A Note on Semiparametric Estimation of Finite Mixtures of Discrete Choice Models with Application to Game Theoretic Models

Patrick Bajari, Jinyong Hahn, Han Hong and Geert Ridder[1]

Shortened Title: "Semiparametric Estimation of Mixtures"

Abstract: We view a game abstractly as a semiparametric mixture distribution, and study the semiparametric efficiency bound of this model. Our results suggest that a key issue for inference is the number of equilibria compared to the number of outcomes. If the number of equilibria is sufficiently large compared to the number of outcomes, root-n consistent estimation of the model will not be possible. We also provide a simple estimator in the case when the efficiency bound is strictly above zero.

[1]Helpful comments by Bryan Graham and Stephane Bonhomme are greatly acknowledged. Financial support for this research was generously provided through NSF grants SES 0921187, 0819638 and 0721015.

1 Introduction

Recently, there has been considerable interest in the econometric analysis of static, discrete games. In this literature, a game is a generalization of a standard discrete choice model, such as a binary logit or probit. As in a discrete choice model, an agent's payoffs are a parametric function of exogenous covariates and i.i.d. preference shocks. In a game, however, payoffs are also allowed to depend on the actions of other agents.

Multiplicity of equilibria is a common feature of many game theoretic models. As noted by Berry and Tamer (2006), the multiplicity of equilibrium can cause difficulties for both the identification and estimation of these models. In the literature, models with multiple equilibrium have been treated as mixtures. That is, a particular equilibrium gives the probability with which the agents actions should be observed and the likelihood is formed by parameterizing the probability over alternative equilibria. See for example Ackerberg and Gowrisankaran (1996), Sweeting (2005) and Bajari, Hong and Ryan (2009). Following this earlier work, we view a game as a semiparametric mixture model. Each equilibrium characterizes the probability distribution of the outcome variable which is parametrically specified. Econometricians usually do not observe which equilibrium to the model was selected and hence generated the outcome. Economic theory does not give guidance on how to choose between multiple equilibria in general. Therefore, it is natural to characterize the game theoretic model with multiple equilibria as a semiparametric mixture, with the equilibrium selection mechanism being the nonparametric component.

We consider two classes of semiparametric mixture models. The first model is motivated by incomplete information games (see Seim (2006) and Aradillas-Lopez (2005)). Our main contribution there is to derive a simple counting condition, which can help us understand identification of the structural parameters. The condition compares the number of possible outcomes with the number of possible equilibria. If there are too many equilibria, then identification and estimation of parameters are difficult in general. We also provide a simple moment condition that provides an intuition behind this counting condition. The moment condition can be used to estimate the structural parameters as well.

The second mixture model is motivated by complete information games (see Bresnahan and Reiss (1990, 1991), Berry (1992), Tamer (2003), Bajari, Hong and Ryan (2009) and Ciliberto and Tamer (2007)). In the first model, the mixing probability usually only depends on variable(s) observed by econometricians. In the second model, though, the mixing probability might also depend on variable(s) unobserved by econometricians. The unobserved variable in the mixing probability effectively increases the dimension of the nonparametric component for the same amount of sample information, and the structural parameter is unlikely to have positive information. Exclusion restrictions or other parametric restrictions on equilibrium selection will probably be necessary for identification and estimation as a result. Partial identification approach is another possibility if exclusion restrictions are not credible. See Aradillas-Lopez and Tamer (2008) and Beresteanu, Molchanov, and Molinari (2008), e.g., for recent application of the partial identification approach to game theoretic models.

The identification of discrete games has been studied previously by numerous researchers including Bresnahan and Reiss (1991), Tamer (2003), Sweetings (2005), Bajari, Hong and Ryan (2009) and Pesendorfer and Schmidt-Dengler (2003). Previous work on private information games, for example, takes as given the ability of the researcher to directly observe the equilibrium probabilities. As is well known, this literature makes an implicit assumption that the data implicitly "selects a unique equilibrium". We consider the more general, and possibly more realistic, situation in which the observed actions in a discrete game are a mixture of different equilibria to the model.

This paper contributes to this literature by viewing a game abstractly as a semiparametric mixture distribution and by studying the semiparametric efficiency bound to this model. If the bound is zero, $\sqrt{n}$ estimation of the parametric part of the model is not possible, which suggests that using the model in empirical work may be problematic. Our results suggest that a key issue for inference is the number of equilibria compared to the number of outcomes. If the number of equilibria is sufficiently large compared to the number of outcomes, $\sqrt{n}$ estimation of the model will not be possible. This can be the case in models of complete information. Other restrictions on the model, such as exclusion restrictions on how equilibria are selected or parameterizing equilibrium selection mechanisms, will be necessary to estimate the model parameters at the parametric rate. Our second contribution is to provide a simple estimator in the case that the efficiency bound is strictly above zero. As we discuss in the paper, this will often be the case for games of incomplete information. To the best of our knowledge, these results are new to the literature.

2 Simple Examples

In the literature, there have been two main approaches to modeling discrete games. The first is to assume that agents have complete information. The second is to assume that the preference shocks are private information to the agents. As we shall demonstrate, the identification and estimation differs markedly across these two classes of models. Therefore, it is useful to provide a simple example of both models to fix ideas and illustrate these differences.

In what follows, we shall describe a simple entry game in both the incomplete and complete information cases. We deliberately consider the simplest possible examples. These simple examples will illustrate some key differences between both models and help us to illustrate the more general framework in the next sections.

2.1 Private Information Games

In the existing literature, entry has been perhaps the most widely studied application of discrete games. See for example Seim (2006), Sweetings (2005) and Aradillas-Lopez (2005). In dynamic games, the private information assumption is made by Aguirregabiria and Mira (2007) and Pesendorfer and Schmidt-Dengler (2003). In the model, there are two firms, $i = 1, 2$. Each firm must make a decision whether to enter a market given a set of exogenous variables x and its private profit shock. Following the literature, we let $a_i = 1$ $(= 0)$ denote the decision to enter (not enter) the market. The profits to firm i can be written as:

$$u_i = \begin{cases} x'\beta + \delta a_{-i} + \varepsilon_i \text{ if } a_i = 1 \\ 0 \text{ if } a_i = 0 \end{cases} \tag{1}$$

In equation (1), i's profits depends on three terms. The first term $x'\beta$ captures the effects of exogenous covariates x given a vector of parameters β. In applied work, these exogenous covariates include variables which proxy for consumer demand and the costs of firm i. The term δa_{-i} captures the impact of entry by the competing firm $-i$ on firm i's profits. Oligopoly models typically imply that entry by $-i$ lowers i's profits. Therefore, it is reasonable to expect that $\delta < 0$ in most settings. The term ε_i is an iid random shock to firm i's profits. This term is meant to capture shocks to firm i's profits which are observed by the firm, but not the econometrician. In this example, we shall assume that these shocks are independent of x and distributed extreme value so that we can exploit well known properties of the logit model.

In the private information game, it is assumed that firm i knows ε_i but it does not know ε_{-i}. As a result, firm i does not know the payoffs of firm $-i$ and typically will not be able to infer $-i$'s equilibrium strategy as a result.

In this environment, it is conventional to focus on Bayes-Nash equilibrium. Let $\sigma_{-i}(a_{-i} = 1|x)$ denote player i's beliefs about the probability that player $-i$ will enter with x the vector of of all variables in the x for the potential entrants. In a Bayes-Nash equilibrium, firm i makes a best response to these beliefs. This implies that

$$a_i = 1 \Leftrightarrow x'\beta + \delta\sigma_{-i}(a_{-i} = 1|x) + \varepsilon_i > 0. \tag{2}$$

That is, firm i will enter the market if and only if the expected profits from entering are greater than zero.

Assuming that the error terms are distributed extreme value, equation (2) implies that

$$\sigma_1(a_1 = 1|x) = \frac{\exp(x'\beta + \delta\sigma_2(a_2 = 1|x))}{1 + \exp(x'\beta + \delta\sigma_2(a_2 = 1|x))}, \tag{3}$$

$$\sigma_2(a_2 = 1|x) = \frac{\exp(x'\beta + \delta\sigma_1(a_1 = 1|x))}{1 + \exp(x'\beta + \delta\sigma_1(a_1 = 1|x))}. \tag{4}$$

Note that the probability that 1 enters, $\sigma_1(a_1 = 1|x)$, depends on the probability that 2 enters, $\sigma_2(a_2 = 1|x)$ through a well known expression from the binary logit model. The equilibrium strategies to this game, $\sigma_1(a_1 = 1|x)$ and $\sigma_2(a_2 = 1|x)$, can therefore be viewed as a solution to these two non-linear equations.

Next, we derive the likelihood function implied by our model (3) and (4). It is well known that discrete games of incomplete information, such as the simple entry example above, need not have a single equilibrium. This can cause difficulties for both the identification and estimation of the model. Therefore, the likelihood function needs to account for the possibility of multiple equilibrium.

In this two by two entry game, there are 4 possible outcomes. These outcomes are $(a_1 = 1, a_2 = 1)$,$(a_1 = 0, a_2 = 1)$, $(a_1 = 1, a_2 = 0)$ and $(a_1 = 0, a_2 = 0)$. We index these outcomes by $k = 1, ..., K + 1$. Since probabilities must sum to one, one of these outcomes is redundant. Therefore we can focus on K $(= 3)$ outcomes without loss of generality.

Let $\theta_0 = (\beta_0, \delta_0)$ denote the true parameters. Given the parameters and covariates x, there are $j = 1, ..., J + 1$ equilibria to the model. These equilibria correspond to all possible solutions to the equations (3) and (4). These equations are nonlinear, and therefore, more than one solution is possible, as demonstrated by Brock and Durlauf (2001), Sweetings (2005) and Bajari, Hong, Krainer and Nekipelov (2006). In general, the number of solutions to (3) and (4), and hence the number of equilibria, $J + 1$, will depend on θ_0 and x. Although the J should in principle be written J_x, we shall frequently suppress this dependence. Let $p_k^{(j)}(x, \theta_0)$ denote the probability of outcome k in equilibrium j. This can be derived in the obvious way using σ_1 and σ_2.

Since there is more than one equilibrium, a complete econometric model must specify how these outcomes are chosen. We let $\lambda_0^{(j)}(x)$ denote the probability that equilibrium j

is selected. Researchers have proposed a number of mechanisms for defining $\lambda_0^{(j)}(x)$. For example, Berry (1992) restricts attention to equilibria which are profit maximizing for an incumbent firm in an entry game. Ackerberg and Gowrisankaran (2006) restrict attention to "extremal" equilibria in a supermodular game. In our analysis, we will let $\lambda_0^{(j)}(x)$ be a general nonparametrically specified function of x.[2] This admits the earlier approaches as special cases and allows for a general dependence between covariates and how the equilibrium is selected. Tamer (2003) has argued that it is desirable to be agnostic about how equilibrium is selected given that economic theory does not provide guidance to picking out a unique equilibrium to games.

Then, the probability $p_k(x, \theta_0)$ of observing outcome k conditional on x is equal to

$$p_k(x, \theta_0) = \sum_{j=1}^{J+1} p_k^{(j)}(x, \theta_0) \lambda_0^{(j)}(x) \tag{5}$$

In our analysis of identification, we shall be concerned with whether we can uniquely recover θ from observing the probability $p_k(x, \theta)$. The difficulty, as seen in equation (5), is that we do not directly observe the equilibrium of the game $p_k^{(j)}(x, \theta)$. Instead, we see a mixture of the equilibria, marginalizing out their probabilities using $\lambda^{(j)}(x)$. Thus a complication in the analysis of identification is whether we can separate the equilibrium selection mechanism, $\lambda_0^{(j)}(x)$, from the probability that outcome k is observed in a particular equilibrium, $p_k^{(j)}(x, \theta_0)$.

2.2 Complete Information Game

A second case that the literature has studied is games of complete information (see Tamer (2003), Pakes, Porter, Ho and Ishii (2005), Bajari, Hong and Ryan (2009) and Ciliberto and Tamer (2007)). In these games, the random preference shocks are observed by all of the agents, so that the payoffs are common knowledge.

As in the previous example, assume that there are two players in an entry game. As before, the actions are $a_i \in \{0, 1\}$ for $i = 1, 2$. Payoffs depend on actions $a = (a_1, a_2)$,covariates (x) and on iid random preference shocks. In the complete information case, we let the preference shocks depend on the actions of both agents and we write $\varepsilon_i(a_1, a_2)$. In contrast, the preference shock in the incomplete information case only depends on i's own actions. This

[2] Implicitly, λ_0 depends on the value of the true parameters θ_0. Because the $\lambda_0^{(j)}(\cdot)$ is nonparametrically specified, we drop this dependence.

means that we can write payoffs in the following manner:

$$u_i = \begin{cases} x'\beta + \delta a_{-i} + \varepsilon_i(a) \text{ if } a_i = 1 \\ 0 \text{ if } a_i = 0 \end{cases}$$

Analogous to (1), i's utility depends on three factors. The term $x'\beta$ captures the impact of exogenous covariates on profits, δa_{-i} the impact of entry by $-i$ on i's profits and $\varepsilon_i(a)$ the impact of variables seen by the agents but not by the econometrician on profits.

In this model, payoffs are observable to all agents. Therefore, the payoffs can be summarized in the following matrix form:

players 1/2	$a_2 = 1$	$a_2 = 0$
$a_1 = 1$	$\begin{pmatrix} x'\beta - \delta + \varepsilon_1(1,1), \\ x'\beta - \delta + \varepsilon_2(1,1) \end{pmatrix}$	$\begin{pmatrix} x'\beta + \varepsilon_1(1,0), \\ 0 \end{pmatrix}$
$a_1 = 0$	$\begin{pmatrix} 0, \\ x'\beta + \varepsilon_2(0,1) \end{pmatrix}$	$\begin{pmatrix} 0,0 \end{pmatrix}$

(6)

In the above, we write, for example, $\varepsilon_1(1,1)$ to denote the preference shock to player 1 corresponding to the event where player 1 and player 2 both choose to enter.

In the econometric model, we assume that the observed behavior is a Nash equilibrium. That is, agent i chooses a probability distribution $\pi_i(a_i)$ which maximizes his utility given the equilibrium strategy $\pi_{-i}(a_{-i})$ of the other agent. It is well known that generically a 2×2 game has either a unique equilibrium or three equilibria. There are two cases in which there are three equilibria to the game. In the first case, $(a_1 = 1, a_2 = 1)$ and $(a_1 = 0, a_2 = 0)$ are both pure strategy equilibria and there is a mixed strategy equilibrium, where all strategies have strictly positive probability. In the second case, $(a_1 = 1, a_2 = 0)$ and $(a_1 = 0, a_2 = 1)$ are both pure strategy equilibria and there is a mixed strategy equilibrium with strictly positive probability on all strategies.

As in the incomplete information game, the number of equilibria $J + 1$ and equilibrium strategies π_i depend on the exogenous variables x and the parameters $\theta = (\beta, \delta)$. Similarly, the potential outcomes are $(a_1 = 1, a_2 = 1)$,$(a_1 = 0, a_2 = 1)$, $(a_1 = 1, a_2 = 0)$ and $(a_1 = 0, a_2 = 0)$. We index these outcomes by $k = 1, ..., K + 1$ and focus on K $(= 3)$ outcomes without loss of generality.

We shall let $\lambda_0^{(j)}(x, \varepsilon)$ denote the probability that the j equilibrium is chosen. Note that this will in general depend on the preference shocks ε, which was not true in the incomplete information case. Then, the probability $p_k(x,\theta_0)$ of observing outcome k conditional on x is equal to

$$p_k(x,\theta_0) = \int \sum_{j=1}^{J+1} p_k^{(j)}(x,\varepsilon,\theta_0)\,\lambda_0^{(j)}(x,\varepsilon)\,f_\varepsilon(\varepsilon)\mathrm{d}\varepsilon \tag{7}$$

In equation (7), $f_\varepsilon(\varepsilon)$ is the density for $\varepsilon = (\varepsilon_1, \varepsilon_2)$. Notice that (7) is significantly more complicated than (5). This is because the preference shocks enter both $p_k^{(j)}(x,\varepsilon,\theta_0)$ and $\lambda_0^{(j)}(x,\varepsilon)$. We therefore need to marginalize out ε using the density f_ε. As a result, complete information games will require stronger conditions for identification and estimation.

3 Mixture of Discrete Response Model

In this section, we analyze a general model that includes the incomplete information game as a special case. Consider a mixture of discrete response models with $K+1$ possible outcomes $\kappa_1, \ldots, \kappa_{K+1}$. Because of redundancy of the last outcome, we will, without loss of generality, consider the K-dimensional random vector of outcome indicators y as the vector of dependent variables. Let x denote the vector of exogenous variables, and θ the d-dimensional parameter vector.

The probability distribution of the dependent variable y, which can be described by the K-dimensional vector of response probabilities, differs depending on the "equilibrium" selected. Let $J+1$ denote the number of possible equilibria, which is assumed to be known. As discussed in the previous section, the number $J+1$ of equilibria can in principle depend on x, but we suppress the dependence for notational simplicity.

Let θ_0 denote the true value of the parameter θ. Each equilibrium j is characterized by a vector of probabilities

$$p^{(j)}(x,\theta_0) = \begin{bmatrix} p_1^{(j)}(x,\theta_0) \\ \vdots \\ p_K^{(j)}(x,\theta_0) \end{bmatrix} = \begin{bmatrix} \Pr\left[y = \kappa_1 \middle| x, \text{ equilibrium is } j\right] \\ \vdots \\ \Pr\left[y = \kappa_K \middle| x, \text{ equilibrium is } j\right] \end{bmatrix}$$

We will adopt a technical regularity condition that the $p^{(j)}(x,\theta)$ are continuously differentiable in θ throughout the rest of the paper. Inspection of equations (3) and (4) demonstrates this condition is satisfied in the entry game example.

We only observe the outcome y, i.e. we do not know which equilibrium has generated that outcome. The probability that equilibrium j generates the outcome given x is $\lambda_0^{(j)}(x)$,

$j = 1, \ldots, J+1$. Because the $\lambda_0^{(j)}(x)$ add up to one, we can succinctly summarize them in the J-dimensional vector $\lambda_0(x) = \left(\lambda_0^{(1)}(x), \ldots, \lambda_0^{(J)}(x)\right)'$.

Let $p(x)$ denote the K-dimensional vector that characterizes the distribution of the random vector y. Analogous to equation (5), we can express it in the general model as

$$p(x) = \sum_{j=1}^{J+1} p^{(j)}(x, \theta_0) \lambda_0^{(j)}(x), \tag{8}$$

where $\lambda_0^{(J+1)}(x) = 1 - \lambda_0^{(1)}(x) - \cdots - \lambda_0^{(J)}(x)$. In matrix terms, we can write

$$p(x) - p^{(J+1)}(x, \theta_0) = Q(x, \theta_0) \lambda_0(x) \tag{9}$$

where $Q(x, \theta_0)$ is a $K \times J$ matrix such that

$$Q(x, \theta_0) = \left[p^{(1)}(x, \theta_0) - p^{(J+1)}(x, \theta_0), \ldots, p^{(J)}(x, \theta_0) - p^{(J+1)}(x, \theta_0)\right].$$

Our objective is to come up with a semiparametric estimator of θ_0. As suggested by the notation, the conditional outcome probabilities $p^{(j)}(x, \theta)$ are parametric, but the conditional equilibrium selection probabilities $\lambda_0^{(j)}(x)$ are nonparametrically specified.

3.1 Information Bound

We present our result on the semiparametric information bound for θ_0. Our result shows that the condition for zero information can be easily characterized by comparing J and K.[3] To simplify our analysis, we assume that the number of equilibria is invariant with respect to x. The more general case, where the number of equilibria depends on x, is a simple generalization of the results below. We note however, that characterizing the number of equilibria in incomplete information discrete games is a difficult issue. For a simple two by two game under the logistic error distribution assumption without exogenous covariates, it is known that the number of equilibria is either one or three. With more than two players, the number of equilibria is in general odd under the same set of assumptions. However, in general, not much is known about characterizing the number of equilibria in incomplete information games, which can depend on the covariate variables.

Theorem 1 *Suppose that the number of equilibria is fixed at J. Also suppose that $J \geq K$, and the rank of the matrix $Q(x, \theta_0)$ is K. Then the semiparametric information bound for θ is zero.*

[3] Our result is mathematically similar to Chamberlain's (1992, Section 4) analysis of some panel models.

Proof. See Appendix A. ■

In practice the rank of the matrix $Q(x, \theta_0)$ is likely to be equal to $\min(J, K)$, and in that case Theorem 1 simply means that the information is zero if $J \geq K$.

In our example (3) and (4), the information is zero if the number of equilibria is greater than or equal to 4. In those equations, the number possible outcomes is equal to 4, so $K = 3$. Therefore the information is zero if $J \geq 3$, or if the number of equilibria is equal to or larger than 4.

Zero semiparametric information does not necessarily imply that the parameter θ is not identified. Manski (1985) considered a binary response model with median restrictions, and showed that the parameter is identified and consistently estimated by the maximum score estimator, but Chamberlain (1986) proved that its semiparametric information bound is zero. In duration analysis, Ridder (1990) demonstrated identification of parameters of a mixed proportional hazard model, but Hahn (1994) showed that the semiparametric information bound is zero for Weibull models. To understand this phenomenon, it is convenient to recall that the Cramer-Rao lower bound is the inverse of the Fisher Information, a parametric analog of the semiparametric information bound. Zero semiparametric information can therefore be intuitively understood to mean that the asymptotic variance of a $\sqrt{n}$-consistent estimator must be infinity. In other words, an estimator in a model with zero information cannot be $\sqrt{n}$-consistent. Zero information is often associated with models identified at infinity, see for example Heckman (1990). When a parameter is identified but has a zero information, it is in principle possible to construct consistent estimators. But the consistent estimators tend to have more complex asymptotic distributions, rendering statistical inference more difficult in practice. A zero information bound may therefore be understood as a useful criterion to measure the feasibility of empirical analysis.

3.2 Intuition

To appreciate the intuition underlying Theorem 1, consider the simple case where x has a multinomial distribution with M support points, and the number of equilibria does not depend on x and is fixed at $J + 1$. Note that the model has $d + MJ$ unknown parameters, where the d comes from the number of components in θ and the MJ comes from the number of $\lambda_0^{(1)}(x), \ldots, \lambda_0^{(J)}(x)$ for M different values of x. Also note that the model gives us K moment restrictions in $p(x)$ for M different values of x. In other words, identification requires that we need to solve for $d + MJ$ unknown parameters from MK equations. We can in general find the unique solution only if $MK \geq d + MJ$. Assuming that M is sufficiently large, the condition $MK \geq d + MJ$ is equivalent to

$$K > J. \tag{10}$$

The discussion can be made a little more rigorous by separately considering two cases: $K > J$ and $K \leq J$. If $K > J$ we can solve for $\lambda_0(x)$ in (9)

$$\lambda_0(x) = \left(Q(x,\theta_0)' Q(x,\theta_0)\right)^{-1} Q(x,\theta_0)' \left(p(x) - p^{(J+1)}(x,\theta_0)\right).$$

Substituting this in (9) gives

$$p(x) - p^{(J+1)}(x,\theta_0) = Q(x,\theta_0)\left(Q(x,\theta_0)' Q(x,\theta_0)\right)^{-1} Q(x,\theta_0)' \left(p(x) - p^{(J+1)}(x,\theta_0)\right).$$

Because $I - Q(x,\theta_0)\left(Q(x,\theta_0)' Q(x,\theta_0)\right)^{-1} Q(x,\theta_0)'$ has rank $K - J$, this corresponds to $K - J$ independent equations[4] in $p_k^{(j)}(x,\theta_0), k = 1,\ldots,K, j = 1,\ldots,J+1$. For instance, if $J = 1$ the equations are

$$p_k(x) - p_k^{(2)}(x,\theta) = \frac{p_k^{(1)}(x,\theta) - p_k^{(2)}(x,\theta)}{p_1^{(1)}(x,\theta) - p_1^{(2)}(x,\theta)}(p_1(x) - p_1^{(2)}(x,\theta)),\ k = 2,\ldots,K.$$

Therefore θ_0 is not identified in general if $K \leq J$. If $K > J$ we have $M(K-J)$ equations in d unknown parameters θ_0. Hence if M is sufficiently large, the parameters in the outcome probabilities are likely to be identified.

3.3 Estimation

Our discussion of the information bound suggests a new estimator based on some moment restrictions when the information is not zero. The estimator allows practitioners to avoid estimating the nonparametric component $\lambda(x)$, which is necessary if a likelihood based approach is adopted. The moment restrictions do not use every aspect of the model, and therefore, the resultant GMM estimator is not expected to be semiparametrically efficient.

When $J \geq K$, we may adopt a slightly different approach to understand the nonidentification result of Theorem 1. Let θ be arbitrary. For each x, we can "solve" for $\lambda(x)$ that satisfies the restriction

$$E\left[\left(y - p^{(J+1)}(x,\theta_0)\right) - Q(x,\theta)\lambda(x)\middle|\, x\right] = 0$$

by choosing

$$\lambda(x) = \left(Q(x,\theta)' Q(x,\theta)\right)^{+1} Q(x,\theta)' \left(p(x) - p^{(J+1)}(x,\theta)\right) \tag{11}$$

[4]If the ranks of $Q(x,\theta_0)$ is less than K, the number of equations is the rank minus J.

if $J \geq K$, or more precisely, if the rank of the matrix $Q(x,\theta)$ is K and $J \geq K$. Here $\left(Q(x,\theta)' Q(x,\theta)\right)^{+1}$ is the generalized inverse. As long as $\lambda(x)$ thus solved belongs to the J-simplex, such arbitrary (θ, λ) cannot be distinguished from (θ_0, λ_0), hence θ_0 is not identified.

The above discussion suggests that the model parameters are identified if the rank of the $K \times J$ matrix $Q(x,\theta)$ is strictly less than K. The above discussion can be used to construct a simple estimator in case the information is not zero. Our proposal is based on the simple intuition that we can view

$$E\left[\left(y - p^{(J+1)}(x,\theta_0)\right) - Q(x,\theta_0)\lambda(x)\middle| x\right] = 0 \tag{12}$$

as a population linear regression problem. We can imagine "estimating" $\lambda_0(x)$ by the "OLS estimator" $\left(Q(x,\theta_0)' Q(x,\theta_0)\right)^{-1} Q(x,\theta_0)' \left(y - p^{(J+1)}(x,\theta_0)\right)$. The "residual" in this imaginary "regression" is equal to

$$M(x,\theta_0)\left(y - p^{(J+1)}(x,\theta_0)\right).$$

with

$$M(x,\theta_0) = I_K - Q(x,\theta_0)\left(Q(x,\theta_0)' Q(x,\theta_0)\right)^{-1} Q(x,\theta_0)'$$

a matrix of rank $K - J$. By construction, this "residual" is guaranteed to have zero mean conditional on x. This implies that we are given the restriction

$$E\left[M(x,\theta_0)\left(y - p^{(J+1)}(x,\theta_0)\right)\middle| x\right] = 0 \tag{13}$$

which can be estimated with the usual simple GMM. We note that our proposal does not involve a practitioner actually involve estimating $\lambda_0(x)$ in practice. The "regression" is just an intuitive device to come up with the moment restriction (13).

To illustrate how our equation (13) can be used, consider a simple 2×2 game, which means that there are four possible outcomes, i.e., $K + 1 = 4$. Assume that the indeterminacy is such that we do not know the equilibrium selection of two outcome combinations, i.e.„ $J + 1 = 2$, and

$$p^{(1)}(x,\theta_0) = \begin{bmatrix} A(x,\theta_0) \\ B(x,\theta_0) \\ 0 \end{bmatrix}, \quad p^{(2)}(x,\theta_0) = \begin{bmatrix} A(x,\theta_0) \\ 0 \\ B(x,\theta_0) \end{bmatrix}$$

It can be shown that the "residual" $M(x,\theta_0)\left(y - p^{(J+1)}(x,\theta_0)\right)$ in this example is

$$\begin{bmatrix} y_1 - A(x,\theta) \\ \frac{1}{2}y_2 + \frac{1}{2}y_3 - \frac{1}{2}B(x,\theta) \\ \frac{1}{2}y_2 + \frac{1}{2}y_3 - \frac{1}{2}B(x,\theta) \end{bmatrix}$$

and the corresponding moment equation can be written (after some manipulations) as

$$E\left[\begin{array}{c|c} y_1 - A(x,\theta) & \\ y_2 + y_3 - B(x,\theta) & \end{array} x\right] = 0.$$

Because $y_2 + y_3$ is a binary variable, this moment equation is nothing but the method of moment counterpart of Maddala's (1983) maximum likelihood estimation of the trinomial distribution $(y_1, y_2 + y_3, y_4)$.

3.4 Dependence of J on x

So far, we have not explicitly treated the dependence of J on x and θ_0. In general, the number of possible equilibria will depend on the (x,θ_0) combination. This implies that the number of columns in $Q(x,\theta_0)$ may change as a function of (x,θ_0). Therefore, the $M(x,\theta_0)$ operator should be interpreted cautiously. For example, it is possible that there may exist only one equilibrium for some (x,θ_0), i.e., $J = 0$. Then the $M(x,\theta_0)$ should be interpreted to be an identity matrix.

If there are sufficiently many values of x such that $J = 0$ (unique equilibrium), we could in principle identify and estimate θ_0 by conditioning on such x without using our proposal. If the set of such x has a positive probability, as in the example discussed in Bajari, Hong, Krainer, and Nekipelov (2006) and Aradillas-Lopez (2006), then $\sqrt{n}$-consistent estimation should be feasible just by exploiting the set of unique equilibrium. In general, if there are sufficiently many values of x such that the information is positive, then $\sqrt{n}$-consistent estimation is possible just by conditioning on such set of x. Even when this is the case, our equation (13) can be a potential source of efficiency, by exploiting the entire support of x. The potential gain in efficiency may be substantial if the set of unique equilibrium has a positive yet small probability.

The number of equilibria to an incomplete information game must be checked on a case by case basis. One approach to this is analytical. For example, Sweeting (2005) and Brock and Durlauf (2001) analytically derive the number of equilibria to their models. A second approach is numerical. Bajari, Hong, Krainer, and Nekipelov (2006) discuss a method for computing the entire set of equilibrium to an incomplete information game using the all solution homotopy. In either case, if one finds that with positive probability $J < K$, then the model has non-zero information.

For many other cases, the set of such x has a zero probability, which implies that estimation based on a unique equilibrium cannot produce $\sqrt{n}$-consistent estimation. Our equation (13) can be potentially used to produce a $\sqrt{n}$-consistent estimator by using additional information.

When J is allowed to depend on x, Theorem 1 should be interpreted to be a local condition. The (global) information for θ_0 is the average of the "local" information over x. If the "local" information is zero for every x, then the global information would be equal to zero as a consequence. For example, if the number of equilibria is larger than K at every point x, then the global information is equal to zero and $\sqrt{n}$-consistent estimation is infeasible.

4 Mixtures with an Unobserved Variable

In this section, we consider another general discrete response model, which includes the complete information game as a special case. In this model, we have

$$p(x,\varepsilon) = \sum_{j=1}^{J+1} p^{(j)}(x,\varepsilon,\theta_0)\,\lambda_0^{(j)}(x,\varepsilon), \tag{14}$$

with x observed and ε unobserved. The mathematical structure of the underlying model (14) is similar to (8) except that the ε is not unobserved. We argue that $\sqrt{n}$-consistent estimation is impossible in general for this sort of situation.[5] In general, the number of equilibria $J+1$ will depend on both x and ε, but in order to give a simple intuition, we will assume that the number of equilibria $J+1$ does not depend on (x,ε,θ_0).

We first note that, although the moment restriction

$$E\left[\left(y - p^{(J+1)}(x,\varepsilon,\theta_0)\right) - Q(x,\varepsilon,\theta_0)\,\lambda(x,\varepsilon)\right| x,\varepsilon\Big] = 0 \tag{15}$$

is still valid, it is not directly usable because we cannot condition on ε. Does the moment restriction

$$E\left[\left(y - p^{(J+1)}(x,\varepsilon,\theta_0)\right) - Q(x,\varepsilon,\theta_0)\,\lambda(x,\varepsilon)\right| x\Big] = 0,$$

which is implied by (15), help estimation of θ_0? We are pessimistic for reasons outlined below.

[5]An important exception is discussed in Appendix B.

Our pessimism can be explained by exploiting the order condition discussed before. Assume for simplicity that ε has a discrete distribution with m support points $\{\varepsilon_1, \ldots, \varepsilon_m\}$ and corresponding probabilities $\{f_1, \ldots, f_m\}$. We will assume that $\{\varepsilon_1, \ldots, \varepsilon_m\}$ and $\{f_1, \ldots, f_m\}$ are known to the econometrician. Our moment restriction can then be rewritten as

$$E\left[y - \sum_{j=1}^{m} f_j p^{(J+1)}\left(x, \varepsilon_j, \theta_0\right) - \sum_{j=1}^{m} f_j Q\left(x, \varepsilon_j, \theta_0\right) \lambda_0\left(x, \varepsilon_j\right) \middle| x\right] = 0$$

Now write

$$\begin{aligned}
\mathbb{Q}\left(x, \theta_0\right) &= \left[f_1 Q\left(x, \varepsilon_1, \theta_0\right), \ldots, f_m Q\left(x, \varepsilon_m, \theta_0\right)\right] \\
\overline{p}^{(J+1)}\left(x, \theta_0\right) &= \sum_{j=1}^{m} f_j p^{(J+1)}\left(x, \varepsilon_j, \theta_0\right)
\end{aligned}$$

and

$$\mu_0(x) = \begin{bmatrix} \lambda_0\left(x, \varepsilon_1\right) \\ \vdots \\ \lambda_0\left(x, \varepsilon_m\right) \end{bmatrix}$$

Then our moment restriction can be succinctly written

$$E\left[y - \overline{p}^{(J+1)}\left(x, \theta_0\right) - \mathbb{Q}\left(x, \theta_0\right) \mu_0(x) \middle| x\right] = 0$$

We now recognize that, $\mathbb{Q}(x, \theta)$ plays the same mathematical role as $Q(x, \theta)$ in (12). In light of Theorem 1, we can conclude that the information for θ is zero in general if the rank of the matrix $\mathbb{Q}(x, \theta)$ is K. In order to verify this condition, it suffices to count the rank of

$$\left[Q\left(x, \varepsilon_1, \theta_0\right), \ldots, Q\left(x, \varepsilon_m, \theta_0\right)\right].$$

Note that the number mJ of columns of $\mathbb{Q}(x, \theta)$ is in general a lot bigger than the number J of columns of $Q(x, \theta)$. Therefore, we anticipate that $\sqrt{n}$-consistent estimation of θ_0 is impossible in general.[6] Note that the ε is usually assumed to have a continuous distribution, and the problem explained here can only be more difficult there.

In the previous two sections we compared complete information models with incomplete information models, and concluded that complete information models are more difficult to

[6] We can relax our assumption a little bit and allow that the number of equilibria may depend on ε. This implies that the number of columns of $Q\left(x, \varepsilon_j, \theta_0\right)$ may depend on ε_j. Other than that, the order condition remains.

identify in the presence of a nonparametric latent equilibrium selection mechanism. These two types of models cover the extreme ends of the spectrum of models regarding the trade-off between the degree of unobserved heterogeneity that can be accommodated and the amount of asymmetric information. Unobserved heterogeneity in discrete games refers to variables that are common knowledge to the players but unobserved by the econometrician. Complete information games incorporates a substantial amount of unobserved heterogeneity but requires that all information is symmetric among players. A standard incomplete information game that we study allows for asymmetric information between players but rules out unobserved heterogeneity. A more realistic model which incorporates both of these elements is necessarily even more difficult to identify without imposing additional assumptions. While some progress has been in the literature in understanding unobserved heterogeneity in dynamic discrete choice models, such as the results by Kasahara and Shimotsu (2008) and Hu and Shum (2009), the focus has been on models with a unique equilibrium outcome.

The cardinality of equilibrium and the difficulty of equilibrium computation also differ substantially across these two classes of models. In complete information games, the number of equilibria is generically finite and depends only on the number of players and the number of actions they can take. This implies in particular a universal upper bound on the number of equilibria that is invariant with respect to the observed covariates. In addition, the system of equalities and inequalities for determining the equilibria usually involves polynomials. A variety of algorithms developed by McKelvey and McLennan for computing both pure and mixed strategy nash equilibria are discussed in Bajari, Hong and Ryan (2009). In incomplete information games, while the number of equilibria is finite, it can depend on the value of the covariates. The system of equations for determining the equilibrium is usually transcendental, making the use of polynomial-based all-solution homotopy methods more difficult to implement.

5 Exclusion Restriction

The discussion in the previous section suggests that complete information games will require stronger conditions for identification. Even in private information games, the estimability of parameters depends on the relative magnitudes of K and J. In this section, we explore the role of exclusion restrictions in this regard. Without such restrictions, the researcher would have to adopt a partial identification approach. See Aradillas-Lopez and Tamer (2008) and Beresteanu, Molchanov, and Molinari (2008), e.g., for recent applications of this approach.

Our point of departure is that the difficulty of identification disappears as long as researchers are willing to make some exclusion restriction on $\lambda_0(x)$. To gain intuition, consider the private information game and suppose that $\lambda_0(x) = \lambda_0$, i.e., the equilibrium selection does not depend on x. The moment restriction then becomes

$$E\left[\left(y - p^{(J+1)}(x,\theta_0)\right) - Q(x,\theta_0)\lambda_0 \middle| x\right] = 0$$

which is just a parametric GMM problem. See, e.g., Sweetings (2005).

There is a long and distinguished literature in economic theory on equilibrium selection in games. This literature may provide some guidance in imposing restrictions on λ_0 to identify the model. For example, Bresnahan and Reiss (1991) and Tamer (2003) rule out mixed strategy equilibria in entry games. Berry (1992) focuses on the equilibrium that maximizes payoffs for the incumbent firm. Ackerberg and Gowrisankaran (2006) focus on extremal equilibria in supermodular games. Bajari, Hong and Ryan (2009) estimate equilibrium selection in an auction entry example. They allow the selection mechanism to depend on whether the equilibrium is in pure strategies, is Pareto dominated, maximizes industry profits and whether the equilibrium is risk dominant. Classifying equilibria in this fashion can generate restrictions on λ_0 which can allow for identification. As our analysis makes clear, it is not possible to identify complete information games if λ_0 is a completely general function of both covariates x and the random preference shocks ε.

Similar intuition suggests a matching method to gain information when $x = (u, v)$, and $\lambda_0(x)$ only depends on v. We have

$$E\left[\left(y - p^{(J+1)}(x,\theta_0)\right) - Q(x,\theta_0)\lambda_0(v) \middle| x\right] = 0$$

Now randomly choose two observations (y_t, u_t, v_t) and (y_s, u_s, v_s) such that $v_t = v_s$. If we stack the two observations together, we will obtain the moment restriction

$$E\left[\begin{pmatrix} y_t - p^{(J+1)}(u_t, v_t, \theta_0) \\ y_s - p^{(J+1)}(u_s, v_s, \theta_0) \end{pmatrix} - \widetilde{Q}(u_t, u_s, v_t, v_s, \theta_0)\lambda_0(v_t) \middle| v_t = v_s, u_t, u_s\right] = 0$$

where

$$\widetilde{Q}(u_t, u_s, v_t, v_s, \theta_0) = \begin{pmatrix} Q(u_t, v_t, \theta_0) \\ Q(u_s, v_s, \theta_0) \end{pmatrix}$$

Note now that the $\widetilde{Q}(u_t, u_s, v_t, v_s, \theta_0)$ plays the same mathematical role as $Q(x,\theta)$ in (12). Although $\widetilde{Q}(u_t, u_s, v_t, v_s, \theta_0)$ has the same number J of columns as $Q(x,\theta)$, it has twice as many rows $2K$. This implies that the condition for identification is more likely to be satisfied.

As for the complete information game, a convenient exclusion restriction would take the form $\lambda_0(x,\varepsilon) = \lambda_0(x)$, i.e., the equilibrium selection only depends on observed characteristics. If so, we can write the model as

$$E\left[y - \left(\int P(x,\varepsilon,\theta_0) F_\varepsilon(d\varepsilon) \right) \lambda_0(x) \middle| x \right] = 0.$$

Writing $P(x,\theta_0) = \int P(x,\varepsilon,\theta_0) F_\varepsilon(d\varepsilon)$, we can easily see that the mathematical structure is now identical to the private information game. The difficulty in the complete information game was that the $\mathbb{Q}(x,\theta)$ had too many columns to accommodate all the support points of ε. The exclusion restriction here effectively eliminates the problem.

6 Summary

We considered two types of mixed discrete response models. The models were motivated by game theoretic models. The multiplicity of equilibria naturally leads to the mixture characterization. In the first model, which includes the incomplete information game as a special case, we argued that the number of equilibria relative to the number of possible outcome plays an important role in parameter estimation. We derived a moment equation that may be helpful to estimate the parameter. The moment equation can be exploited by the usual GMM estimation. In the second model, which includes the complete information game as a special case, we argued that the information for the structural parameter is zero in general. This suggests that, for complete information games, some restriction on the equilibrium selection probabilities seems to be unavoidable.

Patrick Bajari (Minnesota & NBER, USA), Jinyong Hahn (UCLA, USA), Han Hong (Stanford, USA) and Geert Ridder (USC, USA)

A Proof of Theorem 1

Let $y_k = 1\left(y = \kappa_k\right)$ denote the binary variable indicating whether $y = \kappa_k$ or not. We note that

$$\begin{aligned}
\Pr\left[y_k = 1 \middle| x\right] &= \sum_{j=1}^{J} p_k^{(j)}\left(x,\theta\right)\lambda^{(j)}\left(x\right) + p_k^{(J+1)}\left(x,\theta\right)\left(1 - \sum_{j=1}^{J}\lambda^{(j)}\left(x\right)\right) \\
&= \sum_{j=1}^{J}\left(p_k^{(j)}\left(x,\theta\right) - p_k^{(J+1)}\left(x,\theta\right)\right)\lambda^{(j)}\left(x\right) + p_k^{(J+1)}\left(x,\theta\right)
\end{aligned}$$

for $k = 1,\ldots,K$ and

$$\begin{aligned}
\Pr\left[y_{K+1} = 1 \middle| x\right] &= \sum_{j=1}^{J} p_{K+1}^{(j)}\left(x,\theta\right)\lambda^{(j)}\left(x\right) + p_{K+1}^{(J+1)}\left(x,\theta\right)\left(1 - \sum_{j=1}^{J}\lambda^{(j)}\left(x\right)\right) \\
&= \sum_{j=1}^{J}\left(p_{K+1}^{(j)}\left(x,\theta\right) - p_{K+1}^{(J+1)}\left(x,\theta\right)\right)\lambda^{(j)}\left(x\right) + p_{K+1}^{(J+1)}\left(x,\theta\right) \\
&= \sum_{j=1}^{J}\left(\left(1 - \sum_{k=1}^{K} p_k^{(j)}\left(x,\theta\right)\right) - \left(1 - \sum_{k=1}^{K} p_k^{(J+1)}\left(x,\theta\right)\right)\right)\lambda^{(j)}\left(x\right) \\
&\quad + \left(1 - \sum_{k=1}^{K} p_k^{(J+1)}\left(x,\theta\right)\right) \\
&= \sum_{j=1}^{J}\sum_{k=1}^{K}\left(p_k^{(J+1)}\left(x,\theta\right) - p_k^{(j)}\left(x,\theta\right)\right)\lambda^{(j)}\left(x\right) + \left(1 - \sum_{k=1}^{K} p_k^{(J+1)}\left(x,\theta\right)\right)
\end{aligned}$$

Our log-likelihood is given by

$$\begin{aligned}
\ell = &\sum_{k=1}^{K} y_k \log\left[\sum_{j=1}^{J}\left(p_k^{(j)}\left(x,\theta\right) - p_k^{(J+1)}\left(x,\theta\right)\right)\lambda^{(j)}\left(x\right) + p_k^{(J+1)}\left(x,\theta\right)\right] \\
&+ y_{K+1}\log\left[\sum_{j=1}^{J}\sum_{k=1}^{K}\left(p_k^{(J+1)}\left(x,\theta\right) - p_k^{(j)}\left(x,\theta\right)\right)\lambda^{(j)}\left(x\right) + \left(1 - \sum_{k=1}^{K} p_k^{(J+1)}\left(x,\theta\right)\right)\right]
\end{aligned}$$

Consider one-dimensional parametrization of $\lambda^{(j)}\left(x\right)$, i.e., $\lambda^{(j)}\left(x;\eta\right)$. Let

$$\begin{aligned}
s_\theta = &\frac{\partial \ell}{\partial \theta} \\
= &\sum_{k=1}^{K}\frac{y_k}{p_k\left(x;\theta,\lambda\right)}\left[\sum_{j=1}^{J}\left(\frac{\partial p_k^{(j)}\left(x,\theta\right)}{\partial\theta} - \frac{\partial p_k^{(J+1)}\left(x,\theta\right)}{\partial\theta}\right)\lambda^{(j)}\left(x\right) + \frac{\partial p_k^{(J+1)}\left(x,\theta\right)}{\partial\theta}\right] \\
&+ \frac{y_{K+1}}{p_{K+1}\left(x;\theta,\lambda\right)}\left[\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\frac{\partial p_k^{(J+1)}\left(x,\theta\right)}{\partial\theta} - \frac{\partial p_k^{(j)}\left(x,\theta\right)}{\partial\theta}\right)\lambda^{(j)}\left(x\right) - \sum_{k=1}^{K}\frac{\partial p_k^{(J+1)}\left(x,\theta\right)}{\partial\theta}\right]
\end{aligned}$$

and

$$
\begin{aligned}
s_\eta =& \frac{\partial \ell}{\partial \eta} \\
=& \sum_{k=1}^{K} \frac{y_k}{p_k(x;\theta,\lambda)} \left[\sum_{j=1}^{J} \left(p_k^{(j)}(x,\theta) - p_k^{(J+1)}(x,\theta) \right) \frac{\partial \lambda^{(j)}(x;\eta)}{\partial \eta} \right] \\
&+ \frac{y_{K+1}}{p_{K+1}(x;\theta,\lambda)} \left[\sum_{j=1}^{J} \sum_{k=1}^{K} \left(p_k^{(J+1)}(x,\theta) - p_k^{(j)}(x,\theta) \right) \frac{\partial \lambda^{(j)}(x;\eta)}{\partial \eta} \right]
\end{aligned}
$$

with

$$
p_k(x;\theta,\lambda) = \sum_{j=1}^{J} \left(p_k^{(j)}(x,\theta) - p_k^{(J+1)}(x,\theta) \right) \lambda^{(j)}(x) + p_k^{(J+1)}(x,\theta)
$$

$$
p_{K+1}(x;\theta,\lambda) = \sum_{j=1}^{J} \sum_{k=1}^{K} \left(p_k^{(J+1)}(x,\theta) - p_k^{(j)}(x,\theta) \right) \lambda^{(j)}(x) + \left(1 - \sum_{k=1}^{K} p_k^{(J+1)}(x,\theta) \right)
$$

Note that we have the simplification

$$
\begin{aligned}
s_\theta =& \sum_{k=1}^{K} \sum_{j=1}^{J} \left(\frac{y_k}{p_k(x;\theta,\lambda)} - \frac{y_{K+1}}{p_{K+1}(x;\theta,\lambda)} \right) \left(\frac{\partial p_k^{(j)}(x,\theta)}{\partial \theta} - \frac{\partial p_k^{(J+1)}(x,\theta)}{\partial \theta} \right) \lambda^{(j)}(x) \\
&+ \sum_{k=1}^{K} \left(\frac{y_k}{p_k(x;\theta,\lambda)} - \frac{y_{K+1}}{p_{K+1}(x;\theta,\lambda)} \right) \frac{\partial p_k^{(J+1)}(x,\theta)}{\partial \theta} \\
=& \sum_{k=1}^{K} \left(\frac{y_k}{p_k(x;\theta,\lambda)} - \frac{y_{K+1}}{p_{K+1}(x;\theta,\lambda)} \right) \left(\sum_{j=1}^{J} \left(\frac{\partial p_k^{(j)}(x,\theta)}{\partial \theta} - \frac{\partial p_k^{(J+1)}(x,\theta)}{\partial \theta} \right) \lambda^{(j)}(x) + \frac{\partial p_k^{(J+1)}(x,\theta)}{\partial \theta} \right)
\end{aligned}
$$

and

$$
s_\eta = \sum_{k=1}^{K} \sum_{j=1}^{J} \left(\frac{y_k}{p_k(x;\theta,\lambda)} - \frac{y_{K+1}}{p_{K+1}(x;\theta,\lambda)} \right) \left(p_k^{(j)}(x,\theta) - p_k^{(J+1)}(x,\theta) \right) \frac{\partial \lambda^{(j)}(x;\eta)}{\partial \eta}
$$

Now consider the tangent set $\mathcal{T}$, which consists of

$$
\sum_{k=1}^{K} \left(\frac{y_k}{p_k(x;\theta,\lambda)} - \frac{y_{K+1}}{p_{K+1}(x;\theta,\lambda)} \right) \sum_{j=1}^{J} \left(p_k^{(j)}(x,\theta) - p_k^{(J+1)}(x,\theta) \right) t^{(j)}(x)
$$

We want to show that the projection of s_θ on $\mathcal{T}$ is in fact equal to s_θ when J is sufficiently large. For this purpose, it suffices to show that there exists $t^{(1)}(x), \ldots, t^{(J)}(x)$ such that

$$
\sum_{j=1}^{J} \left(\frac{\partial p_k^{(j)}(x,\theta)}{\partial \theta} - \frac{\partial p_k^{(J+1)}(x,\theta)}{\partial \theta} \right) \lambda^{(j)}(x) + \frac{\partial p_k^{(J+1)}(x,\theta)}{\partial \theta} = \sum_{j=1}^{J} \left(p_k^{(j)}(x,\theta) - p_k^{(J+1)}(x,\theta) \right) t^{(j)}(x)
$$

or

$$\sum_{j=1}^{J} \frac{\partial p_k^{(j)}(x,\theta)}{\partial\theta} \lambda^{(j)}(x) + \frac{\partial p_k^{(J+1)}(x,\theta)}{\partial\theta} \left(1 - \sum_{j=1}^{J} \lambda^{(j)}(x)\right) = \sum_{j=1}^{J} \left(p_k^{(j)}(x,\theta) - p_k^{(J+1)}(x,\theta)\right) t^{(j)}(x)$$

for all $k = 1, \ldots, K$. This can be verified by taking $t^{(j)}(x) = 0$ for all $j > K$, and solving for $t^{(1)}(x), \ldots, t^{(K)}(x)$ from

$$\begin{aligned} &\sum_{j=1}^{K} \left(p_k^{(j)}(x,\theta) - p_k^{(J+1)}(x,\theta)\right) t^{(j)}(x) \\ &\quad = \sum_{j=1}^{J} \left(\frac{\partial p_k^{(j)}(x,\theta)}{\partial\theta} - \frac{\partial p_k^{(J+1)}(x,\theta)}{\partial\theta}\right) \lambda^{(j)}(x) + \frac{\partial p_k^{(J+1)}(x,\theta)}{\partial\theta}, \quad k = 1, \ldots, K \end{aligned}$$

B Complete Information Game: An Exceptional Example

Consider the complete information case in a 2×2 game, which means that there are four possible outcomes. Suppose that ε is not observed. Suppose that the outcome is such that

$$\begin{aligned} y = \kappa_1 \quad &\Leftrightarrow \quad \varepsilon \in \mathbb{A}(x,\theta) \\ y = \kappa_2 \text{ or } \kappa_3 \quad &\Leftrightarrow \quad \varepsilon \in \mathbb{B}(x,\theta) \\ y = \kappa_4 \quad &\Leftrightarrow \quad \varepsilon \in \mathbb{C}(x,\theta) \end{aligned}$$

where $\mathbb{A}(x,\theta)$, $\mathbb{B}(x,\theta)$, and $\mathbb{C}(x,\theta)$ are subsets of the support of ε that depends on (x,θ). Because of redundancy, we will ignore the $y = \kappa_4$ case.

Using our earlier notation, this can be written as

$$p^{(1)}(x,\varepsilon,\theta_0) = \begin{bmatrix} 1(\varepsilon \in \mathbb{A}(x,\theta)) \\ 1(\varepsilon \in \mathbb{B}(x,\theta)) \\ 0 \end{bmatrix}, \quad p^{(2)}(x,\varepsilon,\theta_0) = \begin{bmatrix} 1(\varepsilon \in \mathbb{A}(x,\theta)) \\ 0 \\ 1(\varepsilon \in \mathbb{B}(x,\theta)) \end{bmatrix}$$

and the usable moment restriction is

$$E\left[y - p^{(2)}(x,\varepsilon,\theta_0) - \left(p^{(1)}(x,\varepsilon,\theta_0) - p^{(1)}(x,\varepsilon,\theta_0)\right) \lambda(x,\varepsilon) \middle| x \right] = 0$$

or

$$E\left[y - \left(\int p^{(2)}(x,\varepsilon,\theta_0) f_\varepsilon(\varepsilon)\, \mathrm{d}\varepsilon\right) - \left(\int \left(p^{(1)}(x,\varepsilon,\theta_0) - p^{(2)}(x,\varepsilon,\theta_0)\right) \lambda(x,\varepsilon) f_\varepsilon(\varepsilon)\, \mathrm{d}\varepsilon\right) \middle| x \right] = 0$$

It turns out that we can without loss of generality assume that $\lambda(x, \varepsilon)$ does not depend on ε in this special case. To be more precise, the model is mathematically equivalent to the model where the equilibrium selection probability $\lambda(x, \varepsilon)$ does not depend on ε in this special case.

Let $p_k^{(1)}(x, \varepsilon, \theta_0)$ and $p_k^{(2)}(x, \varepsilon, \theta_0)$ denote the k-th components of $p^{(1)}(x, \varepsilon, \theta_0)$ and $p^{(2)}(x, \varepsilon, \theta_0)$. It can be shown that

$$\int \left(p_1^{(1)}(x, \varepsilon, \theta_0) - p_1^{(2)}(x, \varepsilon, \theta_0)\right) \lambda(x, \varepsilon) f_\varepsilon(\varepsilon) \mathrm{d}\varepsilon = 0$$
$$\int \left(p_2^{(1)}(x, \varepsilon, \theta_0) - p_2^{(2)}(x, \varepsilon, \theta_0)\right) \lambda(x, \varepsilon) f_\varepsilon(\varepsilon) \mathrm{d}\varepsilon = \int 1(\varepsilon \in \mathbb{B}(x, \theta)) \lambda(x, \varepsilon) f_\varepsilon(\varepsilon) \mathrm{d}\varepsilon$$
$$\int \left(p_3^{(1)}(x, \varepsilon, \theta_0) - p_3^{(2)}(x, \varepsilon, \theta_0)\right) \lambda(x, \varepsilon) f_\varepsilon(\varepsilon) \mathrm{d}\varepsilon = -\int 1(\varepsilon \in \mathbb{B}(x, \theta)) \lambda(x, \varepsilon) f_\varepsilon(\varepsilon) \mathrm{d}\varepsilon$$

Let

$$\mathbf{p}^{(1)}(x, \theta_0) \equiv \left(\int p^{(1)}(x, \varepsilon, \theta_0) f_\varepsilon(\varepsilon) \mathrm{d}\varepsilon\right) \equiv \begin{bmatrix} A(x, \theta_0) \\ B(x, \theta_0) \\ 0 \end{bmatrix},$$

$$\mathbf{p}^{(2)}(x, \theta_0) \equiv \left(\int p^{(2)}(x, \varepsilon, \theta_0) f_\varepsilon(\varepsilon) \mathrm{d}\varepsilon\right) = \begin{bmatrix} A(x, \theta_0) \\ 0 \\ B(x, \theta_0) \end{bmatrix},$$

and

$$\zeta(x) \equiv \frac{\int 1(\varepsilon \in \mathbb{B}(x, \theta)) \lambda(x, \varepsilon) f_\varepsilon(\varepsilon) \mathrm{d}\varepsilon}{\mathbf{p}_2^{(1)}(x, \theta_0) - \mathbf{p}_2^{(2)}(x, \theta_0)}.$$

Note that

$$\zeta(x) = -\frac{\int 1(\varepsilon \in \mathbb{B}(x, \theta)) \lambda(x, \varepsilon) f_\varepsilon(\varepsilon) \mathrm{d}\varepsilon}{\mathbf{p}_3^{(1)}(x, \theta_0) - \mathbf{p}_3^{(2)}(x, \theta_0)}.$$

Because $\lambda(x, \varepsilon)$ is nonparametrically specified, we can without loss of generality ignore the dependence of ζ on θ_0. We then automatically have

$$\int \left(p^{(1)}(x, \varepsilon, \theta_0) - p^{(2)}(x, \varepsilon, \theta_0)\right) \lambda(x, \varepsilon) f_\varepsilon(\varepsilon) \mathrm{d}\varepsilon = \left(\mathbf{p}^{(1)}(x, \theta_0) - \mathbf{p}^{(2)}(x, \theta_0)\right) \zeta(x)$$

or our moment restriction becomes

$$E\left[y - \mathbf{p}^{(2)}(x, \theta_0) - \left(\mathbf{p}^{(1)}(x, \theta_0) - \mathbf{p}^{(2)}(x, \theta_0)\right) \zeta(x) \middle| x\right] = 0$$

We conclude that this particular complete information game is mathematically equivalent to the model where the equilibrium selection probability $\lambda(x, \varepsilon)$ does not depend on ε. (The source of the equivalence is the existence of $\zeta(x)$ such that

$$\int \left(p_k^{(1)}(x, \varepsilon, \theta_0) - p_k^{(2)}(x, \varepsilon, \theta_0)\right) \lambda(x, \varepsilon) f_\varepsilon(\varepsilon) \mathrm{d}\varepsilon = \left(\mathbf{p}_k^{(1)}(x, \theta_0) - \mathbf{p}_k^{(2)}(x, \theta_0)\right) \zeta(x) \qquad \forall k$$

The equivalence does not seem to hold in more general "complete information" games.)

References

[1] Ackerberg, D., "Empirically distinguishing informative and prestige effects of advertising," mimeo, Boston University, 1996.

[2] Ackerberg, D. and G. Gowrisankaran, "Quantifying equilibrium network externalities in the ACH banking industry," RAND Journal of Economics 37 (2006), 738-761.

[3] Aguirregabiria, V. and P. Mira, "Sequential Estimation of Dynamic Discrete Games," Econometrica 75 (2007), 1-53.

[4] Aradillas-Lopez, A., "Semiparametric estimation of a simultaneous game with incomplete information," 2005, forthcoming, Journal of Econometrics.

[5] Aradillas-Lopez, A. and E. Tamer, "The identification power of equilibrium in simple games," Journal of Business and Economic Statistics 26 (2008), 261-283.

[6] Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2006), "Estimating static models of strategic interactions," forthcoming in Journal of Business and Economic Statistics

[7] Bajari, P., H. Hong, and S. Ryan (2009), "Identification and estimation of discrete games of complete information," forthcoming in Econometrica.

[8] Beresteanu, A., I. Molchanov, and F. Molinari (2008). Sharp identification regions in games. forthcoming in Econometrica.

[9] Berry, S., "Estimation of a model of entry in the airline industry," Econometrica 60 (1992), 889-917.

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

[11] Bresnahan, T. and P. Reiss, "Entry in monopoly markets," Review of Economic Studies 57 (1990), 531-553.

[12] Bresnahan, T. and P. Reiss, "Empirical models of discrete games," Journal of Econometrics 48 (1991), 57-81.

[13] Brock, W. and S. Durlauf, "Discrete choice with social interactions," Review of Economic Studies 68 (2001), 235-260.

[14] Chamberlain, G., "Asymptotic efficiency in semiparametric models with censoring," Journal of Econometrics 32 (1986), 189-218.

[15] Chamberlain, G., "Efficiency Bounds for Semiparametric Regression," Econometrica 60 (1992), 567-596.

[16] Ciliberto, F. and E. Tamer (2007), "Market Structure and Multiple Equilibria in Airline Markets," forthcoming in Econometrica.

[17] Hahn, J., "The efficiency bound of the mixed proportional hazard model," Review of Economic Studies 61 (1994), 607-629.

[18] Heckman, J., "Varieties of selection bias," American Economic Review 80 (1990), 313-318.

[19] Hu, Y. and M. Shum (2009), "Identifying Dynamic Games with Serially-Correlated Unobservables, ," working paper, John Hopkins University.

[20] Kasahara, H. and K. Shimotsu, "Nonparametric identification of finite mixture models of dynamic discrete choices, ," Econometrica 77 (2008), 135-176.

[21] Maddala, G.S., Limited-dependent and qualitative variables in econometrics (Cambridge: Cambridge University Press, 1983)

[22] Manski, C., "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator," Journal of Econometrics 27 (1985), 313-333.

[23] Pakes, A., J. Porter, K. Ho, and J. Ishii, "Moment inequalities and their application," mimeo, Harvard University, 2005.

[24] Pesendorfer, M. and P. Schmidt-Dengler, "Identification and estimation of dynamic games," NBER working paper No. w9726.

[25] Ridder, G., "The non-parametric identification of generalized accelerated failure-time models," Review of Economic Studies 57 (1990), 167-181.

[26] Seim, K., "An empirical model of firm entry with endogenous product-type choices," RAND Journal of Economics 37 (2006), 619-640.

[27] Sweetings, A., "Coordination games, multiple equilibria and the timing of radio commercials," mimeo, Northwestern University, 2005.

[28] Tamer, E., "Incomplete simultaneous discrete response model with multiple equilibria," Review of Economic Studies 70 (2003), 147-167.