Identification and Estimation of Games with Incomplete Information Using Excluded Regressors[1]

Arthur Lewbel, Boston College[2]

Xun Tang, Rice University and University of Pennsylvania[3]

Original: November 2010. Current version: November 2014

Abstract

We show structural components in binary games with incomplete information are nonparametrically identified using variation in player-specific excluded regressors. An excluded regressor for a player i is a sufficiently varying state variable that does not affect other players' utility and is additively separable from other components in i's payoff. Such excluded regressors arise in various empirical contexts. Our identification method is constructive, and provides a basis for nonparametric estimators. For a semiparametric model with linear payoffs, we propose root-N consistent and asymptotically normal estimators for players' payoffs. We also discuss extension to the case with multiple Bayesian Nash equilibria in the data-generating process without assuming equilibrium selection rules.

Keywords: *Games with Incomplete Information, Excluded Regressors, Nonparametric Identification, Semiparametric Estimation, Multiple Equilibria.*

JEL Codes: C14, C51, D43

[1] We thank Arie Beresteanu, Aureo de Paula, Stefan Hoderlein, Zhongjian Lin, Pedro Mira, Jean-Marc Robin, Frank Schorfheide, Steven Stern and Kenneth Wolpin for helpful discussions. We also thank the editor and two anonymous referees and seminar participants at Asian Econometric Society Summer Meeting 2011, Brown U, Duke U, London School of Economics, Rice U, Penn State, U British Columbia, U Connecticut, U Pennsylvania, U Pittsburgh and U Rochester for feedbacks and comments. All errors are our own.

[2] Corresponding author. Address: Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467. Email: lewbel@bc.edu

[3] Address: Department of Economics MS-22, Rice University, P.O.Box 1892, Houston, TX 77251-1892. Email: xt9@rice.edu

1 Introduction

We show identification of, and provide estimators for, a class of static game models where players simultaneously choose between binary actions based on imperfect information about each other's payoffs. Such models are widely applied in industrial organization. Examples include firms' entry decisions in Seim (2006), decisions on the timing of commercials by radio stations in Sweeting (2009), firms' choices of capital investment strategies in Aradillas-Lopez (2010) and interactions between equity analysts in Bajari, Hong, Krainer and Nekipelov (2010).

• **Preview of Our Method**

We introduce a new approach to identify and estimate such models by assuming players' payoffs depend on a vector of "excluded regressors." An excluded regressor associated with a given player i is a state variable that does not enter payoffs of the other players, and is additively separable from other components in i's payoff. We show that if excluded regressors are orthogonal to players' private information given other observed states, then interaction effects between players and marginal effects of excluded regressors on players' payoffs can be nonparametrically identified. If in addition the excluded regressors vary sufficiently relative to the support of private information, then the distribution of payoffs for all players is also nonparametrically identified. No distributional assumption on the private information is necessary for these results. Our identification proofs are constructive, so consistent nonparametric estimators can be readily based on them.

We provide several examples of economic models where excluded regressors arise. For example, consider an entry game where profit maximizing firms decide simultaneously whether to enter a market or not, knowing that after entry they will compete through their choices of quantities. The fixed cost of entry for a firm i does not affect its perception of the interaction effects with other firms, or the difference between its monopoly profits (when i is the sole entrant) and its oligopoly profits (when i is one of the multiple entrants). This is precisely the independence restriction required of excluded regressors.

The intuition for this result is that fixed costs drop out of the first-order conditions in the profit maximization problems both for a monopolist and for oligopolists engaging in Cournot competition. As a result, while fixed costs affect the decision to enter, they do not otherwise affect the quantities supplied by any of the players. In general, only some components of a firm's fixed cost are likely to be observable by econometricians, while the remaining, unobserved part of the fixed cost is private information known only to the firm. Our method can be applied in this case, with the observed component of fixed costs playing the role of excluded regressors. Other excluded regressor conditions that contribute to identification (such as conditional independence from private information and sufficient variation relative to the conditional support of private information given observed states) also have economic interpretations in the context of entry games. We discuss those interpretations in greater details in Section 4.

We also discuss how to extend our approach to obtain point identification when there are multiple Bayesian Nash equilibria (BNE) in the data generating process (DGP), without imposing restrictions on the distribution of private information or assuming knowledge of the equilibrium selection mechanism. In this case, identification can be obtained as before, but only using variation of excluded regressors within a set of states where the choices observed in the data are rationalized by a single BNE. We build on De Paula and Tang (2012), showing that if players' private information is independent conditional on observed states then the players' actions are correlated given these states if and only if choices observed under those states are rationalized by multiple BNE. Thus one can infer from the data whether a given set of states have a single BNE, and base identification only on those such states.

• **Relation to Existing Literature**

This paper contributes to the existing literature on structural estimation of Bayesian games in several ways. First, we identify the complete structure of the model under a set of mild nonparametric restrictions on primitives. Bajari, Hong, Nekipelov and Krainer (2010) identify the players' payoffs in a general model of multinomial choices, without assuming any particular functional form for payoffs, but they require the researcher to have full knowledge of the distribution of private information. In comparison, we allow the distribution of private information to be unknown and nonparametric. We instead obtain identification by imposing economically motivated exclusion restrictions.

Sweeting (2009) proposes a maximum-likelihood estimator where players' payoffs are linear indices and the distribution of players' private information is known to belong to a parametric family. Aradillas-Lopez (2010) estimates a model with linear payoffs where players' private information is independent of states observed in the data. He also proposes a corresponding root-N consistent estimator by extending the semi-parametric estimator for single-agent decision models in Klein and Spady (1993) to the game-theoretic setup. Tang (2010) shows identification of a similar model to Aradillas-Lopez (2010), with the independence between private information and observed states replaced by a weaker median independence assumption. He proposes consistent estimators for linear coefficients without establishing rates of convergence. Wan and Xu (2014) consider a class of Bayesian games with index utilities under median restrictions. They allow players' private information to be positively correlated in a particular way, and focus on a class of monotone, threshold crossing pure-strategy equilibria. Unlike these papers, we show non-parametric identification of players' payoffs, with no functional forms imposed. We also allow private information to be correlated with non-excluded regressors in unrestricted ways, though we do require them to be independent across players conditional on observable states.

The second contribution of our paper is to propose consistent and asymptotically normal estimators for semi-parametric models with linear payoffs and heteroskedastic private information. If in addition the support of private information and excluded regressors are bounded, our estimators for the linear coefficients in baseline payoffs attain the parametric rate. If these supports are unbounded, then the rate of convergence of estimators for these coefficients can be parametric or slower, depending on properties of the tails of the distribution of the excluded regressors. The estimators for interaction effects converge at the parametric rate regardless of the boundedness of these supports.[4]

When multiple BNE exist in the data, a player's choice probability is an unknown mixture of the choice probabilities implied under each single equilibrium. Hence, in that case the structural link between the distribution of actions and the model primitives is unknown. The existing literature offers several solutions to this multiple BNE issue. These include parameterizing the equilibrium selection mechanism; maintaining that only a single BNE is followed by the players (Bajari, Hong, Krainer and Nekipelov (2010)); introducing sufficient restrictions on primitives to ensure that the system characterizing the equilibria have a unique solution (Aradillas-Lopez (2010)); or obtaining partial identification results (Beresteanu, Molchanov, Molinari (2011)).[5] For a model of Bayesian games with linear payoffs and correlated bivariate normal private information, Xu (2014) addressed the multiplicity issue by focusing on states for which the model only admits a unique, monotone equilibrium. Xu (2014) showed that with these parametric assumptions, a subset of these states can be inferred using the observed distribution of choices, which can then be used for estimating parameters in payoffs. While different, as explained above our approach for dealing with multiple BNE is based on a similar idea.

Our use of excluded regressors is related to the use of “special regressors" in single-agent, qualitative response models (See Dong and Lewbel (2014) and Lewbel (2014) for surveys of special regressor estimators). Lewbel (1998, 2000) studied nonparametric identification and estimation of transformation, binary, ordered, and multinomial choice models using a special regressor that is additively separable from other components in decision-makers' payoffs, is independent from unobserved heterogeneity conditional on other covariates, and has a large support. Magnac and Maurin (2008) study partial identification of the model when the special regressor is discrete or measured within intervals. Magnac and Maurin (2007) remove the large support requirement on the special regressor in such a model, and replace it with an alternative tail symmetry restriction on the distribution of latent utility. Lewbel, Linton, and McFadden (2011) estimate features of

[4]Khan and Nekipelov (2011) studied the Fisher information for the interaction parameter in a similar model of Bayesian games, where private information is correlated across players, but the baseline payoff functions are assumed known. Unlike our paper, the goal of their paper is not to jointly identify and estimate the full structure of the model. Rather, they focus on showing that the Fisher information of the interaction parameter is positive while assuming the functional form of the rest of the payoffs as well as the equilibrium selection mechanism are known.

[5]Aguirregabiria and Mira (2005) proposed an estimator for payoff parameters in games with multiple equilibria where identification is assumed. Their estimator combines a pseudo-maximum likelihood procedure with an algorithm that searches globally over the space of possible combinations of multiple equilibria in the data.

willingness-to-pay models with a special regressor constructed by experimental design. Berry and Haile (2009, 2010) extend the use of special regressors to nonparametrically identify the market demand for differentiated goods.

Special regressors and exclusion restrictions have also been used for estimating social interaction models and games with complete information. Brock and Durlauf (2007) and Blume, Brock, Durlauf and Ioannides (2010) used exclusion restrictions of instruments to identify a linear model of social interactions. Bajari, Hong and Ryan (2010) used exclusion restrictions and identification at infinity, while Tamer (2003) and Ciliberto and Tamer (2009) used some special regressors to identify games with complete information. Our identification differs fundamentally from that of complete information games because in our context of incomplete information, each player-specific excluded regressor enters the strategies of all other players' through its impact on the equilibrium choice probabilities. In a similar context of Quantal Response Models, Haile, Hortaçsu and Kosenok (2008) suggested that, if each player's choice probability is invariant to others' baseline payoffs, then one can derive some testable restrictions for the model.

• **Roadmap**

The rest of the paper is organized as follows. Section 2 introduces our model of discrete Bayesian games. Section 3 establishes nonparametric identification of the model. Section 4 motivates our key identifying assumptions, by providing three different examples of games in which economic behavior gives rise to the presence of observable excluded regressors. Sections 5 proposes semiparametric estimators for a model with linear payoffs. Section 6 provides some evidence of the finite sample performance of this estimator. Section 7 extends our identification method to models with multiple equilibria, including testing whether observed choices are rationalizable by a single BNE over a subset of states in the data-generating process. Section 8 concludes. Some proofs and additional discussions are provided in an online supplement to this paper.

2 The Model

Consider a simultaneous game between N players making binary actions under incomplete information. Let N denote both the set and the number of players. Each player i chooses $D_i \in \{1, 0\}$. A vector of states $X \in \mathbb{R}^J$ with $J > N$ is observed by all players and the econometrician. Let $\epsilon^* \equiv (\epsilon_i^*)_{i=1}^N \in \mathbb{R}^N$ denote players' private information, where ϵ_i^* is observed by i but not anyone else. Throughout the paper, we will use upper cases (e.g. X, ϵ_i^*) for random vectors and lower cases (e.g. x, ε_i^*) for their realizations. For any pair of random vectors R and R', let $\Omega_{R'|R}$, $f_{R'|R}$ and $F_{R'|R}$ denote the support, the density and the distribution of R' conditional on R respectively. We use $\Omega_{R'|R=r}$, $F_{R'|R=r}$ and $f_{R'|R=r}$ (or simply $\Omega_{R'|r}$, $F_{R'|r}$ and $f_{R'|r}$) to denote the conditional support, distribution and density of R' given $R = r$ respectively.

The payoff for i from choosing action 1 is $[u_i^*(X,\epsilon_i^*) + h_i(D_{-i})\delta_i^*(X)]\, D_i$ where $D_{-i} \equiv \{D_j\}_{j\neq i}$ and $h_i(.)$ is a bounded function known to all players and the econometrician and satisfies $h_i(\mathbf{0}) = 0$. When $N = 2$, we can set $h_i(D_j) \equiv D_j$ without loss of generality.[6] Payoffs from choosing action 0 are state-independent and thus normalized to 0 for all players. Here $u_i^*(X,\epsilon_i^*)$ denotes the baseline payoff for i from choosing 1 when all others choose 0, while $h_i(D_{-i})\delta_i^*(X)$ captures the interaction effect, or how others' actions affect i's payoff.

For example, Sweeting (2009) and Aradillas-Lopez (2010) maintained $h_i(D_{-i}) \equiv \sum_{j\neq i} D_j$, thus making the interaction effect proportional to the number of competitors choosing 1. Other possible examples of $h_i(D_{-i})$ include $\min_{j\neq i} D_{-i}$ or $\max_{j\neq i} D_{-i}$. We require the interaction effect $h_i(D_{-i})\delta_i^*(X)$ to be multiplicatively separable in X and D_{-i}, but δ_i^* and u_i^* are allowed to depend on X. The model primitives $\{u_i^*,\delta_i^*\}_{i=1}^N$ and $F_{\epsilon^*|X}$ are common knowledge among players but are unknown to the econometrician. We maintain the following restrictions on $F_{\epsilon^*|X}$.

A1 *(Conditional Independence) Given any x, ϵ_i^* is independent from $\{\epsilon_j^*\}_{j\neq i}$ for all i, and $F_{\epsilon_i^*|x}$ is absolutely continuous with respect to the Lebesgue measure and has bounded positive Radon-Nikodym density almost everywhere over $\Omega_{\epsilon_i^*|x}$.*

This assumption allows X to be correlated with players' private information. It is commonly maintained in the existing literature on the estimation of Bayesian games. Examples include Aradillas-Lopez (2010), Bajari, Hahn, Hong and Ridder (2010), Bajari, Hong, Krainer, and Nekipelov (2010), Berry and Tamer (2007), Brock and Durlauf (2007), Seim (2006), Sweeting (2009) and Tang (2010).[7] Throughout this paper, we assume players adopt pure strategies only. A pure strategy for i in a game with X is a mapping $s_i(X,.) : \Omega_{\epsilon_i^*|X} \to \{0,1\}$. Let $s_{-i}(X,\epsilon_{-i}^*)$ represent a profile of $N-1$ strategies of i's competitors $\{s_j(X,\epsilon_j^*)\}_{j\neq i}$. Under A1, any Bayes Nash equilibrium (BNE) under states X must be characterized by a profile of strategies $\{s_i(X,.)\}_{i\leq N}$ such that for all i and ϵ_i^*,

$$s_i(X,\epsilon_i^*) = 1\left\{u_i^*(X,\epsilon_i^*) + \delta_i^*(X)E[h_i(s_{-i}(X,\epsilon_{-i}^*)) \mid X,\epsilon_i^*] \geq 0\right\}. \tag{1}$$

The expectation in (1) is conditional on i's information (X,ϵ_i^*) and is taken with respect to private information of his rivals ϵ_{-i}^*. By A1, $E[h_i(D_{-i}) \mid X,\epsilon_i^*]$ must be a function of X alone. Thus in any BNE, the joint distribution of D_{-i} conditional on i's information (x,ε_i^*) and the profile of strategies takes the form

$$\Pr\{D_{-i} = d_{-i} \mid X = x, \epsilon_i^* = \varepsilon_i^*\} = \Pr\left\{s_{-i}(X,\epsilon_{-i}^*) = d_{-i} \mid X = x\right\} = \Pi_{j\neq i}\,(p_j(x))^{d_j}\,(1-p_j(x))^{1-d_j}, \tag{2}$$

[6] With $N \geq 3$, assuming h_i is known is restrictive. Bajari, Hong, Krainer, and Nekipelov (2010) avoid this assumption and allow the payoff matrix to be completely nonparametric. In comparison, they assume the knowledge of the distribution of the unobserved states.

[7] An exception is Wan and Xu (2010), who restricted equilibrium strategies to be monotone with a threshold-crossing property, and required conditions on the magnitude of the strategic interaction parameters.

for all $d_{-i} \in \{0,1\}^{N-1}$, where $p_i(x) \equiv \Pr\{u_i^*(X, \epsilon_i^*) + \delta_i^*(X) E[h_i(s_{-i}(X, \epsilon_{-i}^*)) \mid X] \geq 0 \mid X = x\}$ is i's probability of choosing 1 given x.

For the special case where $h_i(D_{-i}) = \sum_{j \neq i} D_j$ for all i, the profile of BNE strategies in a game with states x can be characterized by a vector of choice probabilities $\{p_i(x)\}_{i \leq N}$ such that:

$$p_i(x) = \Pr\left\{ u_i^*(X, \epsilon_i^*) + \delta_i^*(X) \textstyle\sum_{j \neq i} p_j(X) \geq 0 \middle| \; X = x \right\} \text{ for all } i \in N. \tag{3}$$

Because of the conditional independence in A1, this characterization only restricts a vector of marginal probabilities p_i as opposed to joint probabilities. The existence of pure-strategy BNE given any x follows from an application of the Brouwer's Fixed Point Theorem and the continuity of $F_{\epsilon^*|X}$ in A1.

In general, the model could admit multiple BNE, because the system (3) admits multiple solutions. So in the data-generating process (DGP), players' choices observed could potentially be rationalized by more than one BNE. In Sections 3 and 5, we first identify and estimate the model, assuming players' choices are rationalized by a single BNE for each state x in the data (either because their specific model only admits a unique equilibrium, or because agents employ some degenerate equilibrium selection mechanism). Later in Section 7, we discuss how to extend our use of excluded regressors to identify the model when there are multiple BNE in the DGP. This organization of the paper allows us to separate our main identification strategy from the added complications involved in dealing with multiple equilibria. It also permits direct comparison with existing results that assume a single BNE, such as Bajari, Hong, Krainer and Nekipelov (2010), Aradillas-Lopez (2010) and Wan and Xu (2014).

3 Nonparametric Identification

We consider nonparametric identification assuming a large cross-section of independent games, each involving N players, with the fixed structure $\{u_i^*, \delta_i^*\}_{i \in N}$ and $F_{\epsilon^*|X}$ underlying all games observed in the data.[8] The econometrician observes players' decisions and the states in X, but does not observe $\{\varepsilon_i^*\}_{i \in N}$.

A2 *(Unique Equilibrium) Under each state x in the data, players' choices are rationalized by a single BNE.*

This assumption, which we maintain for now, will be relaxed later. For each state x, A2 could hold either because the model only admits a unique BNE under that state, or because the equilibrium selection mechanism is degenerate at a single BNE under that state. Under A2,

[8] It is not necessary for all games in the data to have the same group of players. Rather, it is only required that the N players in each game in the data have the same set of preferences $(u_i^*, \delta_i^*)_{i \in N}$.

the vector of conditional expectations of $h_i(D_{-i})$ given X (which is directly identifiable from data) satisfies the system in (1)-(2). When $h(D_{-i}) = \sum_{j \neq i} D_j$, A2 ensures that the identifiable conditional choice probabilities of each player, denoted by $\{p_i^*(x)\}_{i \in N}$, solve (3). As such, (3) reveals how the variation in excluded regressors affects the observed choice probabilities. This provides the foundation of our identification strategy.

A3 *(Excluded Regressors) The J-vector of states in X is partitioned into two vectors $X_e \equiv (X_i)_{i \in N}$ and $\tilde{X}$ such that (i) for all i and (x, ε_i^*), $\delta_i^*(x) = \delta_i(\tilde{x})$ and $u_i^*(x, \varepsilon_i^*) = \alpha_i x_i + u_i(\tilde{x}, \varepsilon_i^*)$ for $\alpha_i \in \{-1, 1\}$ and some unknown functions u_i and δ_i, where u_i is continuous in ε_i^* at all $\tilde{x}$; (ii) for all i, ϵ_i^* is independent from X_e given any $\tilde{x}$; and (iii) given any $\tilde{x}$, the distribution of X_e is absolutely continuous w.r.t. the Lebesgue measure and has positive Radon-Nikodym densities a.e. over the support.*

Hereafter we refer to the N-vector $X_e \equiv (X_i)_{i \in N}$ as *excluded regressors*, and the remaining *(J–N)*-vector of regressors $\tilde{X}$ as *non-excluded regressors*. The main restriction on excluded regressors is that they do not appear in the interaction effects $\delta_i^*(x)$; each of them only appears additively in one baseline payoff $u_i^*(x, \varepsilon_i^*)$; and $\{\epsilon_i^*\}_{i \in N}$ is independent from X_e given $\tilde{X}$. In practice, such conditional independence becomes more likely to hold as the dimension of the conditioning variables in $\tilde{X}$ increases. Note that A1 and A3-(ii) together imply $\epsilon_i^* \perp (\{\epsilon_j^*\}_{j \neq i}, X_e)$ conditional on $\tilde{X}$. The scale of each coefficient α_i in the baseline payoff $u_i^*(x, \varepsilon_i^*) = \alpha_i x_i + u_i(\tilde{x}, \varepsilon_i^*)$ is normalized to $|\alpha_i|$ $= 1$. This is the free scale normalization that is available in all binary choice threshold-crossing models.

A3 extends the notion of continuity and conditional independence of a scalar special regressor in single-agent models as in Lewbel (1998, 2000) to the context with strategic interactions between multiple decision-makers. Identification using excluded regressors in X_e in the game-theoretic setup differs qualitatively from the use of special regressors in single-agent cases, because in any equilibrium X_e affects *all* players' decisions through its impact on the joint distribution of actions.[9]

3.1 Overview and intuition

To summarize our identification strategy, consider the special case of $N = 2$ players indexed by $i \in \{1, 2\}$. When we refer to one player as i, the other will be denoted j. For this special case assume $h_i(D_{-i}) = D_j$. Here the vector x consists of the excluded regressors x_1 and x_2 along with the vector of remaining, non-excluded regressors $\tilde{x}$. Let $\mathcal{S}_i \equiv -u_i(\tilde{X}, \epsilon_i^*)$, where $u_i(\tilde{X}, \epsilon_i^*)$ is the part of player i's payoff that does not depend on excluded regressors, and let $F_{\mathcal{S}_i|\tilde{X}=\tilde{x}}$ (or simply

[9] The exclusion restriction in A3 differs from that mentioned in Haile, Hortaçsu and Kosenok (2008) for testing Quantal Response Models. Specifically, A3 allows a player i's choice probability to move together with other players' baseline payoffs through the non-excluded regressors in $\tilde{X}$ as well as $(X_j)_{j \neq i}$.

$F_{\mathcal{S}_i|\tilde{x}}$) denote the conditional distribution of $\mathcal{S}_i$ given $\tilde{X} = \tilde{x}$.

Given A1,2,3, the equation (3) simplifies in this two-player case to

$$p_i^*(x) = F_{\mathcal{S}_i|\tilde{x}}\left[\alpha_i x_i + \delta_i(\tilde{x})p_j^*(x)\right]. \tag{4}$$

Let $p_{i,j}^*(x) \equiv \partial p_i^*(X)/\partial X_j|_{X=x}$. Taking partial derivatives of equation (4) with respect to x_i and x_j gives:

$$p_{i,i}^*(x) = \left[\alpha_i + \delta_i(\tilde{x})p_{j,i}^*(x)\right]\tilde{f}_i\left(x\right) \tag{5}$$

$$p_{i,j}^*(x) = \delta_i(\tilde{x})\tilde{f}_i\left(x\right)p_{j,j}^*(x) \tag{6}$$

where $\tilde{f}_i\left(x\right)$ is the density of $\mathcal{S}_i$ given $\tilde{x}$ evaluated at $\alpha_i x_i + \delta_i(\tilde{x})p_j^*(x)$. Since choices are observable, the choice probabilities $p_i^*(x)$ and their derivatives $p_{i,i}^*(x)$, $p_{i,j}^*(x)$ are identified.

The first goal is to identify the signs of the marginal effects of the excluded regressor α_i (recall that their magnitudes are normalized to one) and the interaction terms $\delta_i(\tilde{x})$. Solve equation (6) for $\delta_i(\tilde{x})$, substitute the result into equation (5), and solve for α_i to obtain

$$\alpha_i\tilde{f}_i\left(x\right) = p_{i,i}^*(x) - p_{i,j}^*(x)p_{j,i}^*(x)/p_{j,j}^*(x). \tag{7}$$

Because $\tilde{f}_i\left(x\right)$ is positive under A1, the sign α_i is over-identified as the sign of the right-hand side of (7), which only consists of marginal effects of excluded regressors on identifiable choice probabilities. Then divide (5) by (6) and over-identify the interaction term as:

$$\delta_i(\tilde{x}) = \alpha_i p_{i,j}^*(x)/\left[p_{i,i}^*(x)p_{j,j}^*(x) - p_{j,i}^*(x)p_{i,j}^*(x)\right]. \tag{8}$$

A technical condition called n.d.s. (see Definition 1 below) defines a set ω of values of x for which these constructions can be done avoiding division by zero. Estimation of α_i and $\delta_i(\tilde{x})$ directly follows these constructions, based on nonparametric regression estimates of $p_i^*(x)$. We can then average the expression for α_i over all $x \in \omega$ and average the expression for $\delta_i(\tilde{x})$ over the values that excluded regressors take on in ω conditional on $\tilde{x}$. This averaging exploits over-identifying information in our model to increase efficiency.

The remaining step is to identify the distribution function $F_{\mathcal{S}_i|\tilde{x}}$ and its features. For each i and x, define a "generated special regressor" as $v_i \equiv V_i(x) \equiv \alpha_i x_i + \delta_i(\tilde{x})p_j^*(x)$. These $V_i(x)$ are identified since α_i and $\delta_i(\tilde{x})$ have been identified. It then follows from the BNE and equation (1) that each player i's strategy is to choose

$$D_i = 1\{\alpha_i X_i + \delta_i(\tilde{X})p_j^*(X) + u_i(\tilde{X}, \epsilon_i^*) \geq 0\} = 1\{\mathcal{S}_i \leq V_i\}.$$

This has the form of a binary choice threshold crossing model, with the unknown function $\mathcal{S}_i = -u_i(\tilde{X}, \epsilon_i^*)$ and the error ϵ_i^* having an unknown distribution with unknown heteroskedasticity. It follows from the excluded regressors X_e being conditionally independent of ϵ_i that V_i and $\mathcal{S}_i$ are conditionally independent of each other conditioning on $\tilde{x}$. The main insight is that V_i takes the form of a special regressor as in Lewbel (1998, 2000), which provides the required identification. In particular, the conditional independence makes $E\left(D_i \mid V_i = v_i, \tilde{X} = \tilde{x}\right) = F_{\mathcal{S}_i|\tilde{x}}(v_i)$, so $F_{\mathcal{S}_i|\tilde{x}}(v)$ is identified at every value of v on the support of V_i given $\tilde{X} = \tilde{x}$, and is identified everywhere on its full support if V_i has sufficiently large support.

Define $\tilde{u}_i(\tilde{x}) \equiv E[u_i(\tilde{X}, \epsilon_i^*) \mid \tilde{x}] = E(-\mathcal{S}_i \mid \tilde{x})$. Then by definition and the conditional independence in A3, the conditional mean of player i's baseline payoff is $E[u_i^*(X, \epsilon_i^*) \mid x] = \alpha_i x_i + \tilde{u}_i(\tilde{x})$, and all other moments of the distribution of payoff functions likewise depend only on $\alpha_i x_i$ and on conditional moments of $\mathcal{S}_i$. So a goal is to estimate $\tilde{u}_i(\tilde{x})$, and more generally estimate conditional moments or quantiles of $\mathcal{S}_i$. Nonparametric estimators for conditional moments of $\mathcal{S}_i$ and their limiting distribution could be obtained as in Lewbel, Linton, and McFadden (2011), except that they assumed an ordinary binary choice model where V_i is observed whereas in our case V_i is an estimated generated regressor, constructed from estimates of α_i, $\delta_i(\tilde{x})$, and $p_j^*(x)$. We provide alternative estimators that take the estimation error in V_i into account. Note that we cannot avoid this problem by taking the excluded regressor x_i itself to be the special regressor because x_i also appears in $p_j^*(x)$.

The rest of this section formalizes these constructions, and generalizes them to games with more than two players.

3.2 Identifying interaction effects

First consider identifying the direction of the marginal effects of excluded regressors α_i, and the interaction effects $\delta_i(\tilde{x})$. We start with the derivation of the general game analog to (4).

For any x, let $\phi_i^*(x)$ denote the conditional expectation of $h_i(D_{-i})$ given x in equilibrium, which is identifiable directly from the data (with h_i known to all i and econometricians). If $h_i(D_{-i}) = \sum_{j\neq i} D_j$, then $\phi_i^*(x) \equiv \sum_{j\neq i} p_j^*(x)$. For all i,j, let $\phi_{i,j}^*(x) \equiv \partial\phi_i^*(X)/\partial X_j|_{X=x}$. Let $\mathcal{A}$ denote a N-vector with ordered coordinates $(\alpha_1, \alpha_2, .., \alpha_N)$. For any x, define four N-vectors $\mathcal{W}_1(x) \equiv (p_{1,1}^*(x),\ p_{2,2}^*(x),\ ..,\ p_{N,N}^*(x))$; $\mathcal{V}_1(x) \equiv (\phi_{1,1}^*(x), \phi_{2,2}^*(x), .., \phi_{N,N}^*(x))$; $\mathcal{W}_2(x) \equiv (p_{1,2}^*(x),\ p_{2,3}^*(x),\ ..,\ p_{N-1,N}^*(x),\ p_{N,1}^*(x))$; and $\mathcal{V}_2(x) \equiv (\phi_{1,2}^*(x),\ \phi_{2,3}^*(x),\ ..,\ \phi_{N-1,N}^*(x),\ \phi_{N,1}^*(x))$ respectively. For any two vectors R and R', let "$R.R'$" and "$R./R'$" denote component-wise multiplication and division (i.e. $R.R' = (r_1 r_1', r_2 r_2', .., r_N r_N')$ and $R./R' = (r_1/r_1', r_2/r_2', .., r_N/r_N')$ where r_n and r_n' are the n-th components of R and R' respectively). Under A1,2,3,

$$p_i^*(x) = F_{\mathcal{S}_i|\tilde{x}}(\alpha_i x_i + \delta_i(\tilde{x})\phi_i^*(x))$$

for all $i \in N$ in a Bayesian Nash equilibrium.

Definition 1. *A state x is non-degenerate and non-singular (n.d.s.) if $0 < p_i^*(x) < 1$ for all $i \in N$ and if all components in $\mathcal{W}_2(x)$, $\mathcal{V}_2(x)$ and $\mathcal{V}_2(x).\mathcal{W}_1(x) - \mathcal{V}_1(x).\mathcal{W}_2(x)$ are non-zero. A set ω is n.d.s. if it is a bounded set in the interior of the support of X, and all x in ω are n.d.s.*

It is instructive to explain the n.d.s. condition using the simple example of two players i, j with $h_i(D_j) \equiv D_j$ and $h_j(D_i) \equiv D_i$. The Bayesian Nash equilibrium in this case is characterized as in (4). That $F_{\epsilon^*|\tilde{x}}$ is smooth with bounded density given any $\tilde{x}$ (A1) and that u_i is continuous in ϵ_i^* for all $\tilde{x}$ (A3) imply the partial derivatives of p_i^*, p_j^* with respect to x_i, x_j exist in the interior of the support of X.

In this two-player case, $\phi_i^*(x)$ and $\phi_j^*(x)$ reduce to $p_j^*(x)$ and $p_i^*(x)$, and a state x in the interior of the support of X is n.d.s. if $p_i^*(x), p_j^*(x) \in (0,1)$; $p_{i,i}^*$, $p_{j,i}^*$, $p_{i,j}^*$, $p_{j,i}^*$ at x are all non-zero; and the Jacobian of (p_i^*, p_j^*) with respect to the pair of excluded regressors is non-singular at x. An intuitive and sufficient condition for a realized state $x \equiv (x_i, x_j, \tilde{x})$ to be n.d.s. is as follows:

"For $k \in \{i, j\}$, $\delta_k(\tilde{x}) \neq 0$ and both $\alpha_k x_k$ and $\alpha_k x_k + \delta_k(\tilde{x})$ are in the interior of the support of $\mathcal{S}_k$ given $\tilde{x}$; and the conditional density $f_{\mathcal{S}_k|\tilde{x}}(.)$ is bounded away from 0 over the interval between $\alpha_k x_k$ and $\alpha_k x_k + \delta_k(\tilde{x})$".

This sufficient condition holds, for example, if for both $k = i, j$, the excluded regressor x_k takes a moderate value while ϵ_k^* varies sufficiently to create a large support of $\mathcal{S}_k$ conditional on $\tilde{x}$. To see this, first note that under these conditions the support of $\mathcal{S}_i$ includes the support of $\alpha_i x_i + D_j \delta_i(\tilde{x})$ in its interior conditional on x. Hence $0 < p_i^*(x)$, $p_j^*(x) < 1$, which in turn implies the partial derivatives $p_{i,i}^*(x)$, $p_{j,j}^*(x)$ are nonzero under A1,2,3. Second, with $\delta_i(\tilde{x}) \neq 0$ and the density of $\mathcal{S}_i$ given $\tilde{x}$ being strictly positive between $\alpha_i x_i$ and $\alpha_i x_i + \delta_i(\tilde{x})$, the equilibrium relation between $p_{i,j}^*$ and $p_{j,j}^*$ in (6) implies that $p_{i,j}^*(x) \neq 0$ whenever $p_{j,j}^*(x) \neq 0$. By symmetric argument, $p_{j,i}^*(x)$ is also nonzero. Finally, that $f_{\mathcal{S}_k|\tilde{x}}(.)$ is strictly positive between $\alpha_k x_k$ and $\alpha_k x_k + \delta_k(\tilde{x})$ allows us to take the ratio of both sides of the equations (5) and (6) to get

$$p_{i,i}^*(x) p_{j,j}^*(x) - p_{i,j}^*(x) p_{j,i}^*(x) = \tfrac{\alpha_i}{\delta_i(\tilde{x})} p_{i,j}^*(x) \neq 0.$$

Hence the Jacobian is non-singular at such a x.[10]

Theorem 1 *Suppose A1,2,3 hold and ω is n.d.s.. Then (i) $(\alpha_1, ., \alpha_N)$ are identified as the signs of the components of*

$$E\left[\left(\mathcal{W}_1(X) - \mathcal{W}_2(X).\mathcal{V}_1(X)./\mathcal{V}_2(X)\right) 1\{X \in \omega\}\right]. \tag{9}$$

[10] In order for all coordinates in $\mathcal{W}_2(x)$, $\mathcal{V}_2(x)$ and $\mathcal{V}_2(x).\mathcal{W}_1(x) - \mathcal{V}_1(x).\mathcal{W}_2(x)$ to be non-zero, $\delta_i(\tilde{x})$ must necessarily be non-zero for all i. This rules out uninteresting cases with no strategic interaction between players. To see this, note if $\delta_i(\tilde{x}) = 0$ for some i, then $p_i^*(x)$ would be independent from x_j for all $j \neq i$ (and $p_{i,j}(x)$ would be zero) for all x in equilibrium.

(ii) For any $\tilde{x}$ such that there exists x_e with $(x_e, \tilde{x}) \in \omega$, the vector $(\delta_1(\tilde{x}), ., \delta_N(\tilde{x}))$ is identified as:

$$\mathcal{A}.E\left[\mathcal{W}_2(X)./\left(\mathcal{V}_2(X).\mathcal{W}_1(X) - \mathcal{V}_1(X).\mathcal{W}_2(X)\right) \mid X \in \omega, \tilde{X} = \tilde{x}\right]. \tag{10}$$

A few remarks are in order. First, for the simple case of two players with $h_i(D_j) = D_j$ and $h_j(D_i) = D_i$, Theorem 1 is simplified to the argument in Section 3.1. Second, Theorem 1 does not adopt an identification-at-infinity approach. That is, it does not send x_i to positive or negative infinity to identify α_i and $\delta_i(\tilde{x})$ from the limits of i's choice probability. Hence identification does not require observing states far out in the tails of the distribution. Third, if $\delta_i(\tilde{x}) = \delta_i$ for all i and some constants δ_i, then $(\delta_1, ., \delta_N)$ are identified as

$$\mathcal{A}.E[\mathcal{W}_2(x)./\left(\mathcal{V}_2(x).\mathcal{W}_1(x) - \mathcal{V}_1(x).\mathcal{W}_2(x)\right) \mid X \in \omega]$$

by pooling data from all games with states in ω. In a two-player game,

$$\delta_1 = \alpha_1 E\left[\left(p^*_{1,1}(X)p^*_{2,2}(X) - p^*_{2,1}(X)p^*_{1,2}(X)\right)^{-1} p^*_{1,2}(X) \mid X \in \omega\right]; \tag{11}$$

and likewise for δ_2.

3.3 Identifying the distribution of baseline payoffs

For any i and $x \equiv (x_e, \tilde{x})$, define a "generated special regressor":

$$V_i(x) \equiv \alpha_i x_i + \delta_i(\tilde{x})\phi^*_i(x).$$

Such a generated regressor can be constructed from the distribution of actions and states in the data once α_i and $\delta_i(\tilde{x})$ is identified. The support of V_i conditional on $\tilde{X} = \tilde{x}$ is the set of $t \in \mathbb{R}$ such that there exists x_e on the support of X_e given $\tilde{X} = \tilde{x}$ with $\alpha_i x_i + \delta_i(\tilde{x})\phi^*_i(x) = t$.

Theorem 2 *Suppose A1,2,3 hold and let α_i and $\delta_i(\tilde{x})$ be identified at $\tilde{x}$. Then the distribution of $\mathcal{S}_i$ given $\tilde{X} = \tilde{x}$ is identified over the support of V_i given $\tilde{X} = \tilde{x}$.*

To prove Theorem 2, observe that for a fixed $\tilde{x}$, A3 ensures the variation in X_e does not affect the distribution of $\mathcal{S}_i$ given $\tilde{X} = \tilde{x}$. Given that the data is rationalized by a unique BNE (A2), the choice probabilities identified from the data is related to the distribution of $\mathcal{S}_i$ given $\tilde{x}$ for all i as:

$$\Pr\left(D_i = 1 \mid X = (x_e, \tilde{x})\right) = \Pr(\mathcal{S}_i \leq V_i(x_e, \tilde{x}) \mid \tilde{X} = \tilde{x}) \tag{12}$$

where the equality follows from the conditional independence between X_e and ϵ^*_i in A3. Such conditional independence allows us to identify the distribution of $\mathcal{S}_i$ given $\tilde{X} = \tilde{x}$ as

$$\Pr\left(\mathcal{S}_i \leq t \mid \tilde{X} = \tilde{x}\right) = E(D_i \mid \tilde{X} = \tilde{x}, V_i = t) \tag{13}$$

for all t on the support of $V_i(X)$ given $\tilde{X} = \tilde{x}$. Therefore for any $\tau \in (0,1)$ such that there exists some x with $\Pr(D_i = 1 \mid x) = \tau$, the τ-quantile of $\mathcal{S}_i$ given $\tilde{X} = \tilde{x}$ is identified as

$$\inf\{t \in \mathbb{R} : E(D_i \mid \tilde{X} = \tilde{x}, V_i = t) \geq \tau\}.$$

Since α_i is identified, the identification of $F_{\mathcal{S}_i|\tilde{x}}$ over the support of $V_i(X)$ given $\tilde{X} = \tilde{x}$ implies the distribution of the baseline payoff $u_i^*(x, \epsilon_i^*) \equiv \alpha_i x_i + u_i(\tilde{x}, \epsilon_i^*) = \alpha_i x_i - \mathcal{S}_i$ conditional on $X = x$ is also identified over a relevant part of its support.To recover the baseline payoff distributions over the full support, we need the following large support condition.

A4 *(Large Support) For any i and $\tilde{x}$, the support of $\mathcal{S}_i$ given $\tilde{x}$ is a subset of the support of V_i given $\tilde{x}$.*

A4 is analogous to the large support condition for special regressors in single-agent models in Lewbel (2000). In the special case with the two players considered above, it holds if the support of X_i given $\tilde{x}$ is large enough relative to both $\delta_i(\tilde{x})$ and the support of ϵ_i^* given $\tilde{x}$. We provide primitive conditions that are sufficient for this large support condition in the online supplement to this paper. Note A4 can also be checked using the distribution of actions and states from the data. Specifically, A4 holds for some i and $\tilde{x}$ if and only if the supreme and the infimum of $E(D_i \mid \tilde{X} = \tilde{x}, V_i = v)$ as v vary over the support of V_i given $\tilde{X} = \tilde{x}$ are 1 and 0 respectively.

Corollary 1 *(Theorem 2) Suppose A1,2,3,4 hold. The distribution of $u_i^*(X, \epsilon_i^*)$ conditional on X is identified.*

To prove this corollary, note the distribution of the baseline payoff $u_i^*(X, \epsilon_i^*)$ conditional on $X = x = (x_e, \tilde{x})$ is

$$\Pr(u_i^*(X, \epsilon_i^*) \leq u \mid X = x) = \Pr(\mathcal{S}_i \geq \alpha_i x_i - u \mid X = x) = E(1 - D_i \mid \tilde{X} = \tilde{x}, V_i = \alpha_i x_i - u)$$

for all u. The first equality above follows from the definition of $\mathcal{S}_i$ and the second from the independence between excluded regressors X_e and ϵ_i^* conditional on $\tilde{X}$. Under A4, the distribution of $\mathcal{S}_i$ given $\tilde{X} = \tilde{x}$ is identified over its full support for all $\tilde{x}$. It then follows that the distribution of $u_i^*(X, \epsilon_i^*)$ given $X = x$ is identified over its full support for all x.

It is important to clarify the relation between A4 and the existence of a *n.d.s.* set that is used to identify α_i and $\delta_i(.)$ in Section 3.2. For any fixed $\tilde{x}$, the large support condition in A4 is consistent with the existence of such a *n.d.s.* set ω. To see this, recall the set of sufficient conditions for a realized $x = (x_i, x_j, \tilde{x})$ to be *n.d.s.* in the two-player example in Section 3.2. Those conditions require that for $k \in \{i, j\}$, both $\alpha_k x_k$ and $\alpha_k x_k + \delta_k(\tilde{x})$ fall in the interior of the support of $\mathcal{S}_k \equiv -u_k(\tilde{X}, \epsilon_k^*)$ once conditional on $\tilde{X} = \tilde{x}$. In other words, the existence of a *n.d.s.* set ω when $\tilde{X}$ is fixed at $\tilde{x}$ requires there exists some (x_i, x_j) such that $\alpha_i x_i, \alpha_j x_j$ and $\alpha_i x_i + \delta_i(\tilde{x}), \alpha_j x_j + \delta_j(\tilde{x})$

are all in the support of $\mathcal{S}_k$ given $\tilde{X} = \tilde{x}$. In comparison, the large support condition in A4 requires that there is sufficient variation in the *random vector* (X_i, X_j) once conditioning on $\tilde{X} = \tilde{x}$. It is therefore consistent with the existence of a *n.d.s.* set of states ω.

3.4 Shortcuts for recovering mean baseline payoffs

For the rest of this article, we assume the moment $E[u_i(\tilde{X}, \epsilon_i^*) \mid \tilde{X} = \tilde{x}]$ exists for all i and $\tilde{x}$. Without loss of generality, $u_i(\tilde{X}, \epsilon_i^*)$ can be reparametrized as $\tilde{u}_i(\tilde{X}) - \epsilon_i$, with $\tilde{u}_i(\tilde{X}) \equiv E[u_i(\tilde{X}, \epsilon_i^*) \mid \tilde{X}]$ and $\epsilon_i \equiv \tilde{u}_i(\tilde{X}) - u_i(\tilde{X}, \epsilon_i^*)$. By construction $E[\epsilon_i \mid \tilde{X} = \tilde{x}] = 0$ for all i and $\tilde{x}$; and $F_{\epsilon|X}$ satisfies the identifying restrictions on $F_{\epsilon^*|X}$ (that is, A1 and A3-(ii) hold with ϵ^* replaced by ϵ). Under A1,2,3,4, the distribution of $\mathcal{S}_i$ conditional on $\tilde{X} = \tilde{x}$ is identified over its full support for all $\tilde{x}$ and i. This implies $\tilde{u}_i(\tilde{x})$ is identified as $E\left[-\mathcal{S}_i \mid \tilde{x}\right]$ for all $\tilde{x}$ and i.

The argument for identifying $\tilde{u}_i(\tilde{x})$ above is intuitive. However, an estimator based on this argument and the sample analog principle is cumbersome to implement. It involves the integration of the estimate of $F_{\mathcal{S}_i|\tilde{x}}$, which itself conditions on a generated special regressor V_i (see (13)). For the rest of this subsection, we develop alternative "shortcut" methods for identifying the mean baseline payoff that leads to simpler forms of estimators whose asymptotic properties can be established using standard arguments.

A5 (*Positive Density) For any $\tilde{x}$, the density of V_i given $\tilde{X} = \tilde{x}$ is positive almost surely w.r.t. the Lebesgue measure over its support.*

Let c^* be any constant known to be in the interior of the support of V_i given $\tilde{x}$. For all i and $x = (x_e, \tilde{x})$ define:

$$Y_i^* \equiv \frac{D_i - 1\{V_i(x) \geq c^*\}}{f_{V_i|\tilde{X}=\tilde{x}}(V_i(x))}. \tag{14}$$

The choice of c^* is feasible in estimation, because the support of V_i given $\tilde{x}$ is identified once α_i and $\delta_i(\tilde{x})$ are identified. The choice of such a c^* is allowed to depend on $\tilde{x}$.

Theorem 3 *Suppose A1,2,3,4,5 hold, and α_i and $\delta_i(\tilde{x})$ are both identified at $\tilde{x}$. Then:*

$$\tilde{u}_i(\tilde{x}) = E\left(Y_i^* \mid \tilde{X} = \tilde{x}\right) - c^*$$

for any c^* in the interior of the support of V_i conditional on $\tilde{X} = \tilde{x}$.

The proof of Theorem 3 is similar to that of Theorem 1 in Lewbel (2000), and so is relegated to this paper's online supplement. Using sample analogs of (14) to estimate $\tilde{u}_i$ requires estimating the density of V_i conditional on $\tilde{x}$, which itself needs to use the estimates for $\alpha_i, \delta_i(\tilde{x})$ and $p_j(x)$.

Furthermore, the estimates for V_i enter the indicator in the numerator, thus making the derivation of asymptotic properties of the estimator even harder.

Our next result shows these complications in estimation can be avoided by implementing two modifications in the regressand constructed for (14): First, replace the conditional density of V_i given $\tilde{X}$ in the denominator with the density of X_i given X_{-i} by changing variables between V_i and X_i given X_{-i}. Identification then requires the monotonicity of V_i in X_i and a strengthened version of the support condition in A4. These conditions are presented in A4' below. Second, replace the indicator function in the numerator of (14) with a smooth distribution satisfying certain properties. Under an argument using integration-by-parts, these properties help to establish identification.

Specifically, let H be a continuously differentiable cumulative distribution function with a bounded interval support which is a subset of the support of V_i conditional on X_{-i}.[11] Let μ denote the expectation of H, which is in the interior of the support of H. For any i and x, modify the definition of Y_i^* in (14) to:

$$Y_i \equiv \frac{[D_i - H(V_i(x))]\left[1 + \alpha_i \delta_i(\tilde{x})\phi_{i,i}^*(x)\right]\alpha_i}{f_{X_i|X_{-i}=x_{-i}}(x_i)}. \tag{15}$$

Note $\alpha_i + \delta_i(\tilde{x})\phi_{i,i}^*(x)$ is the marginal effect of X_i on V_i evaluated at x; and the multiplicative factor $1 + \alpha_i\delta_i(\tilde{x})\phi_{i,i}^*(x) = \alpha_i\left[\alpha_i + \delta_i(\tilde{x})\phi_{i,i}^*(x)\right]$ is a correction term needed to offset the impact of using $f_{X_i|x_{-i}}$ in the denominator.

A4' *For any i and $x_{-i} \equiv ((x_j)_{j\neq i}, \tilde{x})$, (i) the support of $\mathcal{S}_i$ given $\tilde{X} = \tilde{x}$ is a subset of the support of V_i given $X_{-i} = x_{-i}$; (ii) the sign of $\partial V_i(X)/\partial X_i|_{X=(x_i,x_{-i})}$ equals the sign of α_i for all x_i; and (iii) the density of X_i given x_{-i} exists and is positive almost everywhere.*

Both A4 and A4' require V_i to vary over some conditional support which is sufficiently large relative to the support of $\mathcal{S}_i$ given $\tilde{x}$. In A4', such variation arises solely from X_i while X_{-i} is fixed. In comparison, A4 only fixes $\tilde{X}$ and the variation in V_i can arise from *all* other excluded regressors X_j with $j \neq i$. The monotonicity of V_i in X_i holds when X_i's *direct* marginal effect on i's latent payoff is never offset by its *indirect* marginal effect through the interaction effect given X_{-i}. While such monotonicity is not required for Theorem 3, it is instrumental for the identification shortcut proposed in Corollary 2, which uses a change-of-variable between X_i and V_i.[12] Both

[11] The choice of H may depend on i and X_{-i}, which are suppressed in the notation for simplicity. Choice of such an H is feasible with α_i, $\delta_i(\tilde{x})$ and the support of $V_i(X)$ given $\tilde{x}$ already identified. Section 5.2 provides an example of the choice of H in estimation.

[12] V_i is monotone in X_i given $x_{-i} = ((x_j)_{j\neq i}, \tilde{x})$ when components in $(x_j)_{j\neq i}$ are extremely large or small. Consider the case with $N = 2$ and $h_i(D_j) = D_j$. The marginal effect of X_i on V_i at $x = (x_i, x_j, \tilde{x})$ is $\alpha_i + \delta_i(\tilde{x})p_{j,i}^*(x)$, where α_i and $\delta_i(\tilde{x})$ are both finite constants once conditional on $\tilde{x}$. With x_j extremely large or small, the impact of x_i on p_j^* in BNE is close to 0. Thus $sign(\alpha_i + \delta_i(\tilde{x})p_{j,i}^*(x)) = sign(\alpha_i)$ for all x_i conditional on $\tilde{x}$ and on such extremely large or small x_j. In fact, such a conditional monotonicity holds uniformly over the support of X if $\delta_i(\tilde{x})$ and the density $f_{\epsilon_i|\tilde{x}}$ are both bounded by small constants.

the large support and the monotonicity in A4' are testable, because the marginal effect of X_i on $\Pr(D_i = 1 \mid X)$ is identifiable and shares the same sign as the marginal effect of X_i on V_i under the exclusion restriction in A3.

Corollary 2 *(Theorem 3) Suppose A1,2,3,4' hold and α_i, $\delta_i(\tilde{x})$ are identified at $\tilde{x}$. Then*

$$\tilde{u}_i(\tilde{x}) = E\left(Y_i - \mu \mid \tilde{X} = \tilde{x}\right). \tag{16}$$

The definition of Y_i simplifies the limiting distribution theory of our estimator using generated special regressors. In particular, a smooth H allows us to take standard Taylor expansions to linearize the estimator as a function of the estimates for V_i around the true V_i. Dividing by the density of X_i instead of the density of V_i only involves estimating the density of regressors reported in the data, as opposed to that of generated regressors.

3.5 Discussion: specification and over-identification

Our specification allows the distribution of ϵ_i^* to be heteroskedastic and the baseline payoffs $u_i^*(X, \epsilon_i^*)$ to be unrestricted except for the additive separability and conditional independence of excluded regressors. It also allows the multiplicative factor $\delta_i(.)$ in the interaction effects to be state-dependent. On the other hand, our specification requires private information to be excluded from interaction effects.

In the entry game considered in Section 4 below, the interaction effect reflects the difference between a firm's profit under alternative market structures (monopoly vs Cournot oligopoly). In this game, the aforementioned exclusion restriction essentially says that variable costs depend on commonly known states while fixed costs depend on firms' private information. This assumption is plausible in industries where the production technology is standard and common knowledge across firms, and where the quantity produced is determined by market or firm level states. Unobserved idiosyncrasies across firms then capture quantity-invariant cost factors. An example would be the airline industry, where the quantity produced by a carrier refers to the publicly observable traffic flow accommodated on the routes it operates, which in turn depends on demographics and airport presence in each market. In contrast, fixed costs depend in part on unobserved errors, possibly including measures of opportunity costs. For instance, both Berry (1992) and Ciliberto and Tamer (2009) set the interaction effect δ_i to be constant, which is consistent with the assumption that interaction effects are invariant to errors and to states reported in the data. Strictly speaking, our model and that in Berry (1992) and Ciliberto and Tamer (2009) are non-nested because of different assumptions on the information structure. However, the main analogy we draw here is that the unobserved idiosyncrasies on the firm level are excluded from the interaction effect both in their model and in ours.

Although plausible in some contexts as above, the exclusion restrictions we use for identification do limit the scope of applicability of our method. A model where interaction effects depend on players' private information would give rise to additional technical challenges. In particular, a Bayesian Nash equilibrium may not be monotone in private information, and can not be characterized through the fixed-point equations in the choice probabilities as in (4). Hence robust identification would require a qualitatively different approach.

The assumptions we use also provide over-identifying information that can be used to test our model specification and identifying conditions. For example, our identification of α_i and $\delta_i(.)$ in the two-player example assumes that the sign of the right-hand side of (7) does not vary over the support of X and that the right-hand side of (8) is invariant over the support of excluded regressors X_e for any given $\tilde{x}$. One could test this by estimating the the right-hand side of (8) at different values of X_e when the non-excluded regressor is fixed at x, and testing whether the resulting estimates are equal.

In addition, with more than two players we can also recover α_i and $\delta_i(.)$ under the maintained assumptions using other partial derivatives than those in $\mathcal{W}_2$, $\mathcal{V}_2$. For example, the proof of Theorem 1 can be modified by replacing $\mathcal{W}_2$ with $(p^*_{1,N}, p^*_{2,1}, p^*_{3,2}, ..., p^*_{N-1,N-2}, p^*_{N,N-1})$ and replacing the components in $\mathcal{V}_2$ with other partial derivatives accordingly. The interaction effects recovered through these different methods must be identical, thus providing a testable implication for the identifying conditions.

4 Potential Applications

In this section, we provide examples of how excluded regressors of the type we use for identification arise in various games considered in the literature.

4.1 Fixed costs in oligopoly markets

Consider a two-stage entry game between two firms $i \in \{1, 2\}$ and $j = 3 - i$. First, each firms decides whether to enter a market. If only one enters, it chooses its monopoly quantity and price, while if both enter the firms compete as a duopoly by choosing quantities to produce as in a Cournot game.

Let q_i denote the quantity produced by i after entry, and $\tilde{x}$ denote a vector of market-level or firm-level characteristics observed by both firms and reported in the data. The inverse market demand is $\psi^M(q_i; \tilde{x})$ if i is the sole entrant on a monopoly market; and it is $\psi^D(q_i, q_j; \tilde{x})$ otherwise.

That the inverse market demand only depends on $\tilde{x}$ implies researchers obtain the same information set as consumers on the market.

The cost for firm i is $c_i(q_i;\tilde{x})+x_i+\varepsilon_i$, where c_i is i's variable costs, the scalar x_i is a commonly observed component of i's fixed cost, and ε_i is an idiosyncratic component of i's fixed cost known only to itself. The fixed costs $x_i+\varepsilon_i$ (and any other fixed cost components that may be absorbed in c_i) are paid *after* entry but *before* choosing production quantities. Both (x_1,x_2) and $\tilde{x}$ are known to the firms when they make entry decisions in the first stage, and are reported in the data. A firm that decides not to enter receives zero profit.

If i is the only entrant, then its monopoly profit is $\psi^M(q^*;\tilde{x})q^* - c_i(q^*;\tilde{x}) - x_i - \varepsilon_i$, where q^* solves

$$\tfrac{\partial}{\partial q}\psi^M(q^*;\tilde{x})q^* + \psi^M(q^*;\tilde{x}) = \tfrac{\partial}{\partial q}c_i(q^*;\tilde{x})$$

assuming an interior solution exists. Otherwise, i's profit is $\psi^D(q_i^*,q_j^*;\tilde{x})q_i^* - c_i(q_i^*;\tilde{x}) - x_i - \varepsilon_i$ where (q_i^*,q_j^*) solves

$$\begin{aligned}
\tfrac{\partial}{\partial q_i}\psi^D(q_i^*,q_j^*;\tilde{x})q_i + \psi^D(q_i^*,q_j^*;\tilde{x}) &= \tfrac{\partial}{\partial q_i}c_i(q_i^*;\tilde{x}) \text{ and} \\
\tfrac{\partial}{\partial q_j}\psi^D(q_i^*,q_j^*;\tilde{x})q_j + \psi^D(q_i^*,q_j^*;\tilde{x}) &= \tfrac{\partial}{\partial q_j}c_j(q_j^*;\tilde{x}),
\end{aligned}$$

assuming interior solutions exist. Choices of output quantities upon entry depend only on $\tilde{x}$ in both cases. This is because by definition the fixed cost components x_i and ε_i do not vary with output. Hence an entrant's profit under either structure takes the form $\pi_i^s(\tilde{x}) - x_i - \varepsilon_i$, with $s \in \{D,M\}$ and π_i^s is i's gross profit (defined by subtracting its variable cost component $c_i(q_i^*;\tilde{x})$ from its gross revenue) under market structure s.

In the first stage, firms make entry decisions simultaneously based on their expected profit from entry conditional on the public information $X \equiv (\tilde{X},X_1,X_2)$ and private information ε_i. The following matrix summarizes profits in all scenarios, taking into account firms' Cournot Nash strategies under duopoly and profit maximization under monopoly (Firm 1 is the row player, whose profit is the first entry in each cell):

	Enter	Stay out
Enter	$(\pi_1^D(\tilde{X}) - X_1 - \epsilon_1 \ , \ \pi_2^D(\tilde{X}) - X_2 - \epsilon_2)$	$(\pi_1^M(\tilde{X}) - X_1 - \epsilon_1 \ , \ 0)$
Stay out	$(0 \ , \ \pi_2^M(\tilde{X}) - X_2 - \epsilon_2)$	$(0 \ , \ 0)$

It follows that i's ex ante profits from entry are:

$$\pi_i^M(\tilde{x}) + p_j(x)\delta_i(\tilde{x}) - x_i - \varepsilon_i \tag{17}$$

where $\delta_i(\tilde{x}) \equiv \pi_i^D(\tilde{x}) - \pi_i^M(\tilde{x})$, and $p_j(x)$ is the probability that firm j decides to enter conditional on i's information set. That p_j does not depend on ε_i is due to the independence between private information conditional on $\tilde{x}$. Profits from not entering are zero. Hence, in any Bayesian Nash equilibrium, i decides to enter if and only if (17) is greater than zero, and $\Pr(D_1 = d_1,\ D_2 = d_2 \mid x) = \prod_{i=1,2} p_i(x)^{d_i}(1 - p_i(x))^{1-d_i}$, where

$$p_i(x) = \Pr\left\{ \epsilon_i \le \pi_i^M(\tilde{x}) + p_j(x)\delta_i(\tilde{x}) - x_i \;\middle|\; x \right\}$$

for $i = 1, 2$. This fits in the class of games considered above, with $V_i(X) \equiv p_j(X)\delta_i(\tilde{X}) - X_i$.

Examples of fixed cost components (X_1, X_2) that could be public knowledge include lump-sum expenses such as costs of purchasing land, buildings, and equipment, maintenance fees and patent or license fees, while (ϵ_1, ϵ_2) could include idiosyncratic administrative overhead and costs of lost opportunities of firm specific alternatives to entry. The non-excluded regressors $\tilde{X}$ would include market-level or industry-level variables that shift demand or cost. In these examples the fixed cost components (ϵ_1, ϵ_2) and (X_1, X_2) are incurred by different aspects of the production process, making it plausible that they are independent of each other (after conditioning on the given market and industry conditions $\tilde{X}$), thereby satisfying the conditional independence in A3.

The large support assumption required for identifying the mean baseline payoffs holds if the support for publicly known fixed costs is large relative to that of privately known fixed cost components. For example, a set of sufficient (but not necessary) conditions for the large support assumption to hold at a given $\tilde{x}$ are: (a) the support of ϵ_i given $\tilde{x}$ is included in $[0, \bar{\varepsilon}]$ with $\bar{\varepsilon} \le \pi_i^M(\tilde{x})$; (b) there exists (x_i, x_j) with $p_j(x_i, x_j, \tilde{x}) = 0$ and $\pi_i^M(\tilde{x}) - \bar{\varepsilon} \ge x_i$; and (c) there exists $(\hat{x}_i, \hat{x}_j)$ with $p_j(\hat{x}_i, \hat{x}_j, \tilde{x}) = 1$ and $\pi_i^D(\tilde{x}) \le \hat{x}_i$. In economic terms, (a) means the gross profits under monopoly are sufficiently high; while (b) and (c) require the observed fixed cost components X_i and X_j to vary sufficiently.[13]

More generally, the large support condition has an immediate economic interpretation. It means that, conditional on demand shifters $\tilde{X}$, a firm i may not find it profitable to enter even when the fixed cost component ϵ_i is as low as possible. This happens if X_i (the observed component of fixed cost) is sufficiently high, or if the competition is too intense (in the sense that the other firm's entry probability p_j is sufficiently high). Large support also means that firm i could decide to enter even when ϵ_i is high, because X_i and $p_j(X)$ can be sufficiently low while the market demand is sufficiently high to make entry attractive. These conditions on the supports of X_i and ϵ_i are plausible in applications where variation in publicly known components of fixed costs are relatively large. For example, the fixed costs in acquiring land, buildings, and equipment and acquiring license/patents could be much more substantial than the fixed costs incurred by idiosyncratic supplies and overhead. Economic implications that suffice for large support are verified by checking whether firms' entry probabilities are degenerate at 0 and 1 for sufficiently high or low levels of the publicly known components of fixed costs.

[13] We are grateful to an anonymous referee for pointing this out.

4.2 Entry in airline markets

In the U.S. airline industry, a market is defined as a pair of origin and destination cities, and the set of potential entrants on a market consists of carriers (airlines) that operate in the U.S. Several recent papers study how firm heterogeneity determines the market structure in this industry. Leading examples include Berry (1992) and Ciliberto and Tamer (2009).[14]

Empirical models for airline entry typically adopt a linear specification for a carrier (firm) i's profits on a market m:

$$\pi_{i,m} = S_m \alpha_i + Z_{i,m}\beta_i + W_{i,m}\gamma_i + \delta_i h(D_{-i,m}) + \epsilon_{i,m} \tag{18}$$

where S_m and $Z_{i,m}$ are respectively market and firm characteristics; $W_{i,m}$ is a firm-level variable that affects i's profits only; and $D_{-i,m}$ is a vector of decisions by all potential entrants except i. The term δ_i captures heterogeneous interaction effects, and $h(.)$ is a known function that captures how airline's entry decisions affect competitors' profits. For example, Berry (1992) let $h(D_{-i,m}) = \ln(\sum_{j\neq i} D_{j,m} + 1)$ while Ciliberto and Tamer (2009) let $h(D_{-i,m}) = \sum_{j\neq i} D_{j,m}$.

Berry (1992) and Ciliberto and Tamer (2009) make a strong complete information assumption regarding market participants, while we assume firms only have incomplete information about others' payoff structure. Nonetheless, the type of data used in these papers lends itself to an application of our method. As explained below, the firm-specific profit factors $W_{i,m}$, defined as a proxy of fixed costs in Ciliberto and Tamer (2009), can be used as excluded regressors in our context.

Both Berry (1992) and Ciliberto and Tamer (2009) constructed their samples from the *Origin and Destination Survey of Air Passenger Traffic*. Collected by the U.S. Department of Transportation, the survey is composed of a random sample of 10% of all passenger tickets issued by U.S. carriers / airlines every quarter since 1980s. Market-level variables S_m reported in this data source include the total population of each market; the non-stop distance between the origin and the destination; per capita income and the average rate of income growth; and the distance from each airport to the closest alternative airport. This last variable controls for the possibility that passengers may choose to fly to a nearby airport so as to get to the same destination. For each carrier-market pair, Ciliberto and Tamer (2009) included a measure of airport presence $Z_{i,m}$ to capture a part of carrier heterogeneity.

To proxy carrier i's fixed costs on market m, Ciliberto and Tamer (2009) constructed $W_{i,m}$ as the percentage of nonstop distance that the airline must travel in excess of the nonstop distance if

[14] Berry (1992) considered the population of all airlines operating in the U.S. as the set of potential entrants. Ciliberto and Tamer (2009) focused on the strategic interaction between the five largest carriers in the U.S. .

the airline uses a connecting instead of a nonstop flight.[15] They argue that this is a good proxy for fixed costs for serving a market, even when a carrier already serves that market on a nonstop basis. This is because this measure reflects the cost of the best alternative to nonstop service, which is a connecting flight through the closest hub. It is associated with the fixed cost of providing airline service because it is a function of the total capacity of a plane, and does not depend on the number of passengers transported on a particular flight.

In the linear specification in (18), $W_{i,m}$ is by construction additively separable in the profit function, and excluded from the constant interaction effect δ_i. It is also plausible that an airline's private shock ϵ_i is orthogonal to the proxy for fixed costs after controlling for market characteristics and other idiosyncratic firm factors such as hub presence. In fact, Berry (1992) and Ciliberto and Tamer (2009) both maintained a stronger condition of the full independence between unobserved shocks and the vector of observed states. The support condition on $W_{i,m}$ is also justifiable in this context, because the entry probabilities for the airlines are known to vary greatly when $W_{i,m}$ take on quite large or small values, holding other factors fixed.

4.3 Provision of care for elderly parents

Checkovich and Stern (2002) studied a model of care-giving decisions by adult children in a household. In their model the amount of care (measured, e.g., in hours) provided to elderly parents by an adult child i in a household h is determined by

$$Y_{i,h}^* = X_{i,h}\beta_i - \lambda_i \sum_{j \in \mathcal{C}_h \setminus \{i\}} Y_{j,h}^* + \epsilon_{i,h}, \tag{19}$$

where $X_{i,h}$ includes both household and individual characteristics; $\mathcal{C}_h$ denotes the set of adult children in household h; and λ_i reflects how much a child i views the care provided by siblings as substitutes. The term $\epsilon_{i,h}$ captures remaining idiosyncratic factors that affect i's care-giving decisions but are unknown to i's siblings or to the econometrician. Checkovich and Stern (2002) show how such a specification could arise if adult children maximize a concave utility function that depends on siblings' decisions. They estimate an ordered response model, using a discretized measure of care $Y_{i,h}$ in the data (the number of days per week when a child provided care).

The method we propose in this paper could be used to estimate similar models where the provision of care is further discretized into a binary decision. That is,

$$D_{i,h} = 1\left\{X_{i,h}\beta_i - \lambda_i \sum_{j \in \mathcal{C}_h \setminus \{i\}} D_{j,h} + \epsilon_{i,h} \geq 0\right\} \tag{20}$$

[15] They defined $W_{i,m}$ as follows: First, compute the sum of the geographical distances between market m's endpoints and the closest hub of carrier i as a proxy for the cost that carrier i has to face to serve that market. Then they compute the difference between this distance and the nonstop distance between two airports, and divide this difference by the nonstop distance.

where $D_{i,h} = 1$ if the child i provides regular care to the elderly parents in household h. The model in (20) is based on an assumption that an individual's decision is affected mostly by the incidence (rather than the exact amount) of the care provided by siblings. Alternatively, one could view (20) as a discrete approximation to equation (19), which is not estimable due to a lack of data on the precise measure of $Y_{i,h}^*$.[16]

A potential source of data that would permit estimation of the Bayesian game model in (20) using excluded regressors is that used in Byrne, Goeree, Hiedemann and Stern (2009). Byrne et al (2009) constructed a data set based on the 1993 wave of the Assets and Health Dynamics Among the Oldest Old (AHEAD) survey and the Current Population Survey (CPS), and used it to estimate a model of siblings' care-giving decisions. Byrne et al's data include household characteristics such as the gender, age, marital status and health status of the elderly. It also contains characteristics of the adult children in the household such as age, gender, education, marital status, number of their own children and a discretized measure of how far they lived from the residence of the elderly parent. In addition, Byrne et al (2009) imputed the children's wages using the Current Population Survey, by regressing log-wages on demographic characteristics available for the children of AHEAD respondents.

The vector $X_{i,h}$ in (20) includes characteristics of the household h as well as those of the adult child i (including i's imputed wage). The imputed wage of an adult child in a household could serve as an excluded regressor in our method. A child i's wage is a continuous random variable excluded from the payoffs for his siblings ($j \neq i$) and the constant interaction effect λ_i, and is additively separable in the linear specification. After conditioning on the child i's and the parent's demographics in $X_{i,h}$, the child's wage is plausibly orthogonal to other unobservable idiosyncratic factors in the care-giving decisions (such as the degree of distaste for elderly care-giving activities). Byrne et al (2009) also provide evidence that children's wages, as a proxy to the opportunity costs for providing care for elderly parents, have a substantial effect on their care-giving decisions. The patterns reported in the paper are also consistent with the assumption that children's propensity scores vary sufficiently with excluded regressors to satisfy the necessary support assumptions.

[16] The data used in Checkovich and Stern (2002) was collected from the National Long Term Care Survey (NLTCS) in 1982 and 1984, and as such is not fit for estimating (20) using excluded regressors. That data provided information about households with elderly individuals suffering from some impairment of daily living activities. It also contained demographic information about the elderly (e.g., medical status, income and assets, use of healthcare service and sources of payment) and characteristics for each adult child (e.g., age, gender, marital status, employment status, and the number of their children). A candidate for the excluded regressors would have been the reported distance between the residence of the elderly parents and those of their adult children. However, the NLTCS only reported a *discretized* measure of such distances (i.e. brackets), thus failing to satisfy the support condition required for excluded regressors. Galichon and Henry (2011) used the same source of data to illustrate their approach of set identification in a similar model with multiple equilibria.

5 Semiparametric Estimation

In this section, we estimate a two-player semi-parametric model. For $i = 1, 2$ and $j \equiv 3 - i$, $h_i(D_j) = D_j$, $\tilde{u}_i(\tilde{X}) = \tilde{X}'\beta_i$ and $\delta_i(\tilde{X}) = \delta_i$, where β_i, δ_i are constant parameters to be inferred. Assume $E\left(\tilde{X}\tilde{X}'\right)$ has full rank. The specification is similar to Aradillas-Lopez (2010), except the independence between (ϵ_1, ϵ_2) and X therein is replaced by the conditional independence between (ϵ_1, ϵ_2) and (X_1, X_2) given $\tilde{X}$.

Consider any *n.d.s.* set ω. For $i = 1, 2$,

$$\alpha_i = sign\left\{E\left[1\{X \in \omega\}\left(p^*_{i,i}(X) - p^*_{i,j}(X)p^*_{j,i}(X)/p^*_{j,j}(X)\right)\right]\right\} \tag{21}$$

and δ_i are identified as in equation (11). For any i and x, let Y_i be defined as in (15) where $\delta_i(\tilde{x})$ is a constant δ_i and $\phi^*_{i,i}(x)$ is $p^*_{j,i}(x)$. The choice of H in (15) is specified in Section 5.2 below. Corollary 2 implies:

$$\beta_i = \left[E\left(\tilde{X}\tilde{X}'\right)\right]^{-1} E\left[\tilde{X}\left(Y_i - \mu\right)\right].$$

Sections 5.1 and 5.2 define estimators for $\alpha_i, \beta_i, \delta_i$ and derive their asymptotic distributions. Recovering these parameters requires conditioning on a *n.d.s.* set ω in (21) and (11). Because marginal effects of X_i on (p^*_i, p^*_j) are identified from the data, this set can be estimated using sample analogs of the choice probabilities and the generated special regressors. For instance, ω can be estimated by $\hat{\omega} = \{x : \hat{p}_i(x),\ \hat{p}_j(x) \in (c, 1-c);\ |\hat{p}_{i,i}(x)| > c,\ |\hat{p}_{j,j}(x)| > c,\ |\hat{p}_{i,i}(x)\hat{p}_{j,j}(x) - \hat{p}_{i,j}(x)\hat{p}_{j,i}(x)| > c\}$, where $\hat{p}_i, \hat{p}_j, \hat{p}_{i,i}, \hat{p}_{j,j}$ are kernel estimators to be defined in Section 5.1 and c is some constant close to 0. As c is strictly above 0, $\Pr\{\hat{\omega} \subset \omega\} \to 1$ by construction. Thus estimation errors in $\hat{\omega}$ has no bearing on the limiting distribution of estimators for the other parameters. In what follows, we will take ω as known in the estimation. This simplifies the derivation of asymptotic properties of the estimators, and allows us to focus on dealing with the estimation errors in $\hat{p}_i, \hat{p}_j, \hat{p}_{i,i}, \hat{p}_{i,j}$ and $\hat{\delta}_i$ in our multi-step estimators below.

5.1 Estimation of α_i and δ_i

Let X include J_c continuous components X_c and J_d discrete components X_d. Assumption A3 implies the vector of excluded regressors $X_e \equiv (X_i)_{i\in N}$ is a subvector of X_c. Hence we can write $X_c \equiv (X_e, \tilde{X}_c)$ and $X_d \equiv \tilde{X}_d$, with $\tilde{X}_c, \tilde{X}_d$ denoting the continuous and discrete components in $\tilde{X}$ respectively. The data contains G independent games, each of which is indexed by g. First, we show α_i can be estimated at an arbitrarily fast rate and δ_i estimated at the root-N rate.

In what follows, we use $f(x_c \mid \tilde{x}_d)$ as a shorthand notation for $f_{X_c|\tilde{X}_d=\tilde{x}_d}(x_c)$. Let $\gamma_0(x) \equiv f(x_c \mid \tilde{x}_d)f_d(\tilde{x}_d)$ and $\gamma_k(x) \equiv E(D_k \mid x)f(x_c \mid \tilde{x}_d)f_d(\tilde{x}_d)$ for $k = 1, 2$, with $E(.|.)$ being a conditional expectation and $f_d(.)$ the probability mass function for $\tilde{X}_d$. Let $\gamma_{k,i}(x) \equiv \partial\gamma_k(X)/\partial X_i|_{X=x}$ for $k \in \{0,1,2\}$ and $i \in \{1,2\}$. Define

$$\boldsymbol{\gamma} \equiv \left(\gamma_0, \gamma_1, \gamma_2, \gamma_{0,1}, \gamma_{0,2}, \gamma_{1,1}, \gamma_{1,2}, \gamma_{2,1}, \gamma_{2,2}\right).$$

Let $\boldsymbol{\gamma}^*$ and $\hat{\boldsymbol{\gamma}}$ denote true parameters in the DGP and the kernel estimates respectively. To simplify notation, we suppress X in $\boldsymbol{\gamma}^*$ and $\hat{\boldsymbol{\gamma}}$ below when there is no ambiguity. We use the sup-norm for the space of functions defined over the support of X.

For $i = 1, 2$, let $j \equiv 3 - i$ and define

$$A^i \equiv E\left[m_A^i\left(X; \boldsymbol{\gamma}^*\right) 1\{X \in \omega\}\right] \text{ and } \Delta^i \equiv E\left[m_\Delta^i\left(X; \boldsymbol{\gamma}^*\right) \mid X \in \omega\right],$$

where for a generic vector of functions in $\boldsymbol{\gamma}$, $m_A^i(x; \boldsymbol{\gamma})$ is the right-hand side of (7) and $m_\Delta^i(x; \boldsymbol{\gamma})$ is the right-hand side of (8) times α_i, with the choice probability p_i and its partial derivatives $p_{i,i}, p_{j,i}$ with respect to (X_i, X_j) expressed in terms of components in $\boldsymbol{\gamma}$. Let $\rho_0 \equiv \Pr\{X \in \omega\}$.

Estimate A^i, Δ^i by their sample analogs as follows:

$$\hat{A}^i \equiv \tfrac{1}{G}\textstyle\sum_g w_g m_A^i\left(x_g; \hat{\boldsymbol{\gamma}}\right) \text{ and } \hat{\Delta}^i \equiv \left(\sum_g w_g\right)^{-1} \sum_g w_g m_\Delta^i\left(x_g; \hat{\boldsymbol{\gamma}}\right)$$

where $w_g \equiv 1\{x_g \in \omega\}$. Let $K(.)$ be a product kernel for continuous coordinates in X_c, and σ be a sequence of bandwidths with $\sigma \to 0$ as $G \to \infty$. To simplify notation, let $d_{g,0} \equiv 1$ for all g. For $k = 0, 1, 2$, the kernel estimator for γ_k^* is:

$$\hat{\gamma}_k(x_c, \tilde{x}_d) \equiv \tfrac{1}{G}\textstyle\sum_g d_{g,k} K_\sigma\left(x_{g,c} - x_c\right) 1\{\tilde{x}_{g,d} = \tilde{x}_d\} \tag{22}$$

where $K_\sigma(.) \equiv \sigma^{-J_c} K(./\sigma)$. Besides, for $k = 0, 1, 2$ and $i = 1, 2$, the kernel estimator for $\gamma_{k,i}^*$, denoted by $\hat{\gamma}_{k,i}$, is defined in a similar fashion by replacing $K_\sigma(.)$ in (22) with $K_{\sigma,i}(.) \equiv \sigma^{-J_c}\partial K(./\sigma)/\partial X_i$.[17] Our estimators for α_i, δ_i are

$$\hat{\alpha}_i = sign(\hat{A}^i) \text{ and } \hat{\delta}_i = \hat{\alpha}_i \hat{\Delta}^i.$$

We now derive the limiting behavior of $\hat{\alpha}_i$ and $\hat{\delta}_i$. Let F_Z^*, f_Z^* denote the true distribution and the density of $Z \equiv (X, D_1, D_2)$ and $f^*(x_c \mid \tilde{x}_d)$ denote the true conditional density in the DGP. Additional discussion and interpretation of the following assumptions is provided in the online supplement to this paper.

Assumption S. *(i) For $k \in \{0,1,2\}$, γ_k^* is continuously differentiable in x_c up to an order of $\bar{m} \geq 2$ given any $\tilde{x}_d$, and the derivatives are continuous and uniformly bounded over an open set*

[17] Alternatively, we can also replace the indicator functions in $\hat{\gamma}_k$ and $\hat{\gamma}_{k,i}$ with smoothing product kernels for the discrete covariations as well. Denote such a joint product kernel for all coordinates in X as $\tilde{K}$. Bierens (1985) showed uniform convergence of $\hat{\gamma}_k$ in probability of can be established as long as $\sqrt{G}\int \sup_{|u_d|>\zeta/\sigma} |\tilde{K}(u_c, u_d)| du_c \to 0$ for all $\zeta > 0$.

that covers ω. (ii) All components in $\boldsymbol{\gamma}^*$ are bounded away from 0 over ω. (iii) Let $VAR_{u_i}(x)$ be the variance of $u_i = d_i - p_i^*(x)$ given x. There exists $\xi > 0$ such that $[VAR_{u_i}(x)]^{1+\xi} f^*(x_c \mid \tilde{x}_d)$ is uniformly bounded over ω. For any $\tilde{x}_d$, both $p_i^*(x)^2 f^*(x_c \mid \tilde{x}_d)$ and $VAR_{u_i}(x) f^*(x_c \mid \tilde{x}_d)$ are continuous in x_c and uniformly bounded over ω.

Assumption W. *The set ω is convex, and there exists an open set ω^o on the support of X with $\omega \subseteq \omega^o$.*

Assumption K. *(i) $K(.)$ is bounded and differentiable in X_e of order $\bar{m}$ and the partial derivatives are bounded over ω. (ii) $\int |K(t)|dt < \infty$, $\int K(t)dt = 1$, $K(.)$ has zero moments up to the order of m (where $m \leq \bar{m}$), and $\int ||u||^m |K(t)|dt < \infty$. (iii) $K(.)$ is zero outside a bounded set.*

Assumption B. $\sqrt{G}\sigma^m \to 0$ *and* $(\sqrt{G}/\ln G)\sigma^{(J_c+2)} \to \infty$ *as* $G \to \infty$.

To specify the correction terms in the limiting distribution, we need to introduce additional notation. Let $w(x) \equiv 1\{x \in \omega\}$ for the convex, *n.d.s.* set ω. Let $\tilde{D}_{i,A}(x)$ (and $\tilde{D}_{i,\Delta}(x)$) denote a nine-by-one vector consisting of derivatives of $m_A^i(x;\boldsymbol{\gamma})$ (and $m_\Delta^i(x;\boldsymbol{\gamma})$) with respect to the ordered vector (γ_0, γ_i, γ_j, $\gamma_{0,i}$, $\gamma_{0,j}$, $\gamma_{i,i}$, $\gamma_{i,j}$, $\gamma_{j,i}$, $\gamma_{j,j}$) evaluated at $\boldsymbol{\gamma}^*$. Let $\tilde{D}_{i,A,t}$ (and $\tilde{D}_{i,\Delta,t}$) denote the t-th coordinate in $\tilde{D}_{i,A}$ (and $\tilde{D}_{i,\Delta}$). For $s = i, j$, let $\tilde{D}^{(s)}_{i,A,t}(x)$ (and $\tilde{D}^{(s)}_{i,\Delta,t}(x)$) denote the derivative of $\tilde{D}_{i,A,t}(X)\gamma_0^*(X)$ (and $\tilde{D}_{i,\Delta,t}(X)\gamma_0^*(X)$) with respect to the excluded regressor X_s at x. We provide closed form expressions of $\tilde{D}_{i,A}, \tilde{D}_{i,\Delta}, \tilde{D}^{(s)}_{i,A}, \tilde{D}^{(s)}_{i,\Delta}$ in the online supplement to this paper.

Assumption D. *(i) For both i and $s = 1, 2$, $\tilde{D}_{i,A}$, $\tilde{D}_{i,\Delta}$, $\tilde{D}^{(s)}_{i,A}$, $\tilde{D}^{(s)}_{i,\Delta}$ are continuous and bounded over the open set ω^o with $\omega \subseteq \omega^o$. (ii) The second order derivatives of $m_A^i(x;\boldsymbol{\gamma})$ and $m_\Delta^i(x;\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}(x)$ evaluated at $\boldsymbol{\gamma}^*(x)$ are bounded for all $x \in \omega^o$.*

Under S,W,D, there exists $\psi_\Delta^i \equiv (\psi_{\Delta,0}^i, \psi_{\Delta,1}^i, \psi_{\Delta,2}^i)$ such that:

$$\int w(x)\boldsymbol{\gamma}(x)'\tilde{D}_{i,\Delta}(x)dF_X^*(x) = \int \textstyle\sum_{k=0,1,2} \psi_{\Delta,k}^i(x)\gamma_k(x)dx \tag{23}$$

for any γ twice continuously differentiable in X_e. Similarly there exists $\psi_A^i \equiv (\psi_{A,0}^i, \psi_{A,1}^i, \psi_{A,2}^i)$ such that an equation similar to (23) holds with $\tilde{D}_{i,\Delta}$ replaced by $\tilde{D}_{i,A}$. The closed forms of ψ_A^i, ψ_Δ^i are provided in the online supplement. Define $\tilde{\nu}_\Delta^i(z) \equiv \psi_\Delta^i(x)d - E\left[\psi_\Delta^i(X)D\right]$ and define $\tilde{\nu}_A^i(z)$ similarly by replacing ψ_Δ^i in $\tilde{\nu}_\Delta^i(z)$ with ψ_A^i.

Assumption R. *For each i, (i) there exist constants $c_A^i, c_\Delta^i > 0$ such that $E\left[\sup_{\|\eta\| \leq c_A^i} \left\|\psi_{A,k}^i(x+\eta)\right\|^4\right] < \infty$ and $E\left[\sup_{\|\eta\| \leq c_\Delta^i} \left\|\psi_{\Delta,k}^i(x+\eta)\right\|^4\right] < \infty$ for $k = 0, 1, 2$; (ii) there exists an open neighborhood $\mathcal{N}$ around γ^* with $E[\sup_{\boldsymbol{\gamma} \in \mathcal{N}} ||w(X)m_\Delta^i(X;\boldsymbol{\gamma})||] < \infty$; and (iii) $E[||w(X)m_A^i(X;\boldsymbol{\gamma}^*) + \tilde{\nu}_A^i(Z)||^2] < \infty$ and $E[||w(X)[m_\Delta^i(X;\boldsymbol{\gamma}^*) - \Delta^i] + \tilde{\nu}_\Delta^i(Z)||^2] < \infty$.*

Theorem 4 *Suppose A1,2,3 and Assumptions S, K, B, W, D and R hold. Then for* $i = 1, 2$, $\Pr(\hat{\alpha}_i = \alpha_i) \rightarrow 1$ *and* $\sqrt{G}(\hat{\delta}_i - \delta_i) \xrightarrow{d} \mathcal{N}(0, \Sigma^i_\Delta)$, *where*

$$\Sigma^i_\Delta = Var\left\{w(X)[m^i_\Delta(X; \boldsymbol{\gamma}^*) - \Delta^i] + \tilde{\nu}^i_\Delta(Z)\right\}/\rho_0^2.$$

Proof of Theorem 4 follows from the steps in Section 8 of Newey and McFadden (1994), and is provided in the online supplement to this paper. The estimator for the interaction effect $\hat{\delta}_i$ takes the form of a sample moment involving sufficiently regular preliminary nonparametric estimates, and converges at the parametric rate.

5.2 Estimation of β_i

We derive asymptotic properties of the estimator for β_i treating α_i as known. As seen in the previous section, α_i can be estimated at an arbitrarily fast rate, so first order asymptotics for estimating β_i are unaffected by estimation error in α_i. Furthermore, in some applications, α_i is known a priori to the econometrician. One example is an entry game where an increase in X_i (components of fixed costs) reduces profits. Without loss of generality, let $\alpha_i = -1$ in what follows as would be the case in an entry game.

Let $v^l_i < v^h_i$ be points chosen on the support of V_i given $x_{-i} = ((x_j)_{j\neq i}, \tilde{x})$. The dependence of v^l_i, v^h_i on x_{-i} is suppressed to simplify the notation below. Choices of v^l_i, v^h_i are feasible as the support of V_i given x_{-i} can be estimated by the infimum and the supremum of $-t + \hat{\delta}_i\hat{p}_{j,i}(t, x_{-i})$ for t on the support of X_i given x_{-i}. We take v^l_i, v^h_i as known while deriving the asymptotic distribution of $\hat{\beta}_i$. Define a smooth distribution by

$$H(v) \equiv \mathcal{K}\left(2\left(\frac{v - v^l_i}{v^h_i - v^l_i}\right) - 1\right) \text{ for all } v \in [v^l_i, v^h_i],$$

where $\mathcal{K}(u) = 0$ if $u < -1$; $\mathcal{K}(u) = 1$ if $u > 1$; and $\mathcal{K}(u) = \frac{1}{16}(8 + 15u - 10u^3 + 3u^5)$ for $-1 \leq u \leq 1$. Note $\mathcal{K}$ is continuously differentiable with its derivative $\mathcal{K}'(u) = 0$ if $|u| > 1$, and $\mathcal{K}'(u) = \frac{15}{16}(1 - u^2)^2$ otherwise. Note H depends on x_{-i} through its support $[v^l_i, v^h_i]$, and is symmetric around its expectation $\mu \equiv \frac{1}{2}(v^h_i + v^l_i)$. Such a choice of H satisfies the required conditions stated in Section 3.4.

For $i = 1, 2$ and $j \equiv 3 - i$, define:

$$\widehat{\mathcal{H}}_i(x) \equiv H(-x_i + \hat{\delta}_i\hat{p}_j(x)) \text{ and } \widehat{\mathcal{V}}_{(i)}(x) \equiv 1 - \hat{\delta}_i\hat{p}_{j,i}(x)$$

where $\hat{p}_j$ and $\hat{p}_{j,i}$ are kernel estimates for p^*_j and $p^*_{j,i}$ respectively. We use $\widehat{\mathcal{H}}_i(x)$ as an estimator for $H(V_i(x))$ and $\widehat{\mathcal{V}}_{(i)}(x)$ as an estimator for $-\partial V_i(x)/\partial X_i$ when $\alpha_i = -1$. For each game indexed by

g, define

$$\hat{y}_{g,i} \equiv \frac{\left[d_{g,i} - \widehat{\mathcal{H}}_i(x_g)\right] \widehat{\mathcal{V}}_{(i)}(x_g) \int \hat{\gamma}_0(t, x_{g,-i})dt}{\hat{\gamma}_0(x_g)}.$$

Our estimator for β_i is given by $\hat{\beta}_i \equiv \left(\sum_g \tilde{x}_g \tilde{x}_g'\right)^{-1} \left(\sum_g \tilde{x}_g \left(\hat{y}_{g,i} - \mu\right)'\right)$.

To write $\hat{\beta}_i$ in terms of sample moments, define:

$$m_B^i\left(z; \delta_i, \boldsymbol{\gamma}\right) \equiv \tilde{x}' \left\{ \left[d_i - H\left(-x_i + \delta_i p_j(x)\right)\right] \frac{1-\delta_i p_{j,i}(x)}{f_{X_i|x_{-i}}(x_i)} - \mu \right\}$$

with $z \equiv (d_1, d_2, x)$ and $f_{X_i|x_{-i}}, p_j, p_{j,i}$ determined by $\boldsymbol{\gamma}$. Then by definition

$$\hat{\beta}_i = \left(\sum_g \tilde{x}_g \tilde{x}_g'\right)^{-1} \sum_g m_B^i\left(z_g; \hat{\delta}_i, \hat{\boldsymbol{\gamma}}\right).$$

Similar to the case with $\hat{\delta}_i$, we derive the limiting distribution of $\hat{\beta}_i$ using the steps in Newey and McFadden (1994). Under the assumptions for Corollary 2 (A1, 2, 3, 4') and some regularity conditions analogous to those in Section 5.1, $\sqrt{G}(\hat{\beta}_i - \beta_i) \xrightarrow{d} N(0, \Sigma_B^i)$ for $i = 1, 2$, where

$$\Sigma_B^i \equiv \left[E\left(\tilde{X}\tilde{X}'\right)^{-1}\right] Var\left[\Psi_B^i(Z) - \tilde{X}\tilde{X}'\beta_i\right] \left[E\left(\tilde{X}\tilde{X}'\right)^{-1}\right]$$

where $\Psi_B^i(z)$ consists of the moment function $m_B^i\left(z; \delta_i, \boldsymbol{\gamma}^*\right)$ and two additive terms that are functions of z and that correct the estimation errors in $\hat{\delta}_i$ and $\hat{\boldsymbol{\gamma}}$ respectively.

Among the regularity conditions for root-N convergence of $\hat{\beta}_i$ is a condition that the density of X_i given X_{-i} is bounded away from zero over the support of X. With excluded regressors distributed over a bounded support, the large support condition in A4' requires that the support of reparametrized errors ϵ_i given $\tilde{x}$ is bounded. All regularity conditions and the proof for the asymptotic property of $\hat{\beta}_i$ (including the form of correction terms in $\Psi_B^i(z)$) are included in the online supplement to this paper.

We conclude this subsection with several remarks about the relation between the rate of convergence of $\hat{\beta}_i$ and these support restrictions. Lewbel (2000) shows for single-agent binary decision models that the linear coefficient estimators using special regressors can converge at parametric rates, even when the support of the error term and the support of the special regressor are unbounded. Khan and Tamer (2010) shows in this case the distribution of a special regressor would need to have sufficiently thick tails to make its variance infinite.[18] In our case, if the support of

[18]Specifically, Khan and Tamer (2010) showed for a single-agent, binary regression with a special regressor that the semiparametric efficiency bound for linear coefficients is not finite when the second moment of all regressors (including the special regressors) is finite. For a simplified model $Y = 1\{\alpha + v + \varepsilon \geq 0\}$ (where $v \perp \varepsilon$, α is a constant and both v, ε are distributed over the real-line), they showed the rate of convergence for an inverse-density-weighted estimator can vary between $N^{-1/4}$ and the parametric rate, depending on the relative tail behaviors of v and ε.

excluded regressors is allowed to be unbounded, then our estimator needs to be modified to incorporate asymptotic trimming or some similar device to deal with the diminishing denominator in Y_i as V_i takes extreme values in the tails.

More generally, without bounded support, the rate of convergence of $\hat{\beta}_i$ depends on the tail behavior of the distributions of V_i given X_{-i}. Robust inference of $\hat{\beta}_i$ independent of the tail behaviors might be conducted in our context using a "rate-adaptive" approach as discussed in Andrews and Schafgans (1993) and Khan and Tamer (2010). This would involve performing inference on a studentized version of $\hat{\beta}_i$. More speculatively, it may also be possible to attain parametric rates of convergence without these tail thickness constraints by adapting some version of tail symmetry as in Magnac and Maurin (2007) to our game context. We leave these topics to future research.

6 Monte Carlo Simulation

In this section we present evidence for the performance of our estimators in 2-by-2 entry games. For $i = 1, 2$, the payoff for Firm i from entry ($D_i = 1$) is $\beta_i^0 + \beta_i^1 \tilde{X} - X_i + \delta_i D_{3-i} - \epsilon_i$, where D_{3-i} is the entry decision of i's competitor. The payoff from staying out is 0 for both $i = 1, 2$. The non-excluded regressor $\tilde{X}$ is discrete with $\Pr(\tilde{X} = 1/2) = \Pr(\tilde{X} = 1) = 1/2$. The true parameters in the data-generating process (DGP) are $[\beta_1^0, \beta_1^1] = [1.8, 0.5]$, $[\beta_2^0, \beta_2^1] = [1.6, 0.8]$ and $[\delta_1, \delta_2] = [-1.3, -1.3]$. For both $i = 1, 2$, the support of the observable part of the fixed costs (i.e. excluded regressors X_i) is $[0, 5]$ and the supports for the unobservable part of the fixed costs (i.e. ϵ_i) is $[-2, 2]$. All variables $(\tilde{X}, X_1, X_2, \epsilon_1, \epsilon_2)$ are mutually independent. To illustrate how the semiparametric estimator performs under various distributions of unobservable states, we experiment with two designs of the DGP: one in which both X_i and ϵ_i are uniformly distributed (*Uniform Design*); and one in which X_i is distributed with symmetric, bell-shaped density $f_{X_i}(t) = \frac{3}{8}(1 - (\frac{2t}{5} - 1)^2)^2$ over $[0, 5]$ while ϵ_i is distributed with symmetric, bell-shaped density $f_{\epsilon_i}(t) = \frac{15}{32}(1 - \frac{t^2}{4})^2$ over $[-2, 2]$. That is, both X_i, ϵ_i are linear transformation of a random variable whose density is given by the quartic (bi-weight) kernel, so we will refer to the second design as the BWK design. By construction, the conditional independence, additive separability, large support, and monotonicity conditions are all satisfied by these designs.

For each design, we simulate $S = 300$ samples and calculate summary statistics from empirical distributions of estimators $\hat{\delta}_i$ across these samples. These statistics include the mean, standard deviation (*Std.Dev*), 25% percentile (*LQ*), median, 75% percentile (*HQ*), root of mean squared error (*RMSE*) and median of absolute error (*MAE*). Both *RMSE* and *MAE* are estimated using the empirical distribution of estimators and the knowledge of our true parameters in the design.

Table 1(a) reports performance of $(\hat{\delta}_1, \hat{\delta}_2)$ under the uniform design. The two statistics reported in each cell of Table 1(a) correspond to those of $[\hat{\delta}_1, \hat{\delta}_2]$ respectively. Each row of Table 1(a) reports corresponds to a different combination of sample size (either $G = 1500$ or $G = 3000$) and choice of bandwidth σ for estimating p_i^*, p_j^* and their partial derivatives w.r.t. X_i, X_j. We use the tri-weight kernel function (i.e. $K(t) = \frac{35}{32}(1-t^2)^3 1\{|t| \leq 1\}$) in estimation. We construct a bandwidth b through cross-validation by minimizing the *Expectation of Average Squared Error* (MASE) in the estimation of p_i^* and p_j^*. This is done by choosing the bandwidths that minimize the *Estimated Predication Error* (EPE) for estimating p_i^*, p_j^*.[19] We then do estimation using the bandwidth $\sigma = b$, and also report summary statistics based on under-smoothing (with a bandwidth $\sigma = b/2$) and over-smoothing (with a bandwidth $\sigma = 3b/2$). To implement our estimator, we estimate the *n.d.s.* set of states ω by $\{x : \hat{p}_i(x),\ \hat{p}_j(x) \in (0,1),\ \hat{p}_{i,i}(x),\ \hat{p}_{j,j}(x) \neq 0 \textit{ and } \hat{p}_{i,i}(x)\hat{p}_{j,j}(x) \neq \hat{p}_{i,j}(x)\hat{p}_{j,i}(x)\}$.

Table 1(a) shows that the RMSE of estimation diminishes as the sample size G increases, and the MAE also generally improves. The trade-off between variance and bias of the estimator under different choices of the bandwidth is evident from the first two columns in Table 1(a). The choice of bandwidth seem to have a moderate effect on estimator performance.

[Insert Table 1(a) and Table 1(b) here.]

Table 1(b) reports the same statistics for the BWK design where both X_i, ϵ_i are distributed with bell-shaped densities. Again, there is strong evidence for convergence of the estimator in terms of RMSE. A comparison between panels (a) and (b) in Table 1 suggests the estimators for δ_i, δ_j perform better (most clearly in terms of RMSE) under the BWK design. This difference is due in part to the fact that, for a given sample size, the BWK design puts less probability mass towards tails of the density of X_i. To see this, note that by construction, the large support property of the model implies p_i, p_j could hit the boundaries (0 and 1) for large or small values of X_i, X_j in the tails. Therefore states with such extreme values of x_i, x_j are likely to be trimmed as we estimate δ_i, δ_j. The uniform distribution assigns higher probability mass towards the tails than bi-weight kernel densities. Thus, for a fixed N, the uniform design tends to trim away more observations than BWK design while estimating δ_i, δ_j. Note that while these trimmed-out tail observations do not contribute towards the estimation of $\hat{\delta}_i, \hat{\delta}_j$, they do contribute to attaining the parametric rate convergence of $\hat{\beta}_i^0, \hat{\beta}_i^1$.

Next, we report performance of estimators $(\hat{\beta}_i^0, \hat{\beta}_i^1)_{i=1,2}$ in Table 2, where we experiment with the same set of bandwidths as in Table 1. In addition, we report in Table 2 the performance of an "infeasible version" of our estimator where the preliminary estimates for p_i, p_j and their partial derivatives with respect to excluded regressors are replaced by true values in the DGP. Panels (a) and (b) show summary statistics of $S = 300$ simulated samples under the uniform design.

[Insert Table 2(a) and Table 2(b) here.]

[19] See Page 119 of Pagan and Ullah (1999) for the definitions of EPE and MASE.

Similar to Table 1, these panels in Table 2 show some evidence that the estimators converge (with RMSE diminishing) as the sample size increases. In both panels in Table 2, the choice of bandwidths seems to have impact on estimator performance. The impact of bandwidth on the variances of $\hat{\beta}_i^0, \hat{\beta}_i^1$ is also evident. The performance of the feasible estimators are roughly comparable to that of the infeasible estimators in terms of RMSE and MAE, especially in a larger sample. This is reassuring, though not unexpected because the final estimator for β_i^0, β_i^1 entails further averaging of sample moments involving preliminary estimates.

The two panels (a) and (b) in Table 3 report estimator performance for the BWK design, which exhibit similar patterns to Table 2. The performance of $(\hat{\beta}_i^0, \hat{\beta}_i^1)$ under the BWK design is better than that in the uniform design in terms of both RMSE and MAE. The uniform design has more observations in the data that are closer to the boundary of supports than the BWK, which, for a given accuracy of first stage estimates, would tend to make estimation of β^0, β_i^1 more accurate under the uniform design. However, this effect appears to be offset by the impact of distributions of X_i, ϵ_i on first-stage estimators $\hat{\delta}_i, \hat{\delta}_j$, which were more precisely estimated under the BWK design.

[Insert Table 3(a) and Table 3(b) here.]

7 Multiple Equilibria

So far we have maintained that the choices reported in the data are rationalized by a single BNE in all states (as assumed in A2). Now suppose A2 is removed and multiple BNE can arise for some states in the data. Then, from the perspective of an econometrician, the set of *all* states where the choices are rationalized by a single equilibrium, denoted ω^* henceforth, becomes an unknown (and potentially strict) subset of the support of X. In this section we extend our identification and estimation strategies to allow for the possibility of multiple BNE in the data. The main idea is to condition the identification argument in Section 3 on sets of states that belong to ω^*.

7.1 Identification under multiple equilibria

Let ω' denote a subset of ω^*. Note ω' (and ω^*) includes states where the system of equations (1) admits a unique solution, and states where (1) admits multiple solutions but the equilibrium selection mechanism is degenerate at only one of them. For now, we assume such a set ω' is known. Later, in Section 7.2, we show how one could test the requirement that, in a given candidate set ω' choices are almost everywhere rationalized by a single BNE, as required.

We first modify our single equilibrium identification of α_i and $\delta_i(\tilde{x})$ to handle data containing multiple BNE, by conditioning our earlier arguments in Section 3.2 on a set $\omega' \subseteq \omega^*$. Specifically, suppose assumptions A1 and A3 hold, and ω' is open and *n.d.s.*. It follows that α_i is identified for all i as in Theorem 1, only with ω in that theorem replaced by ω'. Besides, for all $\tilde{x}$ such that $\Pr(X \in \omega' \mid \tilde{X} = \tilde{x}) > 0$, $\delta_i(\tilde{x})$ is also identified for all i as in Theorem 1, again with ω therein replaced by ω'. In the online supplement to this paper, we present sufficient conditions for the existence of such a n.d.s. set of states with a single equilibrium ω'. Essentially these conditions require the density of $\mathcal{S}_i$ given $\tilde{x}$ to be bounded away from zero over intervals between $\alpha_i x_i + \delta_i(\tilde{x})$ and $\alpha_i x_i$ for any x_e such that $(x_e, \tilde{x}) \in \omega'$.[20]

Next, the argument for identifying the mean baseline payoff $\tilde{u}_i(\tilde{x})$ in Section 3.4 can be extended similarly, provided excluded regressors X_e vary sufficiently within the set of states with a single BNE ω^*. To see this, let $\alpha_i, \delta_i(\tilde{x})$ be identified at $\tilde{x}$, and suppose there exists $x_{e,-i} \equiv (x_j)_{j \neq i}$ and a connected set ω_i such that all states in the set $\{(x_i, x_{e,-i}, \tilde{x}) : x_i \in \omega_i\} \subseteq \omega^*$. Also assume: *(i)* the support of V_i given $X_i \in \omega_i$ and $X_{-i} = x_{-i}$ includes the support of $\mathcal{S}_i$ given $\tilde{x}$; *(ii)* the density of X_i given $X_i \in \omega_i$ and $X_{-i} = x_{-i}$ is positive almost everywhere on ω_i; and *(iii)* the sign of $\partial V_i(X)/\partial X_i|_{X=(x_i,x_{-i})}$ is identical to the sign of α_i for all $x \in \{(x_i, x_{e,-i}, \tilde{x}) : x_i \in \omega_i\}$. It is important to note that this set does not have to be a subset of ω' that was used for identifying $\alpha_i, \delta_i(\tilde{x})$ in the paragraph above.

Define $\tilde{Y}_i$ by replacing $f_{X_i|X_{-i}=x_{-i}}$ with the density of X_i conditional on $X_{-i} = x_{-i}$ and $X_i \in \omega_i$ in the definition of Y_i in (15). Choose a continuous distribution H as in (15), but now require it to have an interval support contained in the support of V_i conditional on "$X_{-i} = x_{-i}$ and $X_i \in \omega_i$". Let μ denote the expectation under the distribution H as before. Then, under A1 and A3, an argument similar to that in Section 3.4 shows

$$\tilde{u}_i(\tilde{x}) = E\left[\tilde{Y}_i - \mu \mid X_i \in \omega_i, X_{-i} = x_{-i}\right]. \tag{24}$$

The main idea here is that, once we condition on $X_i \in \omega_i$ and $X_{-i} = x_{-i}$, the link between the model elements and the distribution of states and choices from the data is summarized by (1).

Just as in Section 3.4, it is possible to exploit the over-identification in (24) to improve the efficiency in estimation by applying the Law of Iterated Expectation and integrating out $X_{e,-i}$ from the right-hand side of (24) conditional on $\tilde{X} = \tilde{x}$.

A sufficient condition for the existence of $x_{e,-i}$ and ω_i satisfying (i)-(iii) above is that the other excluded regressors $x_{e,-i}$ take on extremely large or small values so that the system (3) admits a unique solution.[21] Condition (iii) holds when both $\delta_i(\tilde{x})$ and the density of ϵ_i given $\tilde{x}$ are bounded

[20] This condition is sufficient for the system in (3) to admit only one solution for all $x \in \omega'$ when $h_i(D_{-i}) = \sum_{j \neq i} D_j$ and $u_i(x, \epsilon_i)$ is additively separable in x and ϵ_i for both i. It is stronger than what we need because there can be a unique BNE in the DGP even when the system admits multiple solutions at a state x.

[21] To see this, consider the example of entry game. Suppose the observed component of fixed costs, or excluded regressors, for all but one firm take on extremely large values. Then the probability of entry for all but one of the firms

above. We provide primitive conditions that imply the existence of such x_{-i} and ω_i in the online supplement to this paper.

7.2 Test for multiple equilibria

As discussed above, the identification of $\alpha_i, \delta_i(\tilde{x})$ under multiple equilibria requires a set $\omega' \subseteq \omega^*$; and that of $\tilde{u}_i(\tilde{x})$ needs to use a set of states from ω^*. We now show how to test the hypothesis that a generic set of states $\bar{\omega}$ is a subset of ω^*. For any $\bar{\omega}$ from the support of X, we modify the method in De Paula and Tang (2012) to propose a test for the null that almost all states in $\bar{\omega}$ belong to ω^*. To fix ideas we focus on the case with $N = 2$ and $\delta_i(\tilde{x}) = \delta_i$. Define:

$$T(\bar{\omega}) \equiv \; E\left[1\{X \in \bar{\omega}\}\left(D_1 D_2 - p_1^*(X)p_2^*(X)\right)\right].$$

The following lemma is a special case of Proposition 1 in De Paula and Tang (2012), which builds on earlier results in Manski (1993) and Sweeting (2009).

Lemma 1 *Suppose $\delta_i(\tilde{x}) = \delta_i \neq 0$ for all $\tilde{x}$ and $i = 1, 2$. Under A1,*

$$\Pr\{X \notin \omega^* \mid X \in \bar{\omega}\} > 0 \textit{ if and only if } T(\bar{\omega}) \neq 0 \tag{25}$$

The intuition for (25) is that, if players' private information are independent conditional on x, then their actions must be uncorrelated if the data are rationalized by a single BNE. Any correlations between actions in the data can only occur as players simultaneously move between strategies prescribed in different equilibria.

Lemma 1 implies the null and the alternative hypotheses

$$H_0 : \Pr\{X \in \omega^* \mid X \in \bar{\omega}\} = 1 \quad \textit{vs.} \quad H_A : \Pr\{X \in \omega^* \mid X \in \bar{\omega}\} < 1;$$

are equivalent to:

$$H_0 : T(\bar{\omega}) = 0 \quad \textit{vs.} \quad H_A : T(\bar{\omega}) \neq 0,$$

which can be tested using a test statistic based on the analog principle.

To simplify exposition, let $\tilde{X}$ be discrete. For a product kernel K and a sequence of bandwidth σ (with $\sigma \to 0$ as $G \to \infty$), define $K_\sigma(.) \equiv \sigma^{-2}K(./\sigma)$. Let $D_g \equiv (D_{g,0}, D_{g,1}, D_{g,2})$ where $D_{g,0} \equiv 1$,

are practically zero, yielding a unique BNE where the remaining one player with a non-extreme fixed cost component X_i enters if and only if his monopoly profit exceeds his fixed cost of entry. Again, note that $X_{e,-i}$ takes on extreme values is stronger than necessary. This is because the equilibrium selection mechanism may well be degenerate at a single BNE even when the states do not take on extreme values.

$X_g \equiv (X_{g,e}, \tilde{X}_g)$ and $X_{g,e} \equiv (X_{g,1}, X_{g,2})$ denote actions and states observed in independent games, each of which is indexed by g. Define the statistic

$$\hat{T}(\bar{\omega}) \equiv \frac{1}{G} \sum_{g \leq G} 1\{x_g \in \bar{\omega}\} \left[d_{1,g} d_{2,g} - \hat{p}_1(x_g) \hat{p}_2(x_g) \right],$$

where $\hat{p}_i$ are kernel estimates of propensity score p_i^* using $K_\sigma(.)$ as before. Under some mild regularity conditions,

$$\sqrt{G} \left(\hat{T}(\bar{\omega}) - \mu_{\bar{\omega}} \right) \xrightarrow{d} N\left(0, \Sigma_{\bar{\omega}}\right) \tag{26}$$

where $\mu_{\bar{\omega}} \equiv E\left[1\{X \in \bar{\omega}\} \left(D_1 D_2 - p_1^*(X) p_2^*(X)\right)\right]$, and the limiting covariance matrix $\Sigma_{\bar{\omega}}$ is

$$Var\left[1\{X \in \bar{\omega}\} \left(D_1 D_2 - p_1^*(X) p_2^*(X) + \frac{2 p_1^*(X) p_2^*(X) - D_1 p_2^*(X) - D_2 p_1^*(X)}{\gamma_0^*(X)} \right) \right],$$

where γ_0^* is defined in Section 5.1. The covariance matrix $\Sigma_{\bar{\omega}}$ is expressed in terms of functions that are directly identifiable from the data, and so can be consistently estimated by its sample analog under standard conditions. We provide details about the regularity conditions and the derivation of the limiting distribution in (26) in the online supplement to this paper.

8 Conclusion

We provide conditions for nonparametric identification of binary games with incomplete information, and introduce semiparametric estimators that converge at a parametric rate to the coefficients in linear payoffs. The point identification result can be extended when there are multiple Bayesian Nash equilibria in the data. We motivate the identifying conditions using several examples, and provide economic interpretation of these conditions in classical contexts such as market entry games between firms.

Our method uses excluded regressors, which are state variables that affect just one player's payoffs, and do so in an additively separable way. Like all payoff-related public information, excluded variables affect all players' choice probabilities. We show the interactive parts of the players' payoffs can be identified using variation in excluded regressors. Thus these parts can be constructed from the data, and then used to identify the remaining model components in a way similar to the use of special regressor for identifying single-agent models. Identification of the full model requires relatively large variation in excluded regressors to ensure the constructed special regressors also have a large support.

In the example of entry games, observed components of fixed costs are natural examples of excluded regressors. An obstacle to applying our results in practice is that data on components of fixed costs, while public, is often not collected, perhaps because existing empirical applications of entry games (which generally require far more stringent assumptions than our model) do not require observation of fixed cost components for identification. We hope our results will provide the incentive to collect such data in the future.

Appendix

Proof of Theorem 1. For any $\tilde{x}$, let $\mathcal{D}(\tilde{x})$ denote a N-vector with ordered coordinates $(\delta_1(\tilde{x}), \delta_2(\tilde{x}),$ .., $\delta_N(\tilde{x}))$. For any $x \equiv (x_e, \tilde{x})$, let $\mathcal{F}(x)$ denote a N-vector with the i-th coordinate being $f_{\mathcal{S}_i|\tilde{X}=\tilde{x}}(\alpha_i x_i + \delta_i(\tilde{x})\phi_i^*(x))$. Under assumptions A1-3, for any x and all i,

$$p_i^*(x) = \Pr\{\mathcal{S}_i \leq \alpha_i X_i + \delta_i(\tilde{X})\phi_i^*(X) \mid X = x\}. \tag{27}$$

Now for any x, fix $\tilde{x}$ and differentiate both sides of (27) with respect to X_i for all i. This gives:

$$\mathcal{W}_1(x) = \mathcal{F}(x). (\mathcal{A} + \mathcal{D}(\tilde{x}).\mathcal{V}_1(x)) \tag{28}$$

where "." denotes component-wise multiplication. Differentiate both sides of (27) w.r.t. X_{i+1} for $i \leq N-1$ and differentiate both sides of (27) w.r.t. X_1 for $i = N$. Stacking N equations resulted from such differentiation at x,

$$\mathcal{W}_2(x) = \mathcal{F}(x).\mathcal{D}(\tilde{x}).\mathcal{V}_2(x) \tag{29}$$

For all $x \in \omega$ (where ω is a n.d.s. set), none of the coordinates in $\mathcal{F}(x)$, $\mathcal{V}_2(x)$ and $\mathcal{V}_2(x).\mathcal{W}_1(x) - \mathcal{V}_1(x).\mathcal{W}_2(x)$ are zero. It then follows that for any $x \in \omega$,

$$\mathcal{A}.\mathcal{F}(x) = \mathcal{W}_1(x) - \mathcal{W}_2(x).\mathcal{V}_1(x)./\mathcal{V}_2(x); \tag{30}$$

$$\mathcal{D}(\tilde{x}) = \mathcal{A}.\mathcal{W}_2(x)./\left[\mathcal{V}_2(x).\mathcal{W}_1(x) - \mathcal{V}_1(x).\mathcal{W}_2(x)\right]. \tag{31}$$

Because $\mathcal{F}(x)$ is a vector with strictly positive coordinates under A1, the signs of α_i's are identical to the signs of coordinates on the right-hand side in (30) for any $x \in \omega$. Integrating out X over ω using the marginal distribution of X gives (9). (31) (over-)identifies $\mathcal{D}(\tilde{x})$ for all $x \in \omega$. Integrating out X over ω using the distribution of X conditional on $\tilde{x}$ and $X \in \omega$ gives (10). *Q.E.D.*

Proof of Corollary 2. To prove the corollary, it suffices to show that for any i and $x_{-i} = (x_{e,-i}, \tilde{x})$, the mean baseline payoff is (over-)identified as

$$\tilde{u}_i(\tilde{x}) = E(Y_i \mid X_{-i} = x_{-i}) - \mu. \tag{32}$$

The claim of the corollary would then follow from an application of the Law of Iterated Expectation. Throughout the proof, suppress x_{-i} in the notation for the support of V_i given $X_{-i} = x_{-i}$ (denoted $[\underline{v}_i, \overline{v}_i]$) and the expectation according to H (denoted μ). Then

$$E(Y_i \mid X_{-i} = x_{-i}) = \int \frac{[E(D_i|X=x) - H(V_i(x))][\alpha_i + \delta_i(\tilde{x})\phi_{i,i}^*(x)]\alpha_i}{f_{X_i|x_{-i}}(x_i)} f_{X_i|x_{-i}}(x_i) dx_i$$

$$= \int \left[E\left(D_i \mid V_i = V_i(x), \tilde{x}\right) - H(V_i(x))\right]\left[\alpha_i + \delta_i(\tilde{x})\phi_{i,i}^*(x)\right]\alpha_i dx_i \tag{33}$$

where the first equality follows from the law of iterated expectation and A4'-(iii); and the second from A3 and that $\epsilon_i \perp V_i$ given $\tilde{X} = \tilde{x}$. The integrand in (33) is continuous in V_i, and V_i as a function of X is continuously differentiable in X_i given $X_{-i} = x_{-i}$ under A1 and A3, with $\partial V_i(X)/\partial X_i|_{X=x} = \alpha_i + \delta_i(\tilde{x})\phi^*_{i,i}(x)$.

Consider the case with $\alpha_i = 1$. Then V_i is increasing in x_i given $X_{-i} = x_{-i}$ according to A4'-(ii). Changing variables between v_i and x_i while fixing x_{-i} gives:

$$E(Y_i \mid X_{-i} = x_{-i}) = \int_{\underline{v}_i}^{\overline{v}_i} [E(D_i \mid V_i = v, \tilde{x}) - H(v)]\, dv = \int_{\underline{v}_i}^{\overline{v}_i} [E(D_i \mid V_i = v, \tilde{x}) - 1\{v \geq \mu\}]\, dv,$$

where the second equality uses integration by parts and the properties of H. It also uses that μ is in the interior of the support of H. Then

$$\begin{aligned} E(Y_i \mid X_{-i} = x_{-i}) &= \int_{\underline{v}_i}^{\overline{v}_i} \left(\int_{\Omega_{\epsilon_i|\tilde{x}}} 1\{\varepsilon_i \leq \tilde{u}_i(\tilde{x}) + v\} dF_{\epsilon_i|\tilde{x}}(\varepsilon_i) - 1\{v \geq \mu\} \right) dv \\ &= \int_{\Omega_{\epsilon_i|\tilde{x}}} \left(\int_{\underline{v}_i}^{\overline{v}_i} 1\{v \geq s_i\} - 1\{v \geq \mu\}\, dv \right) dF_{\epsilon_i|\tilde{x}}(\varepsilon_i), \end{aligned}$$

where s_i is $-\tilde{u}_i(\tilde{x}) + \varepsilon_i$ and the second equality follows changing of the order of integration as allowed by the support condition in A4'-(i) and Fubini's Theorem. Then $E(Y_i \mid X_{-i} = x_{-i})$ becomes

$$\begin{aligned} & \int_{\Omega_{\epsilon_i|\tilde{x}}} \int_{\underline{v}_i}^{\overline{v}_i} 1\{s_i \leq v < \mu\} 1\{s_i \leq \mu\} - 1\{\mu < v \leq s_i\} 1\{s_i > \mu\} dv dF(\varepsilon_i|\tilde{x}) \\ = & \int_{\Omega_{\epsilon_i|\tilde{x}}} \left(1\{s_i \leq \mu\} \int_{s_i}^{\mu} dv - 1\{s_i > \mu\} \int_{\mu}^{s_i} dv \right) dF(\varepsilon_i|\tilde{x}) \\ = & \int_{\Omega_{\epsilon_i|\tilde{x}}} (\mu - s_i)\, dF(\varepsilon_i|\tilde{x}) = \mu + \tilde{u}_i(\tilde{x}). \end{aligned}$$

Thus (32) holds. If instead $\alpha_i = -1$ then V_i is decreasing in x_i given $X_{-i} = x_{-i}$ according to A4'-(ii). By construction $[1 + \delta_i(\tilde{x})\phi^*_{i,i}(x)]\alpha_i = -[\alpha_i + \delta_i(\tilde{x})\phi^*_{i,i}(x)] = -\frac{\partial V_i(X)}{\partial X_i}|_{X=x}$. Then (32) holds using a symmetric argument and changing variables between V_i and X_i. *Q.E.D.*

Proof of Lemma 1. Under A1,

$$x \in \omega^* \text{ if and only if } E[D_1 D_2 \mid X = x] = p^*_1(x) p^*_2(x); \tag{34}$$

and $sign(E[D_1 D_2 \mid X = x] - p^*_1(x)p^*_2(x)) = sign(\delta_i)$ whenever $x \notin \omega^*$. (See Proposition 1 in De Paula and Tang (2012).)

First, suppose $\Pr\{X \in \omega^* \mid X \in \bar{\omega}\} = 1$. Integrating out X in the equation in (34) given $X \in \bar{\omega}$ implies $E[D_1 D_2 - p^*_1(X)p^*_2(X) \mid X \in \bar{\omega}] = 0$, which in turn implies $T(\bar{\omega}) = 0$.

Next, suppose $\Pr\{X \notin \omega^* \mid X \in \bar{\omega}\} > 0$, which necessarily requires $sign(\delta_1) = sign(\delta_2)$. (Otherwise, there can't be multiple BNE at any state x.) The set of states in $\bar{\omega}$ can be partitioned into its intersection with ω^* and its intersection with the complement of ω^* in the support of X. For all x in the intersection of $\bar{\omega}$ and the complement of ω^*, $E[D_1D_2 \mid X = x] - p_1^*(x)p_2^*(x)$ is non-zero and its sign equals $sign(\delta_i)$ for $i = 1, 2$. For all x in the intersection of $\bar{\omega}$ and ω^*, $E[D_1D_2 \mid X = x] - p_1^*(x)p_2^*(x)$ equals 0. Applying the Law of Iterated Expectations based on these partitioning intersections shows $E\left[D_1D_2 - p_1^*(X)p_2^*(X) \mid X \in \bar{\omega}\right] \neq 0$ and its sign equals $sign(\delta_i)$. This implies $T(\bar{\omega}) \neq 0$ and $sign(T(\bar{\omega})) = sign(\delta_i)$ for $i = 1, 2$. *Q.E.D.*

References

[1] Aguirregabiria, V. and P. Mira, 2005. A Hybrid Genetic Algorithm for the Maximum Likelihood Estimation of Models with Multiple Equilibria: A First Report. New Mathematics and Natural Computation, 1(2), pp: 295-303.

[2] Aradillas-Lopez, A., 2010. Semiparametric Estimation of a Simultaneous Game with Incomplete Information. Journal of Econometrics, Volume 157 (2), pp: 409-431

[3] Andrews, D.K. and M. Schafgans, 1998. Semiparametric Estimation of the Intercept of a Sample Selection Model. Review of Economic Studies, Vol. 65, pp:497-517

[4] Bajari, P., J. Hahn, H. Hong and G. Ridder, 2011. A Note on Semiparametric Estimation of Finite Mixtures of Discrete Choice Models with Application to Game Theoretic Models. International Economic Review, Vol 52, 807-824.

[5] Bajari, P., Hong, H., Krainer, J. and D. Nekipelov, 2010. Estimating Static Models of Strategic Interactions. Journal of Business and Economic Statistics 28(4), 469-482

[6] Bajari, P., Hong, H. and S. Ryan, 2010. Identification and Estimation of a Discrete Game of Compete Information. Econometrica, Vol. 78, No. 5, 1529-1568.

[7] Beresteanu, A., Molchanov, I. and F. Molinari, 2011. Sharp Identification Regions in Models with Convex Moment Predictions. Econometrica, Vol 166, 1785-1821

[8] Berry, S., 1992. Estimation of a Model of Entry in the Airline Industry. Econometrica, Vol. 60, No. 4., pp. 889-917

[9] Berry, S. and P. Haile, 2014. Identification in Differentiated Products Markets Using Market Level Data. Econometrica, Vol 82, 1749-1797

[10] Berry, S. and P. Haile, 2009. Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers. Cowles Foundation Discussion Paper No.1718

[11] Blume, L., Brock, W., Durlauf, S. and Y. Ioannides, 2010. Identification of Social Interactions. Handbook of Social Economics, J. Benhabib, A. Bisin, and M. Jackson, eds., Amsterdam: North Holland, 2010.

[12] Bierens, H., 1987. Kernel Estimators of Regression Functions, in: Truman F.Bewley (ed.), Advances in Econometrics: Fifth World Congress, Vol.I, New York: Cambridge University Press, 99-144.

[13] Brock, W., and S. Durlauf, 2007. Identification of Binary Choice Models with Social Interactions. Journal of Econometrics, 140(1), pp 52-75

[14] Byrne, D., Goeree, M., Hiedemann, B. and S. Stern, 2009. Formal Home Health Care, Informal Care and Family Decision Making. International Economic Review, Vol. 50, No. 4

[15] Checkovich, T. and S. Stern, 2002. Shared Caregiving Responsibilities of Adult Siblings with Elderly Parents. Journal of Human Resources, Vol. 37, No.3, pp. 441-478

[16] Ciliberto, F., and E. Tamer, 2009. Market Structure and Multiple Equilibria in Airline Markets. Econometrica, Vol. 77, No. 6, 1791–1828

[17] De Paula, A. and X. Tang, 2012. Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information. Econometrica, Vol. 80, Issue 1.

[18] Dong, Y. and A. Lewbel, 2015. A Simple Estimator for Binary Choice Models with Endogenous Regressors. Econometric Reviews, Vol. 34, 82-105

[19] Galichon, A. and M. Henry, 2011. Set Identification in Models with Multiple Equilibria. Review of Economic Studies, Vol. 78, No. 4, pp 1264-1298

[20] Haile, P., Hortaçsu, A. and G. Kosenok, 2008. On the Empirical Content of Quantal Response Equilibrium. American Economic Review, 98(1), pp 180-200

[21] Khan, S. and D. Nekipelov, 2011. Information Structure and Statistical Information in Discrete Response Models. Working paper, Duke University and UC Berkeley

[22] Khan, S. and E. Tamer, 2010. Irregular Identification, Support Conditions and Inverse Weight Estimation. Econometrica, Vol 78, 2021-2042

[23] Klein, R., and W. Spady, 1993. An Efficient Semiparametric Estimator of Binary Response Models. Econometrica, 61, 387–421, 1993

[24] Magnac, T. and E. Maurin, 2007. Identification and Information in Monotone Binary Models. Journal of Econometrics, Vol 139, pp 76-104

[25] Magnac, T. and E. Maurin, 2008. Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data. Review of Economic Studies, Vol 75, pp 835-864

[26] Lewbel, A., 1998. Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors. Econometrica, Vol.66, pp.105-121

[27] Lewbel, A., 2000. Semiparametric Qualitative Response Model Estimation with Unknown Heteroskaedasticity and Instrumental Variables. Journal of Econometrics, Vol 97, 145-177

[28] Lewbel, A., 2014. An Overview of the Special Regressor Method, in the Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics, Co-edited by Aman Ullah, Jeffrey Racine, and Liangjun Su, Oxford University Press, pp. 38-62.

[29] Lewbel, A., O. Linton, and D. McFadden, 2011. Estimating Features of a Distribution From Binomial Data. Journal of Econometrics, 162, 170-188

[30] Lewbel, A. and X. Tang, 2014. Supplement for "Identification and Estimation of Games with Incomplete Information Using Excluded Regressors". Unpublished manuscript, Boston College and U Penn.

[31] Manski, C., 1993. Identification of Endogenous Social Effects: The Selection Problem. Review of Economic Studies, 60(3), pp 531-542.

[32] Newey, W. and D. McFadden, 1994. Large Sample Estimation and Hypothesis Testing, Handbook of Econometrics, Vol 4, ed. R.F.Engle and D.L.McFadden, Elsevier Science

[33] Pagan, A. and A. Ullah, 1999. Nonparametric Econometrics. Cambridge University Press.

[34] Seim, K., 2006. An Empirical Model of Firm Entry with Endogenous Product-Type Choices. The RAND Journal of Economics, 37(3), pp 619-640.

[35] Sweeting, A., 2009. The Strategic Timing of Radio Commercials: An Empirical Analysis Using Multiple Equilibria. RAND Journal of Economics, 40(4), 710-742

[36] Tamer, E., 2003. Incomplete Simultaneous Discrete Response Model with Multiple Equilibria. Review of Economic Studies, Vol 70, No 1, pp. 147-167.

[37] Tang, X., 2010. Estimating Simultaneous Games with Incomplete Information under Median Restrictions. Economics Letters, 108(3)

[38] Wan, Y., and H. Xu, 2014. Semiparametric Identification and Estimation of Binary Decision Games of Incomplete Information with Correlated Private Signals. Journal of Econometrics, Vol 182, 235-246

[39] Xu, H., 2014. Estimation of Discrete Game with Correlated Private Signals. The Econometrics Journal, forthcoming.

Table 1(a): Estimator for (δ_1, δ_2) (*Uniform Design*)

G	σ	*Mean*	*Std.Dev.*	*LQ*	*Median*	*HQ*	*RMSE*	*MAE*
1500	b	[-1.488,-1.480]	[0.303,0.299]	[-1.698,-1.661]	[-1.497,-1.469]	[-1.290,-1.291]	[0.357,0.349]	[0.253,0.217]
	$\frac{b}{2}$	[-1.465,-1.497]	[0.497,0.462]	[-1.769,-1.784]	[-1.470,-1.518]	[-1.107,-1.221]	[0.524,0.502]	[0.337,0.319]
	$\frac{3b}{2}$	[-1.563,-1.561]	[0.268,0.251]	[-1.716,-1.713]	[-1.546,-1.547]	[-1.365,-1.381]	[0.375,0.362]	[0.267,0.262]
3000	b	[-1.452,-1.423]	[0.214,0.212]	[-1.580,-1.551]	[-1.445,-1.411]	[-1.310,-1.284]	[0.262,0.245]	[0.176,0.156]
	$\frac{b}{2}$	[-1.447,-1.449]	[0.367,0.362]	[-1.668,-1.685]	[-1.437,-1.432]	[-1.201,-1.212]	[0.395,0.391]	[0.288,0.211]
	$\frac{3b}{2}$	[-1.508,-1.482]	[0.179,0.199]	[-1.651,-1.616]	[-1.483,-1.467]	[-1.375,-1.349]	[0.274,0.270]	[0.187,0.188]

Table 1(b): Estimator for (δ_1, δ_2) (*BWK Design*)

G	σ	*Mean*	*Std.Dev.*	*LQ*	*Median*	*HQ*	*RMSE*	*MAE*
1500	b	[-1.415,-1.409]	[0.201,0.191]	[-1.548,-1.532]	[-1.421,-1.408]	[-1.278,-1.265]	[0.232,0.220]	[0.161,0.142]
	$\frac{b}{2}$	[-1.406,-1.412]	[0.231,0.257]	[-1.559,-1.594]	[-1.415,-1.406]	[-1.258,-1.228]	[0.255,0.280]	[0.156,0.199]
	$\frac{3b}{2}$	[-1.496,-1.503]	[0.165,0.172]	[-1.607,-1.622]	[-1.492,-1.505]	[-1.374,-1.376]	[0.256,0.266]	[0.163,0.172]
3000	b	[-1.394,-1.379]	[0.129,0.132]	[-1.484,-1.471]	[-1.380,-1.378]	[-1.307,-1.287]	[0.160,0.154]	[0.101,0.093]
	$\frac{b}{2}$	[-1.364,-1.368]	[0.164,0.159]	[-1.475,-1.465]	[-1.367,-1.364]	[-1.255,-1.257]	[0.176,0.173]	[0.122,0.126]
	$\frac{3b}{2}$	[-1.428,-1.472]	[0.118,0.121]	[-1.503,-1.505]	[-1.434,-1.474]	[-1.349,-1.368]	[0.174,0.210]	[0.129,0.157]

Table 2(a): Estimator for (β_i^0, β_i^1) (Uniform Design, $G = 1500$)

		Mean	*Std.Dev.*	*LQ*	*Median*	*HQ*	*RMSE*	*MAE*
Infeasible	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.954,0.640]	[0.245,0.269]	[1.779,0.450]	[1.936,0.628]	[2.106,0.820]	[0.289,0.303]	[0.181,0.224]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.736,0.940]	[0.290,0.277]	[1.548,0.747]	[1.734,0.936]	[1.944,1.138]	[0.320,0.310]	[0.209,0.231]
$\sigma = b$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[2.059,0.686]	[0.316,0.314]	[1.796,0.487]	[2.049,0.675]	[2.279,0.885]	[0.408,0.364]	[0.256,0.249]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.775,0.991]	[0.356,0.340]	[1.550,0.769]	[1.762,0.998]	[2.006,1.219]	[0.396,0.390]	[0.267,0.275]
$\sigma = \frac{b}{2}$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[2.039,0.674]	[0.455,0.464]	[1.749,0.366]	[2.027,0.657]	[2.358,0.969]	[0.514,0.495]	[0.311,0.339]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.757,0.986]	[0.482,0.418]	[1.450,0.686]	[1.743,0.999]	[2.058,1.262]	[0.507,0.457]	[0.368,0.327]
$\sigma = \frac{3b}{2}$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[2.129,0.749]	[0.247,0.288]	[1.959,0.549]	[2.136,0.747]	[2.296,0.952]	[0.411,0.381]	[0.292,0.238]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.861,1.059]	[0.340,0.316]	[1.639,0.841]	[1.844,1.070]	[2.103,1.271]	[0.428,0.409]	[0.310,0.261]

Table 2(b): Estimator for (β_i^0, β_i^1) (Uniform Design, $G = 3000$)

		Mean	*Std.Dev.*	*LQ*	*Median*	*HQ*	*RMSE*	*MAE*
Infeasible	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.939,0.573]	[0.139,0.219]	[1.849,0.433]	[1.941,0.506]	[2.033,0.711]	[0.197,0.231]	[0.142,0.172]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.654,0.880]	[0.231,0.203]	[1.548,0.746]	[1.663,0.882]	[1.800,1.015]	[0.237,0.218]	[0.152,0.139]
$\sigma = b$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.989,0.599]	[0.227,0.245]	[1.826,0.443]	[1.989,0.521]	[2.151,0.774]	[0.296,0.264]	[0.202,0.188]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.688,0.902]	[0.269,0.257]	[1.546,0.722]	[1.684,0.895]	[1.870,1.073]	[0.283,0.277]	[0.193,0.178]
$\sigma = \frac{b}{2}$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.938,0.565]	[0.312,0.398]	[1.734,0.354]	[1.950,0.497]	[2.158,0.830]	[0.341,0.403]	[0.222,0.302]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.672,0.896]	[0.380,0.341]	[1.413,0.674]	[1.661,0.910]	[1.919,1.137]	[0.388,0.354]	[0.288,0.251]
$\sigma = \frac{3b}{2}$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[2.201,0.635]	[0.171,0.241]	[1.908,0.481]	[2.023,0.548]	[2.142,0.790]	[0.279,0.276]	[0.182,0.196]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.745,0.965]	[0.289,0.214]	[1.525,0.814]	[1.735,0.958]	[1.904,1.116]	[0.289,0.270]	[0.204,0.189]

Table 3(a): Estimator for (β_i^0, β_i^1) (BWK Design, $G = 1500$)

		Mean	*Std.Dev.*	*LQ*	*Median*	*HQ*	*RMSE*	*MAE*
Infeasible	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.917,0.606]	[0.205,0.257]	[1.784,0.432]	[1.926,0.595]	[2.058,0.771]	[0.236,0.278]	[0.146,0.171]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.703,0.887]	[0.223,0.194]	[1.556,0.754]	[1.694,0.885]	[1.842,1.016]	[0.245,0.213]	[0.161,0.142]
$\sigma = b$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.953,0.651]	[0.238,0.288]	[1.791,0.467]	[1.944,0.641]	[2.117,0.850]	[0.283,0.325]	[0.207,0.218]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.746,0.934]	[0.258,0.239]	[1.575,0.773]	[1.758,0.933]	[1.912,1.096]	[0.297,0.274]	[0.193,0.175]
$\sigma = \frac{b}{2}$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.946,0.633]	[0.376,0.425]	[1.692,0.368]	[1.964,0.622]	[2.199,0.926]	[0.403,0.446]	[0.248,0.331]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.745,0.916]	[0.396,0.321]	[1.480,0.695]	[1.741,0.904]	[2.027,1.145]	[0.422,0.341]	[0.278,0.210]
$\sigma = \frac{3b}{2}$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.994,0.676]	[0.221,0.271]	[1.836,0.503]	[1.983,0.683]	[2.138,0.852]	[0.294,0.323]	[0.226,0.211]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.801,1.002]	[0.242,0.226]	[1.630,0.859]	[1.798,1.004]	[1.973,1.150]	[0.315,0.303]	[0.200,0.184]

Table 3(b): Estimator for (β_i^0, β_i^1) (BWK Design, $G = 3000$)

		Mean	*Std.Dev.*	*LQ*	*Median*	*HQ*	*RMSE*	*MAE*
Infeasible	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.845,0.555]	[0.137,0.186]	[1.758,0.429]	[1.851,0.551]	[1.940,0.674]	[0.144,0.194]	[0.098,0.135]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.640,0.839]	[0.155,0.143]	[1.535,0.745]	[1.641,0.846]	[1.751,0.941]	[0.160,0.148]	[0.115,0.096]
$\sigma = b$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.868,0.557]	[0.171,0.221]	[1.750,0.421]	[1.863,0.560]	[1.977,0.713]	[0.184,0.228]	[0.124,0.147]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.662,0.855]	[0.190,0.178]	[1.525,0.732]	[1.665,0.856]	[1.798,0.973]	[0.200,0.186]	[0.133,0.129]
$\sigma = \frac{b}{2}$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.859,0.562]	[0.274,0.325]	[1.688,0.386]	[1.860,0.549]	[2.036,0.782]	[0.280,0.331]	[0.193,0.221]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.659,0.848]	[0.296,0.235]	[1.468,0.684]	[1.655,0.834]	[1.871,1.010]	[0.302,0.240]	[0.199,0.171]
$\sigma = \frac{3b}{2}$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.924,0.624]	[0.159,0.210]	[1.814,0.478]	[1.925,0.630]	[2.031,0.772]	[0.201,0.244]	[0.131,0.172]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.735,0.925]	[0.177,0.161]	[1.620,0.823]	[1.730,0.917]	[1.853,1.038]	[0.223,0.204]	[0.141,0.135]