Estimating Static Models of Strategic Interactions

Patrick Bajari, Han Hong, John Krainer and Denis Nekipelov [1]

University of Michigan and NBER

Duke University

Federal Reserve Bank of San Francisco

Duke University

September 13, 2005

Abstract

We propose a method for estimating static games of incomplete information. A static game is a generalization of a discrete choice model, such as a multinomial logit or probit, which allows the actions of a group of agents to be interdependent. Unlike most earlier work, the method we propose is semiparametric and does not require the covariates to lie in a discrete set. While the estimator we propose is quite flexible, we demonstrate that in many cases it can be implemented using a simple two-stage least squares procedure in a standard statistical package. We also propose an algorithm for simulating the model which finds all equilibria to the game. As an application of our estimator, we study recommendations for high technology stocks between 1998-2003. We find that strategic motives, typically ignored in the empirical literature, appear to be an important consideration in the recommendations submitted by equity analysts.

1 Introduction

Game theory is one of the most commonly applied tools in economic theory, with substantive applications in all major fields in economics. In some fields, particularly industrial organization, game theory has not only transformed the analysis of market interactions, but also serves as an important basis for policy recommendations. Given the importance of gaming in economic theory, it is not surprising that the empirical analysis of games has been the focus of a recent literature in econometrics and industrial organization.

In much of the literature, a discrete game is modeled much like a standard discrete choice problem, such as the multinomial logit. An agent's utility is often assumed to be a linear function of covariates and a random preference shock. However, unlike a discrete choice model, utility is also allowed to depend on the actions of other agents. A discrete game strictly generalizes a standard random utility model, but does not impose the often strong assumption that agents act in isolation. Early attempts at the econometric analysis of such games included Vuong and Bjorn (1984) and Bresnahan and Reiss (1990,1991). Other recent examples include Aradillas-Lopez (2005), Ho (2005), Ishii (2005), Pakes, Porter, Ho and Ishii (2005), Rysman (2004), Seim (2004), Sweeting (2004) and Tamer (2003).

An important insight in the recent literature is that it is often most straightforward to estimate discrete games in two steps. For examples see Aguirregabiria and Mira (2003), Bajari, Benkard and Levin (2004), Pakes, Ostrovsky and Berry (2004) and Pesendorfer and Schmidt-Dengler (2004). In a first step, the economist estimates the reduced forms implied by the model. This often boils down to using standard econometric methods to estimate the probability that one, out of a finite number of possible choices, is observed conditional on the relevant covariates. In the second step, the economist estimates a single agent random utility model, including as controls the equilibrium beliefs about the behavior of others from the first step.

In this paper, we propose an estimator that can be applied to static games of strategic interaction. Like the two-step approach discussed above, we estimate the reduced form choice probabilities in a first stage in order to simplify the estimation of the model. The approach that we propose, however, differs from earlier work in four ways. First, much of the earlier literature on two step estimation considered fully dynamic games. This made

it difficult for researchers unfamiliar with dynamic programming to understand how to estimate a game. Also, we note that to date, almost all empirical applications of discrete games have been static. Therefore, proposing methods for static games is of practical importance for many researchers.

Second, the approach that we propose can be done nonparametrically or semiparametrically. Much of the earlier literature on two-step estimation of games considered the case where the set of regressors was discrete or the first stage was a correctly specified parametric model. In this paper, we allow for continuous covariates and a fully nonparametric first stage. We establish two useful properties of this estimator. First, as in Bajari and Hong (2005), despite the fact the first stage is nonparametric and might converge at a slow rate, the structural parameters estimated in the second stage have normal asymptotics and converge at a rate proportional to the square root of the sample size. This follows from arguments based on Newey (1994). Second, we demonstrate in many cases our model can be estimated, with correct standard errors, using a two stage least squares procedure in a standard statistical package like STATA. We hope that the simplicity of this approach will make the estimation of these models accessible to a larger audience of researchers.

Third, we consider the problem of identification in games with continuous state variables. Our results, as in Bajari and Hong (2005), show that a sufficient condition for identification is to exclude payoff relevant covariates for a particular player i from the utilities of the other players. For instance, in an entry model, if the productivity shock of firm i influences its own entry decision, but only indirectly influences the entry decisions of other players, then our results imply that the model is identified. An alternative identification strategy is to search for events that change which equilibrium to the game is played, but otherwise do not influence payoffs. Our results can be interpreted as standard rank conditions for an appropriately defined linear system. We note that Pesendorfer and Schmidt-Dengler (2004) demonstrate that exclusion restrictions are sufficient for identification in a particular set of entry games with discrete states. Sweeting (2004) demonstrates that multiplicity of equilibrium can assist with identification in a symmetric location game.

Finally, we consider the problem of simulating the model, which is required to study predictions of the model such as counterfactuals. It is widely known that models of the

3

form that we consider can generate multiple solutions. However, outside of certain specific examples (e.g., those studied in Sweeting (2004)), it is not possible to analytically derive all of the solutions of the model or even determine the number of possible solutions. Therefore, we propose an algorithm that can compute all of the equilibrium to the model. This algorithm uses the "all solutions homotopy", which is available in standard numerical libraries such as hompack. Therefore, we can use this to find the entire set of equilibrium actions at our estimated parameter values. We discuss the potential uses of this algorithm in our application.

As an application of our methods, we model the determination of stock recommendations (e.g. strong buy, buy, hold, sell) issued by equity analysts for high technology stocks listed in the NASDAQ 100 between 1998 and 2003. The determination of recommendations during this time period is of particular interest in the wake of the sharp stock price declines for technology firms in 2000. Not only did recommended stocks vastly underperform the market as a whole during this period, but highly-publicized allegations of conflicts of interest have called into question whether analysts were more concerned with helping their firms win investment banking business than with producing accurate assessments of the prospects for the firms under scrutiny. While there is a fairly large literature in finance on recommendations, we are not aware of any papers that formally consider the simultaneity of recommendations due to strategic motives.

In our model, recommendations submitted by analysts depend on four factors. First, recommendations must depend on fundamentals and commonly shared expectations about the future profitability of the firm. These expectations will be embedded in the stock price. Second, analysts are heterogeneous, both in terms of talent and perhaps in terms of access to information. We try to capture an individual analyst's private belief about the stock by looking at the difference between the quarterly earnings forecast submitted by the analyst (or the analyst's brokerage firm) and the distribution of forecasts from other firms. Mindful of the large number of inquiries into possible conflicts of interest among research analysts, we include as a third factor a dummy variable for an investment banking relationship between the firm and the analyst's employer.

Finally, we consider the influence of peers on the recommendation decision. Peer effects

4

can impact the recommendation in different ways. Individual analysts have incentive to condition their recommendation on the recommendations of their peers, because even if their recommendations turn out to be unprofitable ex-post, performance evaluation is typically a comparison against the performance of peers. More subtly, recommendations are relative rankings of firms and are not easily quantifiable (or verifiable) objects. As such, ratings scales usually reflect conventions and norms. The phenomenon is similar to the college professor's problem of assigning grades. If a professor were to award the average student with a C while other faculty give a B+ to the average student, the professor might incorrectly signal his views of student performance. Even while there is heterogeneity in how individual professors feel about grading, most conform to norms if only to communicate clearly with students (and their potential employers) about their performance. Similarly, analysts might have an incentive to benchmark their recommendations against perceived industry norms.

The paper is organized as follows. In section 2 we outline the general economic environment. For purposes of exposition, we develop many of the key formulae within the context of a simple entry model. In section 3 we discuss the problem of nonparametric identification. In section 4 we show how to derive nonparametric and semiparametric estimates of the structural parameters for our class of models. Section 5 describes the all solutions homotopy algorithm for simulating the model. Section 6 contains the empirical application to equity analyst recommendations. Section 7 concludes the paper.

2   The model

In the model, there are a finite number of players, $i = 1, ..., n$ and each player simultaneously chooses an action $a_i \in \{0, 1, \dots, K\}$ out of a finite set. We restrict players to have the same set of actions for notational simplicity. However, all of our results will generalize to the case where all players have different finite sets of actions. Let $A = \{0, 1, \dots, K\}^n$ denote the vector of possible actions for all players and let $a = (a_1, ..., a_n)$ denote a generic element of A. As is common in the literature, we shall let $a_{-i} = (a_1, ...a_{i-1}, a_{i+1}, ..., a_n)$ denote a vector of strategies for all players excluding player i. We will abstract from mixed strategies since in our model, with probability one each player will have a unique best response.

Let $s_i \in S_i$ denote the state variable for player i. Let $S = \Pi_i S_i$ and let $s = (s_1, ..., s_n) \in$

S denote a vector of state variables for all n players. We will assume that s is common knowledge to all players in the game and in our econometric analysis, we will assume that s is observable to the econometrician. The state variable is assumed to be a real valued vector, but S_i is not required to be a finite set. Much of the previous literature assumes that the state variables in a discrete games lie in a discrete set. While this assumption simplifies the econometric analaysis of the estimator and identification, it is a strong assumption that may not be satisfied in many applications.

For each agent, there are also $K + 1$ state variables which we label as $\epsilon_i(a_i)$ which are private information to each agent. These state variables are distributed i.i.d. across agents and actions. Let ϵ_i denote the $1 \times (K + 1)$ vector of the individual $\epsilon_i(a_i)$. The density of $\epsilon_i(a_i)$ will be denoted as $f(\epsilon_i(a_i))$, however, we shall sometimes abuse notation and denote the density for $\varepsilon_i = (\varepsilon_i(0), ..., \varepsilon_i(K))$ as $f(\epsilon_i)$.

The period utility function for player i is:

$$u_i(a, s, \epsilon_i; \theta) = \Pi_i(a_i, a_{-i}, s; \theta) + \epsilon_i(a_i). \tag{1}$$

The utility function in our model is similar to a standard random utility model such as a multinomial logit. Each player i receives a stochastic preference shock, $\epsilon_i(a_i)$, for each possible action a_i. In many applications, this will be drawn from an extreme value distribution as in the logit model. In the literature, the preference shock is alternatively interpreted as an unobserved state variable (see Rust (1994)). Utility also depends on the vector of state variables s and actions a through $\Pi_i(a_i, a_{-i}, s; \theta)$. For example, in the literature, this part of utility is frequently parameterized as a simple linear function of actions and states. Unlike a standard discrete choice model, however, note that the actions a_{-i} of other players in the game enter into i's utility. A standard discrete choice model typically assumes that agents i act in isolation in the sense that a_{-i} is omitted from the utility function. In many applications, this is an implausible assumption.

In this model, player i's decision rule is a function $a_i = \delta_i(s, \epsilon_i)$. Note that i's decision does not depend on the ϵ_{-i} since these shocks are private information to the other $-i$ players in the game and hence unobservable to i. Define $\sigma_i(a_i|s)$ as:

$$\sigma_i(a_i = k|s) = \int 1\{\delta_i(s, \epsilon_i) = k\} f(\epsilon_i) d\epsilon_i. \tag{2}$$

6

In the above expression, $1\{\delta_i(s,\epsilon_i)=k\}$ is the indicator function that player i's action is k given the vector of state variable (s,ϵ_i). Therefore, $\sigma_i(a_i=k|s)$ is the probability that i chooses action k conditional on the state variables s that are public information. We will define the distribution of a given s as $\sigma(a|s) = \Pi_{i=1}^n \sigma(a_i|s)$.

Next define $\pi_i(a_i,s,\epsilon_i;\theta)$ as:

$$\pi_i(a_i,s,\epsilon_i;\theta) = \sum_{a_{-i}} \Pi_i(a_i,a_{-i},s;\theta)\sigma_{-i}(a_{-i}|s) + \epsilon_i(a_i) \qquad (3)$$

$$\text{where } \sigma_{-i}(a_{-i}|s) = \Pi_{j\neq i}\sigma_j(a_j|s). \qquad (4)$$

In (3), $\pi_i(a_i,s,\epsilon_i;\theta)$ is player i's expected utility from choosing a_i when the vector of parameters is θ. Since i does not know the private information shocks, ϵ_j for the other players, i's beliefs about their actions are given by $\sigma_{-i}(a_{-i}|s)$. The term $\sum_{a_{-i}} \Pi_i(a_i,a_{-i},s,\theta)\sigma_{-i}(a_{-i}|s)$ is the expected value of $\Pi_i(a_i,a_{-i},s;\theta)$, marginalizing out the strategies of the other players using $\sigma_{-i}(a_{-i}|s)$. The structure of payoffs in (3) is quite similar to standard random utility models, except that the probability distribution over other agent's actions enter into the formula for agent i's utility. Note that if the error term has an atomless distribution, then player i's optimal action is unique with probability one. This is an extremely convenient property and eliminates the need to consider mixed strategies as in a standard normal form game.

We also define the deterministic part of the expected payoff as

$$\Pi_i(a_i,s;\theta) = \sum_{a_{-i}} \Pi_i(a_i,a_{-i},s,\theta)\sigma_{-i}(a_{-i}|s). \qquad (5)$$

It follows immediately then that the optimal action for player i satisfies:

$$\sigma_i(a_i|s) = Prob\{\epsilon_i|\Pi_i(a_i,s;\theta)+\epsilon_i(a_i) > \Pi_i(a_j,s;\theta)+\epsilon_i(a_j) \text{ for } j\neq i.\} \qquad (6)$$

2.1 A Simple Example.

For expositional clarity, it is worthwhile to consider a simple example of a discrete game. Perhaps the most commonly studied example of a discrete game in the literature is a static

entry game (see Bresnahan and Reiss (1990,1991), Berry (1992), Tamer (2002), Ciliberto and Tamer (2003), and Manuszak and Cohen (2004)). In the empirical analysis of entry games, the economist typically has data on a cross section of markets and observes whether a particular firm, i chooses to enter a particular market. In Berry (1992) and Ciliberto and Tamer (2003), for example, the firms are major U.S. airlines such as American, United and Northwest and the markets are large, metropolitan airports. The state variables, s_i might include the population in the metropolitan area surrounding the airport and measures of an airline's operating costs. Let $a_i = 1$ denote the decision to enter a particular market and $a_i = 0$ denote the decision not to enter the market. In many applications, $\Pi_i(a_i, a_{-i}, s; \theta)$ is assumed to be a linear function, e.g.:

$$\Pi_i(a_i, a_{-i}, s) = \begin{cases} s' \cdot \beta + \delta \sum_{j \neq i} 1\{a_j = 1\} \text{ if } a_i = 1 \\ 0 \text{ if } a_i = 0 \end{cases} \tag{7}$$

In equation (7), the mean utility from not entering is set equal to zero.[2] The term δ measures the influence of j's choice on i's entry decision. If profits decrease from having another firm enter the market then $\delta < 0$. The parameters β measure the impact of the state variables on $\Pi_i(a_i, a_{-i}, s)$.

The random error terms $\varepsilon_i(a_i)$ are thought to capture shocks to the profitability of entry that are private information to firm i. Suppose that the error terms are distributed exteme value. Then, utility maximization by firm i implies that:

$$\sigma_i(a_i = 1|s) = \frac{\exp(s' \cdot \beta + \delta \sum_{j \neq i} \sigma_j(a_j = 1|s))}{1 + \exp(s' \cdot \beta + \delta \sum_{j \neq i} \sigma_j(a_j = 1|s))} \text{ for } i = 1, ..., n \tag{8}$$

In the system of equations above, applying the formula in equation (5) implies that $\Pi_i(a_i, s; \theta) = s' \cdot \beta + \delta \sum_{j \neq i} \sigma_j(a_j = 1|s)$. Since the error terms are distributed extreme value, equation (6) implies that the choice probabilities, $\sigma_i(a_i = 1|s)$ take a form similar to a single agent

---

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

multinomial logit model. We note in passing that it can easily be shown using Brouwer's fixed point theorem an equilibrium to this model exists for any finite s (see McKelvy and Palfrey (1995)).

We shall exploit the convenient representation of equilibrium in equation (8) in our econometric analysis. Suppose that the econometrician observes $t = 1, ..., T$ repetitions of the game. Let $a_{i,t}$ denote the entry decision of firm i in repetition t and let the value of the state variables be equal to s_t. By observing entry behavior in a large number of markets, the econometrician could form a a consistent estimate $\widehat{\sigma}_i(a_i = 1|s)$ of $\sigma_i(a_i = 1|s)$ for $i = 1, ..., n$. In an application, this simply boils down to flexibly estimating the probability that a binary response, a_i is equal to one conditional on a given set of covariates. This could be done using any one of a number of standard techniques.

We let $L(\beta, \delta)$ denote the psueodo likelihood function defined as:

$$L(\beta, \delta) = \prod_{t=1}^{T}\prod_{i=1}^{n}\left(\frac{\exp(s'\cdot\beta+\delta\sum_{j\neq i}\hat{\sigma}_j(a_j=1|s))}{1+\exp(s'\cdot\beta+\delta\sum_{j\neq i}\hat{\sigma}_j(a_j=1|s))}\right)^{1\{a_{i,t}=1\}}\left(1-\frac{\exp(s'\cdot\beta+\delta\sum_{j\neq i}\hat{\sigma}_j(a_j=1|s))}{1+\exp(s'\cdot\beta+\delta\sum_{j\neq i}\hat{\sigma}_j(a_j=1|s))}\right)^{1\{a_{i,t}=0\}}$$

Given first stage estimates of $\widehat{\sigma}_i(a_i = 1|s)$, we could then estimate the structural parameters of the payoff, β and δ by maximzing the above psuedo-likelihood function. There are two attractive features of this strategy. The first is that it not demanding computationally. First stage estimates of choice probabilities could be done using a strategy as simple as a linear probability model. The computational burden of the second stage is also light since we only need to estimate a logit model. A second attractive feature is that it allows us to view a game as a generalization of a standard discrete choice model. Thus, techniques from the voluminous econometric literature on discrete choice models can be imported into the study of strategic interaction. While the example considered above is simple, it nonetheless illustrates many of the key ideas that will be essential in what follows.

We can also see a key problem with identification in the simple example above. Both the first stage estimates $\widehat{\sigma}_i(a_i = 1|s)$ and the term $s' \cdot \beta$ depend on the vector of state variables s. This suggests that we will suffer from a colinearity problem in order to seperately identify the effects of β and δ on the observed choices. The standard solution to this type of problem in many settings is to impose an exclusion restriction. Suppose, for instance, a firm specific productivity shock is included in s. In most oligopoly models, absent technology spillovers, the productivity shocks of firms $-i$ would not directly enter into firm i's profits. These shocks only enter indirectly through the endogenously determined actions of firms $-i$, e.g. price, quantity or entry decisions. Therefore, if we exclude the productivity shocks of other firms from the term $s' \cdot \beta$, we would no longer suffer

from a colinearity problem. While this idea is quite simple, as we shall discover in the next section, similar restrictions are required to identify more general models.

3 Nonparametric Identification

In this section, we consider the problem of identifying the deterministic part of payoffs, without making particular assumptions about its functional form (e.g. that it is a linear index as in the previous example). In the context of nonparametric identification, we let θ be completely nonparametric and write $\Pi_i(a_i, a_{-i}, s)$ instead of $\Pi_i(a_i, a_{-i}, s; \theta)$.

Definition We will say that $\Pi_i(a_i, a_{-i}, s)$ is identified if $\Pi_i(a_i, a_{-i}, s) \neq \Pi_i'(a_i, a_{-i}, s)$ for some $i = 1, ..., n$ implies that for the corresponding equilibrium choice probabilities $\sigma_i(a_i = 1|s) \neq \sigma_i'(a_i = 1|s)$ for some $i = 1, ..., n$.

Formally, identification requires that different values of the primitives generate different choice probabilites. If this condition is not satisfied, then it will be impossible for us to uniquely recover the structural parameters $\Pi_i(a_i, a_{-i}, s)$ (for $i = 1, ..., n$) from knowledge of the observed choice probabilities, $\sigma_i(a_i = 1|s)$.

It is well known that even in a single agent problem, it is not possible to nonparametrically identify both $\Pi_i(a_i, a_{-i}, s)$ and the joint distribution of the error terms $F(\epsilon_i)$. To take the simplest possible example, consider a simple binary response model and assume that the error terms are normally distributed, as in the probit model. Let $\sigma_i(a_i = 1|s)$ denote the probability that the response is equal to one in the data conditional on s. Define $\Pi_i(a_i = 0|s) = 0$ and $\Pi_i(a_i = 0|s) = F^{-1}(\sigma_i(a_i = 1|s))$ where F^{-1} denotes the normal cdf. It can easily be verified that this definition of Π_i perfectly rationalizes any set of choice probabilities $\sigma_i(a_i = 1|s)$. Since our model is not identified without a parameteric assumption on the error term in even a simple probit model, assumptions at least as strong will be required in the more general set up that we consider here. In Bajari, Hong and Ryan (2005) we demonstrate that when we allow for nonparametric $\Pi_i(a_i, a_{-i}, s)$ an independence assumption is required to identify our model in even single agent problems. In what follows, we will typically impose the assumption that the error terms are distributed i.i.d. extreme value since both an independence and parametric form assumption on the error terms are required for identification.

Based on the discussion above, we shall impose the following assumption in order to identify the model.

A1 The error terms $\epsilon_i(a_i)$ are distributed i.i.d. across actions a_i and agents i. Furthermore, the parametric form of the distribution F comes from a known family.

Analogous to the notation in the previous section, define $\Pi_i(k, s) = \sum_{a_{-i}} \Pi_i(a_i = k, a_{-i}, s, \theta)\sigma_{-i}(a_{-i}|s)$. It is straightforward to show that the equilibrium in this model must satisfy:

$$\delta_i(s, \epsilon_i) = k \text{ if and only if } \Pi_i(k, s) + \epsilon_i(k) > \Pi_i(k', s) + \epsilon_i(k') \text{ for all } k' \neq k. \tag{9}$$

That is, action k is chosen if and only if the deterministic expected payoff and error term associated with action k is greater than the analogous varlues of $k' \neq k$. An implication of (9) is that the equilibrium choice probabilities $\sigma_i(a|s)$ must satisfy:

$$\sigma_i(a_i|s) = \Pr\{\epsilon_i(a_i) + \Pi_i(a_i, s) - \Pi_i(0, s) > \epsilon_i(k) + \Pi_i(k, s) - \Pi_i(0, s), \forall k = 0, \dots, K, k \neq a_i\} \tag{10}$$

Equation (10) is a simple consequence of (9) where we can subtract $\Pi_i(0, s)$ from both sides fo the inequality.

Suppose we generate $\epsilon_i(a_i)$ from an extreme value distribution as in the logit model. Then (10) implies that:

$$\sigma_i(a_i|s) = \frac{\exp(\Pi_i(a_i, s) - \Pi_i(0, s))}{\sum_{k=0}^{K} \exp(\Pi_i(k, s) - \Pi_i(0, s))} \tag{11}$$

A key insight due to Hotz and Miller (1993) is that equation (11) implies that the equilibrium choice probabilities, $\sigma_i(a_i|s)$, have a one to one relationship to the "choice specific value functions", $\Pi_i(a_i, s) - \Pi_i(0, s)$. To see why, note that taking logarithms of boths sides of (11) implies that for any k, k':

$$\log(\sigma_i(k|s)) - \log(\sigma_i(k'|s)) = \Pi_i(k, s) - \Pi_i(k', s).$$

The one to one mapping between choice probabilities and choice specific value functions holds more generally than in just the logit model. It is obvious that we should expect it in any model where the distribution of ϵ_i has full support. We let $\Gamma : \{0, ..., K\} \times S \to [0, 1]$ denote the map in general from choice specific value functions to choice probabilities, i.e.

$$(\sigma_i(0|s), ..., \sigma_i(K|s)) = \Gamma_i(\Pi_i(1, s) - \Pi_i(0, s), ..., \Pi_i(K, s) - \Pi_i(0, s)). \tag{12}$$

We will denote the inverse mapping by Γ^{-1}:

$$(\Pi_i(1, s) - \Pi_i(0, s), ..., \Pi_i(K, s) - \Pi_i(0, s)) = \Gamma_i^{-1}(\sigma_i(0|s), ..., \sigma_i(K|s)). \tag{13}$$

The above analysis implies that we can invert the equilibrium choice probabilities to nonparametrically recover $\Pi_i(1, s) - \Pi_i(0, s), ..., \Pi_i(K, s) - \Pi_i(0, s)$. However, the above analysis implies that we will not be able to seperately identify $\Pi_i(1, s)$ and $\Pi_i(0, s)$, we can only identify the difference between these two terms. Therefore, we shall impose the following assumption:

11

A2 For all i and all a_{-i} and s, $\Pi_i(a_i = 0, a_{-i}, s) = 0$.

The above assumption is similar to the "outside good" assumption in a single agent model where the mean utility from a particular choice is set equal to zero. In the context of our entry model, this assumption is satisfied if the profit from not entering the market is equal to zero regardless of the actions of other agents. Just as in the single agent model, there are alternative normalizations that we could use to identify the $\Pi_i(a_i, a_{-i}, s)$ just as in a single agent model. However, for expositional simplicity we shall restrict attention to the normalization A2.

Given assumption A2 and knowlege of the equilibrium choice probabilities, $sigma_i(a_i|s)$, we can then apply the mapping in (13) to recover $\Pi_i(a_i, s)$ for all i, a_i and s. Recall that the definition of $\Pi_i(a_i, s)$ implies that:

$$\Pi_i(a_i, s) = \sum_{a_{-i}} \sigma_{-i}(a_{-i}|s)\Pi_i(a_i, a_{-i}, s), \forall i = 1, \dots, n, a_i = 1, \dots, K. \tag{14}$$

Even if we know the values of $\Pi_i(a_i, s)$ and $\sigma_{-i}(a_{-i}|s)$ in the above equation, it is not possible to uniquely determine the values of $\Pi_i(a_i, a_{-i}, s)$ are not identified. To see why, hold the state vector s fixed, determining the utilities of all agents involves solving for $n \times K \times (K + 1)^{n-1}$ unknowns. That is, there are n agents, for each action $k = 1, ..., K$, utility depends on the $(K + 1)^{n-1}$ possible actions of the other agents. However, the left hand side of (14) only contains information about $n \times (K+1)$ scalars holding s fixed. It is clearly not possible to invert this system in order to identify $\Pi_i(a_i, a_{-i}, s)$ for all i, all $k = 1, ..., K$ and all $a_{-i} \in A_{-i}$. Related nonidentification results have been found by Bresnahan and Reiss (1991,1992) and Pesendorfer and Schmidt-Dengler (2003) in the context of dynamic games with discrete state spaces.

Obviously, there must be cross equation restrictions across either i or k in order to identify the system. An obvious way to identify the system is to impose exclusion restrictions. Partition $s = (s_i, s_{-i})$, and suppose $\Pi_i(a_i, a_{-i}, s) = \Pi_i(a_i, a_{-i}, s_i)$ depends only on the subvector s_i. An example of this might be in an entry model. In this type of model the state is usually a vector of productivity shocks. While we might expect the profit of firm i to depend on the entry decisions of other agents, it should not depend on the productivity shocks of other agents. See Bajari, Hong and Ryan (2004) for other examples of possible exclusion restrictions that can be used in applications. If such an exclusion restriction is possible, we can then write

$$\Pi_i(a_i, s_{-i}, s_i) = \sum_{a_{-i}} \sigma_{-i}(a_{-i}|s_{-i}, s_i)\Pi_i(a_i, a_{-i}, s_i). \tag{15}$$

Clearly, a sufficient identification condition is that for each s_i, there exists $(K + 1)^{n-1}$ points in the support of the conditional distribution of s_{-i} given s_i, such that this system of equations form by

12

these $(K+1)^{n-1}$ points given s_i is invertible. In other words, Let $s_{-i}^1, \dots, s_{-i}^{(K+1)^{n-1}}$ denote these points, then identification requires that the matrix

$$\left[\sigma(a_{-i}|s_{-i}^j, s_i), a_{-i} = 1, \dots, (K+1)^{n-1}, j = 1, \dots, (K+1)^{n-1} \right]$$

be nonsingular and invertible. Note that this assumption will be satisfied as long as s_{-i} contains a continuously distributed variable with $\Pi_i(a_i, a_{-i}, s_i)$ sufficient variability.

Theorem 1 *Suppose that A1 and A2 hold. Also suppose that for each s_i, there exists $(K+1)^{n-1}$ points in the support of the conditional distribution of s_{-i} given s_i so that (15) is invertible. Then the latent utilities $\Pi_i(a_i, a_{-i}, s_i)$ are identified.*

Another approach to identification is to exploit the multiplicity of equilbrium. Suppose that there is some state variable z which shifts which equilibrium to the game is played, but otherwise does not enter into the payoffs. Then we can write the payoffs are $\sigma_i(a_i|s, z)$ but the utilities do not directly depend on the variable z. We shall give a detailed example of such a variable in our application which is due to the intervention in the market by a regulator which we interpret as shifting the equilibrium to the game but not directly entering payoffs. This generates varation in the right hand size of (15) which allows us to check similar rank conditions. We note that Sweeting (2005) has also pointed out that the multiplicity of equilibrium can help in identifying a special case of the model above.

4   Estimation

In the previous section, we demonstrated that there is a nonparametric inversion between choice probabilities and the choice specific value functions, $\Pi(a_i, s)$. Furthermore, we demonstrated that the structural parameters of our model are identified if appropriate exclusion restrictions are made on payoffs. In this section, we exploit this inversion to construct nonparametric and semiparametric estimates of our structural parameters.

Step 1: Estimation of Choice Probabilities. Suppose the economist has access to data on $t = 1, \dots, T$ repetitions of the game. For each repetition, the economist observes the actions and state variables for each agent $(a_{i,t}, s_{i,t})$. In the first step we form an estimate $\hat{\sigma}_i(k|s)$ of $\sigma_i(k|s)$ using sieve series expansions (see Newey (1990) and Ai and Chen (2003)). We note, however, that we could alternatively estimate the first stage using other nonparametric regression methods such as kernel smoothing or local polynomial regressions.

13

Let $\{q_l(s), l = 1, 2, \dots\}$ denote a sequence of known basis functions that can approximate a real valued measurable function of s arbitrarily well for a sufficiently large value of l. The sieve could be formed using splines, Fourier Series or orthogonal polynomials. We let the basis become increasingly flexible as the number of repetitions of the game T becomes large. Let $\kappa(T)$ denote the number of basis functions to be used when the sample size is T. We shall assume that $\kappa(T) \to \infty$, $\kappa(T)/T \to 0$ at an appropriate rate to be specified below. Denote the $1 \times \kappa(T)$ vector of basis functions as

$$q^{\kappa(T)}(s) = (q_1(s), \dots, q_{\kappa(T)}(s)), \tag{16}$$

and its collection into a regressor data matrix as

$$Q_T = (q^{\kappa(T)}(s_1), \dots, q^{\kappa(T)}(s_T).$$

One potential sieve estimator for $\widehat{\sigma}_i(k|s), k = 1, \dots, K$ is a linear probability model, i.e.:

$$\widehat{\sigma}_i(k|s) = \sum_{t=1}^{T} 1(a_{it} = k) q^{\kappa(T)}(s_t)(Q_T' Q_T)^{-1} q^{\kappa(T)}(s). \tag{17}$$

Equation (17) is the standard formula for a linear probability model where the regressors are the sieve functions $\kappa(T)$ in equation (16). We note that in the presence of continuous state variables, the sieve estimator $\widehat{\sigma}_i(k|s)$ will converge to the true $\sigma_i(k|s)$ at a nonparametric rate which is slower than $\sqrt{T}$.

Second Step: Inversion   In our second step, we take as given our estimates $\widehat{\sigma}_i(k|s)$ of the equilibrium choice probabilities. We then form an estimate of the expected deterministic utility functions, $\hat{\Pi}_i(k, s_t) - \hat{\Pi}_i(0, s_t)$ for $k = 1, \dots, K$ and $t = 1, \dots, T$. This can be done by evaluating (13) using $\widehat{\sigma}_i(k|s)$ in place of $\sigma_i(k|s)$. That is:

$$\left(\widehat{\Pi}_i(1, s_t) - \widehat{\Pi}_i(0, s_t), \dots \widehat{\Pi}_i(K, s_t) - \widehat{\Pi}_i(0, s_t) \right) = \Gamma_i^{-1} \left(\widehat{\sigma}_i(0|s_t), \dots, \widehat{\sigma}_i(K|s_t) \right) \tag{18}$$

In the specific case of the logit model, this inversion would simply be:

$$\hat{\Pi}_i(k, s_t) - \hat{\Pi}_i(0, s_t) = \log\left(\widehat{\sigma}_i(k|s_t) \right) - \log\left(\widehat{\sigma}_i(0|s_t) \right) \tag{19}$$

In an alternative model, such as one with normal shocks, we would need to solve a nonlinear system. In what follows, we shall impose A2 so that $\hat{\Pi}_i(0, s) = \hat{\Pi}_i(0, a_{-i}, s) = 0$ for all a_{-i}.

14

Third Step: Recovering The Structural Parameters In the first step we recovered an estimate of $\hat{\sigma}_i(a_i, s)$ and in the second step we recovered an estimate of the choice specific value function $\hat{\Pi}_i(k, s)$. In our third step, we use the empirical analogue of (14) to form an estimate of $\Pi(a_i, a_{-i}, s_i)$. We shall assume that we have made a sufficient number of exclusion restrictions, as discussed in the previous section, so that the model is identified. For a given value of s_i, for a given $a = (a_i, a_{-i})$, we estimate $\Pi_i(a_i, a_{-i}, s_i)$ by minimizing the following weighted least square function $\Pi_i(a_i, a_{-i}, s_i)$, which are taken to be a vector of coefficients:

$$\sum_{t=1}^{T} \left(\hat{\Pi}_i(a_i, s_{-it}, s_i) - \sum_{a_{-i}} \hat{\sigma}_{-i}(a_{-i}|s_{-it}, s_i) \Pi_i(a_i, a_{-i}, s_i) \right)^2 w(t, s_i), \tag{20}$$

where the nonparametric weights $w(t, s_i)$ can take a variety of forms. For example,

$$w(t, s_i) = k\left(\frac{s_{it} - s_i}{h}\right) / \sum_{\tau=1}^{T} k\left(\frac{s_{i\tau} - s_i}{h}\right) \tag{21}$$

uses kernel weights, and other local weights are also possible. The identification condition in the previous section ensures that the regressor matrix in this weighted least square regression is nonsingular asymptotically.

4.1   A Linear Model of Utility

The nonparametric estimation procedure described in the previous section follows the identification arguments closely and offers the advantage of flexibility and robustness against misspecification. However, without a huge amount of data nonparametric estimation methods can be subject to severe curve of dimensionality when we intend to control for a large dimension of state variables s. Also, in small samples, different implementations of nonparametric procedures may lead to drastically different point estimates. Therefore in the following we consider a semiparametric estimation where the deterministic utility components $\Pi_i(a_i, a_{-i}, s)$ are specified to be a linear function of a finite dimensional parameter vector θ. This is the typical econometric specification that is commonly used in the empirical literature. In this section we describe a straightforward estimation and inference procedure for this model.

The deterministic part of the utility is assumed to take the form of

$$\Pi_i(a_i, a_{-i}, s_i) = \Phi_i(a_i, a_{-i}, s_i)' \theta. \tag{22}$$

In the above expression, the deterministic part of utility is a linear combination of a vector of basis functions, $\Phi_i(a_i, a_{-i}, s_i)$. For instance, we might let utility be a linear index as in our simple entry

game example of the previous section. Alternatively, we might choose $\Phi_i\left(a_i, a_{-i}, s_i\right)$ to be a standard flexible functional form, such as a high order polynomial, spline function or orthogonal polynomial. The estimator we discuss below can easily be generalized to allow for the possibility that θ enters the utility nonlinearly. However, the exposition of the properties of the estimator is facilitated by the linearity assumption. Also, most applications of discrete choice models and discrete games usually are linear in the structural parameters of interest.

This linearity assumption implies that the choice specific value function, given a_i and s, takes the convenient form:

$$\Pi_i\left(a_i, s\right) = E\left[\Pi_i\left(a_i, a_{-i}, s_i\right)|s, a_i\right] = \Phi_i\left(a_i, s\right)' \theta. \tag{23}$$

where $\Phi_i\left(a_i, s\right)$ is defined as:

$$\Phi_i\left(a_i, s\right) = \qquad E\left[\Phi_i\left(a_i, a_{-i}, s_i\right)|a_i, s\right] \tag{24}$$
$$= \quad \sum_{a_{-i}} \Phi_i\left(a_i, a_{-i}, s_i\right) \prod_{j \neq i} \sigma(a_j = k_j | s).$$

Arguing as in Section 2, it follows that if the error terms are distributed extreme value, the equilibrium choice probabilities must satisfy:

$$\sigma_i\left(a_i | s\right) = \frac{\exp\left(\Phi_i\left(a_i, s\right)' \theta\right)}{1 + \sum_{k=1}^{K} \exp\left(\Phi_i\left(k, s\right)' \theta\right)}. \tag{25}$$

More generally, if the error terms are not distributed extreme value, it may not be possible to express the equilibrium choice probabilities in closed form as above. However, equation (6) implies that each $\sigma_i\left(a_i | s\right)$ depends on $\sigma_j\left(a_j | s\right), j \neq i$ through (24). We denote this mapping as:

$$\sigma_i\left(a_i | s\right) = \Gamma_{i, a_i}\left(s, \sigma_j\left(k | s\right), j \neq i, k = 1, \dots, K\right). \tag{26}$$

If we define $\sigma\left(s\right)$ to be the stacked vector of choice probabilities $\sigma_i\left(k | s\right)$ for all $k = 1, \dots, K, i = 1, \dots, n$, then we can collect (26) into a fixed point mapping:

$$\sigma\left(s\right) = \Gamma\left(\sigma\left(s\right)\right). \tag{27}$$

To emphasize the dependence on the parameter θ, we can also write

$$\sigma\left(s; \theta\right) = \Gamma\left(s, \theta; \sigma\left(s; \theta\right)\right). \tag{28}$$

4.2 Semiparametric Estimation.

Step 1: Estimation of Choice Probabilities. The simple semiparametric procedure we propose proceeds in two steps. We begin by forming a nonparametric estimate of the choice probabilities, $\hat{\sigma}_i(k|s)$. We will do this like above using a sieve approach, however, one could alternatively use kernels or a local polynomial method.

$$\hat{\sigma}_i(k|s) = q^{\kappa(T)}(s)'(Q'_T Q_T)^{-1} \sum_{\tau=1}^{T} q^{\kappa(T)}(s_\tau) 1(a_i = k). \tag{29}$$

Given our estimates of the choice probabilities, we can then estimate $\Phi_i(k, s)$ correspondingly by

$$\hat{\Phi}_i(k, s) = \sum_{a_{-i}} \Phi_i(a_i = k, a_{-i}, s_i) \prod_{j \neq i} \hat{\sigma}(a_j|s) \tag{30}$$

For instance, take the example presented in (7). In this example, $\Pi_i(a_i = 1, a_{-i}, s) = (s, \sum_{j \neq i} 1\{a_j = 1\}) \cdot (\beta, \delta)$ where " $\cdot$ " denotes an inner product. Thus, in the above formual, $\Phi'_i(a_i = 1, a_{-i}) = (s, \sum_{j \neq i} 1\{a_j = 1\})$ and $\theta = (\beta, \delta)$. Then, given our first stage estimates of the choice probabilities, $\hat{\Phi}'(a_i = 1, a_{-i}) = (s, \sum_{j \neq i} 1\{a_j = 1\} \sigma_j(a_j|s))$.

Then for each parameter value θ, we can evaluate the empirical analogue of (28). For example, in the multinomial logit case,

$$\sigma_i\left(a_i|s, \hat{\Phi}, \theta\right) = \frac{\exp\left(\hat{\Phi}_i(a_i, s)' \theta\right)}{1 + \sum_{k=1}^{K} \exp\left(\hat{\Phi}_i(k, s)' \theta\right)} \tag{31}$$

Step 2: Parameter Estimation. In the second stage a variety of estimators can be used to recover the value of θ. Most of these estimators can be written as GMM estimators with a properly defined set of instruments. To describe the second stage, define $y_{ikt} = 1$ if $a_{it} = k$ and $y_{ikt} = 0$ otherwise, for $k = 0, \ldots, K$. Define $y_{it} = (y_{i1t}, \ldots, y_{iKt})$ and the vector

$$\sigma_i\left(s_t, \hat{\Phi}, \theta\right) = \left(\sigma_i\left(k|s_t, \hat{\Phi}, \theta\right), k = 1, \ldots, K\right) \tag{32}$$

Furthermore, collect $y_{it}, i = 1, \ldots, n$ into a long vector y_t with $n \times K$ elements, and similarly collect $\sigma_i\left(s_t, \hat{\Phi}, \theta\right), i = 1, \ldots, n$ into a long vector $\sigma\left(s_t, \hat{\Phi}, \theta\right)$ with corresponding $n \times K$ elements. Then for any dimension $\dim(\theta) \times (nK)$ matrix of instruments $\hat{A}(s_t)$, a GMM estimator $\hat{\theta}$ can be defined by solving the sample equations:

$$\frac{1}{T} \sum_{t=1}^{T} \hat{A}(s_t)\left(y_t - \sigma\left(s_t, \hat{\Phi}, \hat{\theta}\right)\right) = 0. \tag{33}$$

17

The instrument matrix $\hat{A}(s_t)$ may be known as $A(s_t)$, may be estimated in the first stage (such as two-step optimally weighted GMM), or may be estimated simultaneously (such as pseudo MLE). For example, with a pseudo MLE method, $\hat{\theta}$ is defined by maximizing the pseudo likelihood function, which can be written as the following because of the independent error terms across agents:

$$\sum_{t=1}^{T}\sum_{i=1}^{n}\left[\sum_{k=1}^{K} y_{ikt}\log\sigma_i\left(k|s_t,\hat{\Phi},\theta\right) + \left(1-\sum_{k=1}^{K} y_{ikt}\right)\log\left(1-\sum_{k=1}^{K}\sigma_i\left(k|s_t,\hat{\Phi},\theta\right)\right)\right]. \tag{34}$$

In the following we let $\hat{\sigma}_{ikt}(\theta) = \sigma_i\left(k|s_t,\hat{\Phi},\theta\right)$. The first order condition for the "pseudo" MLE estimator is then given by

$$\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{k=1}^{K}\left[\frac{y_{ikt}-\hat{\sigma}_{ikt}(\theta)}{\hat{\sigma}_{ikt}(\theta)} + \frac{\sum_{j=1}^{K}\left(y_{ijt}-\hat{\sigma}_{ijt}(\theta)\right)}{1-\sum_{j=1}^{K}\hat{\sigma}_{ijt}(\theta)}\right]\frac{\partial}{\partial\theta}\sigma_i\left(k|s_t,\hat{\Phi},\hat{\theta}\right), \tag{35}$$

In term of expressing it as an instrumental variable estimator, this corresponds to (33) with $\hat{A}(s_t)$ equal to the $\frac{\partial}{\partial\theta}\sigma\left(s_t,\hat{\Phi},\hat{\theta}\right)\times\hat{W}(s_t,\theta)$, where $\hat{W}(s_t,\theta)$ is the $nK\times nK$ diagonal matrix, such that the ith diagonal block, for $i=1,\ldots,n$, is given by

$$\begin{bmatrix}\frac{1}{\hat{\sigma}_{i1t}(\theta)} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\hat{\sigma}_{iKt}(\theta)}\end{bmatrix} + \begin{bmatrix}\frac{1}{\hat{\sigma}_{i0t}(\theta)} \\ \vdots \\ \frac{1}{\hat{\sigma}_{i0t}(\theta)}\end{bmatrix}\begin{bmatrix}1\cdots 1\end{bmatrix}, \tag{36}$$

where $\hat{\sigma}_{i0t}(\theta) = 1 - \sum_{j=1}^{K}\hat{\sigma}_{ijt}(\theta)$. If we define $W(s_t,\theta)$ in analog to $\hat{W}(s_t,\theta)$, except with $\hat{\Phi}$ replaced by the true unknown Φ in each $\hat{\sigma}_{ikt}(\theta)$, then it is easy to verify that $W(s_t,\theta_0) = Var\left(y_t - \sigma(s_t,\Phi,\theta_0)|s_t\right)^{-1}$, where the ith diagonal block of the block diagonal matrix $Var\left(y_t - \sigma(s_t,\Phi,\theta_0)|s_t\right)$ is given by

$$\begin{bmatrix}\sigma_{i1t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{iKt}\end{bmatrix} - \begin{bmatrix}\sigma_{i1t} \\ \vdots \\ \sigma_{iKt}\end{bmatrix}\begin{bmatrix}\sigma_{i1t}\cdots\sigma_{iKt}\end{bmatrix}. \tag{37}$$

It is well known that the estimation errors in $\hat{A}(s_t)$ will not affect the asymptotic distribution of $\hat{\theta}$ defined by (33), regardless of whether $\hat{A}(s_t)$ is estimated in a preliminary step or is estimated simultaneously with $\hat{\theta}$. Therefore, the asymptotic distribution of the pseudo MLE is the same as that of an infeasible GMM IV estimator with an instrument matrix $A(s_t)$ which is given by

$$\frac{\partial}{\partial\theta}\sigma(s_t,\Phi_0,\theta_0)\times W(s_t,\theta_0) \tag{38}$$

18

Here note that in terms of (28)

$$\frac{\partial}{\partial\theta}\sigma\left(s_t,\Phi_0,\theta_0\right) = \frac{\partial}{\partial\theta_1}\Gamma\left(s_t,\theta_1,\sigma\left(s_t;\theta_2\right)\right)\Big|_{\theta_1=\theta_2=\theta} \tag{39}$$

Therefore, assuming all Φ are known, the MLE corresponds to the effecient choice of instruments $\hat{A}\left(s_t\right)$ in (33), under the efficiency framework of [7]. Therefore, next we will focus on deriving the large sample properties of $\hat{\theta}$ defined by (33) where $A(s_t)$ is known.

The estimator that we consider falls within the class of semiparametric estimators considered by Newey (1994). A somewhat surprising conclusion is that even though the first stage is estimated nonparametrically and can be expected to converge at a rate slower than $\sqrt{T}$, the structural parameters will be asymptotically normal and will converge at a rate of $\sqrt{T}$. Moreover, under appropriate regularity conditions, the second stage asymptotic variance will be independent of the particular choice of nonparametric method used to estimate the first stage (e.g. sieve or kernel). As a practical matter, these results justify the use of the bootstrap to calculate standard errors for our model.

In the appendix, we derive the following result, applying the general framework developed by [5]. Under appropriate regularity conditions, the asymptotic distribution of $\hat{\theta}$ defined in 33 satisfies the following:

$$\sqrt{T}\left(\hat{\theta}-\theta\right) \overset{d}{\longrightarrow} N\left(0, G^{-1}\Omega G^{-1'}\right),$$

where

$$G = EA\left(s_t\right)\frac{\partial}{\partial\theta}\sigma\left(s_t,\Phi_0,\theta_0\right), \tag{40}$$

and Ω is defined such that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}A\left(s_t\right)\left(y_t - \sigma\left(s_t,\hat{\Phi},\theta_0\right)\right) \overset{d}{\longrightarrow} N\left(0,\Omega\right). \tag{41}$$

In the appendix, we also compare the asymptotic variance of alternative estimators.

4.3   Linear Probability Model

Next, we demonstrate that in many cases it is possible to estimate our model using a simple two stage least squares method. While the formal econometrics in this section are quite straightforward, we believe that this observation is useful for researchers. Much of the previous literature on estimating games has concentrated on the case of fully dynamic models and required an understanding of dynamic programming. In this section, we demonstrate that the methods in their simplest version only require invoking the ivreg command in STATA.

To simplify notation, we will focus on a binary choice model in the following. The extension to multinomial models is straightforward. Therefore consider a special case of (1) where $K = 1$, $\Pi_i(0, a_{-i}, s) = \epsilon_i(0) \equiv 0$, θ, $\Phi_i(1, a_{-i}, s; \theta) = \Phi_i^1(s_i)' \beta + \Phi_i^1(a_{-i}, s)' \gamma$ so that

$$u_i(1, a_{-i}, s; \theta) = \Phi_i^1(s_i)' \beta + \Phi_i^2(a_{-i}, s)' \gamma + \epsilon_i(1),$$

and $\epsilon_i(1)$ is uniformly distributed on the interval $(-1, 0)$. The decision rule of agent i is such that action 1 is chosen if and only if

$$E[u_i(1, a_{-i}, s; \theta) | s] = \Phi_i^1(s_i)' \beta + E[\Phi_i^2(a_{-i}, s) | s]' \gamma + \epsilon_i(1) > 0.$$

Because of the uniform distribution assumption of $\epsilon_i(1)$, this implies that

$$P(a_i = 1 | s) = \Phi_i^1(s_i)' \beta + E[\Phi_i^2(a_{-i}, s) | s]' \gamma,$$

which can be written as

$$a_i = \Phi_i^1(s_i)' \beta + E[\Phi_i^2(a_{-i}, s) | s]' \gamma + \eta_{it}(1), \tag{42}$$

such that $E(\eta_{it}(1) | s_t) = 0$.

In this model $\Phi_i^1(s_i)$ and $\Phi_i^2(a_{-i}, s)$ are known functions, but β and γ are unknown parameters. $E[\Phi_i^2(a_{-i}, s) | s]$ needs to be estimated nonparametrically. As before, this can be consistently estimated by a linear least square regression with a sufficiently large number of sieve approximating functions in the first step:

$$\hat{E}[\Phi_i^2(a_{-i}, s) | s] = q^{\kappa(T)}(s)' (Q_T' Q_T)^{-1} \sum_{\tau=1}^{T} q^{\kappa(T)}(s_\tau) \Phi_i^2(a_{-i\tau}, s_\tau).$$

In the second step, a linear regression is used to estimate β and γ:

$$\left(\hat{\beta}, \hat{\gamma}\right) = \arg\min_{\beta, \gamma} \sum_{t=1}^{T} \left(a_{it} - \Phi_i^1(s_{it})' \beta - \hat{E}[\Phi_i^2(a_{-it}, s_t) | s_t]' \gamma\right)^2.$$

Note that this is precisely a two stage least square estimator. Also note that this is a special case of (33) where

$$\hat{A}_i(s_t) = \begin{pmatrix} \Phi_i^1(s_{it}) \\ \hat{E}[\Phi_i^2(a_{-it}, s_t) | s_t] \end{pmatrix},$$

and $\sigma\left(s_t, \hat{\Phi}, \hat{\theta}\right) = \Phi_i^1(s_{it})' \beta + \hat{E}[\Phi_i^2(a_{-it}, s_t) | s_t]'$ and $y_{it} - \sigma_i(s_t, \theta_0) = \eta_{it}(1)$.

In the following we will demonstrate that the ivreg procedure with the robust standard error options in stata will give correct standard errors for this model, when a_i is entered as the dependent

variable, $\Phi_i^1(s)$ is entered as the exogenous regressor, $\Phi_i^2(a_{-i}, s)$ is entered as the endogenuous regressor, and $q^{\kappa(T)}(s)$ is used as the set of instruments to instrument for $\Phi_i^2(a_{-i}, s)$. In other words, $\hat{\beta}$ and $\hat{\gamma}$ are computed as

$$\hat{\theta} = \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \widehat{\mathcal{A}}_T^{-1} \frac{1}{T} \sum_{t=1}^{T} \hat{A}_i(s_t) a_{it},$$

where

$$\widehat{\mathcal{A}}_T = \frac{1}{T} \sum_{t=1}^{T} \hat{A}_i(s_t) \hat{A}_i(s_t)'.$$

Then it follows from a standard manipulation that

$$\sqrt{T} \begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} = \widehat{\mathcal{A}}_T^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{m}_i(s_t, \beta, \gamma),$$

where

$$\hat{m}_i(s_t, \beta, \gamma) = \hat{A}_i(s_t) \left(a_{it} - \Phi_i^1(s_{it})' \beta - \hat{E}\left[\Phi_i^2(a_{-it}, s_t)|s_t\right]' \gamma \right).$$

The asymptotic variance is then a special case of section (C). We will use a slightly different notation here to facilitate comparison with 2SLS. In particular, it follows from [6] that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{m}_i(s_t, \beta, \gamma) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} m_i(s_t, \beta, \gamma) + o_p(1),$$

where $m_i(s_t, \beta, \gamma) = A_i(s_t) (\eta_{it}(1) - v_{it}(1)\gamma)$, and

$$v_{it}(1) = \Phi_i^2(a_{-it}, s_t) - E\left[\Phi_i^2(a_{-it}, s_t)|s_t\right].$$

The asymptotic distribution of the 2SLS estimator is then given by, letting $\mathcal{A} = E A_i(s_t) A_i(s_t)'$:

$$\sqrt{T}\left(\hat{\theta} - \theta\right) \xrightarrow{d} N\left(0, \mathcal{A}^{-1} E A_i(s_t) \underbrace{Var\left(\eta_{it}(1) - v_{it}(1)' \gamma | s_t\right)}_{\Sigma(s_t)} A_i(s_t) \mathcal{A}^{-1}\right).$$

To obtain a consistent estimate of this asymptotic variance matrix, both $\mathcal{A}$ and $\Sigma(s_t)$ needs to be estimated. $\mathcal{A}$ can be estimated by $\mathcal{A}_T$. The key in estimating $\Sigma(s_t)$ is to estimate $\eta_{it}(1) - v_{it}(1)$ consistently. This can be estimated by

$$e_{it} = a_{it} - \Phi_i^1(s_{it})' \hat{\beta} - \Phi_i^2(a_{-it}, s_t)' \hat{\gamma}$$
$$= a_{it} - \Phi_i^1(s_{it})' \hat{\beta} - E\left[\Phi_i^2(a_{-it}, s_t)|s_t\right]' \hat{\gamma} - \left(\Phi_i^2(a_{-it}, s_t) - E\left[\Phi_i^2(a_{-it}, s_t)|s_t\right]\right)' \hat{\gamma}.$$

Then $EA_i(s_t)\Sigma(s_t)A_i(s_t)'$ can be consistently estimated by

$$\frac{1}{T}\sum_{t=1}^{T}\hat{A}_i(s_t)e_{it}e_{it}'\hat{A}_i(s_t)'.$$

Now note that this is precisely how *Stata ivreg* calculates robust standard errors for 2SLS estimators.

5 Fixed effect model of unobserved heterogeneity

5.1 Nonparametric identification

Unobserved heterogeneity can be a potentially important source of misspecification if it is not accounted for in the model. Specifically, we can introduce fixed effects as a method of controlling for unobserved heterogeneity into our model. For each market $t = 1, \dots, T$, all the players in that market observe a market specific profit shifter α_t that is not observed by the econometrician. We will assume that α_t is an unknown but smooth function of the state variables $s_t = (s_{t1}, \dots, s_{tn})$ in that market, which we will denote as $\alpha(s_t)$. In other words, the mean period utility function in (1) for player i in market t is now modified to

$$\Pi_i(a_{it}, a_{-it}, s_t; \theta) = \alpha(a_i, s_t) + \tilde{\Pi}_i(a_{it}, a_{-it}, s_{it}; \theta). \tag{43}$$

In the following we may omit the market specific subscript and simply denote this as

$$\Pi_i(a_i, a_{-i}, s; \theta) = \alpha(a_i, s) + \tilde{\Pi}_i(a_i, a_{-i}, s_i; \theta).$$

Under the normalization assumption that $\Pi_i(0, a_{-i}, s) \equiv 0$ for all $i = 1, \dots, n$, our previous results show that, as in (14), the expected utilities given each i's information (the left hand side of the following equation) are nonparametrically identified, $\forall i = 1, \dots, n, a_i = 1, \dots, K$:

$$
\begin{aligned}
\Pi_i(a_i, s) &= \sum_{a_{-i}} \sigma_{-i}(a_{-i}|s)\Pi_i(a_i, a_{-i}, s) \\
&= \alpha(a_i, s) + \sum_{a_{-i}} \sigma_{-i}(a_{-i}|s)\tilde{\Pi}_i(a_i, a_{-i}, s_i).
\end{aligned} \tag{44}
$$

Obviously, since $\alpha(a_i, s)$ is unknown but is the same function across all market participants, they can be difference out by looking at the difference of $\Pi_i(k, s)$ and $\Pi_j(k, s)$ between different players i and j. This allows us to identify some aspect of all the pairwise differences $\tilde{\Pi}_i(a_i, a_{-i}, s_i)$ and $\tilde{\Pi}_j(a_j, a_{-j}, s_j)$ between all the possible pairs i, j of players. By differencing (44) between i and j one obtains

$$\Pi_i(k, s) - \Pi_j(k, s) = \sum_{a_{-i}} \sigma_{-i}(a_{-i}|s)\tilde{\Pi}_i(a_i, a_{-i}, s_i) - \sum_{a_{-j}} \sigma_{-j}(a_{-j}|s)\tilde{\Pi}_j(a_j, a_{-j}, s_j)$$

Here we can treat $\tilde{\Pi}_i(a_i, a_{-i}, s_i)$ and $\tilde{\Pi}_j(a_j, a_{-j}, s_j)$ as coefficients, and $\sigma_{-i}(a_{-i}|s)$ and $\sigma_{-j}(a_{-j}|s)$ as regressors in a linear regression. As long as for given values of s_i, s_j, there are sufficient variations in the remaining state variables s_{-ij} to generate sufficient variation in the regressor matrix $\sigma_{-i}(a_{-i}|s)$ and $\sigma_{-j}(a_{-j}|s)$, the coefficients $\tilde{\Pi}_i(a_i, a_{-i}, s_i)$ and $\tilde{\Pi}_j(a_j, a_{-j}, s_j)$ can be nonparametrically identified.

More specific assumptions on $\tilde{\Pi}_j(a_j, a_{-j}, s_j)$, such as symmetry assumptions or parametric assumptions, can provide more transparent identification arguments. **Symmetry:** If the mean payoff functions are symmetric between any two given players i and j, then the choice set of competitors a_{-i} and a_{-j} are identical and can be commonly denoted as $a_{-ij} \equiv a_{-i} = a_{-j}$. In this case, for each given $k = 1, \ldots, K$, one can difference (44) between i and j to obtain

$$\Pi_i(k, s) - \Pi_j(k, s) = \sum_{a_{-ij}} \sigma_{-ij}(a_{-ij}|s) \left[\tilde{\Pi}_{ij}(k, a_{-ij}, s_i) - \tilde{\Pi}_{ij}(k, a_{-ij}, s_j) \right].$$

Therefore can trace out the above difference in the mean utility function for all the possible pairs s_i and s_j. In particular, by moving s_i and s_j close to each other this identifies the derivative of the entire function $\tilde{\Pi}_{ij}(k, a_{-ij}, s_{ij})$ and hence identifies the entire function up to a constant term.

5.2 Parametric mean utilities

Alternative estimation strategies are available under the parametric specification of the mean utility function given in (22), which is now modified to apply to the individual heterogenous payoff component:

$$\tilde{\Pi}_i(a_i, a_{-i}, s_i) = \Phi_i(a_i, a_{-i}, s_i)' \theta. \tag{45}$$

In this case one can estimate the parameters θ using a panel data rank estimation type procedure ([4]) or a conditional logit estimator in the case of a multinomial logit model. Note that while a semiparametric rank estimator is free of distribution assumptions of the error terms, to compute a simulated equilibrium give a prespecified set of fixed effects one still needs to impose a distribution assumption on the error terms. In the following we will focus on models with binary action choices such as entry games, where either $a_{it} = 1$ or $a_{it} = 0$.

5.2.1 Conditional Logit Estimation

If the error terms are specified to form a multinomial choice model as in (11), one can form a conditional logit likelihood function ([2]):

$$\log L = \sum_{t=1}^{T} \left(\log \left[\exp \left(\theta' \sum_{i=1}^{n} a_{it} \hat{\Phi}_i(1, s_{it}) \right) \right] - \log \left[\sum_{d_t \in B_t} \exp \left(\theta' \sum_{i=1}^{n} d_{it} \hat{\Phi}_i(1, s_{it}) \right) \right] \right)$$

where $B_t \equiv \{d_t : \sum_{i=1}^n d_{it} = \sum_{i=1}^n a_{it}\}$ is the set of actions that give rise to the same number of entrants as observed in the data. In other words, the likelihood is formulated conditional on the number of entrants. One can follow similar steps as in previous derivations to prove consistency and demonstrate asymptotic normality of this semipametric conditional logit estimator where $\hat{\Phi}_i(1, s_{it})$ is estimated in the first stage nonparametrically.

5.2.2 Rank Estimation

A rank estimator can be formulated as minimizing over θ the objective function

$$\sum_{t=1}^T \sum_{i=1}^n \sum_{j \neq i} 1\left(a_{it} > a_{jt}\right) \rho_- \left(\left(\left(\hat{\Phi}_i(1, s_{it}) - \hat{\Phi}_j(1, s_{jt}) \right)' \theta \right) \right).$$

In the above $\rho_-(x)$ is a penalty function that penalizes the negative part of x. One could use a constant penalty function

$$\rho_-(x) = 1\left(x < 0\right)$$

which is also used in [4]). Alternatively one can also use a quadratic penalty function

$$\rho_-(x) = 1\left(x < 0\right) x^2.$$

to smooth out the objective function with respect to the parameter θ.

6 Computing Models with Multiple Equilibria.

In the previous sections, we have either assumed that the model has a unique equilibrium (which can be the case, for example, for a linear probability interaction model), or that only a single equilibrium outcome out of several possible multiple equilibria is being observed in the data set. However, in many static game models, multiple equilibria are possible. The importance of multiple equilibria in empirical research is emphasized by many authors, including [1] and [8]. In the rest of this manuscript we present a method for estimating parametric models of interactions in the presence of possible multiple equilibria.

In the previous sections we have considered a model with known distribution $F(\epsilon_i)$ of the error terms and a parametric model for the mean utility functions $\Pi_i(a_i, a_{-i}, \theta)$. At every possible parameter value θ, given the known distribution $F(\epsilon_i)$, equations (10), (12) and (14) defined a fixed point mapping in the conditional choice probabilities:

$$\sigma_i(a_i|s) = \Gamma_i \left(\sum_{-i} \sigma_{a_{-i}}(a_{-i}|s) \left[\Pi_i(k, a_{-i}, s; \theta) - \Pi_i(0, a_{-i}, s; \theta), k = 1, \ldots, K \right] \right). \tag{46}$$

For example, under the linear mean utility specification (5), this system of fixed point mappings in the choice probabilities takes the form of

$$\sigma_i(a_i|s) = \Gamma_i \left(\sum_{a_{-i}} \sigma_{-i}(a_{-i}|s)\Phi_i(a_i, a_{-i}, s)'\theta, \ a_i = 1, \ldots, K \right), \ i = 1, \ldots, n. \qquad (47)$$

In previous sections, we have assumed that either there is a unique solution to this system of fixed mapping with $K \times n$ equations and $K \times n$ unknown variables

$$\sigma_i(a_i|s), \forall a_i = 1, \ldots, K, i = 1, \ldots, n,$$

or that only one particular fixed point of this system gets realized in the observed data. However, this system of fixed point mapping can potentially have multiple solutions, leading to the possibility of multiple equilibria.

In the following of this section, we will first give a brief introduction to the use of Homotopy methods for finding multiple solutions to a system of nonlinear equations. Secondly, we will discuss how the Homotopy method of nonlinear equation solving can be used to compute multiple equilibria for our model of static interactions. In the next section, we will describe an innovative estimation method that we develop for estimating the parameters in static interaction models with multiple equilibria.

6.1   The Homotopy method

The Homotopy continuation method (which be simply be referred to as the Homotopy method in the rest of the paper) is a well known generic algorithm for looking for a fixed point to a system of nonlinear equations. Based on the particular nonlinear system of equations to be solved, a well designed Homotopy system is capable of finding multiple solutions of the nonlinear system, and in some cases, all solutions of these nonlinear systems. A concise and clear introduction to the Homotopy method can be found in the book by [11]. Algorithms for implementing the homotopy method can be found in several papers including [9] and [3].

Our goal is to find, for all possible parameter values and realized state variables s, the solutions for the fixed point system (28):

$$\sigma - \Gamma(\sigma) = 0,$$

where we have used $\sigma = \sigma(s)$ to simply the dependence of choice probabilities on the state variables. A homotopy is a linear mapping between the two topological spaces of functions of the form

$$H(\sigma, \tau) = \tau G(\sigma) + (1 - \tau)(\sigma - \Gamma(\sigma)), \quad \tau \in [0, 1], \qquad (48)$$

where each of $H(\sigma, \tau)$ and $G(\sigma)$ are vectors of functions with $n \times K$ component functions:

$$H_{i,a_i}(\sigma, \tau) \quad \text{and} \quad G_{i,a_i}(\sigma) \quad \text{for} \quad i = 1, \ldots, n \quad \text{and} \quad a_i = 1, \ldots, K.$$

As one can see, for $\tau = 0$ we obtain $H(\sigma, 0) = \Gamma(\sigma)$ and for $\tau = 1$ we get $H(\sigma, 0) = G(\sigma)$ so varying τ from 1 to 0 maps the function $G(\cdot)$ into the function $\Gamma(\cdot)$. If for each $0 \leq \tau < 1$, we can solve for the nonlinear equations, $H(\sigma, \tau) = 0$, then by moving along the path in the direction of $\tau = 1$ to $\tau = 0$, at the end of the path we should be able to reach a solution of the original nonlinear equations $\sigma - \Gamma(\sigma) = 0$. This path then constructs a mapping between a solution of the initial system $G(\sigma) = 0$ to a solution of the original nonlinear system $\sigma - G(\sigma) = 0$. Typically $G(\sigma)$ is a system of equations that are very easy to solve to obtain all the solutions.

Once a solution $\sigma(1)$ for the initial system $G(\sigma) = 0$ is found, algorithms based on solving differential equations can be used to trace the path from $\tau = 1$ to $\tau = 0$. At each τ, we denote the solution along a particular path by $\sigma(\tau)$:

$$H(\sigma(\tau), \tau) = 0.$$

By differentiating this homotopy function with respect to τ:

$$\frac{d}{d\tau} H(\sigma(\tau), \tau) = \frac{\partial H}{\partial \tau} + \frac{\partial H}{\partial \sigma} \cdot \frac{\partial \sigma}{\partial \tau} = 0.$$

This defines a system of differential equations for $\sigma(\tau)$ with initial condition $\sigma(1)$ calculated from the solution of the (easy) initial system $G(\sigma(1)) = 0$. A number of computer algorithms that are available to compute numerical solutions of nonlinear systems of differential equations can then be used to trace this differential equation system to reach an end point in the path of $\tau = 1$ to $\tau = 0$ in order to obtain a solution $\sigma(0)$ of the original system $\sigma - \Gamma(\sigma) = 0$.

A regularity condition is necessary to insure the stability and the proper behavior of the homotopy differential equation system.

Condition 1 (Regularity) *Let $\nabla(\tau)$ denote the Jacobian of the Homotopy functions with respect to σ at the solution path $\sigma(\tau)$:*

$$\nabla(\tau) = \frac{\partial}{\partial \sigma} Re\{H(\sigma, \tau)\}\bigg|_{\sigma = \sigma(\tau)},$$

where $Re\{H(\sigma, \tau)\}$ denotes the real component of the homotopy functions. The jacobian $\sigma(\tau)$ has full rank for almost all τ.

This condition ensures the smoothness and differentiability of the paths. It rules out cases of bifurcation, branching and infinite spiraling. The mapping between $G(\sigma)$ and $\sigma - \Gamma(\sigma)$ is called

a conformal one if the path that links them is free of these complications. If a homotopy system satisfies the regularity condition, it will either reach a solution or drift off to infinity.

A convenient way of generating a conformal mapping, or a homotopy that satisfies the above regularity condition, is to extend the original homotopy into the complex space. A homotopy, as defined in equation (48) is extended to the complex space by allow for the argument σ to take on complex values, which will result in complex values of the homotopy. When the real and complex components of σ are considered two different sets of arguments, and the real and complex components of the output are considered two different sets of components of the homotopy, this defines a real value homotopy with $2nK$ inputs and $2nK$ outputs. If the original homotopy (48) is an analytic function in the complex space, then the Cauchy-Riemann conditions will ensure that the extended real-value $2nK \times 2nK$ homotopy system satisfies the regularity condition above.

The all solution homotopy is one where the initial system $G(\sigma)$ is chosen such that, if we follow the paths originating from each of the solution from $G(\sigma) = 0$, we will reach all solutions of the original system $\sigma = \Gamma(\sigma)$ at the end of the path. The extension of a real homotopy system into the complex space is essential to the idea of all-solution homopoty. It is related with the property of the complex space that conformal mappings in the complex space do not change the algebraic properties of the sets. [3] The concept of conformal mapping is widely used in complex analysis. It implies that we can conformally map a function with a complicated set of roots to another function with a simple set of roots, then the results from an analysis of the function with the simple set of roots should apply to the function with the complicate set of roots. An all solution homotopy has to satisfy an additional path finiteness condition:

Condition 2 (Path Finiteness) *Define $H^{-1}(\tau)$ to be the set of solutions $\sigma(\tau)$ to the homotopy system at τ. $H^{-1}(\tau)$ is bounded for all $0 \leq \tau < 1$. In other words, for all $\tau > 0$.*

$$\lim_{||\sigma|| \to \infty} H(\sigma, \tau) \neq 0.$$

6.2 Multiple equilibria in static discrete games

As we noted in the previous section, the issue of multiple equilibria in static interaction models amounts to the issue of computing all the fixed points to the system of equations of choice probabilities defined in equation (46). Note that the argument to the mapping from expected utility to choice probabilities, $\Gamma(\cdot)$, is linear in the choice probabilities of competing agents $\sigma_{-i}(a_{-i}|s)$.

---

[3]The conformal mapping is a mapping in the complex space that does not locally change the argument of a complex number. It is known that if the function is analytic and does not have stationary points then the associated mapping is conformal.

Therefore, the question of possible multiplicity of equilibria depends crucially on the functional form of Γ, which in turn depends exclusively on the assumed joint distribution of the error terms.

Interestingly, if we are content with the linear probability model where Γ is a linear function of the individual choice probabilities, then the equilibrium will be guaranteed to be unique and the issue of multiple equlibiria is not relevant. Γ is a linear function of the individual choice probabilities if the underlying utility functions depend linear on the indicator function of whether individual competitors make a particular choice or not. For example, this would be the case if the profit of entering a market depends on the total number of competitors who also enter the market. On the other hand, if we have nonlinear interactions of the individual choice probabitilities in the linear probability model, or if the joint distribution of the error term in the multinomial choice model is specified such that Γ_i is a polynomial function for each $i = 1, \dots, n$, then all the equilibria can be found by choosing a homotopy system where the initial system of equation

$$G_{i,a_i}(\sigma), \quad i = 1, \dots, n \quad \text{and} \quad a_i = 1, \dots, K.$$

takes the following simple polynomial form:

$$G_{i,a_i}(\sigma) = \sigma_i(a_i)^{q_{i,a_i}} - 1 = 0 \quad \text{for} \quad i = 1, \dots, n \quad \text{and} \quad a_i = 1, \dots, K, \tag{49}$$

where q_{i,a_i} is an integer that exceeds the degree of the polynomial of Γ_{i,a_i} as a function of $\sigma_{-i}(a_{-i})$. This results in a homotopy mapping

$$H_{i,a_i}(\sigma, \tau) = \tau\{\sigma_i(a_i)^{q_{i,a_i}} - 1\} + (1-\tau)(\sigma_i(a_i) - \Gamma_{i,a_i}(\sigma)), \ \tau \in [0,1]. \tag{50}$$

For $\tau = 0$ the system (48) coincides with the original system while for $\tau = 1$ it is equal to the 'simple' system (49).

It is elementary knowledge that there are exactly q_{i,a_i} complex roots to $G_{i,a_i}(\sigma)$ that are evenly distributed on the unit circle. It is known that nondegenerate polynomial functions are analytic and the regularity condition of the resulting homotopy system is automatically satisfied. The particular choice of q_{i,a_i} also ensures the path finiteness property of the homotopy system (c.f. [11]).

While a polynomial model for $\Gamma(\cdot)$ is convenient for calculating multiple equilibria, it is rarely used in practice because it is not clear what parametric utility specification will give rise a polynomial choice probability function. The most popular multinomial choice probability function is probably the multinomial logit, which is well known to derive from i.i.d. extreme value distributions in the unobserved component of the latent utilities. The logit choice model is given in (25). Our analysis in the following will consist of three steps. First we will establish that this model has a finite number of equilibria represented by real solutions to (25). Secondly, we will shown that letting the degree

of polynomial homotopy increase to infinity at an appropriate rate, we will be able to find all the equilibria for the multinomial logit choice model (25) using the homotopy method. We prove it by first verifying that the homotopy mapping is regular in the complex space when the discontinuity points of the original function are isolated, and then providing a method to make homotopy work in the small vicinity of discontinuity points.

It is not difficult to show that the fixed point system (25) has a finite number of solutions in the real line. To see this, note that in general, this function is clearly continuous and infinitely differentiable with nonsingular derivatives. In particular, it is easy to verify this condition for the multinomial logit and probit models that are commonly used in practice. Consider a compact ball B_R in $\mathbb{R}^{nK}$ with radius larger than 1. By Sard's theorem the set of its irregular values has measure zero. In particular, it can be verified by differentiation through the implicit function theorem that zero is its regular value. This implies that the submanifold of σ satisfying $\sigma = \Gamma(\sigma)$ is compact as a closed subset of this ball B_R. Thus it contains a finite number of points. This verifies that the set of solutions in B_R in finite. It is also obvious that all the solutions must satisfy $0 \leq \sigma_i(a_i) \leq 1$ because of the choice probability functional form. Therefore there can not be solutions outside B_R.

While we have just shown that there are in general a finite number of multiple equilibria, to compute these equilibria we need to make use of an all solution homotopy system. In particular, we choose to use the all solution homotopy defined in (50). In the following we will show that there exists an homotopy system of the form of (50) with a sufficiently high orders of the initial system q_{i,a_i}'s, such that we will find all the solution to the original system of choice probabilities. As described in the previous subsection, verifying the validity of the all solution homotopy requires checking the regularity condition and the path finiteness condition, which in terms require extension of the real homotopy system into the complex space. Verifying these condions require us to specify the particular functional form of the joint distribution of the error terms in the latent utilities. In the following we will focus on the multinomial logit case which is the most widely used discrete choice model in the empirical literature.

The following Theorem 2 and Theorem 2 formally state this result. In the statement of the theorems, $\sigma = \{\sigma_r, \sigma_i\}$ denotes more generally a vector of the real part and the imaginary part of complex numbers which extend the real choice probabilities we considered early into the complex space. Theorem 1 first establishes the regularity properties of the homotopy outside the imaginary subspace.

Theorem 2 *Define the sets $H^{-1} = \{(\sigma_r, \sigma_i, \tau) \mid H(\sigma_r, \sigma_i, \tau) = 0\}$ amd*

$$H^{-1}(\tau) = \{(\sigma_r, \sigma_i) \mid H(\sigma, \tau) = 0\} \quad for \quad \sigma_r \in \mathbb{R}^{nK}, \quad and \quad \sigma_i \in \mathbb{R}^{nK}.$$

Note that H is a homotopy of dimension R^{2nK} that include both real and imaginary parts separately. Also define, for any small ϵ, $\wp_\epsilon = \cup_{i,a_i}\{|\sigma_{r,i,a_i}| \leq \epsilon\}$ to be the area around the imaginary axis. Then:

1) The set $H^{-1} \cap \{\mathbb{R}^{2nK} \setminus \wp_\epsilon \times [0,1]\}$ consists of closed disjoint paths.

2) For any $\tau \in (0,1]$ there exists a bounded set such that $H^{-1}(\tau) \cap \mathbb{R}^{2nK} \setminus \wp_\epsilon$ is in that set.

3) For $(\sigma_r, \sigma_i, \tau) \in H^{-1} \cap \{\mathbb{R}^{2nK} \setminus \wp_\epsilon \times [0,1]\}$ the homotopy system allows parametrization $H(\sigma_r(s), \sigma_i(s), \tau(s)) = 0$. Moreover, $\tau(s)$ is a monotone function.

Remark: Theorem 2 establishes the regularity and path finiteness conditions for the homotopy (50) for the multinomial logit model in areas that are not close to the pure imaginary subspace in the complex domain $\mathbb{C}^{nK}$. The homotopy system can become irregular along the pure imaginary subspace, because the denominator in the system can approach zero and the system will become nonanalytic in the case. However, the next theorem shows that there exists a sequence q such that homotopies with initial system of order q will have paths that stay away from the imaginary subspace. Homotopies with these orders will be able to trace out all the solutions of the original multinomial logit system.

Theorem 3 *For given τ one can pick the power q_{i,a_i} of the initial function (49) such that the homotopy system is regular and path finite given some sequence of converging polyhedra $\wp_\epsilon$, $\epsilon \to 0$.*

Theorem 3 implies that if we continue to increase the power q_{i,a_i} of the initial system (49) of the homotopy, we will eventually be able to find all the solutions to the original multinomial logit system. This also implies, however, we might lose solutions when we continue to increase q_{i,a_i}. But Theorem 3 does imply that for sufficiently large q_{i,a_i}, no new solutions will be added for larger powers. In the monte carlo simulation that we will report in the next section, we do find this to be the case.

6.3 Monte Carlo Analysis

We perform several monte-Carlo simulations for an entry game with a small number of potential entrants. Player's payoff functions for each player i were constructed as linear functions of the indicator of the rival's entry ($a_i = 1$), market covariates and a random term:

$$U_i(a_i = 1, a_{-i}) = \theta_1 - \theta_2 \left(\sum_{j \neq i} \mathbf{1}(a_j = 1) \right) + \theta_3 x_1 + \theta_4 x_2 + \epsilon_i(a), \ i = 1, \dots, n. \qquad (51)$$

The payoff of staying out is equal to $U_i(a_i = 0, a_{-i}) = \epsilon_i(a)$, where the $\epsilon_i(a)$ have i.i.d extreme value distributions across both a and i. The coefficients in the modela are interpreted as: θ_1 is the fixed benefit of entry, θ_2 is the loss of utility when one other player enters, θ_3, θ_4 are the sensitivities of the benefit of entry to market covariates.

The game can be solved to obtain *ex-ante* probabbilities of entry in the market. The solution to this problem is given by:

$$P_i = \frac{e^{\theta_1 - \theta_2\left(\sum_{j \neq i} P_j\right) + \theta_3 x_1 + \theta_4 x_2}}{1 + e^{\theta_1 - \theta_2\left(\sum_{j \neq i} P_j\right) + \theta_3 x_1 + \theta_4 x_2}}, \ i = 1, \dots, n. \tag{52}$$

Here P_i is the *ex-ante* probability of entry for the player i, $P_i = p(a_i = 1|x)$.

Both coefficients of the model and market covariates were taken from independent Monte-Carlo draws. The parameters of generated random variables are presented in the table below.

Table 1: Characteristics of the parameters

Parameter	Mean	Variance	Distribution
θ_1	2.45	1	Normal
θ_2	5.0	1	Normal
θ_3	1.0	1	Normal
θ_4	-1.0	1	Normal
x_1	1.0	0.33	Uniform
x_2	1.0	0.33	Uniform

The means and variances of parameter values and market covariates were chosen so to have a fair percentage of cases with more then one equilibrium.

For the games with 3,4 and 5 players 400 independent parameter combinations for every player were taken. The modification of HOMPACK algorithm was run to solve for all equilibria in each game.

Throughout the Monte-Carlo runs both coefficients and covariates x_1 and x_2 were changing. So, basically every equilibrium was calculated for a specific set of parameters.

Summary statistics for the results of computations are presented below.

It is possible to see that in the constructed games the players have approximately same average parameters in every type of game. This agrees with the symmetric form of underlying data generating process for the coefficients and market covariates.

Characteristics of the estimates

	Mean	Std Dev	Max	Min
n = 3				
Parameter	Mean	Std Dev	Max	Min
# of equilibria	1.513	1.035	7	1
P_1	0.369	0.364	0.999	0.000
P_2	0.357	0.367	0.995	0.000
P_3	0.364	0.346	0.993	0.003
n = 4				
# of equilibria	1.285	0.682	4	1
P_1	0.280	0.328	0.980	0
P_2	0.236	0.316	0.980	0
P_3	0.276	0.341	1.000	0
P_4	0.297	0.345	0.990	0
n = 5				
# of equilibria	1.126	0.475	4	1
P_1	0.102	0.198	0.96	0
P_2	0.146	.260	0.98	0
P_3	0.315	.343	0.99	0
P_4	0.336	0.375	0.98	0
P_5	0.315	0.349	0.98	0

7 Application to stock market analysts' recommendations and peer effects

Many of the ideas developed in this paper can be applied to the problem of analyzing the behavior of equity market analysts and the stock recommendations that they issue. Clearly, the set of recommendations on a stock can be viewed as the outcome of a game. Analysts make choices among a different set of actions, such as whether to recommend the buying or the selling of a stock. Analysts probably have different information about a given company's prospects, and it is well-understood in the profession that these information asymmetries exist. Most importantly, payoffs to the individual analyst depend, to a large extent, on the actions taken by competitor analysts. Accurate forecasts and recommendations are highly valued, of course. But the penalty for issuing a poor recommendation depends on whether competitor analysts also made the same poor recommendation.

There has been a revival of interest on the determinants of analyst recommendations as researchers have tried to explain the remarkable behavior of the analysts in the run-up and subsequent collapse of the NASDAQ in 2000. [4] The focus in this paper is on the recommendations generated for firms in the high tech sector, which includes the firms most affected by the excitement surrounding the development of e-commerce and the spread of the Internet. Given the great uncertainty surrounding the demand for new products and new business models, the late 1990's would seem to have been the perfect environment for equity analysts to add value. Yet analyst recommendations were not particularly helpful or profitable during this period. For example, the analysts were extremely slow to downgrade stocks, even as it was apparent that the market had substantially revised its expectations about the technology sector's earnings potential. Barber, Lehavey, McNichols, and Trueman (2001) show that the least recommended stocks earned an average abnormal return of 13% in 2000-2001, while the most highly recommended stocks earned average abnormal returns of -7%. Observations like this have led commentators to wonder whether the analysts had ulterior motives for keeping their recommendations unjustifiably optimistic, such as the pressure to win investment banking business. Allegedly, this conflict of interest took the form of analysts keeping recommendations on stocks high in order to appease firms, who would then reward the analyst's company by granting it underwriting business or other investment advisory fees.[5] Indeed, these suspicions

---

[4]See for example Barber, Lehavey, McNichols, and Trueman (2001) and Chan, Karceski, and Lakonishok (2003). Research prior to the NASDAQ collapse includes Womack (1996), Lin and McNichols (1998), and Michaely and Womack (1999).

[5]In 1998, Goldman Sachs estimated that Jack Grubman, a prominent telecommunications industry analyst, would bring in $100 to $150 million in investment banking fees. This estimate was based on the fees generated by 32 of the stocks he covered that also had banking relationships with Citigroup, including

came to a head when New York State Attorney General Elliot Spitzer launched an investigation into conflicts of interest in the securities research business.

In this application we develop an empirical model of the recommendations generated by stock analysts from the framework outlined in section 1. We quantify the relative importance of four factors influencing the production of recommendations in a sample of high technology stocks during the time period between 1998 and 2003.

First, recommendations must depend on expectations about the future profitability of a firm. There should be some systematic component to these expectations common across all analysts and investors that will be embedded in the current stock price.

Second, analysts are heterogeneous, both in terms of talent and perhaps in terms of access to information. We try to capture an individual analyst's private belief about the stock by looking at the difference between the quarterly earnings forecast submitted by the analyst (or the analyst's brokerage firm) and the distribution of forecasts from other firms.

Mindful of the large number of inquiries into possible conflicts of interest among research analysts, we include as a third factor a dummy variable for an investment banking relationship between the firm and the analyst's employer.

Finally, we consider the influence of peers on the recommendation decision. Peer effects can impact the recommendation in different ways. Individual analysts have incentive to condition their recommendation on the recommendations of their peers, because even if their recommendations turn out to be unprofitable ex-post, performance evaluation is typically a comparison against the performance of peers. More subtly, recommendations are relative rankings of firms and are not easily quantifiable (or verifiable) objects. As such, ratings scales usually reflect conventions and norms. The phenomenon is similar to the college professor's problem of assigning grades. If a professor were to award the average student with a C while other faculty give a B+ to the average student, the professor might incorrectly signal his views of student performance. Even while there is heterogeneity in how individual professors feel about grading, most conform to norms if only to communicate clearly with students (and their potential employers) about their performance. Similarly, analysts have an incentive to benchmark their recommendations against perceived industry norms.

7.1 Data

Our data consist of the set of recommendations on firms that made up the NASDAQ 100 index as of year-end 2001. The recommendations were collected from Thomson Firstcall. Firstcall is one of the most comprehensive historical data sources for analysts' recommendations and earnings forecasts,

WorldCom, Global Crossing and Winstar Communications. (Wall Street Journal, October 11, 2002).

containing real-time recommendations and forecasts from hundreds of analysts. It is common for analysts to rate firms on a 5 point scale, with 1 denoting the best recommendation and 5 denoting the worst. When this is not the case, Firstcall converts the recommendations to the 5 point scale (see Table 1).

We have 12,719 recommendations from analysts at 185 brokerage firms over this time period (see Table 2). The dependent variable in our data set is a recommendation submitted between January of 1998 and June of 2003 for a firm in the NASDAQ 100. The data set was formed by merging the earnings and recommendations files from Firstcall. In a given quarter, for a given stock, we merge a quarterly earnings forecast with a recommendation from the same brokerage.[6] This will allow us to determine if analysts that are more optimistic than the consensus tend to give higher recommendations. In the Firstcall data, quarterly earnings forecasts are frequently made more than a year in advance. In order to have a consistent time frame, we limit analysis to forecasts that were made within the quarter that the forecast applies.[7] Note that not every recommendation can be paired with an earnings forecast made within the contemporaneous quarter. Recommendations that could not be paired with an earnings forecast were dropped from the results that we report. However, qualitatively similar results were found for a data set where this censoring was not performed. We choose not to report these results in the interests of brevity. The variables in our data include:

- REC- Recommendation from 1-5 for a stock listed in the NASDAQ 100 recorded by I/B/E/S.

- QUARTER- Quarter during which the recommendation was submitted.

- STOCK-Name of the stock for which the recommendation applies.

- BROKERAGE-The brokerage employing the analyst.

- EPS-Earnings per share forecast submitted by the analyst's brokerage associated with the recommendation. Submitted during the same quarter as the recommendation.

- AEPS-Average of the earnings per share forecasts submitted for that quarter.

- RELATION-A dummy variable that is one if the analyst's brokerage engages in investment banking business with the company to which the recommendation applies.

---

[6]When there were multiple recommendations by the same analyst within a quarter, we chose to use the last recommendation in the results that we report.

[7]We chose to merge the brokerage field, instead of the analysts field, because the names and codes in the analysts field were not recorded consistently across IBES data sets for recommendations. It was possible to merge at the level of the brokerage.

- IBANK-A dummy variable that is equal to one if the brokerage does any investment banking business with stocks in the NASDAQ 100.

- SPITDUM-A dummy variable that is equal to one after the quarter starting in June of 2001. Based on a comprehensive search of Wall Street Journal articles, this is when Elliot Spitzer began making very public criticisms of industry practices.

- SBANK-the share of analysts that issued recommendations for a particular stock during a particular quarter where IBANK was one.

The investment banking relationship was identified from several different sources. First, we checked form 424 filings in the SEC's database for information on the lead underwriters and syndicate members of debt issues. When available, we used SEC form S-1 for information on financial advisors in mergers. We also gathered information on underwriters of seasoned equity issues from Securities Data Corporation's Platinum database. To be sure, transaction advisory services (mergers), and debt and equity issuance are not the only services that investment banks provide. However, these sources contribute the most to total profitability of the investment banking side of a brokerage firm.

The average recommendation in our data set is 2.2, which is approximately a buy recommendation (see Table 2). The mean value of RELATION is 0.035. The mean value of IBANK is 0.81. That is, 3.5 percent of the analyst-company pairs in our data set were identified as having a potential conflict of interest due to some kind of investment banking activity for the stock in question. Eighty-one percent of the recommendations in our data were generated by firms engaging in investment banking with some firm list in the NASDAQ 100. Both of these variables are potentially useful measures of potential conflict of interest. The variable RELATION is more direct, since it indicates that the brokerage is engaged in investment banking with the company during the quarter the recommendation was issued. However, brokerages might view any company it is giving a recommendation to as a potential client, particularly in the NASDAQ 100, where many of the companies generated considerable investment banking fees.

The variable earnings was formed by merging the recommendations and earnings files in Firstcall. In a given quarter, for a given stock, we merge the quarterly earnings forecast with the recommendation from the same brokerage. This allows us to determine if analysts that are more or less optimistic than the consensus tend to give higher recommendations. In the Firstcall data, quarterly earnings forecasts are frequently made more than a year in advance. In order to have a consistent time frame, we limit analysis to forecasts that were made within the quarter for which the forecast applies. We chose to merge the brokerage field, instead of the analysts field, because the names and codes in the analysts field were not recorded consistently across Firstcall data sets for recommendations. It was

possible to merge at the level of the brokerage. Note that not every recommendation can be paired with an earnings forecast made in the contemporaneous quarter.

7.2 Empirical model

An observation is a recommendation submitted for a particular stock during a specific quarter. We will let $t = 1, ..., T$ denote a quarter, $s = 1, ..., S$ a stock and and $i = 1, ..., I$ an analyst. We will denote a particular recommendation by $r_{i,s,t}$. The recommendation can take on integer values between 1 and 5, where 1 is the highest recommendation and 5 the lowest. Since the dependent variable can be naturally ranked from highest to lowest, we will assume that the utilities come from an ordered probit. Let $x(i, s, t)$ denote a set of covariates that influence the recommendation for analyst i for stock s during quarter t. Let $x(s, t)$ denote a vector of $(x(i, s, t))$ of payoff relevant covariates that enter into the utility of all the analysts who submit a recommendation for stock s during quarter q. Let $z(s, t)$ denote a set of covariates that shift the equilibrium, but which do not influence payoffs.

Define the utility or payoff to analyst i for a recommendation on stock s in quarter t to be,

$$y_{i,s,t} = \beta' x(i, s, t) + \eta E(r|x(s, t), z(s, t)) + \varepsilon_{i,s,t} \tag{53}$$

In equation (53), the term $E(r|x(s, t), z(s, t))$ is the expected recommendation for stock s during quarter t and $\varepsilon_{i,s,t}$ is a normal error term. Thus, conforming to the expected actions of peers enters into an individual analyst's utility. If we were to assume that the error term is normally distributed, then the model would become the familiar ordered probit, where the probability that a particular recommendation is observed is determined as follows:

$$P(r = 1) = \Phi(-\beta'x(i,s,t) - \eta E(r|x(s,t), z(s,t))) \tag{54}$$

$$P(r = 2) = \Phi(\mu_1 - \beta'x(i,s,t) - \eta E(r|x(s,t), z(s,t))) \tag{55}$$

$$- \Phi(-\beta'x(i,s,t) - \eta E(r|x(s,t), z(s,t))) \tag{56}$$

$$P(r = 3) = \Phi(\mu_2 - \beta'x(i,s,t) - \eta E(r|x(s,t), z(s,t))) \tag{57}$$

$$- \Phi(\mu_1 - \beta'x(i,s,t) - \eta E(r|x(s,t), z(s,t))) \tag{58}$$

$$P(r = 4) = \Phi(\mu_3 - \beta'x(i,s,t) - \eta E(r|x(s,t), z(s,t))) \tag{59}$$

$$- \Phi(\mu_2 - \beta'x(i,s,t) - \eta E(r|x(s,t), z(s,t))) \tag{60}$$

$$P(r = 5) = 1 - \Phi(\mu_3 - \beta'x(i,s,t) - \eta E(r|x(s,t), z(s,t))) \tag{61}$$

In equations (53)-(61), the likelihood that determines the probability that the recommendation is r depends on the latent estimated covariates β and η along with the cut points $\mu_1 - \mu_3$.

7.3 Exclusion restrictions and identification

The analysis of the previous section suggests that identification depends crucially on having appropriate exclusion restrictions. First, we need covariates that influence the payoffs of one particular agent, but not other agents. In our analysis, the covariates will include broker fixed effects and covariates that reflect broker-specific levels of investment banking activity. This assumption would imply, for instance, that the amount of investment banking done by Merrill Lynch should not directly influence the recommendations submitted by analysts working for Goldman Sachs. We believe that this is a reasonable assumption.

Second, identification can be achieved by using covariates $z(s,t)$ that influence the selection of equilibrium, but which do not directly enter into payoffs. The first set of covariates is based on the actions of market regulators. 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 descriptive analysis in the earlier section suggests that this intervention by the regulator encouraged the industry to focus on an equilibrium where more "sell" and "hold" recommendations were issued. We divide time into pre and post "Spitzer" eras with the mid point being the quarter starting in June 2001. Based on coverage in the Wall Street Journal, we believe this is an appropriate starting point for the very vocal and public criticism of the industry by the regulator. Therefore, a first instrument is this dummy variable marking the timing of this

regulatory regime shift. We believe that the actions of Spitzer helped to shift the equilibrium to a more conservative set of recommendations. It is plausible to argue that SPITDUM can be excluded from payoffs. The analysts could scarcely be accused of conflicts of interest for recommendations issued at this time because there was very little investment banking work being done after June of 2001 due to the downturn high technology markets.

The variable SPITDUM only displays time-series variation. We will interact SPITDUM with SBANK in order to get cross sectional variation in the instrument as well. This would be an appropriate instrument if the criticism of the regulators had more impact for stocks where a larger proportion of the analysts were affiliated with brokerages that had traditionally done investment banking work with high technology firms.

A second set of variables that could enter z include average recommendations for the stock submitted in previous quarters. As we can see from Table 3, there were no holds or sell recommendations issued by analysts in the first quarter of 2000 in our data set. If an individual analyst were to deviate from this practice, he might run the risk of miscommunicating to investors his views about the relative desirability of stock s. Therefore, we include behavior in the most recent period since this will communicate information about the norm. Normally, we would be concerned that this instrument is itself arguably endogenous. However, we note that we can control for time and stock effects since we have a panel data set. Also, since we also have access to the instruments based on the actions of the regulator, we can check the robustness of our results both with and without using lagged behavior as an instrument.

7.4 Results

7.4.1 Fundamentals

The first question that we ask is the extent to which recommendations were determined by publicly observable information about the stocks. In our data, these fundamentals correspond to time fixed effects, stock fixed effects and the difference between an individual analyst's beliefs about earnings and beliefs in the market as whole. In Table 4, we run an ordered probit to explore these questions. The variable %DEV is the percentage deviation of an analyst's recommendation from the average recommendation and ABS. DEV is the algebraic difference. In both cases, a more optimistic earnings forecast has the anticipated sign, but is not significant at conventional levels in any of the specifications that we have tried. On the other hand, quarterly and stock fixed effects are almost all statistically significant (not reported in this Table). If quarter and stock fixed effects proxy for publicly available information about the stock, then this information is considerably more important than measures of an individual analyst's optimism. We note, however, that there could be other

interpretations of these variables.

In Table 5, we report the estimated quarterly fixed effects and compare these estimates to the NASDAQ Index and NASDAQ 100. The quarterly effects are labeled qdum2-qdum22 for the 2nd quarter (1998.Q2) through the 22nd quarter (2003.Q2) of our data set. Several points emerge. First, the quarterly effects are typically significant. Second, the quarterly effects are highly correlated with the NASDAQ index and with the QQQ. We take these results as evidence that the quarterly effects can reasonably be interpreted as reflecting publicly observed information about the the firms that is embedded in the share prices as opposed to some other latent effects. The movement in share prices can explain most of the movement in the recommendations in the previous tables. In Table 6, we regress the quarterly dummies on these indexes. We find that the indexes are statistically significant at conventional levels and that the measures of goodness of fit are quite high.

7.4.2 Conflicts of Interest

In Table 7 we run an ordered logit model of recommendations as a function of our conflict of interest measures. The coefficient on RELATION indicates that potential conflicts of interest are statistically significant at conventional levels, except for the third column where quarterly and stock fixed effects are included, and the fourth column where the full set of fixed effects are included along with the more inclusive IBANK variable. The coefficient sign on RELATION is also consistent with our a priori beliefs that conflicts of interest could lead to issuing more favorable recommendations. Consistent with the work of prior researchers and market regulators, these results do suggest that a potential conflict of interest exists when a brokerage engaged in investment banking business with a company that quarter. However, these results must be interpreted with some caution. Since brokerage firms are expected to cover companies with whom they have significant investment banking business, the firms have an incentive to select brokerages that already view them favorably. It would be hard to imagine that a rational manager would want to hire an investment banking firm that views her company in an unfavorable manner.

Interestingly, we note (column 4 of table 7) that when we include the IBANK variable denoting that the brokerage firm has investment banking relationships with at least one firm in the index, the coefficient is positive and significant. Evidently, analysts at general investment banking firm tend to issue slightly more conservative recommendations. This is consistent with a view of the world that companies tended to select investment banking firms that were more favorably disposed towards them. Alternatively, our results could be interpreted as suffering from a bias due to some omitted variable.

Our results suggest that even though investment banking relationships may generate potential

conflicts of interest for equity analysts, the magnitude of the effects recommendations may be small in practice. Notice that measures of the goodness of fit are very low when only investment banking is included. Also, the marginal effects of engaging in investment banking are small. In our data, the expected recommendation is -0.08 when we allow for quarterly fixed effects, but not stock fixed effects. This finding is not consistent with the prosecutors belief that "unbiased" research, separate from investment banking, will generate recommendations less tainted by potential conflicts of interest. However, the behavior of analysts after large settlements have been paid and significant damage has been done to their brokerage's reputation, in some cases, may be more conservative.

7.4.3 Peer Effects

The final question we consider is whether there are peer effects in submitting recommendations. We do this by using the two stage procedure described in the previous section. First, we flexibly model the expectations of individual analysts about the recommendations that will be submitted by other analysts. We begin by regressing the average recommendation submitted for a particular stock within a given quarter on the following covariates:

1. time dummies

2. stock dummies

3. SBANK- the fraction of analysts that have IBANK equal to one

4. SPITDUM- a dummy for whether the quarter follows the quarter in which Elliot Spitzer began publicly questioning the research quality of Wall Street analysts

5. ESHIFT-the interaction of spitdum and sbank

6. LAG-the lagged value of the average recommendation for the stock

In our analysis, we will assume that variables 4-6 above are valid exclusion restrictions in the sense that they shift the equilibrium that is played in the market, but do not directly enter into an analyst's utility function. The excluded variables are significant in this first stage regression. We will let IVBELIEF denote the fitted value of the regression. We will suppose that the fitted value of the regression corresponds to their beliefs.

In all of the specifications that we examine in Table 8, peer effects seem to be important. An individual analyst will raise his recommendation proportionally to the recommendation that he expects from other analysts. This is intuitive. A given recommendation does not make senses in isolation, but only relative to the recommendations of other analysts. If no one else in the market is

issuing recommendations of "market underperform" or "sell", an individual analysts may give the wrong signal by issuing such a recommendation even if he believes the recommendation is literally true.

It is worth noting that the results for our measure of peer effects are not only statistically significant, but peer effects also explain the results quite well compared to the other covariates. The Pseudo-R^2 using only the fitted values from the first stage is virtually the same as in the specifications including time and stock fixed effects. The other variables are not significant in our final column and do not increase measures of goodness of fit. We obtain nearly identical results when the first stage is done with a sieve estimator formed by using splines (Table 9). The estimated coefficients on IVBELIEF range from roughly 1.8 to 2.3 in both tables, depending primarily on the covariates included in the second stage. The fit of the ordered logits in Tables 8 and 9 changes little when the fixed effects are included as covariates.

With estimates of the model's key parameters in hand, it is possible to simulate the model and find the set of equilibria to the analyst recommendation game. In keeping with the static nature of our analysis, we solve for all possible equilibra at two points in time: the first just prior to the regulatory regime change in 2000.Q1, and then again in 2003.Q1 after the Spitzer inquiry was well-underway. These results are based on a simplified case of a 2-player game (see the appendix for details on the computation procedure).

Evidently, the pre-Spitzer era was characterized by two different equilibria (see Tables 1 and 2 in the appendix and Figures 1 and 2 below). In the post-Spitzer era, there was only one equilibrium, given our estimated parameter values reported in the ordered logit specification in Table 8.

8   Conclusion

In this paper we propose a method for estimating static games of incomplete information. The method we propose is semiparametric and does not require the covariates to lie in a discrete set. Perhaps most importantly, the method is both flexible and easy to implement using standard statistical packages. We also introduce an algorithm for computing all equilibria to a game, which is useful for policy simulations using the estimated model.

We apply these methods to the problem of determining the factors that govern the assignment of stock recommendations by equity analysts for a set of high tech stocks between 1998 and 2003. Two factors seem to be most important for explaining the production of stock recommendations. First, publicly observable information about the stocks under recommendation, as reflected in our time and quarter dummies, plays a large role in explaining the distribution of recommendations. As

we saw in Table 4.3, these variables explained a large fraction of the variation in the data and were highly correlated with market indexes such as the NASDAQ or QQQ. Simply put, recommendations improved in 1999-2000 as the stock market rose. The second and most important factor for explaining recommendations is the peer group effect. Individual analysts raise their recommendations proportionally to the recommendations they expect from their peers. Investment banking relationships are shown to be statistically significant in the recommendations regressions, but the economic effect of the investment banking relationship is estimated to be small, and the variable does not add much to the total fit of the model.

References

[1] William Brock and Steven Durlauf. Discrete choice with social interactions. *Review of Economic Studies*, 68:235–260, 2001.

[2] G. Chamberlain. Panel data. In Z. Griliches and M. Intriligator, editors, *Handbook of Econometrics, Vol. 2*. North Holland, 1984.

[3] M. M. Kostreva and L. A. Kinard. A differential homotopy approach for solving polynomial optimization problems and noncoorperative games. *Computers and Mathematics with Applications*, 21:135–143, 1991.

[4] Charles F. Manski. Semiparametric analysis of random effect linear models from binary panel data. *Econometrica*, 55(2):357–362, 1987.

[5] W. Newey. Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5(2):99–135, 1990.

[6] W. Newey. The asymptotic variance of semiparametric estimators. *Econometrica*, 62:1349–82, 1994.

[7] W. Newey and D. McFadden. Large sample estimation and hypothesis testing. In R. Engle and D. McFadden, editors, *Handbook of Econometrics, Vol. 4*, pages 2113–2241. North Holland, 1994.

[8] Andrew Sweeting. Coordination games, multiple equilibria and the timing of radio commercials. Northwestern University Working Paper, 2005.

[9] Layne T. Watson, Stephen C. Billups, and Alexander P. Morgan. Algorithm 652 hompack: A suite of codes of globally convergent homotopy algorithms. *ACM Transactions on Mathematical Software*, 13(3):281–310, 1987.

[10] W. I. Zangwill and C. B. Garcia. Global continuation methods for finding all solutions to polynomial systems of equations in n variables. In A. V. Fiacco and K. O. Kortanek, editors, *Extremal Methods and Systems Analysis, Lecture Notes in Economics and Mathematical Systems*, pages 481–498. Springr-Verlag, 1980.

[11] W. I. Zangwill and C. B. Garcia. *Pathways to Solutions, Fixed Points, and Equilibria*. Prentice-Hall, 1981.

A Proof for Theorem 2

Before we set out to prove the theorem we need to introduce some notations. Collapse the indexation for $i = 1, \dots, n$ and $a_i = 1, \dots, K$ to a single index $j = 1, \dots, nK$. In other words, each j represents a (i, a_i) pair. First we will rewrite the expression (47) for the case of multinomial choice probability as:

$$\sigma_j = \frac{\exp(\mathcal{P}_j(\sigma))}{1 + \sum_{k \in I_i} \exp(\mathcal{P}_k(\sigma))}, \tag{62}$$

where $I_i = \{(i, a_i), a_i = 1, \dots, K\}$ is the set of all indices $j = (i, a_i)$ that corresponds to the set of strategies available to player i, and $P_j(\sigma)$ is the expected utility associated with player i for playing a_i when $j = (i, a_i)$:

$$\mathcal{P}_j(\sigma) = P_{i,a_i}(\sigma) = \sum_{a_{-i}} \sigma_{-i}(a_{-i}|s) \Phi_i(a_i, a_{-i}, s)' \theta,$$

which is in general a polynomial function in σ_j. Let $\mathcal{P}(\cdot)$ denote the vector-function of polynomials of size $nK \times 1$ that collects all the elements $\mathcal{P}_j(\cdot)$ for $j = 1, \dots, nK$. Let Q be the product of the degrees of the polynomial over all elements of the vector $\mathcal{P}(\cdot)$. In other words, $Q = \prod_{j=1}^{nK} Q_j$ where Q_j is the degree of polynomial $\mathcal{P}_j(\cdot)$. For each complex argument $\xi \in \mathbb{C}^{nK}$ the system of polynomials has Q solutions has exactly Q solutions. Because of this, for each $\xi \in \mathbb{C}^{nK}$ we can find Q vectors σ^* such that $\mathcal{P}(\sigma^*) = \xi$. Let us denote each particular vector σ^* by $\mathcal{P}_{(k)}^{-1}(\xi)$.

The complex-valued vector $\mathcal{P}(\cdot)$ of dimension $nK \times 1$ can be transformed into a real-valued vector of dimension $2nK \times 1$ by considering separately real and complex part of vector $\mathcal{P}(\cdot)$. Because of the polynomial property, each $\mathcal{P}_{(k)}^{-1}(\xi)$ is a continuously differentiable function of ξ for almost all ξ. It is possible that for some range of the argument ξ, two (or more) solution paths $\mathcal{P}_{(k)}^{-1}(\xi)$ and $\mathcal{P}_{(k')}^{-1}(\xi)$ for $k \neq k'$ might coincide with each other. In this case we will relabel the paths k so that the merged paths to replicated to create a total Q of smooth solution paths $\mathcal{P}_{(k)}^{-1}(\xi)$.

44

The following analysis will apply to each individual branch $\mathcal{P}^{-1}_{(k)}(\xi)$, which we will just denote by $\mathcal{P}^{-1}(\xi)$ without explicit reference to the path indice k. For $j = 1, \dots, nK$ introduce the following notations: $\xi_j = x_j + iy_j$, $\rho_j = \|\xi_j\|$, $\varphi_j = \arctan\left(\frac{y_j}{x_j}\right)$. Then a homotopy system can be constructed for (62) as:

$$H_{1j}(\xi, \tau) = \{\rho_j^q \cos(q\varphi_j) - 1\}\tau + (1 - \tau)\left\{Re\{\mathcal{P}^{-1}(\xi)\} - \right.$$
$$\left. -\frac{e^{2x_j} + e^{x_j}\cos(y_j) + \sum_{k \neq j} e^{x_j + x_k}\cos(y_j - y_k)}{1 + \sum_{k \in I_i} e^{2x_k} + 2\sum_{k \in I_i} e^{x_k}\cos(y_k) + \sum_{l \in I_i}\sum_{k \neq l} e^{x_k + x_l}\cos(y_l - y_k)}\right\}$$

and

$$H_{2j}(\xi, \tau) = \{\rho_j^q \sin(q\varphi_j) - 1\}\tau + (1 - \tau)\left\{Im\{\mathcal{P}^{-1}(\xi)\} - \right.$$
$$\left. -\frac{e^{x_j}\sin(y_j) + \sum_{k \neq j} e^{x_k + x_j}\sin(y_j - y_k)}{1 + \sum_{k \in I_i} e^{2x_k} + 2\sum_{k \in I_i} e^{x_k}\cos(y_k) + \sum_{l \in I_i}\sum_{k \neq l} e^{x_k + x_l}\cos(y_l - y_k)}\right\}$$

If the system $\mathcal{P}(\cdot)$ is polynomial, $\mathcal{P}^{-1}(\xi)$ is smooth and has a Jacobian of full rank for almost all ξ. Therefore, we can locally linearize it so that $\mathcal{P}^{-1}(\xi) \approx \Lambda\xi + \mathcal{C}$. The homotopy system can then be written as:

$$H_{1j}(\xi, \tau) = \{\rho_j^q \cos(q\varphi_j) - 1\}\tau + (1 - \tau)\left\{\Lambda^j x_j - \right.$$
$$\left. -\frac{e^{2x_j} + e^{x_j}\cos(y_j) + \sum_{k \neq j} e^{x_j + x_k}\cos(y_j - y_k)}{1 + \sum_{k \in I_i} e^{2x_k} + 2\sum_{k \in I_i} e^{x_k}\cos(y_k) + \sum_{l \in I_i}\sum_{k \neq l} e^{x_k + x_l}\cos(y_l - y_k)}\right\} \tag{63}$$

and

$$H_{2j}(\xi, \tau) = \{\rho_j^q \sin(q\varphi_j) - 1\}\tau + (1 - \tau)\left\{\Lambda^j y_j - \right.$$
$$\left. -\frac{e^{x_j}\sin(y_j) + \sum_{k \neq j} e^{x_k + x_j}\sin(y_j - y_k)}{1 + \sum_{k \in I_i} e^{2x_k} + 2\sum_{k \in I_i} e^{x_k}\cos(y_k) + \sum_{l \in I_i}\sum_{k \neq l} e^{x_k + x_l}\cos(y_l - y_k)}\right\} \tag{64}$$

where Λ^j is the jth row of the $nK \times nK$ matrix Λ. Without loss of generality we will let $\mathcal{C} = 0$ in subsequent analysis for the sake of brevity because all the results will hold for any other given $\mathcal{C}$. To simplify notation we will denote:

$$\Theta_i(x, y) = \sum_{k \in I_i} e^{2x_k} + 2\sum_{k \in I_i} e^{x_k}\cos(y_k) + \sum_{l \in I_i}\sum_{k \neq l} e^{x_k + x_l}\cos(y_l - y_k)$$

Now given some index $k \in \{1, \dots, Q\}$, we consider the solutions of the system $\{H(x, y, \tau)\} = 0$ for all possible real values of the vectors of x and y.

Now we set out to prove the statements of Theorem 2. First we will prove statement (2). Define $\rho = \|\xi\|$ to be the euclidean norm of the entire $nK \times 1$ vector ξ. We need to prove that there will not be a sequence of solutions along a path where $\rho \to \infty$. We will show this by contradiction. Consider

a path where $\rho \to \infty$. Choose the component j of the homotopy system for which $\rho_j^q \cos(q\,\varphi_j) \to \infty$ at the fastest rate among all the possible indexes j where $\rho_j \to \infty$. [8]

Consider the real part of the homotopy function, $H_{1j}(\cdot, \cdot, \cdot)$. The equation $H_{1j}(x, y, \tau) = 0$ is equivalent to the equation $\frac{H_{1j}(x,y,\tau)}{\tau\left(\rho_j^q \cos(q\,\varphi_j)-1\right)} = 0$ for $\rho_j > 1$. The last equation can be rewritten as:

$$1 + \frac{(1-\tau)}{\tau\left(\rho_j^q \cos(q\,\varphi_j)-1\right)} \left\{\Lambda^j x - \frac{e^{2x_j} + e^{x_j}\cos(y_j) + \sum_{k \neq j} e^{x_j + x_k}\cos(y_j - y_k)}{1 + \Theta_i(x,y)}\right\} = 0. \qquad (65)$$

We will show that the second term in the curly bracket of the previous equation is uniformly bounded from above in absolute terms:

$$\left|\frac{e^{2x_j} + e^{x_j}\cos(y_j) + \sum_{k \neq j} e^{x_j + x_k}\cos(y_j - y_k)}{1 + \Theta_i(x,y)}\right| \leq C \quad \text{and} \quad \text{for a constant } C, \qquad (66)$$

where the constant C can depend on ϵ. Therefore the term in the curly bracket in the homotopy (65) will grow at most at a linear rate $|x| \leq C\rho_j$. On the other hand, denominator $\tau\left(\rho_j^q \cos\left(q\varphi_j\right)-1\right)$ outside the curly bracket grows at a much faster polynomial rate for large q. Hence the second term in (65) is close to 0 for large q for large values of ξ, and equation (65) can not have a sequence of solutions that tends to infinity.

In other words, there exists $R_0 > 0$ such that for any $\xi = (x, y)$ outside $\wp_\epsilon$ with $\|\xi\| \geq R_0$ and any $\tau \in (0, 1]$ we have that $H_1(x, y, \tau) \neq 0$, that is, homotopy system does not have solutions. This implies that

$$H^{-1}(\tau) \cap \wp_\epsilon \subset B_{R_0}^\tau = \{(x, y, \tau) \in \mathbb{R}^{2nK} \setminus \wp_\epsilon \times (0, 1] \cap \|\xi\| < R_0\}.$$

This proves the second statement 2).

Finally, we will prove both statements 1) and 3) of Theorem 2. Again we consider the above homotopy system on the compact set $B_{R_0}^\tau$. The homotopy function is analytic in this set so Cauchy - Riehmann theorem holds. This implies that

$$\frac{\partial H_{1j}}{\partial x_k} = \frac{\partial H_{2j}}{\partial y_k} \quad \text{and} \quad \frac{\partial H_{1j}}{\partial y_k} = -\frac{\partial H_{2j}}{\partial x_k}, \qquad \text{for all } j, k = 1, \ldots, 2nK.$$

---

[8]In case when instead of $\rho_j^q \cos(q\,\varphi_j) \to \infty$ we have that $\rho_j^q \sin(q\,\varphi_j) \to \infty$, the proof can be appropriately modified by considering the imaginary part of the j-th element of the homotopy system without any further changes. The logic of the proof can be seen to hold as long as there is a slower growing element of x or y. In case when all components of x and y grow at the same rate to infinity in such a way that the second terms inside the curly brackets of (63) and (64) explode to infinity, one can take a Laurent expansion around the values of y_k's such that the denominators are close to zero. Then one can see that these terms in (63) and (64) explode to infinity at quadratic and linear rates in $1/(y - y^*)$, respectively. Therefore (63) and (64) can not both be zero simultaneously for large x and y.

46

This means that if the Jacobian is considered:

$$J_i = \begin{pmatrix} \frac{\partial H_{1j}}{\partial x'} & \frac{\partial H_{1j}}{\partial y'} & \frac{\partial H_{1j}}{\partial \tau} \\ \frac{\partial H_{2j}}{\partial x'} & \frac{\partial H_{2j}}{\partial y'} & \frac{\partial H_{2j}}{\partial \tau} \end{pmatrix},$$

then it contains at least one 2×2 submatrix with nonegative determinant $\left[\frac{\partial H_{1j}}{\partial x_k}\right]^2 + \left[\frac{\partial H_{1j}}{\partial y_k}\right]^2$. Calculating the derivatives directly due to the fact that $\epsilon < \rho < R_0$ this determinant is strictly positive for all $(x, y, \tau) \in B_{R_0}^\tau$. Therefore, the implicit function theorem verifies that the pair (x, y) can be locally parameterized by τ. Moreover, this representation is locally unique and continuous. This proves the first statement. The same arguments above, which show that the determinant is positively almost everywhere, also immediately implies the third statement. $\square$

Proof of equation (66):

We are to bound the left hand side of equation (66) by a given constant. First of all we can bound the denominator from below by

$$\|1 + \Theta_i(x,y)\| \geq \left\|1 + \sum_{k \in I_i} e^{2x_k}\right\| - \left\|2\sum_{k \in I_i} e^{x_k}\cos(y_k) + \sum_{l \in I_i}\sum_{k \neq l} e^{x_k + x_l}\cos(y_l - y_k)\right\|,$$

as $\|a + b\| \geq \|a\| - \|b\|$. Then we can continue to bound:

$$\|1 + \Theta_i(x,y)\| \geq 1 + \sum_{k \in I_i} e^{2x_k} - 2\sum_{k \in I_i} e^{x_k} - \sum_{l \in I_i}\sum_{k \neq l} e^{x_k + x_l}. \tag{67}$$

The last expression was obtained taking into account the fact that

$$\max_{y_k,\, k \in I_i} \|2\sum_{k \in I_i} e^{x_k}\cos(y_k) + \sum_{l \in I_i}\sum_{k \neq l} e^{x_k + x_l}\cos(y_l - y_k)\|$$

is attained at the point $\cos(y_k) \equiv \cos(y_k - y_l) = 1, \forall k, l \in I_i$.
For the same reason, we can bound the numerator from above by

$$\left\|e^{2x_j} + e^{x_j}\cos(y_j) + \sum_{k \neq j} e^{x_j + x_k}\cos(y_j - y_k)\right\| \leq e^{2x_j} + \|e^{x_j}\cos(y_j)$$
$$+ \sum_{k \neq j} e^{x_j + x_k}\cos(y_j - y_k)\| \leq e^{2x_j} + e^{x_j} + \sum_{k \neq j} e^{x_j + x_k}.$$

Recall that j - th component was assumed to be the fastest growing x component as $\rho \to \infty$. Then from equation (67) for some small but positive constant ψ we can write:

$$\|1 + \Theta_i(x,y)\| \geq 1 + \psi e^{2x_j}$$

Collecting terms we have that:

$$\frac{\left\|e^{2x_j} + e^{x_j}\cos(y_j) + \sum_{k \neq j} e^{x_j + x_k}\cos(y_j - y_k)\right\|}{\|1 + \Theta_i(x,y)\|} \leq \frac{1 + e^{2x_j} + e^{x_j} + \sum_{k \neq j} e^{x_j + x_k}}{1 + \psi e^{2x_j}}.$$

which is clearly uniformly bounded from above by a large constant.

The same arguments can be used by looking at the imaginary part of the homotopy system when there exists a y_j that converges to infinity at the fastest rate. □

B Proof for Theorem 3

For the clarify of exposition we will first present the proof in the case of two strategies for each player. Then we will discuss the modifications that are needed when there are more than two strategies for each player. Consider the two strategy case first. In this case we can rewrite the homopoty system (63) and (64) as

$$H_{1j}(\xi, \tau) = \{\rho_j^q \cos(q\varphi_j) - 1\}\tau + (1 - \tau)\left\{\Lambda^j x_j - \frac{e^{2x_j} + e^{x_j}\cos(y_j)}{1 + e^{2x_j} + 2e^{x_j}\cos(y_j)}\right\},$$

and

$$H_{2j}(\xi, \tau) = \{\rho_j^q \sin(q\varphi_j) - 1\}\tau + (1 - \tau)\left\{\Lambda^j y_j - \frac{e^{x_j}\sin(y_j)}{1 + e^{2x_i} + 2e^{x_j}\cos(y_j)}\right\}.$$

We need to check the presence of solutions in the small vicinity of the imaginary axis. Now consider positive increments of x_j such that x_j is equal to some small value ϵ. If we linear the above homotopy system around $x_j = 0$, we can approximate them linearly by

$$
\begin{aligned}
H_{1j} &= \tau q \epsilon y_j^{q-1} - \frac{(1-\tau)\epsilon}{2}\frac{1}{1+\cos(y_j)} - \frac{1+\tau}{2} + \lambda_{jj}(1-\tau)\epsilon + \sum_{k \neq j}\lambda_{jk}x_k(1-\tau) \\
H_{2j} &= \tau y_j^q + (1-\tau)\sum_k \lambda_{jk}y_k - \frac{1-\tau}{2}\frac{sin(y_j)}{1+cos(y_j)} - \tau
\end{aligned}
\tag{68}
$$

where λ_{jj} is the j,jth element of the Λ matrix.

One can see that these two functions are continuous everywhere except for the set of points $\{y_j = \pi + 2\pi k, \ k \in \mathbb{Z}\}$ where $cos(y_j) = -1$.

We will prove that for appropriate large values of q this system has no solutions in the vicinity of this set. First of all note that if we take a second order expansion of $1 + cos(y_j)$ around some $y_j^* = \pi + 2\pi k$ we can approximate $1 + cos(y_j) \approx \frac{1}{2}(y_j - y_j^*)^2$. Then we can further linearize these two equations in (68) to:

$$
\begin{aligned}
H_{1j} &= \tau q \epsilon y_j^{*\ q-1} - \lambda_{jj}(1-\tau)\epsilon + \sum_{k \neq j}\lambda_{kj}x_k(1-\tau) - \frac{1+\tau}{2} - \frac{(1-\tau)\epsilon}{(y_j - y_j^*)^2} \\
H_{2j} &= \tau y_j^{*\ q} + (1-\tau)\lambda_{jj}y_j^* - \sum_{k \neq j}\lambda_{kj}y_k(1-\tau) - \tau + (1-\tau)\frac{1}{(y_i - y_i^*)}
\end{aligned}
\tag{69}
$$

48

where we have also used $sin(y_j) \approx - \left(y_j - y_j^* \right)$.

Now we can construct a sequence of homotopies with the order q increasing to infinity at appropriate rate such that these homotopies do not have solutions with extraneous solution of $|y_j| \to \infty$. This sequence of q is constructed by letting $q = 1 + 1/\epsilon$, as $\epsilon \to 0$. Along this sequence, we will see below that the solutions $y_j - y_j^*$ to H_{1j} and H_{2j} will be of different orders of magnitude. Therefore there can not solutions $y_j - y_j^*$ that simultaneously satisfy both equations $H_{1j} = 0$ and $H_{2j} = 0$.

To see this, consider the first part $H_{1j} = 0$ of (69). For small ϵ only the first term $\tau q \epsilon y_j^{* \, q-1} = O\left(y_j^{* \frac{1}{\epsilon}} \right)$ and the last term $\frac{(1-\tau)\epsilon}{(y_j - y_j^*)^2}$ dominate. Therefore the solution $y_j - y_j^*$ has to have the order of magnitude $O\left(\sqrt{\frac{1}{\epsilon}} y_j^{* - \frac{1}{2\epsilon}} \right)$. On the other hand, for the second part $H_{2j} = 0$ of (69). For small ϵ only the first term $\tau y_j^{* \, q} = O\left(y_j^{* \frac{1}{\epsilon}} \right)$ and the last term $(1 - \tau) \frac{1}{(y_i - y_i^*)}$ dominate. Therefore the solution $y_j - y_j^*$ has to have the order of magnitude $O\left(y_j^{* - \frac{1}{\epsilon}} \right)$ which increases to ∞ much slower than $O\left(\sqrt{\frac{1}{\epsilon}} y_j^{* - \frac{1}{2\epsilon}} \right)$ as $\epsilon \to \infty$. Therefore there can be no solution y_j to both H_{1j} and H_{2j} simultaneously for the sequence of q chosen above. This proves that the homotopy is path finite along that sequence of q.

The considered homotopy function is analytic outside the balls of fixed radius around the members of countable set of points $\{x_j = 0, y_j = \pi + 2\pi k\}$, $k \in \mathbb{Z}$ [9]. Therefore a monotone smooth parametrization is available except for the interior of these balls because the determinant of the Jacobian is strictly positive everywhere else.

This establishes regularity of the homotopy and conludes the proof. $\square$

Proof of equation (68): We consider each term individually. First of all

$$\varphi = \arctan(y/\epsilon) = \frac{\pi}{2} - \arctan(\epsilon/y) \approx \frac{\pi}{2} - \frac{\epsilon}{y}.$$

Hence, as long as q is chosen so that $q\pi/2$ is $2k\pi + \frac{\pi}{2}$ for some k,

$$cos(q\varphi) = cos(q \arctan(y/\epsilon)) \approx cos\left(q\frac{\pi}{2} - q\frac{\epsilon}{y} \right) = sin\left(\frac{q\epsilon}{y} \right) \approx \frac{q\epsilon}{y}.$$

Together with $\rho^q \approx y_j^q$, this gives the first term in H_{1j}.

Secondly, a first order expansion around $\epsilon = 0$ gives

$$\frac{e^{x_j} \sin(y_j)}{1 + e^{2x_i} + 2e^{x_j} \cos(y_j)} \approx \frac{1}{2} + \frac{1}{2}\epsilon \frac{1}{1 + cos(y_j)}.$$

[9] Moreover, it is possible to check that the homotopy system has no solutions when all arguments are purely imaginary in case if q is an arbitrary odd number

Therefore the H_{1j} is proved in (68).

The second part of H_{2j} follows similarly, noting that given the choice of q where $sin(q\varphi) = 1$, and $\rho_j^q \approx y_j^q$, and the first taylor expansion term for $\frac{e^{x_j} \sin(y_j)}{1+e^{2x_i}+2e^{x_j} \cos(y_j)}$ vanishes.

End of proof for equation (68).

C   Semiparametric Variance

To derive Ω, we need to follow [5] and derive the asymptotic linear influence function of the left hand side of the above relation. For this purpose, note that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} A(s_t) \left(y_t - \sigma\left(s_t, \hat{\Phi}, \theta_0\right) \right)$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} A(s_t) (y_t - \sigma(s_t, \Phi_0, \theta_0)) - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} A(s_t) \left(\sigma\left(s_t, \hat{\Phi}, \theta_0\right) - \sigma(s_t, \Phi_0, \theta_0) \right).$$

Since $\hat{\Phi}$ depends only on the nonparametric estimates of choice probabilities $\hat{\sigma}_j(k|s)$, $j = 1, \ldots, n, k = 1, \ldots, K$ in (29) through (24), the second part can also be written as

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} A(s_t) \left(\Gamma(s_t, \theta_0; \hat{\sigma}(s)) - \Gamma(s_t, \theta_0; \sigma_0(s)) \right),$$

where $\hat{\sigma}(s)$ is the collection of all $\hat{\sigma}_j(k|s)$ for $j = 1, \ldots, n$ and $k = 1, \ldots, K$, and the function $\Gamma(\cdot)$ is defined in (28). Then using the semiparametric influence function representation of [6], as long as $\Gamma(s_t, \theta, \sigma(s))$ is sufficiently smooth in $\sigma(s)$ and as long as the nonparametric first stage estimates satisfy certain regularity conditions regarding the choice of the smoothing parameters, we can write this second part as

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} A(s_t) \left(\Gamma(s_t, \theta_0; \hat{\sigma}(s)) - \Gamma(s_t, \theta_0; \sigma_0(s)) \right)$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} A(s_t) \frac{\partial}{\partial \sigma} \Gamma(s_t, \theta_0; \sigma_0(s)) (y_t - \sigma(s_t, \theta_0)) + o_p(1).$$

In other words, if we write $\Gamma_\sigma(s) = \frac{\partial}{\partial \sigma} \Gamma(s_t, \theta_0; \sigma_0(s))$, we can write

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} A(s_t) \left(y_t - \sigma\left(s_t, \hat{\Phi}, \theta_0\right) \right)$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} A(s_t) (I - \Gamma_\sigma(s_t)) (y_t - \sigma(s_t, \theta_0)) + o_p(1).$$

Therefore we can derive the asymptotic distribution of the two-step semiparametric $\hat{\theta}$ defined through (33) as

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) = -\left(EA\left(s_t\right)\Gamma_\theta\left(s_t\right)\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} A\left(s_t\right)\left(I - \Gamma_\sigma\left(s_t\right)\right)\left(y_t - \sigma\left(s_t, \theta_0\right)\right) + o_p\left(1\right).$$

Hence

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \Sigma\right)$$

where Σ is equal to

$$E\left(A\left(s_t\right)\Gamma_\theta\left(s_t\right)\right)^{-1} \left[EA\left(s_t\right)\left(I - \Gamma_\sigma\left(s_t\right)\right)\Omega\left(s_t\right)\left(I - \Gamma_\sigma\left(s_t\right)\right)' A\left(s_t\right)'\right] E\left(\Gamma_\theta\left(s_t\right)' A\left(s_t\right)'\right)^{-1}.$$

In the above, we have defined

$$\Gamma_\theta\left(s_t\right) = \left.\frac{\partial}{\partial\theta_1}\Gamma\left(s_t, \theta_1, \sigma\left(s_t; \theta_2\right)\right)\right|_{\theta_1 = \theta_2 = \theta_0},$$

and

$$\Omega\left(s_t\right) = Var\left(y_t - \sigma\left(s_t, \theta_0\right)|s_t\right).$$

The efficient choice of the instrument matrix (which can be feasibly estimated in preliminary steps without affecting the asymptotic variance) is then given by

$$A\left(s_t\right) = \Gamma_\theta\left(s_t\right)'\left(I - \Gamma_\sigma\left(s_t\right)\right)^{-1}\Omega\left(s_t\right)^{-1}\left(I - \Gamma_\sigma\left(s_t\right)\right)^{-1'}.$$

With this efficient choice of the instrument matrix, the asymptotic variance of $\hat{\theta}$ becomes

$$\left(E\Gamma_\theta\left(s_t\right)'\left(I - \Gamma_\sigma\left(s_t\right)\right)^{-1}\Omega\left(s_t\right)^{-1}\left(I - \Gamma_\sigma\left(s_t\right)\right)^{-1'}\Gamma_\theta\left(s_t\right)\right)^{-1}. \tag{70}$$

C.1 Efficiency Considerations

We present two efficiency results in this section. First of all, we show that with the above efficient choice of the instrument matrix $A\left(s_t\right)$, the semiparametric two step estimation procedure above is as efficient as the full maximum likelihood estimator where the fixed point mapping in (28) is solved for every parameter value θ which is then nested inside maximum likelihood optimization to obtain choice probabilities as a function of θ. Secondly, we show that estimating $\hat{\sigma}\left(s_t\right)$ may even improves efficiency over the hypothesis case where $\sigma\left(s_t\right)$ is known and the pseudo MLE in (34) is used to estimate θ (but using Φ_0 instead of $\hat{\Phi}$).

C.1.1 Efficiency comparison with full maximum likelihood

Consider a full maximum likelihood approach where a fixed point calculation (assuming the solution is unique) of (28) is nested inside the numerical optimization. For each θ, (28) is solved to obtain $\sigma(s_t, \theta)$ as a function of θ, which is then used to form the likelihood function as in (34). Define the total derivative of (28) as

$$\frac{d}{d\theta}\sigma(s_t, \theta_0) = \frac{d}{d\theta}\Gamma(s_t, \theta, \sigma(s_t; \theta))\bigg|_{\theta=\theta_0}$$

$$= \Gamma_\theta(s_t) + \Gamma_\sigma(s_t)\frac{d}{d\theta}\sigma(s_t, \theta_0)$$

which can be used to solve for

$$\frac{d}{d\theta}\sigma(s_t, \theta_0) = (I - \Gamma_\sigma(s_t))^{-1}\Gamma_\theta(s_t). \tag{71}$$

Following the same logic as the discussions of pseudo MLE after (34), but with the pseudo log likelihood function replaced by the full maximum likelihood function, it is easy to show that the asymptotic distribution of the full maximum likelihood estimator, which is the same as an iv estimator with the instruments chosen optimally, satisfies

$$\sqrt{T}\left(\hat{\theta}_{FMLE} - \theta_0\right) \xrightarrow{d} N(0, \Sigma_{FMLE})$$

where

$$\Sigma_{FMLE} = \left(E\frac{d}{d\theta}\sigma(s_t, \theta_0)'\,\Omega(s)^{-1}\,\frac{d}{d\theta}\sigma(s_t, \theta_0)'\right)^{-1}.$$

Using (71), we can also write

$$\Sigma_{FMLE} = \left[E\Gamma_\theta(s_t)'(I - \Gamma_\sigma(s_t))^{-1}\Omega(s_t)^{-1}(I - \Gamma_\sigma(s_t))^{-1}\Gamma_\theta(s_t)\right]^{-1}.$$

This is identical to (70) for the asymptotic variance of the two step semiparametric iv estimator when the instrument matrix is chosen optimally.

C.1.2 Efficiency comparison with infeasible pseudo MLE

Consider an infeasible pseudo MLE, which is similar to (34) except with $\hat{\Phi}$ replaced by the true but unknown Φ_0:

$$\sum_{t=1}^{T}\sum_{i=1}^{n}\left[\sum_{k=1}^{K}y_{ikt}\log\sigma_i(k|s_t, \Phi_0, \theta) + \left(1 - \sum_{k=1}^{K}y_{ikt}\right)\log\left(1 - \sum_{k=1}^{K}\sigma_i(k|s_t, \Phi_0, \theta)\right)\right]. \tag{72}$$

The asymptotic variance of this estimator is similar to that of Σ_{FMLE} except with $\frac{d}{d\theta}\sigma\left(s_t,\theta_0\right)'$ replaced by $\Gamma_\theta\left(s_t\right)$. In other words,

$$\Sigma_{IPMLE} = \left[E\Gamma_\theta\left(s_t\right)'\Omega\left(s_t\right)^{-1}\Gamma_\theta\left(s_t\right)\right]^{-1}.$$

where $IPMLE$ stands for infeasible pseudo MLE.

The relation between Σ_{FMLE} and Σ_{IPMLE} is obviously ambiguous and depends on the response matrix $\Gamma_\sigma\left(s_t\right)$. It is clear possible that $\Sigma_{FMLE} < \Sigma_{IPMLE}$, in which case estimating $\hat{\Phi}$ may improve efficiency over the case where Φ_0 is known.

D Preliminary results for multiple equilibria

Procedure setup.

In this example we compute equilibria in the example of the game where choice - specific probabilities are described by the ordered logit model. We consider a case with 2 players in the asymmetric setup so that, for instance, the choice probabilities for player 1 can be described in terms of choice probabilities of player 2 and a vector of covariates x_1 in the following way:

$$
\begin{aligned}
P_{11} &= \Lambda\left(\mu_1 - x_1'\beta - \eta\sum_{i=1}^{5} i\,P_{2i}\right) \\
P_{12} &= \Lambda\left(\mu_2 - x_1'\beta - \eta\sum_{i=1}^{5} i\,P_{2i}\right) - \Lambda\left(\mu_1 - x_1'\beta - \eta\sum_{i=1}^{5} i\,P_{2i}\right) \\
P_{13} &= \Lambda\left(\mu_3 - x_1'\beta - \eta\sum_{i=1}^{5} i\,P_{2i}\right) - \Lambda\left(\mu_2 - x_1'\beta - \eta\sum_{i=1}^{5} i\,P_{2i}\right) \\
P_{14} &= \Lambda\left(\mu_4 - x_1'\beta - \eta\sum_{i=1}^{5} i\,P_{2i}\right) - \Lambda\left(\mu_3 - x_1'\beta - \eta\sum_{i=1}^{5} i\,P_{2i}\right) \\
P_{j5} &= 1 - \sum_{k=1}^{4} P_{jk}, \quad \text{for } j = 1, 2.
\end{aligned}
\tag{73}
$$

where $\Lambda(\cdot)$ is the logistic function, the first index stands for the number of player and the second index stands for the recommendation strategy.

To calculate the equilibria in the game where ex-ante probabilities of choices are described by the ordered logit model, we take the cutoff points for the ordered logit procedure which are reported as (2.987521, 4.779925, 7.533645, 8.800422) and use them as values μ_1, μ_2, μ_3 and μ_4 to substitute for the cutoff values in ordered logit in the formula for the ordered estimated in the empirical section of the paper. We use the results of the regression with quarterly and stock dummies.

According to the setup of the ordered logit model in or case I put stock dummy to its average over periods as well as $\%DEV$ variable. This adds to all the cutoff points. I use 2 quarter; one is quarter

Table 2: Adjusted cutoff points for choice probabilities.

	$RELATION = 0$	$RELATION = 1$
Quarter 9	(3.239, 5.0323, 7.7860, 9.0528)	(3.2242, 5.0166, 7.7703, 9.0371)
Quarter 21	(2.9655, 4.7579, 7.5116, 8.7784)	(2.9499, 4.7423, 7.4960, 8.7628)

9 when NASDAQ is 4,732 which shifts cutoff points by -0.15528; the other quarter is quarter 21 when NASDAQ is 1,374, which shifts the cutoff points by +0.11905. I consider a case with 2 bidders. For the $RELATION$ dummy we consider 2 cases: one is when it is 0 for both players and the other when it is 1 for both players. Asymmetric case shifts the cutoff points by .0156 for one player and symmetric case when both players have $RELATION = 1$ shifts cutoff points by .0156 for both players. The adjusted cutoff points for choice probabilities are formed in the following way. We take values of %DEV variable at its mean and compute $\beta_{\%DEV} \cdot \text{mean}(\%DEV) = .048048.0022158 = .000106$, then we compute mean(stock dummy)=-0.09711. We consider then cases with different values of quarter dummies and $RELATION$ variable. Namely, we take quarter 9 dummy= -0.15528 quarter 21 dummy=0.11905. The total value of $x'\beta$ is: -0.2523 for quarter 9 with $RELATION = 0$, -0.2367 for quarter 9 with $RELATION = 1$, 0.0219 for quarter 21 with $RELATION = 0$, 0.0375 for quarter 21 with $RELATION = 1$. The set of corresponding cutoff values:

We use then the estimated values of η coefficient (fixing it equal to 1.96) to solve the obtained system of equations for choice probabilities.

Results.

We used HOMPACK package to solve for multiple equilibria in the considered game, namely its part originally designed to find all roots of the polynomial systems. To do that we modified the function routine to represent the system, characterizing the ordered logit. The initial system in the homotopy was wet equal to 80,000. This means that to solve for equilibria the algorithm computed 80,000 paths for the solutions of a system of 16 differential equations formed by real and complex parts of the choice probability functions for 2 players (so that each player has 4 choice probabilities and one choice probability id complimentary).

The following results are preliminary as in theory the degree of initial system should go to infinity to make it possible to isolate all real solutions to the system of choice probabilities[10], so there is a

---

[10]For example, in our Monte-Carlo experiments for the entry game a characteristic degree of the initial system was 150,000. A characteristic CPU time to solve for all equilibria in a single game for the 5 player

Table 3: Computed choice probabilities for choices 1-4.	
$RELATION = 0$	(0.74876E-04; 0.37660E-03; 0.66071E-02; 0.17543E-01)
Quarter 9	(0.21574E-01; 0.95336E-01; 0.55826E-0; 0.20545E-0)
$RELATION = 1$	(0.16503E-03; 0.82499E-03; 0.14331E-01; 0.37016E-01)
Quarter 9	(0.21124E-01; 0.93580E-01; 0.55573E-0; 0.207916E-0)
$RELATION = 0$	(0.12220E-03; 0.61102E-03; 0.10656E-01; 0.27896E-01)
Quarter 21	
$RELATION = 1$	(0.12005E-03; 0.60027E-03; 0.10471E-01; 0.27429E-01)
Quarter 21	

need in the further increase in the power of the initial system. In the table below we represent the computation results for the game with symmetric players. No asymmetric equilibria were found, so we report only the choice probabilities for one player as for the other player they are the same.

As one can see from the table D, 2 equilibria were found for the choice probabilities in quarter 9, while only one was found for quarter 21.

---

entry game was about 4 hours. Taking into account that the dimensionality in the considered ordered logit game is 8 for the 2-player case, the characteristic CPU time for the ordered logit case is 8 hours.