

# Supplement to: “Estimation of Games with Ordered Actions: An Application to Chain-Store Entry”

Andrés Aradillas-López and Amit Gandhi

February 24, 2016

## Abstract

This document includes supplementary material for the paper “Estimation of Games with Ordered Actions: An Application to Chain-Store Entry”. It includes a step-by-step proof of Theorem 2, an analysis of the uniform asymptotic properties of our confidence sets (CS), a description of the kernels and bandwidths used in our empirical application, a detailed discussion on identification and nontrivial CS, a discussion of identification when there exists correlation across players’ unobserved payoff shocks (i.e, when our assumption of Independent Private Shocks fails), and additional Monte Carlo experiment results when our assumptions are violated.

## Labeling Conventions

Every result, equation, assumption, table, etc., introduced in this supplement will be labeled starting with an ‘S’. Specifically, equations will be labeled (S.1), (S.2), etc. Every equation referenced here that is not of that type refers to an equation in the main paper. Sections in this supplement will be labeled S-A, S-B, and so on. Sections referenced here that are not of that format refer to sections in the main paper. Similarly, all assumptions, claims, propositions, theorems and results introduced here will be labeled S1, S2, S3, etc. Any other labeling refers to results, equations, tables, assumptions, etc. in the main paper.

## S-A Step-by-Step Proof of Theorem 2

Here we present a detailed proof of Theorem 2, which can be found in Appendix B of the paper. Throughout the proof we will refer to the assumptions of the theorem (Assumptions B1, B2, B3 and B4), which are described in detail in Appendix B of the paper.

In Assumption B1 we described  $\mathcal{W}$  as

$$\mathcal{W} = \left\{ (x, y) \in \text{Supp}(X, Y) : x \in \mathcal{X} \right\},$$

where  $\mathcal{X} \subset \text{Supp}(X)$  is a prespecified set such that  $\mathcal{X} \cap \text{Supp}(X^c) \subset \text{int}(\text{Supp}(X^c))$ . We maintain the assumption that  $f_X(x) \geq \underline{f} > 0$  for all  $x \in \mathcal{X}$ . We will split the proof in three steps.

## Step 1

Our first step is to show that under our assumptions, there exist  $D_1 > 0$ ,  $D_2 > 0$  and  $D_3 > 0$  such that

$$\begin{aligned} & Pr \left( \sup_{(x, y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \widehat{\tau}^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p) \right| \geq b_n \right) \\ & \leq D_1 \exp \left\{ -\sqrt{n} h_n^q (D_2 \cdot b_n - D_3 \cdot h_n^M) \right\}. \end{aligned}$$

Fix  $y^p$ ,  $x$  and  $\theta^p$  and let  $Q_{F_{Y^p}}^p(y^p|x)$ ,  $Q_{\lambda^p}(x; \theta^p)$  and  $Q_{\mu^p}(y^p|x; \theta^p)$  be as defined in Assumption B1. We estimate these functionals with

$$\begin{aligned} \widehat{Q}_{F_{Y^p}}(y^p|x) &= (nh_n^q)^{-1} \sum_{i=1}^n \mathbb{1}\{Y_i^p \leq y^p\} \cdot \mathcal{H}(X_i - x; h_n), \\ \widehat{Q}_{\lambda^p}(x; \theta^p) &= (nh_n^q)^{-1} \sum_{i=1}^n \eta^p(Y_i^{-p}; x|\theta^p) \cdot \mathcal{H}(X_i - x; h_n), \\ \widehat{Q}_{\mu^p}(y^p|x; \theta^p) &= (nh_n^q)^{-1} \sum_{i=1}^n \mathbb{1}\{Y_i^p \leq y^p\} \cdot \eta^p(Y_i^{-p}; x|\theta^p) \cdot \mathcal{H}(X_i - x; h_n). \end{aligned}$$

Using an  $M^{\text{th}}$  order approximation, our smoothness restrictions in Assumption B1 imply the existence of a finite constant  $\overline{M}$  such that,

$$\begin{aligned} \sup_{x \in \mathcal{X}} \left| E \left[ \widehat{f}_X(x) \right] - f_X(x) \right| &\leq \overline{M} \cdot h_n^M, \\ \sup_{(x, y^p) \in \mathcal{W}} \left| E \left[ \widehat{Q}_F^p(y^p|x) \right] - Q_{F_{Y^p}}(y^p|x) \right| &\leq \overline{M} \cdot h_n^M, \\ \sup_{x \in \mathcal{X}, \theta^p \in \Theta} \left| E \left[ \widehat{Q}_{\lambda^p}(x; \theta^p) \right] - Q_{\lambda^p}(x; \theta^p) \right| &\leq \overline{M} \cdot h_n^M, \\ \sup_{\substack{(x, y^p) \in \mathcal{W} \\ \theta^p \in \Theta}} \left| E \left[ \widehat{Q}_{\mu^p}(y^p|x; \theta^p) \right] - Q_{\mu^p}(y^p|x; \theta^p) \right| &\leq \overline{M} \cdot h_n^M. \end{aligned} \tag{S.1}$$

Invoking Lemma 22 in Nolan and Pollard (1987) and Lemmas 2.4 and 2.14 in Pakes and Pollard (1989), having a kernel of bounded variation implies that the class of functions

$$\mathcal{G} = \left\{ g : g(x) = \mathcal{H}(x - v; h) \text{ for some } v \in \mathbb{R}^{\dim(X)} \text{ and some } h > 0 \right\}$$

is *Euclidean*<sup>1</sup> with respect to the constant envelope  $\overline{K}$ . Lemma 2.4 in Pakes and Pollard (1989) also implies that the class of functions

$$\mathcal{G} = \left\{ g : g(y^p) = \mathbb{1}\{y^p \leq v\} \text{ for some } v \in \mathbb{R} \right\}$$

<sup>1</sup>See Definition 2.7 in Pakes and Pollard (1989).

is Euclidean with respect to the envelope 1. Combined with Assumption B4(i) and Lemma 2.14 in Pakes and Pollard (1989) we have that the classes of functions

$$\begin{aligned}\mathcal{F}_1 &= \{f : f(y^{-p}, x) = \eta^p(y^{-p}; u|\theta^p) \cdot \mathcal{H}(x - u; h) \text{ for some } u \in \mathcal{X} \text{ and } \theta^p \in \Theta\}, \\ \mathcal{F}_2 &= \{f : f(y, x) = \mathbb{1}\{y^p \leq v\} \cdot \eta^p(y^{-p}; u|\theta^p) \cdot \mathcal{H}(x - u; h) \text{ for some } v \in \mathbb{R}, u \in \mathcal{X} \text{ and } \theta^p \in \Theta\}\end{aligned}$$

are Euclidean with respect to the envelope  $\bar{K} \cdot \bar{\eta}^p(\cdot)$ . Since this envelope has a moment generating function by Assumption B4(i), the maximal inequality results in Chapter 7 of Pollard (1990) combined with the bias conditions in S.1 imply that there exist positive constants  $A_1$ ,  $A_2$  and  $A_3$  such that for any  $\delta > 0$ ,

$$\begin{aligned}Pr \left( \sup_{x \in \mathcal{X}} \left| \widehat{f}_X(x) - f_X(x) \right| \geq \delta \right) &\leq A_1 \cdot \exp \left\{ -(\sqrt{n} \cdot h_n^q (A_2 \cdot \delta - A_3 \cdot h_n^M))^2 \right\}, \\ Pr \left( \sup_{(x, y^p) \in \mathcal{W}} \left| \widehat{Q}_{F_{Y^p}}(y^p|x) - Q_{F_{Y^p}}(y^p|x) \right| \geq \delta \right) &\leq A_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^q (A_2 \cdot \delta - A_3 \cdot h_n^M) \right\}, \\ Pr \left( \sup_{x \in \mathcal{X}, \theta^p \in \Theta} \left| \widehat{Q}_{\lambda^p}(x; \theta^p) - Q_{\lambda^p}(x; \theta^p) \right| \geq \delta \right) &\leq A_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^q (A_2 \cdot \delta - A_3 \cdot h_n^M) \right\}, \\ Pr \left( \sup_{\substack{(x, y^p) \in \mathcal{W} \\ \theta^p \in \Theta}} \left| \widehat{Q}_{\mu^p}(y^p|x; \theta^p) - Q_{\mu^p}(y^p|x; \theta^p) \right| \geq \delta \right) &\leq A_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^q (A_2 \cdot \delta - A_3 \cdot h_n^M) \right\}.\end{aligned}\tag{S.2}$$

For any  $x$  such that  $f_X(x) > 0$  define

$$\begin{aligned}\psi_{F_{Y^p}}(Y_i^p, X_i, y^p, x; h) &= \frac{(\mathbb{1}\{Y_i^p \leq y^p\} - F_{Y^p}(y^p|x))}{f_X(x)} \cdot \mathcal{H}(X_i - x; h), \\ \psi_{\lambda^p}(Y_i^{-p}, X_i, x, \theta^p; h) &= \frac{(\eta^p(Y_i^{-p}; x|\theta^p) - \lambda^p(x; \theta^p))}{f_X(x)} \cdot \mathcal{H}(X_i - x; h), \\ \psi_{\mu^p}(Y_i, X_i, y^p, x, \theta^p; h) &= \frac{(\mathbb{1}\{Y_i^p \leq y^p\} \cdot \eta^p(Y_i^{-p}; x|\theta^p) - \mu^p(y^p|x; \theta^p))}{f_X(x)} \cdot \mathcal{H}(X_i - x; h).\end{aligned}\tag{S.3}$$

And let

$$\begin{aligned}\widehat{\zeta}_{F_{Y^p}}(y^p, x) &= \left( \left[ \widehat{Q}_{F_{Y^p}}(y^p|x) - Q_{F_{Y^p}}(y^p|x) \right] \left[ \widehat{f}_X(x) - f_X(x) \right] \right)', \\ \widehat{\zeta}_{\lambda^p}(x, \theta^p) &= \left( \left[ \widehat{Q}_{\lambda^p}(x; \theta^p) - Q_{\lambda^p}(x; \theta^p) \right] \left[ \widehat{f}_X(x) - f_X(x) \right] \right)', \\ \widehat{\zeta}_{\mu^p}(y^p, x, \theta^p) &= \left( \left[ \widehat{Q}_{\mu^p}(y^p|x; \theta^p) - Q_{\mu^p}(y^p|x; \theta^p) \right] \left[ \widehat{f}_X(x) - f_X(x) \right] \right)'.\end{aligned}$$

Note that (S.2) implies that for any  $\delta > 0$ ,

$$\begin{aligned}
& Pr \left( \sup_{(x,y^p) \in \mathcal{W}} \left| \widehat{\zeta}_{F_{Y^p}}(y^p, x) \right| \geq \delta \right) \leq Pr \left( \sup_{(x,y^p) \in \mathcal{W}} \left| \widehat{Q}_{F_{Y^p}}(y^p|x) - Q_{F_{Y^p}}(y^p|x) \right| \geq \frac{\delta}{\sqrt{2}} \right) \\
& + Pr \left( \sup_{x \in \mathcal{X}} \left| \widehat{f}_X(x) - f_X(x) \right| \geq \frac{\delta}{\sqrt{2}} \right) \\
& \leq A_1 \cdot \exp \left\{ - \left( \sqrt{n} \cdot h_n^q \left( A_2 \cdot \frac{\delta}{\sqrt{2}} - A_3 \cdot h_n^M \right) \right)^2 \right\} + A_1 \cdot \exp \left\{ - \sqrt{n} \cdot h_n^q \left( A_2 \cdot \frac{\delta}{\sqrt{2}} - A_3 \cdot h_n^M \right) \right\} \\
& \leq 2 \cdot A_1 \cdot \exp \left\{ - \sqrt{n} \cdot h_n^q \left( A_2 \cdot \frac{\delta}{\sqrt{2}} - A_3 \cdot h_n^M \right) \right\}
\end{aligned}$$

Similarly (S.2) yields

$$\begin{aligned}
& Pr \left( \sup_{x \in \mathcal{X}, \theta^p \in \Theta} \left| \widehat{\zeta}_{\lambda^p}(x, \theta^p) \right| \geq \delta \right) \leq 2 \cdot A_1 \cdot \exp \left\{ - \sqrt{n} \cdot h_n^q \left( A_2 \cdot \frac{\delta}{\sqrt{2}} - A_3 \cdot h_n^M \right) \right\}, \\
& Pr \left( \sup_{\substack{(x,y^p) \in \mathcal{W} \\ \theta^p \in \Theta}} \left| \widehat{\zeta}_{\mu^p}(y^p, x, \theta^p) \right| \geq \delta \right) \leq 2 \cdot A_1 \cdot \exp \left\{ - \sqrt{n} \cdot h_n^q \left( A_2 \cdot \frac{\delta}{\sqrt{2}} - A_3 \cdot h_n^M \right) \right\}.
\end{aligned}$$

Whenever  $\widehat{f}_X(x) > 0$  and  $f_X(x) > 0$ , a second order approximation yields the following results:

$$\begin{aligned}
& \widehat{F}_{Y^p}(y^p|x) - F_{Y^p}(y^p|x) = \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{F_{Y^p}}(Y_i^p, X_i, y^p, x; h_n) + \xi_n^{F_{Y^p}}(y^p, x), \\
& \text{where } \xi_n^{F_{Y^p}}(y^p, x) = \frac{1}{2} \widehat{\zeta}_{F_{Y^p}}(y^p, x)' \begin{pmatrix} 0 & -\frac{1}{\widehat{f}_X^2(x)} \\ -\frac{1}{\widehat{f}_X^2(x)} & \frac{2\widetilde{Q}_{F_{Y^p}}(y^p|x)}{\widehat{f}_X^3(x)} \end{pmatrix} \widehat{\zeta}_{F_{Y^p}}(y^p, x)
\end{aligned}$$

where  $(\widetilde{f}_X(x), \widetilde{Q}_{F_{Y^p}}(y^p|x))$  belongs in the line segment connecting  $(\widehat{f}_X(x), \widehat{Q}_{F_{Y^p}}(y^p|x))$  and  $(f_X(x), Q_{F_{Y^p}}(y^p|x))$ .

$$\begin{aligned}
& \widehat{\lambda}^p(x; \theta^p) - \lambda^p(x; \theta^p) = \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{\lambda^p}(Y_i^{-p}, X_i, x, \theta^p; h_n) + \xi_n^{\lambda^p}(x, \theta^p), \\
& \text{where } \xi_n^{\lambda^p}(x, \theta^p) = \frac{1}{2} \widehat{\zeta}_{\lambda^p}(x, \theta^p)' \begin{pmatrix} 0 & -\frac{1}{\widehat{f}_X^2(x)} \\ -\frac{1}{\widehat{f}_X^2(x)} & \frac{2\widetilde{Q}_{\lambda^p}(x; \theta^p)}{\widehat{f}_X^3(x)} \end{pmatrix} \widehat{\zeta}_{\lambda^p}(x, \theta^p)
\end{aligned}$$

where  $(\widetilde{f}_X(x), \widetilde{Q}_{\lambda^p}(x; \theta^p))$  belongs in the line segment connecting  $(\widehat{f}_X(x), \widehat{Q}_{\lambda^p}(x; \theta^p))$  and  $(f_X(x), Q_{\lambda^p}(x; \theta^p))$ .

$$\begin{aligned}
& \widehat{\mu}^p(y^p|x; \theta^p) - \mu^p(y^p|x; \theta^p) = \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{\mu^p}(Y_i, X_i, y^p, x, \theta^p; h_n) + \xi_n^{\mu^p}(y^p, x, \theta^p), \\
& \text{where } \xi_n^{\mu^p}(y^p, x, \theta^p) = \frac{1}{2} \widehat{\zeta}_{\mu^p}(y^p, x, \theta^p)' \begin{pmatrix} 0 & -\frac{1}{\widehat{f}_X^2(x)} \\ -\frac{1}{\widehat{f}_X^2(x)} & \frac{2\widetilde{Q}_{\mu^p}(y^p|x; \theta^p)}{\widehat{f}_X^3(x)} \end{pmatrix} \widehat{\zeta}_{\mu^p}(y^p, x, \theta^p)
\end{aligned}$$

where  $(\ddot{f}_X(x), \ddot{Q}_{\mu^p}(y^p|x; \theta^p))$  belongs in the line segment connecting  $(\widehat{f}_X(x), \widehat{Q}_{\mu^p}(y^p|x; \theta^p))$  and  $(f_X(x), Q_{\mu^p}(y^p|x; \theta^p))$ . Let  $\overline{Q}$  be as described in Assumption B1. For any  $0 < \underline{f}^* < \underline{f}$ , define

$$D(\underline{f}^*) = \left\| \begin{array}{cc} 0 & -\frac{1}{(\underline{f}^*)^2} \\ -\frac{1}{(\underline{f}^*)^2} & \frac{3\overline{Q}}{(\underline{f}^*)^3} \end{array} \right\|. \quad (\text{S.4})$$

Let  $0 < \underline{f}^* < \underline{f}$  and  $D(\underline{f}^*)$  be as described in (S.4). Combining our previous results, for any  $\delta > 0$ ,

$$\begin{aligned} & Pr \left( \sup_{(x, y^p) \in \mathcal{W}} |\xi_n^{F_{Y^p}}(y^p, x)| \geq \delta \right) \leq Pr \left( \sup_{(x, y^p) \in \mathcal{W}} \left| \widehat{Q}_{F_{Y^p}}(y^p|x) - Q_{F_{Y^p}}(y^p|x) \right| \geq \overline{Q} \right) \\ & + Pr \left( \sup_{x \in \mathcal{X}} \left| \widehat{f}_X(x) - f_X(x) \right| \geq \underline{f} - \underline{f}^* \right) + Pr \left( \sup_{(x, y^p) \in \mathcal{W}} \left| \widehat{\zeta}_{F_{Y^p}}(y^p, x) \right| \geq \sqrt{\frac{2\delta}{D(\underline{f}^*)}} \right) \\ & \leq 4A_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^q \left( A_2 \cdot \min \left\{ \sqrt{\frac{\delta}{D(\underline{f}^*)}}, \overline{Q}, \underline{f} - \underline{f}^* \right\} - A_3 \cdot h_n^M \right) \right\}. \end{aligned}$$

And the same bound holds for

$$Pr \left( \sup_{x \in \mathcal{X}, \theta^p \in \Theta} \left| \xi_n^{\lambda^p}(x, \theta^p) \right| \geq \delta \right) \quad \text{and} \quad Pr \left( \sup_{\substack{(x, y^p) \in \mathcal{W} \\ \theta^p \in \Theta}} \left| \xi_n^{\mu^p}(y^p, x, \theta^p) \right| \geq \delta \right).$$

Assumption B4 and Lemma 2.14 in Pakes and Pollard (1989) we have that the classes of functions

$$\begin{aligned} \mathcal{G}_1 &= \{g: g(y^p, x) = \psi_{F_{Y^p}}(y^p, x, v^p, u; h): (v^p, u) \in \mathcal{W}, h > 0\}, \\ \mathcal{G}_2 &= \{g: g(y^{-p}, x) = \psi_{\lambda^p}(y^{-p}, x, u, \theta^p): u \in \mathcal{X}, \theta^p \in \Theta, h > 0\}, \\ \mathcal{G}_3 &= \{g: g(y, x) = \psi_{\mu^p}(y, x, v^p, u, \theta^p; h): (v^p, u) \in \mathcal{W}, \theta^p \in \Theta, h > 0\} \end{aligned}$$

are Euclidean with respect to envelopes  $\frac{2\overline{K}}{\underline{f}}$ ,  $\frac{2\overline{K}\overline{\eta}^p(\cdot)}{\underline{f}}$  and  $\frac{2\overline{K}\overline{\eta}^p(\cdot)}{\underline{f}}$ , respectively. The existence of moments feature of  $\overline{\eta}^p(\cdot)$  in Assumption B4 and the results in Chapter 7 of Pollard (1990) combined with the bias conditions in S.1 imply that there exist positive constants  $A'_1$ ,  $A'_2$  and  $A'_3$  such that for any  $\delta > 0$ , the probabilities

$$\begin{aligned} & Pr \left( \sup_{(x, y^p) \in \mathcal{W}} \left| \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{F_{Y^p}}(Y_i^p, X_i, y^p, x; h_n) \right| \geq \delta \right), \\ & Pr \left( \sup_{x \in \mathcal{X}, \theta^p \in \Theta} \left| \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{\lambda^p}(Y_i^{-p}, X_i, x, \theta^p; h_n) \right| \geq \delta \right), \\ & Pr \left( \sup_{(x, y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{\mu^p}(Y_i, X_i, y^p, x, \theta^p; h_n) \right| \geq \delta \right), \end{aligned}$$

are bounded above by

$$A'_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^q \left( A'_2 \cdot \delta - A'_3 \cdot h_n^M \right) \right\}.$$

Let  $0 < \underline{f}^* < \underline{f}$  and  $D(\underline{f}^*)$  be as described in (S.4). Combining our results, for any  $\delta > 0$  we have

$$\begin{aligned}
& Pr \left( \sup_{(x, y^p) \in \mathcal{W}} \left| \widehat{F}_{Y^p}(y^p|x) - F_{Y^p}(y^p|x) \right| \geq \delta \right) \leq Pr \left( \sup_{x \in \mathcal{X}} \left| \widehat{f}_X(x) - f_X(x) \right| \geq \underline{f} - \underline{f}^* \right) \\
& + Pr \left( \sup_{(x, y^p) \in \mathcal{W}} \left| \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{F_{Y^p}}(Y_i^p, X_i, y^p, x; h_n) \right| \geq \frac{\delta}{2} \right) \\
& + Pr \left( \sup_{(x, y^p) \in \mathcal{W}} \left| \xi_n^{F_{Y^p}}(y^p, x) \right| \geq \frac{\delta}{2} \right) \\
& \leq A_1 \cdot \exp \left\{ -(\sqrt{n} \cdot h_n^q (A_2 \cdot (\underline{f} - \underline{f}^*) - A_3 \cdot h_n^M))^2 \right\} \\
& + A'_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^q \left( A'_2 \cdot \frac{\delta}{2} - A'_3 \cdot h_n^M \right) \right\} \\
& + 4A_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^q \left( A_2 \cdot \min \left\{ \sqrt{\frac{\delta}{2D}}, \overline{Q}, \underline{f} - \underline{f}^* \right\} - A_3 \cdot h_n^M \right) \right\} \\
& \leq B_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^q \left( B_2 \cdot \min \left\{ \frac{\delta}{2}, \sqrt{\frac{\delta}{2D}}, \overline{Q}, \underline{f} - \underline{f}^* \right\} - B_3 \cdot h_n^M \right) \right\}
\end{aligned}$$

where  $B_1 = 6 \cdot \max \{A_1, A'_1\}$ ,  $B_2 = \min \{A_2, A'_2\}$  and  $B_3 = \max \{A_3, A'_3\}$ . The same type of bound is valid for

$$\begin{aligned}
& Pr \left( \sup_{x \in \mathcal{X}, \theta^p \in \Theta} \left| \widehat{\lambda}^p(x; \theta^p) - \lambda^p(x; \theta^p) \right| \geq \delta \right), \\
& Pr \left( \sup_{(x, y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \widehat{\mu}^p(y^p|x; \theta^p) - \mu^p(y^p|x; \theta^p) \right| \geq \delta \right).
\end{aligned}$$

The previous results allow us now to turn our attention to  $\widehat{\tau}^p(y^p|x; \theta^p)$ . For  $h > 0$  let

$$\begin{aligned}
& \psi_{\tau^p}(Y_i, X_i, y^p, x, \theta^p; h) \\
& = \lambda^p(x; \theta^p) \cdot \psi_{F_{Y^p}}(Y_i^p, X_i, y^p, x; h) + F_{Y^p}(y^p|x) \cdot \psi_{\lambda^p}(Y_i^p, X_i, x, \theta^p; h) - \psi_{\mu^p}(Y_i, X_i, y^p, x, \theta^p; h) \\
& = \left[ \lambda^p(x; \theta^p) \cdot (\mathbb{1}\{Y_i^p \leq y^p\} - F_{Y^p}(y^p|x)) + F_{Y^p}(y^p|x) \cdot (\eta^p(Y_i^p; x|\theta^p) - \lambda^p(x; \theta^p)) \right. \\
& \quad \left. - (\mathbb{1}\{Y_i^p \leq y^p\} \cdot \eta^p(Y_i^p; x|\theta^p) - \mu^p(y^p|x; \theta^p)) \right] \cdot \frac{\mathcal{H}(X_i - x; h)}{f_X(x)}
\end{aligned} \tag{S.5}$$

From our previous results we have

$$\widehat{\tau}^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p) = \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{\tau^p}(Y_i, X_i, y^p, x, \theta^p; h_n) + \xi_n^{\tau^p}(y^p, x, \theta^p), \tag{S.6}$$

where

$$\begin{aligned}
\xi_n^{\tau^p}(y^p, x, \theta^p) & = \lambda^p(x; \theta^p) \cdot \xi_n^{F_{Y^p}}(y^p, x) + F_{Y^p}(y^p|x) \cdot \xi_n^{\lambda^p}(x, \theta^p) - \xi_n^{\mu^p}(y^p, x, \theta^p) \\
& + \left( \widehat{F}_{Y^p}(y^p|x) - F_{Y^p}(y^p|x) \right) \cdot \left( \widehat{\lambda}^p(x; \theta^p) - \lambda^p(x; \theta^p) \right).
\end{aligned}$$

Let

$$\sup_{x \in \mathcal{X}, \theta^p \in \Theta} |\lambda^p(x; \theta^p)| = \bar{\lambda}^p.$$

For any  $\delta > 0$ ,

$$\begin{aligned} & Pr \left( \sup_{(x, y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \xi_n^{\tau^p}(y^p, x, \theta^p) \right| \geq \delta \right) \leq Pr \left( \sup_{(x, y^p) \in \mathcal{W}} \left| \xi_n^{F_{Y^p}}(y^p, x) \right| \geq \frac{\delta}{4\bar{\lambda}^p} \right) \\ & + Pr \left( \sup_{(x, y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \xi_n^{\lambda^p}(x, \theta^p) \right| \geq \frac{\delta}{4} \right) + Pr \left( \sup_{(x, y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \xi_n^{\mu^p}(y^p, x, \theta^p) \right| \geq \frac{\delta}{4} \right) \\ & + Pr \left( \sup_{(x, y^p) \in \mathcal{W}} \left| \widehat{F}_{Y^p}(y^p|x) - F_{Y^p}(y^p|x) \right| \geq \frac{\sqrt{\delta}}{2} \right) \\ & + Pr \left( \sup_{x \in \mathcal{X}, \theta^p \in \Theta} \left| \widehat{\lambda}^p(x; \theta^p) - \lambda^p(x; \theta^p) \right| \geq \frac{\sqrt{\delta}}{2} \right) \end{aligned}$$

Let  $0 < \underline{f}^* < \underline{f}$  and  $D(\underline{f}^*)$  be as described in (S.4), the previous expression is bounded above by

$$\begin{aligned} & 4A_1 \exp \left\{ -\sqrt{n}h_n^q \left( A_2 \min \left\{ \frac{1}{2} \sqrt{\frac{\delta}{D(\underline{f}^*)\bar{\lambda}^p}}, \bar{Q}, \underline{f} - \underline{f}^* \right\} - A_3 \cdot h_n^M \right) \right\} \\ & + 8A_1 \exp \left\{ -\sqrt{n}h_n^q \left( A_2 \min \left\{ \frac{1}{2} \sqrt{\frac{\delta}{D(\underline{f}^*)}}, \bar{Q}, \underline{f} - \underline{f}^* \right\} - A_3 \cdot h_n^M \right) \right\} \\ & + 2B_1 \exp \left\{ -\sqrt{n}h_n^q \left( B_2 \min \left\{ \frac{1}{2} \sqrt{\delta}, \frac{1}{2} \frac{\delta^{1/4}}{\sqrt{D(\underline{f}^*)}}, \bar{Q}, \underline{f} - \underline{f}^* \right\} - B_3 h_n^M \right) \right\} \end{aligned}$$

Let  $\underline{B} = \frac{1}{2} \cdot \min \left\{ \frac{1}{\sqrt{D\bar{\lambda}^p}}, \frac{1}{\sqrt{D}}, 1, 2\bar{Q}, 2(\underline{f} - \underline{f}^*) \right\}$  and define  $C_1 \equiv 4 \cdot B_1$ ,  $C_2 \equiv B_2 \cdot \underline{B}$ ,  $C_3 \equiv B_3$ . We have

$$\begin{aligned} & Pr \left( \sup_{(x, y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \xi_n^{\tau^p}(y^p, x, \theta^p) \right| \geq \delta \right) \\ & \leq C_1 \exp \left\{ -\sqrt{n}h_n^q \left( C_2 \cdot \min \left\{ \delta^{1/2}, \delta^{1/4}, 1 \right\} - C_3 \cdot h_n^M \right) \right\} \end{aligned}$$

By Assumption B4 and Lemma 2.14 in Pakes and Pollard (1989), the class of functions

$$\mathcal{G}_4 = \{g: g(y, x) = \psi_{\tau^p}(y, x, v^p, u, \theta^p; h) : (v^p, u) \in \mathcal{W}, \theta^p \in \Theta, h > 0\}$$

is Euclidean with respect to the envelope  $\frac{2\bar{\lambda}^p \bar{K}}{\underline{f}} + \frac{4\bar{K}\bar{\eta}^p(\cdot)}{\underline{f}}$ . The existence of moments feature of  $\bar{\eta}^p(\cdot)$  in Assumption B4 and the results in Chapter 7 of Pollard (1990) combined with the bias conditions in S.1 imply that there exist positive constants  $C'_1$ ,  $C'_2$  and  $C'_3$  such that for any  $\delta > 0$ ,

$$\begin{aligned} & Pr \left( \sup_{(x, y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{\tau^p}(Y_i, X_i, y^p, x, \theta^p; h_n) \right| \geq \delta \right) \\ & \leq C'_1 \cdot \exp \left\{ -\sqrt{n} \cdot h_n^q \left( C'_2 \cdot \delta - C'_3 \cdot h_n^M \right) \right\}. \end{aligned}$$

As before, if we let  $0 < \underline{f}^* < \underline{f}$  be as described in (S.4)

$$\begin{aligned}
& Pr \left( \sup_{(x,y^p) \in \mathcal{W}, \theta^p \in \Theta} |\widehat{\tau}^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p)| \geq \delta \right) \\
& \leq Pr \left( \sup_{x \in \mathcal{X}} \left| \widehat{f}_X(x) - f_X(x) \right| \geq \underline{f} - \underline{f}^* \right) \\
& + Pr \left( \sup_{(x,y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{\tau^p}(Y_i, X_i, y^p, x, \theta^p; h_n) \right| \geq \frac{\delta}{2} \right) \\
& + Pr \left( \sup_{(x,y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \xi_n^{\tau^p}(y^p, x, \theta^p) \right| \geq \frac{\delta}{2} \right).
\end{aligned}$$

From here, putting our results together we have that for any  $\delta > 0$ ,

$$\begin{aligned}
& Pr \left( \sup_{(x,y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \widehat{\tau}^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p) \right| \geq \delta \right) \\
& \leq D_1 \exp \left\{ -\sqrt{n} h_n^q \left( D_2 \cdot \min \left\{ \delta, \delta^{1/2}, \delta^{1/4}, 1 \right\} - D_3 \cdot h_n^M \right) \right\},
\end{aligned} \tag{S.7}$$

where  $D_1 = 3 \cdot \max \{A_1, C'_1, C_1\}$ ,  $D_2 = \frac{1}{2} \cdot \min \{C'_2, C_2, 2A_2(\underline{f} - \underline{f}^*)\}$ ,  $D_3 = \max \{A_3, C_3, C'_3\}$ . Our results also imply

$$\begin{aligned}
& \widehat{\tau}^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p) = \frac{1}{nh_n^q} \sum_{i=1}^n \psi_{\tau^p}(Y_i, X_i, y^p, x, \theta^p; h_n) + \xi_n^{\tau^p}(y^p, x, \theta^p), \\
& \text{where } \sup_{(x,y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \xi_n^{\tau^p}(y^p, x, \theta^p) \right| = O_p \left( \frac{\log(n)^2}{nh_n^q} \right), \\
& \text{and } \sup_{(x,y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \widehat{\tau}^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p) \right| = O_p \left( \frac{\log(n)}{\sqrt{nh_n^q}} \right).
\end{aligned} \tag{S.8}$$

Let  $b_n$  be the sequence used in our construction. For  $n$  large enough we have  $\min \{b_n, b_n^{1/2}, b_n^{1/4}, 1\} = b_n$  and therefore (S.7) yields

$$\begin{aligned}
& Pr \left( \sup_{(x,y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \widehat{\tau}^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p) \right| \geq b_n \right) \\
& \leq D_1 \exp \left\{ -\sqrt{n} h_n^q \left( D_2 \cdot b_n - D_3 \cdot h_n^M \right) \right\},
\end{aligned} \tag{S.9}$$

This concludes Step 1 of our proof.

## Step 2

Here we use the results from Step 1 to show that

$$\begin{aligned}
& \widehat{T}_X^p(\theta^p) = \frac{1}{n} \sum_{i=1}^n \widehat{\tau}^p(Y_i^p|X_i; \theta^p) \cdot \mathbb{1} \{ \tau^p(Y_i^p|X_i; \theta^p) \geq 0 \} \cdot \mathbb{1}_X(X_i) + \varphi_n^p(\theta^p), \\
& \text{where } \sup_{\theta^p \in \Theta} |\varphi_n^p(\theta^p)| = O_p \left( n^{-1/2-\epsilon} \right) \text{ for some } \epsilon > 0.
\end{aligned}$$



We begin by noting that we can express

$$\widehat{T}_{\mathcal{X}}^p(\theta^p) = \frac{1}{n} \sum_{i=1}^n \widehat{\tau}^p(Y_i^p | X_i; \theta^p) \cdot \mathbb{1} \{ \tau^p(Y_i^p | X_i; \theta^p) \geq 0 \} \cdot \mathbb{I}_{\mathcal{X}}(X_i) + \varphi_n^p(\theta^p),$$

where

$$\begin{aligned} |\varphi_n^p(\theta^p)| &\leq \underbrace{\left| \frac{1}{n} \sum_{i=1}^n \widehat{\tau}^p(Y_i^p | X_i; \theta^p) \cdot \mathbb{1} \{ -2b_n \leq \tau^p(Y_i^p | X_i; \theta^p) < 0 \} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \right|}_{\equiv |\varphi_n^{p,1}(\theta^p)|} \\ &\quad + \underbrace{\left| \frac{2}{n} \sum_{i=1}^n \widehat{\tau}^p(Y_i^p | X_i; \theta^p) \cdot \mathbb{1} \{ |\widehat{\tau}^p(Y_i^p | X_i; \theta^p) - \tau^p(Y_i^p | X_i; \theta^p)| \geq b_n \} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \right|}_{\equiv |\varphi_n^{p,2}(\theta^p)|}. \end{aligned}$$

We begin by examining  $\varphi_n^{p,2}$ . Using (S.8),  $\sup_{(x,y^p) \in \mathcal{W}, \theta^p \in \Theta} \widehat{\tau}^p(y^p | x; \theta^p) = O_p(1)$ . Therefore,

$$\sup_{\theta^p \in \Theta} |\varphi_n^{p,2}(\theta^p)| \leq O_p(1) \cdot \sup_{\theta^p \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1} \{ |\widehat{\tau}^p(Y_i^p | X_i; \theta^p) - \tau^p(Y_i^p | X_i; \theta^p)| \geq b_n \} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \right|$$

Take any  $\alpha > 0$  and any  $\varepsilon > 0$ . Then,

$$\begin{aligned} &Pr \left( n^\alpha \cdot \sup_{\theta^p \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1} \{ |\widehat{\tau}^p(Y_i^p | X_i; \theta^p) - \tau^p(Y_i^p | X_i; \theta^p)| \geq b_n \} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \right| > \varepsilon \right) \\ &\leq Pr \left( \mathbb{1} \left\{ \sup_{\theta^p \in \Theta} |\widehat{\tau}^p(Y_i^p | X_i; \theta^p) - \tau^p(Y_i^p | X_i; \theta^p)| \geq b_n \right\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \neq 0 \text{ for some } i = 1, \dots, n \right) \\ &\leq \sum_{i=1}^n Pr \left( \mathbb{1} \left\{ \sup_{\theta^p \in \Theta} |\widehat{\tau}^p(Y_i^p | X_i; \theta^p) - \tau^p(Y_i^p | X_i; \theta^p)| \geq b_n \right\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \neq 0 \right) \\ &\leq n \cdot Pr \left( \sup_{(x,y^p) \in \mathcal{W}, \theta^p \in \Theta} |\widehat{\tau}^p(y^p | x; \theta^p) - \tau^p(y^p | x; \theta^p)| \geq b_n \right) \\ &\leq n \cdot D_1 \exp \left\{ -\frac{1}{2} \sqrt{n} h_n^q (D_2 \cdot b_n - D_3 \cdot h_n^M) \right\} = D_1 \exp \left\{ -\frac{1}{2} \sqrt{n} h_n^q (D_2 \cdot b_n - D_3 \cdot h_n^M) + \log(n) \right\} \rightarrow 0 \end{aligned}$$

Therefore,  $\sup_{\theta^p \in \Theta} |\varphi_n^{p,2}(\theta^p)| = o_p(n^{-\alpha})$ . In particular, the following much weaker (but useful for our purposes) result holds,

$$\sup_{\theta^p \in \Theta} |\varphi_n^{p,2}(\theta^p)| = O_p \left( n^{-1/2-\epsilon} \right) \text{ for some } \epsilon > 0.$$

We move on to  $\varphi_n^{p,1}(\theta^p)$ . Note that

$$\widehat{\tau}^p(Y_i^p | X_i; \theta^p) = \sum_{j=0}^1 (\tau^p(Y_i^p | X_i; \theta^p))^{1-j} \cdot (\widehat{\tau}^p(Y_i^p | X_i; \theta^p) - \tau^p(Y_i^p | X_i; \theta^p))^j.$$

Therefore,

$$\begin{aligned}
& |\varphi_n^{p,1}(\theta^p)| \\
& \leq \frac{1}{n} \sum_{i=1}^n \left[ \sum_{j=0}^1 |\tau^p(Y_i^p|X_i; \theta^p)|^{1-j} \cdot |\widehat{\tau}^p(Y_i^p|X_i; \theta^p) - \tau^p(Y_i^p|X_i; \theta^p)|^j \right] \cdot \mathbb{1}\{-2b_n \leq \tau^p(Y_i^p|X_i; \theta^p) < 0\} \mathbb{I}_{\mathcal{X}}(X_i) \\
& \leq \frac{1}{n} \sum_{i=1}^n \left[ \sum_{j=0}^1 |2b_n|^{1-j} \cdot |\widehat{\tau}^p(Y_i^p|X_i; \theta^p) - \tau^p(Y_i^p|X_i; \theta^p)|^j \right] \cdot \mathbb{1}\{-2b_n \leq \tau^p(Y_i^p|X_i; \theta^p) < 0\} \mathbb{I}_{\mathcal{X}}(X_i).
\end{aligned}$$

Using (S.8) we have

$$\begin{aligned}
\sup_{(x, y^p) \in \mathcal{W}, \theta^p \in \Theta} \left| \sum_{j=0}^1 |2b_n|^{1-j} \cdot |\widehat{\tau}^p(y^p|x; \theta^p) - \tau^p(y^p|x; \theta^p)|^j \right| &= \sum_{j=0}^1 O(b_n^{1-j}) \cdot O_p \left( \left( \frac{\log(n)}{\sqrt{n}h_n^q} \right)^j \right) \\
&= O_p(b_n),
\end{aligned}$$

where the last equality follows from the bandwidth convergence restrictions in Assumption B2 since they imply that  $\frac{\log(n)}{\sqrt{n} \cdot h_n^q \cdot b_n} \rightarrow 0$ . Therefore,

$$\sup_{\theta^p \in \Theta} |\varphi_n^{p,1}(\theta^p)| \leq O_p(b_n) \cdot \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{-2b_n \leq \tau^p(Y_i^p|X_i; \theta^p) < 0\} \mathbb{I}_{\mathcal{X}}(X_i)$$

For a given  $b > 0$  denote

$$g_i^{p,1}(\theta^p, b) = \mathbb{1}\{-b \leq \tau^p(Y_i^p|X_i; \theta^p) < 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i).$$

And let

$$\nu_n^{p,1}(\theta^p) = \frac{1}{n} \sum_{i=1}^n \left( g_i^{p,1}(\theta^p, 2b_n) - E \left[ g_i^{p,1}(\theta^p, 2b_n) \right] \right).$$

Let  $A$  and  $\bar{b}$  be the constants described in Assumption B3. For large enough  $n$  we have  $2b_n \leq \bar{b}$  and therefore we can express

$$\frac{1}{n} \sum_{i=1}^n \mathbb{1}\{-2b_n \leq \tau^p(Y_i^p|X_i; \theta^p) < 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) = \nu_n^{p,1}(\theta^p) + \xi_n^{p,1}(\theta^p),$$

where

$$\sup_{\theta^p \in \Theta} |\xi_n^{p,1}(\theta^p)| = 2Ab_n = O(b_n) \quad \text{and} \quad \sup_{\theta^p \in \Theta} \text{Var}(\mathbb{1}\{-2b_n \leq \tau^p(Y_i^p|X_i; \theta^p) < 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i)) = O(b_n).$$

by Assumption B3. Using part (ii) of Assumption B4(ii),

$$\sup_{\theta^p \in \Theta} |\nu_n^{p,1}(\theta^p)| = O_p \left( \sqrt{\frac{b_n}{n}} \right) = O_p(b_n).$$

Combining these results, we have

$$\sup_{\theta^p \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{-2b_n \leq \tau^p(Y_i^p|X_i; \theta^p) < 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \right| = O_p(b_n).$$

And therefore

$$\sup_{\theta^p \in \Theta} |\varphi_n^{p,1}(\theta^p)| \leq O(b_n) \times O_p(b_n) = O_p(b_n^2) = O_p(n^{-1/2-\epsilon}) \quad \text{for some } \epsilon > 0.$$

Where the last line follows from the bandwidth convergence restrictions in Assumption B2. Combining the results for  $\varphi_n^{p,1}$  and  $\varphi_n^{p,2}$ ,

$$\widehat{T}_{\mathcal{X}}^p(\theta^p) = \frac{1}{n} \sum_{i=1}^n \widehat{\tau}^p(Y_i^p | X_i; \theta^p) \cdot \mathbb{1}\{\tau^p(Y_i^p | X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) + \varphi_n^p(\theta^p), \quad (\text{S.10})$$

$$\text{where } \sup_{\theta^p \in \Theta} |\varphi_n^p(\theta^p)| = O_p(n^{-1/2-\epsilon}) \quad \text{for some } \epsilon > 0.$$

### Step 3

This is the last step in the proof. We take the results from Step 2 to show that

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n (\widehat{\tau}^p(Y_i^p | X_i; \theta^p) - \tau^p(Y_i^p | X_i; \theta^p)) \cdot \mathbb{1}\{\tau^p(Y_i^p | X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \\ &= \frac{1}{n^2} \sum_{j \neq i} \sum_{i=1}^n g_{\tau^p}(X_i, Y_i, X_j, Y_j; \theta^p, h_n) + \varrho_n^{p,1}(\theta^p), \\ & \text{where } \sup_{\theta^p \in \Theta} |\varrho_n^{p,1}(\theta^p)| = O_p(n^{-1/2-\epsilon}) \quad \text{for some } \epsilon > 0. \end{aligned}$$

We then examine the Hoeffding decomposition of the U-statistic described above and, using our assumptions, we obtain the result in Theorem 2. We have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \widehat{\tau}^p(Y_i^p | X_i; \theta^p) \cdot \mathbb{1}\{\tau^p(Y_i^p | X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) = \\ & \frac{1}{n} \sum_{i=1}^n \max\{\tau^p(Y_i^p | X_i; \theta^p), 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \\ & + \frac{1}{n} \sum_{i=1}^n (\widehat{\tau}^p(Y_i^p | X_i; \theta^p) - \tau^p(Y_i^p | X_i; \theta^p)) \cdot \mathbb{1}\{\tau^p(Y_i^p | X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \end{aligned} \quad (\text{S.11})$$

Let  $\psi_{\tau^p}$  be as defined in (S.5). For any pair of observations  $i, j$  in  $1, \dots, n$  and  $h > 0$  let

$$g_{\tau^p}(X_i, Y_i, X_j, Y_j; \theta^p, h) = \frac{1}{h^q} \cdot \psi_{\tau^p}(Y_j, X_j, Y_i^p, X_i, \theta^p; h) \cdot \mathbb{1}\{\tau^p(Y_i^p | X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i). \quad (\text{S.12})$$

Note that

$$\sup_{\theta^p \in \Theta} \left| \frac{1}{n^2} \sum_{i=1}^n g_{\tau^p}(X_i, Y_i, X_i, Y_i; \theta^p, h_n) \right| = O_p\left(\frac{1}{nh_n^q}\right) = O_p(n^{-1/2-\epsilon}) \quad \text{for some } \epsilon > 0.$$

Combined with (S.8), this yields

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n (\hat{\tau}^p(Y_i^p|X_i; \theta^p) - \tau^p(Y_i^p|X_i; \theta^p)) \cdot \mathbb{1}\{\tau^p(Y_i^p|X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \\ &= \frac{1}{n^2} \sum_{j \neq i} \sum_{i=1}^n g_{\tau^p}(X_i, Y_i, X_j, Y_j; \theta^p, h_n) + \varrho_n^{p,1}(\theta^p), \end{aligned} \quad (\text{S.13})$$

$$\text{where } \sup_{\theta^p \in \Theta} |\varrho_n^{p,1}(\theta^p)| = O_p\left(\frac{\log(n)^2}{nh_n^q}\right) + O_p\left(\frac{1}{nh_n^q}\right) = O_p\left(n^{-1/2-\epsilon}\right) \text{ for some } \epsilon > 0.$$

We will examine the U-statistic in (S.13). Using (S.5) we can express

$$g_{\tau^p}(X_i, Y_i, X_j, Y_j; \theta^p, h) = g_{\tau^p}^a(X_i, Y_i, X_j, Y_j; \theta^p, h) + g_{\tau^p}^b(X_i, Y_i, X_j, Y_j; \theta^p, h) + g_{\tau^p}^c(X_i, Y_i, X_j, Y_j; \theta^p, h),$$

where

$$\begin{aligned} g_{\tau^p}^a(X_i, Y_i, X_j, Y_j; \theta^p, h) &= \\ & \frac{1}{h^q} \cdot \lambda^p(X_i; \theta^p) \cdot (\mathbb{1}\{Y_j^p \leq Y_i^p\} - F_{Y^p}(Y_i^p|X_i)) \cdot \mathbb{1}\{\tau^p(Y_i^p|X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \cdot \frac{\mathcal{H}(X_j - X_i; h)}{f_X(X_i)}, \\ g_{\tau^p}^b(X_i, Y_i, X_j, Y_j; \theta^p, h) &= \\ & \frac{1}{h^q} \cdot F_{Y^p}(Y_i^p|X_i) \cdot (\eta^p(Y_j^{-p}; X_i|\theta^p) - \lambda^p(X_i; \theta^p)) \cdot \mathbb{1}\{\tau^p(Y_i^p|X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \cdot \frac{\mathcal{H}(X_j - X_i; h)}{f_X(X_i)}, \\ g_{\tau^p}^c(X_i, Y_i, X_j, Y_j; \theta^p, h) &= \\ & \frac{1}{h^q} \cdot (\mathbb{1}\{Y_j^p \leq Y_i^p\} \cdot \eta^p(Y_j^{-p}; X_i|\theta^p) - \mu^p(Y_i^p|X_i; \theta^p)) \cdot \mathbb{1}\{\tau^p(Y_i^p|X_i; \theta^p) \geq 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) \cdot \frac{\mathcal{H}(X_j - X_i; h)}{f_X(X_i)}, \end{aligned}$$

Let  $\gamma_p^I$ ,  $\gamma_p^{II}$  and  $\gamma_p^{III}$  be as defined in Assumption B1. By the smoothness conditions in Assumption B1, there exists a  $C < \infty$  such that

$$\begin{aligned} \sup_{(x,y) \in \mathcal{W}, \theta^p \in \Theta} |E[g_{\tau^p}^a(x, y, X, Y; \theta^p, h)]| &\leq C \cdot h^M, \\ \sup_{(x,y) \in \mathcal{W}, \theta^p \in \Theta} |E[g_{\tau^p}^b(x, y, X, Y; \theta^p, h)]| &\leq C \cdot h^M, \\ \sup_{(x,y) \in \mathcal{W}, \theta^p \in \Theta} |E[g_{\tau^p}^c(x, y, X, Y; \theta^p, h)]| &\leq C \cdot h^M. \end{aligned}$$

And

$$\begin{aligned} E[g_{\tau^p}^a(X, Y, x, y; \theta^p, h)] &= (\gamma_p^I(y^p, x; \theta^p) - \gamma_p^{II}(x; \theta^p)) \cdot \mathbb{I}_{\mathcal{X}}(x) + \varsigma_p^a(y, x; \theta^p, h), \\ E[g_{\tau^p}^b(X, Y, x, y; \theta^p, h)] &= (\eta^p(y^{-p}; x|\theta^p) - \lambda^p(x; \theta^p)) \cdot \gamma_p^{II}(x; \theta^p) \cdot \mathbb{I}_{\mathcal{X}}(x) + \varsigma_p^b(y, x; \theta^p, h), \\ E[g_{\tau^p}^c(X, Y, x, y; \theta^p, h)] &= (\gamma_p^I(y^p, x; \theta^p) \cdot \eta^p(y^{-p}; x|\theta^p) - \gamma_p^{III}(x; \theta^p)) \cdot \mathbb{I}_{\mathcal{X}}(x) + \varsigma_p^c(y, x; \theta^p, h), \end{aligned}$$

where

$$\begin{aligned} \sup_{(x,y) \in \mathcal{W}, \theta^p \in \Theta} |\varsigma_p^a(y, x; \theta^p, h)| &\leq C \cdot h^M, \\ \sup_{(x,y) \in \mathcal{W}, \theta^p \in \Theta} |\varsigma_p^b(y, x; \theta^p, h)| &\leq C \cdot h^M, \\ \sup_{(x,y) \in \mathcal{W}, \theta^p \in \Theta} |\varsigma_p^c(y, x; \theta^p, h)| &\leq C \cdot h^M, \end{aligned}$$

In particular, this implies that

$$\sup_{\theta^p \in \Theta} \left| E [g_{\tau^p}(X_i, Y_i, X_j, Y_j; \theta^p, h_n) | X_i, Y_i] \right| \leq C \cdot h_n^M,$$

and if we define

$$\begin{aligned} \psi_U^p(Y, X; \theta^p) = & \\ & \left[ (\gamma_p^I(Y^p, X; \theta^p) - \gamma_p^{II}(X; \theta^p)) \cdot \lambda^p(X; \theta^p) + (\eta^p(Y^{-p}; X | \theta^p) - \lambda^p(X; \theta^p)) \cdot \gamma_p^{II}(X; \theta^p) \right. \\ & \left. + (\gamma_p^I(Y^p, X; \theta^p) \cdot \eta^p(Y^{-p}; X | \theta^p) - \gamma_p^{III}(X; \theta^p)) \right] \cdot \mathbb{I}_{\mathcal{X}}(X), \end{aligned} \quad (\text{S.14})$$

then

$$E [g_{\tau^p}(X_i, Y_i, X_j, Y_j; \theta^p, h_n) | X_j, Y_j] = \psi_U^p(Y_j, X_j; \theta^p) + \varsigma_{p,n}(\theta^p), \quad \text{where} \quad \sup_{\theta^p \in \Theta} |\varsigma_{p,n}(\theta^p)| = O_p(h_n^M)$$

Combining Assumptions B1, B2 and B4 we can show that the class of functions

$$\mathcal{F} = \{f : \mathcal{W} \times \mathcal{W} \rightarrow \mathbb{R} : f(x_1, y_1, x_2, y_2) = g_{\tau^p}(x_1, y_1, x_2, y_2; \theta^p, h) \text{ for some } \theta^p \in \Theta \text{ and some } h > 0\}$$

is Euclidean with respect to an envelope with finite second moment. Combining this with our previous results, a Hoeffding decomposition (Serfling (1980)) and Corollary 4 in Sherman (1994) imply that (S.13) can be expressed as

$$\frac{1}{n} \sum_{i=1}^n (\widehat{\tau}^p(Y_i^p | X_i; \theta^p) - \tau^p(Y_i^p | X_i; \theta^p)) \cdot \mathbb{1}_{\{\tau^p(Y_i^p | X_i; \theta^p) \geq 0\}} \cdot \mathbb{I}_{\mathcal{X}}(X_i) = \frac{1}{n} \sum_{i=1}^n \psi_U^p(Y_i, X_i; \theta^p) + \vartheta_{p,n}(\theta^p),$$

where

$$\sup_{\theta^p \in \Theta} |\vartheta_{p,n}(\theta^p)| = O_p \left( \frac{\log(n)^2}{nh_n^q} \right) + O_p \left( \frac{1}{nh_n^q} \right) + O_p(h_n^M) = O_p(n^{-1/2-\epsilon}) \quad \text{for some } \epsilon > 0,$$

where the last line follows from our bandwidth convergence conditions. Going back to (S.10) and (S.11) we obtain

$$\begin{aligned} \widehat{T}_{\mathcal{X}}^p(\theta^p) &= T_{\mathcal{X}}^p(\theta^p) + \frac{1}{n} \sum_{i=1}^n \psi^p(Y_i, X_i; \theta^p) + \varepsilon_{p,n}(\theta^p), \\ \text{where } \psi^p(Y_i, X_i; \theta^p) &= (\max\{\tau^p(Y_i^p | X_i; \theta^p), 0\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) - T_{\mathcal{X}}^p(\theta^p)) + \psi_U^p(Y_i, X_i; \theta^p), \\ \text{and } \sup_{\theta^p \in \Theta} |\varepsilon_{p,n}(\theta^p)| &= O_p(n^{-1/2-\epsilon}) \quad \text{for some } \epsilon > 0. \end{aligned} \quad (\text{S.15})$$

This concludes Step 3 and finishes the proof of Theorem 2.  $\square$

### S-A.0.1 Two key properties of $\psi^p$

The ‘‘influence function’’  $\psi^p$  has two key properties:

$$(i) \ E[\psi^p(Y_i, X_i; \theta^p)] = 0 \quad \forall \theta^p \in \Theta.$$

$$(ii) \ \psi^p(Y_i, X_i; \theta^p) = 0 \quad \forall \theta^p : \tau^p(Y^p|X; \theta^p) < 0 \quad \text{w.p.1.}$$

Part (ii) is obvious by inspection. To see why (i) is true we can show how it holds for each one of the summands that comprise  $\psi^p$ . Note first that by definition,

$$E[\max\{\tau^p(Y_i^p|X_i; \theta^p), 0\} \cdot \mathbb{I}_{\mathcal{X}}(X) - T_{\mathcal{X}}^p(\theta^p)] = 0.$$

We will show how each of the three summands that comprise  $\psi_U^p$  has mean zero. We begin with the first term. Exchanging the order of integration, we have

$$\begin{aligned} & E\left[\left(\gamma_p^I(Y_i^p, X_i; \theta^p) - \gamma_p^{II}(X_i; \theta^p)\right) \cdot \lambda^p(X_i; \theta^p) \cdot \mathbb{I}_{\mathcal{X}}(X_i)\right] \\ &= E_{X_i} \left[ E_{Y_j|X_j} \left[ E_{Y_i|X_i} \left[ (\mathbb{1}\{Y_i^p \leq Y_j^p\} - F_{Y^p}(Y_j^p|X_i)) |X_i, Y_j, X_j\right] \cdot \mathbb{1}\{\tau^p(Y_j^p|X_i; \theta^p) \geq 0\} |X_j = X_i, X_i\right] \right. \\ &\quad \left. \times \lambda^p(X_i; \theta^p) \cdot \mathbb{I}_{\mathcal{X}}(X_i) \right] \\ &= E_{X_i} \left[ E_{Y_j|X_j} \left[ E_{Y_i|X_i} \left[ (F_{Y^p}(Y_j^p|X_i) - F_{Y^p}(Y_j^p|X_i)) |X_i, Y_j, X_j\right] \cdot \mathbb{1}\{\tau^p(Y_j^p|X_i; \theta^p) \geq 0\} |X_j = X_i, X_i\right] \right. \\ &\quad \left. \times \lambda^p(X_i; \theta^p) \cdot \mathbb{I}_{\mathcal{X}}(X_i) \right] = 0 \end{aligned}$$

For the second term we have

$$\begin{aligned} & E\left[\left(\eta^p(Y_i^{-p}; X_i|\theta^p) - \lambda^p(X_i; \theta^p)\right) \cdot \gamma_p^{II}(X_i; \theta^p) \cdot \mathbb{I}_{\mathcal{X}}(X_i)\right] \\ &= E_{X_i} \left[ (\lambda^p(X_i; \theta^p) - \lambda^p(X_i; \theta^p)) \cdot \gamma_p^{II}(X_i; \theta^p) \cdot \mathbb{I}_{\mathcal{X}}(X_i) \right] = 0, \end{aligned}$$

where we simply used the fact that  $\lambda^p(X_i; \theta^p) = E_{Y^{-p}|X}[\eta^p(Y_i^{-p}; X_i|\theta^p)|X_i]$ . For the third term, exchanging the order of integration we have

$$\begin{aligned} & E\left[\left(\gamma_p^I(Y_i^p, X_i; \theta^p) \cdot \eta^p(Y_i^{-p}; X_i|\theta^p) - \gamma_p^{III}(X_i; \theta^p)\right) \cdot \mathbb{I}_{\mathcal{X}}(X_i)\right] \\ &= E_{X_i} \left[ E_{Y_j|X_j} \left[ E_{Y_i|X_i} \left[ (\mathbb{1}\{Y_i^p \leq Y_j^p\} \cdot \eta^p(Y_i^{-p}; X_i|\theta^p) - \mu^p(Y_j|X_i; \theta^p)) |X_i, Y_j, X_j\right] \right. \right. \\ &\quad \left. \left. \times \mathbb{1}\{\tau^p(Y_j^p|X_i; \theta^p) \geq 0\} |X_j = X_i, X_i\right] \times \mathbb{I}_{\mathcal{X}}(X_i) \right] \\ &= E_{X_i} \left[ E_{Y_j|X_j} \left[ (\mu^p(Y_j|X_i; \theta^p) - \mu^p(Y_j|X_i; \theta^p)) \times \mathbb{1}\{\tau^p(Y_j^p|X_i; \theta^p) \geq 0\} |X_j = X_i, X_i\right] \times \mathbb{I}_{\mathcal{X}}(X_i) \right] = 0 \end{aligned}$$

Combining these results we have  $E[\psi^p(Y_i, X_i; \theta^p)] = 0 \quad \forall \theta^p \in \Theta$ , as claimed.

### S-A.1 Constructing a confidence set

Let  $\kappa_n$  denote any sequence of positive numbers such that  $\kappa_n \rightarrow 0$  and  $n^\epsilon \kappa_n \rightarrow \infty$  for any  $\epsilon > 0$ .

For each  $\theta \in \Theta$  define  $t_n(\theta) = \frac{\sqrt{n} \cdot \hat{T}_{\mathcal{X}}(\theta)}{\max\{\kappa_n, \sigma(\theta)\}}$ . By Theorem 2 and (B-3),

$$t_n(\theta) = \frac{\sqrt{n} \cdot T_{\mathcal{X}}(\theta)}{\max\{\kappa_n, \sigma(\theta)\}} + \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\psi(Y_i, X_i; \theta)}{\max\{\kappa_n, \sigma(\theta)\}} + \varsigma_n(\theta).$$

By Theorem 2 and (B-3),  $\sup_{\theta \in \Theta} |\varsigma_n(\theta)| = o_p(1)$  since

$$\sup_{\theta \in \Theta} |\varsigma_n(\theta)| = \sup_{\theta \in \Theta} \left| \frac{\sqrt{n} \cdot \varepsilon_n(\theta)}{\max\{\kappa_n, \sigma(\theta)\}} \right| = O_p\left(\frac{1}{n^\epsilon \cdot \kappa_n}\right) \text{ for some } \epsilon > 0,$$

and  $n^\epsilon \kappa_n \rightarrow \infty$  for any  $\epsilon > 0$ . Let

$$\bar{\Theta}_{\mathcal{X}}^I = \{\theta \in \Theta : \tau^p(Y^p|X; \theta^p) < 0 \text{ w.p.1. } \forall p = 1, \dots, P\}$$

$\bar{\Theta}_{\mathcal{X}}^I$  is the collection of parameter values that satisfy our inequalities as *strict inequalities* w.p.1 over our inference range. Inspecting the terms that comprise  $\psi(Y_i, X_i; \theta)$ , we can see that  $\psi(Y_i, X_i; \theta) = 0$  w.p.1  $\forall \theta \in \bar{\Theta}_{\mathcal{X}}^I$ . On the other hand, inspecting the terms that comprise  $\psi_U^p(Y, X; \theta^p)$  we can verify that  $P(\psi_U^p(Y, X; \theta^p) \neq 0) > 0$  for any  $\theta \in \Theta_{\mathcal{X}}^I \setminus \bar{\Theta}_{\mathcal{X}}^I$  and therefore  $\sigma^2(\theta) > 0$  for any such  $\theta$ . Therefore,

- (i) If  $\theta \in \Theta \setminus \Theta_{\mathcal{X}}^I$ , then  $T_{\mathcal{X}}(\theta) > 0$  and therefore  $t_n(\theta) \rightarrow +\infty$  w.p.1.
- (ii) If  $\theta \in \bar{\Theta}_{\mathcal{X}}^I$ , then  $t_n(\theta) = o_p(1)$ .
- (iii) If  $\theta \in \Theta_{\mathcal{X}}^I \setminus \bar{\Theta}_{\mathcal{X}}^I$ , then  $t_n(\theta) \xrightarrow{d} \mathcal{N}(0, 1)$ .

$t_n(\theta)$  is unfeasible because  $\sigma^2(\theta)$  is unknown. However it can be estimated, we use  $\hat{t}_n(\theta) = \frac{\sqrt{n}\hat{T}_{\mathcal{X}}(\theta)}{\max\{\kappa_n, \hat{\sigma}(\theta)\}}$ , where

$$\begin{aligned} \hat{\psi}_U^p(Y_i, X_i; \theta^p) &= \frac{1}{(n-1)} \sum_{j \neq i} \hat{g}_{\tau^p}(X_j, Y_j, X_i, Y_i; \theta^p, h_n), \\ \hat{\psi}^p(Y_i, X_i; \theta^p) &= \hat{\tau}^p(Y_i^p|X_i; \theta^p) \cdot \mathbb{1}\{\hat{\tau}^p(Y_i^p|X_i; \theta^p) \geq -b_n\} \cdot \mathbb{I}_{\mathcal{X}}(X_i) - \hat{T}_{\mathcal{X}}^p(\theta^p) + \hat{\psi}_U^p(Y_i, X_i; \theta^p), \\ \hat{\sigma}^2(\theta) &= \frac{1}{n} \sum_{i=1}^n \hat{\psi}(Y_i, X_i; \theta)^2. \end{aligned} \tag{S.16}$$

$g_{\tau^p}$  is as described in (S.12). Under our assumptions we have  $\hat{\sigma}^2(\theta) \xrightarrow{p} \sigma^2(\theta)$  for each  $\theta \in \Theta$ .

### Confidence set and pointwise asymptotic properties

For a desired coverage probability  $1 - \alpha$ , our confidence set (CS) for  $\theta_0$  is of the form

$$CS_n(1 - \alpha) = \{\theta \in \Theta : \hat{t}_n(\theta) \leq c_{1-\alpha}\}, \tag{S.17}$$

where  $c_{1-\alpha}$  is the Standard Normal critical value for  $1 - \alpha$ . By the features outlined above our CS will have correct pointwise coverage properties. Namely,

$$\inf_{\theta \in \Theta: \theta = \theta_0} \liminf_{n \rightarrow \infty} P(\theta \in CS_n(1 - \alpha)) \geq 1 - \alpha$$

And if  $\Theta_{\mathcal{X}}^I \setminus \bar{\Theta}_{\mathcal{X}}^I \neq \emptyset$ , then

$$\inf_{\theta \in \Theta: \theta = \theta_0} \liminf_{n \rightarrow \infty} P(\theta \in CS_n(1 - \alpha)) = 1 - \alpha$$

Our CS will also satisfy

$$\lim_{n \rightarrow \infty} P(\theta \in CS_n(1 - \alpha)) = 0 \quad \forall \theta \in \Theta \setminus \Theta_{\mathcal{X}}^I.$$

By the design of our CS, its pointwise properties have the potential to hold uniformly (i.e., over sequences of parameter values and distributions) under appropriate assumptions about the underlying space of distributions. We describe those assumptions next and we characterize the asymptotic properties that would follow from them.

## S-B Analysis of uniform properties of our CS

Let us generalize our basic setup and assume that  $\left\{ \left( (Y_i^p)_{p=1}^P, X_i \right) : 1 \leq i \leq n, n \geq 1 \right\}$  is a triangular array, row-wise iid with distribution  $F_n \in \mathcal{F}$ . For a given  $F \in \mathcal{F}$  we will now index all the objects that depend on the distribution of the data by  $F$ . Thus, we denote  $\psi(Y, X; \theta, F)$ ,  $\sigma^2(\theta, F)$ ,  $\Theta_{\mathcal{X}}^I(F)$ ,  $\bar{\Theta}_{\mathcal{X}}^I(F)$ , and so on. We assume the following conditions about  $\mathcal{F}$ .

**Assumption S1.** *The space of distributions  $\mathcal{F}$  has common support and satisfies  $P_F(X \in \mathcal{X}) \geq p > 0$  for all  $F \in \mathcal{F}$ . In addition:*

(i) *The conditions in Assumptions B1, B3 and B4 are satisfied by every  $F \in \mathcal{F}$ .*

(ii) *For some  $\delta > 0$  and  $b < \infty$ ,*

$$\sup_{\substack{\theta \in \Theta \setminus \bar{\Theta}_{\mathcal{X}}^I(F) \\ F \in \mathcal{F}}} E_F \left[ \frac{|\psi(Y, X; \theta, F)|^{2+\delta}}{\sigma^{2+\delta}(\theta, F)} \right] \leq b$$

### S-B.0.1 Coverage properties

Part (i) of Assumption S1 is meant to ensure that the linear representation in (B-3) holds uniformly over  $\mathcal{F}$ . Part (ii) is sufficient to ensure the Lindeberg condition,

$$\lim_{\lambda \rightarrow \infty} \sup_{\substack{\theta \in \Theta \setminus \bar{\Theta}_{\mathcal{X}}^I(F) \\ F \in \mathcal{F}}} E_F \left[ \frac{|\psi(Y, X; \theta, F)|^2}{\sigma^2(\theta, F)} \cdot \mathbb{1} \left\{ \frac{|\psi(Y, X; \theta, F)|}{\sigma(\theta, F)} > \lambda \right\} \right] = 0.$$

To see why, note that for any  $\tilde{\lambda} > 0$  and  $\delta > 0$ ,  $\tilde{\lambda}^\delta \cdot \psi(Y, X; \theta, F)^2 \cdot \mathbb{1} \left\{ |\psi(Y, X; \theta, F)| > \tilde{\lambda} \right\} \leq |\psi(Y, X; \theta, F)|^{2+\delta}$ . Therefore  $E \left[ \psi(Y, X; \theta, F)^2 \cdot \mathbb{1} \left\{ |\psi(Y, X; \theta, F)| > \tilde{\lambda} \right\} \right] \leq \frac{E \left[ |\psi(Y, X; \theta, F)|^{2+\delta} \right]}{\tilde{\lambda}^\delta}$ . The Lindeberg condition follows by using the  $\delta$  described in Assumption S1, letting  $\tilde{\lambda} = \sigma(\theta, F)$  and dividing both sides of the inequality by  $\sigma^2(\theta, F)$ . Combined with the kernel and bandwidth conditions in Assumption B2, part (i) and the Lindeberg condition implied by part (ii) of Assumption S1 imply that for any sequence  $(F_n, \theta_n)$  such that  $F_n \in \mathcal{F}$  and  $\theta_n \in \Theta_{\mathcal{X}}^I(F_n) \setminus \bar{\Theta}_{\mathcal{X}}^I(F_n)$ ,

$$\frac{\sqrt{n} \cdot \hat{T}_{\mathcal{X}}(\theta_n)}{\sigma(\theta_n, F_n)} \xrightarrow{d} \mathcal{N}(0, 1).$$



And for any sequence  $(F_n, \theta_n)$  such that  $F_n \in \mathcal{F}$  and  $\theta_n \in \bar{\Theta}_{\mathcal{X}}^I(F_n)$ ,

$$\frac{\sqrt{n} \cdot \widehat{T}_{\mathcal{X}}(\theta_n)}{\max\{\kappa_n, \sigma(\theta_n, F_n)\}} \xrightarrow{p} 0.$$

Let  $t_n(\theta) = \frac{\sqrt{n} \widehat{T}_{\mathcal{X}}(\theta)}{\max\{\kappa_n, \sigma(\theta, F_n)\}}$  denote the unfeasible test-statistic that uses  $\sigma(\theta, F_n)$  instead of  $\widehat{\sigma}(\theta)$ .

Combined, parts (i) and (ii) of Assumption S1 would yield

$$\liminf_{n \rightarrow \infty} \inf_{\substack{\theta \in \Theta: \theta = \theta_0 \\ F \in \mathcal{F}}} P_F(t_n(\theta) \leq c_{1-\alpha}) \geq 1 - \alpha, \quad (\text{S.18})$$

with

$$\liminf_{n \rightarrow \infty} \inf_{\substack{\theta \in \Theta: \theta = \theta_0 \\ F \in \mathcal{F}}} P_F(t_n(\theta) \leq c_{1-\alpha}) = 1 - \alpha \quad \text{if } \Theta_{\mathcal{X}}^I(F) \setminus \bar{\Theta}_{\mathcal{X}}^I(F) \neq \emptyset \text{ for some } F \in \mathcal{F}.$$

Of course, our CS is based on  $\widehat{t}_n(\theta) = \frac{\sqrt{n} \widehat{T}_{\mathcal{X}}(\theta)}{\max\{\kappa_n, \widehat{\sigma}(\theta)\}}$ , where  $\widehat{\sigma}^2(\theta)$  is estimated as described in (S.16). We need to endow  $\mathcal{F}$  with conditions that ensure that the necessary Laws of Large Numbers for triangular arrays hold in a way that ensures that  $|\widehat{\sigma}^2(\theta_n) - \sigma^2(\theta_n, F_n)| \xrightarrow{p} 0$  over sequences  $(F_n, \theta_n) \in \mathcal{F} \times \Theta$ . For this we can look at the type of sufficient conditions found in Romano (2004, Lemma 2). To this end we impose the following conditions.

**Assumption S2.** Let  $\psi_{F_{Y^p}}$ ,  $\psi_{\lambda^p}$ ,  $\psi_{\mu^p}$  and  $\psi_{\tau^p}$  and  $g_{\tau^p}$  be as described in (S.3), (S.5) and (S.12). Then, for some  $\delta > 0$  and  $b < \infty$  the following holds for each  $p = 1, \dots, P$ ,

$$\begin{aligned} & \sup_{\substack{F \in \mathcal{F} \\ (y^p, x) \in \mathcal{W} \\ h > 0}} E_F \left[ \left| \frac{1}{h^q} \psi_{F_{Y^p}}(Y_i^p, X_i, y^p, x; h, F) - E \left[ \frac{1}{h^q} \psi_{F_{Y^p}}(Y_i^p, X_i, y^p, x; h, F) \right] \right|^{1+\delta} \right] \leq b, \\ & \sup_{\substack{F \in \mathcal{F} \\ x \in \mathcal{X} \\ \theta^p \in \Theta \\ h > 0}} E_F \left[ \left| \frac{1}{h^q} \psi_{\lambda^p}(Y_i^{-p}, X_i, x, \theta^p; h) - E \left[ \frac{1}{h^q} \psi_{\lambda^p}(Y_i^{-p}, X_i, x, \theta^p; h) \right] \right|^{1+\delta} \right] \leq b, \\ & \sup_{\substack{F \in \mathcal{F} \\ (y^p, x) \in \mathcal{W} \\ \theta^p \in \Theta \\ h > 0}} E_F \left[ \left| \frac{1}{h^q} \psi_{\mu^p}(Y_i, X_i, y^p, x, \theta^p; h) - E \left[ \frac{1}{h^q} \psi_{\mu^p}(Y_i, X_i, y^p, x, \theta^p; h) \right] \right|^{1+\delta} \right] \leq b, \\ & \sup_{\substack{F \in \mathcal{F} \\ (y^p, x) \in \mathcal{W} \\ \theta^p \in \Theta \\ h > 0}} E_F \left[ \left| \frac{1}{h^q} \psi_{\tau^p}(Y, X, y^p, x, \theta^p; h, F) - E \left[ \frac{1}{h^q} \psi_{\tau^p}(Y, X, y^p, x, \theta^p; h, F) \right] \right|^{1+\delta} \right] \leq b, \\ & \sup_{\substack{F \in \mathcal{F} \\ (y, x) \in \mathcal{W} \\ \theta^p \in \Theta \\ h > 0}} E_F \left[ \left| g_{\tau^p}(Y, X, x, y; \theta^p, h, F) - E[g_{\tau^p}(Y, X, x, y; \theta^p, h, F)] \right|^{1+\delta} \right] \leq b, \end{aligned}$$

Assumption S2 is sufficient to satisfy the conditions for the Law of Large Numbers for triangular arrays in Romano (2004, Lemma 2). Combined with Assumption S1, the smoothness conditions in

Assumption B1 and the linear representation in (S.6), Assumption S2 and Romano (2004, Lemma 2) can be used to show that for any sequence  $(F_n, \theta_n) \in \mathcal{F} \times \Theta$ ,

$$|\widehat{\sigma}^2(\theta_n) - \sigma^2(\theta_n, F_n)| \xrightarrow{p} 0.$$

Combining Assumptions S1 and S2, our confidence sets would inherit the coverage properties in (S.18). Namely,

$$\liminf_{n \rightarrow \infty} \inf_{\substack{\theta \in \Theta: \theta = \theta_0 \\ F \in \mathcal{F}}} P_F(\theta \in CS_n(1 - \alpha)) \geq 1 - \alpha,$$

with

$$\liminf_{n \rightarrow \infty} \inf_{\substack{\theta \in \Theta: \theta = \theta_0 \\ F \in \mathcal{F}}} P_F(\theta \in CS_n(1 - \alpha)) = 1 - \alpha \quad \text{if } \Theta_{\mathcal{X}}^I(F) \setminus \overline{\Theta}_{\mathcal{X}}^I(F) \neq \emptyset \text{ for some } F \in \mathcal{F}.$$

### S-B.1 Power properties

The linear representation in (B-3) facilitates the study of the power features of our procedure. Take a sequence  $(F_n, \theta_n)$  such that  $F_n \in \mathcal{F}$  and  $\theta_n \in \Theta \setminus \Theta_{\mathcal{X}}^I(F_n)$ . By Assumption S1(ii), for any  $c$  we have

$$\lim_{n \rightarrow \infty} P_{F_n} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\psi(Y_i, X_i; \theta_n, F_n)}{\sigma(\theta_n, F_n)} > c \right) = 1 - \Phi(c).$$

The key to the power properties of our test over such a sequence is the behavior of  $\sigma^2(\theta_n, F_n) = \text{Var}_{F_n}(\psi(Y, X; \theta_n, F_n))$ . Recall that  $T_{\mathcal{X}}(\theta_n, F_n) = \sum_{p=1}^P E[\max\{\tau^p(Y^p|X; \theta_n^p, F_n), 0\} \cdot \mathbb{1}_{\mathcal{X}}(X)]$ . By Assumption S1,  $\lim_{n \rightarrow \infty} P_{F_n}(X \in \mathcal{X}) \geq \underline{p} > 0$  for any sequence  $F_n \in \mathcal{F}$ . Therefore we have  $T_{\mathcal{X}}(\theta_n, F_n) \rightarrow 0$  if and only if  $P_{F_n}(\tau^p(Y^p|X; \theta_n, F_n) > 0 | X \in \mathcal{X}) \rightarrow 0$  for each  $p = 1, \dots, P$ . If we inspect the structure of  $\psi(Y, X; \theta_n, F_n)$  we will see that the key will be the behavior of the sequence

$$P_{F_n}(\tau^p(Y^p|X; \theta_n, F_n) = 0 \text{ for some } p = 1, \dots, P | X \in \mathcal{X}) \equiv \Delta_{\mathcal{X}}(\theta_n, F_n).$$

$\Delta_{\mathcal{X}}(\theta_n, F_n)$  is the probability that the inequalities are binding for some  $p$  over our inference range. We have the following:

- (i) If  $T_{\mathcal{X}}(\theta_n, F_n) \rightarrow 0$  and  $\Delta_{\mathcal{X}}(\theta_n, F_n) \rightarrow 0$ , then  $\sigma(\theta_n, F_n) \rightarrow 0$ .
- (ii) If  $T_{\mathcal{X}}(\theta_n, F_n) \rightarrow 0$  but  $\Delta_{\mathcal{X}}(\theta_n, F_n) \not\rightarrow 0$ , then  $\sigma(\theta_n, F_n) \not\rightarrow 0$ .
- (iii) If  $T_{\mathcal{X}}(\theta_n, F_n) \not\rightarrow 0$ , then  $\sigma(\theta_n, F_n) \not\rightarrow 0$ .

The asymptotic power of our approach will be determined by the behavior of the following two sequences,

$$s_{1,n}(\theta_n, F_n) = \frac{\max\{\kappa_n, \sigma(\theta_n, F_n)\}}{\sigma(\theta_n, F_n)}, \quad \text{and} \quad s_{2,n}(\theta_n, F_n) = \frac{\sqrt{n} \cdot T_{\mathcal{X}}(\theta_n, F_n)}{\max\{\kappa_n, \sigma(\theta_n, F_n)\}}$$

Suppose  $s_{1,n}(\theta_n, F_n) \rightarrow s_1$  and  $s_{2,n}(\theta_n, F_n) \rightarrow s_2$ . Note that  $s_1 \geq 1$  by construction. If Assumptions S1 and S2 hold, the conditions in Romano (2004, Theorem 5) are satisfied and we can use this to show that

$$\lim_{n \rightarrow \infty} P_{F_n}(\hat{t}(\theta_n) > c_{1-\alpha}) = 1 - \Phi(s_1 \cdot (c_{1-\alpha} - s_2)).$$

From here we conclude that our procedure will have asymptotic power of 1 if either:

- (a)  $s_2 = \infty$ : This includes as a special case any sequence such that  $T_{\mathcal{X}}(\theta_n, F_n) = O(n^{-\alpha})$  for some  $\alpha < 1/2$ . In this case we would have  $s_{2,n}(\theta_n, F_n) = O\left(\frac{n^{1/2-\alpha}}{\kappa_n}\right) \rightarrow \infty$  by the convergence restrictions of  $\kappa_n$ .
- (b)  $s_1 = \infty$  and  $s_2 > c_{1-\alpha}$ : Firstly, our discussion above implies that  $s_1 = \infty$  can occur only if  $\Delta_{\mathcal{X}}(\theta_n, F_n) \rightarrow 0$  and  $T_{\mathcal{X}}(\theta_n, F_n) \rightarrow 0$ . The additional condition  $s_2 > c_{1-\alpha}$  forbids  $T_{\mathcal{X}}(\theta_n, F_n)$  from converging to zero “too fast”.

Part (a) shows that our procedure will have asymptotic power of 1 whenever  $T_{\mathcal{X}}(\theta_n, F_n) = O(n^{-\alpha})$  for some  $\alpha < 1/2$ . Suppose  $T_{\mathcal{X}}(\theta_n, F_n) = O(n^{-\alpha})$  for some  $\alpha > 1/2$ . Then we have  $s_2 = 0$  by the bandwidth convergence restrictions of  $\kappa_n$ . In this case our approach will have asymptotic power of zero if  $s_1 = \infty$  (i.e, if  $\sigma(\theta_n, F_n)/\kappa_n \rightarrow 0$ ). On the other hand if  $\sigma(\theta_n, F_n)/\kappa_n \rightarrow \infty$  then the asymptotic power will be  $\alpha$ . This will be the case, for example, for any sequence such that  $T_{\mathcal{X}}(\theta_n, F_n) = O(n^{-\alpha})$  for some  $\alpha > 1/2$  but  $\lim_{n \rightarrow \infty} \Delta_{\mathcal{X}}(\theta_n, F_n) > 0$ . On the other hand, our asymptotic power would be zero if  $\lim_{n \rightarrow \infty} \Delta_{\mathcal{X}}(\theta_n, F_n) = 0$ . If  $\sigma(\theta_n, F_n) \propto \kappa_n$ , the power will be bounded between zero and  $\alpha$ . Finally, suppose  $T_{\mathcal{X}}(\theta_n, F_n) = O(n^{-1/2})$ . Our procedure will have asymptotic power of 1 for any such sequence as long as  $\lim_{n \rightarrow \infty} \Delta_{\mathcal{X}}(\theta_n, F_n) = 0$ , as this would yield  $s_2 = \infty$ . If  $\lim_{n \rightarrow \infty} \Delta_{\mathcal{X}}(\theta_n, F_n) \neq 0$ , then  $s_2 < \infty$ . In this case our asymptotic power will be 1 if  $s_2 > c_{1-\alpha}$  but it will be zero if  $s_2 < c_{1-\alpha}$ . Thus, our asymptotic power for any sequence  $T_{\mathcal{X}}(\theta_n, F_n) = O(n^{-1/2})$  will be determined by the limit of the sequence  $\Delta_{\mathcal{X}}(\theta_n, F_n)$ . Note that –as one should expect– choosing the maximum rate of convergence for  $\kappa_n$  that is consistent with our assumptions is beneficial for power. Given our bandwidth convergence restrictions, this rate is  $\kappa_n \propto \log(n)$ . Our analysis shows the power advantages of our approach vis-a-vis using a test-statistic based on a least-favorable configuration, as this would be based on normalizing our test statistic by a standard deviation that does not converge to zero when  $T_{\mathcal{X}}(\theta_n, F_n) \rightarrow 0$ .

## S-C Kernels and bandwidths used in our empirical application

Our covariate vector  $X$  includes  $q = 8$  continuous random variables. The smallest kernel order  $M$  compatible with Assumption B2 is  $M = 2 \cdot q + 1 = 17$ . We employed a multiplicative kernel

$K(\psi_1, \dots, \psi_8) = k(\psi_1) \cdot k(\psi_2) \cdots k(\psi_8)$ , where each  $k(\cdot)$  is a bias-reducing Biweight-type kernel of order  $M = 18$  of the form,

$$k(u) = \sum_{j=1}^9 c_j \cdot (1 - u^2)^{2j} \cdot \mathbb{1}\{|u| \leq s\},$$

where  $c_1, \dots, c_9$  were chosen to satisfy the restriction of a bias-reducing kernel of order 18. As in Aradillas-López, Gandhi, and Quint (2013) we set  $s = 30$ . Following the guidelines in Assumption B2 we employed a bandwidth of the form  $h_n = c \cdot \hat{\sigma}(X) \cdot n^{-\alpha_h}$  (note that each  $X$  has its own bandwidth), where  $\alpha_h = \frac{1}{2M} + \bar{\epsilon}$  and  $\bar{\epsilon} = 10^{-5}$ . As a guidance to select the constant ‘ $c$ ’ we used the “rule of thumb” formula (Silverman (1986)), using the Normal distribution as the reference distribution. We set

$$c = 2 \cdot \left( \frac{\pi^{1/2} (M!)^3 \cdot R_k}{(2M) \cdot (2M)! \cdot (k_M^2)} \right)^{\frac{1}{2M+1}}, \quad \text{where } R_k = \int_{-1}^1 k^2(u) du, \quad k_M = \int_{-1}^1 u^M k(u) du.$$

This yielded  $c \approx 0.2$  and therefore  $h_n \approx 0.16 \cdot \hat{\sigma}(X)$  (for our sample size  $n = 954$ ). Let  $\bar{\Omega} = \max_{\theta \in \Theta} |\hat{\sigma}(\theta)|$ . We used  $b_n = c_b \cdot \bar{\Omega} \cdot n^{-\alpha_b}$  where  $\alpha_b = \frac{1}{4} + \bar{\epsilon}$  and  $\kappa_n = c_\kappa \cdot \bar{\Omega} \cdot \log(n)^{-1}$  with  $c_b = 10^{-6}$  and  $c_\kappa = 10^{-8}$ . We chose these tuning parameters proportional to  $\bar{\Omega}$  to ensure our procedure is scale-invariant. These bandwidth choices satisfy Assumption B2. For our sample size  $n = 954$  this resulted in  $b_n \approx 10^{-5}$  and  $\kappa_n \approx 10^{-7}$ . The inference range used was

$$\mathcal{X} = \left\{ x : \hat{f}_X(x) \geq \hat{f}_X^{(0.15)}, \quad POP < 5 \text{ Million} \right\},$$

where  $\hat{f}_X^{(0.15)}$  denotes the estimated 15th percentile of the density  $\hat{f}_X$ . Our main findings were qualitatively robust to moderate changes in these tuning parameters. Our results were qualitatively robust to moderate changes in the constants  $c$ ,  $c_b$ ,  $c_\kappa$ ,  $\alpha_h$  and  $\alpha_b$  used to construct our bandwidths.

## S-D Identification and nontrivial confidence sets

In this section we outline in more detail the type of DGP features that can lead us to reject parameter values and therefore produce nontrivial confidence sets (CS) of a parametric specification of the strategic index  $\eta^p(Y^{-p}; X|\theta^p)$ . In an effort to relate this to our empirical application, we focus on a game with three players (i.e.,  $P = 3$ ) and on the type of parameterizations of the strategic index we used there. Without loss of generality take player  $p = 1$  and assume that we focus on parameterizations of the strategic index of the form

$$\eta^1(y^2, y^3|\theta^1) = \theta^{12} \cdot y^2 + \theta^{13} \cdot y^3,$$

where  $\theta^{12}$  and  $\theta^{13}$  are constant parameters<sup>2</sup>, with  $\frac{\theta^{12}}{\theta^{13}} \equiv k_0 < \infty$ . Note that

$$Cov(\mathbb{1}\{Y^1 \leq y^1\}, \eta^1(Y^2, Y^3|\theta^1)|X) = \theta^{13} \cdot (k_0 \cdot Cov(\mathbb{1}\{Y^1 \leq y^1\}, Y^2|X) + Cov(\mathbb{1}\{Y^1 \leq y^1\}, Y^3|X)).$$

As in our empirical application, suppose we maintain that actions are strategic substitutes, and therefore  $\theta^{12} \geq 0$  and  $\theta^{13} \geq 0$ . Suppose there exists a range  $\mathcal{X}^* \subset \text{Supp}(X)$  with  $Pr(X \in \mathcal{X}^*) > 0$  where the possible outcomes for  $Y^2$  and  $Y^3$  that can be chosen by the equilibrium selection mechanism are concentrated over a set  $\mathcal{Y}_{2,3}^* \subset \mathcal{A}^2 \times \mathcal{A}^3$  that satisfies

$$\left( \frac{y^{3'} - y^3}{y^2 - y^{2'}} \right) > k_0 \quad \forall (y^2, y^3) \neq (y^{2'}, y^{3'}) : (y^2, y^3), (y^{2'}, y^{3'}) \in \mathcal{Y}_{2,3}^* \quad (\text{S.19})$$

Note that (S.19) implies a negative relationship between  $y^2$  and  $y^3$  everywhere in  $\mathcal{Y}_{2,3}^*$ , a natural condition in a strategic substitutes setting. If (S.19) holds, then for any pair  $(y^2, y^3)$  and  $(y^{2'}, y^{3'})$  in  $\mathcal{Y}_{2,3}^*$ ,

$$\eta^1(y^2, y^3|\theta^1) > \eta^1(y^{2'}, y^{3'}|\theta^1) \iff y^2 < y^{2'}.$$

Therefore, if  $X \in \mathcal{X}^*$  there exists a negative relationship between  $\eta^1(Y^2, Y^3|\theta^1)$  and  $Y^2$ . Note that (S.20) is entirely compatible with  $Y^2$  being a strategic substitute of  $Y^1$ . From this negative relationship and Theorem 1 in the paper we would obtain

$$X \in \mathcal{X}^* \implies Cov(\mathbb{1}\{Y^1 \leq y^1\}, Y^2|X) < 0 \quad \forall y^1 \in \mathcal{A}^1. \quad (\text{S.20})$$

We stress once again that (S.20) is entirely compatible with  $Y^2$  being a strategic substitute<sup>3</sup> of  $Y^1$ . It is simply the reflection of the range of values that  $(Y^2, Y^3)$  can take when  $X \in \mathcal{X}^*$ . Now take any  $\tilde{k} > k_0$  such that

$$Pr\left(\underbrace{\tilde{k} \cdot Cov(\mathbb{1}\{Y^1 \leq y^1\}, Y^2|X)}_{<0 \text{ when } X \in \mathcal{X}^*} + \underbrace{Cov(\mathbb{1}\{Y^1 \leq y^1\}, Y^3|X)}_{>0 \text{ when } X \in \mathcal{X}^*} < 0 \mid X \in \mathcal{X}^*\right) > 0,$$

for some  $y^1 \in \mathcal{A}^1$ . Then any  $\tilde{\theta}^1 \in \Theta$  with  $\frac{\tilde{\theta}^{12}}{\tilde{\theta}^{13}} \geq \tilde{k}$  would violate our inequalities and would therefore be rejected. Thus, a condition like (S.19) has potential identification power to rule out large values of  $k \equiv \frac{\theta^{12}}{\theta^{13}}$ . Note that (S.20) can be satisfied even if some outcomes in  $\mathcal{Y}_{2,3}^*$  violate (S.19) as long as those outcomes that satisfy (S.19) are selected with sufficiently high probability.

What restrictions does (S.19) imply on the range of equilibrium outcomes  $\mathcal{Y}_{2,3}^*$ ? That depends on the value of  $k_0$ . Suppose  $k_0 = 2$ . Then the following set satisfies the conditions in (S.19),

$$\mathcal{Y}_{2,3}^* = \{(1, 6), (2, 3), (3, 0)\}.$$

<sup>2</sup>If  $\theta^{13} = 0$ , simply re-define  $k_0 \equiv \frac{\theta^{13}}{\theta^{12}}$ . We focus on the case  $\|\theta^1\| \neq 0$  since, as we stated previously, the case  $\theta^{12} = \theta^{13} = 0$  (i.e., the case of no strategic interaction effect) can never be rejected given our assumptions.

<sup>3</sup>Recall that  $Y^q$  is a strategic substitute of  $Y^p$  if the index  $\eta^p(Y^q, Y^r|\theta^p)$  is increasing in  $Y^q$  keeping  $Y^r$  fixed, but  $Y^r$  is not fixed when we talk about  $Cov(\mathbb{1}\{Y^p \leq y^p\}, Y^q|X)$ .

Note that in this case the range of equilibrium values that can be chosen for  $Y^3$  is richer than that of  $Y^2$ , since  $Y^3 \in \{0, 3, 6\}$  while  $Y^2 \in \{0, 1, 2\}$ . The degree of asymmetry in equilibrium outcomes needed to satisfy (S.19) depends on how large  $k_0$  is.

What conditions can lead to rejecting small values of  $k \equiv \frac{\theta^{12}}{\theta^{13}}$  different from  $k_0$ ? Suppose there exist  $\mathcal{X}^{**}$  with  $Pr(X \in \mathcal{X}^{**}) > 0$  such that, whenever  $X \in \mathcal{X}^{**}$  the equilibrium selection mechanism concentrates over a set  $\mathcal{Y}_{2,3}^{**} \subset \mathcal{A}^2 \times \mathcal{A}^3$  such that

$$\left( \frac{y^{2'} - y^2}{y^3 - y^{3'}} \right) > \frac{1}{k_0} \quad \forall (y^2, y^3) \neq (y^{2'}, y^{3'}) : (y^2, y^3), (y^{2'}, y^{3'}) \in \mathcal{Y}_{2,3}^{**}. \quad (19')$$

Under (19') there would now exist a negative relationship between  $\eta^1(Y^2, Y^3 | \theta^1)$  and  $Y^3$  whenever  $X \in \mathcal{X}^{**}$ . By Theorem 1, this would now yield

$$X \in \mathcal{X}^{**} \implies Cov(\mathbb{1}\{Y^1 \leq y^1\}, Y^3 | X) < 0 \quad \forall y^1 \in \mathcal{A}^1. \quad (20')$$

While (S.19)-(S.20) can help us reject values of  $k \equiv \frac{\theta^{12}}{\theta^{13}}$  larger than  $k_0$ , (19')-(20') can help us reject values smaller than  $k_0$ . Take any  $\tilde{k} < k_0$  such that

$$Pr\left(\underbrace{\tilde{k} \cdot Cov(\mathbb{1}\{Y^1 \leq y^1\}, Y^2 | X)}_{>0 \text{ when } X \in \mathcal{X}^{**}} + \underbrace{Cov(\mathbb{1}\{Y^1 \leq y^1\}, Y^3 | X)}_{<0 \text{ when } X \in \mathcal{X}^{**}} < 0 \mid X \in \mathcal{X}^{**}\right) > 0,$$

for some  $y^1 \in \mathcal{A}^1$ . Then any  $\tilde{\theta}^1 \in \Theta$  with  $\frac{\tilde{\theta}^{12}}{\tilde{\theta}^{13}} \leq \tilde{k}$  would violate our inequalities and would therefore be rejected.

What restrictions does (19') imply on the range of equilibrium outcomes  $\mathcal{Y}_{2,3}^{**}$ ? Once again suppose  $k_0 = 2$ , then the following set satisfies (20'),

$$\mathcal{Y}_{2,3}^{**} = \{(0, 3), (1, 2), (2, 1), (3, 0)\}.$$

Let us compare this set with the example given above for  $\mathcal{Y}_{2,3}^*$ . In both instances we have a negative relationship between  $y^2$  and  $y^3$  (a natural condition with strategic substitutes), but while  $\mathcal{Y}_{2,3}^*$  required a fundamentally richer range of equilibrium outcomes for  $Y^3$  compared with  $Y^2$ , this is not the case in  $\mathcal{Y}_{2,3}^{**}$  (both  $Y^2$  and  $Y^3$  can take on the values  $\{0, 1, 2, 3\}$ ). It is easy to anticipate that this comparison would be reversed if we had  $k_0 < 1$ . The main insight of this example is that, if  $k_0$  is “large” (larger than 1), rejecting false values of  $\theta^1$  where  $\frac{\theta^{12}}{\theta^{13}}$  is larger than  $k_0$  requires the existence of regions where equilibrium outcomes are fundamentally asymmetric between players 2 and 3, while this may not be required in order to reject false values of  $\theta^1$  where  $\frac{\theta^{12}}{\theta^{13}}$  is smaller than  $k_0$ . This qualitative feature would be reversed if  $k_0 < 1$ . The main insight however is that, having two-sided identification power for  $k_0$  would require some degree of asymmetry in the range of equilibrium outcomes available to the players, and the extent of this asymmetry would depend on the specific value of  $k_0$ .

Note that if we have a more general form of the strategic index,

$$\eta^1(y^2, y^3 | X; \theta^1) = \phi^{12}(X; \theta^{12}) \cdot y^2 + \phi^{13}(X; \theta^{13}) \cdot y^3,$$

where  $\phi^{12}(\cdot; \theta^{12}) \geq 0$  and  $\phi^{13}(\cdot; \theta^{13}) \geq 0$  (pairwise strategic substitutes once again), then the results in (S.20) and (20') could be obtained even under weaker conditions than (S.19) and (19') respectively, since the relative strategic effects would now be allowed to vary with  $X$ .

### S-D.1 The presence of player-specific payoff shifters

In principle, a condition like (S.19) could arise entirely from the properties of the underlying equilibrium selection mechanism. However, an argument in favor of (S.19) would be stronger if there exist elements in  $X$  that shift individual players' payoffs asymmetrically in such a way that can generate intrinsically different ranges of equilibrium choices across players over certain regions of  $X$ . One natural way this can occur is when  $X$  includes player-specific payoff shifters, such as in our empirical application. In this context, (S.19) is more plausible to hold in markets where the nearest distribution center of player 3 (Walgreens) is much closer than that of player 2 (Rite Aid), and the regions  $\mathcal{X}^*$  could be (partially) characterized by this feature. Here we are thinking of markets where this difference in relative distance is such that the range of profitable number of stores is fundamentally different for players 2 and 3. Table (1) describes the difference in distance between pairs of firms in our empirical example in an effort to illuminate whether markets with marked asymmetries are prevalent in our data.

As we can see in Table 1, there exists a nontrivial proportion of markets with significant asym-

Table 1:  $DIST^p - DIST^q$ : Difference in distance (measured in miles) between the market and the nearest distribution center for firms  $p$  and  $q$ . Player 1 =CVS, 2 =Rite Aid and 3 =Walgreens

	$DIST^1 - DIST^2$	$DIST^1 - DIST^3$	$DIST^2 - DIST^3$
5 <sup>th</sup> quantile	-469	-117	-202
10 <sup>th</sup> quantile	-358	-91	-114
90 <sup>th</sup> quantile	121	300	445
95 <sup>th</sup> quantile	380	376	511
Median	-21	10	41

metries in relative distance; this is true for each pair of firms and in each direction. For example, the difference  $DIST^{CVS} - DIST^{Walgreens}$  is at least 30 times greater than its median value in 10% of the markets in our sample, and at least 42 times greater in 5% of markets. The availability of player-specific payoff shifters in our data, and the richness of such data lead us to believe that conditions such as the one described in (S.19) are plausible for the underlying DGP. As we outlined previously, depending on the true parameter values, conditions like (S.19) or (19') can hold without

any significant degree of asymmetry, and nontrivial confidence sets (CS) can result even without player-specific payoff shifters. Nevertheless, the availability of player-specific payoff shifters in our specific application and the richness displayed by such data lead us to believe that conditions such as the one described in (S.19) and (19') are very plausible for the underlying DGP in our application and can help explain why we obtained nontrivial<sup>4</sup>

## S-E Nonzero correlation across players' unobserved payoff shocks

Independent private shocks (Assumption 3) is a condition widely imposed in econometric work on incomplete information games. Nevertheless, it is an important restriction whose validity depends on the richness of the observable covariates  $X$  present in the data. It can be violated in many ways, but one that is particularly interesting is the case where there exist market-level unobserved shocks. In our context, we can model their presence by partitioning  $X$  as  $X = (X^O, X^U)$ , where  $X^O$  denotes the elements in  $X$  observed by the econometrician and  $X^U$  denotes the unobserved market-level shocks. For simplicity suppose  $X^U \in \mathbb{R}$  and also suppose that the strategic index  $\eta^p$  does not depend directly on  $X^U$  and is correctly specified by the econometrician. Also, maintain that, conditional on the entire vector  $X$ , our assumptions hold and therefore,

$$\text{Cov}(\mathbb{1}\{Y^p \leq y^p\}, \eta^p(Y^{-p}; X^O) | X) \geq 0 \quad \text{w.p.1 in } X. \quad (\text{S.21})$$

In this scenario the econometrician effectively misspecifies the model, excluding  $X^U$  from the vector of covariates  $X$  and using only the incomplete vector  $X^O$ , basing inference on the (possibly incorrect) restriction

$$\text{Cov}(\mathbb{1}\{Y^p \leq y^p\}, \eta^p(Y^{-p}; X^O) | X^O) \geq 0 \quad \text{w.p.1 in } X. \quad (21')$$

Our results would be inconsistent<sup>5</sup> if the above inequality is violated with positive probability at the true strategic index function. Using the so-called Law of Total Covariance,

$$\begin{aligned} \text{Cov}(\mathbb{1}\{Y^p \leq y^p\}, \eta^p(Y^{-p}; X^O) | X^O) &= \underbrace{E[\text{Cov}(\mathbb{1}\{Y^p \leq y^p\}, \eta^p(Y^{-p}; X^O) | X) | X^O]}_{\geq 0 \text{ from (S.21)}} \\ &+ \underbrace{\text{Cov}(E[\mathbb{1}\{Y^p \leq y^p\} | X], E[\eta^p(Y^{-p}; X^O) | X]) | X^O}_{\text{sign undetermined}} \end{aligned} \quad (22)$$

Our results would be inconsistent only if the second term is negative. This, in turn, will be determined by the way in which the unobserved shock  $X^U$  shifts players' payoff functions, and by the

<sup>4</sup>For example, the relatively large prevalence of markets where the distribution center of Walgreens was much closer than that of CVS may help explain why the CS obtained for Rite Aid was more informative than those of the two other firms.

<sup>5</sup>Inconsistency here occurs when our CS excludes the true strategic index function.



properties of the equilibrium selection mechanism. Suppose the common shock  $X^U$  shifts all players' payoffs in the same direction. To make matters more precise, *suppose all players' equilibrium choices are almost surely nondecreasing in  $X^U$*  (the conclusion to follow will also hold if equilibrium choices are nonincreasing in  $X^U$ , all that matters is that they are all affected in the same direction). Then the presence of  $X^U$  will have very different implications if actions are strategic complements vs. substitutes. Suppose all actions  $Y^{-p}$  are strategic complements of  $Y^p$ . Then, for a.e.  $X^O$ , both  $E[\mathbb{1}\{Y^p \leq y^p\} | X^U, X^O]$  and  $E[\eta^p(Y^{-p}; X^O) | X^U, X^O]$  are nonincreasing in  $X^U$ , which would lead to  $Cov(E[\mathbb{1}\{Y^p \leq y^p\} | X], E[\eta^p(Y^{-p}; X^O) | X] | X^O)$  being nonnegative. In this scenario (21') would be true and our results would not be inconsistent. On the other hand, if all actions  $Y^{-p}$  are strategic substitutes of  $Y^p$ , then  $E[\eta^p(Y^{-p}; X^O) | X^U, X^O]$  would be nondecreasing in  $X^U$ , leading to the possibility that  $Cov(E[\mathbb{1}\{Y^p \leq y^p\} | X], E[\eta^p(Y^{-p}; X^O) | X] | X^O)$  is negative. This in turn could lead to a violation of (21') and inconsistency of our results.

Under the presence of an unobserved common shock that shifts all payoffs in the same direction, a game of strategic complements has a better possibility of preserving consistency. In general, one way to guard against inconsistency would be to choose an inference range for the observable covariates where the magnitude of the second term in (22) may be mitigated. While the functional form of the strategic index can help guide the choice of such a range, it would also require, in general, more precise assumptions about the direction in which payoff functions shift with the observable covariates included.

## S-F Monte Carlo experiments continued: performance under violations to our assumptions

Our last goal in this section is to investigate the extent to which the properties of our CS break down when some of our key assumptions are violated. Specifically we want to study what happens when two key conditions are violated:

- (i) Violations to Assumption 3 introducing correlation in players' private shocks.
- (ii) Violations to Assumption 1. Specifically, to the assumption that the strategic index  $\eta^p$  can be expressed as a function solely of observable payoff shifters  $X$ .

To modify our design in a way that violates both assumptions, the demand system is now given by

$$\begin{aligned} \mathcal{P}^1 &= \zeta^1 \cdot X_a - (\lambda^1 + \delta^1 \cdot X_b) \cdot Y^1 - (\beta^{12} + \gamma^{12} \cdot X_b + \rho \cdot \zeta) \cdot Y^2 - (\beta^{13} + \gamma^{13} \cdot X_b + \rho \cdot \zeta) \cdot Y^3, \\ \mathcal{P}^2 &= \zeta^2 \cdot X_a - (\lambda^2 + \delta^2 \cdot X_b) \cdot Y^2 - (\beta^{21} + \gamma^{21} \cdot X_b + \rho \cdot \zeta) \cdot Y^1 - (\beta^{23} + \gamma^{23} \cdot X_b + \rho \cdot \zeta) \cdot Y^3, \\ \mathcal{P}^3 &= \zeta^3 \cdot X_a - (\lambda^3 + \delta^3 \cdot X_b) \cdot Y^3 - (\beta^{31} + \gamma^{31} \cdot X_b + \rho \cdot \zeta) \cdot Y^1 - (\beta^{32} + \gamma^{32} \cdot X_b + \rho \cdot \zeta) \cdot Y^2. \end{aligned}$$

Where  $\zeta$  is unobserved by the econometrician but perfectly observed by all three firms and  $\rho$  is a parameter that measures the importance of  $\zeta$  as a payoff shifter. Since the latter is a common

component of players' private shocks,  $\rho$  provides also a measure of the correlation between players' private shocks. We generate  $\zeta \sim U[0, 1]$ , independent of all other covariates in the model. With this to the demand system, it is no longer possible to express the strategic index  $\eta^p$  as a function only of observables. There is also correlation in payoff shocks unobserved by the econometrician, violating Assumption 3. As a result of these violations, the main result in Theorem 1 is no longer valid. For finite samples, our a-priori conjectures are the following,

- (a) The asymptotic predictions of our approach should retain some of their validity for small values of  $\rho$  (i.e, small correlation between private shocks).
- (b) For increasingly larger values of  $\rho$  (i.e, larger correlation between players' private shocks), our approach has the potential of producing empty confidence sets, which would in turn reveal that the model is misspecified.

To investigate the validity of our conjectures we repeat two of the exercises done previously in Tables 3 and 4 in the paper. We generated 1,000 samples of size  $n = 2,000$  and we tried different values of  $\rho$ . For each one we computed the frequency with which our CS included  $\theta_0$  and excluded  $\theta_b$  (as defined above). According to our conjectures, our approach should still lead us to reject the fake value  $\theta_b$  and, for increasingly larger values of  $\rho$ , it should also lead us to reject  $\theta_0$ . Our results are summarized in Tables 2 and 3 and are directly comparable to those in Table 3 and in the second panel in Table 4.

Table 2: Observed frequency with which  $\theta_0$  was INCLUDED in our CS when Assumptions 1 and 3 are violated

Value of $\rho$	Target coverage: 95% ( $c_{1-\alpha} = 1.645$ )	Target coverage: 99% ( $c_{1-\alpha} = 2.33$ )	95 <sup>th</sup> percentile observed value of test-statistic	Maximum observed value of test-statistic
$\rho = 0.25$	95.2%	100%	1.627	2.216
$\rho = 0.50$	83.3%	96.2%	2.142	3.398
$\rho = 1$	48.3	66.3%	4.686	7.638
$\rho = 2$	30%	40.9%	8.082	13.965

1,000 simulated samples of size  $n = 2,000$ .

The results in Tables 2 and 3 are in line with our previous conjectures. Firstly, the ability of our approach to reject the false value  $\theta_b$  is not affected by misspecification (if anything, the propensity

Table 3: Observed frequency with which  $\theta_b$  was EXCLUDED from our CS when Assumptions 1 and 3 are violated

Value of $\rho$	Target coverage: 95% ( $c_{1-\alpha} = 1.645$ )	Target coverage: 99% ( $c_{1-\alpha} = 2.33$ )	95 <sup>th</sup> percentile observed value of test-statistic	Maximum observed value of test-statistic
$\rho = 0.25$	77.1%	70.4%	14.693	22.551
$\rho = 0.50$	71.5%	64.3%	13.292	24.154
$\rho = 1$	79.6%	74.5%	15.951	21.753
$\rho = 2$	85.4%	81.4%	20.797	29.267

1,000 simulated samples of size  $n = 2,000$ .

to reject  $\theta_b$  is increased by the misspecification). Regarding the inclusion of  $\theta_0$  in our CS, when our model is only slightly misspecified and  $\rho$  is relatively small our results remain very much in line with the asymptotic predictions. As the model becomes increasingly misspecified and  $\rho$  is larger, our procedure rejects  $\theta_0$  because Theorem 1 is no longer true.

## References

- Aradillas-López, A., A. Gandhi, and D. Quint (2013). Identification and inference in ascending auctions with correlated private values. Forthcoming, *Econometrica*.
- Nolan, D. and D. Pollard (1987). U-processes: Rates of convergence. *Annals of Statistics* 15, 780–799.
- Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* 57(5), 1027–1057.
- Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics.
- Romano, J. (2004). On non-parametric testing, the uniform behaviour of the t-test and related problems. *Scandinavian Journal of Statistics* 31, 567–584.
- Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley. New York, NY.
- Sherman, R. (1994). Maximal inequalities for degenerate u-processes with applications to optimization estimators. *Annals of Statistics* 22, 439–459.
- Silverman, B. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall/CRC.