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

Arthur Lewbel, Boston College

Xun Tang, University of Pennsylvania

Original: November 25, 2010. Revised: March 5, 2013

Abstract

The existing literature on binary games with incomplete information assumes that either payoff functions or the distribution of private information are finitely parameterized to obtain point identification. In contrast, we show that, given excluded regressors, payoff functions and the distribution of private information can both be nonparametrically point identified. An excluded regressor for 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. We show how excluded regressors satisfying these conditions arise in contexts such as entry games between firms, as variation in observed components of fixed costs. Our identification proofs are constructive, so consistent nonparametric estimators can be readily based on them. For a semiparametric model with linear payoffs, we propose root-N consistent and asymptotically normal estimators for parameters in players' payoffs. Finally, we extend our approach to accommodate the existence of 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   Introduction

We estimate a class of static game models where players simultaneously choose between binary actions based on imperfect information about each other's payoffs. Such a class of models lends itself to wide applications in industrial organization. Examples include firms' entry decisions in Seim (2006), decisions on the timing of commercials by radio stations in Sweeting (2009), firms' choice 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 when 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 observed states, then interaction effects between players and marginal effects of excluded regressors on players' payoffs can be identified. If in addition the excluded regressors vary sufficiently relative to the support of private information, then the entire payoff functions of each player and the distribution of players' private information are 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 an example of an economic application in which profit maximization by players gives rise to excluded regressors with the required properties. Consider the commonly studied game in which firms each simultaneously decide whether to enter a market or not, knowing that after entry they will compete through choices of quantities. The fixed cost for a firm i (to be paid upon entry) 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 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 result, while fixed costs affect the decision to enter, they do not otherwise affect the quantities supplied by any of the players. In practice, only some components of a firm's fixed cost are observed by econometricians, while the remaining, unobserved part of the fixed cost is the private information only known to this firm. Our method can thus be applied, with the observed component of fixed costs playing the role of excluded regressors.

Other excluded regressor conditions that contribute to identification (such as conditional in-

dependence from private information and sufficient variation relative to the conditional support of private information given observed states) also have immediate economic interpretations in the context of entry games. We discuss those interpretations in greater details in Section 4.

We extend our excluded regressor approach to obtain point identification while allowing for multiple Bayesian Nash equilibria (BNE) in the data generating process (DGP), without assuming knowledge of an equilibrium selection mechanism. Identification is obtained as before, but only using variation of excluded regressors within a set of states with a single BNE in the data-generating process. We build on De Paula and Tang (2011), showing that when players' private information is independent conditional on observed states, then the players' actions are correlated given these states if and only if those states give rise to multiple BNE. Thus it is possible to infer from the data whether a given set of states give rise to single BNE, and base identification only on those states, without imposing parametric assumptions on equilibrium selection mechanisms or on the distribution of private information.

• 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. This is in contrast to the existing literature, which requires that either the payoff functions or the distribution of private information be finitely parameterized to obtain point identification.

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 contrast 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 when players' private information are independent of states observed by researchers. He also proposes a 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, and proposes consistent estimators for linear coefficients without establishing rates of convergence. Wan and Xu (2010) consider a class of Bayesian games with index utilities under median restrictions. They allow players' private information to be positively correlated in a particular form, 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 but require them to be independent across players conditional on observable states.

The second contribution of this paper is to propose a two-step, consistent and asymptotically normal estimator for semi-parametric models with linear payoffs and private information correlated with non-excluded regressors. 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 depends on properties of the tails of the distribution of excluded regressors. The estimators for interaction effects converge at the parametric rate regardless of the boundedness of these supports.[2]

Our third contribution is to accommodate multiple Bayesian Nash equilibria without relying on knowledge of any equilibrium selection mechanism. The possibility of multiple BNE in data poses a challenge to structural estimation because in this case observed player choice probabilities are an unknown mixture of those implied by each of the multiple equilibria. Hence, the exact structural link between the observable distribution of actions and model primitives is not known. The existing literature proposes several approaches to address this issue. These include parameterizing the equilibrium selection mechanism, assuming 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 only obtaining partial identification results (Beresteanu, Molchanov, Molinari (2011)).[3]

For a model of Bayesian games with linear payoffs and correlated bivariate normal private information, Xu (2009) addressed the multiplicity issue by focusing on states for which the model only admits a unique, monotone equilibrium. He showed, with these parametric assumptions, a subset of these states can be inferred using the observed distribution of choices and then used for estimating parameters in payoffs.

Our approach for dealing with multiple BNE builds on a related idea. Instead of focusing on states where the model only permits a unique equilibrium, we focus on a set of states where the observed choice probabilities are rationalized by a unique BNE. This set can include both states

---

[2]Khan and Nekipelov (2011) studied the Fisher information for the interaction parameter in a similar model of Bayesian games, where private information are correlated across players, but the baseline payoff fuctions 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 regular identification (i.e. positive Fisher information) of the interaction parameter while assuming knowledge of the functional form of the rest of the payoffs as well as the equilibrium selection mechanism.

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

where the model admits a unique BNE only, and states where multiple equilibria are possible but where the agents in the data always choose the same BNE. We identify such states without parametric restrictions on primitives and without equilibrium selection assumptions. Applying our excluded regressor assumptions to a set of states where the observed choice probabilities is rationalized by a unique BNE then allows us to identify the model nonparametrically.

Our use of excluded regressors is related to the use of "special regressors" in single-agent, qualitative response models (See Dong and Lewbel (2011) and Lewbel (2012) for surveys of special regressor estimators). Lewbel (1998, 2000) studies 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 symmetry restriction on the distribution of the latent utility. Lewbel, Linton, and McFadden (2011) estimate features of willingness-to-pay models with a special regressor chosen 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 a wider class of models including social interaction models and games with complete information. Brock and Durlauf (2007) and Blume, Brock, Durlauf and Ioannides (2010) use exclusion restrictions of instruments to identify a linear model of social interactions. Bajari, Hong and Ryan (2010) use exclusion restrictions and identification at infinity, while Tamer (2003) and Ciliberto and Tamer (2009) use some special regressors to identify games with *complete* information. Our identification differs qualitatively from that of complete information games because with incomplete information, each player-specific excluded regressor can affect the strategies of all other players' through its impact on the equilibrium choice probabilities.

● **Roadmap**

The rest of the paper proceeds as follows. Section 2 introduces the model of discrete Bayesian gams. Section 3 establishes nonparametric identification of the model. Section 4 motivates our key identifying assumptions using an example of entry games, where observed components of each firm's fixed cost lends itself to the role of excluded regressors. Sections 5 and 6 propose estimators for a semiparametric model where players' payoffs are linear. Section 7 provides some evidence of the finite sample performance of this estimator. Section 8 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 9 concludes.

2 The Model

Consider a model of simultaneous games with binary actions and incomplete information that involves N players. We use N to denote both the set and the number of players. Each player i chooses an action $D_i \in \{1, 0\}$. A vector of states $X \in \mathbb{R}^J$ with $J > N$ is commonly 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 use upper cases (e.g. X, ϵ_i^*) for random vectors and lower cases (e.g. x, ε_i^*) for realized values. For any pair of random vectors R and R', let $\Omega_{R'|R}$, $f_{R'|R}$ and $F_{R'|R}$ denote respectively the support, the density and the distribution of R' conditional on R.

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 bounded and common knowledge among all players and known to the econometrician.[4] The payoff from choosing action 0 is normalized to 0 for all players. Here $u_i^*(X, \epsilon_i^*)$ denotes the baseline return for i from choosing 1, while $h_i(D_{-i})\delta_i^*(X)$ captures interaction effects, i.e., how other players' actions affect i's payoff. One example of h_i is $h_i(D_{-i}) \equiv \sum_{j \neq i} D_j$, making the interaction effect proportional to the number of competitors choosing 1, as in Sweeting (2009) and Aradillas-Lopez (2010). Other 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 observable states X and rivals' actions D_{-i}, but δ_i^* and the baseline u_i^* are allowed to depend on X in general ways. The model primitives $\{u_i^*, \delta_i^*\}_{i=1}^N$ and $F_{\epsilon^*|X}$ are common knowledge among players but are not known to the econometrician. We maintain the following restrictions on $F_{\epsilon^*|X}$ throughout the paper.

A1 *(Conditional Independence) Given any $x \in \Omega_X$, 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 positive Radon-Nikodym densities almost everywhere over $\Omega_{\epsilon_i^*|x}$.*

This assumption, which allows X to be correlated with players' private information, is maintained in almost all previous works on the estimation of Bayesian games (e.g. Seim (2006), Aradillas-Lopez (2010), Berry and Tamer (2007), Bajari, Hong, Krainer, and Nekipelov (2010), Bajari, Hahn, Hong and Ridder (2010), Brock and Durlauf (2007), Sweeting (2009) and Tang (2010)).[5] 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

---

[4]With $N = 2$, assuming $h_i(D_j) = D_j$ does not induce any loss of generality. With $N \geq 3$, assuming h_i is known to the researcher is restrictive. Bajari, Hong, Krainer, and Nekipelov (2010) avoid this assumption, and allow the payoff matrix to be completely nonparametric, but they require the assumption that the distribution of unobserved states is completely known to the econometrician.

[5]A recent exception is Wan and Xu (2010), who nevertheless restrict equilibrium strategies to be monotone with a threshold-crossing property, and impose restrictions on the magnitude of the strategic interaction parameters.

be characterized by a profile of strategies $\{s_i(X,.)\}_{i \leq N}$ such that for all i and ϵ_i^*,

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

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, $E[h_i(D_{-i})|X, \epsilon_i^*]$ 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

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

for all $d_{-i} \in \{0,1\}^{N-1}$, where $p_i(x) \equiv \Pr\{u_i^*(X, \epsilon_i) + \delta_i^*(X)E[h_i(s_{-i}(X, \epsilon_{-i}))|X] \geq 0 \mid X = x\}$ 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 can be characterized by a vector of choice probabilities $\{p_i(x)\}_{i \leq N}$ such that:

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

Note this characterization only involves a vector of marginal rather than joint probabilities. The existence of pure-strategy BNE given any x follows from the Brouwer's Fixed Point Theorem and the continuity of $F_{\epsilon^*|X}$ in A1.

In general, the model can admit multiple BNE since the system (3) could admit multiple solutions. So in the DGP players' choices observed under some state x could potentially be rationalized by more than one BNE. In Sections 3 and 6, we first identify and estimate the model assuming players' choices are rationalized by a single BNE for each state x in data (either because their specific model only admits unique equilibrium, or because agents employ some equilibrium selection mechanism possibly unknown to the researcher). Later in Section 8 we extend our use of excluded regressors to obtain identification under multiple BNE. This organization of the paper allows us to separate our main identification strategy from the added complications involved in dealing with multiple equilibria, and 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 (2010).

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 data.[6] The econometrician observes players' actions and the states in x, but does not observe $\{\varepsilon_i^*\}_{i \in N}$.

A2 *(Unique Equilibrium) In data, players' choices under each state x are rationalized by a single BNE at that state.*

This assumption will be relaxed later, essentially by restricting attention to a subset of states for which it holds. 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. Without A2, the joint distribution of choices observable from data would be a mixture of several choice probabilities, with each implied by one of the multiple BNE. The key role of A2 is to ensure the vector of conditional expectations of $h_i(D_{-i})$ given X as directly identified from data satisfies the system in (1)-(2). When $h(D_{-i}) = \sum_{j \neq i} D_j$, A2 ensures the conditional choice probabilities of each player (denoted by $\{p_i^*(x)\}_{i \in N}$) as directly identified from data can solve (3). As such, (3) reveals how the variation in excluded regressors affects the observed choice probabilities, which provides the foundation of our identification strategies.

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 $\delta_i^*(x)$, and each only appears additively in one baseline payoff $u_i^*(x, \varepsilon_i^*)$. 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 still affect *all* players' decisions through its impact on the joint distribution of actions.

---

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

3.1 Identification 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 payoffs that does not depend on excluded regressors, and let $F_{\mathcal{S}_i | \tilde{x}}$ denote the conditional c.d.f. of $\mathcal{S}_i$.

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

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

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

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

The first goal is to identify the signs of the excluded regressor marginal effects α_i (recalling 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 $\alpha_i \tilde{f}_i(x) = p_{i,i}^*(x) - p_{i,j}^*(x) p_{j,i}^*(x) / p_{j,j}^*(x)$. Since the density $\tilde{f}_i(x)$ must be nonnegative and somewhere strictly positive, this identifies the sign α_i in terms of identified choice probability derivatives. Next, the interaction term $\delta_i(\tilde{x})$ is identified by dividing equation (5) by equation (6) to obtain $p_{i,i}^*(x) / p_{i,j}^*(x) = \left[\alpha_i + \delta_i(\tilde{x}) p_{j,i}^*(x) \right] / \delta_i(\tilde{x}) p_{j,j}^*(x)$, and solving this equation for $\delta_i(\tilde{x})$.

A technical condition called NDS that we provide later defines a set ω of values of x for which these constructions can be done avoiding the subtle problem of division by zero. Estimation of α_i and $\delta_i(\tilde{x})$ can directly follow these constructions, based on nonparametric regression estimates of $p_i^*(x)$, then averaging the expression for α_i over all $x \in \omega$ for α_i and averaging the expression for $\delta_i(\tilde{x})$ over values the 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

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

This has the form of a binary choice threshold crossing model, with unknown function $\mathcal{S}_i = -u_i(\tilde{X}, \epsilon_i^*)$ and error ϵ_i^* having an unknown distribution with unknown heteroskaedasticity. 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 therefore 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 $E\left(D_i \mid v_i, \tilde{x}\right) = F_{\mathcal{S}_i|\tilde{x}}\left(v_i\right)$, so $F_{\mathcal{S}_i|\tilde{x}}$ is identified at every value V_i can take on, and is identified everywhere if V_i has sufficiently large support.

Define $\tilde{u}_i(\tilde{x}) \equiv E\left(u_i(\tilde{X}, \epsilon_i^*) \mid \tilde{x}\right) = E\left(-\mathcal{S}_i \mid \tilde{x}\right)$. Then by definition the conditional mean of player i's baseline payoff function is $E\left[u_i^*(X, \epsilon_i^*) \mid x\right] = \alpha_i x_i + \tilde{u}_i(\tilde{x})$, 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 are provided by Lewbel, Linton, and McFadden (2011), except that they assumed an ordinary binary choice model where V_i is observed (even determined by experimental design), 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 will later 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 α_i and δ_i

First consider identifying the marginal effects of excluded regressors (α_i), and the state-dependent component in interaction effects $\delta_i(\tilde{x})$. We begin by deriving the general game analog to equation (4).

For any x, let $\phi_i^*(x)$ denote the conditional expectation of $h_i(D_{-i})$ given x, which is identifiable directly from data (with h_i known to all i and econometricians). When $h_i(D_{-i}) = \sum_{j \neq i} D_j$, $\phi_i^*(x) \equiv \sum_{j \neq i} p_j^*(x)$. 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 $\mathcal{W}_1(x)$, $\mathcal{W}_2(x)$, $\mathcal{V}_1(x)$, $\mathcal{V}_2(x)$ whose ordered coordinates are $(p_{i,i}^*(x))_{i \in N}$, $(p_{1,2}^*(x), p_{2,3}^*(x), .., p_{N-1,N}^*(x), p_{N,1}^*(x))$, $(\phi_{i,i}^*(x))_{i \in N}$ and $(\phi_{1,2}^*(x), \phi_{2,3}^*(x), .., \phi_{N-1,N}^*(x), \phi_{N,1}^*(x))$ respectively. For any two vectors $R \equiv (r_1, r_2, .., r_N)$ and $R' \equiv (r_1', r_2', .., r_N')$, let "$R.R'$" and "$R./R'$" denote component-wise multiplication and divisions (i.e. $R.R' = (r_1 r_1', r_2 r_2', .., r_N r_N')$ and $R./R' = (r_1/r_1', r_2/r_2', .., r_N/r_N')$).

Definition 1 *A set ω satisfies the* **Non-degenerate and Non-singular** *(NDS) condition if it*

is in the interior of Ω_X and for all $x \in \omega$, (i) $0 < p_i^*(x) < 1$ for all i; and (ii) all coordinates in $\mathcal{W}_2(x)$, $\mathcal{V}_2(x)$ and $\mathcal{V}_2(x).\mathcal{W}_1(x) - \mathcal{V}_1(x).\mathcal{W}_2(x)$ are non-zero.

The *NDS* condition holds for ω when the support of $\mathcal{S}_i = -u_i(\tilde{x}, \epsilon_i^*)$ includes the support of $\alpha_i x_i + h_i(D_{-i})\delta_i(\tilde{x})$ in its interior for any $x \equiv (x_e, \tilde{x}) \in \omega$. This happens when x_i takes moderate values while ϵ_i^* varies sufficiently conditional on $\tilde{x}$ to generate a large support of $\mathcal{S}_i$. In any BNE and for any x,

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

for all $i \in N$. For any $\tilde{x}$, the smoothness of $F_{\epsilon^*|\tilde{x}}$ in A1 and the continuity of u_i in ϵ_i^* given $\tilde{x}$ in A3 imply partial derivatives of p_i^*, ϕ_i^* with respect to excluded regressors exist over ω. To have all coordinates in $\mathcal{W}_2(x)$, $\mathcal{V}_2(x)$ and $\mathcal{V}_2(x).\mathcal{W}_1(x) - \mathcal{V}_1(x).\mathcal{W}_2(x)$ be non-zero as NDS requires, $\delta_i(\tilde{x})$ must necessarily be non-zero for all i. This rules out uninteresting cases with no strategic interaction between players.[7]

Theorem 1 *Suppose A1-3 hold and ω satisfies NDS. Then (i) $(\alpha_1, ., \alpha_N)$ are identified as signs of the ordered coordinates of*

$$E\left[(\mathcal{W}_1(X) - \mathcal{W}_2(X).\mathcal{V}_1(X)./\mathcal{V}_2(X))\, 1\{X \in \omega\}\right]. \tag{7}$$

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

$$\mathcal{A}.E\left[\mathcal{W}_2(X)./\left(\mathcal{V}_2(X).\mathcal{W}_1(X) - \mathcal{V}_1(X).\mathcal{W}_2(X)\right) \mid X \in \omega, \tilde{X} = \tilde{x}\right]. \tag{8}$$

Remark 1.1 For the special case of $N = 2$ with with $h_1(D_{-1}) = D_2$ and $h_2(D_{-2}) = D_1$, this theorem is simplified to the identification argument described in the previous subsection.

Remark 1.2 Our identification of $(\alpha_i)_{i \in N}$ in Theorem 1 does not depend on identification at infinity (i.e. does not need to send the excluded regressor x_i to positive or negative infinity in order to recover the sign of α_i from the limit of i's choice probabilities). Thus our argument does not hinge on observing states that are far out in the tails of the distribution. The same is true for identification of $(\delta_i(\tilde{x}))_{i \in N}$.

Remark 1.3 If $(\delta_i(\tilde{x}))_{i \in N}$ consists of a vector of constants (i.e. $\delta_i(\tilde{x}) = \delta_i$ for all $i, \tilde{x}$), 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 the earlier two person game this reduces to

$$\delta_1 = \alpha_1 E\left[\left(p_{1,1}^*(X)p_{2,2}^*(X) - p_{2,1}^*(X)p_{1,2}^*(X)\right)^{-1} p_{1,2}^*(X)|X \in \omega\right]. \tag{9}$$

---

[7]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 $j \neq i$ in any BNE and some of the diagonal entries in $\mathcal{W}_2$ (and $\mathcal{V}_2$) would be zero for any x. Other than this, the NDS condition not restrict how the non-excluded states enter $\delta_i(\tilde{x})$ at all. The condition that all diagonal entries in $\mathcal{V}_2.\mathcal{W}_1 - \mathcal{V}_1.\mathcal{W}_2$ be non-zero can be verified from data, because these entries only involve functions that are directly identifiable.

and similarly for δ_2.

Remark 1.4 With more than two players, there exists additional information that can be used to identify α_i and δ_i, based on using different partial derivatives from those in $\mathcal{W}_2$ and $\mathcal{V}_2$. For example, the proof can be modified to replace $\mathcal{W}_2$ with $(p_{1,N}^*, p_{2,1}^*, p_{3,2}^*, ..., p_{N-1,N-2}^*, p_{N,N-1}^*)$ and $\mathcal{V}_2$ defined accordingly. This is an additional source of over-identification that might be exploited to improve efficiency in estimation.

3.3 Recovering distributions of $\mathcal{S}_i$ given non-excluded regressors

For any i and $x \equiv (x_e, \tilde{x})$, our "generated special regressor" is defined as:

$$V_i(x; \alpha_i, \delta_i, \phi_i^*) \equiv \alpha_i x_i + \delta_i(\tilde{x})\phi_i^*(x).$$

For any $\tilde{x}$ with $(\delta_i(\tilde{x}))_{i \in N}$ identified, such a generated regressor can be constructed from the distribution of actions and states observed from data. Denote the support of V_i conditional on $\tilde{X} = \tilde{x}$ by $\Omega_{V_i|\tilde{x}}$. That is, $\Omega_{V_i|\tilde{x}} \equiv \{t \in \mathbb{R} : \exists x_e \in \Omega_{X_e|\tilde{x}} \; s.t. \; \alpha_i x_i + \delta_i(\tilde{x})\phi_i^*(x) = t\}$.

Theorem 2 *Suppose A1-3 hold and α_i and $\delta_i(\tilde{x})$ are both identified at $\tilde{x}$. Then $F_{\mathcal{S}_i|\tilde{x}}$ is identified over $\Omega_{V_i|\tilde{x}}$.*

To prove Theorem 2, observe that for a fixed $\tilde{x}$, A3 guarantees that variation in X_e does not affect the distribution of $\mathcal{S}_i$ given $\tilde{x}$. Given the unique BNE in data from A2, the choice probabilities identified from data can be related to the distribution of $\mathcal{S}_i$ given $\tilde{x}$ as:

$$\Pr(D_i = 1 | X = (x_e, \tilde{x})) = \Pr(\mathcal{S}_i \leq V_i(x_e, \tilde{x}) | \tilde{X} = \tilde{x}) \tag{10}$$

for all i (where the equality follows from independence between X_e and ϵ_i^* given $\tilde{x}$ in A3). Because the left-hand side of (10) is directly identifiable, we can recover $F_{\mathcal{S}_i|\tilde{x}}(t)$ for all t on the support of $V_i(X)$ given $\tilde{x}$. It also follows that the conditional quantiles of $u_i(\tilde{X}, \epsilon_i^*)$ given $\tilde{X} = \tilde{x}$ can also be identified as $\inf\{t : \Pr(-\mathcal{S}_i \leq t | \tilde{x}) \geq \alpha\}$ for all $\alpha \in \{t \in (0,1) : \exists x \text{ such that } \Pr(D_i = 1 | x) = t\}$. Since α_i is identified, identification of $F_{\mathcal{S}_i|\tilde{x}}$ implies identification of the conditional distribution of baseline probabilities $u_i^*(x, \epsilon_i^*) = \alpha_i x_i + u_i(\tilde{x}, \epsilon_i^*)$ given x.

3.4 Identifying u_i under mean independence

A4 *(Reparametrization) For all i, $E[u_i(\tilde{X}, \epsilon_i^*)|\tilde{x}]$ is bounded for all $\tilde{x}$, so that $u_i(\tilde{X}, \epsilon_i^*)$ can be reparametrized as $\tilde{u}_i(\tilde{X}) - \epsilon_i$, with $\epsilon_i \equiv \tilde{u}_i(\tilde{X}) - u_i(\tilde{X}, \epsilon_i^*)$ and $\tilde{u}_i(\tilde{X}) \equiv E[u_i(\tilde{X}, \epsilon_i^*)|\tilde{X}]$.*

Note that A4 is a free location normalization provided the boundedness condition holds. By construction, $E[\epsilon_i|\tilde{x}] = 0$ for all i and $\tilde{x}$; and the conditional distribution $F_{\epsilon|X}$ satisfies previous restrictions on $F_{\epsilon^*|X}$ (that is, A1 and condition (ii) in A3 holds with ϵ^* replaced by ϵ).

A5 *(Large Support and Positive Density) For any i and $\tilde{x}$, (i) the support of $V_i(X)$ given $\tilde{x}$, denoted $\Omega_{V_i|\tilde{x}}$, includes the support of $\mathcal{S}_i$ given $\tilde{x}$; and (ii) the density of $V_i(X)$ given $\tilde{x}$ is positive almost surely w.r.t. Lebesgue Measure over $\Omega_{V_i|\tilde{x}}$.*

The next corollary establishes the identification of $\tilde{u}_i(\tilde{x})$, and hence of the conditional mean of baseline payoffs. The "large support" assumption (condition (i)) in A5 is key for identification. This is analogous to the large support condition for special regressors in single-agent models in Lewbel (2000). In the special case of the two player game considered earlier, A5-(i) holds if the support of X_i given $\tilde{x}$ is large enough relative to $\delta_i(\tilde{x})$ and the support of ϵ_i given $\tilde{x}$. We include detailed primitive conditions that are sufficient for A5 in Appendix B. Also note the large support condition A5-(i) can be checked using data. Specifically, $\Omega_{\mathcal{S}_i|\tilde{x}} \subseteq \Omega_{V_i|\tilde{x}}$ if and only if the supreme and the infimum of $\Pr(D_i = 1|V_i = v, \tilde{X} = \tilde{x})$ over $v \in \Omega_{V_i|\tilde{x}}$ are respectively 1 and 0.

Corollary 1 *(Theorem 2) Under A1-4 and 5(i), $\tilde{u}_i(\tilde{x})$ is identified as $E[-\mathcal{S}_i|\tilde{x}]$.*

For a given $\tilde{x}$, Theorem 2 implies we can recover the distribution $F_{\mathcal{S}_i|\tilde{x}}$ over its full support, as long as the support of the generated regressor V_i given $\tilde{x}$ covers the support of $\mathcal{S}_i$ given $\tilde{x}$. This then suffices to identify $\tilde{u}_i(\tilde{x})$ since $\tilde{u}_i(\tilde{x}) = E[-\mathcal{S}_i|\tilde{x}]$. Corollary 1 then follows immediately given that the distribution of $\mathcal{S}_i$ given $\tilde{x}$ implies the identification of $E[-\mathcal{S}_i|\tilde{x}]$.

With α_i, $\delta_i(\tilde{x})$ and $\tilde{u}_i(\tilde{x})$ identified at $\tilde{x}$, the distribution of ϵ_i given $\tilde{x}$ is also identified, and can be recovered as

$$E\left(D_i|V_i = v, \tilde{X} = \tilde{x}\right) = \Pr\left(\epsilon_i \leq \tilde{u}_i(\tilde{X}) + V_i|v, \tilde{x}\right) = F_{\epsilon_i|\tilde{x}}\left(\tilde{u}_i(\tilde{x}) + v\right)$$
$$\implies \quad F_{\epsilon_i|\tilde{x}}(e) = E\left[D_i|V_i = e - \tilde{u}_i(\tilde{x}), \tilde{X} = \tilde{x}\right]$$

for all e on the support $\Omega_{\epsilon_i|\tilde{x}}$. By A5, the support of V_i given $\tilde{x}$ is sufficiently large to ensure that $F_{\epsilon_i|\tilde{x}}(e)$ is identified for all e on the support $\Omega_{\epsilon_i|\tilde{x}}$. Thus it is possible for econometricians to predict equilibrium outcomes under known counterfactual changes of the error distribution.

The remainder of this subsection introduces a useful shortcut for constructing $\tilde{u}_i(\tilde{x})$ from observable distributions. Such a shortcut allows us to back out $\tilde{u}_i(\tilde{x})$ without having to recover the whole distribution $F_{\mathcal{S}_i|\tilde{x}}$ under the mild conditions of positive densities in A5-(ii). Let c^* be a constant that lies in the interior of $\Omega_{V_i|\tilde{x}}$. For all i and $x = (x_e, \tilde{x})$ define

$$Y_i^*(d_i, x) \equiv \frac{d_i - 1\{V_i(x) \geq c^*\}}{f_{V_i|\tilde{x}}(V_i(x))}, \tag{11}$$

The choice of c^* is feasible in estimation, for the support $\Omega_{V_i|\tilde{x}}$ is known as long as $\alpha_i, \delta_i(\tilde{x})$ are identified. Such a choice of c^* is also allowed to depend on $\tilde{x}$.

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

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

for any c^* in the interior of $\Omega_{V_i|\tilde{x}}$.

Proof of Theorem 3. For all i and $\tilde{x}$,

$$E\left(Y_i^*|\tilde{x}\right) = E\left[E\left(Y_i^*|V_i, \tilde{X}\right)|\tilde{X} = \tilde{x}\right] = \int_{\Omega_{V_i|\tilde{x}}} \frac{E\left[D_i|v_i, \tilde{x}\right] - 1\{v_i > c^*\}}{f_{V_i}(v_i|\tilde{x})} f_{V_i}(v_i|\tilde{x}) dv_i$$

$$= \int_{\Omega_{V_i|\tilde{x}}} E\left[1\{\epsilon_i \le v_i + \tilde{u}_i(\tilde{x})\} - 1\{v_i > c^*\}|v_i, \tilde{x}\right] dv_i$$

$$= \int_{\Omega_{V_i|\tilde{x}}} \int_{\Omega_{\epsilon_i|\tilde{x}}} 1\{\varepsilon_i \le v_i + \tilde{u}_i(\tilde{x})\} - 1\{v_i > c^*\} dF_{\epsilon_i}(\varepsilon_i|\tilde{x}) dv_i$$

$$= \int_{\Omega_{\epsilon_i|\tilde{x}}} \int_{\Omega_{V_i|\tilde{x}}} 1\{v_i \ge s_i\} - 1\{v_i > c^*\} dv_i dF_{\epsilon_i}(\varepsilon_i|\tilde{x}) \tag{12}$$

where s_i is a shorthand for $-\tilde{u}_i(\tilde{x}) + \varepsilon_i$. The first equality follows from the Law of Iterated Expectation, and the second and the third from the definitions of Y_i^*, V_i and positive densities in A5-(ii) respectively. The fourth follows from the independence between ϵ_i and X_e given $\tilde{x}$, which implies $F_{\epsilon_i}(.|v_i, \tilde{x}) = F_{\epsilon_i}(.|\tilde{x})$ for all $v_i, \tilde{x}$. The last equality follows from a change of the orders of integration, which is due to the Fubini's Theorem. The last expression on the right-hand side of (12) can be written as

$$\int_{\Omega_{\epsilon_i|\tilde{x}}} \int_{\Omega_{V_i|\tilde{x}}} 1\{s_i \le v_i < c^*\} 1\{s_i \le c^*\} - 1\{s_i \ge v_i > c^*\} 1\{s_i > c^*\} dv_i dF_{\epsilon_i}(\varepsilon_i|\tilde{x})$$

$$= \int_{\Omega_{\epsilon_i|\tilde{x}}} \left(1\{s_i \le c^*\} \int_{s_i}^{c^*} dv_i - 1\{s_i > c^*\} \int_{c^*}^{s_i} dv_i\right) dF_{\epsilon_i}(\varepsilon_i|\tilde{x}) = \int_{\Omega_{\epsilon_i|\tilde{x}}} (c^* - s_i) dF_{\epsilon_i}(\varepsilon_i|\tilde{x})$$

$$= c^* + \int_{\Omega_{\epsilon_i|\tilde{x}}} (\tilde{u}_i(\tilde{x}) - \varepsilon_i) dF_{\epsilon_i}(\varepsilon_i|\tilde{x}) = c^* + \tilde{u}_i(\tilde{x})$$

where the second equality uses the large support condition in A5-(i). Hence $E\left[Y_i^*|\tilde{X} = \tilde{x}\right] - c^* = \tilde{u}_i(\tilde{x})$. *Q.E.D.*

4   Fixed Cost Components as Excluded Regressors in Entry Games

In this section, we show how some components in firms' fixed costs in classical entry games satisfy properties required to be excluded regressors in our model. We also provide a context for economic interpretation of these properties.

Consider a two stage entry game with two players $i \in \{1,2\}$. When we refer to one player as i, the other will be denoted j. In the first stage of the game firms decide whether to enter. In the second stage each player observes the other's entry decision and hence knows the market structure (monopoly or duopoly). If duopoly occurs, then in the second stage firms engage in Cournot competition, choosing output quantities. The inverse market demand function face by firm i is denoted $\psi^M(q_i; \tilde{x})$ if there is monopoly, and $\psi^D(q_i, q_j; \tilde{x})$ if there is duopoly, where q_i denotes quantities produced by i and $\tilde{x}$ a vector of market- or firm-level characteristics observed by both firms and the econometrician. That the inverse market demand functions only depend on the commonly observed $\tilde{x}$ reflects the assumption that researchers can obtain the same information about commonly observed state variables as consumers in the market.

Costs for firms are given by $c_i(q_i; \tilde{x}) + x_i + \varepsilon_i$ for $i \in \{1,2\}$, where the function c_i specifies i's variable costs, the scalar x_i is a commonly observed component of i's fixed cost, and ε_i is a component of i's fixed cost that is privately observed by i. The fixed costs $x_i + \varepsilon_i$ (and any other fixed costs that may be absorbed in c_i function) are paid *after* entry but *before* the choice of production quantities. Both (x_1, x_2) and $\tilde{x}$ are known to both firms as they make entry decisions in the first stage, and can be observed by researchers in data. A firm that decides not to enter receives zero profit in the second stage.

If only firm i enters in the first stage, then its profit is given by $\psi^M(q^*; \tilde{x})q^* - c_i(q^*; \tilde{x}) - x_i - \varepsilon_i$, 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. If there is a duopoly in the second stage, then i's profit is $\psi^D(q_i^*, q_j^*)q_i^* - c_i(q_i^*; \tilde{x}) - x_i - \varepsilon_i$ 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. The non-entrant's profit is 0. The entrants' choices of outputs are functions of $\tilde{x}$ alone in both cases. This is because by definition both the commonly observed and the idiosyncratic (and privately known) component of the fixed costs (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 $s \in \{D, M\}$, with π_i^s being the difference between firm i's revenues and variable costs under market structure s.

In the first stage, firms simultaneously decide whether to enter or not, based on a calculation of the 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 expected profits from entry are:

$$\pi_i^M(\tilde{x}) + p_j(x)\delta_i(\tilde{x}) - x_i - \varepsilon_i \tag{13}$$

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 (x, ε_i). That p_j does not depend on ε_i is due to the assumed independence between private information conditional on $\tilde{x}$ in A1. The expected profits from staying out is zero. Hence, in any Bayesian Nash equilibrium, i decides to enter if and only if the expression in (13) is greater than zero, and $\Pr(D_1 = d_1, \ D_2 = d_2 \mid x) = \Pi_{i=1,2} p_i(x)^{d_i}(1 - p_i(x))^{1-d_i}$, where

$$p_i(x) = \Pr\left\{\epsilon_i \leq \pi_i^M(\tilde{x}) + p_j(x)\delta_i(\tilde{x}) - x_i \mid x\right\}$$

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

Examples of fixed cost components (x_1, x_2) that could be public knowledge include lump-sum expenses such as equipment costs, maintenance fees and patent/license fees, while (ϵ_1, ϵ_2) could be idiosyncratic administrative overhead such as firms' office rental or supplies. The non-excluded regressors $\tilde{x}$ would include market or industry level demand and cost shifters. 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 $\tilde{u}_i(\tilde{x})$ in Section 3.4 also holds, as long as 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 $\bar{\varepsilon} \leq \pi_i^M(\tilde{x})$; (b) there exists (x_i, x_j) with $p_j(x_i, x_j, \tilde{x}) = 0$ and $\pi_i^M(\tilde{x}) - \bar{\varepsilon} \geq x_i$; and (c) there exists $(\hat{x}_i, \hat{x}_j)$ with $p_j(\hat{x}_i, \hat{x}_j, \tilde{x}) = 1$ and $\pi_i^D(\tilde{x}) \leq x_i$. In economic terms, (a) means monopoly profits are sufficiently high before subtracting commonly known fixed costs X_i; while (b) and (c) require that x_j can be relatively high for firm j while x_i is relatively low for firm i; and vice versa.[8]

More generally, the large support assumption has an immediate economic interpretation. It means that, conditional on commonly observed demand shifters (non-excluded states in $\tilde{X}$), a firm

---

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

i may not find it profitable to enter even when private fixed costs ϵ_i are as low as possible. This can happen if X_i (the publicly known component of fixed cost) is too high, or the competition is too intense (in the sense that the other firm's equilibrium entry probability p_j is too high). Large support also means that firm i could decide to enter even when ϵ_i is as high as maximum possible, because X_i and $p_j(X)$ can be sufficiently low while the market demand is sufficiently high to make entry attractive. These assumptions on the supports of X_i and ϵ_i are particularly plausible in applications where publicly known components of fixed costs are relatively large, e.g., the fixed costs in acquiring land, buildings, and equipment and acquiring license/patents could be much more substantial than those incurred by idiosyncratic supplies and overhead. Finally, note that the economic implications of the large support condition can in principle be 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.

5   A Simplified Estimator for $\tilde{u}_i$

Estimation for payoffs could be done in principle using sample analogs of Theorem 3. Nonetheless we introduce a refined and simplified argument for identifying $\tilde{u}_i$ in this subsection, in order to simplify the limiting distribution theory for semiparametric estimators. We then build on a sample analog of the approach in this subsection to construct a relatively easier estimator in Section 6, and establish its limiting distribution. For any $i \in N$, let $x_{e,-i}$ denote $(x_j)_{j \in N \setminus \{i\}}$ and let x_{-i} denote $(x_{e,-i}, \tilde{x})$.

A5' *For any i and $\tilde{x}$, there exists some x_{-i} such that (i) the support of V_i given x_{-i} includes the support of S_i given $\tilde{x}$; (ii) the conditional density of X_i given x_{-i} exists and is positive almost everywhere over its full support; and (iii) the sign of $\partial V_i(X)/\partial X_i|_{X=(x_i,x_{-i})}$ equals the sign of α_i for all $x_i \in \Omega_{X_i|x_{-i}}$.*

The large support condition in A5' is stronger than that in A5. To see this, note the support of $F_{S_i|x}$ does not vary with x_e under the conditional independence of excluded regressors in A3. While condition (i) in A5 and A5' both require generated special regressors to vary sufficiently and cover the same support of $F_{S_i|\tilde{x}}$, the latter is more stringent in that it requires such variation be generated only by variation in X_i while x_{-i} is fixed. In comparison, the former fixes $\tilde{x}$ and allow variations in V_i to be generated by *all* excluded regressors in X_e. This stronger support condition (i) in A5' also provides a source of over-identification of $\tilde{u}_i(\tilde{x})$. Similar to A5, the large support condition in A5' can in principle be checked using observable distributions. That is, $\Omega_{S_i|\tilde{x}} \subseteq \Omega_{V_i|x_{-i}}$ if and only if the supreme and infimum of $\Pr(D_i = 1|X_i = t, X_{-i} = x_{-i})$ over $t \in \Omega_{X_i|x_{-i}}$ are 1 and 0 respectively.

The last condition in A5' requires the generated special regressor associated with i to be

monotone in the excluded regressor X_i given $(x_{e,-i}, \tilde{x})$. It holds when X_i's *direct* marginal effect on i's latent utility (i.e. α_i) is not offset by its *indirect* marginal effect (i.e. $\delta_i(\tilde{x})\phi_{i,i}^*(x_i, x_{-i})$) for all $x_i \in \Omega_{X_i|x_{-i}}$. This condition is needed for the alternative method for recovering $\tilde{u}_i$ below (Corollary 2), which builds on changing variables between excluded regressors and generated special regressors. Such a monotonicity condition is *not* needed for the identification argument in Section 3.4 (Theorem 3).

Condition (iii) in A5' holds under several scenarios with mild conditions. One such scenario is when $x_{e,-i}$ takes extreme values. To see this, consider the case with $N = 2$ and $h_i(D_{-i}) = D_j$. The marginal effect of X_i on V_i at $x = (x_i, x_j, \tilde{x})$ is $\alpha_i + \delta_i(\tilde{x}) \left. \partial p_j^*(X)/\partial X_i \right|_{X=x}$, where α_i and $\delta_i(\tilde{x})$ are both finite constants given $\tilde{x}$. With x_j taking extremely large or small values, the impact of x_i on p_j^* in BNE gets close to 0. Thus the sign of $\alpha_i + \delta_i(\tilde{x})p_{j,i}^*(x)$ will be identical to the sign of α_i over $\Omega_{X_i|x_{-i}}$ for such x_j and $\tilde{x}$. In another scenario, A5'-(iii) can even hold uniformly over the entire support of X, provided both $\delta_i(\tilde{x})$ and the conditional density of players' private information $f_{\epsilon_i|\tilde{x}}$ are bounded above by small constants. In Appendix B, we formalize in details these sufficient primitive conditions for the monotonicity in A5'-(iii).

We now show how conditions in A5' can be used to provide a short-cut for identifying $\tilde{u}_i(\tilde{x})$. Fix i and $\tilde{x}$, and consider any $x_{e,-i}$ that satisfies A5' for i and $\tilde{x}$. Let $\underline{v_i}(x_{-i})$ and $\overline{v_i}(x_{-i}) \in \mathbb{R} \cup \{-\infty, +\infty\}$ denote the infimum and the supremum of the conditional support $\Omega_{V_i|x_{-i}}$. Let H be a continuous cumulative distribution (to be chosen by the econometrician) with a bounded interval support that is a subset of $\Omega_{V_i|x_{-i}}$. The choice of H is allowed to depend on both i and x_{-i}, though for simplicity we will suppress any such dependence in notation. Choosing such a distribution H with the required support is feasible in estimation, since α_i and $\delta_i(\tilde{x})$ (and consequently conditional support $\Omega_{V_i|\tilde{x}}$) is identified. Let μ_H denote the expectation of H, which is known because H is chosen by the econometrician, and which could potentially depend on x_{-i}. We discuss how to choose H in estimation in Section 6.2. For any i and x define

$$Y_{i,H} \equiv \frac{[d_i - H(V_i(x))] \left[1 + \alpha_i\delta_i(\tilde{x})\phi_{i,i}^*(x)\right]}{f_{X_i}(x_i|x_{-i})} \tag{14}$$

where $\alpha_i + \delta_i(\tilde{x})\phi_{i,i}^*(x)$ by construction is the marginal effect of the excluded regressor x_i on the generated special regressor $V_i(x)$. Note $Y_{i,H}$ is well-defined under A5', for $f_{X_i|x_{-i}}$ is positive almost everywhere over $\Omega_{X_i|x_{-i}}$.

Corollary 2 *(Theorem 3) Suppose A1-4 hold, and both α_i and $\delta_i(\tilde{x})$ are identified at $\tilde{x}$. For any i and $x_{-i} = (x_{e,-i}, \tilde{x})$ such that A5' holds,*

$$\tilde{u}_i(\tilde{x}) = E\left[Y_{i,H}|x_{-i}\right] - \mu_H. \tag{15}$$

Remark 4.1 If zero is in the support of V_i given x_{-i} then Corollary 2 holds taking H to be the degenerate distribution function $H(V) = I\{V \geq 0\}$ and $\mu_H = 0$. Corollary 2 generalizes the special

regressor results in Lewbel (2000) and Lewbel, Linton, and McFadden (2011) by allowing $H\,(.)$ to be an arbitrary distribution function instead of a degenerate distribution function, and by using the density of an excluded regressor rather than that of the special regressor in the denominator of the endogenous variable constructed. Both modifications simplify the limiting distribution theory in our context where the special regressor is also a generated regressor. In particular, choosing H to be continuously differentiable allows us to take standard Taylor expansions to linearize the estimator as a function of the estimated V_i around the true V_i. Dividing by the density of X_i instead of the density of V_i requires estimating the density of an observed regressor instead of the estimating the density of an estimated regressor.

Remark 4.2 It is clear from Corollary 2 that $\tilde{u}_i(\tilde{x})$ can be recovered using *any* $(N-1)$-vector $x_{e,-i}$ such that the support of V_i given $x_{-i} = (x_{e,-i}, \tilde{x})$ is large enough and $V_i(X)$ is monotone in x_i given x_{-i}. As such, $\tilde{u}_i(\tilde{x})$ is over-identified. The source of over-identification is due to the stronger support condition required in condition (i) of A5'.

Remark 4.3 The large support condition (i) and the monotonicity condition (iii) in A5' could be checked from observable distribution. This is because $\Pr(D_i = 1|V_i = v, x_{-i})$ is identified over $v \in \Omega_{V_i|x_{-i}}$ for any x_{-i}, and the marginal impact of excluded regressors on generated special regressors are identified.

6 Semiparametric Estimation of Index Payoffs

We estimate a two-player semi-parametric model, where players have constant interaction effects and index payoffs. That is, 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. The specification is similar to Aradillas-Lopez (2010), except the independence between $(\epsilon_i)_{i\in N}$ and X therein is replaced here by the weaker assumption of conditional independence between (ϵ_1, ϵ_2) and (X_1, X_2) given $\tilde{X}$.

We first show how to relate $\alpha_i, \delta_i, \beta_i$ to observable distributions in the semiparametric model. Given identification has been established for the general non-parametric case, our focus here is to propose convenient algebraic links between parameters and observable distributions under additional parametric restrictions. Consider any ω such that $0 < p_i^*(X), p_j^*(X) < 1$ and all coordinates in $\mathcal{W}_2(x), \mathcal{V}_2(x)$ and $\mathcal{V}_2(x).\mathcal{W}_1(x) - \mathcal{V}_1(x).\mathcal{W}_2(x)$ are non-zero almost everywhere on ω. For $i = 1, 2$, α_i are identified as

$$\alpha_i = sign\left\{ E\left[\left(p_{i,i}^*(X) - \frac{p_{i,j}^*(X)p_{j,i}^*(X)}{p_{j,j}^*(X)} \right) 1\{X \in \omega\} \right] \right\} \tag{16}$$

and δ_i are identified as in equation (9) respectively. To attain identification of β_i in this semi-

parametric model based on Corollary 2, define for any i and x

$$Y_{i,H} \equiv \frac{[d_i - H(V_i(x))] [1 + \alpha_i \delta_i p^*_{j,i}(x)]}{f_{X_i}(x_i|x_{-i})},$$

where $V_i(x) \equiv \alpha_i x_i + \delta_i p^*_j(x)$, and H is an arbitrary continuously differentiable distribution with a support contained in the support of $V_i(X)$ given x_{-i} and an expectation equal to μ_H. We will choose H explicitly while estimating β_i in Section 6.2 below. We maintain the following assumption throughout this section.

A5" *(i) Given any i and $\tilde{x}$, the support $\Omega_{V_i|x_{-i}}$ includes the support $\Omega_{\mathcal{S}_i|\tilde{x}}$ for all $x_{-i} \in \Omega_{X_{-i}|\tilde{x}}$; (ii) Given any i and $\tilde{x}$, the density of X_i given x_{-i} is positive almost surely for all $x_{-i} \in \Omega_{X_{-i}|\tilde{x}}$. (iii) The sign of $\partial V_i(X)/\partial X_i|_{X=x}$ equals the sign α_i for all $x \in \Omega_X$.*

Condition A5" is more restrictive than A5' (stated in Section 5), in that A5" requires the set of x_{-i} satisfying the three conditions in A5' to be the whole support of X_{-i} given $\tilde{x}$. Intuitively, A5" can be satisfied when *(i)* conditional on any x_{-i}, the support of X_i is sufficiently large relative to that of $\mathcal{S}_i$ given $\tilde{x}$; and *(ii)* $\delta_i(\tilde{x})$ and the density of ϵ_i given $\tilde{x}$ are both bounded above by constants that are relative small (in the sense stated in Appendix B). The condition that $f_{\epsilon_i|\tilde{x}}$ is bounded above by small constant is plausible especially in cases where the support of ϵ and X_e are both large given $\tilde{x}$. We formalize these restrictions on model primitives which imply A5" in Appendix B.

Corollary 3 *(Theorem 3) Suppose A1-4 and A5" hold and $\tilde{u}_i(\tilde{x}) = \tilde{x}'\beta_i$, $\delta_i(\tilde{x}) = \delta_i$ for $i = 1, 2$. Suppose α_i, δ_i are identified for both i. If $E\left(\tilde{X}\tilde{X}'\right)$ has full rank, then*

$$\beta_i = \left[E\left(\tilde{X}\tilde{X}'\right)\right]^{-1} E\left[\tilde{X}(Y_{i,H} - \mu_H)\right]. \tag{17}$$

To prove Corollary 3, apply the proof of Corollary 2 with $\tilde{u}_i(\tilde{X}) = \tilde{X}'\beta_i$ to get $\tilde{x}'\beta_i = E[Y_{i,H}| x_{-i}] - \mu_H$ for any $x_{-i} = (x_{e,-i}, \tilde{x}) \in \Omega_{X_{-i}|\tilde{x}}$. Thus $\tilde{x}\tilde{x}'\beta_i = E[\tilde{x}(Y_{i,H} - \mu_H)|x_{-i}]$ for all such x_{-i}. Integrating out X_{-i} on both sides proves (17).

Sections 6.1 and 6.2 below define semiparametric estimators for $\alpha_i, \beta_i, \delta_i$ and derive their asymptotic distributions. Recovering these parameters requires conditioning on a set of states ω in (16) and (9). Because marginal effects of excluded regressors X_i on the equilibrium choice probabilities (p^*_i, p^*_j) are identified from data, the set can be estimated using sample analogs of the choice probabilities and the generated special regressors. For instance, ω can be estimated by $\hat{\omega} = \{x : \hat{p}_i(x), \hat{p}_j(x) \in (c, 1-c); |\hat{p}_{i,i}(x)|, |\hat{p}_{j,j}(x)| > c \text{ and } |\hat{p}_{i,i}(x)\hat{p}_{j,j}(x) - \hat{p}_{i,j}(x)\hat{p}_{j,i}(x)| > c\}$, where $\hat{p}_i, \hat{p}_j, \hat{p}_{i,i}, \hat{p}_{j,j}$ are kernel estimators to be defined in Section 6.1 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 do not account for estimation error in ω while establishing limiting distributions of our estimators for α_i, δ_i and β_i. Such a task is left for future research. This choice allows us to better focus on dealing with estimation errors in $\hat{p}_i, \hat{p}_j, \hat{p}_{i,i}, \hat{p}_{i,j}$ as well as in $\hat{\delta}_i$ in our multi-step estimators below.

6.1 Estimation of α_i and δ_i

We consider the case where X consists of J_c continuous and J_d discrete coordinates (denoted by $X_c \equiv (X_e, \tilde{X}_c)$ and $\tilde{X}_d$ respectively). Econometricians observe G independent games in data, with each indexed by g. We show α_i can be estimated at an arbitrarily fast rate and δ_i at the root-N rate. Let $\boldsymbol{\gamma}$ denote a vector of functions $(\gamma_0, \gamma_1, \gamma_2)$, where $\gamma_0(x) \equiv f(x_c|\tilde{x}_d)f_d(\tilde{x}_d)$ and $\gamma_k(x) \equiv E[D_k|x]f(x_c|\tilde{x}_d)f_d(\tilde{x}_d)$ for $k = 1, 2$, with $f(.|.)$, $E(.|.)$ being conditional densities and expectations, and $f_d(.)$ being a probability mass function for $\tilde{X}_d$ respectively. Define $\gamma_{k,i}(x) \equiv \frac{\partial \gamma_k(x)}{\partial X_i}$ for $k \in \{0, 1, 2\}$ and $i \in \{1, 2\}$. Define a 9-by-1 vector:

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

Let $\boldsymbol{\gamma}_{(1)}^*$ and $\hat{\boldsymbol{\gamma}}_{(1)}$ denote true parameters in the data-generating process and their corresponding kernel estimates respectively. (For notation simplicity, we sometimes suppress their arguments X when there is no ambiguity.) We use the sup-norm $\sup_{x \in \Omega_X} \|.\|$ for spaces of functions with domain Ω_X.

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

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

where

$$m_A^i\left(x; \boldsymbol{\gamma}_{(1)}\right) \equiv \frac{\gamma_{i,i}(x)\gamma_0(x) - \gamma_i(x)\gamma_{0,i}(x)}{\gamma_0(x)\gamma_0(x)} - \frac{\gamma_{i,j}(x)\gamma_0(x) - \gamma_i(x)\gamma_{0,j}(x)}{\gamma_0(x)\gamma_0(x)} \frac{\gamma_{j,i}(x)\gamma_0(x) - \gamma_j(x)\gamma_{0,i}(x)}{\gamma_{j,j}(x)\gamma_0(x) - \gamma_j(x)\gamma_{0,j}(x)},$$

$$m_\Delta^i\left(x; \boldsymbol{\gamma}_{(1)}\right) \equiv \left(\frac{\gamma_{i,i}(x)\gamma_0(x) - \gamma_i(x)\gamma_{0,i}(x)}{\gamma_{i,j}(x)\gamma_0(x) - \gamma_i(x)\gamma_{0,j}(x)} - \frac{\gamma_{j,i}(x)\gamma_0(x) - \gamma_j(x)\gamma_{0,i}(x)}{\gamma_{j,j}(x)\gamma_0(x) - \gamma_j(x)\gamma_{0,j}(x)}\right)^{-1} \left(\frac{\gamma_{j,j}\gamma_0 - \gamma_j\gamma_{0,j}}{\gamma_0(x)\gamma_0(x)}\right)^{-1}.$$

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

$$\hat{A}^i \equiv \frac{1}{G}\sum_g w_g m_A^i\left(x_g; \hat{\boldsymbol{\gamma}}_{(1)}\right); \tag{18}$$

$$\hat{\Delta}^i \equiv \left(\sum_g w_g\right)^{-1}\sum_g w_g m_\Delta^i\left(x_g; \hat{\boldsymbol{\gamma}}_{(1)}\right), \tag{19}$$

where w_g is a short-hand for $1\{x_g \in \omega\}$. 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 exposition, define $d_{g,0} \equiv 1$ for all $g \leq G$. Kernel estimates for γ_k^* and its partial derivatives with respect to excluded regressors are

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

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

for $k = 0, 1, 2$ and $i = 1, 2$, where $K_\sigma(.) \equiv \sigma^{-J_c} K(./\sigma)$, $K_{\sigma,i}(.) \equiv \sigma^{-J_c} \partial K(./\sigma)/\partial X_i$.[9] Our estimators for α_i, δ_i are

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

Let F_Z^*, f_Z^* denote the true distribution and the density of $Z \equiv (X, D_1, D_2)$ in the data-generating process.

Assumption S *(i) γ^* is continuously differentiable in x_c 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 containing ω. (ii) $\gamma_{(1)}^*$ is 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 $[VAR_{u_i}(x)]^{1+\xi} f^*(x_c|\tilde{x}_d)$ is uniformly bounded over ω. For any $\tilde{x}_d$, both $p_i^*(x)^2 f^*(x_c|\tilde{x}_d)$ and $VAR_{u_i}(x) f^*(x_c|\tilde{x}_d)$ are continuous in x_c and uniformly bounded over ω.*

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

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

A sequence of bandwidths σ that satisfies Assumption B exists, as long as the parameter for smoothness m is sufficiently large relative to J_c. Under Assumptions S, K and B, $\sup_{x \in \omega} ||\hat{\gamma}_{(1)}(x) - \gamma_{(1)}^*(x)|| = o_p(G^{-1/4})$. This ensures the remainder term from linearization of $\frac{1}{G} \sum_g w_g m_A^i(x_g; \hat{\gamma}_{(1)})$ at $\gamma_{(1)}^*$ diminishes at a rate faster than $\sqrt{G}$.

Assumption W *The set ω is convex and contained in an open subset of Ω_X.*

The convexity of ω implies that for all i the the set $\{x_i : (x_i, x_{-i}) \in \omega\}$ must be an interval. Denote end-points of the interval by $l_i(x_{-i})$ and $h_i(x_{-i})$ for i and x_{-i}. This property implies for any function $\mathcal{G}(x)$, the expression $\int 1\{x \in \omega\} \mathcal{G}(x) dF_X^*$ can be written as $\int \int_{l_i(x_{-i})}^{h_i(x_{-i})} \mathcal{G}(x) \gamma_0^*(x) dx_i dx_{-i}$, which comes in handy as we derive the correction terms in the limiting distribution of $\sqrt{G}\left(\hat{A}^i - A^i\right)$ and $\sqrt{G}(\hat{\Delta}^i - \Delta^i)$ due to the use of kernel estimates $\hat{\gamma}_{(1)}$. That ω is contained in an open subset of Ω_X is necessary for derivatives of $\gamma_{(1)}, \gamma_{(1)}^*$ with respect to x to be well-defined over ω.

To specify correction terms in the limiting distribution, we need to introduce additional notations. Let $w_x \equiv 1\{x \in \omega\}$ for the convex set ω satisfying NDS. For $i = 1, 2$, let $\tilde{D}_{i,A}(x)$ (and $\tilde{D}_{i,\Delta}(x)$) denote two 9-by-1 vectors that consist of derivatives of $m_A^i(x; \gamma_{(1)})$ (and $m_\Delta^i(x; \gamma_{(1)})$ respectively) with respect to the ordered vector $(\gamma_0, \gamma_i, \gamma_j, \gamma_{0,i}, \gamma_{0,j}, \gamma_{i,i}, \gamma_{i,j}, \gamma_{j,i}, \gamma_{j,j})$ at $\gamma_{(1)}^*$.

---

[9] 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 $\zeta > 0$.

Let $\tilde{D}_{i,A,k}$ (and $\tilde{D}_{i,\Delta,k}$) denote the k-th coordinate in $\tilde{D}_{i,A}$ (and $\tilde{D}_{i,\Delta}$) for all k. For $s = i, j$, let $\tilde{D}_{i,A,k}^{(s)}(x)$ (and $\tilde{D}_{i,\Delta,k}^{(s)}(x)$) denote the derivative of $\tilde{D}_{i,A,k}(X)\gamma_0^*(X)$ (and $\tilde{D}_{i,\Delta,k}(X)\gamma_0^*(X)$) with respect to the excluded regressor X_s at x. We include the detailed forms of $\tilde{D}_{i,A}, \tilde{D}_{i,\Delta}, \tilde{D}_{i,A}^{(s)}, \tilde{D}_{i,\Delta}^{(s)}$ in the supplement of this article (Lewbel and Tang (2012)).

Let $\psi_A^i \equiv (\psi_{A,0}^i, \psi_{A,i}^i, \psi_{A,j}^i)$, where

$$\psi_{A,0}^i\left(x;\boldsymbol{\gamma}_{(1)}^*\right) \equiv w_x\left[\tilde{D}_{i,A,1}(x)\gamma_0^*(x) - \tilde{D}_{i,A,4}^{(i)}(x) - \tilde{D}_{i,A,5}^{(j)}(x)\right] + \mathcal{I}_{A,1}^i(x);$$

$$\psi_{A,i}^i\left(x;\boldsymbol{\gamma}_{(1)}^*\right) \equiv w_x\left[\tilde{D}_{i,A,2}(x)\gamma_0^*(x) - \tilde{D}_{i,A,6}^{(i)}(x) - \tilde{D}_{i,A,7}^{(j)}(x)\right] + \mathcal{I}_{A,2}^i(x);$$

$$\psi_{A,j}^i\left(x;\boldsymbol{\gamma}_{(1)}^*\right) \equiv w_x\left[\tilde{D}_{i,A,3}(x)\gamma_0^*(x) - \tilde{D}_{i,A,8}^{(i)}(x) - \tilde{D}_{i,A,9}^{(j)}(x)\right] + \mathcal{I}_{A,3}^i(x),$$

with

$$\mathcal{I}_{A,1}^i(x) \equiv \left(\begin{array}{c} 1\{x_i = h_i(x_{-i})\}\tilde{D}_{i,A,4}(x)\gamma_0^*(x) - 1\{x_i = l_i(x_{-i})\}\tilde{D}_{i,A,4}(x)\gamma_0^*(x) \\ +1\{x_j = h_j(x_{-j})\}\tilde{D}_{i,A,5}(x)\gamma_0^*(x) - 1\{x_j = l_j(x_{-j})\}\tilde{D}_{i,A,5}(x)\gamma_0^*(x) \end{array} \right). \tag{20}$$

The term $\mathcal{I}_{A,2}^i(x)$ is defined by replacing $\tilde{D}_{i,A,4}, \tilde{D}_{i,A,5}$ in (20) with $\tilde{D}_{i,A,6}, \tilde{D}_{i,A,7}$ respectively. Likewise, $\mathcal{I}_{A,3}^i(x)$ is defined by replacing $\tilde{D}_{i,A,4}, \tilde{D}_{i,A,5}$ in (20) with $\tilde{D}_{i,A,8}, \tilde{D}_{i,A,9}$ respectively. Similarly, define $\psi_\Delta^i \equiv (\psi_{\Delta,0}^i, \psi_{\Delta,i}^i, \psi_{\Delta,j}^i)$ by replacing $\tilde{D}_{i,A}, \tilde{D}_{i,A}^{(s)}, \mathcal{I}_A^i$ in $\psi_{A,s}^i$ with $\tilde{D}_{i,\Delta}, \tilde{D}_{i,\Delta}^{(s)}, \mathcal{I}_\Delta^i$ respectively, where $\mathcal{I}_\Delta^i \equiv (\mathcal{I}_{\Delta,1}^i, \mathcal{I}_{\Delta,2}^i, \mathcal{I}_{\Delta,3}^i)$ is defined by replacing $\tilde{D}_{i,A,k}$ in the definition of $\mathcal{I}_A^i$ with $\tilde{D}_{i,\Delta,k}$ for all k.

Assumption D *(i) For both i and $s = 1, 2$, $\tilde{D}_{i,A}, \tilde{D}_{i,\Delta}, \tilde{D}_{i,A}^{(s)}, \tilde{D}_{i,\Delta}^{(s)}$ are continuous and bounded over an open set containing ω. (ii) The second order derivatives of $m_A^i(x;\boldsymbol{\gamma}_{(1)})$ and $m_\Delta^i(x;\boldsymbol{\gamma}_{(1)})$ with respect to $\boldsymbol{\gamma}_{(1)}$ are bounded at $\boldsymbol{\gamma}_{(1)}^*$ over an open set containing ω. (iii) For both i, there exists constants $c_A^i, c_\Delta^i > 0$ such that $E\left[\sup_{\|\eta\|\le c_A^i} \left\|\psi_{A,s}^i(x+\eta)\right\|^4\right] < \infty$ and $E\left[\sup_{\|\eta\|\le c_\Delta^i} \left\|\psi_{\Delta,s}^i(x+\eta)\right\|^4\right] < \infty$ for $s = 0, 1, 2$.*

As before, we let $D \equiv [D_0, D_1, D_2]'$ (with $D_0 \equiv 1$) in order to simplify notations, and use lower case d and d_1, d_2 for corresponding realized values. Conditions D-(i),(ii) ensure the remainder term after linearizing the sample moments around the true functions $\boldsymbol{\gamma}_{(1)}^*$ in DGP vanishes at a rate faster than $\sqrt{G}$. Applying the V-statistic projection theorem (Lemma 8.4 in Newey and McFadden (1994)), we also use these conditions to show that the difference between $\frac{1}{G}\sum_g w_g m_A^i(x_g;\hat{\boldsymbol{\gamma}}_{(1)})$ and $\frac{1}{G}\sum_g w_g m_A^i(x_g;\boldsymbol{\gamma}_{(1)}^*)$ can be written as $\int w_x(\hat{\boldsymbol{\gamma}}_{(1)} - \boldsymbol{\gamma}_{(1)}^*)'\tilde{D}_{i,A}dF_X^*$ plus a term that is $o_p(G^{-1/2})$, where F_X^* is the true marginal distribution of X in the DGP. Furthermore, by definition and the smoothness properties in S, for any γ that is twice continuously differentiable in excluded regressors X_e, $\psi_A^i \equiv (\psi_{A,0}^i, \psi_{A,1}^i, \psi_{A,2}^i)$ satisfies:

$$\int w_x(\boldsymbol{\gamma}_{(1)})'\tilde{D}_{i,A}dF_X^* = \int \sum_{s=0,1,2} \psi_{A,s}^i(x)\gamma_s(x)dx. \tag{21}$$

Using (21), the difference $\int w_x(\hat{\boldsymbol{\gamma}}_{(1)} - \boldsymbol{\gamma}_{(1)}^*)'\tilde{D}_{i,A}dF_X^*$ can be written as $\int \tilde{\nu}_A(z)d\tilde{F}_Z$, where $\tilde{\nu}_A(z) \equiv \psi_A^i(x)d - E\left[\psi_A^i(X)D\right]$ and $\tilde{F}_Z$ is some smoothed version of the empirical distribution

of $Z \equiv (X, D_i, D_j)$. Condition (iii) in Assumption D is used for showing the difference between $\int \tilde{\nu}_A(z) d\tilde{F}_Z$ and $\frac{1}{G} \sum_{g \leq G} \tilde{\nu}_A(z_g)$ is $o_p(G^{-1/2})$. Thus the difference between $\frac{1}{G} \sum_g w_g m_A^i(x_g; \hat{\gamma}_{(1)})$ and $\frac{1}{G} \sum_g [w_g m_A^i(x_g; \gamma_{(1)}^*) + \tilde{\nu}_A(z_g)]$ is $o_p(G^{-1/2})$. Under the same conditions and steps for derivations, a similar result holds with m_A^i, $\tilde{D}_{i,A}$ and $\psi_{A,s}^i$ replaced by m_Δ^i, $\tilde{D}_{i,\Delta}$ and $\psi_{\Delta,s}^i$ respectively. For condition (iii) in Assumption D to hold, it is sufficient that for $i = 1, 2$, there exists constants $c_A^i, c_\Delta^i > 0$ with $E[\sup_{\|\eta\| \leq c_A^i} \left\|\tilde{D}_{i,A,s}(x + \eta)\right\|^4] < \infty$ and $E[\sup_{\|\eta\| \leq c_\Delta^i} \left\|\tilde{D}_{i,\Delta,s}(x + \eta)\right\|^4] < \infty$ for $s = 0, 1, 2$, and $E[\sup_{\|\eta\| \leq c_A^i} \|\tilde{D}_{i,A,k}^{(s)}(x + \eta)\|^4] < \infty$ and $E[\sup_{\|\eta\| \leq c_\Delta^i} \|\tilde{D}_{i,\Delta,k}^{(s)}(x + \eta)\|^4] < \infty$ for $s = 1, 2$ and $k \in \{4, 5, 6, 7\}$.

Assumption R (i) There exists an open neighborhood $\mathcal{N}_\gamma$ around $\gamma_{(1)}^*$ with $E[\sup_{\gamma_{(1)} \in \mathcal{N}_\gamma} ||w_X m_\Delta^i(X; \gamma_{(1)})||] < \infty$. (ii) $E[||w_X m_A^i(X; \gamma_{(1)}^*) + \tilde{\nu}_A(Z)||^2] < \infty$ and $E[||w_X [m_\Delta^i(X; \gamma_{(1)}^*) - \Delta^i] + \tilde{\nu}_\Delta(Z)||^2] < \infty$.

Condition (i) in Assumption R ensures $\frac{1}{G} \sum_{g \leq G} w_g m_\Delta^i(X; \hat{\gamma}_{(1)}) \xrightarrow{p} E[w_X m_\Delta^i(X; \gamma_{(1)}^*)]$ when $\hat{\gamma}_{(1)} \xrightarrow{p} \gamma_{(1)}^*$. This is useful for deriving the correction term that results from the use of sample proportions $\frac{1}{G} \sum_g w_g$ instead of the population probability $\rho_0 \equiv E(w_X) \equiv \Pr(X \in \omega)$ while estimating Δ^i. Condition (ii) in Assumption R is necessary for the Central Limit Theorem to apply.

Lemma 1 *Suppose S, K, B, D, W and R hold. Then for $i = 1, 2$,*

$$\sqrt{G} \left(\hat{A}^i - A^i\right) \xrightarrow{d} \mathcal{N} \left(0, \; Var \left[w_X m_A^i \left(X; \gamma_{(1)}^*\right) + \tilde{\nu}_A^i(Z)\right]\right)$$
$$\sqrt{G} \left(\hat{\Delta}^i - \Delta^i\right) \xrightarrow{d} \mathcal{N} \left(0, \; \rho_0^{-2} Var \left[w_X \left(m_\Delta^i \left(X; \gamma_{(1)}^*\right) - \Delta^i\right) + \tilde{\nu}_\Delta^i(Z)\right]\right) \quad (22)$$

where $\tilde{\nu}_A^i(z) \equiv \psi_A^i(x)d - E\left[\psi_A^i(X)D\right]$ and $\tilde{\nu}_\Delta^i(z) \equiv \psi_\Delta^i(x)d - E[\psi_\Delta^i(X)D]$.

Lemma 1 follows from the steps in Section 8 of Newey and McFadden (1994) delineated above. The proof of this lemma is included in the supplement of this paper (Lewbel and Tang (2012)). It follows from Lemma 1 and Slutsky's Theorem that $\hat{\delta}_i$ converges to δ_i at the parametric rate.

Theorem 4 *Suppose A1-4 and Assumptions S, K, B, W, D and R hold. Then for $i = 1, 2$, $\Pr(\hat{\alpha}_i = \alpha_i) \to 1$ and $\sqrt{G}(\hat{\delta}_i - \delta_i) \xrightarrow{d} \mathcal{N}(0, \Sigma_\Delta^i)$, where Σ_Δ^i is the limiting variance of $\sqrt{G} \left(\hat{\Delta}^i - \Delta^i\right)$ in (22).*

Proof of Theorem 4 is included in the supplement of this paper. The rate of convergence for each $\hat{\alpha}_i$ is arbitrarily fast because each is defined as the sign of an estimator that converges at the root-N rate to its population counterpart. The parametric rate can be attained by $\hat{\delta}_i$ because it takes the form of a sample moment involving sufficiently regular preliminary nonparametric estimates. Both properties come in handy as we derive the limiting behavior of our estimator for β_i.

6.2   Estimation of β_i

Since $\Pr(\hat{\alpha}_i = \alpha) \to 1$, we derive asymptotic properties of the estimator for β_i treating α_i as if it is known. In some applications α_i's are in fact known a priori, as in the entry game example where X_i (observed components in fixed costs) must negatively affect profits. Without loss of generality, let $\alpha_i = -1$ in what follows as would be the case in the entry game.

Let $v_i^l < v_i^h$ be any two values that known to be in the support of $V_i(X)$ given $x_{-i} = (x_{e,-i}, \tilde{x})$. The dependence of v_i^l, v_i^h (and the choice of the smooth distribution H) on x_{-i} is suppressed to simplify the notation throughout this section. Choices of v_i^l, v_i^h are feasible in estimation as the support $\Omega_{V_i|x_{-i}}$ can be estimated by inf and sup of $-t + \hat{\delta}_i \hat{p}_{j,i}(t, x_{-i})$ for $t \in \Omega_{X_i|x_{-i}}$. We take v_i^l, v_i^h as known while deriving the asymptotic distribution of $\hat{\beta}_i$. Define a smooth distribution by

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

where $\mathcal{K}$ denotes the integrated bi-weight kernel function. That is, $\mathcal{K}(u) = 0$ if $u < -1$; $\mathcal{K}(u) = 1$ if $u > 1$; and $\mathcal{K}(u) = \frac{1}{16}(8 + 15u - 10u^3 + 3u^5)$ for $-1 \le u \le 1$. The integrated bi-weight kernel is continuously differentiable with its first-order derivative $\mathcal{K}'(u)$ equal to 0 if $|u| > 1$ and equal to the bi-weight kernel $\frac{15}{16}(1 - u^2)^2$ otherwise. This distribution H depends on x_{-i} through its support $[v_i^l, v_i^h]$, and it is symmetric around its expectation $\mu_H \equiv \frac{1}{2}(v_i^h + v_i^l)$. This choice of H is not necessary, but is convenient for our estimator and satisfies all the required conditions.

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

$$\hat{p}_j(x) \equiv \hat{\gamma}_j(x)/\hat{\gamma}_0(x); \quad \hat{p}_{j,i}(x) \equiv [\hat{\gamma}_0(x)]^{-2}\left[\hat{\gamma}_{j,i}(x)\hat{\gamma}_0(x) - \hat{\gamma}_j(x)\hat{\gamma}_{0,i}(x)\right];$$
$$\widehat{\mathcal{H}}_i(x) \equiv H(-x_i + \hat{\delta}_i \hat{p}_j(x)); \text{ and } \widehat{\mathcal{V}}_{(i)}(x) \equiv 1 - \hat{\delta}_i \hat{p}_{j,i}(x).$$

We use $\widehat{\mathcal{H}}_i(x)$ as an estimator for $H(V_i(x))$ and $\widehat{\mathcal{V}}_{(i)}(x)$ an estimator for $-\partial V_i(x)/\partial X_i$ in the case with $\alpha_i = -1$.

For $i = 1, 2$ and $j = 3 - i$, let $p_j \equiv \gamma_j/\gamma_0$ and $p_{j,i} \equiv \left(\gamma_{j,i}\gamma_0 - \gamma_j\gamma_{0,i}\right)/\gamma_0^2$; and $\boldsymbol{\gamma}_{(1,i)} \equiv (\gamma_0, \gamma_j, \gamma_{0,i}, \gamma_{j,i})$ denote a subvector of $\boldsymbol{\gamma}_{(1)}$. For any x, the conditional density $f(x_i|x_{-i})$ is a functional of γ_0 as $\gamma_0(x_i, x_{-i})/\int \gamma_0(t, x_{-i})dt$. Let $\boldsymbol{\gamma}_{(1,i)}^*$, p_j^*, $p_{j,i}^*$ denote true functions in the data-generating process, and $\hat{\boldsymbol{\gamma}}_{(1,i)}$, $\hat{p}_j$, $\hat{p}_{j,i}$ to denote kernel estimates. For each game $g \le G$ and both $i = 1, 2$, 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 $\hat{\beta}_i \equiv \left(\sum_g \tilde{x}_g' \tilde{x}_g\right)^{-1}\left(\sum_g \tilde{x}_g'\left(\hat{y}_{g,i} - \mu_H\right)\right)$.

Assumption S' *(a) Assumption S holds with the set ω therein replaced by Ω_X. (b) the true density $f^*(x_i|x_{-i})$ is bounded above and away from zero by some positive constant over Ω_X.*

Along with kernel and bandwidth conditions in Assumptions K and B, conditions (a) and (b) in Assumption S' ensure $\sup_{x \in \Omega_X} ||\hat{\boldsymbol{\gamma}}_{(1,i)} - \boldsymbol{\gamma}^*_{(1,i)}||$ converges in probability to 0 at rates faster than $G^{1/4}$. Bounding $\gamma^*(x)$ and $f^*(x_i|x_{-i})$ away from zero over the support Ω_X and $\Omega_{X_i|x_{-i}}$ helps attain the stochastic boundedness of $\sqrt{G} \left(\hat{\beta}_i - \beta_i \right)$.

With the excluded regressors X_e continuously distributed with positive densities almost everywhere, this requires the support of excluded regressors given X_{-i} to be bounded. Thus in order for the support of V_i given x_{-i} to cover that of $-\tilde{u}_i(\tilde{x}) + \epsilon_i$, it is necessary that the support of ϵ given $\tilde{X}$ is also bounded. (See the discussion in the next section regarding outcomes if boundedness is violated.) Condition (a) in Assumption S' also implies $p^*_{j,i}(x)$ is bounded above over Ω_X for $i = 1, 2$ and $j = 3 - i$.

Define $m^i_B \left(z; \delta_i, \boldsymbol{\gamma}_{(1,i)} \right)$ for $i = 1, 2$ as

$$m^i_B \left(z; \delta_i, \boldsymbol{\gamma}_{(1,i)} \right) \equiv \tilde{x}' \left\{ [d_i - H \left(-x_i + \delta_i p_j(x) \right)] \frac{1 - \delta_i p_{j,i}(x)}{f(x_i|x_{-i})} - \mu_H \right\},$$

with $z \equiv (d_1, d_2, x)$ and $p_j, p_{j,i}, f_{X_i|x_{-i}}$ being functions of $\boldsymbol{\gamma}_{(1,i)}$ defined above. By definition,

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

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

$$\tilde{D}_{i,B,\delta} \left(z; \delta_i, \boldsymbol{\gamma}_{(1,i)} \right) \equiv \frac{\partial}{\partial \tilde{\delta}_i} m^i_B \left(z; \tilde{\delta}_i, \boldsymbol{\gamma}_{(1,i)} \right) |_{\tilde{\delta}_i = \delta_i}$$

$$= \tilde{x}' \left\{ -H'(-x_i + \delta_i p_j(x)) \frac{p_j(x)[1 - \delta_i p_{j,i}(x)]}{f(x_i|x_{-i})} - [d_i - H(-x_i + \delta_i p_j(x))] \frac{p_{j,i}(x)}{f(x_i|x_{-i})} \right\}.$$

For $i = 1, 2$ and $j = 3 - i$, denote partial derivatives of $m^i_B \left(z; \delta_i, \boldsymbol{\gamma}_{(1,i)} \right)$ with respect to components of $\boldsymbol{\gamma}_{(1,i)}(x)$ at δ_i and $\boldsymbol{\gamma}^*_{(1,i)}(x)$ as:

$$\tilde{D}_{i,B,1}(z) \equiv \frac{\partial m^i_B \left(z; \delta_i, \boldsymbol{\gamma}^*_{(1,i)} \right)}{\partial p^*_j(x)} \frac{\partial p^*_j(x)}{\partial \gamma_0(x)} + \frac{\partial m^i_B \left(z; \delta_i, \boldsymbol{\gamma}^*_{(1,i)} \right)}{\partial p^*_{j,i}(x)} \frac{\partial p^*_{j,i}(x)}{\partial \gamma_0(x)},$$

$$\tilde{D}_{i,B,2}(z) \equiv \frac{\partial m^i_B \left(z; \delta_i, \boldsymbol{\gamma}^*_{(1,i)} \right)}{\partial p^*_j(x)} \frac{\partial p^*_j(x)}{\partial \gamma_j(x)} + \frac{\partial m^i_B \left(z; \delta_i, \boldsymbol{\gamma}^*_{(1,i)} \right)}{\partial p^*_{j,i}(x)} \frac{\partial p^*_{j,i}(x)}{\partial \gamma_j(x)},$$

$$\tilde{D}_{i,B,3}(z) \equiv \frac{\partial m^i_B \left(z; \delta_i, \boldsymbol{\gamma}^*_{(1,i)} \right)}{\partial p^*_{j,i}(x)} \frac{\partial p^*_{j,i}(x)}{\partial \gamma_{0,i}(x)}, \text{ and } \tilde{D}_{i,B,4}(z) \equiv \frac{\partial m^i_B \left(z; \delta_i, \boldsymbol{\gamma}^*_{(1,i)} \right)}{\partial p^*_{j,i}(x)} \frac{\partial p^*_{j,i}(x)}{\partial \gamma_{j,i}(x)}.$$

Let $\tilde{D}_{i,B}(z)$ denote a four-by-one vector with its k-th coordinate being $\tilde{D}_{i,B,k}(z)$ for $k = 1, 2, 3, 4$ and $i = 1, 2$. For any i and z, define

$$M^i_{B,\gamma} \left(z, \boldsymbol{\gamma}_{(1,i)} \right) = \boldsymbol{\gamma}_{(1,i)}(x)' \tilde{D}_{i,B}(z). \tag{23}$$

Note $\tilde{D}_{i,B}$ on the right-hand side of (23) is evaluated at the true parameters δ_i and $\boldsymbol{\gamma}^*_{(1,i)}$.

For both i, let $\bar{D}_{i,B,k}(x) \equiv E\left[\tilde{D}_{i,B,k}(Z)|x\right]$ for all k, and $\bar{D}_{i,B,k}^{(s)}(x) \equiv \frac{\partial[\bar{D}_{i,B,k}(x)\gamma_0^*(x)]}{\partial X_s}$ for $s = 1, 2$ and $k = 3, 4$. Define

$$
\begin{aligned}
\psi_{B,0}^i(x) &\equiv \bar{D}_{i,B,1}(x)\gamma_0^*(x) - \bar{D}_{i,B,3}^{(i)}(x) + \mathcal{I}_{B,1}^i(x); \\
\psi_{B,j}^i(x) &\equiv \bar{D}_{i,B,2}(x)\gamma_0^*(x) - \bar{D}_{i,B,4}^{(i)}(x) + \mathcal{I}_{B,2}^i(x),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{I}_{B,1}^i(x) &\equiv 1\{x_i = h_i^*(x_{-i})\}\bar{D}_{i,B,3}(x)\gamma_0^*(x) - 1\{x_i = l_i^*(x_{-i})\}\bar{D}_{i,B,3}(x)\gamma_0^*(x); \\
\mathcal{I}_{B,2}^i(x) &\equiv 1\{x_i = h_i^*(x_{-i})\}\bar{D}_{i,B,4}(x)\gamma_0^*(x) - 1\{x_i = l_i^*(x_{-i})\}\bar{D}_{i,B,4}(x)\gamma_0^*(x),
\end{aligned}
$$

with $h_i^*(x_{-i}), l_i^*(x_{-i})$ being the supreme and infimum of the support of X_i given x_{-i}.

For $i = 1, 2$ and any h twice continuously differentiable in x_c over Ω_X, define

$$
M_{B,f}^i(z; h) \equiv \frac{C_i^*(z)\int h(t, x_{-i})dt}{\gamma_0^*(x)} - \frac{C_i^*(z)h(x)\int \gamma_0^*(t, x_{-i})dt}{\gamma_0^*(x)^2}
$$

where

$$
C_i^*(z) \equiv \tilde{x}'\left[d_i - H(-x_i + \delta_i p_j^*(x))\right]\left[1 - \delta_i p_{j,i}^*(x)\right].
$$

The functional $M_{B,f}^i(z; h)$ is the Frechet derivative of $m_B^i\left(z; \delta_i, \gamma_{(1,i)}\right)$ with respect to $\gamma_{(1,i)}$ at $\gamma_{(1,i)}^*$ in DGP. Define

$$
\psi_{B,f}^i(x) \equiv E\left(\frac{C_i^*(Z)}{\gamma_0^*(X)}|x_{-i}\right)\gamma_0^*(x_{-i}) - \frac{E\left(C_i^*(Z)|x\right)}{\gamma_0^*(x)}\int \gamma_0^*(t, x_{-i})dt,
$$

where the integral is over the support $\Omega_{X_i|x_{-i}}$, and we also use $\gamma_0^*(x_{-i})$ to denote the true density of X_{-i}. Our proof builds on the observation that $\int M_{B,f}^i(z; h)dF_Z^* = \int \psi_{B,f}^i(x)h(x)dx$.

We now define the major components in the limiting distribution of $\sqrt{G}(\hat{\beta}_i - \beta_i)$:

$$
\begin{aligned}
\tilde{\nu}_B^i(z_g) &\equiv \psi_{B,0}^i(x_g) + \psi_{B,f}^i(x_g) + \psi_{B,j}^i(x_g)d_{g,j} - E\left[\psi_{B,0}^i(X) + \psi_{B,f}^i(X) + \psi_{B,j}^i(X)D_j\right]; \\
\Psi_\delta^i(z_g) &\equiv -\frac{1}{\rho_0}\left[w_g\left(m_\Delta^i\left(x_g; \gamma_{(1,i)}^*\right) - \Delta^i\right) + \tilde{\nu}_\Delta^i(z_g)\right]; \quad \mathcal{M}_{i,\delta}^* \equiv E\left[\tilde{D}_{i,B,\delta}\left(Z; \delta_i, \gamma_{(1,i)}^*\right)\right]; \\
\Psi_B^i(z) &\equiv m_B^i\left(z; \delta_i, \gamma_{(1,i)}^*\right) + \tilde{\nu}_B^i(z) + \mathcal{M}_{i,\delta}^*\Psi_\delta^i(z).
\end{aligned}
$$

The key step in finding the limiting distribution of $\sqrt{G}\left(\hat{\beta}_i - \beta_i\right)$ is to show $\frac{1}{G}\sum_g m_B^i(z_g; \hat{\delta}_i, \hat{\gamma}_{(1,i)})$ $= \frac{1}{G}\sum_g \Psi_B^i(z_g) + o_p(G^{-1/2})$. Specifically, the difference between $\frac{1}{G}\sum_g m_B^i(z_g; \hat{\delta}_i, \hat{\gamma}_{(1,i)})$ and the infeasible moment $\frac{1}{G}\sum_g m_B^i(z_g; \delta_i, \gamma_{(1,i)}^*)$ is a sample average of some correction terms $\tilde{\nu}_B^i(z) + \mathcal{M}_{i,\delta}^*\Psi_\delta^i(z)$ plus $o_p(G^{-\frac{1}{2}})$. The form of these correction terms depends on $\gamma_{(1)}^*$ in the DGP as well as how $\hat{\delta}_i, \hat{\gamma}_{(1)}$ enter the moment function m_B^i.

Assumption D' *(i) For $i = 1, 2$, the second order derivatives of $m_B^i\left(z; \delta_i, \gamma_{(1,i)}\right)$ with respect to $\gamma_{(1,i)}(x)$ at $\gamma_{(1,i)}^*(x)$ are continuous and bounded over the support of Z (i.e. $\{1, 0\}^2 \otimes \Omega_X$). (ii)*

For both i and $s = 1, 2$, $\tilde{D}_{i,B}^{(s)}$, are continuous and bounded over $\{1, 0\}^2 \otimes \Omega_X$. (iii) For $i = 1, 2$ and $j = 3 - i$, there exists $c > 0$ such that $E[\sup_{\|\eta\| \leq c} \|\psi_{B,s}^i(x + \eta)\|^4] < \infty$ for $s = 0, j$ and $E[\sup_{\|\eta\| \leq c} \|\psi_{B,f}^i(x + \eta)\|^4] < \infty$.

Assumption W' *For any i and x_{-i}, the support of X_i given x_{-i} is convex and closed.*

Assumption R' *(i) There exists open neighborhoods $\mathcal{N}_\delta$, $\mathcal{N}_i$ around δ_i^*, $\boldsymbol{\gamma}_{(1,i)}^*$ respectively such that $E[\sup_{\delta_i \in \mathcal{N}_\delta, \boldsymbol{\gamma}_{(1,i)} \in \mathcal{N}_i} \|\tilde{D}_{i,B,\delta}(Z; \delta_i, \boldsymbol{\gamma}_{(1,i)})\|] < \infty$. (ii) $E[\|\Psi_B^i(Z) - \tilde{X}'\tilde{X}\beta_i\|^2] < +\infty$. (iii) $E\left[\tilde{X}'\tilde{X}\right]$ is non-singular.*

Condition R'-(i) is used for showing the difference between $\frac{1}{G} \sum_g m_B^i(z_g; \hat{\delta}_i, \hat{\boldsymbol{\gamma}}_{(1,i)})$ and $\frac{1}{G} \sum_g m_B^i(z_g; \delta_i, \hat{\boldsymbol{\gamma}}_{(1,i)})$ is represented by a sample average of a certain function plus an $o_p(G^{-\frac{1}{2}})$ term. Under R'-(i), this function takes the form of the product of $E[\tilde{D}_{i,B,\delta}(z; \delta_i, \boldsymbol{\gamma}_{(1,i)}^*)]$ and the influence function that leads to the limiting distribution of $\sqrt{G}(\hat{\delta}_i - \delta_i)$.

Similar to the case with $\hat{\Delta}^i$, we can apply the linearization argument and the V-statistic Projection Theorem to show that, under conditions in D'-(i) and (ii), $\frac{1}{G} \sum_g [m_B^i(x_g; \delta_i, \hat{\boldsymbol{\gamma}}_{(1)}) - m_B^i(x_g; \delta_i, \boldsymbol{\gamma}_{(1)}^*)]$ can be written as $\int M_{B,\gamma}^i(z, \hat{\boldsymbol{\gamma}}_{(1,i)} - \boldsymbol{\gamma}_{(1,i)}^*) + M_{B,f}^i(z; \hat{\gamma}_0 - \gamma_0^*) dF_Z^*$ plus $o_p(G^{-\frac{1}{2}})$, where F_Z^* is the true distribution of Z in the DGP. Again by definition and smoothness properties in S', for any $\boldsymbol{\gamma}_{(1,i)}$ that is twice continuously differentiable in excluded regressors, the functions $\psi_{B,0}^i$, $\psi_{B,j}^i$ and $\psi_{B,f}^i$ can be shown to satisfy:

$$\int M_{B,\gamma}^i(z; \boldsymbol{\gamma}_{(1,i)}) + M_{B,f}^i(z; \gamma_0) dF_Z^* = \int [\psi_{B,0}^i(x) + \psi_{B,f}^i(x)]\gamma_0(x) + \psi_{B,j}^i(x)\gamma_j(x) dx$$

using an argument of integration by parts.

Using this equation, the difference $\int M_{B,\gamma}^i(z; \hat{\boldsymbol{\gamma}}_{(1,i)} - \boldsymbol{\gamma}_{(1,i)}^*) + M_{B,f}^i(z; \hat{\boldsymbol{\gamma}}_{(1,i)} - \boldsymbol{\gamma}_{(1,i)}^*) dF_Z^*$ can be expressed as $\int \tilde{\nu}_B^i(z) d\tilde{F}_Z$, where $\tilde{\nu}_B^i(z)$ was defined above and $\tilde{F}_Z$ is some smoothed version of the empirical distribution of Z as mentioned in the proof of asymptotic properties of $\hat{\delta}_i$. Condition D'-(iii) is then used for showing the difference between $\int \tilde{\nu}_B^i(z) d\tilde{F}_Z$ and $\frac{1}{G} \sum_g \tilde{\nu}_B^i(z_g)$ is $o_p(G^{-\frac{1}{2}})$. Thus the difference between $\frac{1}{G} \sum_g m_B^i(z_g; \delta_i, \hat{\boldsymbol{\gamma}}_{(1)})$ and $\frac{1}{G} \sum_g [m_B^i(z_g; \delta_i, \boldsymbol{\gamma}_{(1)}^*) + \tilde{\nu}_B^i(z_g)]$ is $o_p(G^{-1/2})$. For D'-(iii) to hold, it suffices to have that, for $i = 1, 2$, there exists $c_i > 0$ such that $E[\sup_{\|\eta\| \leq c_i} \|\bar{D}_{i,B}(x + \eta)\|^4] < \infty$ and $E[\sup_{\|\eta\| \leq c_i} \|\bar{D}_{i,B}^{(s)}(x + \eta)\|^4] < \infty$. Condition R'-(ii) ensures the Central Limit Theorem can be applied. Condition R'-(iii) is the standard full-rank condition necessary for consistency of regressor estimators. Proof of the following theorem is included in the supplement of this paper.

Theorem 5 *Under A1-4, A5" and S', K, B, D', W', R', $\sqrt{G}(\hat{\beta}_i - \beta_i) \xrightarrow{d} N(0, \Sigma_B^i)$ for $i = 1, 2$, where*

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

Estimation errors in $\hat{\delta}_i$ and $\hat{p}_j$ affect the distribution of $\hat{\beta}_i$ in general. The optimal rate of convergence of $\hat{p}_j$ is generally slower than $\sqrt{G}$ because $\hat{p}_j$ depends on the number of continuous coordinates in X, but $\hat{\beta}_i$ can still converge at the parametric rate because $\hat{\beta}_i$ takes the form of a sample average.

6.3 Discussion of β_i Estimation Rates

To obtain a parametric convergence rate for $\hat{\beta}_i$, Assumption S' assumes that excluded regressors have bounded support, and that support of V_i given some $x_{-i} = (x_{e,-i}, \tilde{x})$ includes that of $-\tilde{x}\beta_i + \epsilon_i$ given $\tilde{x}$. This necessarily requires the support of ϵ_i given $\tilde{x}$ to be bounded. As discussed earlier, such boundedness is a sensible assumption in the entry game example.

Lewbel (2000) shows that special regressor estimates can converge at parametric rates ($\sqrt{G}$ in the current context) with unbounded support, but doing so requires that that the special regressor have very thick tails. It should therefore be possible to relax the bounded support restriction in Assumption S'. However, Khan and Tamer's (2010) impossibility result shows that, in the unboundedness case, the tails would need to be thick enough to give the special regressor infinite variance.[10] Moreover our estimator would need to be modified to incorporate asymptotic trimming or some similar device to deal with the denominator of Y_i going to zero 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.

7 Monte Carlo Simulations

In this section we present evidence for the performance of the estimators in 2-by-2 entry games with linear payoffs. For $i = 1, 2$, the payoff for Firm i from entry ($D_i = 1$) is $\beta_i^0 + \beta_i^1 \tilde{X} - X_i + \delta_i D_{3-i} - \epsilon_i$,

---

[10]More precisely, Khan and Tamer (2010) showed for a single-agent, binary regression with a special regressor that the semiparametric efficiency bound for linear coefficients are not finite as long as the second moment of all regressors (including the special regressors) are finite. For a further simplified model where $Y = 1\{\alpha + v + \varepsilon \geq 0\}$ (with $v \perp \varepsilon$, α constant and both v, ε distributed over the real-line), they showed that 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.

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 set to $[\beta_1^0, \beta_1^1] = [1.8, 0.5]$, $[\beta_2^0, \beta_2^1] = [1.6, 0.8]$ 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 DGP designs: 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 $\hat{\delta}_i$ from 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 Each row relates to a different sample size G (either $5,000$ or $10,000$) and certain choices 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. $K(t) = \frac{35}{32}(1 - t^2)^3 1\{|t| \le 1\}$) in estimation. We choose the bandwidth (*b.w.*) 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^*.[11] To study how robust the performance of our estimator is against various bandwidths, we report summary statistics for our estimator under both intentional under-smoothing (with a bandwidth equal to half of *b.w.*) and over-smoothing (with a bandwidth equal to 1.5 times of *b.w.*). To implement our estimator, we estimate the set of states ω satisfying NDS by $\{x : \hat{p}_i(x), \hat{p}_j(x) \in (0, 1), \hat{p}_{i,i}(x), \hat{p}_{j,j}(x) \ne 0 \text{ and } \hat{p}_{i,i}(x)\hat{p}_{j,j}(x) \ne \hat{p}_{i,j}(x)\hat{p}_{j,i}(x)\}$.

---

[11] See Page 119 of Pagan and Ullah (1999) for definition of EPE and MASE.

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

G		$Mean$	$Std.Dev.$	LQ	$Median$	HQ	$RMSE$	MAE
	$b.w.$	[-1.397,-1.384]	[0.163,0.174]	[-1.514,-1.499]	[-1.387,-1.386]	[-1.294,-1.267]	[0.189,0.193]	[0.123,0.122]
$5k$	$\frac{1}{2}b.w.$	[-1.430,-1.433]	[0.282,0.292]	[-1.606,-1.621]	[-1.408,-1.404]	[-1.257,-1.232]	[0.310,0.320]	[0.196,0.199]
	$\frac{3}{2}b.w.$	[-1.442,-1.429]	[0.141,0.148]	[-1.536,-1.529]	[-1.436,-1.425]	[-1.344,-1.324]	[0.200,0.196]	[0.144,0.139]
	$b.w.$	[-1.380,-1.379]	[0.113,0.119]	[-1.456,-1.457]	[-1.381,-1.374]	[-1.298,-1.294]	[0.138,0.143]	[0.096,0.098]
$10k$	$\frac{1}{2}b.w.$	[-1.390,-1.394]	[0.187,0.185]	[-1.519,-1.520]	[-1.376,-1.386]	[-1.256,-1.254]	[0.207,0.207]	[0.134,0.129]
	$\frac{3}{2}b.w.$	[-1.424,-1.420]	[0.098,0.102]	[-1.488,-1.487]	[-1.421,-1.417]	[-1.352,-1.349]	[0.158,0.157]	[0.121,0.117]

The column of RMSE in Table 1(a) shows that the mean squared error (MSE) in estimation diminishes as the sample size increases from $G = 5000$ to $G = 10000$. There is also evidence for improvement of the estimator in terms median absolute errors as G increases. The trade-off between variance and bias of the estimator as bandwidth changes is also evident from the first two columns in Table 1(a). The choice of bandwidth has a moderate effect on estimator performance.

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

G		$Mean$	$Std.Dev.$	LQ	$Median$	HQ	$RMSE$	MAE
	$b.w.$	[-1.347,-1.342]	[0.095,0.090]	[-1.402,-1.410]	[-1.343,-1.335]	[-1.287,-1.282]	[0.105,0.100]	[0.071,0.062]
$5k$	$\frac{1}{2}b.w.$	[-1.343,-1.343]	[0.120,0.112]	[-1.418,-1.424]	[-1.342,-1.345]	[-1.263,-1.266]	[0.127,0.120]	[0.084,0.081]
	$\frac{3}{2}b.w.$	[-1.406,-1.440]	[0.092,0.090]	[-1.462,-1.464]	[-1.404,-1.391]	[-1.346,-1.344]	[0.141,0.135]	[0.108,0.095]
	$b.w.$	[-1.317,-1.320]	[0.061,0.061]	[-1.361,-1.356]	[-1.312,-1.314]	[-1.276,-1.276]	[0.063,0.064]	[0.042,0.038]
$10k$	$\frac{1}{2}b.w.$	[-1.309,-1.311]	[0.076,0.077]	[-1.361,-1.357]	[-1.305,-1.300]	[-1.264,-1.263]	[0.076,0.077]	[0.048,0.045]
	$\frac{3}{2}b.w.$	[-1.365,-1.368]	[0.058,0.060]	[-1.407,-1.404]	[-1.361,-1.365]	[-1.325,-1.325]	[0.087,0.090]	[0.061,0.066]

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 MSE (and MAE). A comparison between panels (a) and (b) in Table 1 suggests the estimators for δ_i, δ_j perform better under the BWK design. First, for a given sample size, the RMSE is smaller for the BWK design. Second, as sample size increases, improvement of performance (as measured by percentage reductions in RMSE) is larger for 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 construct $\hat{\delta}_i, \hat{\delta}_j$ conditioning on states in ω (i.e. those satisfying the NDS condition). 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 estimation of $\hat{\delta}_i, \hat{\delta}_j$, their presence is required for 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. Besides, 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 w.r.t. excluded regressors are replaced by true values in the DGP. Panels (a) and (b) show summary statistics from the estimates in $S = 300$ simulated samples under the uniform design.

Table 2(a): Estimator for (β_i^0, β_i^1) (Uniform Design, $G = 5k$)

		Mean	Std.Dev.	LQ	Median	HQ	RMSE	MAE
Infeasible	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.818,0.487]	[0.088,0.116]	[1.756,0.408]	[1.819,0.483]	[1.879,0.564]	[0.090,0.117]	[0.062,0.081]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.617,0.786]	[0.081,0.105]	[1.560,0.710]	[1.622,0.785]	[1.677,0.855]	[0.083,0.105]	[0.060,0.073]
b.w.	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.835,0.510]	[0.095,0.120]	[1.774,0.427]	[1.839,0.502]	[1.900,0.599]	[0.101,0.120]	[0.066,0.085]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.636,0.801]	[0.096,0.111]	[1.571,0.713]	[1.635,0.801]	[1.706,0.884]	[0.102,0.111]	[0.079,0.085]
$\frac{1}{2}$ b.w.	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.875,0.490]	[0.104,0.123]	[1.816,0.403]	[1.873,0.482]	[1.946,0.574]	[0.128,0.123]	[0.088,0.090]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.685,0.770]	[0.107,0.114]	[1.604,0.682]	[1.689,0.770]	[1.764,0.853]	[0.136,0.117]	[0.102,0.086]
$\frac{3}{2}$ b.w.	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.815,0.533]	[0.095,0.123]	[1.751,0.446]	[1.812,0.523]	[1.887,0.620]	[0.096,0.127]	[0.068,0.089]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.613,0.826]	[0.097,0.112]	[1.544,0.737]	[1.617,0.826]	[1.684,0.911]	[0.098,0.115]	[0.074,0.079]

Table 2(b): Estimator for (β_i^0, β_i^1) (Uniform Design, $G = 10k$)

		Mean	Std.Dev.	LQ	Median	HQ	RMSE	MAE
Infeasible	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.816,0.492]	[0.059,0.076]	[1.777,0.440]	[1.811,0.496]	[1.855,0.542]	[0.061,0.076]	[0.040,0.052]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.627,0.776]	[0.063,0.078]	[1.587,0.725]	[1.629,0.773]	[1.667,0.826]	[0.069,0.081]	[0.046,0.051]
b.w.	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.823,0.525]	[0.067,0.083]	[1.779,0.477]	[1.822,0.531]	[1.866,0.577]	[0.071,0.087]	[0.047,0.059]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.631,0.811]	[0.073,0.085]	[1.584,0.754]	[1.634,0.809]	[1.676,0.864]	[0.079,0.086]	[0.056,0.053]
$\frac{1}{2}$ b.w.	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.847,0.497]	[0.069,0.082]	[1.801,0.451]	[1.849,0.500]	[1.888,0.548]	[0.083,0.081]	[0.057,0.049]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.658,0.780]	[0.074,0.085]	[1.611,0.730]	[1.658,0.782]	[1.700,0.831]	[0.094,0.087]	[0.067,0.059]
$\frac{3}{2}$ b.w.	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.790,0.575]	[0.070,0.087]	[1.745,0.522]	[1.792,0.582]	[1.837,0.631]	[0.071,0.115]	[0.049,0.090]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.599,0.860]	[0.076,0.089]	[1.548,0.803]	[1.604,0.858]	[1.644,0.918]	[0.075,0.107]	[0.049,0.070]

Similar to Table 1, these panels in Table 2 show some evidence that estimators are converging (with MSE diminishing) as sample size increases. In Table 2(a) and 2(b), the choices of bandwidth now seem to have less impact on estimator performance relative to Table 1. Furthermore, the trade-off between bias and variances of $\hat{\beta}_i^0, \hat{\beta}_i^1$ as bandwidth varies is also less evident than in Table 1. This may be partly explained by the fact that the first-stage estimates $\hat{\delta}_i$ and $\hat{\delta}_j$ (which itself is a sample moment involving preliminary estimates of nuisance parameters $\hat{p}_i, \hat{p}_j, \hat{p}_{i,j}, \hat{p}_{j,i}$ etc) now enter $\hat{\beta}_1^0, \hat{\beta}_1^1$ through sample moments. While estimating β_i^0, β_i^1, these moments are then averaged over *all* observed states in data, *including* those not necessarily satisfying NDS. This second-round averaging, which involves more observations from data, possibly further mitigates the impact of bandwidths for nuisance estimates ($\hat{p}_i, \hat{p}_j$ etc) on final estimates. It is also worth noting that the performance of the feasible estimator is somewhat comparable to that of the infeasible estimators in

terms of MSE and MAE. Again, this is as expected, because the final estimator for β_i^0, β_i^1 requires further averaging of sample moments that involve preliminary estimates.

Table 2(c): Estimator for (β_i^0, β_i^1) (BWK Design, $G = 5k$)

		Mean	Std.Dev.	LQ	Median	HQ	RMSE	MAE
Infeasible	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.811,0.513]	[0.072,0.095]	[1.757,0.448]	[1.813,0.516]	[1.860,0.573]	[0.072,0.096]	[0.051,0.064]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.618,0.804]	[0.064,0.084]	[1.574,0.747]	[1.620,0.801]	[1.661,0.858]	[0.066,0.083]	[0.046,0.055]
b.w.	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.813,0.513]	[0.074,0.096]	[1.773,0.456]	[1.814,0.514]	[1.859,0.576]	[0.075,0.097]	[0.045,0.055]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.619,0.797]	[0.069,0.085]	[1.569,0.739]	[1.618,0.795]	[1.665,0.854]	[0.071,0.085]	[0.051,0.058]
$\frac{1}{2}b.w.$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.835,0.487]	[0.073,0.097]	[1.784,0.421	[1.835,0.494]	[1.881,0.544]	[0.081,0.098]	[0.058,0.061]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.641,0.776]	[0.073,0.089]	[1.593,0.716]	[1.643,0.774]	[1.688,0.831]	[0.083,0.092]	[0.061,0.066]
$\frac{3}{2}b.w.$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.796,0.574]	[0.074,0.098]	[1.742,0.503]	[1.796,0.576]	[1.848,0.637]	[0.074,0.123]	[0.053,0.090]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.619,0.835]	[0.071,0.088]	[1.564,0.775]	[1.620,0.832]	[1.665,0.892]	[0.074,0.095]	[0.053,0.056]

Table 2(d): Estimator for (β_i^0, β_i^1) (BWK Design, $G = 10k$)

		Mean	Std.Dev.	LQ	Median	HQ	RMSE	MAE
Infeasible	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.805,0.521]	[0.052,0.067]	[1.778,0.484]	[1.811,0.523]	[1.842,0.561]	[0.052,0.070]	[0.030,0.044]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.616,0.809]	[0.053,0.062]	[1.587,0.768]	[1.619,0.803]	[1.649,0.847]	[0.055,0.062]	[0.038,0.039]
b.w.	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.810,0.507]	[0.052,0.067]	[1.773,0.469]	[1.812,0.506]	[1.845,0.552]	[0.053,0.067]	[0.037,0.046]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.618,0.796]	[0.052,0.065]	[1.581,0.749]	[1.620,0.793]	[1.654,0.838]	[0.055,0.065]	[0.038,0.044]
$\frac{1}{2}b.w.$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.827,0.487]	[0.054,0.070]	[1.789,0.442]	[1.828,0.490]	[1.865,0.532]	[0.060,0.071]	[0.042,0.045]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.632,0.780]	[0.055,0.070]	[1.592,0.735]	[1.635,0.781]	[1.670,0.825]	[0.063,0.070]	[0.043,0.050]
$\frac{3}{2}b.w.$	$[\hat{\beta}_1^0, \hat{\beta}_1^1]$	[1.808,0.527]	[0.056,0.067]	[1.766,0.489]	[1.808,0.527]	[1.841,0.571]	[0.056,0.072]	[0.036,0.050]
	$[\hat{\beta}_2^0, \hat{\beta}_2^1]$	[1.619,0.807]	[0.052,0.064]	[1.582,0.764]	[1.621,0.804]	[1.655,0.849]	[0.055,0.065]	[0.040,0.043]

The remaining two panels 2(c) and 2(d) report estimator performance for the BWK design. These two panels exhibit similar patterns as noted for Table 2(a) and 2(b). The performance of $(\hat{\beta}_i^0, \hat{\beta}_i^1)$ under the BWK design is slightly better than that in the uniform design in terms of both MSE 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, should have made estimation of β^0, β_i^1 more accurate under the uniform design. However, this effect was 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 in the uniform design.

8 Extension: Multiple Bayesian Nash Equilibria

As discussed in Section 2, the model may admit multiple BNE in general. Preceding sections maintained the assumption that choices under each state are generated by a single BNE in the

data-generating process (A2). This section shows how to extend our identification and estimation strategies to allow for multiple BNE in some states.

Our approach differs qualitatively from other methods such as those surveyed in the introduction, and only requires fairly mild nonparametric restrictions. Consider the possibly unknown set of all states in which choices observed in data are rationalized by a single BNE. Denote this set by $\omega^* \subseteq \Omega_X$. Note ω^* not only includes those states in which the system of equations in (1) has a unique solution, but also includes those in which (1) admits multiple solutions in $\{s_i(x,.)\}_{i \in N}$ but for which some (possibly unknown to the researcher) equilibrium selection mechanism in data is degenerate at only one of them.

We first modify previous arguments to identify α_i and interaction effects $\delta_i(\tilde{x})$ by conditioning on states being in a known subset ω' of ω^*. Furthermore, provided the excluded regressors X_e vary sufficiently over ω', we similarly extend previous arguments to identify the baseline payoffs. These extensions of results in Sections 3.2 and 3.4 are provided in Section 8.1. Also, in Appendix B we provide sufficient conditions to guarantee that a non-empty set ω' exists which contains enough states, and hence is sufficiently large enough, to allow identification of players' payoffs in the model.

Implementation of this identification strategy assumes knowledge of some sufficiently large subset ω' of ω^*. Our final results show that this assumption is testable, that is, given a sufficiently large candidate set of states $\overline{\omega}$, we can use our data to test if $\overline{\omega}$ is a subset of ω^*, and hence test if $\overline{\omega}$ is a valid choice of a set of states ω' to use in our identification method.

8.1 Identification under multiple equilibria in some states

Corollary 4 *(Theorem 1)Suppose A1,3 hold and let ω' be an open subset of ω^* that satisfies NDS. Then (i) $(\alpha_1,.,\alpha_N)$ is identified as in (7) in Theorem 1 with ω replaced by ω'. (ii) For all $\tilde{x}$ such that $\Pr(X \in \omega'|\tilde{X} = \tilde{x}) > 0$, $(\delta_1(\tilde{x}),.,\delta_N(\tilde{x}))$ is identified as in (8) in Theorem 1 with ω replaced by ω'.*

Corollary 4 identifies $\delta_i(\tilde{x})$ at those $\tilde{x}$ for which an open subset $\omega' \subseteq \omega^*$ with $\Pr(X \in \omega'|\tilde{x}) > 0$ is known. In Appendix B, we present sufficient conditions for existence of such a ω' that also satisfies the NDS condition. 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 $(x_e, \tilde{x}) \in \omega'$.[12] The next subsection discusses how to test the assumption that a given set of states belongs to ω^*.

---

[12]This condition is sufficient for the system in (3) to admit only one solution for all $(x_e, \tilde{x}) \in \omega'$ 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.

Now consider extending the arguments that identify $\tilde{u}_i(\tilde{x})$ in Section 3 under the mean independence of ϵ_i to allow for multiple BNE in data. The idea is to use the variation of X_i as before, but now only over some subset states in ω^*.

A5‴ *For any i and $\tilde{x}$, there exists some $x_{-i} = (x_{e,-i}, \tilde{x})$ and $\omega_i \subseteq \Omega_{X_i|x_{-i}}$ such that (i) ω_i is an interval with $\Pr(X_i \in \omega_i | x_{-i}) > 0$ and $\omega_i \otimes x_{-i} \subseteq \omega^*$; (ii) the support of V_i conditional on $X_i \in \omega_i$ and x_{-i} includes the support of $\mathcal{S}_i$ given $\tilde{x}$; (iii) the density of X_i conditional on $X_i \in \omega_i$ and x_{-i} is positive a.e. on ω_i; and (iv) 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 $x_i \in \omega_i$.*

A5‴ is more restrictive than A5', because A5‴ requires the large support condition and the monotonicity condition to hold over a set of states with a unique equilibrium in DGP. The presence of excluded regressors with the assumed support properties in our model generally implies that ω^* is not empty. An intuitive sufficient condition for A5‴ to hold for some i and $\tilde{x}$ is that the other excluded regressors (except x_i) take extremely large or small values so that the system (3) admits a unique solution. (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 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.)

Nonetheless, note that such extreme values are not necessary for ω^* to be non-empty and for A5‴ to hold. This is because the equilibrium selection mechanism may well be such that even for states with non-extreme values the data is rationalized by a single BNE. For instance, in the extreme case of $\omega^* = \Omega_X$ we would be back to the scenario as assumed in A2. Conditions (i) and (iii) in A5‴ are mild regularity conditions. Condition (iv) in A5‴ can be satisfied when $\delta_i(\tilde{x})$ and the density of ϵ_i given $\tilde{x}$ are bounded above. We formalize primitive conditions that are sufficient for the existence of such x_{-i} and ω_i satisfying A5‴ in Appendix B.

For any x_{-i} and ω_i, define $Y'_{i,H}$ by replacing $f(x_i|x_{-i})$ with $f(x_i|x_{-i}, \omega_i)$ in the definition of $Y_{i,H}$ in (14). Choose a continuous distribution H as in (14), 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_H denote the expectation with respect to the chosen distribution H. Again, the potential dependence of H (and μ_H) on (x_{-i}, ω_i) is suppressed to simplify notations.

Corollary 5 *(Theorem 3) Suppose A1,3,4 hold. For any i, x_{-i} and ω_i such that A5‴ holds,*

$$\tilde{u}_i(\tilde{x}) = E\left[Y'_{i,H}|X_i \in \omega_i, x_{-i}\right] - \mu_H.$$

Once we condition on x_{-i} and ω_i with $\omega_i \otimes x_{-i} \subseteq \omega^*$, the link between model primitives and

distribution of states and actions observed from data takes the form of (1). The proof of Corollary 5 is then an extension of Corollary 2, and is omitted for brevity.

Similar to the case with unique BNE, we could also modify the identifying argument in Corollary 5 to exploit the variation of X_e over a set of states with unique BNE and $\tilde{X} = \tilde{x}$. This would be done by replacing the denominator of (11) with $f_{V_i}(V_i(x)|X_e \in \omega_e, \tilde{x})$, where ω_e consists of a set of x_e such that: (a) $\Pr(X_e \in \omega_e|\tilde{x}) > 0$; (b) $\omega_e \otimes \tilde{x} \subseteq \omega^*$; and (c) the support of V_i given "$\tilde{X} = \tilde{x}$ and $X_e \in \omega_e$" covers that of $-\tilde{u}_i(\tilde{x}) + \epsilon_i$ given $\tilde{x}$.

8.2 Testing the assumption of unique equilibrium in a subset of states

When there may be multiple BNE in data under some states, estimation of $\alpha_i, \delta_i(\tilde{x})$ and $\tilde{u}_i(\tilde{x})$ requires knowledge of some subset of ω^*. Specifically, the researcher needs to know some $\omega_e \subseteq \Omega_{X_e|\tilde{x}}$ with $\omega_e \otimes \tilde{x} \subseteq \omega^*$ in order to estimate $\alpha_i, \delta_i(\tilde{x})$, and some $\omega_i \subseteq \Omega_{X_i|x_{-i}}$ with $\omega_i \otimes x_{-i} \subseteq \omega^*$ in order to estimate $\tilde{u}_i(\tilde{x})$. The goal of this subsection is to discuss how to test the assumption that a given set of states the researcher chooses to use for estimation is in fact a subset of ω^*. More specifically, given some $\bar{\omega} \subset \Omega_X$, we modify the method in De Paula and Tang (2011) to propose a test for the null that for almost all states in $\bar{\omega}$ the actions observed in data are rationalized by a single BNE. For brevity we focus on the semiparametric case with $\delta_i(\tilde{x}) = \delta_i$ and $N = 2$.

For any generic $\bar{\omega} \subset \Omega_X$ with $\Pr(X \in \bar{\omega}) > 0$, 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 (2011), which builds on results in Manski (1993) and Sweeting (2009).

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

$$\Pr\{X \notin \omega^*|\ X \in \bar{\omega}\} > 0 \text{ if and only if } T(\bar{\omega}) \neq 0 \tag{24}$$

The intuition for (24) is that, if players' private information sets 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 data can only occur as players simultaneously move between strategies prescribed in different BNE. De Paula and Tang (2011) showed this result in a general setting where both the baseline payoffs $u_i(x, \epsilon_i)$ and the interaction effects $\delta_i(x)$ are unrestricted functions of states x. They also propose the use of a multiple testing procedure to infer the signs of strategic interaction effects in addition to the presence of multiple BNE.

Given Lemma 2, for any $\bar{\omega}$ with $\Pr\{X \in \bar{\omega}\} > 0$, the formulation of the null and alternative as

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

is equivalent to the formulation

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

Exploiting this equivalence, we propose a test statistic based on the analog principle and derive its asymptotic distribution. To simplify exposition, we focus on the case where $\tilde{X}$ is discrete (recall that X_e is continuous). The generalization to the case with continuous coordinates in $\tilde{X}$ is straightforward.

For a product kernel K with support in $\mathbb{R}^2$ and 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 $D_{g,0} \equiv 1$, $X_g \equiv (X_{g,e}, \tilde{X}_g)$ and $X_{g,e} \equiv (X_{g,1}, X_{g,2})$ denotes actions and states observed in independent games, each of which is indexed by g. Define the statistic

$$\hat{T}(\bar{\omega}) \equiv \tfrac{1}{G} \sum_{g \leq G} 1\{x_g \in \bar{\omega}\} \left[d_{1,g} d_{2,g} - \hat{p}_1(x_g) \hat{p}_2(x_g) \right], \tag{25}$$

with

$$\hat{p}_i(x) \equiv \hat{\gamma}_i(x)/\hat{\gamma}_0(x) \text{ and } \hat{\gamma}_k(x) \equiv \tfrac{1}{G} \sum_g d_{g,k} K_\sigma (x_{g,e} - x_e) 1\{\tilde{x}_g = \tilde{x}\}$$

for $i = 1, 2$ and $k = 0, 1, 2$.[13] Define $\hat{\gamma} \equiv [\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2]'$. We then have

$$\hat{T}(\bar{\omega}) = \tfrac{1}{G} \sum_g 1\{x_g \in \bar{\omega}\} m(z_g; \hat{\gamma})$$

where $z \equiv (d_1, d_2, x)$, and $m(z; \gamma) \equiv d_1 d_2 - \gamma_0^{-2}(x) \gamma_1(x) \gamma_2(x)$ for a generic vector $\boldsymbol{\gamma} \equiv [\gamma_0, \gamma_1, \gamma_2]$. For notation simplicity, suppress the dependence of w and m on $(\tilde{x}, \omega_e)$ when there is no ambiguity. Let $\boldsymbol{\gamma}^* \equiv [\gamma_0^*, \gamma_1^*, \gamma_2^*]'$, where γ_0^* denotes the true density of X and $\gamma_i^* \equiv p_i^* \gamma_0^*$ for $i = 1, 2$ (with p_i^* being choice probabilities directly identified from data). Let F^*, f^* denote the true distribution and density of z in the DGP.

The detailed conditions needed to obtain the asymptotic distribution of $\hat{T}(\bar{\omega})$ (*T1-4*) are collected in the supplement of this article (Lewbel and Tang (2012)). These conditions include assumptions on smoothness and boundedness of relevant population moments (*T1, T4*), the properties of kernel functions (*T2*) and properties of the sequence of bandwidth (*T3*).

Theorem 6 *Suppose assumptions T1-4 hold and* $\Pr\{X \in \bar{\omega}\} > 0$. *Then*

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

---

[13] Alternatively, we could replace the indicator function $1\{\tilde{x}_g = \tilde{x}\}$ with smooth kernels while estimating γ. These joint product kernels will be defined over R^J and satisfy $\int |\tilde{K}(t_1, t_2, \mathbf{0})| dt_1 dt_2 < \infty$; $\int \tilde{K}(t_1, t_2, \mathbf{0}) dt_1 dt_2 = 1$; and for any $\eta > 0$, $\sqrt{G} \sup_{||\tilde{t}|| > \eta/\sigma} \int |\tilde{K}(t_1, t_2, \tilde{\mathbf{t}})| dt_1 dt_2 \to 0$. See Bierens (1987) for details.

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 $\Sigma_{\bar{\omega}}$ is defined as

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

The covariance matrix $\Sigma_{\bar{\omega}}$ above is expressed in terms of functions that are observable in the data, and so can be consistently estimated by its sample analog under standard conditions.

9   Conclusions

We have provided conditions for nonparametric identification of static binary games with incomplete information, and associated estimators that converge at parametric rates for a semiparametric specification of the model with linear payoffs. The point identification extends to allow for the presence of multiple Bayesian Nash equilibria in the data-generating process. We show that assumptions required for identification are plausible and have ready economic interpretations in contexts such as simultaneous market entry decisions by firms.

Our identification and estimation methods depend critically on observing excluded regressors, which are state variables that affect just one player's payoffs, and do so in an additively separable way. Like all public information affecting payoffs, excluded variables still affect all player's probabilities of actions. Nevertheless, we show that a function of excluded regressors and payoffs can be identified and constructed that play the role of special regressors for identifying the model. Full identification of the model requires relatively large variation in these excluded regressors, to give these constructed special regressors a large support.

In the entry game example, 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.

Our results may also be applicable to different types of games, e.g. they might be used to identify and estimate the outcomes of household bargaining models, given sufficient data on the components of each spouse's payoffs.

Appendix A: Proofs of Identification Results

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 $f_{\mathcal{S}_i|\tilde{x}}(\alpha_i x_i + \delta_i(\tilde{x})\phi_i^*(x))$. Under Assumptions 1-3, for any x and all i,

$$p_i^*(x) = \Pr\{\mathcal{S}_i \leq \alpha_i X_i + \delta_i(\tilde{X})\phi_i^*(X)|X = x\}. \tag{26}$$

Now for any x, fix $\tilde{x}$ and differentiate both sides of (26) with respect to X_i for all i. This gives:

$$\mathcal{W}_1(x) = \mathcal{F}(x).(\mathcal{A} + \mathcal{D}(\tilde{x}).\mathcal{V}_1(x)) \tag{27}$$

where "." denotes component-wise multiplication. Furthermore, for $i \leq N - 1$, differentiate both sides of (26) w.r.t. X_{i+1} at x. Also differentiate both sides of (26) for $i = N$ with respect to X_1 at x. Stacking N equations resulting from these differentiations, we have

$$\mathcal{W}_2(x) = \mathcal{F}(x).\mathcal{D}(\tilde{x}).\mathcal{V}_2(x) \tag{28}$$

For all $x \in \omega$ (where ω satisfies NDS), 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 $x \in \omega$,

$$\mathcal{A}.\mathcal{F}(x) = \mathcal{W}_1(x) - \mathcal{W}_2(x).\mathcal{V}_1(x)./\mathcal{V}_2(x) \tag{29}$$

$$\mathcal{D}(\tilde{x}) = \mathcal{A}.\mathcal{W}_2(x)./\left[\mathcal{V}_2(x).\mathcal{W}_1(x) - \mathcal{V}_1(x).\mathcal{W}_2(x)\right], \tag{30}$$

where $\mathcal{A}$ is a constant vector and "./" denotes component-wise divisions. Note $\mathcal{F}(x)$ is a vector with strictly positive coordinates under Assumption 1. Hence signs of α_i's must be identical to those of coordinates on the right-hand side in (29) for any $x \in \omega$. Integrating out X over ω using the marginal distribution of X gives (7). The equation (30) suggests $\mathcal{D}_i(\tilde{x})$ is over-identified from the right-hand side for all $x \in \omega$. Integrating out X over ω using the distribution of X conditional on $\tilde{x}$ and $X \in \omega$ gives (8). *Q.E.D.*

Proof of Corollary 2. Fix a $x_{-i} = (x_{e,-i}, \tilde{x})$ that satisfies A5'. Then

$$E[Y_{i,H}|x_{-i}] = \int \frac{[E(D_i|X=x)-H(V_i(x))][1+\alpha_i\delta_i(\tilde{x})\phi_{i,i}^*(x)]}{f_{X_i}(x_i|x_{-i})} f_{X_i}(x_i|x_{-i})dx_i$$

$$= \int \left[E\left(D_i|V_i = V_i(x), \tilde{x}\right) - H(V_i(x))\right]\left[1 + \alpha_i\delta_i(\tilde{x})\phi_{i,i}^*(x)\right]dx_i \tag{31}$$

where the first equality follows from the law of iterated expectations and A5'-(ii); and the second from A3 and that $\epsilon_i \perp V_i(X)$ given $\tilde{x}$. The integrand in (31) is continuous in $V_i(X)$, and $V_i(X)$ is continuously differentiable in X_i given x_{-i} under A1 and A3, with $\frac{\partial V_i(X)}{\partial X_i}|_{X=x} = \alpha_i + \delta_i(\tilde{x})\phi_{i,i}^*(x)$. We suppress the dependence of $\overline{v}_i, \underline{v}_i, \mu_H$ on x_{-i} to simplify notations in the proof. Consider the case

with $\alpha_i = 1$ so that V_i is increasing in x_i given x_{-i} under A5'-(iii). In this case, $1 + \alpha_i \delta_i(\tilde{x}) \phi^*_{i,i}(x) = \alpha_i + \delta_i(\tilde{x}) \phi^*_{i,i}(x)$, and a change of variables between v_i and x_i while holding x_{-i} fixed gives us

$$E[Y_{i,H}|x_{-i}] = \int_{\underline{v}_i}^{\overline{v}_i} [E(D_i|v,\tilde{x}) - H(v)]\,dv = \int_{\underline{v}_i}^{\overline{v}_i} [E(D_i|v,\tilde{x}) - 1\{v \geq \mu_H\}]\,dv,$$

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

$$E[Y_{i,H}|x_{-i}] = \int_{\underline{v}_i}^{\overline{v}_i} \left(\int_{\Omega_{\epsilon_i|\tilde{x}}} 1\{\varepsilon_i \leq \tilde{u}_i(\tilde{x}) + v\} dF(\varepsilon_i|\tilde{x}) \right) - 1\{v \geq \mu_H\}\,dv$$

$$= \int_{\Omega_{\epsilon_i|\tilde{x}}} \int_{\underline{v}_i}^{\overline{v}_i} 1\{v \geq s_i\} - 1\{v \geq \mu_H\}\,dv\,dF(\varepsilon_i|\tilde{x}),$$

where $s_i \equiv -\tilde{u}_i(\tilde{x}) + \varepsilon_i$ and the last equality follows from a change of the order of integration between v and ε_i allowed under the support condition in A5'-(i) and the Fubini's Theorem. Then $E[Y_{i,H}|x_{-i}]$ becomes

$$\int_{\Omega_{\epsilon_i|\tilde{x}}} \int_{\underline{v}_i}^{\overline{v}_i} 1\{s_i \leq v < \mu_H\} 1\{s_i \leq \mu_H\} - 1\{\mu_H < v \leq s_i\} 1\{s_i > \mu_H\} dv\,dF(\epsilon_i|\tilde{x})$$

$$= \int_{\Omega_{\epsilon_i|\tilde{x}}} \left(1\{s_i \leq \mu_H\} \int_{s_i}^{\mu_H} dv - 1\{s_i > \mu_H\} \int_{\mu_H}^{s_i} dv \right) dF(\epsilon_i|\tilde{x})$$

$$= \int_{\Omega_{\epsilon_i|\tilde{x}}} (\mu_H - s_i)\,dF(\epsilon_i|\tilde{x}) = \tilde{u}_i(\tilde{x}) + \mu_H.$$

Next consider the case with $\alpha_i = -1$ so that V_i is decreasing in x_i given x_{-i} under A5'-(iii). By construction $1 + \alpha_i \delta_i(\tilde{x}) \phi^*_{i,i}(x) = -\alpha_i - \delta_i(\tilde{x}) \phi^*_{i,i}(x) = -\frac{\partial V_i(X)}{\partial X_i}|_{X=x}$. The proof then follows from change of variables as above. _Q.E.D._

Proof of Lemma 2. Under A1,

$$x \in \omega^* \text{ if and only if } E[D_1 D_2 | X = x] = p^*_1(x) p^*_2(x), \tag{32}$$

and $sign(E[D_1 D_2 | X = x] - p^*_1(x) p^*_2(x)) = sign(\delta_i)$ for all $x \notin \omega^*$. (See Proposition 1 in De Paula and Tang (2011) for details.)

First suppose $\Pr\{X \in \omega^* | X \in \bar{\omega}\} = 1$. Integrating out X in (32) given "$X \in \bar{\omega}$" shows $E[D_1 D_2 - p^*_1(X) p^*_2(X) | X \in \bar{\omega}]$, which in turn implies $T(\bar{\omega}) = 0$. Next suppose $\Pr\{X \notin \omega^* | X \in \bar{\omega}\} > 0$. Then $sign(\delta_1)$ must necessarily equal $sign(\delta_2)$. (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 states. For all x in the latter intersection, $E[D_1 D_2 | x] - p^*_1(x) p^*_2(x)$ is non-zero and its sign is equal to $sign(\delta_i)$ for $i = 1, 2$. For all x in the

former intersection (with ω^*), $E[D_1D_2|x] - p_1^*(x)p_2^*(x)$ is equal to 0. Applying the Law of Iterated Expectations using these two partitioning intersections shows $E\left[D_1D_2 - p_1^*(X)p_2^*(X)|X \in \bar{\omega}\right] \neq 0$ and its sign is equal to $sign(\delta_i)$. This in turn implies $T(\bar{\omega}) \neq 0$ with $sign(T(\bar{\omega})) = sign(\delta_i)$ for $i = 1, 2$.   *Q.E.D.*

Appendix B: NDS, Large Support and Monotonicity Conditions

In this section, we provide conditions on the model primitives that are sufficient for some of the main identifying restrictions stated in the text. These include the NDS condition for identifying α_i and $\delta_i(\tilde{x})$; the large support condition in A5, A5',A5" and A5''' for identifying component of the baseline payoff $\tilde{u}_i(\tilde{x})$; as well as the monotonicity condition in A5', A5" and A5''' that the sign of marginal effects of excluded regressors on generated special regressors are identical to the sign of α_i.

B1. Sufficient conditions for NDS

We now present fairly mild sufficient conditions for there to exist a subset of states $\omega \subseteq \omega^*$ that satisfies NDS. Consider a game with $N = 2$ (with players denoted by 1 and 2) and $h_i(D_j) = D_j$ for $i = 1, 2$ and $j = 3 - i$. We begin by reviewing the sufficient conditions under which the system characterizing BNE in (3) admits only a unique solution at x. Let $p \equiv (p_1, p_2)$ and define:

$$\varphi(p; X) \equiv \left[\begin{array}{c} \varphi_1(p; X) \\ \varphi_2(p; X) \end{array} \right] \equiv \left[\begin{array}{c} p_1 - F_{\mathcal{S}_1|\tilde{X}}\left(\alpha_1 X_1 + \delta_1(\tilde{X})p_2\right) \\ p_2 - F_{\mathcal{S}_2|\tilde{X}}\left(\alpha_2 X_2 + \delta_2(\tilde{X})p_1\right) \end{array} \right]$$

By Theorem 7 in Gale and Nikaido (1965) and Proposition 1 in Aradillas-Lopez (2010), the solution for p in the fixed point equation $p = \varphi(p; x)$ at any x is unique if none of the principal minors of the Jacobian of $\varphi(p; x)$ with respect to p vanishes on $[0, 1]^2$. Or equivalently, if $\Pi_{i=1,2}\delta_i(\tilde{x})f_{\mathcal{S}_i|\tilde{x}}(\alpha_i x_i + \delta_i(\tilde{x})p_{3-i}(x)) \neq 1$ for all $p \in [0, 1]^2$ at x. Let $I_i(\tilde{x})$ denote the interval between 0 and $\delta_i(\tilde{x})$ for any $\tilde{x}$. That is, $I_i(\tilde{x}) \equiv [0, \delta_i(\tilde{x})]$ if $\delta_i(\tilde{x}) > 0$ and $[\delta_i(\tilde{x}), 0]$ otherwise. For any $\tilde{x}$, define:

$$\omega^a(\tilde{x}) \equiv \left\{ x_e : f_{\mathcal{S}_i|\tilde{x}}(t + \alpha_i x_i) \text{ bounded away from 0 for all } t \in I_i(\tilde{x}) \text{ and } i = 1, 2 \right\}; \text{ and}$$

$$\omega^b(\tilde{x}) \equiv \left\{ x_e : \begin{array}{c} \Pi_i \min_{t \in I_i(\tilde{x})} f_{\mathcal{S}_i|\tilde{x}}(t + \alpha_i x_i) > (\delta_1(\tilde{x})\delta_2(\tilde{x}))^{-1} \quad or \\ \Pi_i \max_{t \in I_i(\tilde{x})} f_{\mathcal{S}_i|\tilde{x}}(t + \alpha_i x_i) < (\delta_1(\tilde{x})\delta_2(\tilde{x}))^{-1} \end{array} \right\}$$

Proposition B1 *Suppose A1,3 hold with $N = 2$. (a) If $\delta_1(\tilde{x})\delta_2(\tilde{x}) < 0$ and $\Pr\left\{X_e \in \omega^a(\tilde{x})|\tilde{x}\right\} > 0$, then $\omega^a(\tilde{x}) \otimes \tilde{x} \subseteq \omega^*$ and satisfies NDS. (b) If $\delta_1(\tilde{x})\delta_2(\tilde{x}) > 0$ and $\Pr\left\{X_e \in \omega^a(\tilde{x}) \cap \omega^b(\tilde{x})|\tilde{x}\right\} > 0$, then $\left(\omega^a(\tilde{x}) \cap \omega^b(\tilde{x})\right) \otimes \tilde{x} \subseteq \omega^*$ and satisfies NDS.*

Proof of Proposition B1. First consider $\tilde{x}$ with $\Pi_{i=1,2}\delta_i(\tilde{x}) < 0$. Then $\Pi_{i=1,2}\delta_i(\tilde{x})\, f_{\mathcal{S}_i|\tilde{x}}(\alpha_i x_i + \delta_i(\tilde{x})p_{3-i}(x)) < 0$ for all $x_e \in \omega^a(\tilde{x})$ and all p on $[0,1]^2$. It then follows that $(x_e, \tilde{x}) \in \omega^*$ for all $x_e \in \omega^a(\tilde{x})$. That $\frac{\partial p_i^*(X)}{\partial X_i}|_{X=x}$, $\frac{\partial p_i^*(X)}{\partial X_{3-i}}|_{X=x}$ exist for $i = 1, 2$ a.e. on $\omega^a(\tilde{x})$ follows from the exclusion restrictions and smoothness conditions in Assumption 3. We now show $\Pr\{p_{i,i}^*(X)p_{j,j}^*(X) \neq p_{i,j}^*(X)p_{j,i}^*(X)$ and $p_{j,j}^*(X) \neq 0 \mid X_e \in \omega^a(\tilde{x}), \tilde{X} = \tilde{x}\} = 1$. Suppose this probability is strictly less than 1 for $i = 1$. Suppose $\Pr\{\frac{\partial p_2^*(X)}{\partial X_2} = 0| X_e \in \omega^a(\tilde{x}), \tilde{X} = \tilde{x}\} > 0$. By (5)-(6) for $i = 1$, this implies there is a set of (x_1, x_2) on $\omega^a(\tilde{x})$ with positive probability such that both $\frac{\partial p_2^*(x)}{\partial X_2}$ and $\frac{\partial p_1^*(x)}{\partial X_2}$ are 0. But then by (5)-(6) for $i = 2$ implies the *same* set of (x_1, x_2) must satisfy $\frac{\partial p_1^*(x_1, x_2, \tilde{x})}{\partial X_2} = -\alpha_2/\delta_1(\tilde{x}) \neq 0$, because the conditional densities of $\mathcal{S}_i$ over the relevant section of its domain is nonzero by definition of $\omega^a(\tilde{x})$. Contradiction. Hence $\Pr\{\frac{\partial p_2^*(X)}{\partial X_2} = 0| X_e \in \omega^a(\tilde{x}), \tilde{X} = \tilde{x}\} = 0$. Symmetric arguments prove the same statement with $\frac{\partial p_2^*(X)}{\partial X_2}$ replaced by $\frac{\partial p_1^*(X)}{\partial X_1}$. Now suppose $\Pr\{\frac{\partial p_i^*(X)}{\partial X_i}\frac{\partial p_{3-i}^*(X)}{\partial X_{3-i}} = \frac{\partial p_i^*(X)}{\partial X_{3-i}}\frac{\partial p_{3-i}^*(X)}{\partial X_i} \mid X_e \in \omega^a(\tilde{x}), \tilde{X} = \tilde{x}\} > 0$. By construction, for all x_e in $\omega^a(\tilde{x})$, the relevant conditional densities of $\mathcal{S}_i$ must be positive for $i = 1, 2$. Hence

$$\frac{\partial p_1^*(x)}{\partial X_1}\frac{\partial p_2^*(x)}{\partial X_2} - \frac{\partial p_1^*(x)}{\partial X_2}\frac{\partial p_2^*(x)}{\partial X_1} = \frac{\alpha_1}{\delta_1(\tilde{x})}\frac{\partial p_1^*(x)}{\partial X_2} \tag{33}$$

for all x_e on $\omega^a(\tilde{x})$. Because $\Pr\{\frac{\partial p_2^*(X)}{\partial X_2} \neq 0 \mid X_e \in \omega^a(\tilde{x}), \tilde{X} = \tilde{x}\} = 1$ as argued above, (5)-(6) for $i = 1$ and the definition of $\omega^a(\tilde{x})$ implies $\Pr\{\frac{\partial p_1^*(X)}{\partial X_2} \neq 0 \mid X_e \in \omega^a(\tilde{x}), \tilde{X} = \tilde{x}\} = 1$. Hence $\Pr\{p_{1,1}^*(X)p_{2,2}^*(X) = p_{1,2}^*(X)p_{2,1}^*(X) \mid X_e \in \omega^a(\tilde{x}), \tilde{X} = \tilde{x}\} = 0$. This proves part (a). Next consider $\tilde{x}$ with $\delta_1(\tilde{x})\delta_2(\tilde{x}) > 0$. Because $\omega^a(\tilde{x}) \cap \omega^b(\tilde{x}) \subseteq \omega^a(\tilde{x})$, arguments in the proof of part (a) applies to show that the non-zero requirements in the NDS condition must hold over $(\omega^a(\tilde{x}) \cap \omega^b(\tilde{x})) \otimes \tilde{x}$ under Assumptions 1,3. Furthermore, by construction, for any (x_1, x_2) in $\omega^b(\tilde{x})$, either $\Pi_i\delta_i(\tilde{x})f_{\mathcal{S}_i|\tilde{x}}(\alpha_i x_i + \delta_i(\tilde{x})p_{3-i}(x)) > 1$ or $\Pi_i\delta_i(\tilde{x})f_{\mathcal{S}_i|\tilde{x}}(\alpha_i x_i + \delta_i(\tilde{x})p_{3-i}(x)) < 1$ for all $p \in [0,1]^2$. Hence the BNE (as solutions to (3)) must be unique for all states in $\omega^b(\tilde{x}) \otimes \tilde{x}$. This proves $(\omega^a(\tilde{x}) \cap \omega^b(\tilde{x})) \otimes \tilde{x}$. *Q.E.D.*

B2. Sufficient conditions for various large support conditions

We begin with sufficient conditions on model primitives that are sufficient for the large support condition in A5‴. We will then give primitive conditions that are sufficient for large support restrictions in A5' and A5". Consider the case with $N = 2$. For any $\tilde{x}$, let s_{il}, s_{ih} denote the infimum and supremum of the conditional support of $\mathcal{S}_i$ given $\tilde{x}$ respectively, and let $\delta_i^+(\tilde{x}) \equiv \max\{0, \delta_i(\tilde{x})\}$ and $\delta_i^-(\tilde{x}) \equiv \min\{0, \delta_i(\tilde{x})\}$.

Proposition B2 *Let $N = 2$ and $h_i(D_j) = D_j$ for both $i = 1, 2$ and $j = 3 - i$, and let A1,3 hold. Suppose for i and $x_{-i} = (x_{e,-i}, \tilde{x})$, there exists an interval $\omega_i \subseteq \Omega_{X_i|x_{-i}}$ such that $\omega_i \otimes x_{-i} \in \omega^*$ and the support of $\alpha_i X_i$ given $X_i \in \omega_i$ and x_{-i} is an interval that contains $[s_{il} - \delta_i^+(\tilde{x}), s_{ih} - \delta_i^-(\tilde{x})]$. Then the set ω_i satisfies (i) ω_i is an interval with $\Pr(X_i \in \omega_i|x_{-i}) > 0$; and (ii) support of S_i given $\tilde{x}$ is included in the support of V_i given x_{-i}.*

Proof of Proposition B2. Because $\omega_i \otimes x_{-i} \in \omega^*$, the smoothness conditions in A1,3 imply for all i that $p_i^*(x)$ is continuous in x_i over ω_i. Since ω_i is an interval (path-wise connected), the support of V_i given x_{-i} and $X_i \in \omega_i(x_{-i})$ must also be an interval. By A2 and the condition in Proposition C2, $\Pr\{\alpha_i X_i + \delta_i^-(\tilde{X}) > s_{ih}| \ x_{-i}, \ X_i \in \omega_i\} > 0$. This implies $\Pr\{V_i > s_{ih}| \ x_{-i}, \ X_i \in \omega_i\} > 0$. Symmetrically, $\Pr\{\alpha_i X_i + \delta_i^+(\tilde{X}) < s_{il} \ | \ x_{-i}, \ X_i \in \omega_i\} > 0$ implies $\Pr\{V_i < s_{il}| \ x_{-i}, \ X_i \in \omega_i\} > 0$. Since the support of V_i conditional on x_{-i} and $X_i \in \omega_i$ is an interval, the definition of s_{il}, s_{ih} suggests the conditional support of V_i must contain the support of $\mathcal{S}_i \equiv -u_i(\tilde{X}) + \epsilon_i$ given $\tilde{x}$. Q.E.D.

The large support condition in A5''' holds if for all $\tilde{x}$ there exists $x_{-i} = (x_{e,-i}, \tilde{x})$ that satisfies conditions in Proposition B2.

Next, we give sufficient primitive conditions for large support restrictions in A5' and A5''. Under A2, $\omega^* = \Omega_X$. Hence with A2 in the background (as is the case with Sections 5 and 6), the large support condition in A5'' holds when conditions in Proposition B2 are satisfied by $\omega_i = \Omega_{X_i|x_{-i}}$ *for all* $x_{-i} \in \Omega_{X_{-i}|\tilde{x}}$. Furthermore, under A2, the large support condition in A5' is satisfied as long as the conditions in Proposition B2 are satisfied by $\omega_i = \Omega_{X_i|x_{-i}}$ *by some* $x_{-i} \in \Omega_{X_{-i}|\tilde{x}}$. The large support condition in A5 is implied by that in A5'. As stated in the text, these various versions of large support conditions are verifiable using observed distributions.

B3. Sufficient conditions for monotonicity conditions

For simplicity in exposition, consider the case with $N = 2$, $h(D_{-i}) = D_j$ and $\alpha_i = -1$. For all $x \in \Omega_X$, we can solve for $p_{j,i}(x)$ using the system of fixed point equation in the marginal effects of excluded regressors on the choice probabilities. By construction, this gives:

$$\partial V_i(X)/\partial X_i|_{X=(x_e,\tilde{x})} = -1 + \delta_i(\tilde{x})p_{j,i}(x) = \frac{1}{\delta_i(\tilde{x})\delta_j(\tilde{x})\tilde{f}_i(x)\tilde{f}_j(x)-1}$$

as $\delta_i(\tilde{x})\delta_j(\tilde{x})\tilde{f}_i(x)\tilde{f}_j(x) \neq 1$ almost surely. Thus the sign of $\partial V_i(X)/\partial X_i|_{X=(x_e,\tilde{x})}$ is identical to the sign of $\delta_i(\tilde{x})\delta_j(\tilde{x})\tilde{f}_i(x)\tilde{f}_j(x) - 1$. For the monotonicity condition in A5'' to hold over $x \in \Omega_X$, it suffices to have two positive constants c_1, c_2 such that $c_1 c_2 < 1$ and $sign(\delta_i(\tilde{x})) = sign(\delta_j(\tilde{x}))$ and $|\delta_i(\tilde{x})|, |\delta_j(\tilde{x})| \leq c_1$ and $f_{\epsilon_i|\tilde{x}}, f_{\epsilon_j|\tilde{x}}$ bounded above by some positive constant c_2 uniformly over $\Omega_{\epsilon_i|\tilde{x}}, \Omega_{\epsilon_i|\tilde{x}}$ and $\tilde{x} \in \Omega_{\tilde{X}}$. This would imply the monotonicity condition in A5' as well.

In A5', the monotonicity condition is only required to hold for some $x_{-i} = (x_{e,-i}, \tilde{x})$. Hence it can also hold without such uniform bound conditions in the previous paragraph. It suffices to have δ_i, δ_j bounded and some $x_{e,-i}$ large or small enough so that the first term in the denominator is smaller than 1 for the $\tilde{x}$ and $x_{-i} = (x_{e,-i}, \tilde{x})$ considered.

References

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

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

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

[4] Bajari, P., J. Hahn, H. Hong and G. Ridder, "A Note on Semiparametric Estimation of Finite Mixtures of Discrete Choice Models with Application to Game Theoretic Models'" forthcoming, International Economic Review

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

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

[7] Beresteanu, A., Molchanov, I. and F. Molinari, "Sharp Identification Regions in Models with Convex Moment Predictions," Econometrica, forthcoming, 2011

[8] Berry, S. and P. Haile, "Identification in Differentiated Products Markets Using Market Level Data", Cowles Foundation Discussion Paper No.1744, Yale University, 2010

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

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

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

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

[13] Ciliberto, F., and E. Tamer, "Market Structure and Multiple Equilibria in Airline Markets," Econometrica, 77, 1791-1828.

[14] De Paula, A. and X. Tang, "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information," working paper, University of Pennsylvania, 2011.

[15] Khan, S. and E. Tamer, "Irregular Identification, Support Conditions and Inverse Weight Estimation," forthcoming, Econometrica, June 2010.

[16] Klein, R., and W. Spady (1993): "An Efficient Semiparametric Estimator of Binary Response Models," Econometrica, 61, 387–421.

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

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

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

[20] Lewbel, A. "Semiparametric Qualitative Response Model Estimation with Unknown Heteroskaedasticity. or Instrumental Variables," Journal of Econometrics, 2000

[21] Lewbel, A. "An Overview of the Special Regressor Method," Boston College Working Paper, 2012.

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

[23] Lewbel, A. and X. Tang, "Supplement for 'Identification and Estimation of Games with Incomplete Information Using Excluded Regressors'", unpublished manuscript, 2012.

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

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

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

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

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

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

[30] Wan, Y., and H. Xu, "Semiparametric Estimation of Binary Decision Games of Incomplete Information with Correlated Private Signals," Working Paper, Pennsylvania State University, 2010

[31] Xu, H., "Estimation of Discrete Game with Correlated Private Signals," Working Paper, Penn State University, 2009

Supplement for "Identification and Estimation of Games with Incomplete Information Using Excluded Regressors"

Arthur Lewbel, Boston College

Xun Tang, University of Pennsylvania

1 Proofs of Asymptotics

Assumptions used in the results below are all listed in the main text of Lewbel and Tang (2012).

1.1 Limiting distribution of $\sqrt{G}\left(\hat{A}_i - A_i\right)$

To simplify notation, for a pair of players i and $j = 3 - i$, we define the following shorthands: $a = \gamma_0^*$, $b = \gamma_i^*$, $c = \gamma_j^*$, $d = \gamma_{0,i}^*$, $e = \gamma_{0,j}^*$, $f = \gamma_{i,i}^*$, $g = \gamma_{i,j}^*$, $h = \gamma_{j,i}^*$ and $k = \gamma_{j,j}^*$, with all these functions evaluated at x. As in the text, let w_g, w_x denote $1\{x_g \in \omega\}$ and $1\{x \in \omega\}$ respectively. For any $x \in \Omega_X$, define a linear functional of $\boldsymbol{\gamma}_{(1)}$:

$$M_A^i\left(x; \boldsymbol{\gamma}_{(1)}\right) \equiv \left[\boldsymbol{\gamma}_{(1)}(x)\right]' \tilde{D}_{i,A}(x), \text{ where}$$

$$\tilde{D}_{i,A}\left(x; \boldsymbol{\gamma}_{(1)}^*\right) \underset{\text{9-by-1}}{\equiv} \left[\begin{array}{c} \frac{-1}{a^2(ce-ak)^2}\left(\begin{array}{c} c^2 fe^2 + a^2 fk^2 - c^2 dge - 2abdk^2 - a^2 ghk \\ -bche^2 + bcdke - 2acfke + 2abhke + 2acdgk \end{array}\right),.. \\ \frac{1}{a(ak-ce)}(he-dk), \ \frac{be-ag}{a(ce-ak)^2}(he-dk), \ \frac{1}{a(ak-ce)}(cg-bk),.. \\ -\frac{1}{a^2 e}\left(bd - af - \frac{(be-ag)(cd-ah)}{ak-ce}\right), \ \frac{1}{a}, \ \frac{1}{a(ak-ce)}(cd-ah),.. \\ \frac{1}{a(ak-ce)}(be-ag), \ \frac{be-ag}{a(ak-ce)^2}(cd-ah) \end{array}\right]'. \quad (1)$$

Lemma 1.1 *Suppose Assumptions S, K, B and D-(i),(ii) hold. Then for* $i = 1, 2$,

$$\frac{1}{G}\sum_g w_g \left[m_A^i\left(x_g; \hat{\boldsymbol{\gamma}}_{(1)}\right) - m_A^i\left(x_g; \boldsymbol{\gamma}_{(1)}^*\right)\right] = \int w_x M_A^i\left(x; \hat{\boldsymbol{\gamma}}_{(1)} - \boldsymbol{\gamma}_{(1)}^*\right) dF_X^* + o_p(G^{-1/2}).$$

Proof of Lemma 1.1. Consider $i = 1$. Under D-(i), (ii), there exists some $b(x)$ such that $E[b(x)] < \infty$ and

$$\left\| w_g \left[m_A^i\left(x_g; \hat{\boldsymbol{\gamma}}_{(1)}\right) - m_A^i\left(x_g; \boldsymbol{\gamma}_{(1)}^*\right) - \left(\hat{\boldsymbol{\gamma}}_{(1)}(x_g) - \boldsymbol{\gamma}_{(1)}^*(x_g)\right)' \tilde{D}_{i,A}(x_g)\right] \right\|$$

$$\leq b(x_g) \sup_{x \in \omega} \left\| \hat{\boldsymbol{\gamma}}_{(1)} - \boldsymbol{\gamma}_{(1)}^* \right\|^2$$

Then Lemma 8.10 in Newey and McFadden (1994) applies under S-(i),(ii),(iii), K and B, with the order of derivatives being 1 in the definition of the Sobolev Norm and with the dimension of continuous coordinates being J_c. Hence $\sup_{x\in\omega}\left\|\hat{\boldsymbol{\gamma}}_{(1)} - \boldsymbol{\gamma}_{(1)}^*\right\| = o_p(G^{-1/4})$, and

$$\frac{1}{G}\left\|\begin{array}{c}\sum_g w_g\left[m_A^i\left(x_g;\hat{\boldsymbol{\gamma}}_{(1)}\right) - m_A^i\left(x_g;\boldsymbol{\gamma}_{(1)}^*\right)\right]\\ -\sum_g w_g\left(\hat{\boldsymbol{\gamma}}_{(1)}(x_g) - \boldsymbol{\gamma}_{(1)}^*(x_g)\right)' \tilde{D}_{i,A}(x_g)\end{array}\right\| = o_p(G^{-1/2}).$$

Next, define $\bar{\boldsymbol{\gamma}}_{(1)} \equiv E\left[\hat{\boldsymbol{\gamma}}_{(1)}\right]$ and note:

$$G^{-1}\sum_g w_g\left(\hat{\boldsymbol{\gamma}}_{(1)}(x_g) - \boldsymbol{\gamma}_{(1)}^*(x_g)\right)' \tilde{D}_{i,A}(x_g) - \int w_x\left(\hat{\boldsymbol{\gamma}}_{(1)}(x) - \boldsymbol{\gamma}_{(1)}^*(x)\right)' \tilde{D}_{i,A}(x)dF_X^*$$

$$= G^{-1}\sum_g w_g\left(\hat{\boldsymbol{\gamma}}_{(1)}(x_g) - \bar{\boldsymbol{\gamma}}_{(1)}(x_g)\right)' \tilde{D}_{i,A}(x_g) - \int w_x\left(\hat{\boldsymbol{\gamma}}_{(1)}(x) - \bar{\boldsymbol{\gamma}}_{(1)}(x)\right)' \tilde{D}_{i,A}(x)dF_X^*$$

$$+ G^{-1}\sum_g w_g\left(\bar{\boldsymbol{\gamma}}_{(1)}(x_g) - \boldsymbol{\gamma}_{(1)}^*(x_g)\right)' \tilde{D}_{i,A}(x_g) - \int w_x\left(\bar{\boldsymbol{\gamma}}_{(1)}(x) - \boldsymbol{\gamma}_{(1)}^*(x)\right)' \tilde{D}_{i,A}(x)dF_X^* \quad (2)$$

Let $a_G(z_g, z_s) \equiv$

$$w_g\left[\begin{array}{l}K_\sigma(x_s - x_g),\ d_{s,1}K_\sigma(x_s - x_g),\ d_{s,2}K_\sigma(x_s - x_g),\ ...\\ K_{\sigma,1}(x_s - x_g),\ d_{s,1}K_{\sigma,1}(x_s - x_g),\ d_{s,2}K_{\sigma,1}(x_s - x_g),\ ...\\ K_{\sigma,2}(x_s - x_g),\ d_{s,1}K_{\sigma,2}(x_s - x_g),\ d_{s,2}K_{\sigma,2}(x_s - x_g)\end{array}\right]' \tilde{D}_{i,A}(x_g).$$

Let $a_{G,1}(z_g) \equiv \int a(z_g, z)\ dF_Z^*(z)$ and $a_{G,2}(z_s) \equiv \int a_G(z, z_s)dF_Z^*(z)$. The first difference on the right-hand side of (2) takes the form of a V-statistic:

$$G^{-2}\sum_g\sum_s a_G(z_g, z_s) - G^{-1}\sum_g a_{G,1}(z_g) - G^{-1}\sum_g a_{G,2}(z_g) + E\left[a_{G,1}(z)\right]$$

By Assumption D-(i) and boundedness of kernels in K, $\left\|w_x M_A^i\left(x;\boldsymbol{\gamma}_{(1)}\right)\right\| \leq c(x)\sup_{x\in\omega}\left\|\boldsymbol{\gamma}_{(1)}\right\|$ for some non-negative function c such that $E[c(X)^2] < \infty$. Consequently, $E[\|a_G(Z_g, Z_g)\|] \leq \sigma^{-J}E\left[c(X_g)\right]C_1$ and $\left\{E[\|a_G(Z_g, Z_s)\|^2]\right\}^{1/2} \leq \sigma^{-J}\left\{E[c^2(X_g)]\right\}^{1/2}C_2$, where C_1, C_2 are finite positive constants. Since $G^{-1}\sigma^{-J} \to 0$, it follows from Lemma 8.4 in Newey and McFadden (1994) that the first difference is $o_p(G^{-1/2})$.

We now turn to the second difference in (2). Since $\left\|w_x M_A^i\left(x;\boldsymbol{\gamma}_{(1)}\right)\right\| \leq c(x)\sup_{x\in\omega}\left\|\boldsymbol{\gamma}_{(1)}\right\|$ and $E\left[c(X)^2\right] < \infty$,

$$E\left[\left\|w_X M_A^i\left(X\ ;\ \bar{\boldsymbol{\gamma}}_{(1)} - \boldsymbol{\gamma}_{(1)}^*\right)\right\|^2\right] \leq E\left[c(X)^2\right]\sup_{x\in\omega}\left\|\bar{\boldsymbol{\gamma}}_{(1)} - \boldsymbol{\gamma}_{(1)}^*\right\|^2.$$

By Lemma 8.9 in Newey and McFadden (1994) and Hardle and Linton (1994), we have $\sup_{x\in\omega}\left\|\bar{\boldsymbol{\gamma}}_{(1)} - \boldsymbol{\gamma}_{(1)}^*\right\|^2 = O(\sigma^m)$ under conditions in S, K and B. Thus, as $G \to +\infty$,

$$E\left[\sup_{x\in\omega}\left\|w_X M_A^i\left(X\ ;\ \bar{\boldsymbol{\gamma}}_{(1)} - \boldsymbol{\gamma}_{(1)}^*\right)\right\|^2\right] \to 0$$

Then by the Chebychev's Inequality, the second difference in (2) is $o_p(G^{-1/2})$. This proves the lemma.   Q.E.D.

Lemma 1.2 *Suppose Assumptions S, K, B, D and W hold. Then*

$$\int w_x M_A^i \left(x; \hat{\boldsymbol{\gamma}}_{(1)} - \boldsymbol{\gamma}_{(1)}^* \right) dF_X^* = G^{-1} \sum_g \tilde{\nu}_A^i(z_g) + o_p(G^{-1/2}),$$

where $\tilde{\nu}_A^i(z_g) \equiv \psi_A^i(x_g)d_g - E\left[\psi_A^i(X)D\right]$.

Proof of Lemma 1.2. For $i = 1, 2$ and $j = 3 - i$, by definition of M_A^i,

$$\int w_x M_A^i \left(x; \boldsymbol{\gamma}_{(1)} \right) dF_X^*$$

$$= \int w_x \begin{bmatrix} \gamma_0(x) \\ \gamma_i(x) \\ \gamma_j(x) \end{bmatrix}' \begin{bmatrix} \tilde{D}_{i,A,1}(x) \\ \tilde{D}_{i,A,2}(x) \\ D_{i,A,3}(x) \end{bmatrix} + w_x \begin{bmatrix} \gamma_{0,i}(x), \gamma_{0,j}(x), .. \\ \gamma_{i,i}(x), \gamma_{i,j}(x), .. \\ \gamma_{j,i}(x), \gamma_{j,j}(x) \end{bmatrix}_{1-by-6} \begin{bmatrix} \tilde{D}_{i,A,4}(x) \\ \vdots \\ \tilde{D}_{i,A,9}(x) \end{bmatrix}_{6-by-1} dx. \qquad (3)$$

Applying integration-by-parts to the second inner product in the integrand of (3) and using the convexity of ω (Assumption W), we have

$$\int w_x M_A^i \left(x; \boldsymbol{\gamma}_{(1)} \right) dF_X^* = \int \begin{bmatrix} \gamma_0(x) \\ \gamma_i(x) \\ \gamma_j(x) \end{bmatrix}' \begin{bmatrix} \psi_{A,0}^i(x) \\ \psi_{A,i}^i(x) \\ \psi_{A,j}^i(x) \end{bmatrix} dx,$$

for any $\boldsymbol{\gamma}_{(1)}$ that is continuously differentiable in excluded regressors x_e. The functions $\psi_{A,0}^i, \psi_{A,i}^i, \psi_{A,j}^i$ are defined in the text. (See the web appendix of this supplement for detailed derivations.) These three functions only depend on $\boldsymbol{\gamma}_{(1)}^*$ but not $\boldsymbol{\gamma}_{(1)}$. By conditions in S and K, both $\hat{\boldsymbol{\gamma}}_{(1)}$ and $\boldsymbol{\gamma}_{(1)}^*$ are continuously differentiable. For the rest of the proof, we focus on the case where all coordinates in $\tilde{x}$ are continuously distributed.[1] It follows that:

$$\int w_x M_A^i \left(x; \hat{\boldsymbol{\gamma}}_{(1)} - \boldsymbol{\gamma}_{(1)}^* \right) dF_X^* = \int \sum_{s=0,i,j} \psi_{A,s}^i(x) \left[\hat{\gamma}_s(x) - \gamma_s^*(x)\right] dx$$

$$= G^{-1} \sum_g \int \sum_{s=0,i,j} \psi_{A,s}^i(x) d_{g,s} K_\sigma(x - x_g) dx - E\left[\psi_A^i(X)D\right]$$

$$= G^{-1} \sum_g \int \left(\sum_{s=0,i,j} \psi_{A,s}^i(x) d_{g,s} - \mu \right) K_\sigma(x - x_g) dx \equiv \int \left(\sum_{s=0,i,j} \psi_{A,s}^i(x) d_s - \mu \right) d\hat{F}_Z$$

where $d_{g,0} \equiv 1$, $\mu \equiv E\left[\psi_A^i(X)D\right]$, and $\hat{F}_Z$ is a "smoothed-version" of empirical distribution of $z \equiv (x, d_1, d_2)$ with the expectation of $a(z)$ with respect to $\hat{F}_Z$ equal to $G^{-1} \sum_g \int a(x, d_{1,g}, d_{2,g}) K_\sigma(x -$

---

[1] General cases involving discrete coordinates only add to complex notations, but do not require any changes in the the arguments. For example, $\int \psi_A^i(x) d_g K_\sigma(x - x_g) dx$ in the proof would be replaced by $\sum_{\tilde{x}_d} \int \psi_A^i(x_c, \tilde{x}_d) d_g K_\sigma(x_c - x_{g,c}) dx_c 1\{\tilde{x}_{g,d} = \tilde{x}_d\}$. Likewise, $\int K_\sigma(x - x_g) dx$ would be replaced by $\sum_{\tilde{x}_d} \int K_\sigma(x_c - x_{g,c}) dx_c 1\{\tilde{x}_{g,d} = \tilde{x}_d\}$ (which equals 1).

$x_g)dx$. By construction, $\frac{1}{G}\sum_g \int \mu K_\sigma(x - x_g)dx = \mu$. Hence by definition:

$$\int \left(\sum_{s=0,i,j} \psi_{A,s}^i(x)d_s - \mu\right) d\hat{F}_Z - G^{-1}\sum_g \left(\sum_{s=0,i,j} \psi_{A,s}^i(x_g)d_{s,g} - \mu\right)$$

$$= \int \left(\sum_{s=0,i,j} \psi_{A,s}^i(x)d_s\right) d\hat{F}_Z - G^{-1}\sum_g \left(\sum_{s=0,i,j} \psi_{A,s}^i(x_g)d_{s,g}\right)$$

$$= G^{-1}\sum_g \left[\sum_{s=0,i,j} d_{s,g}\left(\int \psi_{A,s}^i(x)K_\sigma(x - x_g)dx - \psi_{A,s}^i(x_g)\right)\right] \tag{4}$$

It remains to show that the r.h.s. of (4) is $o_p(G^{-1/2})$. First, note

$$\left\|\sqrt{G}E\left[\sum_{s=0,i,j} d_{s,g}\left(\int \psi_{A,s}^i(x)K_\sigma(x - x_g)dx - \psi_{A,s}^i(x_g)\right)\right]\right\|$$

$$= \sqrt{G}\left\|\sum_{s=0,i,j}\left(\int\left(\int \psi_{A,s}^i(x)K_\sigma(x - \hat{x})dx\right)\gamma_s^*(\hat{x})d\hat{x} - \int \psi_{A,s}^i(x)\gamma_s^*(x)dx\right)\right\|$$

$$= \sqrt{G}\left\|\sum_{s=0,i,j}\left(\int\int \psi_{A,s}^i(\hat{x} + \sigma u)K(u)\gamma_s^*(\hat{x})dud\hat{x} - \int \psi_{A,s}^i(x)\gamma_s^*(x)dx\right)\right\|$$

$$= \sqrt{G}\left\|\sum_{s=0,i,j}\left(\int \psi_{A,s}^i(x)\left(\int K(u)\gamma_s^*(x - \sigma u)du\right)dx - \int \psi_{A,s}^i(x)\gamma_s^*(x)dx\right)\right\| \tag{5}$$

where the equalities follows from a change-of-variable between x and $u \equiv \frac{x-\hat{x}}{\sigma}$ first, and then another change-of-variable between $\hat{x}$ and $x = \hat{x} + \sigma u$. (Such operations can be done due to the smoothness conditions on $\gamma_{(1)}^*$ and the forms of $\psi_{A,s}^i$ for $s = 0, i, j$.) Next, using $\int K(u)du = 1$ and re-arranging terms, we can write (5) as:

$$\sqrt{G}\left\|\sum_{s=0,i,j}\int \psi_{A,s}^i(x)\left\{\int [\gamma_s^*(x - \sigma u) - \gamma_s^*(x)]K(u)du\right\}dx\right\|$$

$$\leq \sqrt{G}\int \left\|\psi_A^i(x)\right\|\left\|\int [\gamma^*(x - \sigma u) - \gamma^*(x)]K(u)du\right\|dx, \tag{6}$$

where the inequality follows from the Cauchy-Schwarz Inequality. By definition, for $s = 0, i, j$ and any fixed x,

$$E[\hat{\gamma}_s(x)] = E[D_s K_\sigma(x - X_g)] = \int E[D_s|x_g]K_\sigma(x - x_g)\gamma_0^*(x_g)dx_g = \int \gamma_s^*(x - \sigma u)K(u)du$$

where the last equality follows from a change-of-variable between x_g and $u \equiv \frac{x-x_g}{\sigma}$. (Recall $D_0 \equiv 1$.) Under S-(i) and K-(i) (smoothness of γ^* and the kernel functions) and K-(ii) (on the order of the kernel), $\sup_{x\in\omega}\|E[\hat{\gamma}(x) - \gamma^*(x)]\| = O(\sigma^m)$. Under the boundedness condition in D-(i), $\int \sup_{x\in\omega}\|\psi_A^i(x)\|dx < \infty$ and the r.h.s. of the inequality in (6) is bounded above by $\sqrt{G}\sigma^m\tilde{C}$, with $\tilde{C}$ being some finite positive constant. Since $\sqrt{G}\sigma^m \longrightarrow 0$, it follows that

$$E\left\|\left[\sum_{s=0,i,j} d_{g,s}\left(\int \psi_{A,s}^i(x)K_\sigma(x - x_g)dx - \psi_{A,s}^i(x_g)\right)\right]\right\| = o_p(G^{-1/2})$$

By definition of ψ_A^i and the smoothness condition in D-(i), ψ_A^i is continuous in x almost everywhere, and $\psi_{A,s}^i(x + \sigma u) \to \psi_{A,s}^i(x)$ for almost all x and u and $s = 0, i, j$. By condition D-(iii), for σ small

enough, $\psi^i_{A,s}(x + \sigma u)$ is bounded above by some function of x for almost all x and $s = 0, i, j$ on the support of $K(.)$. It then follows from the Dominated Convergence Theorem that

$$\int \psi^i_A(x + \sigma u)K(u)du \rightarrow \int \psi^i_A(x)K(u)du = \psi^i_A(x)$$

for almost all x as $\sigma \rightarrow 0$, where $\psi^i_A = [\psi^i_{A,0}, \psi^i_{A,i}, \psi^i_{A,j}]$. Under the boundedness conditions of the kernel in K and S-(iii), the Dominated Convergence Theorem applies again to give

$$E\left[\left\|\int \psi^i_A(\tilde{x})K_\sigma(\tilde{x} - x)d\tilde{x} - \psi^i_A(x)\right\|^4\right] \rightarrow 0.$$

It then follows from the Cauchy-Schwartz Inequality that

$$E\left[\|D\|^2 \left\|\int \psi^i_A(\tilde{x})K_\sigma(\tilde{x} - x)d\tilde{x} - \psi^i_A(x)\right\|^2\right] \rightarrow 0.$$

Hence both the mean and variance of the product of $\sqrt{G}$ and the left-hand side of (4) converge to 0. It then follows from Chebyshev's Inequality that this product converges in probability to 0. Q.E.D.

It follows from Lemma 1.1 and Lemma 1.2 that

$$\frac{1}{G}\sum_g w_g m^i_A\left(x_g; \hat{\gamma}_{(1)}\right) = \frac{1}{G}\sum_g \left[w_g m^i_A\left(x_g; \gamma^*_{(1)}\right) + \tilde{\nu}^i_A(z_g)\right] + o_p(G^{-1/2}),$$

which implies

$$\hat{A}^i - A^i = \frac{1}{G}\sum_g \left\{w_g m^i_A\left(x_g; \gamma^*_{(1)}\right) - E\left[w_X m^i_A\left(X; \gamma^*_{(1)}\right)\right] + \tilde{\nu}^i_A(z_g)\right\} + o_p(G^{-1/2}).$$

Note the summand (the expression in $\{...\}$) has zero mean by construction, and has a finite second moment under the boundedness conditions in Assumptions S and D. Hence Lemma 1 in Section 6.1 of Lewbel and Tang (2012) follows from the Central Limit Theorem.

1.2   Limiting distribution of $\sqrt{G}\left(\hat{\Delta}^i - \Delta^i\right)$

We now examine the limiting behavior of $\sqrt{G}\left(\hat{\Delta}^i - \Delta^i\right)$. Rewrite the definition of $\hat{\Delta}^i$ in equation (18) of Lewbel and Tang (2012) as

$$\hat{\Delta}^i \equiv (\hat{\rho})^{-1}\left[\tfrac{1}{G}\sum_g w_g m^i_\Delta\left(x_g; \hat{\gamma}_{(1)}\right)\right] \tag{7}$$

where $\hat{\rho} \equiv \left[G^{-1}\sum_g w_g\right]^{-1}$. Furthermore, let $w_x \equiv 1\{x \in \omega\}$ as before, and define $\rho_0 \equiv E[1\{X \in \omega\}] = \Pr(X \in \omega)$. The first step in Lemma 2.1 below is to describe the impact of replacing $\hat{\rho}$ with the true population moment ρ_0 in (7).

Lemma 2.1 *Suppose Assumptions S, K, B and D hold. Then:*

$$\hat{\Delta}^i = G^{-1} \sum_g \left[\rho_0^{-1} w_g m_\Delta^i \left(x_g; \hat{\boldsymbol{\gamma}}_{(1)} \right) - \mathcal{M}_{i,\rho}^* \left(w_g - \rho_0 \right) \right] + o_p(G^{-1/2}) \tag{8}$$

where

$$\mathcal{M}_{i,\rho}^* \equiv \rho_0^{-2} E \left[w_X m_\Delta^i \left(X; \boldsymbol{\gamma}_{(1)}^* \right) \right]$$

Proof of Lemma 2.1. Apply the Taylor's Expansion to (7) around ρ_0, we get:

$$\hat{\Delta}^i = \rho_0^{-1} \left[\tfrac{1}{G} \sum_g w_g m_\Delta^i \left(x_g; \hat{\boldsymbol{\gamma}}_{(1)} \right) \right] - \hat{\mathcal{M}}_{i,\rho} \left(\tfrac{1}{G} \sum_g 1\{ x_g \in \omega \} - \rho_0 \right)$$

where

$$\hat{\mathcal{M}}_{i,\rho} \equiv \tilde{\rho}^{-2} \left[\tfrac{1}{G} \sum_g w_g m_\Delta^i \left(x_g; \hat{\boldsymbol{\gamma}}_{(1)} \right) \right]$$

and $\tilde{\rho}$ lies on a line segment between $\hat{\rho}$ and ρ_0. By the Law of Large Numbers, $\hat{\rho} \xrightarrow{p} \rho_0$ and therefore $\tilde{\rho} \xrightarrow{p} \rho_0$. Under Assumptions in S, K and B, $\sup_{x \in \omega} \left\| \hat{\boldsymbol{\gamma}}_{(1)} - \boldsymbol{\gamma}_{(1)}^* \right\| = o_p(G^{-1/4})$. By definition of m_Δ^i, it can be shown that $w_X m_\Delta^i \left(X; \boldsymbol{\gamma}_{(1)}^* \right)$ is continuous at $\boldsymbol{\gamma}_{(1)}^*$ with probability one. Lemma 4.3 in Newey and McFadden (1994) then implies that, under Assumption D-(iv), $\tfrac{1}{G} \sum_g w_g m_\Delta^i \left(x_g; \hat{\boldsymbol{\gamma}}_{(1)} \right) \xrightarrow{p} E \left[w_X m_\Delta^i \left(X; \boldsymbol{\gamma}_{(1)}^* \right) \right]$. By the Slutsky Theorem, $\hat{\mathcal{M}}_{i,\rho} = \mathcal{M}_{i,\rho}^* + o_p(1)$. Furthermore, $\tfrac{1}{G} \sum_g 1\{ x_g \in \omega \} - \rho_0 = O_p(G^{-1/2})$ by the Central Limit Theorem. It then follows that

$$\hat{\Delta}^i = \rho_0^{-1} \left[\tfrac{1}{G} \sum_g w_g m_\Delta^i \left(x_g; \hat{\boldsymbol{\gamma}}_{(1)} \right) \right] - \mathcal{M}_{i,\rho}^* \left(\tfrac{1}{G} \sum_g 1\{ x_g \in \omega \} - \rho_0 \right) + o_p(G^{-1/2}).$$

This proves the lemma. *Q.E.D.*

The remaining steps for deriving the limiting distribution of $\sqrt{G}(\hat{\Delta}^i - \Delta^i)$ are similar to those used for $\sqrt{G}(\hat{A}^i - A^i)$. Let $a, b, c, d, e, f, g, h, k$ be the same short-hands as defined above (i.e. short-hands for coordinates in $\boldsymbol{\gamma}_{(1)}^*$ at some x). The dependence of these short-hands on x are suppressed in notations. First, for any $x \in \Omega_X$, define a linear functional of $\boldsymbol{\gamma}_{(1)}$:

$$M_\Delta^i \left(x; \boldsymbol{\gamma}_{(1)} \right) \equiv \left[\boldsymbol{\gamma}_{(1)}(x) \right]' \tilde{D}_{i,\Delta}(x), \text{ where}$$

$$\tilde{D}_{i,\Delta}\left(x; \boldsymbol{\gamma}_{(1)}^* \right)_{9-by-1} \equiv \begin{bmatrix} \frac{-b^2he^2 - a^2g^2h + b^2dke + 2acdg^2 + a^2fgk + bcfe^2 - bcdge - 2acfge + 2abghe - 2abdgk}{(-cfe + bhe + cdg - agh - bdk + afk)^2}, \\ \frac{a(ce-ak)(fe-dg)}{(cfe - bhe - cdg + agh + bdk - afk)^2}, \frac{-a(be-ag)(fe-dg)}{(cfe - bhe - cdg + agh + bdk - afk)^2}, \frac{a(be-ag)(cg-bk)}{(cfe - bhe - cdg + agh + bdk - afk)^2}, \\ \frac{-a(bd-af)(cg-bk)}{(cdg - agh - bdk + afk - cfe + bhe)^2}, \frac{-a(be-ag)(ce-ak)}{(cfe - bhe - cdg + agh + bdk - afk)^2}, \frac{a(ce-ak)(bd-af)}{(cfe - bhe - cdg + agh + bdk - afk)^2}, \\ \frac{a(be-ag)^2}{(cfe - bhe - cdg + agh + bdk - afk)^2}, \frac{-a(be-ag)(bd-af)}{(cfe - bhe - cdg + agh + bdk - afk)^2} \end{bmatrix}'.$$

The detailed forms of $\tilde{D}_{i,\Delta}^{(s)}$ (partial derivatives of $\tilde{D}_{i,\Delta}^{(s)}$ with respect to special regressors X_s) are given in the web appendix to economize space here. The correction terms $\psi_\Delta^i \equiv [\psi_{\Delta,0}^i, \psi_{\Delta,i}^i, \psi_{\Delta,j}^i]$ are defined in the text.

Lemma 2.2 *Suppose Assumptions S, K, B, D and W hold. Then for $i = 1, 2$,*

$$\frac{1}{G} \sum_g w_g m_\Delta^i \left(x_g; \hat{\gamma}_{(1)} \right) = \frac{1}{G} \sum_g \left[w_g m_\Delta^i \left(x_g; \gamma_{(1)}^* \right) + \tilde{\nu}_\Delta(z_g) \right] + o_p(G^{-1/2}) \tag{9}$$

where $\tilde{\nu}_\Delta^i(z_g) \equiv \psi_\Delta^i(x_g) d_g - E \left[\psi_\Delta^i(X) D \right]$.

The proof of Lemma 2.2 uses the same arguments as those for Lemma 1.1 and Lemma 1.2. We provide some additional details for derivations in the web appendix of this article. Combining results in (8) and (9), we have

$$\begin{aligned}
\hat{\Delta}^i - \Delta^i &= G^{-1} \sum_g \left[w_g m_\Delta^i \left(x_g; \gamma_{(1)}^* \right) / \rho_0 + \tilde{\nu}_\Delta^i(z_g) / \rho_0 - \Delta^i - \mathcal{M}_{i,\rho}^* (w_g - \rho_0) \right] + o_p(G^{-1/2}) \\
&= \frac{1}{\rho_0} \frac{1}{G} \sum_g \left[w_g m_\Delta^i \left(x_g; \gamma_{(1)}^* \right) + \tilde{\nu}_\Delta^i(z_g) - \rho_0 \Delta^i - \Delta^i (w_g - \rho_0) \right] + o_p(G^{-1/2}), \tag{10}
\end{aligned}$$

where the summand in the square bracket has zero mean by construction, and have a finite second moment under the boundedness assumptions in S. Since Δ^i is a constant, it follows from the Central Limit Theorem that:

$$\sqrt{G} \left(\hat{\Delta}^i - \Delta^i \right) \xrightarrow{d} \mathcal{N} \left(0 , \ \rho_0^{-2} Var \left[w_X \left(m_\Delta^i \left(X; \gamma_{(1)}^* \right) - \Delta^i \right) + \tilde{\nu}_\Delta^i(Z) \right] \right).$$

1.3 Limiting distribution of $\sqrt{G} \left(\hat{\beta}_i - \beta_i \right)$

Lemma 3.1 *Suppose Assumptions S'-(a),(b), K, B and D'-(i) hold. Then:*

$$\frac{1}{G} \sum_g m_B^i \left(z_g; \hat{\delta}_i, \hat{\gamma}_{(1,i)} \right) = \frac{1}{G} \sum_g \left[m_B^i \left(z_g; \delta_i, \hat{\gamma}_{(1,i)} \right) + \mathcal{M}_{i,\delta}^* \Psi_\delta^i(z_g) \right] + o_p(G^{-1/2})$$

where

$$\begin{aligned}
\mathcal{M}_{i,\delta}^* &\equiv E \left[\tilde{D}_{i,B,\delta} \left(Z; \delta_i, \gamma_{(1,i)}^* \right) \right] \quad and \\
\Psi_\delta^i(z_g) &\equiv -\frac{1}{\rho_0} \left[w_g m_\Delta^i \left(x_g; \gamma_{(1,i)}^* \right) + \tilde{\nu}_\Delta^i(z_g) - \rho_0 \Delta^i - \Delta^i (w_g - \rho_0) \right] \tag{11}
\end{aligned}$$

Proof of Lemma 3.1. By definition and our choice of the distribution function, conditions in S' and K, $m_B^i \left(z; \delta_i, \hat{\gamma}_{(1,i)} \right)$ is continuously differentiable in δ_i. By the Mean Value Theorem,

$$m_B^i \left(z_g; \hat{\delta}_i, \hat{\gamma}_{(1,i)} \right) = m_B^i \left(z_g; \delta_i, \hat{\gamma}_{(1,i)} \right) + \tilde{D}_{i,B,\delta} \left(z_g; \bar{\delta}_i, \hat{\gamma}_{(1,i)} \right) \left(\hat{\delta}_i - \delta_i \right)$$

where $\tilde{D}_{i,B,\delta} \left(z_g; \bar{\delta}_i, \hat{\gamma}_{(1,i)} \right)$ is the partial derivative of m_B^i with respect to the interaction effect at $\bar{\delta}_i$, which is an intermediate value between $\hat{\delta}_i$ and δ_i. Hence

$$\begin{aligned}
&\frac{1}{\sqrt{G}} \sum_g m_B^i \left(z_g; \hat{\delta}_i, \hat{\gamma}_{(1,i)} \right) \\
&= \frac{1}{\sqrt{G}} \sum_g m_B^i \left(z_g; \delta_i, \hat{\gamma}_{(1,i)} \right) + \frac{1}{G} \sum_g \tilde{D}_{i,B,\delta} \left(z_g; \bar{\delta}_i, \hat{\gamma}_{(1,i)} \right) \sqrt{G} \left(\hat{\delta}_i - \delta_i \right). \tag{12}
\end{aligned}$$

With $\alpha_i = -1$, it follows from (10) that $\sqrt{G}(\hat{\delta}_i - \delta_i) = G^{-1/2}\sum_g \Psi^i_\delta(z_g) + o_p(1)$, where the influence function is defined as in (11). (Recall that $\delta_i = \alpha_i \Delta^i$.) Let $\sup_{x \in \Omega_X} \|.\|$ (where $\|.\|$ is the Euclidean norm) be the norm for the space of generic $\gamma_{(1,i)}$ that are continuously differentiable in x_c up to an order of $\bar{m} \geq 0$ given all $\tilde{x}_d$. Under conditions in S' and K, $\tilde{D}_{i,B,\delta}\left(z\,;\delta_i, \gamma_{(1,i)}\right)$ is continuous in $\delta_i, \gamma_{(1,i)}$ with probability 1. Note $\bar{\delta}_i \xrightarrow{p} \delta_i$ as a consequence of $\hat{\delta}_i \xrightarrow{p} \delta_i$. Under conditions in S'-(a),(b), K, and B, $\sup_{x \in \Omega_X} \left\|\hat{\gamma}_{(1,i)} - \gamma^*_{(1,i)}\right\| \xrightarrow{p} 0$ at the rate faster than $G^{1/4}$. It then follows that $\frac{1}{G}\sum_g \tilde{D}_{i,B,\delta}\left(z_g;\bar{\delta}_i, \hat{\gamma}_{(1,i)}\right) \xrightarrow{p} \mathcal{M}^*_{i,\delta}$ under condition D'-(i). Hence the second term on the right-hand side of (12) is $\left[\mathcal{M}^*_{i,\delta} + o_p(1)\right] * \left[G^{-1/2}\sum_g \Psi^i_\delta(z_g) + o_p(1)\right]$, where $G^{-1/2}\sum_g \Psi^i_\delta(z_g)$ is $O_p(1)$. It then follows that

$$G^{-\frac{1}{2}}\sum_g m^i_B\left(z_g;\hat{\delta}_i, \hat{\gamma}_{(1,i)}\right) = G^{-\frac{1}{2}}\sum_g \left[m^i_B\left(z_g;\delta_i, \hat{\gamma}_{(1,i)}\right) + \mathcal{M}^*_{i,\delta}\Psi^i_\delta(z_g)\right] + o_p(1).$$

This proves Lemma 3.1.   Q.E.D.

Lemma 3.2 *Under Assumptions S', K, B and D',*

$$\frac{1}{G}\sum_g \left[m^i_B\left(z_g;\delta_i, \hat{\gamma}_{(1,i)}\right) - m^i_B\left(z_g;\delta_i, \gamma^*_{(1,i)}\right)\right]$$
$$= \int M^i_{B,\gamma}\left(z, \hat{\gamma}_{(1,i)} - \gamma^*_{(1,i)}\right) + M^i_{B,f}\left(z, \hat{\gamma}_0 - \gamma^*_0\right) dF^*_Z + o_p(G^{-1/2}).$$

Proof of Lemma 3.2. Under conditions S'-(a), (b), we have

$$\left\|\begin{array}{c} m^i_B\left(z_g;\delta_i, \hat{\gamma}_{(1,i)}\right) - m^i_B\left(z_g;\delta_i, \gamma^*_{(1,i)}\right) \\ -M^i_{B,\gamma}\left(z_g, \hat{\gamma}_{(1,i)} - \gamma^*_{(1,i)}\right) - M^i_{B,f}\left(z_g, \hat{\gamma}_0 - \gamma^*_0\right)\end{array}\right\|$$
$$\leq b_{\gamma,f}(z_g)\sup_{x \in \Omega_X}\|\hat{\gamma}_{(1,i)} - \gamma^*_{(1,i)}\|^2, \tag{13}$$

for some non-negative function $b_{\gamma,f}$ (which may depend on true parameters in the DGP δ_i and $\gamma^*_{(1)}$) and $E[b_{\gamma,f}(Z)] < +\infty$. Under Assumptions S'-(a),(b), K and B, $\sup_{x \in \Omega_X}\|\hat{\gamma}_{(1,i)} - \gamma^*_{(1,i)}\|$ is $o_p(G^{-1/4})$. Thus the right-hand side of the inequality in (13) is $o_p(G^{-1/2})$. It follows from the triangular inequality and the Law of Large Numbers that:

$$\frac{1}{G}\sum_g \left[\begin{array}{c} m^i_B\left(z_g;\delta_i, \hat{\gamma}_{(1,i)}\right) - m^i_B\left(z_g;\delta_i, \gamma^*_{(1,i)}\right) \\ -M^i_{B,\gamma}\left(z_g, \hat{\gamma}_{(1,i)} - \gamma^*_{(1,i)}\right) - M^i_{B,f}\left(z_g, \hat{\gamma}_0 - \gamma^*_0\right)\end{array}\right] = o_p(G^{-\frac{1}{2}})$$

By construction,

$$\frac{1}{G}\sum_g M^i_{B,\gamma}\left(z_g, \hat{\gamma}_{(1,i)} - \gamma^*_{(1,i)}\right) - \int M^i_{B,\gamma}\left(z;\hat{\gamma}_{(1,i)} - \gamma^*_{(1,i)}\right) dF^*_Z$$
$$= \frac{1}{G}\sum_g M^i_{B,\gamma}\left(z_g, \hat{\gamma}_{(1,i)} - \bar{\gamma}_{(1,i)}\right) - \int M^i_{B,\gamma}\left(z;\hat{\gamma}_{(1,i)} - \bar{\gamma}_{(1,i)}\right) dF^*_Z +$$
$$\frac{1}{G}\sum_g M^i_{B,\gamma}\left(z_g, \bar{\gamma}_{(1,i)} - \gamma^*_{(1,i)}\right) - \int M^i_{B,\gamma}\left(z;\bar{\gamma}_{(1,i)} - \gamma^*_{(1,i)}\right) dF^*_Z. \tag{14}$$

where $\bar{\gamma}_{(1,i)}(x) \equiv E[\hat{\gamma}_{(1,i)}(x)]$ is a function of x (i.e. the expectation is taken with respect to the random sample used for estimating $\hat{\gamma}_{(1,i)}$ at x). For any $s, g \leq G$, define $a_{B,G}(z_g, z_s) \equiv$

$$
\left[\begin{array}{l} K_\sigma(x_s - x_g), d_{s,j}K_\sigma(x_s - x_g), .. \\ K_{\sigma,i}(x_s - x_g), d_{s,j}K_{\sigma,i}(x_s - x_g) \end{array} \right] \tilde{D}_{i,B}(z_g).
$$

Let $a_{B,G}^2(z_s) \equiv \int a_{B,G}(z, z_s)dF_Z^*(z)$, and

$$
\begin{aligned}
a_{B,G}^1(z_g) &\equiv \int a_{B,G}(z_g, \hat{z})dF_Z^*(\hat{z}) = \int \left[\begin{array}{l} K_\sigma(\hat{x} - x_g), \hat{d}_j K_\sigma(\hat{x} - x_g), .. \\ K_{\sigma,i}(\hat{x} - x_g), \hat{d}_j K_{\sigma,i}(\hat{x} - x_g) \end{array} \right] \tilde{D}_{i,B}(z_g)dF_Z^*(\hat{z}) \\
&= \bar{\gamma}_{(1,i)}(x_g)\tilde{D}_{i,B}(z_g).
\end{aligned}
$$

Then the first difference on the r.h.s. of (14) can be written as:

$$
G^{-2}\sum_g \sum_s a_{B,G}(z_g, z_s) - G^{-1}\sum_g a_{B,G}^1(z_g) - G^{-1}\sum_g a_{B,G}^2(z_g) + E[a_{B,G}^1(Z)].
$$

Then by condition D'-(ii) and the boundedness of kernel function in Assumption K, $E\left[\|a_{B,G}(Z,Z)\|\right]$ $\leq \sigma^{-J_c}C_1'$ and $\left\{ E\left[\|a_{B,G}(Z,Z')\|^2\right] \right\}^{1/2} \leq \sigma^{-J_c}C_2'$, where C_1', C_2' are finite positive constants (whose form depend on the true parameters in $\gamma_{(1,i)}^*$ and the kernel function). Since $\sqrt{G}\sigma^{J_c} \longrightarrow +\infty$ under Assumption B, it follows from Lemma 8.4 in Newey and McFadden (1994) that the first difference is $o_p(G^{-1/2})$. The boundedness of derivatives of m_B^i w.r.t. $\gamma_{(1,i)}$ in D'-(ii) implies $E\left[\left\|M_{B,\gamma}^i\left(Z; \bar{\gamma}_{(1,i)} - \gamma_{(1,i)}^*\right)\right\|^2\right] \leq C_3' \sup_{x \in \Omega_X}\left\|\bar{\gamma}_{(1,i)} - \gamma_{(1,i)}^*\right\|$. Under Assumptions K and S'-(a), $\sup_{x \in \Omega_X}\left\|\bar{\gamma}_{(1,i)} - \gamma_{(1,i)}^*\right\| = o(\sigma^m)$, which converges to 0 as $\sigma \longrightarrow 0$ under Assumption B. It then follows from the Chebychev's Inequality that the second difference in (14) is $o_p(G^{-1/2})$. We can also decompose the linear functional in the second term in a similar manner and have:

$$
\begin{aligned}
&\frac{1}{G}\sum_g M_{B,f}^i(z_g, \hat{\gamma}_0 - \gamma_0^*) - \int M_{B,f}^i(z; \hat{\gamma}_0 - \gamma_0^*)dF_Z^* \\
=\ &\frac{1}{G}\sum_g M_{B,f}^i(z_g, \hat{\gamma}_0 - \bar{\gamma}_0) - \int M_{B,f}^i(z; \hat{\gamma}_0 - \bar{\gamma}_0)dF_Z^* + \\
&\frac{1}{G}\sum_g M_{B,f}^i(z_g, \bar{\gamma}_0 - \gamma_0^*) - \int M_{B,f}^i(z; \bar{\gamma}_0 - \gamma_0^*)dF_Z^*.
\end{aligned}
\tag{15}
$$

Arguments similar to those above suggest both differences in (15) are $o_p(G^{-1/2})$ under conditions in Assumption S', K, B and D'. *Q.E.D.*

Lemma 3.3 *Under Assumptions S', K, B, D' and W', for $i = 1, 2$ and $j = 3 - i$,*

$$
\int M_{B,\gamma}^i(z, \hat{\gamma}_{(1,i)} - \gamma_{(1,i)}^*) + M_{B,f}^i(z, \hat{\gamma}_0 - \gamma_0^*)dF_Z^* = \frac{1}{G}\sum_g \tilde{\nu}_B^i(z_g) + o_p(G^{-1/2})
\tag{16}
$$

where

$$
\tilde{\nu}_B^i(z_g) \equiv \psi_{B,0}^i(x_g) + \psi_{B,f}^i(x_g) + \psi_{B,j}^i(x_g)d_{g,j} - E\left[\psi_{B,0}^i(X) + \psi_{B,f}^i(X) + \psi_{B,j}^i(X)D_j\right].
$$

Proof of Lemma 3.3. By smoothness of the true population function $\gamma^*_{(1,i)}$ in DGP in S'-(a), and that of the kernel function in S', K, D', and using the convexity of $\Omega_{X_i|x_{-i}}$ in W', we can apply integration-by-parts to write the left-hand side of the equation (16) as:

$$\int \begin{bmatrix} \hat{\gamma}_0(x) - \gamma^*_0(x) \\ \hat{\gamma}_j(x) - \gamma^*_j(x) \end{bmatrix}' \begin{bmatrix} \psi^i_{B,0}(x) \\ \psi^i_{B,j}(x) \end{bmatrix} + \psi^i_{B,f}(x) \left[\hat{\gamma}_0(x) - \gamma^*_0(x) \right] dx, \tag{17}$$

where $\psi^i_{B,0}$, $\psi^i_{B,j}$ and $\psi^j_{B,f}$ are defined as in the text. (See the web appendix of this article for detailed derivations.) By arguments similar to those in the proof of Lemma 1.2, the expression in (17) is $\frac{1}{G} \sum_g \nu^i_B(z_g) + o_p(G^{-\frac{1}{2}})$ due to conditions in Assumptions S', K, B and D' and that $\psi^i_{B,f}(x)$ is almost everywhere continuous in x. Q.E.D.

By Lemma 3.1, Lemma 3.2 and Lemma 3.3, we have

$$\hat{\mu}_{i,\Psi} \equiv \frac{1}{G} \sum_g m^i_B \left(z_g; \hat{\delta}_i, \hat{\gamma}_{(1,i)} \right) = \frac{1}{G} \sum_g \Psi^i_B(z_g) + o_p(G^{-\frac{1}{2}}),$$

where

$$\Psi^i_B(z) \equiv m^i_B \left(z; \delta_i, \gamma^*_{(1,i)} \right) + \tilde{\nu}^i_B(z) + \mathcal{M}^*_{i,\delta} \Psi^i_\delta(z).$$

By construction

$$E \left[\Psi^i_B(Z) \right] = E \left[m^i_B \left(Z; \delta_i, \gamma^*_{(1,i)} \right) \right] \equiv \mu_{i,\Psi}.$$

Let $\hat{\lambda} \equiv \frac{1}{G} \sum_g \tilde{x}'_g \tilde{x}_g$ and $\lambda \equiv E \left(\tilde{X}' \tilde{X} \right)$. By definition, $\beta_i = \lambda^{-1} \mu_{i,\Psi}$ and

$$\hat{\beta}_i = \hat{\lambda}^{-1} \left(\frac{1}{G} \sum_g \Psi^i_B(z_g) + o_p(G^{-1/2}) \right),$$

where $\hat{\lambda}^{-1} = \lambda^{-1} + o_p(1)$. Note

$$\hat{\beta}_i = \hat{\lambda}^{-1} \left(\hat{\mu}_{i,\Psi} - \hat{\lambda} \beta_i + \hat{\lambda} \beta_i \right) = \hat{\lambda}^{-1} \left(\hat{\mu}_{i,\Psi} - \hat{\lambda} \beta_i \right) + \beta_i$$

$$\Rightarrow \quad \sqrt{G} \left(\hat{\beta}_i - \beta_i \right) = \hat{\lambda}^{-1} \sqrt{G} \left(\hat{\mu}_{i,\Psi} - \hat{\lambda} \beta_i \right).$$

where

$$\sqrt{G} \left(\hat{\mu}_{i,\Psi} - \hat{\lambda} \beta_i \right) = \frac{1}{\sqrt{G}} \sum_g \left(\Psi^i_B(z_g) - \tilde{x}'_g \tilde{x}_g \beta_i \right) + o_p(1).$$

By construction, $E \left[\Psi^i_B(Z) - \tilde{X}' \tilde{X} \beta_i \right] = 0$. Under conditions in S',D' and R', $E \left[\left\| \Psi^i_B(Z) - \tilde{X}' \tilde{X} \beta_i \right\|^2 \right] < +\infty$. Then the Central Limit Theorem and the Slutsky Theorem can be applied to show that $\sqrt{G}(\hat{\beta}_i - \beta_i) \xrightarrow{d} \mathcal{N}(0, \Sigma^i_B)$ with

$$\Sigma^i_B \equiv \lambda^{-1} Var(\Psi^i_B(Z) - \tilde{X}' \tilde{X} \beta_i) \left(\lambda^{-1} \right)'.$$

This completes the proof of the limiting distribution of $\sqrt{G}(\hat{\beta}_i - \beta_i)$.

2   Limiting distribution of $\sqrt{G}\left(\hat{T}(\bar{\omega}) - \mu_{\bar{\omega}}\right)$

Define

$$v(x) \equiv \frac{1\{x \in \bar{\omega}\}}{\gamma_0^*(x)} \left[2p_1^*(x)p_2^*(x), \; -p_2^*(x), \; -p_1^*(x)\right] \tag{18}$$

For notation simplicity, we suppress dependence of $\hat{T}$ on $\bar{\omega}$ when there is no confusion. We now collect the appropriate conditions for establishing the asymptotic properties of $\hat{T}$.

Assumption T1 *(i) The true density $\gamma_0^*(x)$ is bounded away from zero and $\gamma_k^*(x)$ are bounded above over $\bar{\omega}$ for $k = 1, 2$.   (ii) For any $\tilde{x}$ and $i = 1, 2$, $\gamma_0^*(x_e|\tilde{x})$ and $p_i^*(x)\gamma_0^*(x_e|\tilde{x})$ are continuously differentiable of order $m \geq 2$, and the derivatives of order $m \geq 2$ are uniformly bounded over $\bar{\omega}$. (iii) there exists $\eta > 0$ such that $[VAR_{u_i}(x)]^{1+\eta}\gamma_0^*(x_e|\tilde{x})$ (where $VAR_{u_i}(x)$ denotes variance of $u_i = d_i - p_i^*(x)$ conditional on x) is uniformly bounded; and (iv) both $p_i^*(x)^2\gamma_0^*(x_e|\tilde{x})$ and $VAR_{u_i}(x)\gamma_0^*(x_e|\tilde{x})$ are continuous and bounded over the support of X_e given $\tilde{x}$.*

Assumption T2 *$K(.)$ is a product kernel function with support in $\mathbb{R}^2$ such that (i) $\int |K(t_1, t_2)|dt_1dt_2 < \infty$ and $\int K(t_1, t_2)dt_1dt_2 = 1$; and (ii) $K(.)$ has zero moments up to the order of m, and $\int ||u||^m |K(t_1, t_2)| \, dt_1dt_2 < \infty$.*

Assumption T3 *$\sqrt{G}\sigma^m \to 0$ and $\left(\frac{\sqrt{G}}{\ln G}\right)\sigma^2 \to +\infty$ as $G \to +\infty$.*

Assumption T4 *(i) Given any $\tilde{x}$, v is continuous almost everywhere w.r.t. Lebesgue measure on the support of (X_1, X_2). (ii) There exists $\varepsilon > 0$ such that $E[\sup_{||\eta|| \leq \varepsilon} ||v(x + \eta)||^4] < \infty$.*

Conditions in *T1* ensure v is bounded over any bounded subset of Ω_X. Conditions in *T2, T3* together guarantee that the kernel estimators $\hat{\gamma}$ are well-behaved in the sense $\sup_{x \in \bar{\omega}} ||\hat{\gamma} - \gamma^*|| = o_p(G^{-1/4})$ and $\sup_{x \in \bar{\omega}} ||E(\hat{\gamma}) - \gamma^*|| \longrightarrow 0$ as $G \to \infty$.

Lemma 4.1 *Suppose assumptions T1,2,3 hold. Then*

$$\hat{T}(\bar{\omega}) - \frac{1}{G}\sum_g 1\{x_g \in \bar{\omega}\}m(z_g; \gamma^*) = \int M(z, \hat{\gamma} - \gamma^*)dF_Z^* + o_p(G^{-1/2}) \tag{19}$$

where $M(z, \gamma) \equiv v(x)\gamma(x)$ with $v(.)$ defined in (18) above.

Proof of Lemma 4.1.   We will drop the dependence of $\hat{T}$ and T on ω for notational simplicity. It follows from linearization arguments and *T1-(i)* that there exists a constant $C < \infty$ so that for any $\gamma \equiv [\gamma_0, \gamma_1, \gamma_2]'$ with $\sup_{x \in \bar{\omega}} ||\gamma - \gamma^*||$ small enough,

$$1\{x \in \bar{\omega}\}|m(z; \gamma) - m(z; \gamma^*) - M(z, \gamma - \gamma^*)| \leq C \sup_{x \in \bar{\omega}} ||\gamma - \gamma^*||^2 \tag{20}$$

for all z. Given *T1,2,3*, we have $\sup_{x \in \bar{\omega}} ||\hat{\gamma} - \gamma^*|| = O_p(\sigma^{-m/(2m+4)})$ by Theorem 3.2.1. in Bierens (1987). For $m \geq 2$ (the parameter m captures smoothness properties of population moments in

T1,T2),

$$\tfrac{1}{G} \sum_g 1\{x_g \in \bar{\omega}\}[m(z_g; \hat{\gamma}) - m(z_g; \gamma^*)] - M(z_g, \hat{\gamma} - \gamma^*) = o_p(G^{-1/2}) \tag{21}$$

Next, define $\bar{\gamma} \equiv E(\hat{\gamma})$ and by the linearity of $M(z, \gamma)$ in γ, we can write:

$$G^{-1} \sum_g M(z_g, \hat{\gamma} - \gamma^*) - \int M(z, \hat{\gamma} - \gamma^*) dF^*(z)$$

$$= \; G^{-1} \sum_g M(z_g, \hat{\gamma} - \bar{\gamma}) - \int M(z, \hat{\gamma} - \bar{\gamma}) dF^*(z)$$

$$+ G^{-1} \sum_g M(z_g, \bar{\gamma} - \gamma^*) - \int M(z, \bar{\gamma} - \gamma^*) dF^*(z) \tag{22}$$

Under *T1*, there exists a function c such that $\|M(z, \gamma)\| \le c(z) \sup_{x \in \bar{\omega}} \|\gamma\|$ with $E[c(z)^2] < \infty$. (This is because w is an indicator function, $\gamma_i^*/\gamma_0^* = p_i^*$ are bounded above by 1 for $i = 1, 2$ and $1/\gamma_0^*$ is bounded above over the relevant support in *T1*.) By the Theorem of Projection for V-statistics (Lemma 8.4 of Newey and McFadden (1994)) and that $G\sigma^2 \to \infty$ in *T3*, the first difference in (22) is $o_p(G^{-1/2})$. Next, note by the Chebyshev's inequality,

$$\Pr\left\{ \sqrt{G} \left| G^{-1} \sum_g M(z_g, \bar{\gamma} - \gamma^*) - \int M(z, \bar{\gamma} - \gamma^*) dF^*(z) \right| \ge \varepsilon \right\} \le \frac{E[\|M(z_g, \bar{\gamma} - \gamma^*)\|^2]}{\varepsilon^2}$$

But by construction, $E[\|M(z; \bar{\gamma} - \gamma^*)\|^2] \le c(z) \sup_{x \in \bar{\omega}} \|\bar{\gamma} - \gamma^*\|^2$. By smoothness conditions in *T1-(ii)* and properties of K in *T2*, $\sup_{x \in \bar{\omega}} \|\bar{\gamma} - \gamma^*\| = O(\sigma^m)$. Hence the second difference in (22) is also $o_p(G^{-1/2})$. Therefore,

$$G^{-1} \sum_g M(z_g, \hat{\gamma} - \gamma^*) - \int M(z, \hat{\gamma} - \gamma^*) dF^*(z) = o_p(G^{-1/2}) \tag{23}$$

This and (21) together proves (19) in the lemma. *Q.E.D.*

Lemma 4.2 *Suppose T1,2,3,4 hold. Then $\int M(z, \hat{\gamma} - \gamma^*) dF_Z^* = \tfrac{1}{G} \sum_g \delta(z_g) + o_p(G^{-1/2})$, where $\delta(z) \equiv v(x)y - E[v(X)Y]$.*

Proof of Lemma 4.2. By construction,

$$\int M(z, \hat{\gamma} - \gamma^*) dF_Z^* = \int v(x)[\hat{\gamma}(x) - \gamma^*(x)] dx = \int v(x)\hat{\gamma}(x) dx - \int v(x)\gamma^*(x) dx$$

$$= \; G^{-1} \sum_g \int v(x) y_g K_\sigma (x - x_g) \, dx - \int v(x)[1, p_1^*(x), p_2^*(x)]' \gamma_0^*(x) dx$$

$$= \; G^{-1} \sum_g \int v(x) y_g K_\sigma (x - x_g) \, dx - E[v(X)Y] = G^{-1} \sum_g \int \delta(x, d_{g,1}, d_{g,2}) K_\sigma (x - x_g) dx \tag{24}$$

where the last equality uses $\int K_\sigma(x - x_g) dx = 1$ for all x_g (which follows from change-of-variables). Define a probability measure $\hat{F}$ on the support of X such that

$$\int a(x, d_1, d_2) d\hat{F} \equiv G^{-1} \sum_g \int a(x, d_{g,1}, d_{g,2}) K_\sigma (x - x_g) dx$$

for any function a over the support of X and $\{0,1\}^2$. Thus (24) can be written as $\int M(z, \hat{\gamma} - \gamma^*)dF^*$ $= \int \delta(z)d\hat{F}$.

Let $K_\sigma^*(x - x_g) \equiv \sigma^{-2}\Pi_{i=1,2}K_i\left(\frac{x_i - x_{g,i}}{\sigma}\right)1(\tilde{x} = \tilde{x}_g)$, where $K_i(.)$ denote univariate kernels corresponding to coordinates x_i for $i = 1, 2$. Thus:

$$\int \delta(z)d\hat{F} - G^{-1}\sum_g \delta(z_g)$$

$$= G^{-1} \quad \sum_g \left(\int v(x)y_g K_\sigma(x - x_g) - v(x)y_g K_\sigma^*(x - x_g)dx\right)$$

$$+G^{-1} \quad \sum_g \left(\int v(x)y_g K_\sigma^*(x - x_g)dx - v(x_g)y_g\right) \tag{25}$$

By the definition of $v(x)$ and properties of K and σ in *T1-4*, the first term on the right-hand side of (25) is $o_p(G^{-1/2})$. The second term is also $o_p(G^{-1/2})$ through application of the Dominated Convergence Theorem, using the properties of v in *T4* and its boundedness over any bounded subsets of Ω_X. (See Theorem 8.11 in Newey and McFadden (1994) for details.)   *Q.E.D.*

Proof of Theorem 6 in Lewbel and Tang (2012). By *Lemma 4.1* and *4.2*,

$$\hat{T}(\bar{\omega}) = \tfrac{1}{G}\sum_g \left[1\{x_g \in \bar{\omega}\}m(z_g; \gamma^*) + \delta(z_g)\right] + o_p(G^{-\frac{1}{2}})$$

By the Central Limit Theorem, $\sqrt{G}\left(\hat{T}(\bar{\omega}) - \mu_{\bar{\omega}}\right) \xrightarrow{d} N(0, \Sigma_{\bar{\omega}})$, where

$$\Sigma_{\bar{\omega}} \equiv Var\left[1\{X \in \bar{\omega}\}m(Z; \gamma^*) + \delta(Z)\right]$$
$$= Var\left[1\{X \in \bar{\omega}\}\left(D_1 D_2 - p_1^*(X)p_2^*(X) + \tfrac{2p_1^*(X)p_2^*(X) - D_1 p_2^*(X) - D_2 p_1^*(X)}{\gamma_0^*(X)}\right)\right]$$

This proves the theorem.   *Q.E.D.*

References

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

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

[3] Lewbel, A. and X. Tang, "Identification and Estimation of Games with Incomplete Information Using Excluded Regressors," unpublished manuscript, 2012

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