilibrium weight somewhat less so due to the inherent randomness of the selection procedure. We obtain good performance in estimating the parameters with sample sizes as small as 40, although the gains from doubling the sample size are considerable in estimating the weight of playing a mixed strategy.

7 Extensions.

The above identification results can be easily mapped into a nonparametric estimator for the latent utility.

If assumption A4 is satisfied, it is also possible to construct an explicit algorithm for estimating the utility and equilibrium selection parameters whose complexity does not increase with $\#x$ and $\#u$. To simplify notation, we consider the case of a two by two game. Also, we omit the dependence on x_λ and take $x = x_u$, since all the arguments will be applied to each x_λ. Consider for example (with notation defined as in the appendix proof to Theorem 4),

$$P(TL|x,z) = H_1(f) + H_0^{TL}(f) + \lambda_1(z) H_5(f) + \lambda_{12}(z) H_5^{TL|m}(f),$$

where

$$\lambda_{12}(z) = 1 - \lambda_1(z) - \lambda_2(z),$$

and

$$H_0^{TL}(f) = \int_{R_0} p_m(TL|x,\varepsilon)\, dF(\varepsilon),$$

and

$$H_5^{TL|m}(f) = \int_{R_5} p_m(TL|x,\varepsilon)\, dF(\varepsilon).$$

The mixing probability $p_m(TL|x,\varepsilon)$ was defined in the previous section.

Fix z, consider two different values of x: x^1 and x^2, then if all the $f_i(a,x)$ parameters are known, $\lambda_1(z)$ and $\lambda_{12}(z)$ can be solved from the observable $P(TL|x^1,z)$ and $P(TL|x^2,z)$ by:

$$\begin{matrix} \lambda_1(z) \\ \lambda_2(z) \end{matrix} = \begin{bmatrix} H_5(f^1) & H_5^{TL|m}(f^1) \\ H_5(f^2) & H_5^{TL|m}(f^2) \end{bmatrix}^{-1} \begin{pmatrix} P(TL|x^1,z) - H_1(f^1) - H_0^{TL}(f^1) \\ P(TL|x^2,z) - H_1(f^2) - H_0^{TL}(f^2) \end{pmatrix}$$

where f^1, for examples, indicates the value of the vector $f_i(a,x)$ when $x = x^1$.

Now fix the same z but choose two different values of x: x^3 and x^4, then we can also write

$$\begin{matrix} \lambda_1(z) \\ \lambda_2(z) \end{matrix} = \begin{bmatrix} H_5(f^3) & H_5^{TL|m}(f^3) \\ H_5(f^4) & H_5^{TL|m}(f^4) \end{bmatrix}^{-1} \begin{pmatrix} P(TL|x^3,z) - H_1(f^3) - H_0^{TL}(f^3) \\ P(TL|x^4,z) - H_1(f^4) - H_0^{TL}(f^4) \end{pmatrix}$$

Therefore the following equality holds for x^1, x^2, x^3 and x^4:

$$\begin{aligned} & \begin{bmatrix} H_5(f^3) & H_5^{TL|m}(f^3) \\ H_5(f^4) & H_5^{TL|m}(f^4) \end{bmatrix}^{-1} \begin{pmatrix} P(TL|x^3,z) - H_1(f^3) - H_0^{TL}(f^3) \\ P(TL|x^4,z) - H_1(f^4) - H_0^{TL}(f^4) \end{pmatrix} \\ = & \begin{bmatrix} H_5(f^1) & H_5^{TL|m}(f^1) \\ H_5(f^2) & H_5^{TL|m}(f^2) \end{bmatrix}^{-1} \begin{pmatrix} P(TL|x^1,z) - H_1(f^1) - H_0^{TL}(f^1) \\ P(TL|x^2,z) - H_1(f^2) - H_0^{TL}(f^2) \end{pmatrix}. \end{aligned}$$

Now by varying z, we can generate as many equations as needed to solve for the parameters in f^1, f^2, f^3 and f^4. This argument can be repeated for all combinations of f^1, f^2, f^3 and f^4 so that $f_i\,(a,x)$ is identified for all x.

When the x and z are discrete, the conditional outcome probabilities can be estimated by the sample proportions, for example:

$$\hat{p}(TL|x,z_{(1)}) = \frac{\sum_{t=1}^{T} 1\,(a_t = TL)\,1\left(x_t = x, z_t = z_{(1)}\right)}{\sum_{t=1}^{T} 1\left(x_t = x, z_t = z_{(1)}\right)}$$

In this case all the parameters will be $\sqrt{T}$ consistent and asymptotically normal.

On the other hand, when x and z are continuously distributed, the conditional outcome probabilities can be estimated by a variety of nonparametric estimation techniques. For example, a kernel estimator can be employed:

$$\hat{p}(TL|x,z_{(1)}) = \frac{\sum_{t=1}^{T} 1\,(a_t = TL)\,k\left(\frac{x_t - x}{h}, \frac{z_t - z_{(1)}}{h}\right)}{\sum_{t=1}^{T} 1\,(a_t = TL)\,k\left(\frac{x_t - x}{h}, \frac{z_t - z_{(1)}}{h}\right)}$$

where $k\,(\cdot)$ is a kernel function and h is a sequence of bandwidth parameters such that $h \to 0$ and $Th^d \to \infty$ as $T \to \infty$, where d is the dimension of x and z. In this case the estimates will converge at the slower than of $\sqrt{Th^d}$. It will also be asymptotically normal, as long as the bias term is assumed to be sufficiently small. We plan to pursue this extension in a related paper.

The second extension we are considering is developing analogous methods for dynamic games with private information of the type studied by Aguirreguberia and Mira (2002), Berry, Ostrovsky and Pakes (2003) and Pesendorfer and Schmidt-Dengler (2003). These papers have proposed methods for structurally estimating the mean vNM utilities to these games. However, these authors do not propose methods for estimating how the equilibrium is selected.

We conjecture that methods proposed by Conklin, Judd, Yeltekin (2003) can be used to compute all of the value functions for these games that are consistent with equilibrium. Since there is a one to one mapping between equilibrium choice probabilities and value functions, as established by Hotz and Miller (1993), we can derive all of the equilibrium strategies to the game. This would allow us, in principal, to formulate estimators for equilibrium selection in these games.

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 vNM utilities and the equilibrium selection parameters for normal form games. Our algorithms, unlike previous research, can be applied to general normal form games, not just specific examples such as entry games. Furthermore, unlike previous researchers, our algorithms allow for both mixed and pure strategy Nash equilibrium. 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. We demonstrate that identification is possible through the use of exclusion restrictions. Our results compliment this earlier work, without restrictions on payoffs or how the equilibrium is selected, the structural parameters of games are underidentified. We propose a new approach to identification in games based on exclusion restrictions used in standard treatment effect and sample selection models. If we can find variables that influence the probability that a particular equilibrium is observed and make semiparametric restrictions on the selection process, then both the mean vNM utilities and the equilibrium selection process are identified. In practice, such exclusion restrictions could be generated by differences in regulation across markets and beliefs about which equilibrium will be played based on past repetitions of the game, or practices in neighboring markets. In the examples that we presented, it is possible to flexibly parameterize the equilibrium selection process flexibly with a finite number of parameters. This is because the equilibrium set, with probability one, changes continuously with perturbations to the underling vNM utilities.

While many researchers have focused on estimating the mean vNM utilities, considerably less work has studied how to identify how equilibrium is determined. Estimating the equilibrium selection process is an interesting, and we would argue central, problem in game theory. McLennan (2002) and Berg and McLennan (2002) demonstrate the expected number of Nash equilibrium increases at an exponential rate in the number of players and strategies. Therefore, without knowledge of how equilibrium is determined, game theory has little empirical content without strong a priori restrictions on payoffs. Also, there is no generally accepted method in economic theory for selecting a single, out of the many possible, equilibrium to a normal form game. We believe that this is an important potential application of the methods that we propose.

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 Arizon and Ohlin School of Business, Washington University at St. Louis Working Paper.

[3] Aguirregabiria, V., and Mira, P. (2002): "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] Chernozhukov, V., and Hong , H. (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115(2), 293-346.

[12] Ciliberto, F., and Tamer, E. (2003): "Market Structure and Multiple Equilibria in Airline Markets," Princetone and North Carolina State University Working Paper.

[13] Conklin, J., Judd, K., and Yeltekin, S. (2003): "Computing Supergame Equilibria," *Econometrica*,

71(4), 1239-54.

[14] 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.

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

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

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

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

[19] 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.

[20] McLennan, A. (2002): "The Expected Number of Nash Equilibria of a Normal Form Game", University of Minnesota Working Paper.

[21] 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.

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

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

[24] 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.

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

[26] Sweeting, A. (2004): "Coordination Games, Multiple Equilibria and the Timing of Radio Commer-

cials", MIT Working Paper.

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

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 equlibrium is three. If the number is three, one of them must be mixed, which establishes (i). Arguments analogous demonstrate (ii). Q.E.D.

10.2 Proof of Theorem 3.

In the general case, the number of available conditional outcome probabilities in the data is

$$\left(\prod_i \#A_i - 1\right) \times P \tag{28}$$

We demonstrate that this is less than a lower bound on the number of parameters so that identification is not possible.

First consider the special case when $N = 2$. Given given (x, z), suppose the realizations of the error terms in the utility functions of the two players are such that all except two strategies are strictly dominated.

There are a total of

$$\binom{\#A_1}{2}\binom{\#A_2}{2} \tag{29}$$

possible pairs of such strategies (a_1, a_2). For each pair of strategies, there are four parameters corresponding to $\lambda(x, z)$ are needed to calculate the equilibrium selection probabilities. Therefore the number of parameters in the equilibrium selection mechanism in a general two player game is bounded from below by

$$4 \times \binom{\#A_1}{2}\binom{\#A_2}{2} \times P. \tag{30}$$

Adding in the number of utility parameters that must be estimated, the total number of parameters in a general two player game is therefore bounded from below by

$$4 \times \binom{\#A_1}{2}\binom{\#A_2}{2} \times P + (2 \times \#A_1 \cdot \#A_2 - (\#A_1 + \#A_2)) \times (\#x) \tag{31}$$

This is more than the number of moments available.

With more than $N > 2$ players, given each pair (x, z), suppose the realizations of the error terms in the utility functions of $N - 2$ players are such that it is a dominant strategy for them to play one of the strategies. For each such case the number of equilibrium selection parameters for the remaining two player game is bounded from below by (30). Therefore the total number of equilibrium selection parameters in the general case is bounded from below by

$$4 \sum_{i \neq j} \prod_{k \neq i,j} \#A_k \binom{\#A_i}{2}\binom{\#A_j}{2} \times P, \tag{32}$$

and the total number of parameters is therefore bounded from below by

$$4 \sum_{i \neq j} \prod_{k \neq i,j} (\#A_k) \binom{\#A_i}{2}\binom{\#A_j}{2} \times P + \left(N \cdot \prod_i \#A_i - \sum_i \prod_{j \neq i} \#A_j\right) \times (\#X). \tag{33}$$

This is clearly larger than the number of observable conditional outcome probabilities in the population. For example, in the symmetric case where $\#A_i = \#A$ for all i, the lower bound on the total number of parameters is

$$\binom{N}{2} (\#A)^N \frac{(\#A - 1)^2}{2} \times P + \left(N (\#A)^N - N (\#A)^{N-1}\right) \times \#X, \tag{34}$$

and the total number of conditional outcome probabilities is

$$\left((\#A)^N - 1\right) \times P \tag{35}$$

and it is clearly smaller. Q.E.D.

10.3 Proof of Theorem 4.

To begin with, fix x the vector of covariates that enter the payoff functions $f_i\left(a, x\right)$. These are determined by 4 unknown parameters which we label as

$$f_1\left(BL, x\right), f_1\left(BR, x\right), f_2\left(TR, x\right), f_2\left(BR, x\right).$$

The space of the error terms

$$\varepsilon = (\varepsilon_1(BL), \varepsilon_2(TR), \varepsilon_1(BR), \varepsilon_2(BR))$$

can be partitioned into 7 regions which we label as $R_0, ..., R_6$. Let R_0 be the region with only one mix strategy equilibrium. Let regions 1 through 4 correspond to the those utilities that have pure strategy equilibrium

$$(T, L), (T, R), (B, L), (B, R)$$

respectively. Let region 5 be where the set of equilibria is

$$\{(T, L), (B, R), \textit{mixed strategy equilibrium}\}.$$

Let region 6 be where

$$\{(T, R), (B, L), \textit{mixed strategy equilibrium}\}$$

is an equilibrium. The probability of landing in a given region is a known, parametric function of $f_i\left(a, x\right)$. Let

$$p(R_n) = H_n(f) = P\left(\varepsilon \in R_n | f\right), \quad \text{for} \quad n = 0, \ldots, 6$$

denote this function. In fact,

$$\begin{aligned} R_0 \quad = \quad & \left\{ \begin{array}{c} \varepsilon_1\left(BL\right) < -f_1\left(BL, x\right), \varepsilon_1\left(BR\right) > -f_1\left(BR, x\right), \\ \varepsilon_2\left(TR\right) > -f_2\left(TR, x\right), \varepsilon_2\left(BR\right) < -f_2\left(BR, x\right) \end{array} \right\} \\ & \cup \left\{ \begin{array}{c} \varepsilon_1\left(BL\right) > -f_1\left(BL, x\right), \varepsilon_1\left(BR\right) < -f_1\left(BR, x\right), \\ \varepsilon_2\left(TR\right) < -f_2\left(TR, x\right), \varepsilon_2\left(BR\right) > -f_2\left(BR, x\right) \end{array} \right\}. \end{aligned}$$

$$\begin{aligned} R_1 &= \{\varepsilon_1(BL) < -f_1(BL,x), \varepsilon_2(TR) < -f_2(TR,x)\} \\ &\cap \{\varepsilon_1(BR) > -f_1(BR,x), \varepsilon_2(BR) > -f_2(BR,x)\}^c . \end{aligned}$$

$$\begin{aligned} R_4 &= \{\varepsilon_1(BL) < -f_1(BL,x), \varepsilon_2(TR) < -f_2(TR,x)\}^c \\ &\cap \{\varepsilon_1(BR) > -f_1(BR,x), \varepsilon_2(BR) > -f_2(BR,x)\} . \end{aligned}$$

$$\begin{aligned} R_5 &= \{\varepsilon_1(BL) < -f_1(BL,x), \varepsilon_2(TR) < -f_2(TR,x)\} \\ &\cap \{\varepsilon_1(BR) > -f_1(BR,x), \varepsilon_2(BR) > -f_2(BR,x)\} . \end{aligned}$$

$$\begin{aligned} R_2 &= \{\varepsilon_1(BR) < -f_1(BR,x), \varepsilon_2(TR) > -f_2(TR,x)\} \\ &\cap \{\varepsilon_1(BL) > -f_1(BL,x), \varepsilon_2(BR) < -f_2(BR,x)\}^c . \end{aligned}$$

$$\begin{aligned} R_3 &= \{\varepsilon_1(BR) < -f_1(BR,x), \varepsilon_2(TR) > -f_2(TR,x)\}^c \\ &\cap \{\varepsilon_1(BL) > -f_1(BL,x), \varepsilon_2(BR) < -f_2(BR,x)\} . \end{aligned}$$

$$\begin{aligned} R_6 &= \{\varepsilon_1(BR) < -f_1(BR,x), \varepsilon_2(TR) > -f_2(TR,x)\} \\ &\cap \{\varepsilon_1(BL) > -f_1(BL,x), \varepsilon_2(BR) < -f_2(BR,x)\} . \end{aligned}$$

It is easy to check that these regions are both exclusive of each other and exhaustive of the space of ε. They depend only on the sign of the sum of the deterministic and random utility components.

By our assumptions A4 (i)-(iv), the following system of equations must hold, with slight abuse of notation

$$
\begin{aligned}
\text{If } z &= z_{(1)} \\
p(TL|x, z_{(1)}) &= H_1(f) + \int_{R_0} p_m\left(TL|x, \varepsilon\right) dF\left(\varepsilon\right) + H_5(f) \\
p(TR|x, z_{(1)}) &= H_2(f) + \int_{R_0} p_m\left(TR|x, \varepsilon\right) dF\left(\varepsilon\right) + H_6(f) \\
p(BL|x, z_{(1)}) &= H_3(f) + \int_{R_0} p_m\left(BL|x, \varepsilon\right) dF\left(\varepsilon\right) \\
p(BR|x, z_{(1)}) &= H_4(f) + \int_{R_0} p_m\left(BR|x, \varepsilon\right) dF\left(\varepsilon\right) \\
\text{If } z &= z_{(2)} \\
p(TL|x, z_{(2)}) &= H_1(f) + \int_{R_0} p_m\left(TL|x, \varepsilon\right) dF\left(\varepsilon\right) \\
p(TR|x, z_{(2)}) &= H_2(f) + \int_{R_0} p_m\left(TR|x, \varepsilon\right) dF\left(\varepsilon\right) \\
p(BL|x, z_{(2)}) &= H_3(f) + \int_{R_0} p_m\left(BL|x, \varepsilon\right) dF\left(\varepsilon\right) + H_6(f) \\
p(BR|x, z_{(2)}) &= H_4(f) + \int_{R_0} p_m\left(BR|x, \varepsilon\right) dF\left(\varepsilon\right) + H_5(f)
\end{aligned}
$$

where $p_m\left(TL|x, \varepsilon\right)$ is the probability that TL is being played in a mix strategy equilibrium, where the mixing probabilities are functions of x and ε. It is easy to calculate that

$$p_m\left(TL|x, \varepsilon\right) = p_m\left(T|x, \varepsilon\right) \cdot p_m\left(L|x, \varepsilon\right)$$

such that

$$p_m\left(L|x, \varepsilon\right) = \frac{f_1(BR, x) + \varepsilon_1(BR)}{f_1(BR, x) + \varepsilon_1(BR) - (f_1(BL, x) + \varepsilon_1(BL))}$$

and

$$p_m\left(T|x, \varepsilon\right) = \frac{f_2(BR, x) + \varepsilon_2(BR)}{f_2(BR, x) + \varepsilon_2(BR) - (f_1(TR, x) + \varepsilon_1(TR))}.$$

Similarly

$$p_m\left(BR|x, \varepsilon\right) = p_m\left(B|x, \varepsilon\right) \cdot p_m\left(R|x, \varepsilon\right),$$

where

$$p_m\left(B|x, \varepsilon\right) = 1 - p_m\left(T|x, \varepsilon\right)$$

and

$$p_m\left(R|x,\varepsilon\right) = 1 - p_m\left(L|x,\varepsilon\right).$$

The requirement that all of the probabilities sum up to one makes the above system of equations effectively a system of 6 equations in in 4 unknowns for each x conditional on $z_{(1)}$ and $z_{(2)}$. Therefore, the deterministic part of utilities, $f_i\left(a,x\right), i = 1, 2$, can be solved as a function of the 6 observable outcome probabilities for any given x. Since we can repeat this argument for any x, the function that determines the deterministic part of payoffs, $f_i(x, a)$, can be identified from the conditional outcome probabilities for any given x. Q.E.D.

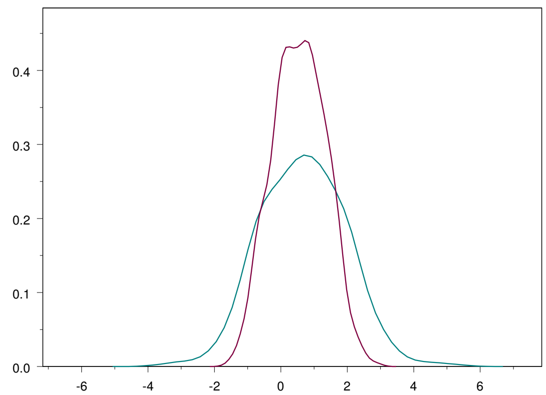

Figure 1: The Posterior Marginal Distribution of Beta

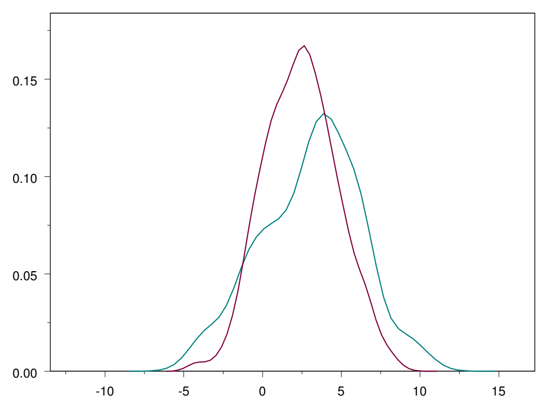

Figure 2: The Posterior Marginal Distribution of Delta

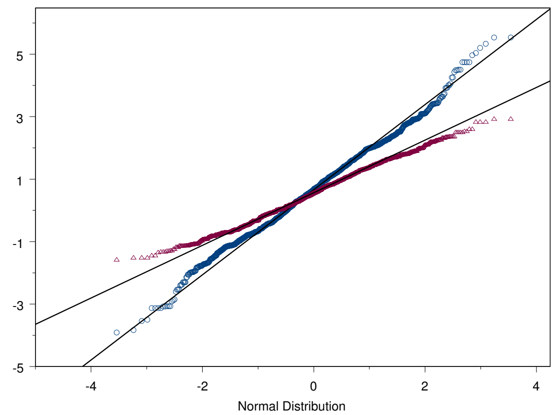

Figure 3: QQ Plot for Beta

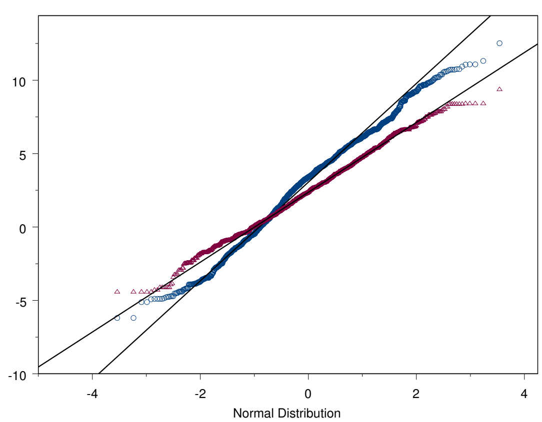

Figure 4: QQ Plot for Delta