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

Arthur Lewbel, Boston College

Xun Tang, University of Pennsylvania

Original: November 2010. Current version: November 2014

This online supplement to Lewbel and Tang (2014) provides proofs and additional results. Section A provides results regarding existence of required sets of states and assumed support conditions. Section B provides further discussion and formal derivations of our asymptotic results, and Section C gives the proof of Theorem 3.

A. N.d.s. Sets and Large Support Conditions

This section provides further primitive conditions on model elements that are sufficient for some of the identifying conditions in the text. These include the existence of a non-degenerate and non-singular (n.d.s.) set for a given $\tilde{x}$; and the large support condition in A4 and A4'.

A.1. Sufficient conditions for n.d.s. sets

In Section 3.2 of Lewbel and Tang (2014), we give conditions under which there exists a n.d.s. set for a generic vector of non-excluded regressors $\tilde{x}$. Here in this section we give stronger conditions that ensure there exists a n.d.s. set ω where the equilibrium is unique. This is useful for the identification strategy discussed in Section 7.1 of Lewbel and Tang (2014).

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 $j = 3-i$. To begin with, we give conditions under which the game admits a unique equilibrium at a generic state x. Let $p \equiv (p_1, p_2)$ and define:

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

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:

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

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

Let ω^* be the set of all states on the support of X where the data is generated by a single Bayesian Nash Equilibrium (BNE).

Proposition A1 *Suppose A1 and A3 hold with $N = 2$. (a) If $\delta_1(\tilde{x})\delta_2(\tilde{x}) < 0$ and $\Pr\{X_e \in \omega^a(\tilde{x}) \mid \tilde{x}\} > 0$, then $\omega^a(\tilde{x}) \otimes \tilde{x} \subseteq \omega^*$ and is n.d.s.. (b) If $\delta_1(\tilde{x})\delta_2(\tilde{x}) > 0$ and $\Pr\left\{X_e \in \omega^a(\tilde{x}) \cap \omega^b(\tilde{x}) \mid \tilde{x}\right\} > 0$, then $\left(\omega^a(\tilde{x}) \cap \omega^b(\tilde{x})\right) \otimes \tilde{x} \subseteq \omega^*$ and is n.d.s..*

Proof of Proposition A1. 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 A3. 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) in Lewbel and Tang (2014), 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

$$\frac{\partial p_1^*(x)}{\partial X_1}\frac{\partial p_2^*(x)}{\partial X_2} - \frac{\partial p_1^*(x)}{\partial X_2}\frac{\partial p_2^*(x)}{\partial X_1} = \frac{\alpha_1}{\delta_1(\tilde{x})}\frac{\partial p_1^*(x)}{\partial X_2} \tag{1}$$

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 definition of a n.d.s. state must hold over $\left(\omega^a(\tilde{x}) \cap \omega^b(\tilde{x})\right) \otimes \tilde{x}$ under A1 and A3. 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 a solution to (3) in Lewbel and Tang (2014) 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.*

A.2. Sufficient conditions for large support assumptions

There are three sets of large support conditions in Lewbel and Tang (2014). These include A4 in Section 3.3; condition (i) in A4' in Section 3.4; and the condition (i) in the third paragraph of Section 7.1. Among these three versions, the last one is the most restrictive because it requires the support of V_i to be large enough conditioning on $X_i \in \omega_i$ and $X_{-i} = x_{-i}$.

We first provide sufficient conditions for the most restrictive version - - condition (i) in Section 7.1 of Lewbel and Tang (2014). Once this is done, we will then proceed to give sufficient conditions A4 and A4'-(i).

Again, 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 A2 *Let $N = 2$ and $h_i(D_j) = D_j$ for both $i = 1, 2$ and $j = 3 - i$, and let A1 and A3 hold. Suppose for i and $x_{-i} = (x_{e,-i}, \tilde{x})$, there exists an interval ω_i on the support of X_i given 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 $\Pr(X_i \in \omega_i \mid x_{-i}) > 0$; and (ii) support of $\mathcal{S}_i$ given $\tilde{x}$ is a subset of the support of V_i given x_{-i}.*

Proof of Proposition A2. Because $\omega_i \otimes x_{-i} \in \omega^*$, A1 and A3 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 condition stated in this proposition, $\Pr\{\alpha_i X_i + \delta_i^-(\tilde{X}) > s_{ih} \mid x_{-i}, X_i \in \omega_i\} > 0$. This implies $\Pr\{V_i > s_{ih} \mid x_{-i}, X_i \in \omega_i\} > 0$. Symmetrically, $\Pr\{\alpha_i X_i + \delta_i^+(\tilde{X}) < s_{il} \mid x_{-i}, X_i \in \omega_i\} > 0$ implies $\Pr\{V_i < s_{il} \mid x_{-i}, X_i \in \omega_i\} > 0$. 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.*

This means condition (i) in Section 7.1 of Lewbel and Tang (2014) holds if for $\tilde{x}$ there exists $x_{-i} = (x_{e,-i}, \tilde{x})$ such that the conditions of this proposition holds. It remains to give sufficient conditions for A4 and A4'-(i) under A2 (the assumption there is a single equilibrium at all states in the data). Under A2, ω^* is identical to the full unconditional support of X. Therefore, A4'-(i) holds when, for all x_{-i}, the conditions in Proposition A2 are satisfied for ω_i equal to the full support of X_i given $X_{-i} = x_{-i}$. If A4'-(i) is satisfied, then A4 is also implied.

A.3. Monotonicity of V_i in X_i

We conclude this section with sufficient conditions for the monotonicity of V_i in excluded regressors. Let $N = 2$, $h(D_{-i}) = D_j$ and $\alpha_i = -1$. For all $x = (x_e, \tilde{x})$, 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:

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

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 $\partial V_i(X)/\partial X_i|_{X=x}$ 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 to hold over the full support of X, it suffices to have two positive constants c_1, c_2 such that $c_1c_2 < 1$ 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 the joint support $(\epsilon_i, \epsilon_j, \tilde{X})$.

B. Derivation of Asymptotic Properties

This section provides further details in the derivation of the limiting distribution of our estimators in Section 5 and that of the test statistic for detecting multiple equilibria in Section 7.2 in Lewbel and Tang (2014). The current section follows the same set of notation as in Lewbel and Tang (2014), and is structured as follows: Sections B.1 and B.2 present the conditions and leading results in the proof of the asymptotic properties of $\hat{\alpha}_i, \hat{\delta}_i$ and $\hat{\beta}_i$ respectively; Section B.3 provides further details in the proofs of these results from Sections B.1 and B.2; Section B.4 provides details in the derivation of the limiting distribution of the test statistic in Section 7.2 of Lewbel and Tang (2014).

We first copy the assumptions from the main text of Lewbel and Tang (2014) here, and provide some additional discussion of these assumptions.

Assumption S. *(i) For $k \in \{0, 1, 2\}$, γ_k^* is continuously differentiable in x_c up to an order of $\bar{m} \geq 2$ given any $\tilde{x}_d$, and the derivatives are continuous and uniformly bounded over an open set that covers ω. (ii) All components in $\boldsymbol{\gamma}^*$ are bounded away from 0 over ω. (iii) Let $VAR_{u_i}(x)$ be the variance of $u_i = d_i - p_i^*(x)$ given x. There exists $\xi > 0$ such that $[VAR_{u_i}(x)]^{1+\xi} f^*(x_c \mid \tilde{x}_d)$ is uniformly bounded over ω. For any $\tilde{x}_d$, both $p_i^*(x)^2 f^*(x_c \mid \tilde{x}_d)$ and $VAR_{u_i}(x) f^*(x_c \mid \tilde{x}_d)$ are continuous in x_c and uniformly bounded over ω.*

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

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

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

A sequence of bandwidths σ that satisfies Assumption B exists, as long as m is sufficiently large relative to J_c. Under Assumptions S, K and B, $\sup_{x\in\omega}||\hat{\boldsymbol{\gamma}}(x)-\boldsymbol{\gamma}^*(x)||=o_p(G^{-1/4})$. This ensures the remainder term from the linearization of $\frac{1}{G}\sum_g w_g m_A^i(x_g;\hat{\boldsymbol{\gamma}})$ around $\boldsymbol{\gamma}^*$ diminishes at a rate faster than $1/\sqrt{G}$. The convexity of ω is needed to derive the variances of the limiting distributions of $\sqrt{G}\left(\hat{A}^i-A^i\right)$ and $\sqrt{G}(\hat{\Delta}^i-\Delta^i)$, which contains a correction term due to the estimation errors in $\hat{\boldsymbol{\gamma}}$.

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

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 the open set ω^o with $\omega\subseteq\omega^o$. (ii) The second order derivatives of $m_A^i(x;\boldsymbol{\gamma})$ and $m_\Delta^i(x;\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}(x)$ evaluated at $\boldsymbol{\gamma}^*(x)$ are bounded for all $x\in\omega^o$.*

Assumption D and an application of the V-statistic Projection Theorem (Lemma 8.4 in Newey and McFadden (1994)) show:

$$\tfrac{1}{G}\textstyle\sum_g w_g\left[m_A^i(x_g;\hat{\boldsymbol{\gamma}})-m_A^i(x_g;\boldsymbol{\gamma}^*)\right]=\displaystyle\int w(x)(\hat{\boldsymbol{\gamma}}(x)-\boldsymbol{\gamma}^*(x))'\tilde{D}_{i,A}(x)dF_X^*(x)+o_p(G^{-1/2})$$

where F_X^* is the true distribution of X in the DGP. Assumptions S,W and D together imply that, for any γ twice continuously differentiable in X_e, there exists $\psi_A^i\equiv(\psi_{A,0}^i,\psi_{A,1}^i,\psi_{A,2}^i)$ such that:

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

Similarly, there exists $\psi_\Delta^i\equiv(\psi_{\Delta,0}^i,\psi_{\Delta,1}^i,\psi_{\Delta,2}^i)$ such that an equation similar to (2) holds when $\tilde{D}_{i,A}$ is replaced by $\tilde{D}_{i,\Delta}$. We present the closed form of ψ_A^i,ψ_Δ^i below.Let $D\equiv[1,D_1,D_2]'$ and let $d\equiv[1,d_1,d_2]'$ denote a realized value of D. Then (2) implies

$$\int w(x)(\hat{\boldsymbol{\gamma}}(x)-\boldsymbol{\gamma}^*(x))'\tilde{D}_{i,A}(x)dF_X^*(x)=\int\tilde{\nu}_A^i(z)d\tilde{F}_Z \tag{3}$$

where $\tilde{\nu}_A^i(z)\equiv\psi_A^i(x)d-E\left[\psi_A^i(X)D\right]$ and $\tilde{F}_Z$ is some smoothed version of the empirical distribution of $Z\equiv(X,D_i,D_j)$. Similarly, define $\tilde{\nu}_\Delta^i(z)\equiv\psi_\Delta^i(x)d-E\left[\psi_\Delta^i(X)D\right]$. Again, an equation similar to (3) holds with $\tilde{D}_{i,A},\tilde{\nu}_A^i$ replaced by $\tilde{D}_{i,\Delta},\tilde{\nu}_\Delta^i$ respectively.

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

R-(i) is instrumental for showing $\int \tilde{\nu}_A(z)d\tilde{F}_Z = \frac{1}{G}\sum_{g\leq G}\tilde{\nu}_A(z_g) + o_p(G^{-1/2})$. Hence, we have

$$\tfrac{1}{G}\sum_g w_g m_A^i(x_g;\hat{\boldsymbol{\gamma}}) = \tfrac{1}{G}\sum_g [w_g m_A^i(x_g;\boldsymbol{\gamma}^*) + \tilde{\nu}_A(z_g)] + o_p(G^{-1/2}).$$

A similar equation holds for $\frac{1}{G}\sum_g w_g m_\Delta^i(x_g;\hat{\boldsymbol{\gamma}})$ with $m_A^i, \tilde{\nu}_A$ replaced by $m_\Delta^i, \tilde{\nu}_\Delta$. R-(ii) ensures $\frac{1}{G}\sum_{g\leq G} w_g m_\Delta^i(x_g;\hat{\boldsymbol{\gamma}}) \xrightarrow{p} E[w(X)m_\Delta^i(X;\boldsymbol{\gamma}^*)]$ when $\hat{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}^*$. This is useful for deriving the adjustment in the limiting variance due to the estimation of $\rho_0 \equiv E[w(X)] \equiv \Pr(X \in \omega)$. R-(iii) is needed to apply the Central Limit Theorem.

B.1. Asymptotic Property of $\hat{\alpha}_i$ and $\hat{\delta}_i$

For $i = 1, 2$, let $j \equiv 3 - i$. Recall

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

where

$$\begin{aligned} m_A^i(x;\boldsymbol{\gamma}) &\equiv \tfrac{\gamma_{i,i}(x)\gamma_0(x)-\gamma_i(x)\gamma_{0,i}(x)}{\gamma_0(x)\gamma_0(x)} - \tfrac{\gamma_{i,j}(x)\gamma_0(x)-\gamma_i(x)\gamma_{0,j}(x)}{\gamma_0(x)\gamma_0(x)}\tfrac{\gamma_{j,i}(x)\gamma_0(x)-\gamma_j(x)\gamma_{0,i}(x)}{\gamma_{j,j}(x)\gamma_0(x)-\gamma_j(x)\gamma_{0,j}(x)}; \text{ and} \\ m_\Delta^i(x;\boldsymbol{\gamma}) &\equiv \left(\tfrac{\gamma_{i,i}(x)\gamma_0(x)-\gamma_i(x)\gamma_{0,i}(x)}{\gamma_{i,j}(x)\gamma_0(x)-\gamma_i(x)\gamma_{0,j}(x)} - \tfrac{\gamma_{j,i}(x)\gamma_0(x)-\gamma_j(x)\gamma_{0,i}(x)}{\gamma_{j,j}(x)\gamma_0(x)-\gamma_j(x)\gamma_{0,j}(x)}\right)^{-1}\left(\tfrac{\gamma_{j,j}\gamma_0-\gamma_j\gamma_{0,j}}{\gamma_0(x)\gamma_0(x)}\right)^{-1}. \end{aligned}$$

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

$$\begin{aligned} \psi_{A,0}^i(x;\boldsymbol{\gamma}^*) &\equiv w(x)\left[\tilde{D}_{i,A,1}(x)\gamma_0^*(x) - \tilde{D}_{i,A,4}^{(i)}(x) - \tilde{D}_{i,A,5}^{(j)}(x)\right] + \mathcal{I}_{A,1}^i(x); \\ \psi_{A,i}^i(x;\boldsymbol{\gamma}^*) &\equiv w(x)\left[\tilde{D}_{i,A,2}(x)\gamma_0^*(x) - \tilde{D}_{i,A,6}^{(i)}(x) - \tilde{D}_{i,A,7}^{(j)}(x)\right] + \mathcal{I}_{A,2}^i(x); \\ \psi_{A,j}^i(x;\boldsymbol{\gamma}^*) &\equiv w(x)\left[\tilde{D}_{i,A,3}(x)\gamma_0^*(x) - \tilde{D}_{i,A,8}^{(i)}(x) - \tilde{D}_{i,A,9}^{(j)}(x)\right] + \mathcal{I}_{A,3}^i(x), \end{aligned}$$

with

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

where the terms $w(x), \tilde{D}_{i,A}, \tilde{D}_{i,\Delta}, \tilde{D}_{i,A}^{(s)}, \tilde{D}_{i,\Delta}^{(s)}$ are defined above (as in Section 5.1 in Lewbel and Tang (2014); and $l_i(x_{-i})$ and $h_i(x_{-i})$ denote the end-points of the interval $\{x_i : (x_i, x_{-i}) \in \omega\}$ for i and x_{-i}, which exist due to the convexity of ω. Note for any function $g(x)$, the expression $\int 1\{x \in \omega\}g(x)dF_X^*$ can be written as $\int\left(\int_{l_i(x_{-i})}^{h_i(x_{-i})} g(x)\gamma_0^*(x)dx_i\right)dx_{-i}$. The closed forms of $\tilde{D}_{i,A}, \tilde{D}_{i,\Delta}$ are provided in Section B.3. below. Define another term $\mathcal{I}_{A,2}^i(x)$ in a similar fashion by replacing $\tilde{D}_{i,A,4}$, $\tilde{D}_{i,A,5}$ in (4) 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 (4) 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.

Proposition B1 *Suppose A1,2,3 and Assumptions S, W, K, B, D and R hold. Then for* $i = 1, 2$,

$$\sqrt{G}\left(\hat{A}^i - A^i\right) \xrightarrow{d} \mathcal{N}\left(0\ ,\ Var\left[w(X)m_A^i\left(X;\boldsymbol{\gamma}^*\right) + \tilde{\nu}_A^i(Z)\right]\right)$$

$$\sqrt{G}\left(\hat{\Delta}^i - \Delta^i\right) \xrightarrow{d} \mathcal{N}\left(0\ ,\ \rho_0^{-2}Var\left[w(X)\left(m_\Delta^i\left(X;\boldsymbol{\gamma}^*\right) - \Delta^i\right) + \tilde{\nu}_\Delta^i(Z)\right]\right) \tag{5}$$

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

Proposition B1 follows from the steps in Section 8 of Newey and McFadden (1994) as explained in Section 5.1 of Lewbel and Tang (2014). The proof of this proposition is collected in Section B.3 below. It follows from Proposition B1 and the Slutsky's Theorem that $\hat{\delta}_i$ converges to δ_i at the parametric rate. Theorem 4 in Lewbel and Tang (2014) follows immediately from this proposition.

B.2. Asymptotic property of $\hat{\beta}_i$

In this subsection we provide the technical conditions for deriving the limiting distribution of $\hat{\beta}_i$ in Section 5.2 in Lewbel and Tang (2014).Let $\boldsymbol{\gamma}_{(i)} \equiv (\gamma_0, \gamma_j, \gamma_{0,i}, \gamma_{j,i})$ denote the subvector of $\boldsymbol{\gamma}$ which enters the definition of the estimator below. Following the convention in Section 5 in Lewbel and Tang (2014), let $\hat{\boldsymbol{\gamma}}_{(i)}, \boldsymbol{\gamma}_{(i)}^*$ denote kernel estimates and the truth in the DGP respectively.

Assumption S' *(i) Assumption S holds over the support of* X. *(ii) The true density* $f_{X_i|x_{-i}}^*(x_i)$ *is bounded above and away from zero by some positive constant over the support of* X.

Along with kernel and bandwidth conditions in Assumptions K and B, conditions (i) and (ii) in Assumption S' ensure $\sup_x ||\hat{\boldsymbol{\gamma}}_{(i)} - \boldsymbol{\gamma}_{(i)}^*||$ converges in probability to 0 at rates faster than $G^{1/4}$. Bounding $\gamma^*(x)$ and $f_{X_i|x_{-i}}^*(x_i)$ away from zero over the support of X and over the support of X_i given x_{-i} respectively 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 the support of X for $i = 1, 2$ and $j = 3 - i$.

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

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

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

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

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

$$\begin{aligned} \tilde{D}_{i,B,\delta}\left(z;\delta_i,\boldsymbol{\gamma}_{(i)}\right) &\equiv \tfrac{\partial}{\partial\tilde{\delta}_i} m_B^i\left(z;\tilde{\delta}_i,\boldsymbol{\gamma}_{(i)}\right)|_{\tilde{\delta}_i=\delta_i} \\ = \quad & \tilde{x}'\left\{-H'(-x_i+\delta_i p_j(x))\tfrac{p_j(x)[1-\delta_i p_{j,i}(x)]}{f_{X_i|x_{-i}}(x_i)} - [d_i - H(-x_i+\delta_i p_j(x))]\tfrac{p_{j,i}(x)}{f_{X_i|x_{-i}}(x_i)}\right\}. \end{aligned}$$

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

$$\begin{aligned} \tilde{D}_{i,B,1}(z) &\equiv \tfrac{\partial m_B^i\left(z;\delta_i,\boldsymbol{\gamma}^*_{(i)}\right)}{\partial p_j^*(x)}\tfrac{\partial p_j^*(x)}{\partial\gamma_0(x)} + \tfrac{\partial m_B^i\left(z;\delta_i,\boldsymbol{\gamma}^*_{(i)}\right)}{\partial p_{j,i}^*(x)}\tfrac{\partial p_{j,i}^*(x)}{\partial\gamma_0(x)}, \\ \tilde{D}_{i,B,2}(z) &\equiv \tfrac{\partial m_B^i\left(z;\delta_i,\boldsymbol{\gamma}^*_{(i)}\right)}{\partial p_j^*(x)}\tfrac{\partial p_j^*(x)}{\partial\gamma_j(x)} + \tfrac{\partial m_B^i\left(z;\delta_i,\boldsymbol{\gamma}^*_{(i)}\right)}{\partial p_{j,i}^*(x)}\tfrac{\partial p_{j,i}^*(x)}{\partial\gamma_j(x)}, \\ \tilde{D}_{i,B,3}(z) &\equiv \tfrac{\partial m_B^i\left(z;\delta_i,\boldsymbol{\gamma}^*_{(i)}\right)}{\partial p_{j,i}^*(x)}\tfrac{\partial p_{j,i}^*(x)}{\partial\gamma_{0,i}(x)}, \text{ and } \tilde{D}_{i,B,4}(z) \equiv \tfrac{\partial m_B^i\left(z;\delta_i,\boldsymbol{\gamma}^*_{(i)}\right)}{\partial p_{j,i}^*(x)}\tfrac{\partial p_{j,i}^*(x)}{\partial\gamma_{j,i}(x)}. \end{aligned}$$

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

$$M_{B,\gamma}^i\left(z,\boldsymbol{\gamma}_{(i)}\right) = \boldsymbol{\gamma}_{(i)}(x)'\tilde{D}_{i,B}(z). \tag{6}$$

Note $\tilde{D}_{i,B}$ on the right-hand side of (6) is evaluated at the true parameters δ_i and $\boldsymbol{\gamma}^*_{(i)}$.

For both i, let $\bar{D}_{i,B,k}(x) \equiv E\left[\tilde{D}_{i,B,k}(Z) \mid x\right]$ for all k, and $\bar{D}_{i,B,k}^{(s)}(x) \equiv \frac{\partial\left[\bar{D}_{i,B,k}(x)\gamma_0^*(x)\right]}{\partial X_s}$ 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 the support of 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 $M^i_{B,f}(z;h)$ is the Frechet derivative of $m^i_B\left(z;\delta_i,\boldsymbol{\gamma}_{(i)}\right)$ with respect to $\boldsymbol{\gamma}_{(i)}$ at $\boldsymbol{\gamma}^*_{(i)}$ in DGP.Define

$$\psi^i_{B,f}(x) \equiv E\left(\left.\frac{C^*_i(Z)}{\gamma^*_0(X)}\right| x_{-i}\right)\gamma^*_0(x_{-i}) - \frac{E\left(C^*_i(Z)|x\right)}{\gamma^*_0(x)}\int \gamma^*_0(t,x_{-i})dt,$$

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 $\int M^i_{B,f}(z;h)dF^*_Z = \int \psi^i_{B,f}(x)h(x)dx$.

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

$$\begin{aligned}
\tilde{\nu}^i_B(z_g) &\equiv \psi^i_{B,0}(x_g)+\psi^i_{B,f}(x_g)+\psi^i_{B,j}(x_g)d_{g,j} - E\left[\psi^i_{B,0}(X)+\psi^i_{B,f}(X)+\psi^i_{B,j}(X)D_j\right];\\
\Psi^i_\delta(z_g) &\equiv -\frac{1}{\rho_0}\left[w_g\left(m^i_\Delta\left(x_g;\boldsymbol{\gamma}^*_{(i)}\right)-\Delta^i\right)+\tilde{\nu}^i_\Delta(z_g)\right];\ \mathcal{M}^*_{i,\delta}\equiv E\left[\tilde{D}_{i,B,\delta}\left(Z;\delta_i,\boldsymbol{\gamma}^*_{(i)}\right)\right];\\
\Psi^i_B(z) &\equiv m^i_B\left(z;\delta_i,\boldsymbol{\gamma}^*_{(i)}\right)+\tilde{\nu}^i_B(z)+\mathcal{M}^*_{i,\delta}\Psi^i_\delta(z).
\end{aligned}$$

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

Assumption D' *(i) For $i=1,2$, the second order derivatives of $m^i_B\left(z;\delta_i,\boldsymbol{\gamma}_{(i)}\right)$ with respect to $\boldsymbol{\gamma}_{(i)}(x)$ at $\boldsymbol{\gamma}^*_{(i)}(x)$ are continuous and bounded over the support of Z. (ii) For both i and $s=1,2$, $\tilde{D}^{(s)}_{i,B}$, are continuous and bounded over the support of Z. (iii) For $i=1,2$ and $j=3-i$, there exists $c>0$ such that $E[\sup_{\|\eta\|\le c}\left\|\psi^i_{B,s}(x+\eta)\right\|^4]<\infty$ for $s=0,j$ and $E[\sup_{\|\eta\|\le c}||\psi^i_{B,f}(x+\eta)||^4]$ $<\infty$.*

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

Condition R'-(i) is used for showing the difference between $\frac{1}{G}\sum_g m^i_B(z_g;\,\hat{\delta}_i,\hat{\boldsymbol{\gamma}}_{(i)})$ and $\frac{1}{G}\sum_g m^i_B(z_g;\,\delta_i,\hat{\boldsymbol{\gamma}}_{(i)})$ 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 $E[\tilde{D}_{i,B,\delta}(z;\delta_i,\boldsymbol{\gamma}^*_{(i)})]$ 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), $\frac{1}{G}\sum_g[m^i_B(x_g;\delta_i,\hat{\boldsymbol{\gamma}}_{(i)})-m^i_B(x_g;\delta_i,\boldsymbol{\gamma}^*_{(i)})]$ can be written as $\int M^i_{B,\gamma}(z,\hat{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}^*_{(i)})+M^i_{B,f}(z;\hat{\gamma}_0-\gamma^*_0)dF^*_Z$ 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 $\boldsymbol{\gamma}_{(i)}$ 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:

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

using an argument of integration by parts.

Using this equation, the difference $\int M_{B,\gamma}^i(z;\hat{\boldsymbol{\gamma}}_{(i)} - \boldsymbol{\gamma}_{(i)}^*) + M_{B,f}^i(z;\hat{\boldsymbol{\gamma}}_{(i)} - \boldsymbol{\gamma}_{(i)}^*)dF_Z^*$ 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 $\frac{1}{G}\sum_g m_B^i(z_g;\delta_i,\hat{\boldsymbol{\gamma}}_{(i)})$ and $\frac{1}{G}\sum_g [m_B^i(z_g;\delta_i,\boldsymbol{\gamma}_{(i)}^*)+\tilde{\nu}_B^i(z_g)]$ 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 $E[\sup_{\|\eta\|\leq c_i} \left\|\bar{D}_{i,B}(x+\eta)\right\|^4] < \infty$ 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.

Proposition B2 *Under A1, 2, 3, 4' and Assumptions S', K, B, D', W', R',* $\sqrt{G}(\hat{\beta}_i - \beta_i) \overset{d}{\to} N(0,\Sigma_B^i)$ *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. Proof of the proposition is included in the next subsection.

B.3. Proofs for Propositions B1 and B2

The first two subsections (B.3.1. and B.3.2) below prove Proposition B1 in Section B.1.; the last subsection proves Proposition B2 in Section B.2.

B.3.1. 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. Let w_g, $w(x)$ be defined as in the text. For any x, define a linear functional of $\boldsymbol{\gamma}$:

$$\begin{aligned} M_A^i\left(x;\boldsymbol{\gamma}\right) &\equiv [\boldsymbol{\gamma}(x)]'\,\tilde{D}_{i,A}\left(x\right), \text{ where} \\ \underset{\text{9-by-1}}{\tilde{D}_{i,A}\left(x;\boldsymbol{\gamma}^*\right)} &\equiv \begin{bmatrix} \frac{-1}{a^2(ce-ak)^2}\begin{pmatrix} c^2fe^2+a^2fk^2-c^2dge-2abdk^2-a^2ghk \\ -bche^2+bcdke-2acfke+2abhke+2acdgk \end{pmatrix},.. \\ \frac{1}{a(ak-ce)}\left(he-dk\right),\ \frac{be-ag}{a(ce-ak)^2}\left(he-dk\right),\ \frac{1}{a(ak-ce)}\left(cg-bk\right),.. \\ -\frac{1}{a^2e}\left(bd-af-\frac{(be-ag)(cd-ah)}{ak-ce}\right),\ \frac{1}{a},\ \frac{1}{a(ak-ce)}\left(cd-ah\right),.. \\ \frac{1}{a(ak-ce)}\left(be-ag\right),\ \frac{be-ag}{a(ak-ce)^2}\left(cd-ah\right) \end{bmatrix}'. \end{aligned} \tag{7}$$

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

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

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

$$\begin{aligned} &\left\|w_g\left[m_A^i\left(x_g;\hat{\boldsymbol{\gamma}}\right)-m_A^i\left(x_g;\boldsymbol{\gamma}^*\right)-\left(\hat{\boldsymbol{\gamma}}(x_g)-\boldsymbol{\gamma}^*(x_g)\right)'\tilde{D}_{i,A}\left(x_g\right)\right]\right\| \\ \leq\ & b(x_g)\sup\nolimits_{x\in\omega}\left\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^*\right\|^2 \end{aligned}$$

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.[1] Hence $\sup_{x\in\omega}\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^*\|=o_p(G^{-1/4})$, and

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

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

$$\begin{aligned} & G^{-1}\sum\nolimits_g w_g\left(\hat{\boldsymbol{\gamma}}(x_g)-\boldsymbol{\gamma}^*(x_g)\right)'\tilde{D}_{i,A}(x_g)-\int w(x)\left(\hat{\boldsymbol{\gamma}}(x)-\boldsymbol{\gamma}^*(x)\right)'\tilde{D}_{i,A}(x)dF_X^* \\ =\ & G^{-1}\sum\nolimits_g w_g\left(\hat{\boldsymbol{\gamma}}(x_g)-\bar{\boldsymbol{\gamma}}(x_g)\right)'\tilde{D}_{i,A}(x_g)-\int w(x)\left(\hat{\boldsymbol{\gamma}}(x)-\bar{\boldsymbol{\gamma}}(x)\right)'\tilde{D}_{i,A}(x)dF_X^* \\ +\ & G^{-1}\sum\nolimits_g w_g\left(\bar{\boldsymbol{\gamma}}(x_g)-\boldsymbol{\gamma}^*(x_g)\right)'\tilde{D}_{i,A}(x_g)-\int w(x)\left(\bar{\boldsymbol{\gamma}}(x)-\boldsymbol{\gamma}^*(x)\right)'\tilde{D}_{i,A}(x)dF_X^* \end{aligned} \tag{8}$$

[1] Recall the boundedness of ω is included in the definition of the n.d.s. set. However, in order to show $\sup_{x\in\omega}\|\hat{\gamma}-\gamma^*\|=o_p(G^{-1/4})$, it is not necessary for ω to be bounded. Specifically, when ω is unbounded, we could establish the $G^{1/4}$-convergence of $\hat{\gamma}$ using elements from the empirical processes theory (e.g., Lemma A1 in Volgushev (2013)) upon some functional class such as $F_n=\{K(\frac{x-\cdot}{\sigma_n}):x\in\omega,\sigma_n>0\}$ under a suitable choice of the kernel function. We are grateful to anonymous referee for pointing this out.

Let $a_G(z_g, z_s) \equiv$

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

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 (8) takes the form of a V-statistic:

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

By Assumption D-(i) and boundedness of kernels in K, $\left\|w(x) M_A^i\left(x; \boldsymbol{\gamma}\right)\right\| \leq c(x) \sup_{x\in\omega} \|\boldsymbol{\gamma}\|$ 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 (8). Since $\left\|w(x) M_A^i\left(x; \boldsymbol{\gamma}\right)\right\| \leq c(x) \sup_{x\in\omega} \|\boldsymbol{\gamma}\|$ and $E\left[c(X)^2\right] < \infty$,

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

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

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

Then by the Chebychev's Inequality, the second difference in (8) is $o_p(G^{-1/2})$. This proves the lemma. *Q.E.D.*

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

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

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

$$\begin{aligned} & \int w(x) M_A^i\left(x; \boldsymbol{\gamma}\right) dF_X^* \\ = & \int w(x) \begin{bmatrix} \gamma_0(x) \\ \gamma_i(x) \\ \gamma_j(x) \end{bmatrix}' \begin{bmatrix} \tilde{D}_{i,A,1}(x) \\ \tilde{D}_{i,A,2}(x) \\ D_{i,A,3}(x) \end{bmatrix} + w(x) \underbrace{\begin{bmatrix} \gamma_{0,i}(x), \gamma_{0,j}(x), .. \\ \gamma_{i,i}(x), \gamma_{i,j}(x), .. \\ \gamma_{j,i}(x), \gamma_{j,j}(x) \end{bmatrix}}_{1-by-6} \underbrace{\begin{bmatrix} \tilde{D}_{i,A,4}(x) \\ \vdots \\ \tilde{D}_{i,A,9}(x) \end{bmatrix}}_{6-by-1} dx. \qquad (9) \end{aligned}$$

Applying integration-by-parts to the second inner product in the integrand of (9) and using the convexity of ω (Assumption W), we have

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

for any $\boldsymbol{\gamma}$ 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 as in Section B1 above. These three functions only depend on $\boldsymbol{\gamma}^*$ but not $\boldsymbol{\gamma}$. By conditions in S and K, both $\hat{\boldsymbol{\gamma}}$ and $\boldsymbol{\gamma}^*$ are continuously differentiable. For the rest of the proof, we focus on the case where all coordinates in $\tilde{x}$ are continuously distributed.[2] It follows that:

$$\begin{aligned} & \int w(x)M_A^i\left(x;\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^*\right)dF_X^* = \int \textstyle\sum_{s=0,i,j}\psi_{A,s}^i\left(x\right)\left[\hat{\gamma}_s(x)-\gamma_s^*(x)\right]dx \\ = \; & G^{-1}\textstyle\sum_g \int \sum_{s=0,i,j}\psi_{A,s}^i(x)d_{g,s}K_\sigma(x-x_g)dx - E\left[\psi_A^i(X)D\right] \\ = \; & G^{-1}\textstyle\sum_g \int \left(\sum_{s=0,i,j}\psi_{A,s}^i(x)d_{g,s}-\mu\right)K_\sigma(x-x_g)dx \equiv \int\left(\sum_{s=0,i,j}\psi_{A,s}^i(x)d_s-\mu\right)d\hat{F}_Z \end{aligned}$$

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 $G^{-1}\sum_g \int a(x,d_{1,g},d_{2,g})K_\sigma(x-x_g)dx$. By construction, $\frac{1}{G}\sum_g\int\mu K_\sigma(x-x_g)dx=\mu$. Hence by definition:

$$\begin{aligned} & \int\left(\textstyle\sum_{s=0,i,j}\psi_{A,s}^i(x)d_s-\mu\right)d\hat{F}_Z - G^{-1}\sum_g\left(\sum_{s=0,i,j}\psi_{A,s}^i(x_g)d_{s,g}-\mu\right) \\ = \; & \int\left(\textstyle\sum_{s=0,i,j}\psi_{A,s}^i(x)d_s\right)d\hat{F}_Z - G^{-1}\sum_g\left(\sum_{s=0,i,j}\psi_{A,s}^i(x_g)d_{s,g}\right) \\ = \; & G^{-1}\textstyle\sum_g\left[\sum_{s=0,i,j}d_{s,g}\left(\int\psi_{A,s}^i(x)K_\sigma(x-x_g)dx-\psi_{A,s}^i(x_g)\right)\right] \end{aligned} \qquad (10)$$

It remains to show that the r.h.s. of (10) is $o_p(G^{-1/2})$. First, note

$$\begin{aligned} & \left\|\sqrt{G}E\left[\textstyle\sum_{s=0,i,j}d_{s,g}\left(\int\psi_{A,s}^i(x)K_\sigma(x-x_g)dx-\psi_{A,s}^i(x_g)\right)\right]\right\| \\ = \; & \sqrt{G}\left\|\textstyle\sum_{s=0,i,j}\left(\int\left(\int\psi_{A,s}^i(x)K_\sigma(x-\hat{x})dx\right)\gamma_s^*(\hat{x})d\hat{x}-\int\psi_{A,s}^i(x)\gamma_s^*(x)dx\right)\right\| \\ = \; & \sqrt{G}\left\|\textstyle\sum_{s=0,i,j}\left(\int\int\psi_{A,s}^i(\hat{x}+\sigma u)K(u)\gamma_s^*(\hat{x})dud\hat{x}-\int\psi_{A,s}^i(x)\gamma_s^*(x)dx\right)\right\| \\ = \; & \sqrt{G}\left\|\textstyle\sum_{s=0,i,j}\left(\int\psi_{A,s}^i(x)\left(\int K(u)\gamma_s^*(x-\sigma u)du\right)dx-\int\psi_{A,s}^i(x)\gamma_s^*(x)dx\right)\right\| \end{aligned} \qquad (11)$$

[2] General cases involving discrete coordinates only add to complex notations, but do not require any changes in the the arguments. For example, $\int\psi_A^i(x)d_gK_\sigma(x-x_g)dx$ in the proof would be replaced by $\sum_{\tilde{x}_d}\int\psi_A^i(x_c,\tilde{x}_d)d_gK_\sigma(x_c-x_{g,c})dx_c1\{\tilde{x}_{g,d}=\tilde{x}_d\}$. Likewise, $\int K_\sigma(x-x_g)dx$ would be replaced by $\sum_{\tilde{x}_d}\int K_\sigma(x_c-x_{g,c})dx_c1\{\tilde{x}_{g,d}=\tilde{x}_d\}$ (which equals one).

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 $\boldsymbol{\gamma}^*$ and the forms of $\psi^i_{A,s}$ for $s = 0, i, j$.) Next, using $\int K(u)du = 1$ and re-arranging terms, we can write (11) as:

$$\begin{aligned} & \sqrt{G} \left\| \sum\nolimits_{s=0,i,j} \int \psi^i_{A,s}(x) \left\{ \int \left[\gamma^*_s(x - \sigma u) - \gamma^*_s(x) \right] K(u)du \right\} dx \right\| \\ \leq \ & \sqrt{G} \int \left\| \psi^i_A(x) \right\| \left\| \int \left[\gamma^*(x - \sigma u) - \gamma^*(x) \right] K(u)du \right\| dx, \end{aligned} \tag{12}$$

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

$$E[\hat{\gamma}_s(x)] = E[D_s K_\sigma(x - X_g)] = \int E[D_s|x_g] K_\sigma(x - x_g) \gamma^*_0(x_g) dx_g = \int \gamma^*_s \left(x - \sigma u \right) K(u)du$$

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 $\boldsymbol{\gamma}^*$ 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), $\int \sup_{x\in\omega} \left\| \psi^i_A(x) \right\| dx < \infty$ and the r.h.s. of the inequality in (12) 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

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

By definition of ψ^i_A and the smoothness condition in D-(i), ψ^i_A is continuous in x almost everywhere, and $\psi^i_{A,s}(x + \sigma u) \to \psi^i_{A,s}(x)$ for almost all x and u and $s = 0, i, j$. By condition R-(i), 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 domain of $K(.)$. It then follows from the Dominated Convergence Theorem that

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

for almost all x as $\sigma \to 0$, 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

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

It then follows from the Cauchy-Schwartz Inequality that

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

Hence both the mean and variance of the product of $\sqrt{G}$ and the left-hand side of (10) 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 B.1.1 and Lemma B.1.2 that

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

which implies

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

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.

B.3.2. 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 Section 5.1 of Lewbel and Tang (2014) as

$$\hat{\Delta}^i\equiv(\hat{\rho})^{-1}\left[\tfrac{1}{G}\sum_g w_g m_\Delta^i\left(x_g;\hat{\boldsymbol{\gamma}}\right)\right]\tag{13}$$

where $\hat{\rho}\equiv\left[G^{-1}\sum_g w_g\right]^{-1}$. Let $w_g, w(x)$ be defined as in the text of Lewbel and Tang (2014). Define $\rho_0\equiv\Pr(X\in\omega)$. The first step in Lemma B.2.1 below is to describe the impact of replacing $\hat{\rho}$ with the true population moment ρ_0 in (13).

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

$$\hat{\Delta}^i=G^{-1}\sum_g\left[\rho_0^{-1}w_g m_\Delta^i\left(x_g;\hat{\gamma}\right)-\mathcal{M}_{i,\rho}^*\left(w_g-\rho_0\right)\right]+o_p(G^{-1/2})\tag{14}$$

where

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

Proof of Lemma B.2.1. Apply the Taylor's Expansion to (13) around ρ_0, we get:

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

where

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

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

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

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 $\boldsymbol{\gamma}^*$ at some x). The dependence of these short-hands on x are suppressed in notations. Similar to the definition of M_A^i above, for any x, define a linear functional of $\boldsymbol{\gamma}$:

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

The correction terms $\psi_{\Delta}^i \equiv [\psi_{\Delta,0}^i, \psi_{\Delta,i}^i, \psi_{\Delta,j}^i]$ are defined in Section B1.

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

$$\frac{1}{G}\sum_g w_g m_{\Delta}^i(x_g;\hat{\boldsymbol{\gamma}}) = \frac{1}{G}\sum_g \left[w_g m_{\Delta}^i(x_g;\boldsymbol{\gamma}^*) + \tilde{\nu}_{\Delta}(z_g)\right] + o_p(G^{-1/2}) \tag{15}$$

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 B.2.2 uses the same arguments as those for Lemma B.1.1 and Lemma B.1.2. Combining results in (14) and (15), we have

$$\begin{aligned} \hat{\Delta}^i - \Delta^i &= G^{-1}\sum_g \left[w_g m_{\Delta}^i(x_g;\boldsymbol{\gamma}^*)/\rho_0 + \tilde{\nu}_{\Delta}^i(z_g)/\rho_0 - \Delta^i - \mathcal{M}_{i,\rho}^*(w_g - \rho_0)\right] + o_p(G^{-1/2}) \\ &= \frac{1}{\rho_0}\frac{1}{G}\sum_g \left[w_g m_{\Delta}^i(x_g;\boldsymbol{\gamma}^*) + \tilde{\nu}_{\Delta}^i(z_g) - \rho_0\Delta^i - \Delta^i(w_g - \rho_0)\right] + o_p(G^{-1/2}), \end{aligned} \tag{16}$$

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

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

B.3.3. Limiting distribution of $\sqrt{G}\left(\hat{\beta}_i - \beta_i\right)$

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

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

where

$$\begin{aligned} \mathcal{M}_{i,\delta}^* &\equiv E\left[\tilde{D}_{i,B,\delta}\left(Z;\delta_i,\boldsymbol{\gamma}_{(i)}^*\right)\right] \text{ and} \\ \Psi_{\delta}^i(z_g) &\equiv -\frac{1}{\rho_0}\left[w_g m_{\Delta}^i\left(x_g;\boldsymbol{\gamma}_{(i)}^*\right) + \tilde{\nu}_{\Delta}^i(z_g) - \rho_0\Delta^i - \Delta^i(w_g - \rho_0)\right] \end{aligned} \tag{17}$$

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

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

where $\tilde{D}_{i,B,\delta}\left(z_g;\bar{\delta}_i,\hat{\boldsymbol{\gamma}}_{(i)}\right)$ 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

$$\begin{aligned}&\tfrac{1}{\sqrt{G}}\textstyle\sum_g m_B^i\left(z_g;\hat{\delta}_i,\hat{\boldsymbol{\gamma}}_{(i)}\right)\\ =\;&\tfrac{1}{\sqrt{G}}\textstyle\sum_g m_B^i\left(z_g;\delta_i,\hat{\boldsymbol{\gamma}}_{(i)}\right)+\tfrac{1}{G}\sum_g\tilde{D}_{i,B,\delta}\left(z_g;\bar{\delta}_i,\hat{\boldsymbol{\gamma}}_{(i)}\right)\sqrt{G}\left(\hat{\delta}_i-\delta_i\right).\end{aligned}\tag{18}$$

With $\alpha_i=-1$, it follows from (16) that $\sqrt{G}(\hat{\delta}_i-\delta_i)=G^{-1/2}\sum_g\Psi_\delta^i(z_g)+o_p(1)$, where the influence function is defined as in (17). (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 $\boldsymbol{\gamma}_{(i)}$ 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, $\tilde{D}_{i,B,\delta}\left(z\,;\delta_i,\boldsymbol{\gamma}_{(i)}\right)$ is continuous in $\delta_i,\boldsymbol{\gamma}_{(i)}$ with probability one. Note $\bar{\delta}_i\xrightarrow{p}\delta_i$ as a consequence of $\hat{\delta}_i\xrightarrow{p}\delta_i$. Under conditions in S'-(i),(ii), K, and B, $\sup_x\left\|\hat{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}_{(i)}^*\right\|\xrightarrow{p}0$ at the rate faster than $G^{1/4}$. It then follows that $\frac{1}{G}\sum_g\tilde{D}_{i,B,\delta}\left(z_g;\bar{\delta}_i,\hat{\boldsymbol{\gamma}}_{(i)}\right)\xrightarrow{p}\mathcal{M}_{i,\delta}^*$ under condition D'-(i). Hence the second term on the right-hand side of (18) is $\left[\mathcal{M}_{i,\delta}^*+o_p(1)\right]*\left[G^{-1/2}\sum_g\Psi_\delta^i(z_g)+o_p(1)\right]$, where $G^{-1/2}\sum_g\Psi_\delta^i(z_g)$ is $O_p(1)$. It then follows that

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

This proves Lemma B.3.1. *Q.E.D.*

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

$$\begin{aligned}&\tfrac{1}{G}\textstyle\sum_g\left[m_B^i\left(z_g;\delta_i,\hat{\boldsymbol{\gamma}}_{(i)}\right)-m_B^i\left(z_g;\delta_i,\boldsymbol{\gamma}_{(i)}^*\right)\right]\\ =\;&\textstyle\int M_{B,\gamma}^i\left(z,\hat{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}_{(i)}^*\right)+M_{B,f}^i\left(z,\hat{\gamma}_0-\gamma_0^*\right)dF_Z^*+o_p(G^{-1/2}).\end{aligned}$$

Proof of Lemma B.3.2. Under conditions S'-(i), (ii), we have

$$\begin{aligned}&\left\|\begin{array}{l}m_B^i\left(z_g;\delta_i,\hat{\boldsymbol{\gamma}}_{(i)}\right)-m_B^i\left(z_g;\delta_i,\boldsymbol{\gamma}_{(i)}^*\right)\\-M_{B,\gamma}^i\left(z_g,\hat{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}_{(i)}^*\right)-M_{B,f}^i\left(z_g,\hat{\gamma}_0-\gamma_0^*\right)\end{array}\right\|\\ \leq\;&b_{\gamma,f}(z_g)\sup_{x\in\Omega_X}||\hat{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}_{(i)}^*||^2,\end{aligned}\tag{19}$$

for some non-negative function $b_{\gamma,f}$ (which may depend on true parameters in the DGP δ_i and $\boldsymbol{\gamma}^*$) and $E[b_{\gamma,f}(Z)]<+\infty$. Under Assumptions S'-(i),(ii), K and B, $\sup_x||\hat{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}_{(i)}^*||$ is $o_p(G^{-1/4})$.

Thus the right-hand side of the inequality in (19) is $o_p(G^{-1/2})$. It follows from the triangular inequality and the Law of Large Numbers that:

$$\frac{1}{G}\sum_g \left[\begin{array}{l} m_B^i\left(z_g;\delta_i,\hat{\boldsymbol{\gamma}}_{(i)}\right) - m_B^i\left(z_g;\delta_i,\boldsymbol{\gamma}^*_{(i)}\right) \\ -M^i_{B,\gamma}\left(z_g,\hat{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}^*_{(i)}\right) - M^i_{B,f}\left(z_g,\hat{\gamma}_0-\gamma_0^*\right) \end{array}\right] = o_p(G^{-\frac{1}{2}})$$

By construction,

$$\begin{aligned} & \tfrac{1}{G}\textstyle\sum_g M^i_{B,\gamma}\left(z_g,\hat{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}^*_{(i)}\right) - \int M^i_{B,\gamma}\left(z;\hat{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}^*_{(i)}\right)dF_Z^* \\ = \; & \tfrac{1}{G}\textstyle\sum_g M^i_{B,\gamma}\left(z_g,\hat{\boldsymbol{\gamma}}_{(i)}-\bar{\boldsymbol{\gamma}}_{(i)}\right) - \int M^i_{B,\gamma}\left(z;\hat{\boldsymbol{\gamma}}_{(i)}-\bar{\boldsymbol{\gamma}}_{(i)}\right)dF_Z^* + \\ & \tfrac{1}{G}\textstyle\sum_g M^i_{B,\gamma}\left(z_g,\bar{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}^*_{(i)}\right) - \int M^i_{B,\gamma}\left(z;\bar{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}^*_{(i)}\right)dF_Z^*. \end{aligned} \tag{20}$$

where $\bar{\boldsymbol{\gamma}}_{(i)}(x) \equiv E[\hat{\boldsymbol{\gamma}}_{(i)}(x)]$ is a function of x (i.e. the expectation is taken with respect to the random sample used for estimating $\hat{\boldsymbol{\gamma}}_{(i)}$ 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 $a^2_{B,G}(z_s) \equiv \int a_{B,G}(z,z_s)dF_Z^*(z)$, and

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

Then the first difference on the r.h.s. of (20) can be written as:

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

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 $\boldsymbol{\gamma}^*_{(i)}$ 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^i_B w.r.t. $\boldsymbol{\gamma}_{(i)}$ in D'-(ii) implies $E\left[\left\|M^i_{B,\gamma}\left(Z;\bar{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}^*_{(i)}\right)\right\|^2\right] \leq C_3'\sup_x\left\|\bar{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}^*_{(i)}\right\|$. Under K and S'-(a), $\sup_x\left\|\bar{\boldsymbol{\gamma}}_{(i)}-\boldsymbol{\gamma}^*_{(i)}\right\| = o(\sigma^m)$, which converges to 0 as $\sigma \longrightarrow 0$ under Assumption B. It then follows from the Chebychev's Inequality that the second difference in (20) is $o_p(G^{-1/2})$. We can also decompose the linear functional in the second term in a similar manner and have:

$$\begin{aligned} & \tfrac{1}{G}\textstyle\sum_g M^i_{B,f}\left(z_g,\hat{\gamma}_0-\gamma_0^*\right) - \int M^i_{B,f}\left(z;\hat{\gamma}_0-\gamma_0^*\right)dF_Z^* \\ = \; & \tfrac{1}{G}\textstyle\sum_g M^i_{B,f}\left(z_g,\hat{\gamma}_0-\bar{\gamma}_0\right) - \int M^i_{B,f}\left(z;\hat{\gamma}_0-\bar{\gamma}_0\right)dF_Z^* + \\ & \tfrac{1}{G}\textstyle\sum_g M^i_{B,f}\left(z_g,\bar{\gamma}_0-\gamma_0^*\right) - \int M^i_{B,f}\left(z;\bar{\gamma}_0-\gamma_0^*\right)dF_Z^*. \end{aligned} \tag{21}$$

Arguments similar to those above suggest both differences in (21) are $o_p(G^{-1/2})$ under conditions in Assumption S', K, B and D'. *Q.E.D.*

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

$$\int M^i_{B,\gamma}(z, \hat{\boldsymbol{\gamma}}_{(i)} - \boldsymbol{\gamma}^*_{(i)}) + M^i_{B,f}\left(z, \hat{\gamma}_0 - \gamma^*_0\right) dF^*_Z = \tfrac{1}{G}\textstyle\sum_g \tilde{\nu}^i_B(z_g) + o_p(G^{-1/2}) \tag{22}$$

where

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

Proof of Lemma B.3.3. By smoothness of the true population function $\boldsymbol{\gamma}^*_{(i)}$ in DGP in S'-(a), and that of the kernel function in S', K, D', and using W', we can apply integration-by-parts to write the left-hand side of the equation (22) as:

$$\int \begin{bmatrix} \hat{\gamma}_0(x) - \gamma^*_0(x) \\ \hat{\gamma}_j(x) - \gamma^*_j(x) \end{bmatrix}' \begin{bmatrix} \psi^i_{B,0}(x) \\ \psi^i_{B,j}(x) \end{bmatrix} + \psi^i_{B,f}(x)\left[\hat{\gamma}_0(x) - \gamma^*_0(x)\right] dx, \tag{23}$$

where $\psi^i_{B,0}$, $\psi^i_{B,j}$ and $\psi^j_{B,f}$ are defined as in Section B2 above. By arguments similar to those in the proof of Lemma B.1.2, the expression in (23) is $\frac{1}{G}\sum_g \nu^i_B(z_g) + o_p(G^{-\frac{1}{2}})$ due to Assumptions S', K, B, D' and R', and that $\psi^i_{B,f}(x)$ is almost everywhere continuous in x. *Q.E.D.*

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

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

where

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

By construction

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

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

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

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

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

where

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

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

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

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

B.4. Limiting distribution of $\sqrt{G}\left(\hat{T}(\bar{\omega}) - \mu_{\bar{\omega}}\right)$

In this subsection, we derive the limiting distribution of the test statistic in Section 7.2 in Lewbel and Tang (2014). Define

$$v(x) \equiv \frac{1\{x \in \bar{\omega}\}}{\gamma_0^*(x)} [2p_1^*(x)p_2^*(x),\ -p_2^*(x),\ -p_1^*(x)] \tag{24}$$

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 the unconditional support of 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 B.4.1 *Suppose assumptions T1,2,3 hold. Then*

$$\hat{T}(\bar{\omega}) - \tfrac{1}{G}\textstyle\sum_g 1\{x_g \in \bar{\omega}\}m(z_g; \gamma^*) = \displaystyle\int M(z, \hat{\gamma} - \gamma^*)dF_Z^* + o_p(G^{-1/2}) \tag{25}$$

where $M(z,\gamma) \equiv v(x)\gamma(x)$ with $v(.)$ defined in (24) above.

Proof of Lemma B.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,

$$1\{x \in \bar{\omega}\}|m(z;\gamma) - m(z;\gamma^*) - M(z, \gamma - \gamma^*)| \leq C \sup_{x\in\bar{\omega}} ||\gamma - \gamma^*||^2 \tag{26}$$

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*),

$$\tfrac{1}{G}\textstyle\sum_g 1\{x_g \in \bar{\omega}\}[m(z_g;\hat{\gamma}) - m(z_g;\gamma^*)] - M(z_g, \hat{\gamma} - \gamma^*) = o_p(G^{-1/2}) \tag{27}$$

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

$$\begin{aligned} & G^{-1}\textstyle\sum_g M(z_g, \hat{\gamma} - \gamma^*) - \displaystyle\int M(z, \hat{\gamma} - \gamma^*) dF^*(z) \\ = \; & G^{-1}\textstyle\sum_g M(z_g, \hat{\gamma} - \bar{\gamma}) - \displaystyle\int M(z, \hat{\gamma} - \bar{\gamma}) dF^*(z) \\ & + G^{-1}\textstyle\sum_g M(z_g, \bar{\gamma} - \gamma^*) - \displaystyle\int M(z, \bar{\gamma} - \gamma^*) dF^*(z) \end{aligned} \tag{28}$$

Under *T1*, there exists a function c such that $\|M(z,\gamma)\| \leq c(z) \sup_{x\in\bar{\omega}} \|\gamma\|$ 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 (28) is $o_p(G^{-1/2})$. Next, note by the Chebyshev's inequality,

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

But by construction, $E[||M(z;\bar{\gamma} - \gamma^*)||^2] \leq c(z) \sup_{x\in\bar{\omega}} ||\bar{\gamma} - \gamma^*||^2$. 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 (28) is also $o_p(G^{-1/2})$. Therefore,

$$G^{-1}\textstyle\sum_g M(z_g, \hat{\gamma} - \gamma^*) - \displaystyle\int M(z, \hat{\gamma} - \gamma^*) dF^*(z) = o_p(G^{-1/2}) \tag{29}$$

This and (27) together proves (25) in the lemma. *Q.E.D.*

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

Proof of Lemma B.4.2. By construction,

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

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

$$= G^{-1}\sum_g \int v(x)y_g K_\sigma(x-x_g)\,dx - E[v(X)Y] = G^{-1}\sum_g \int \delta(x,d_{g,1},d_{g,2})K_\sigma(x-x_g)dx \quad (30)$$

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 (30) can be written as $\int M(z,\hat{\gamma}-\gamma^*)dF^*$ $= \int \delta(z)d\hat{F}$ with δ defined in the statement of the lemma.

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:

$$\begin{aligned}
&\int \delta(z)d\hat{F} - G^{-1}\sum_g \delta(z_g) \\
=G^{-1} \quad &\sum_g \left(\int v(x)y_g K_\sigma(x-x_g) - v(x)y_g K_\sigma^*(x-x_g)dx\right) \\
+G^{-1} \quad &\sum_g \left(\int v(x)y_g K_\sigma^*(x-x_g)dx - v(x_g)y_g\right) \qquad (31)
\end{aligned}$$

By the definition of $v(x)$ and properties of K and σ in *T1-4*, the first term on the right-hand side of (31) 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 the unconditional support of X. (See Theorem 8.11 in Newey and McFadden (1994) for details.) *Q.E.D.*

By Lemma B.4.1 and Lemma B.4.2,

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

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

$$\begin{aligned}
\Sigma_{\bar{\omega}} &\equiv Var\left[1\{X\in\bar{\omega}\}m(Z;\gamma^*) + \delta(Z)\right] \\
&= Var\left[1\{X\in\bar{\omega}\}\left(D_1D_2 - p_1^*(X)p_2^*(X) + \frac{2p_1^*(X)p_2^*(X) - D_1p_2^*(X) - D_2p_1^*(X)}{\gamma_0^*(X)}\right)\right]
\end{aligned}$$

This proves completes the derivation of the limiting distribution in Section 7.2 of Lewbel and Tang (2014).

C. Proof of Theorem 3

For all i and $\tilde{x}$,

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

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

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

$$= \int_{\Omega_{\epsilon_i|\tilde{x}}} \left(\int_{\Omega_{V_i|\tilde{x}}} \left[1\{v_i \geq s_i\} - 1\{v_i > c^*\}\right] dv_i \right) dF_{\epsilon_i|\tilde{x}}\left(\varepsilon_i\right) \tag{32}$$

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 the positive density in A5 respectively. The fourth follows from the independence between ϵ_i and X_e given $\tilde{x}$, which implies $F_{\epsilon_i|V_i=v_i,\tilde{X}=\tilde{x}} = F_{\epsilon_i|\tilde{X}=\tilde{x}}$ for all $v_i, \tilde{x}$. The last equality follows from changing the order of integration due to the Fubini's Theorem and the support condition in A4. The last expression on the right-hand side of (32) can be written as

$$\int_{\Omega_{\epsilon_i|\tilde{x}}} \left(\int_{\Omega_{V_i|\tilde{x}}} \left[1\{s_i \leq v_i < c^*\}1\{s_i \leq c^*\} - 1\{s_i \geq v_i > c^*\}1\{s_i > c^*\}\right] dv_i \right) dF_{\epsilon_i|\tilde{x}}\left(\varepsilon_i\right)$$

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

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

where the second equality uses the large support condition in A4. Hence $E\left(Y_i^* \mid \tilde{X} = \tilde{x}\right) - c^* = \tilde{u}_i(\tilde{x})$. *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] Gale, D., Nikaido, H., 1965. The Jacobian matrix and global univalence of mappings. Mathematische Annalen 159, 81–93.

[4] Lewbel, A. and X. Tang, “Identification and Estimation of Games with Incomplete Information Using Excluded Regressors," unpublished manuscript, 2014

[5] 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

[6] Volgushev, S., “Smoothed Quantile Regression Processes for Binary Response Models,” Arxiv paper.