# Appendix Supplement for the paper "Pairwise-Difference Estimation of Incomplete Information Games"

Andres Aradillas-Lopez*

May, 2010

**Abstract**

We present a step-by-step proof of Theorem 2 in the paper *Pairwise-Difference Estimation of Incomplete Information Games.* We also describe the bandwidth-selection criterion used in the Monte Carlo experiments.

## 1 Step-by-step proof of Theorem 2

The proof relies on showing that the statistic $U_{p_N}$ can be expressed as

$$U_{p_N} = \binom{N}{2}^{-1} \sum_{i<j} \frac{\varepsilon_{p_i}\varepsilon_{p_j}\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right) + \alpha_{p_0}\left(\mathcal{U}_{p_N}^{B_I} + \mathcal{U}_{p_N}^{B_{II}}\right) + \mathcal{U}_{p_N}^{B_{III}} + \mathcal{U}_{p_N}^{B_{IV}} + \mathcal{R}_{p_N}, \qquad \text{(S-1)}$$

where $\varepsilon_{p_i} \equiv Y_{pi} - F_p(t_{pi})$, $N\widetilde{b}^{L/2}\mathcal{R}_{p_N} \xrightarrow{p} 0$, $N\widetilde{h}_c^{3/2}\mathcal{U}_{p_N}^{B_I} \xrightarrow{d} \mathcal{N}\big(0, V_p^{B_I}\big)$, $N\widetilde{h}_b^{L/2}\mathcal{U}_{p_N}^{B_{II}} \xrightarrow{d} \mathcal{N}\big(0, V_p^{B_{II}}\big)$, $N\widetilde{h}_c^{1/2}\mathcal{U}_{p_N}^{B_{IV}} \xrightarrow{d} \mathcal{N}\big(0, V_p^{B_{IV}}\big)$, and $N\mathcal{U}_{p_N}^{B_{III}} \xrightarrow{d} 3\mathcal{Y}$, where $\mathcal{Y} = O_p(1)$. From here, the bandwidth conditions in Assumption (B2) will yield

$$N\widetilde{b}^{L/2}U_{p_N} = N\widetilde{b}^{L/2}\binom{N}{2}^{-1} \sum_{i<j} \frac{\varepsilon_{p_i}\varepsilon_{p_j}\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right) + \xi_N, \quad \text{where } \xi_N \xrightarrow{p} 0.$$

And the main result in Theorem 2 will follow from Theorem 1 in Hall (1984). Without loss of generality we will focus on $p = 1$, with all proofs for $p = 2$ following identical steps. Let

$$t_{1i} = W_{1i} + V_{1i}'\gamma_1 + \alpha_1\mu_{2i} \equiv W_{1i} + Z_{1i}'\theta_1, \quad E[Y_1|t_{1i}] \equiv F_1(t_{1i});\ E[Y_p|X_i] \equiv \mu_p(X_i),\ p = 1, 2.$$

*Department of Economics, University of Wisconsin, Madison WI 53706; aaradill@ssc.wisc.edu .Financial support from the National Science Foundation Grant No. SES−0718409 and the Gregory Chow Econometrics Research Program at Princeton University are gratefully acknowledged.

The density of $t_{1i}$ is denoted by $f_{t_1}(\cdot)$. We will use $\mu_p(X_i)$ and $\mu_{pi}$ interchangeably. If the model is correct, then $\mu_1(X_i) = F_1(t_{1i})$. From now on, we will use the notation $\Delta\zeta_{ij} \equiv \zeta_i - \zeta_j$. Let

$$\widetilde{\mu}_{pi} = \frac{1}{N\widetilde{h}_b^L}\sum_{j=1}^{N} Y_{pi}\mathcal{H}_b\left(\frac{\Delta X_{ji}}{\widetilde{h}_b}\right) \Bigg/ \frac{1}{N\widetilde{h}_b^L}\sum_{j=1}^{N}\mathcal{H}_b\left(\frac{\Delta X_{ji}}{\widetilde{h}_b}\right); \quad \widehat{t}_{1i} \equiv \underbrace{W_{1i} + V_{1i}'\widehat{\gamma}_1 + \widehat{\alpha}_1\widetilde{\mu}_{2i}}_{\equiv W_{1i} + \widetilde{Z}_{1i}'\widehat{\theta}_1}.$$

Define now

$$T_{1_N}(\widehat{t}_{1j}) = \frac{1}{N\widetilde{h}_c}\sum_{i=1}^{N}\mathcal{H}_c\left(\frac{\Delta\widehat{t}_{1ij}}{\widetilde{h}_c}\right)\varphi(X_i), \quad S_{1_N}(\widehat{t}_{1j}) = \frac{1}{N\widetilde{h}_c}\sum_{i=1}^{N}Y_{1i}\mathcal{H}_c\left(\frac{\Delta\widehat{t}_{1ij}}{\widetilde{h}_c}\right)\varphi(X_i), \quad \widehat{F}_1(\widehat{t}_{1j}) = \frac{S_{1_N}(\widehat{t}_{1j})}{T_{1_N}(\widehat{t}_{1j})},$$
$$T_{1_N}(t_{1j}) = \frac{1}{N\widetilde{h}_c}\sum_{i=1}^{N}\mathcal{H}_c\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right)\varphi(X_i), \quad S_{1_N}(t_{1j}) = \frac{1}{N\widetilde{h}_c}\sum_{i=1}^{N}Y_{1i}\mathcal{H}_c\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right)\varphi(X_i), \quad \widehat{F}_1(t_{1j}) = \frac{S_{1_N}(t_{1j})}{T_{1_N}(t_{1j})}$$

First, we will find asymptotic linear representations for $T_{1_N}(\widehat{t}_{1j}) - T_{1_N}(t_{1j})$ and $S_{1_N}(\widehat{t}_{1j}) - S_{1_N}(t_{1j})$ conditional on $X_j$. Note that $\Delta\widehat{t}_{1ij} - \Delta t_{1ij} = \big(Z_{1i} - Z_{1j}\big)'\big(\widehat{\theta}_1 - \theta_1\big) + \big(\widehat{\alpha}_1 - \alpha_1\big)\big[(\widetilde{\mu}_{2i} - \mu_{2i}) - (\widetilde{\mu}_{2j} - \mu_{2j})\big] + \alpha_1\big[(\widetilde{\mu}_{2i} - \mu_{2i}) - (\widetilde{\mu}_{2j} - \mu_{2j})\big]$; where, from our previous notation, $\big(Z_{1i} - Z_{1j}\big)'\big(\widehat{\theta}_1 - \theta_1\big) \equiv \big(V_{1i} - V_{1j}\big)'\big(\widehat{\gamma}_1 - \gamma_1\big) + \big(\mu_{2i} - \mu_{2j}\big)\big(\widehat{\alpha}_1 - \alpha_1\big)$. Given our assumptions, we can use Theorem A-1 in Aradillas-Lopez (2008) or Lemma 3 in Collomb and Hardle (1986) to show that for any compact set $\mathcal{X} \in \mathbb{S}(X)$ and any $\delta > 0$,

$$\sup_{x_i, x_j \in \mathcal{X}}\big|\Delta\widehat{t}_{1ij} - \Delta t_{1ij}\big| = O_p(N^{-1/2}) + O_p(N^{\delta-1}\widetilde{h}_b^{-L})^{1/2}\big[1 + O_p(N^{-1/2})\big] = O_p(N^{\delta-1}\widetilde{h}_b^{-L})^{1/2}.$$

Let $\Delta Z_{1ij} \equiv Z_{1i} - Z_{1j}$. This yields

$$\begin{aligned}
&\underbrace{\frac{1}{N\widetilde{h}_c}\sum_{i=1}^{N}\mathcal{H}_c\left(\frac{\Delta\widehat{t}_{1ij}}{\widetilde{h}_c}\right)\varphi(X_i) - \frac{1}{N\widetilde{h}_c}\sum_{i=1}^{N}\mathcal{H}_c\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right)\varphi(X_i)}_{=T_{1N}(\widehat{t}_{1j}) - T_{1N}(t_{1j})} = \underbrace{\frac{1}{N\widetilde{h}_c^2}\sum_{i=1}^{N}\mathcal{H}_c^{(1)}\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right)\varphi(X_i)\Delta Z_{1ij}'\big(\widehat{\theta}_1 - \theta_1\big)}_{\equiv A_{1N}(X_j)} \\
&+ \alpha_1\underbrace{\frac{1}{N\widetilde{h}_c^2}\sum_{i=1}^{N}\mathcal{H}_c^{(1)}\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right)\Big[(\widetilde{\mu}_{2i} - \mu_{2i}) - (\widetilde{\mu}_{2j} - \mu_{2j})\Big]\varphi(X_i)}_{\equiv B_{1N}(X_j)} \qquad \text{(S-2)} \\
&+ \underbrace{\frac{1}{N\widetilde{h}_c^2}\sum_{i=1}^{N}\mathcal{H}_c^{(1)}\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right)\big(\widehat{\alpha}_1 - \alpha_1\big)\Big[(\widetilde{\mu}_{2i} - \mu_{2i}) - (\widetilde{\mu}_{2j} - \mu_{2j})\Big]\varphi(X_i)}_{\equiv C_{1N}(X_j)} + \xi_N^a(X_j),
\end{aligned}$$

where $\sup_{x\in\mathcal{X}}\left|C_{1N}(x)\right| = O_p(N^{\delta/2-1}\widetilde{h}_b^{-L/2}\widetilde{h}_c^{-2})$ and $\sup_{x\in\mathcal{X}}\left|\xi_N^a(x)\right| = O_p(N^{\delta-1}\widetilde{h}_b^{-L}\widetilde{h}_c^{-3})$ for any compact set $\mathcal{X}\in\mathbb{S}(X)$ and any $\delta>0$.

$$A_{1N}(X_j) = \frac{1}{N^2\widetilde{h}_c^2}\sum_{\ell=1}^N\sum_{i=1}^N \mathcal{H}_c^{(1)}\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right)\varphi(X_i)\Delta Z_{1ij}'\psi_\ell^{\theta_1} + \xi_N^b(X_j)$$
$$B_{1N}(X_j) = \frac{1}{N^2\widetilde{h}_c^2\widetilde{h}_b^L}\sum_{\ell=1}^N\sum_{i=1}^N \mathcal{H}_c^{(1)}\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right)\varphi(X_i)\left[\frac{\left[Y_{2\ell}-\mu_{2i}\right]}{f_x(X_i)}\mathcal{H}_b\left(\frac{\Delta X_{\ell i}}{\widetilde{h}_b}\right) - \frac{\left[Y_{2\ell}-\mu_{2j}\right]}{f_x(X_j)}\mathcal{H}_b\left(\frac{\Delta X_{\ell j}}{\widetilde{h}_b}\right)\right] + \xi_N^c(X_j), \tag{S-3}$$

where $\sup_{x\in\mathcal{X}}\left|\xi_N^b(x)\right| = o_p(N^{-1/2}\widetilde{h}_c^{-2})$ and $\sup_{x\in\mathcal{X}}\left|\xi_N^c(x)\right| = O_p(N^{\delta-1}\widetilde{h}_b^{-L}\widetilde{h}_c^{-2})$. Grouping terms, we can reexpress $B_{1N}(X_j)$ as

$$B_{1N}(X_j) = \frac{1}{N^2\widetilde{h}_c^2\widetilde{h}_b^L}\sum_{i<\ell}\Bigg\{\mathcal{H}_c^{(1)}\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right)\varphi(X_i)\left[\frac{\left[Y_{2\ell}-\mu_{2i}\right]}{f_x(X_i)}\mathcal{H}_b\left(\frac{\Delta X_{\ell i}}{\widetilde{h}_b}\right) - \frac{\left[Y_{2\ell}-\mu_{2j}\right]}{f_x(X_j)}\mathcal{H}_b\left(\frac{\Delta X_{\ell j}}{\widetilde{h}_b}\right)\right]$$
$$+ \mathcal{H}_c^{(1)}\left(\frac{\Delta t_{1\ell j}}{\widetilde{h}_c}\right)\varphi(X_\ell)\left[\frac{\left[Y_{2i}-\mu_{2\ell}\right]}{f_x(X_\ell)}\mathcal{H}_b\left(\frac{\Delta X_{i\ell}}{\widetilde{h}_b}\right) - \frac{\left[Y_{2i}-\mu_{2j}\right]}{f_x(X_j)}\mathcal{H}_b\left(\frac{\Delta X_{ij}}{\widetilde{h}_b}\right)\right]\Bigg\} + O_p(N^{-1}\widetilde{h}_c^{-2}\widetilde{h}_b^{-L}) + \widetilde{\xi}_N^c(X_j),$$

where $\sup_{x\in\mathcal{X}}\left|\widetilde{\xi}_N^c(x)\right| = O_p(N^{-1}\widetilde{h}_c^{-2}\widetilde{h}_b^{-L}) + O_p(N^{\delta-1}\widetilde{h}_b^{-L}\widetilde{h}_c^{-2}) = O_p(N^{\delta-1}\widetilde{h}_b^{-L}\widetilde{h}_c^{-2})$ for any $\delta>0$ and any compact set $\mathcal{X}\in\text{int}\big(\mathbb{S}(X)\big)$. The first term in the above equation is a symmetric, third order U-statistic which satisfies the assumptions of Lemma A.3 in Ahn and Powell (1993). Given our assumptions, taking the projection of this U-statistic *conditional on* $X_j$ yields

$$B_{1N}(X_j) = \frac{1}{N\widetilde{h}_c^2}\sum_{i=1}^N\left[Y_{2i}-\mu_{2i}\right]\varphi(X_i)\mathcal{H}_c^{(1)}\left(\frac{\Delta t_{ij}}{\widetilde{h}_c}\right) + R_1(t_{1j})\frac{1}{N\widetilde{h}_b^L}\sum_{i=1}^N\frac{\left[Y_{2i}-\mu_{2j}\right]}{f_x(X_j)}\mathcal{H}_b\left(\frac{\Delta X_{ij}}{\widetilde{h}_b}\right) + \overline{\xi}_N^c(X_j)$$

where $R_1(t_1)\equiv\nabla_{t_1}\Big(E\big[\varphi(X)\big|t_1\big]f_{t_1}(t_1)\Big)\in\mathbb{R}$ and

$$\sup_{x\in\mathcal{X}}\left|\overline{\xi}_N^c(x)\right| = O_p\left(\frac{\widetilde{h}_c^{\overline{\mathcal{M}}}}{\widetilde{h}_b^L}\right) + O_p\left(\frac{\widetilde{h}_b^{\overline{\mathcal{M}}}}{\widetilde{h}_c^{2+\overline{\mathcal{M}}}}\right) + O_p(N^{\delta-1}\widetilde{h}_b^{-L}\widetilde{h}_c^{-2}) + o_p(N^{-1/2})$$

Let us move on to $A_{1N}(X_j)$, grouping terms we have

$$A_{1N}(X_j) = \frac{1}{N\widetilde{h}_c^2}\sum_{i<\ell}\left\{\mathcal{H}_c^{(1)}\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right)\varphi(X_i)\Delta Z_{1ij}'\psi_\ell^{\theta_1} + \mathcal{H}_c^{(1)}\left(\frac{\Delta t_{1\ell j}}{\widetilde{h}_c}\right)\varphi(X_\ell)\Delta Z_{1\ell j}'\psi_i^{\theta_1}\right\} + \widetilde{\xi}_N^b(X_j),$$

where $\sup_{x\in\mathcal{X}}\left|\widetilde{\xi}_N^b(x)\right| = O_p(N^{-1}\widetilde{h}_c^{-2}) + o_p(N^{-1/2}\widetilde{h}_c^{-2}) = o_p(N^{-1/2}\widetilde{h}_c^{-2})$ for any compact set $\mathcal{X}\in\text{int}(\mathbb{S}(X))$. Taking the projection of the third-order U-statistic conditional on $X_j$ yields

$$A_{1N}(X_j) = \big(R_1(t_{1j})Z_{1j} - Q_1(t_{1j})\big)'\frac{1}{N}\sum_{i=1}^N\psi_i^{\theta_1} + \overline{\xi}_N^b(X_j), \tag{S-4}$$

where $Q_1(t_1)\equiv\nabla_{t_1}\Big(E\big[\varphi(X)Z_1\big|t_1\big]f_{t_1}(t_1)\Big)\in\mathbb{R}^{\dim(Z_1)}$, and $\sup_{x\in\mathcal{X}}\left|\widetilde{\xi}_N^b(x)\right| = O_p(\widetilde{h}_c^{\overline{\mathcal{M}}}) + o_p(N^{-1/2}\widetilde{h}_c^{-2})$ for any compact set $\mathcal{X}\in\text{int}(\mathbb{S}(X))$. Eqs. (S-2) – (S-4) yield a linear representation result for $T_{1N}(\widehat{t}_{1j}) - T_{1N}(t_{1j})$

conditional on $X_j$:

$$\begin{aligned} T_{1N}(\widehat{t}_{1j}) - T_{1N}(t_{1j}) =& \big(R_1(t_{1j})Z_{1j} - Q_1(t_{1j})\big)' \frac{1}{N}\sum_{i=1}^{N}\psi_i^{\theta_1} + \frac{\alpha_1}{N\widetilde{h}_c^2}\sum_{i=1}^{N}\big[Y_{2i} - \mu_{2i}\big]\varphi(X_i)\mathcal{H}_c^{(1)}\left(\frac{\Delta t_{ij}}{\widetilde{h}_c}\right) \\ &+ R_1(t_{1j})\frac{\alpha_1}{N\widetilde{h}_b^L}\sum_{i=1}^{N}\frac{[Y_{2i} - \mu_{2j}]}{f_x(X_j)}\mathcal{H}_b\left(\frac{\Delta X_{ij}}{\widetilde{h}_b}\right) + \xi_N^{T_1}(X_j), \end{aligned} \tag{S-5}$$

where

$$\sup_{x\in\mathcal{X}}\big|\xi_N^{T_1}(x)\big| = O_p\left(\frac{\widetilde{h}_c^{\overline{\mathcal{M}}}}{\widetilde{h}_b^L}\right) + O_p\left(\frac{\widetilde{h}_b^{\overline{\mathcal{M}}}}{\widetilde{h}_c^{2+\overline{\mathcal{M}}}}\right) + O_p(N^{\delta-1}\widetilde{h}_b^{-L}\widetilde{h}_c^{-3}) + o_p(N^{-1/2}\widetilde{h}_c^{-2}) \tag{S-6}$$

for any compact set $\mathcal{X} \in \text{int}(\mathbb{S}(X))$ and any $\delta > 0$. Following analogous steps, we can arrive at an expression for $S_{1N}(\widehat{t}_{1j}) - S_{1N}(t_{1j})$ equivalent to (S-5). Define[1]

$$\widetilde{R}_1(t_1) = \nabla_{t_1}\Big(E[\varphi(X)|t_1]F_1(t_1)f_{t_1}(t_1)\Big), \quad \widetilde{Q}_1(t_1) = \nabla_{t_1}\Big(E[\varphi(X)Z_1|t_1]F_1(t_1)f_{t_1}(t_1)\Big).$$

We have

$$\begin{aligned} S_{1N}(\widehat{t}_{1j}) - S_{1N}(t_{1j}) =& \big(\widetilde{R}_1(t_{1j})Z_{1j} - \widetilde{Q}_1(t_{1j})\big)' \frac{1}{N}\sum_{i=1}^{N}\psi_i^{\theta_1} + \frac{\alpha_1}{N\widetilde{h}_c^2}\sum_{i=1}^{N}\big[Y_{2i} - \mu_{2i}\big]F_1(t_{1i})\varphi(X_i)\mathcal{H}_c^{(1)}\left(\frac{\Delta t_{ij}}{\widetilde{h}_c}\right) \\ &+ \widetilde{R}_1(t_{1j})\frac{\alpha_1}{N\widetilde{h}_b^L}\sum_{i=1}^{N}\frac{[Y_{2i} - \mu_{2j}]}{f_x(X_j)}\mathcal{H}_b\left(\frac{\Delta X_{ij}}{\widetilde{h}_b}\right) + \xi_N^{S_1}(X_j), \end{aligned}$$

where $\xi_N^{S_1}(X_j)$ is of the same order of magnitude as $\xi_N^{T_1}(X_j)$ (see (S-6)). If the model is correct, $E[Y_1|X, t_1] = F_1(t_1)$, this yields

$$S_{1N}(t_{1j}) \xrightarrow{p} \underbrace{F_1(t_{1j})E[\varphi(X)|t_{1j}]f_{t_1}(t_{1j})}_{\equiv S_1(t_{1j})},\ T_{1N}(t_{1j}) \xrightarrow{p} \underbrace{E[\varphi(X)|t_{1j}]f_{t_1}(t_{1j})}_{\equiv T_1(t_{1j})},\ \frac{S_{1N}(t_{1j})}{T_{1N}(t_{1j})} \equiv \widehat{F}_1(t_{1j}) \xrightarrow{p} F_1(t_{1j}).$$

Given our assumptions, we have in fact

$$\sup_{x_j\in\mathcal{X}}\big|\widehat{F}_1(t_{1j}) - F_1(t_{1j})\big| = O_p\big(N^{(\delta-1)/2}\widetilde{h}_c^{-1/2}\big) \quad \text{for any compact } \ \mathcal{X} \in \text{int}(\mathbb{S}(X)) \text{ and any } \delta > 0. \tag{S-7}$$

Using (S-5) – (S-7), adding and subtracting $\widehat{F}_1(t_{1j})$, we obtain

$$\begin{aligned} \widehat{F}_1(\widehat{t}_{1j}) - F_1(t_{1j}) = \frac{1}{T_1(t_{1j})}\big[S_{1N}(\widehat{t}_{1j}) - S_{1N}(t_{1j})\big] - \frac{F_1(t_{1j})}{T_1(t_{1j})}\big[T_{1N}(\widehat{t}_{1j}) - T_{1N}(t_{1j})\big] \\ + \frac{1}{T_1(t_{1j})}\big[S_{1N}(t_{1j}) - F_1(t_{1j})T_{1N}(t_{1j})\big] + \xi_N^{F_1}(X_j). \end{aligned}$$

[1] Recall that $F_1(t_{1i}) = E[Y_{1i}|t_{1i}]$, and if the model is correct then $\mu_1(X_i) = F_1(t_{1i})$.

Using our previous results, simplifying the above equation yields the following asymptotic linear representation for $\widehat{F}_1(\widehat{t}_{1j}) - F_1(t_{1j})$ conditional on $X_j$,

$$\begin{aligned}
&\widehat{F}_1(\widehat{t}_{1j}) - F_1(t_{1j}) = \frac{F_1^{(1)}(t_{1j})}{E[\varphi(X)|t_{1j}]}\big(\widetilde{R}_1(t_{1j})Z_{1j} - \widetilde{Q}_1(t_{1j})\big)'\frac{1}{N}\sum_{i=1}^{N}\psi_i^{\theta_1} \\
&+ \frac{\alpha_1}{N\widetilde{h}_c^2}\sum_{i=1}^{N}\frac{[Y_{2i}-\mu_{2i}][\mu_{1i}-\mu_{1j}]}{T_1(t_{1j})}\varphi(X_i)\mathcal{H}_c^{(1)}\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right) + F_1^{(1)}(t_{1j})\frac{\alpha_1}{N\widetilde{h}_b^L}\sum_{i=1}^{N}\frac{[Y_{2i}-\mu_{2j}]}{f_x(X_j)}\mathcal{H}_b\left(\frac{\Delta X_{ij}}{\widetilde{h}_b}\right) \\
&+ \frac{1}{N\widetilde{h}_c}\sum_{i=1}^{N}\frac{[Y_{1i}-F_1(t_{1j})]}{T_1(t_{1j})}\varphi(X_i)\mathcal{H}_c\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right) + \xi_N^{F_1}(X_j), \quad \text{where} \\
&\sup_{x\in\mathcal{X}}\big|\xi_N^{F_1}(x)\big| = O_p\left(\Big[(\widetilde{h}_c^{\overline{\mathcal{M}}}\widetilde{h}_b^{-L}) + (\widetilde{h}_b^{\overline{\mathcal{M}}}\widetilde{h}_c^{-(2+\overline{\mathcal{M}})}) + (N^{\delta-1}\widetilde{h}_b^{-L}\widetilde{h}_c^{-3}) + (N^{-1/2}\widetilde{h}_c^{-2}) + \big(N^{(\delta-1)/2}\widetilde{h}_c^{-1/2}\big)\Big]^2\right) \\
&\text{for any compact } \mathcal{X} \in \text{int}\big(\mathbb{S}(X)\big) \text{ and any } \delta > 0.
\end{aligned} \tag{S-8}$$

Given our assumptions, this implies in particular that, conditional on $X_j$ [2]

$$\begin{aligned}
\sup_{x_j\in\mathcal{X}}\big|\widehat{F}_1(\widehat{t}_{1j}) - F_1(t_{1j})\big| &= O_p(N^{-1/2}\widetilde{h}_c^{-1}) + O_p(N^{(\delta-1)/2}\widetilde{h}_b^{-L/2}) + O_p(N^{(\delta-1)/2}\widetilde{h}_c^{-1/2}) + O_p(\widetilde{h}_c^{\overline{\mathcal{M}}}) + O_p(\widetilde{h}_b^{\overline{\mathcal{M}}}) \\
&+ O_p\left(\Big[(\widetilde{h}_c^{\overline{\mathcal{M}}}\widetilde{h}_b^{-L}) + (\widetilde{h}_b^{\overline{\mathcal{M}}}\widetilde{h}_c^{-(2+\overline{\mathcal{M}})}) + (N^{\delta-1}\widetilde{h}_b^{-L}\widetilde{h}_c^{-3}) + (N^{-1/2}\widetilde{h}_c^{-2}) + \big(N^{(\delta-1)/2}\widetilde{h}_c^{-1/2}\big)\Big]^2\right)
\end{aligned} \tag{S-9}$$

for any compact $\mathcal{X} \in \text{int}\big(\mathbb{S}(X)\big)$ and any $\delta > 0$.

Denote

$$\varepsilon_{1j} \equiv Y_{1j} - F_1(t_{1j}), \quad \widehat{\varepsilon}_{1j} \equiv Y_{1j} - \widehat{F}_1(\widehat{t}_{1j}), \quad \widehat{\nu}_1(X_j) \equiv \widehat{F}_1(\widehat{t}_{1j}) - F_1(t_{1j}).$$

Note that $\widehat{\varepsilon}_{1j} - \varepsilon_{1j} = -\widehat{\nu}_1(X_j)$. Now, let us go back to the statistic

$$\begin{aligned}
U_{1_N} &= \binom{N}{2}^{-1}\sum_{i<j}\frac{\widehat{\varepsilon}_{1j}\widehat{\varepsilon}_{1i}\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right) = \underbrace{\binom{N}{2}^{-1}\sum_{i<j}\frac{\varepsilon_{1j}\varepsilon_{1i}\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)}_{\equiv U_{1_N}^A} \\
&- \underbrace{\binom{N}{2}^{-1}\sum_{i<j}\frac{\big[\varepsilon_{1j}\widehat{\nu}_1(X_i) + \varepsilon_{1i}\widehat{\nu}_1(X_j)\big]\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)}_{\equiv U_{1_N}^B} \\
&+ \underbrace{\binom{N}{2}^{-1}\sum_{i<j}\frac{\widehat{\nu}_1(X_j)\widehat{\nu}_1(X_i)\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)}_{\equiv U_N^C}
\end{aligned} \tag{S-10}$$

[2]Notice that the expectation of the first two terms in the linear representation are exactly equal to zero conditional on $X_j$.

We deal first with $U_N^C$. Recall that the trimming function $\overline{\phi}(\cdot)$ is nonzero only in a compact set $\mathcal{X} \in \operatorname{int}(\mathbb{S}(X))$.

$$U_N^C \leq \sup_{x\in\mathcal{X}} \left|\widehat{\nu}_1(x)\right|^2 \overline{\overline{\phi}}^2 \cdot \underbrace{\binom{N}{2}^{-1} \sum_{i<j} \frac{1}{\widetilde{b}^L}\left|\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)\right|}_{=O_p(1)} = O_p\left(\sup_{x\in\mathcal{X}}\left|\widehat{\nu}_1(x)\right|^2\right) \tag{S-11}$$

The order of magnitude of $\sup_{x\in\mathcal{X}}\left|\widehat{\nu}_1(x)\right|$ is given by (S-9). We now move on to $U_{1_N}^B$. Let

$$\begin{gathered}
\mathcal{W}_i \equiv \left(Y_{1i}, Y_{2i}, \varepsilon_{1i}, \varepsilon_{2i}, X_i', \psi_i^{\theta_1'}\right)'; \\
\overline{\Delta}(X_i, X_j) = \frac{\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ji}}{\widetilde{b}}\right); \quad \overline{\delta}^I(\mathcal{W}_i, \mathcal{W}_j) = \frac{1}{\widetilde{h}_c^2}\frac{\varepsilon_{2j}[\mu_{1j}-\mu_{1i}]}{E[\varphi(X)|t_{1i}]f_{t_1}(t_{1i})}\varphi(X_j)\mathcal{H}_c^{(1)}\left(\frac{\Delta t_{1ji}}{\widetilde{h}_c}\right); \\
\overline{\delta}^{II}(\mathcal{W}_i, \mathcal{W}_j) = \frac{1}{\widetilde{h}_b^L}\frac{[Y_{2j}-\mu_{2i}]}{f_x(X_i)}\mathcal{H}_b\left(\frac{\Delta X_{ji}}{\widetilde{h}_b}\right)F_1^{(1)}(t_{1i}); \quad \overline{\delta}^{III}(\mathcal{W}_i, \mathcal{W}_j) = \frac{F_1^{(1)}(t_{1i})\left(\widetilde{R}_1(t_{1i})Z_{1i} - \widetilde{Q}_1(t_{1i})\right)'\psi_j^{\theta_1}}{E[\varphi(X)|t_{1i}]}; \\
\overline{\delta}^{IV}(\mathcal{W}_i, \mathcal{W}_j) = \frac{1}{\widetilde{h}_c}\frac{[Y_{1j}-F_1(t_{1i})]\varphi(X_j)}{E[\varphi(X)|t_{1i}]f_{t_1}(t_{1i})}\mathcal{H}_c\left(\frac{\Delta t_{ji}}{\widetilde{h}_c}\right).
\end{gathered}$$

Note that $\overline{\Delta}(X_i, X_j)$ is symmetric in its arguments. Now, for $\mathcal{J} \in \{I, II, III, IV\}$ let

$$\begin{aligned}
\Lambda^{\mathcal{J}}(\mathcal{W}_i, \mathcal{W}_j, \mathcal{W}_k) = {} & \varepsilon_{1i}\left[\overline{\delta}^{\mathcal{J}}(\mathcal{W}_j, \mathcal{W}_k)\overline{\Delta}(X_i, X_j) + \overline{\delta}^{\mathcal{J}}(\mathcal{W}_k, \mathcal{W}_j)\overline{\Delta}(X_i, X_k)\right] + \varepsilon_{1j}\left[\overline{\delta}^{\mathcal{J}}(\mathcal{W}_i, \mathcal{W}_k)\overline{\Delta}(X_i, X_j)\right. \\
& \left. + \overline{\delta}^{\mathcal{J}}(\mathcal{W}_k, \mathcal{W}_i)\overline{\Delta}(X_j, X_k)\right] + \varepsilon_{1k}\left[\overline{\delta}^{\mathcal{J}}(\mathcal{W}_i, \mathcal{W}_j)\overline{\Delta}(X_i, X_k) + \overline{\delta}^{\mathcal{J}}(\mathcal{W}_j, \mathcal{W}_i)\overline{\Delta}(X_j, X_k)\right]; \\
\Upsilon^{\mathcal{J}}(\mathcal{W}_i, \mathcal{W}_j) = {} & \left(\varepsilon_{1i}\left[\overline{\delta}^{\mathcal{J}}(\mathcal{W}_j, \mathcal{W}_i) + \overline{\delta}^{\mathcal{J}}(\mathcal{W}_j, \mathcal{W}_j)\right] + \varepsilon_{1j}\left[\overline{\delta}^{\mathcal{J}}(\mathcal{W}_i, \mathcal{W}_j) + \overline{\delta}^{\mathcal{J}}(\mathcal{W}_i, \mathcal{W}_i)\right]\right)\overline{\Delta}(X_i, X_j).
\end{aligned}$$

Note that these two objects are symmetric in their arguments. In addition, if the model is correct, $E\left[\Lambda^{\mathcal{J}}(\mathcal{W}_i, \mathcal{W}_j, \mathcal{W}_k)\right] = 0$ because in this case, $E[\varepsilon_{1a}|X_a, \mathcal{W}_b, \mathcal{W}_c] = 0$ for $a, b, c \in \{i, j, k\}$. Let

$$\mathcal{U}_{1_N}^{B_{\mathcal{J}}} = \frac{1}{3}\frac{(N-2)}{N}\binom{N}{3}^{-1}\sum_{i<j<k}\Lambda^{\mathcal{J}}(\mathcal{W}_i, \mathcal{W}_j, \mathcal{W}_k); \quad \mathcal{V}_{1_N}^{B_{\mathcal{J}}} = \underbrace{\frac{1}{N}\underbrace{\binom{N}{2}^{-1}\sum_{i<j}\Upsilon^{\mathcal{J}}(\mathcal{W}_i, \mathcal{W}_j)}_{=O_p(1)}}_{=O_p(N^{-1})}.$$

Going back to the $U_{1_N}^B$ component in (S-10), we can express it as

$$\begin{aligned}
U_{1_N}^B &= \alpha_1\left(\mathcal{U}_{1_N}^{B_I} + \mathcal{U}_{1_N}^{B_{II}}\right) + \mathcal{U}_{1_N}^{B_{III}} + \mathcal{U}_{1_N}^{B_{IV}} + \alpha_1\left(\mathcal{V}_{1_N}^{B_I} + \mathcal{V}_{1_N}^{B_{II}}\right) + \mathcal{V}_{1_N}^{B_{III}} + \mathcal{V}_{1_N}^{B_{IV}} + R_N^B \\
&= \alpha_1\left(\mathcal{U}_{1_N}^{B_I} + \mathcal{U}_{1_N}^{B_{II}}\right) + \mathcal{U}_{1_N}^{B_{III}} + \mathcal{U}_{1_N}^{B_{IV}} + O_p(N^{-1}) + R_N^B, \quad \text{where} \\
R_N^B &\leq \left(\sup_{x\in\mathcal{X}}\left|\xi_N^{F_1}(x)\right| + \sup_{x\in\mathcal{X}}\left|\xi_N^{F_1}(x)\right|\right)\overline{\overline{\phi}}^2 \cdot \underbrace{\binom{N}{2}^{-1}\sum_{i<j}\left(|\varepsilon_{1i}| + |\varepsilon_{1j}|\right)\frac{1}{\widetilde{b}^L}\left|\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)\right|}_{=O_p(1)}.
\end{aligned} \tag{S-12}$$

The order of magnitude of $\sup_{x\in\mathcal{X}}\left|\xi_N^{F_1}(x)\right|$ is characterized in (S-8). Define

$$\begin{aligned}
\overline{\Lambda}_a^{\mathcal{J}}(\mathcal{W}_i) &= E\big[\Lambda^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k)\big|\mathcal{W}_i\big]\\
\overline{\Lambda}_b^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j) &= E\big[\Lambda^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k)\big|\mathcal{W}_i,\mathcal{W}_j\big] - E\big[\Lambda^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k)\big|\mathcal{W}_i\big] - E\big[\Lambda^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k)\big|\mathcal{W}_j\big]\\
\overline{\Lambda}_c^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k) &= \Lambda^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k) - E\big[\Lambda^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k)\big|\mathcal{W}_i,\mathcal{W}_j\big] - E\big[\Lambda^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k)\big|\mathcal{W}_i,\mathcal{W}_k\big]\\
&\quad - E\big[\Lambda^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k)\big|\mathcal{W}_j,\mathcal{W}_k\big] + E\big[\Lambda^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k)\big|\mathcal{W}_i\big] + E\big[\Lambda^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k)\big|\mathcal{W}_j\big]\\
&\quad + E\big[\Lambda^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k)\big|\mathcal{W}_k\big].
\end{aligned}$$

If the model is correct, then $E\big[\Lambda^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k)\big]=0$ for the reasons outlined above. In this case the Hoeffding Decomposition (see Lemma 5.1.4.A in Serfling (1980)) of $\mathcal{U}_{1_N}^{B_{\mathcal{J}}}$ is given by

$$\mathcal{U}_{1_N}^{B_{\mathcal{J}}} = \frac{(N-2)}{3N}\left[\frac{3}{N}\sum_{i=1}^{N}\overline{\Lambda}_a^{\mathcal{J}}(\mathcal{W}_i) + 3\binom{N}{2}^{-1}\sum_{i<j}\overline{\Lambda}_b^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j) + \binom{N}{3}^{-1}\sum_{i<j<k}\overline{\Lambda}_c^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k)\right]. \tag{S-13}$$

For each $\mathcal{J}\in\{I,II,III,IV\}$, the third term in the right hand side of (S-13) is a symmetric, third order U-statistic that is degenerate of order 2 as, by construction, $E\big[\overline{\Lambda}_c^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k)\big|\mathcal{W}_i,\mathcal{W}_j\big]=0$. Given our assumptions, it has a finite variance and therefore (see Lemma 5.2.1.A in Serfling (1980))

$$\begin{aligned}
&\binom{N}{3}^{-1}\sum_{i<j<k}\overline{\Lambda}_c^{I}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k) = O_p\left(\frac{1}{N^{3/2}\widetilde{b}^L\widetilde{h}_c^2}\right); \quad \binom{N}{3}^{-1}\sum_{i<j<k}\overline{\Lambda}_c^{II}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k) = O_p\left(\frac{1}{N^{3/2}\widetilde{b}^L\widetilde{h}_b^L}\right);\\
&\binom{N}{3}^{-1}\sum_{i<j<k}\overline{\Lambda}_c^{III}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k) = O_p\left(\frac{1}{N^{3/2}\widetilde{b}^L}\right); \quad \binom{N}{3}^{-1}\sum_{i<j<k}\overline{\Lambda}_c^{IV}(\mathcal{W}_i,\mathcal{W}_j,\mathcal{W}_k) = O_p\left(\frac{1}{N^{3/2}\widetilde{b}^L\widetilde{h}_c}\right).
\end{aligned} \tag{S-14}$$

We now move on to the first term in the right hand side of (S-13). Computing the relevant expectations, our assumptions yield

$$\frac{1}{N}\sum_{i=1}^{N}\overline{\Lambda}_a^{I}(\mathcal{W}_i)=0;\quad \frac{1}{N}\sum_{i=1}^{N}\overline{\Lambda}_a^{II}(\mathcal{W}_i)=O_p\left(\frac{\widetilde{h}_b^{\mathcal{M}}}{\widetilde{b}^L}\right);\quad \frac{1}{N}\sum_{i=1}^{N}\overline{\Lambda}_a^{III}(\mathcal{W}_i)=0;\quad \frac{1}{N}\sum_{i=1}^{N}\overline{\Lambda}_a^{IV}(\mathcal{W}_i)=O_p\left(\frac{\widetilde{h}_c^{\mathcal{M}}}{\widetilde{b}^L}\right).$$

Finally, let us deal with the second term on the right-hand side of (S-13). Define

$$\begin{aligned}
\Phi^{I}(\mathcal{W}_i,\mathcal{W}_j) &= \left[\frac{\varepsilon_{1j}\varepsilon_{2i}[\mu_{1i}-\mu_{1j}]\varphi(X_i)\overline{\phi}(X_j)^2}{E[\varphi(X)|t_{1j}]f_{t_1}(t_{1j})} - \frac{\varepsilon_{1i}\varepsilon_{2j}[\mu_{1j}-\mu_{1i}]\varphi(X_j)\overline{\phi}(X_i)^2}{E[\varphi(X)|t_{1i}]f_{t_1}(t_{1i})}\right]\frac{1}{\widetilde{h}_c^2}\mathcal{H}_c^{(1)}\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right)\\
\Phi^{II}(\mathcal{W}_i,\mathcal{W}_j) &= \left[\varepsilon_{1i}\frac{[Y_{2j}-\mu_{2i}]}{f_x(X_i)}F_1^{(1)}(t_{1i})\overline{\phi}(X_i)^2 + \varepsilon_{1j}\frac{[Y_{2i}-\mu_{2j}]}{f_x(X_j)}F_1^{(1)}(t_{1j})\overline{\phi}(X_j)^2\right]\frac{1}{\widetilde{h}_b^L}\mathcal{H}_b\left(\frac{\Delta X_{ij}}{\widetilde{h}_b}\right)\\
\Phi^{III}(\mathcal{W}_i,\mathcal{W}_j) &= \left[\frac{\varepsilon_{1i}F_1^{(1)}(t_{1i})\overline{\phi}(X_i)^2 f_x(X_i)\big(\widetilde{R}_1(t_{1i})Z_{1i}-\widetilde{Q}_1(t_{1i})\big)'\psi_j^{\theta_1}}{E[\varphi(X)|t_{1i}]}\right.\\
&\qquad\left. + \frac{\varepsilon_{1j}F_1^{(1)}(t_{1j})\overline{\phi}(X_j)^2 f_x(X_j)\big(\widetilde{R}_1(t_{1j})Z_{1j}-\widetilde{Q}_1(t_{1j})\big)'\psi_i^{\theta_1}}{E[\varphi(X)|t_{1j}]}\right]\\
\Phi^{IV}(\mathcal{W}_i,\mathcal{W}_j) &= \left[\frac{\varepsilon_{1i}\big[Y_{1j}-F_1(t_{1i})\big]\varphi(X_j)f_x(X_i)\overline{\phi}(X_i)^2}{E[\varphi(X)|t_{1i}]f_{t_1}(t_{1i})}\right.\\
&\qquad\left. + \frac{\varepsilon_{1j}\big[Y_{1i}-F_1(t_{1j})\big]\varphi(X_i)f_x(X_j)\overline{\phi}(X_j)^2}{E[\varphi(X)|t_{1j}]f_{t_1}(t_{1j})}\right]\frac{1}{\widetilde{h}_c}\mathcal{H}_c\left(\frac{\Delta t_{1ij}}{\widetilde{h}_c}\right)
\end{aligned} \tag{S-15}$$

We have

$$\binom{N}{2}^{-1}\sum_{i<j}\overline{\Lambda}_b^{I}(\mathcal{W}_i,\mathcal{W}_j)=\binom{N}{2}^{-1}\sum_{i<j}\Phi^{I}(\mathcal{W}_i,\mathcal{W}_j)+O_p\left(\frac{\widetilde{b}^{\overline{\mathcal{M}}}}{\widetilde{h}_c^{2+\overline{\mathcal{M}}}}\right)$$

$$\binom{N}{2}^{-1}\sum_{i<j}\overline{\Lambda}_b^{II}(\mathcal{W}_i,\mathcal{W}_j)=\binom{N}{2}^{-1}\sum_{i<j}\Phi^{II}(\mathcal{W}_i,\mathcal{W}_j)+O_p\left(\frac{\widetilde{b}^{\overline{\mathcal{M}}}}{\widetilde{h}_b^{L+\overline{\mathcal{M}}}}\right)+O_p\left(\frac{\widetilde{h}_b^{\overline{\mathcal{M}}}}{\widetilde{b}^{L}}\right)$$

$$\binom{N}{2}^{-1}\sum_{i<j}\overline{\Lambda}_b^{III}(\mathcal{W}_i,\mathcal{W}_j)=\binom{N}{2}^{-1}\sum_{i<j}\Phi^{III}(\mathcal{W}_i,\mathcal{W}_j)+O_p\left(\widetilde{b}^{\overline{\mathcal{M}}}\right)$$

$$\binom{N}{2}^{-1}\sum_{i<j}\overline{\Lambda}_b^{IV}(\mathcal{W}_i,\mathcal{W}_j)=\binom{N}{2}^{-1}\sum_{i<j}\Phi^{IV}(\mathcal{W}_i,\mathcal{W}_j)+O_p\left(\frac{\widetilde{b}^{\overline{\mathcal{M}}}}{\widetilde{h}_c^{1+\overline{\mathcal{M}}}}\right)+O_p\left(\frac{\widetilde{h}_c^{\overline{\mathcal{M}}}}{\widetilde{b}^{L}}\right).$$

For $\mathcal{J}\in\{I,III\}$, it is immediate to verify that if our model is correct, $\binom{N}{2}^{-1}\sum_{i<j}\Phi^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j)$ is a symmetric, degenerate second order U-statistic[3]. For $\mathcal{J}\in\{II,IV\}$, define

$$\widetilde{\Phi}^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j)=\Phi^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j)-E\Big[\Phi^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j)\Big|\mathcal{W}_i\Big]-E\Big[\Phi^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j)\Big|\mathcal{W}_j\Big].$$

Computing the expectations in question, our assumptions yield

$$\binom{N}{2}^{-1}\sum_{i<j}\Phi^{II}(\mathcal{W}_i,\mathcal{W}_j)=\binom{N}{2}^{-1}\sum_{i<j}\widetilde{\Phi}^{II}(\mathcal{W}_i,\mathcal{W}_j)+O_p\left(\widetilde{h}_b^{\overline{\mathcal{M}}}\right)$$

$$\binom{N}{2}^{-1}\sum_{i<j}\Phi^{IV}(\mathcal{W}_i,\mathcal{W}_j)=\binom{N}{2}^{-1}\sum_{i<j}\widetilde{\Phi}^{IV}(\mathcal{W}_i,\mathcal{W}_j)+O_p\left(\widetilde{h}_c^{\overline{\mathcal{M}}}\right)$$

The first object on the right hand side is exactly a second-order, degenerate U-statistic. We will now verify the basic condition in Theorem 1 of Hall (1984) for the following degenerate U-statistics

$$\binom{N}{2}^{-1}\sum_{i<j}\Phi^{I}(\mathcal{W}_i,\mathcal{W}_j);\quad \binom{N}{2}^{-1}\sum_{i<j}\widetilde{\Phi}^{II}(\mathcal{W}_i,\mathcal{W}_j);\quad \binom{N}{2}^{-1}\sum_{i<j}\widetilde{\Phi}^{IV}(\mathcal{W}_i,\mathcal{W}_j).$$

Define

$$\begin{aligned}\mathcal{G}^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j)&=E\Big[\Phi^{\mathcal{J}}(\mathcal{W}_k,\mathcal{W}_i)\Phi^{\mathcal{J}}(\mathcal{W}_k,\mathcal{W}_j)\Big|\mathcal{W}_i,\mathcal{W}_j\Big]\ (\text{for } \mathcal{J}=I);\\ \widetilde{\mathcal{G}}^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j)&=E\Big[\widetilde{\Phi}^{\mathcal{J}}(\mathcal{W}_k,\mathcal{W}_i)\widetilde{\Phi}^{\mathcal{J}}(\mathcal{W}_k,\mathcal{W}_j)\Big|\mathcal{W}_i,\mathcal{W}_j\Big]\ (\text{for } \mathcal{J}\in\{II,IV\}).\end{aligned}\tag{S-16}$$

[3] Note that $\mathcal{H}_c^{(1)}(a)=-\mathcal{H}_c^{(1)}(-a)$.

Given our assumptions, there is a scalar $\underline{\kappa} > 0$ such that

$$\frac{E\big[\mathcal{G}^{I}(\mathcal{W}_i,\mathcal{W}_j)^2\big] + N^{-1}E\big[\Phi^{I}(\mathcal{W}_i,\mathcal{W}_j)^4\big]}{\Big\{E\big[\Phi^{I}(\mathcal{W}_i,\mathcal{W}_j)^2\big]\Big\}^2} \leq \frac{O\big(\widetilde{h}_c^{-5}\big) + O\big(N^{-1}\widetilde{h}_c^{-7}\big)}{\underline{\kappa}\cdot\widetilde{h}_c^{-6}} = \frac{O\big(\widetilde{h}_c\big) + O\big(N^{-1}\widetilde{h}_c^{-1}\big)}{\underline{\kappa}} \longrightarrow 0$$

$$\frac{E\big[\widetilde{\mathcal{G}}^{II}(\mathcal{W}_i,\mathcal{W}_j)^2\big] + N^{-1}E\big[\widetilde{\Phi}^{II}(\mathcal{W}_i,\mathcal{W}_j)^4\big]}{\Big\{E\big[\widetilde{\Phi}^{II}(\mathcal{W}_i,\mathcal{W}_j)^2\big]\Big\}^2} \leq \frac{O\big(\widetilde{h}_b^{-L}\big) + O\big(\widetilde{h}_b^{\overline{\mathcal{M}}-L}\big) + O\big(N^{-1}\widetilde{h}_b^{-3L}\big) + O\big(N^{-1}\widetilde{h}_b^{\overline{\mathcal{M}}-3L}\big)}{\underline{\kappa}\cdot\big(\widetilde{h}_b^{-2L} + \widetilde{h}_b^{\overline{\mathcal{M}}-2L}\big)}$$

$$= \frac{O\big(\widetilde{h}_b^{L}\big) + O\big(\widetilde{h}_b^{\overline{\mathcal{M}}+L}\big) + O\big(N^{-1}\widetilde{h}_b^{-L}\big) + O\big(N^{-1}\widetilde{h}_b^{\overline{\mathcal{M}}-L}\big)}{\underline{\kappa}\cdot\big(1 + \widetilde{h}_b^{\overline{\mathcal{M}}}\big)} \to 0$$

$$\frac{E\big[\widetilde{\mathcal{G}}^{IV}(\mathcal{W}_i,\mathcal{W}_j)^2\big] + N^{-1}E\big[\widetilde{\Phi}^{IV}(\mathcal{W}_i,\mathcal{W}_j)^4\big]}{\Big\{E\big[\widetilde{\Phi}^{IV}(\mathcal{W}_i,\mathcal{W}_j)^2\big]\Big\}^2} \leq \frac{O\big(\widetilde{h}_c^{-1}\big) + O\big(\widetilde{h}_c^{\overline{\mathcal{M}}-1}\big) + O\big(N^{-1}\widetilde{h}_c^{-3}\big) + O\big(N^{-1}\widetilde{h}_c^{\overline{\mathcal{M}}-3}\big)}{\underline{\kappa}\cdot\big(\widetilde{h}_c^{-2} + \widetilde{h}_c^{\overline{\mathcal{M}}-2}\big)}$$

$$= \frac{O\big(\widetilde{h}_c\big) + O\big(\widetilde{h}_c^{\overline{\mathcal{M}}+1}\big) + O\big(N^{-1}\widetilde{h}_c^{-1}\big) + O\big(N^{-1}\widetilde{h}_c^{\overline{\mathcal{M}}-1}\big)}{\underline{\kappa}\cdot\big(1 + \widetilde{h}_c^{\overline{\mathcal{M}}}\big)} \to 0.$$

Given this result, Theorem 1 in Hall (1984) yields

$$\begin{aligned}
&\sum_{i<j}\Phi^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j) \xrightarrow{d} \mathcal{N}\Big(0, \frac{N^2}{2}E\big[\Phi^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j)^2\big]\Big) \text{ for } \mathcal{J} = I;\\
&\sum_{i<j}\widetilde{\Phi}^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j) \xrightarrow{d} \mathcal{N}\Big(0, \frac{N^2}{2}E\big[\widetilde{\Phi}^{\mathcal{J}}(\mathcal{W}_i,\mathcal{W}_j)^2\big]\Big). \text{ for } \mathcal{J} \in \{II, IV\}.
\end{aligned} \tag{S-17}$$

In particular, this implies that

$$\binom{N}{2}^{-1}\sum_{i<j}\Phi^{I}(\mathcal{W}_i,\mathcal{W}_j) = O_p\big(N^{-1}\widetilde{h}_c^{-3/2}\big); \quad \binom{N}{2}^{-1}\sum_{i<j}\widetilde{\Phi}^{II}(\mathcal{W}_i,\mathcal{W}_j) = O_p\big(N^{-1}\widetilde{h}_b^{-L/2}\big);$$

$$\binom{N}{2}^{-1}\sum_{i<j}\widetilde{\Phi}^{IV}(\mathcal{W}_i,\mathcal{W}_j) = O_p\big(N^{-1}\widetilde{h}_c^{-1/2}\big)$$

Given our assumptions, $E\big[\Phi^{III}(\mathcal{W}_i,\mathcal{W}_j)^2\big] < \infty$. Using this, Theorem 5.5.2 in Serfling (1980) yields

$$N \times \binom{N}{2}^{-1}\sum_{i<j}\Phi^{III}(\mathcal{W}_i,\mathcal{W}_j) \xrightarrow{d} 3\mathcal{Y}, \tag{S-18}$$

where $\mathcal{Y} = \sum_{j=1}^{\infty}\lambda_j\big(\chi^2_{1_j} - 1\big)$, with $\chi^2_{1_1}, \chi^2_{1_2}, \ldots$ are independent $\chi^2_1$ variables. The weights $(\lambda_j)_{j=1}^{\infty}$ are the solutions (in $\lambda$) to $Ag - \lambda g = 0$, where $Ag(\mathcal{W}) = \int_{-\infty}^{\infty}\Phi^{III}(\mathcal{W},\mathcal{W}_2)g(\mathcal{W}_2)dF_{\mathcal{W}}(\mathcal{W}_2)$. Immediately, this implies that

$$\binom{N}{2}^{-1}\sum_{i<j}\Phi^{III}(\mathcal{W}_i,\mathcal{W}_j) = O_p(N^{-1}). \tag{S-19}$$

Going back to the Hoeffding decomposition in (S-13) and the results in (S-14) – (S-19), the expression for $U^B_{1_N}$ in (S-12) becomes

$$U^B_{1_N} = O_p\left(\sup_{x\in\mathcal{X}}\left|\xi^{F_1}_N(x)\right|\right) + O_p\left(\frac{1}{N^{3/2}\widetilde{b}^L\widetilde{h}_c^2}\right) + O_p\left(\frac{1}{N^{3/2}\widetilde{b}^L\widetilde{h}_b^L}\right) + O_p\left(\frac{\widetilde{h}_b^{\overline{\mathcal{M}}}}{\widetilde{b}^L}\right) + O_p\left(\frac{\widetilde{h}_c^{\overline{\mathcal{M}}}}{\widetilde{b}^L}\right) + O_p\left(\frac{\widetilde{b}^{\overline{\mathcal{M}}}}{\widetilde{h}_c^{2+\overline{\mathcal{M}}}}\right)$$
$$+ O_p\left(\frac{\widetilde{b}^{\overline{\mathcal{M}}}}{\widetilde{h}_b^{L+\overline{\mathcal{M}}}}\right) + O_p\left(\frac{1}{N\widetilde{h}_c^{3/2}}\right) + O_p\left(\frac{1}{N\widetilde{h}_b^{L/2}}\right) + O_p\left(\frac{1}{N\widetilde{h}_c^{1/2}}\right) + O_p\left(\frac{1}{N}\right),$$

where the order of magnitude of $\sup_{x\in\mathcal{X}}\left|\xi^{F_1}_N(x)\right|$ is given in the last line of (S-8). Note that the last two terms are redundant, given the rest. We will drop them from now on. Using this result along with (S-11) and (S-10) implies that our test-statistic satisfies

$$U_{1_N} = \binom{N}{2}^{-1}\sum_{i<j}\frac{\varepsilon_{1j}\varepsilon_{1i}\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right) + O_p\left(\sup_{x\in\mathcal{X}}\left|\widehat{\nu}_1(x)\right|^2\right) + O_p\left(\sup_{x\in\mathcal{X}}\left|\xi^{F_1}_N(x)\right|\right) + O_p\left(\frac{1}{N^{3/2}\widetilde{b}^L\widetilde{h}_c^2}\right)$$
$$+ O_p\left(\frac{1}{N^{3/2}\widetilde{b}^L\widetilde{h}_b^L}\right) + O_p\left(\frac{\widetilde{h}_b^{\overline{\mathcal{M}}}}{\widetilde{b}^L}\right) + O_p\left(\frac{\widetilde{h}_c^{\overline{\mathcal{M}}}}{\widetilde{b}^L}\right) + O_p\left(\frac{\widetilde{b}^{\overline{\mathcal{M}}}}{\widetilde{h}_c^{2+\overline{\mathcal{M}}}}\right) + O_p\left(\frac{\widetilde{b}^{\overline{\mathcal{M}}}}{\widetilde{h}_b^{L+\overline{\mathcal{M}}}}\right) + O_p\left(\frac{1}{N\widetilde{h}_c^{3/2}}\right) + O_p\left(\frac{1}{N\widetilde{h}_b^{L/2}}\right),$$

where the order of magnitude of $\sup_{x\in\mathcal{X}}\left|\widehat{\nu}_1(x)\right|$ is given by (S-9). Using our assumptions about the relative rates of convergence of the three bandwidths involved, and the "smoothness measure" $\overline{\mathcal{M}}$, the above equation becomes

$$U_{1_N} = \underbrace{\binom{N}{2}^{-1}\sum_{i<j}\frac{\varepsilon_{1j}\varepsilon_{1i}\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)}_{\equiv U^A_{1_N}} + o_p\left(\frac{1}{N\widetilde{b}^{L/2}}\right) \tag{S-20}$$

If our model is correct, $U^A_{1_N}$ is a symmetric, degenerate second order U-statistic. We will verify that it satisfies Hall's condition. As in (S-16), let

$$\Phi(\mathcal{W}_i,\mathcal{W}_j) = \frac{\varepsilon_{1j}\varepsilon_{1i}\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right); \quad \mathcal{G}(\mathcal{W}_i,\mathcal{W}_j) = E\Big[\Phi(\mathcal{W}_i,\mathcal{W}_k)\Phi(\mathcal{W}_j,\mathcal{W}_k)\Big|\mathcal{W}_i,\mathcal{W}_j\Big].$$

Given our assumptions, there exists a scalar $\underline{c} > 0$ such that

$$\frac{E\big[\mathcal{G}(\mathcal{W}_i,\mathcal{W}_j)^2\big] + N^{-1}E\big[\Phi(\mathcal{W}_i,\mathcal{W}_j)^4\big]}{\Big\{E\big[\Phi(\mathcal{W}_i,\mathcal{W}_j)^2\big]\Big\}^2} = \frac{O\big(\widetilde{b}^{-L}\big) + O\big(N^{-1}\widetilde{b}^{-3L}\big)}{\underline{c}\cdot\widetilde{b}^{-2L}} = \frac{O\big(\widetilde{b}^{L}\big) + O\big(N^{-1}\widetilde{b}^{-L}\big)}{\underline{c}} \to 0$$

Therefore the conditions Theorem 1 in Hall (1984) are satisfied and they yield

$$\Sigma_p^{-1}N\widetilde{b}^{L/2}\binom{N}{2}^{-1}\sum_{i<j}\frac{\varepsilon_{p_i}\varepsilon_{p_j}\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right) \xrightarrow{d} \mathcal{N}(0,1), \quad \text{where} \quad \Sigma_p = E\left[\frac{\varepsilon_{pj}^2\varepsilon_{pi}^2\overline{\phi}(X_i)^2\overline{\phi}(X_j)^2}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)^2\right].$$

This yields $\Sigma_p^{-1}N\widetilde{b}^{L/2}U_{p_N} \equiv \Sigma_p^{-1}\mathcal{T}_{p_N} \xrightarrow{d} \mathcal{N}(0,1)$. Moreover, using Theorem 1 in Hall (1984) we can show that for any pair of constants $\tau_1$, $\tau_2 \in \mathbb{R}$,

$$\tau_1\mathcal{T}_{1_N} + \tau_2\mathcal{T}_{2_N} \xrightarrow{d} \mathcal{N}\big(0, \tau_1^2\Sigma_1 + \tau_2^2\Sigma_2 + 2\tau_1\tau_2\Sigma_{1,2}\big), \quad \text{where} \quad \Sigma_{1,2} = E\left[\frac{\varepsilon_{1j}\varepsilon_{1i}\varepsilon_{2j}\varepsilon_{2i}\overline{\phi}(X_i)^2\overline{\phi}(X_j)^2}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)^2\right].$$

From here, the Cramer-Wold device and the properties of the normal distribution imply

$$\big(\mathcal{T}_{1_N}\ \mathcal{T}_{2_N}\big)' \xrightarrow{d} N\big(0,\Sigma\big), \quad \text{where} \quad \Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2} & \Sigma_2 \end{pmatrix}.$$

Note that $\Sigma$ is invertible under the conditions of Theorem 2 which implies (via the continuous mapping theorem) that $N^2\widetilde{b}^L\big(U_{1_N}\ ,\ U_{2_N}\big)\Sigma^{-1}\big(U_{1_N}\ ,\ U_{2_N}\big)' \xrightarrow{d} \chi^2_2$. Next, the uniform convergence result in (S-9) is more than enough to yield

$$\begin{aligned}
\widehat{\Sigma}_p &\equiv \binom{N}{2}^{-1}\sum_{i<j}\frac{\widehat{\varepsilon}^2_{pj}\widehat{\varepsilon}^2_{pi}\overline{\phi}(X_i)^2\overline{\phi}(X_j)^2}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)^2 = \binom{N}{2}^{-1}\sum_{i<j}\frac{\varepsilon^2_{pj}\varepsilon^2_{pi}\overline{\phi}(X_i)^2\overline{\phi}(X_j)^2}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)^2 + o_p(1) \\
&\xrightarrow{p} \Sigma_p, \\
\widehat{\Sigma}_{1,2} &\equiv \binom{N}{2}^{-1}\sum_{i<j}\frac{\widehat{\varepsilon}_{1_i}\widehat{\varepsilon}_{1_j}\widehat{\varepsilon}_{2_i}\widehat{\varepsilon}_{2_j}\overline{\phi}(X_i)^2\overline{\phi}(X_j)^2}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)^2 = \binom{N}{2}^{-1}\sum_{i<j}\frac{\varepsilon_{1_i}\varepsilon_{1_j}\varepsilon_{2_i}\varepsilon_{2_j}\overline{\phi}(X_i)^2\overline{\phi}(X_j)^2}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)^2 \\
&\quad + o_p(1) \xrightarrow{p} \Sigma_{1,2}.
\end{aligned}$$

This yields part 1 of the theorem. Namely, if the model is correctly specified,

$$\mathcal{T}_N \equiv N^2\widetilde{b}^L\big(U_{1_N}\ ,\ U_{2_N}\big)\widehat{\Sigma}^{-1}\big(U_{1_N}\ ,\ U_{2_N}\big)' \xrightarrow{d} \chi^2_2.$$

To prove part 2 of Theorem 2, we let

$$\begin{aligned}
D_p &= E\Big[\Big(E\big[Z_pZ_p'\overline{\phi}(X)\big|\mu_p\big]E\big[\overline{\phi}(X)\big|\mu_p\big] - E\big[Z_p\overline{\phi}(X)\big|\mu_p\big]E\big[Z_p\overline{\phi}(X)\big|\mu_p\big]'\Big)f_{\mu_p}(\mu_p)\Big] \\
C_p &= E\Big[\Big(E\big[Z_pW_p\overline{\phi}(X)\big|\mu_p\big]E\big[\overline{\phi}(X)\big|\mu_p\big] - E\big[Z_p\overline{\phi}(X)\big|\mu_p\big]E\big[W_p\overline{\phi}(X)\big|\mu_p\big]\Big)f_{\mu_p}(\mu_p)\Big].
\end{aligned}$$

Given our assumptions, a dominated convergence argument easily yields

$$\widehat{\theta}_p \xrightarrow{p} -D_p^{-1}C_p \equiv \theta^*_p.$$

Under the conditions of part 2 of Theorem 2, $\theta^*_p$ is well-defined even if the model is incorrect and Equation in AL-09 is violated with positive probability. Conversely, if the model is correctly specified we know that $W_p = F_p^{-1}(\mu_p) - Z_p'\theta_p$ and $\theta^*_p = \theta_p$ (the true structural parameter value). Let $t^*_{pi} = W_{pi} + Z'_{pi}\theta^*_p$. Given our assumptions, conditional on $X_j$ we have[4]

$$\widehat{F}_p(\widehat{t}_{pj}) \xrightarrow{p} \frac{E\big[\mu_p(X_i)\varphi(X_i)\big|t^*_{pi} = t^*_{pj}\big]}{E\big[\varphi(X_i)\big|t^*_{pi} = t^*_{pj}\big]} \equiv F^*_p(t^*_{pj}).$$

[4] We use the notation $\mu_{pi}$ and $\mu_p(X_i)$ interchangeably.

Moreover, this convergence is uniform over $\mathcal{X}$. This yields

$$U_{p_N} = \binom{N}{2}^{-1} \sum_{i<j} \frac{\varepsilon^*_{pi}\varepsilon^*_{pj}\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L}\overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right) + o_p(1), \text{ with } \varepsilon^*_{pi} \equiv Y_{pi} - F^*_p(t^*_{pi}).$$
$$= E\Big[\big(\mu_p(X_i) - F^*_p(t^*_{pi})\big)^2\overline{\phi}(X_i)^2\Big] + o_p(1)$$

Suppose Equation (1) in AL-09 is violated with positive probability for player $p$ in the set $\mathcal{X}$ and

$$\Pr\Big[\mu_p(X_i)E\big[\varphi(X_i)\big|t^*_{pi}\big] \neq E\big[\mu_p(X_i)\varphi(X_i)\big|t^*_{pi}\big]\Big|X_i \in \mathcal{X}\Big] > 0.$$

Then, under the conditions of part 2 of Theorem 2 we would have

$$\Pr\Big[\mu_p(X_i) \neq F_p\big(t^*_p(X_i)\big)\Big|X_i \in \mathcal{X}\Big] > 0.$$

Consequently, $E\Big[\big(\mu_{pi} - F^*_p(t^*_{pi})\big)^2\overline{\phi}(X_i)^2\Big] > 0$. It follows that if Equation (1) in AL-09 is violated with positive probability,

$$\Pr\Big[N\widetilde{b}^{L/2}\big|U_{p_N}\big| > m_{_N}\Big] \longrightarrow 1 \text{ for any sequence of scalars such that } m_{_N}\big/\big(N\widetilde{b}^{L/2}\big) \longrightarrow 0.$$

If the exclusion restriction in Assumption (A1) is satisfied and if $Y_1 - F^*_1(t^*_1)$ and $Y_2 - F^*_2(t^*_2)$ are not perfectly correlated conditional on $X \in \mathcal{X}$ (as it is assumed in part 2 of Theorem 2), it is easy to show that $\widehat{\Sigma}^{-1}$ has a well-defined probability-limit. Combined with the previous result, this yields

$$\Pr\big(\big|\mathcal{T}_N\big| > m_{_N}\big) \longrightarrow 1 \text{ for any sequence of scalars such that } m_{_N}\big/\big(N^2\widetilde{b}^{L}\big) \longrightarrow 0.$$

Therefore $\mathcal{T}_N$ diverges w.p.1. This concludes the proof. □

## 1.1 Kernels and bandwidths used in Monte Carlo experiments

For a random variable $\psi$, define

$$\widehat{\mathcal{R}}(\psi) = \min\left\{\widehat{\sigma}(\psi), \frac{\widehat{F}^{-1}_{\psi}(0.75) - \widehat{F}^{-1}_{\psi}(0.25)}{1.34}\right\}.$$

This is the proportionality constant used in Silverman's "rule of thumb" bandwidth $h_N = 0.9\widehat{\mathcal{R}}N^{-1/5}$ (see Equation 3.31, p. 48 in Silverman (1986)). According to the notation in the previous sections, the kernels involved in the estimation of $\widehat{\theta}_p$ are $K_a$ and $K_b$, while those used to construct our specification test-statistic are $\mathcal{H}_b, \mathcal{H}_c$ and $\overline{H}$. We used covariate-specific bandwidths for $K_b$, $\mathcal{H}_b$ and $\overline{H}$ of the form $h_b(W_p) = \mathcal{C}_{h_b}\widehat{\mathcal{R}}(W_p) \cdot N^{-\lambda_{h_b}}$ and $h_b(V_p) = \mathcal{C}_{h_b}\widehat{\mathcal{R}}(V_p) \cdot N^{-\lambda_{h_b}}$ for $K_b$, $\widetilde{h}_b(W_p) = \mathcal{C}_{\widetilde{h}_b}\widehat{\mathcal{R}}(W_p) \cdot N^{-\lambda_{\widetilde{h}_b}}$ and

$\widetilde{h}_b(V_p) = \mathcal{C}_{\widetilde{h}_b}\widehat{\mathcal{R}}(V_p)\cdot N^{-\lambda_{\widetilde{h}_b}}$ for $\mathcal{H}_b$, and $\widetilde{b}(W_p) = \mathcal{C}_{\widetilde{b}}\widehat{\mathcal{R}}(W_p)\cdot N^{-\lambda_{\widetilde{b}}}$ and $\widetilde{b}(V_p) = \mathcal{C}_{\widetilde{b}}\widehat{\mathcal{R}}(V_p)\cdot N^{-\lambda_{\widetilde{b}}}$ for $\overline{H}$. Our asymptotic results still hold with covariate-specific bandwidths as long as the bandwidth convergence rates in our assumptions are satisfied. The bandwidths used for the remaining kernels were of the form $h_a(\widehat{\mu}_p) = \mathcal{C}_{h_a}\widehat{\mathcal{R}}(\widehat{\mu}_p)\cdot N^{-\lambda_{h_a}}$ for $K_a$, and $\widetilde{h}_c(\widehat{t}_p) = \mathcal{C}_{h_c}\widehat{\mathcal{R}}(\widehat{t}_p)\cdot N^{-\lambda_{\widetilde{h}_c}}$ for $\mathcal{H}_c$. All kernels used were multiplicative with a general functional form of the type

$$K(\psi) = \big(a_0 + a_1\psi^2 + a_2\psi^4 + \cdots + a_p\psi^{2p}\big)\phi(\psi),$$

where $\phi(\cdot)$ is the $\mathcal{N}(0,1)$ density function and the coefficients of the polynomial are chosen to satisfy the various bias-reducing conditions in the paper. The constants $\lambda_{h_b},\ldots,\lambda_{\widetilde{b}}$ were chosen to satisfy the convergence rates in our bandwidth assumptions. The values used are specified in each one of the tables presented below. The next section describes how the constants $\mathcal{C}_{h_b},\ldots,\mathcal{C}_{\widetilde{b}}$ were chosen.

### 1.1.1 Choice of bandwidth constants $\mathcal{C}_{h_b}, \mathcal{C}_{h_a}$ used in the estimation of $\widehat{\theta}_p$

Our previous sections results provide no guidance for choosing the different bandwidths involved. Applied researchers could use different criteria to solve this question. We employed a procedure based on the asymptotic approximations of our estimator and specification test-statistic. Let $\widehat{\theta}_p(-j)$ and $\widehat{\psi}_i^p(-j)$ denote the estimator and sample influence-function analog (as described in Equation (**??**) of Theorem **??**) that result when we drop the $j^{th}$ observation in the sample. Let

$$\widehat{\psi}^p_{N}(-j) = \frac{1}{N-1}\sum_{i\neq j}\widehat{\psi}_i^{\theta_p}(-j), \quad \widehat{E}[\widehat{\theta}_p] = \frac{1}{N}\sum_{j=1}^{N}\widehat{\theta}_p(-j), \quad \widehat{\mathcal{S}}_p(-j) \equiv \widehat{\theta}_p(-j) - \widehat{E}[\widehat{\theta}_p],$$

$$\widehat{T}_{\theta_p}(-j) \equiv \widehat{\mathcal{S}}_p(-j)'\widehat{\mathcal{S}}_p(-j), \quad \text{and} \quad \widehat{T}_{\psi_p}(-j) \equiv \widehat{\psi}^p_{N}(-j)'\widehat{\psi}^p_{N}(-j).$$

The constants $\mathcal{C}_{h_b}, \mathcal{C}_{h_a}$ used in the estimation of $\widehat{\theta}_p$ were chosen among a grid of candidate values to minimize the Kolmogorov-Smirnov distance between the empirical distributions of $\widehat{T}_{\theta_p}(-j)$ and $\widehat{T}_{\psi_p}(-j)$. Specifically, the following steps were taken for each design.

1.– For each design, *one* sample of size $N = 600$ was generated.

2.– A grid of points in the set $\big[0.2, 6\big]\times\big[0.2, 6\big]$ was considered as candidate values for $\mathcal{C}_{h_b}, \mathcal{C}_{h_a}$. For $p = 1$, $\big\{\widehat{T}_{\theta_p}(-j)\big\}_{j=1}^{N}$ and $\big\{\widehat{T}_{\psi_p}(-j)\big\}_{j=1}^{N}$ were computed for each point in this grid.

3.– Let $\widehat{F}_{\widehat{T}_{\theta_p}}(\cdot)$ and $\widehat{F}_{\widehat{T}_{\psi_p}}(\cdot)$ denote the corresponding empirical distributions. The values chosen for $\mathcal{C}_{h_b}, \mathcal{C}_{h_a}$ were those that yielded the smallest value of the Kolmogorov-Smirnov distance between $\widehat{F}_{\widehat{T}_{\theta_p}}(\cdot)$ and $\widehat{F}_{\widehat{T}_{\psi_p}}(\cdot)$.

This procedure was performed only once for each design. The resulting constants $\mathcal{C}_{h_b}, \mathcal{C}_{h_a}$ were used in all subsequent simulations.

### 1.1.2 Choice of bandwidth constants $\mathcal{C}_{\widetilde{h}_b}, \mathcal{C}_{\widetilde{h}_c}, \mathcal{C}_{\widetilde{b}}$ used in the construction of the specification test-statistic $\mathcal{T}_N$

As with the previous bandwidths, we employed a simple procedure based on the asymptotic properties of $\mathcal{T}_N$. To simplify computations, in all our experiments we fixed $\mathcal{C}_{\widetilde{h}_b} = \mathcal{C}_{\widetilde{b}}$. According to Theorem **??**, if the model is correct the test-statistic $\mathcal{T}_N$ described there has an asymptotic $\chi^2_2$ distribution. Following the choice of $\mathcal{C}_{h_b}$ and $\mathcal{C}_{h_a}$ and the resulting estimator $\widehat{\theta}$, we took the following steps for the same sample of size $N = 600$ described above

1.– A grid of points in the set $\big[0.2, 6\big] \times \big[0.2, 6\big]$ was considered as candidate values for $\mathcal{C}_{\widetilde{h}_b} = \mathcal{C}_{\widetilde{b}}$ and $\mathcal{C}_{\widetilde{h}_c}$. Let $\mathcal{T}_N(-j)$ denote the test-statistic described in Theorem **??** after we leave out the $j^{th}$ observation. $\big\{\mathcal{T}_N(-j)\big\}_{j=1}^{N}$ were computed for each point in the grid of candidate values.

2.– Let $\widehat{F}_{\mathcal{T}_N}(\cdot)$ denote the resulting empirical distribution. The values for $\mathcal{C}_{\widetilde{h}_b} = \mathcal{C}_{\widetilde{b}}$ and $\mathcal{C}_{h_a}$ that were chosen were those that yielded the smallest value of the Kolmogorov-Smirnov distance between $\widehat{F}_{\mathcal{T}_N}(\cdot)$ and the distribution of a $\chi^2_2$ random variable.

As before, this was done only once for each design. The resulting constants were used in all subsequent simulations of $\mathcal{T}_N$.

## References


Ahn, H., and J. Powell (1993): "Semiparametric Estimation of Censored Selection Models," *Journal of Econometrics*, 58, 3–29.

Aradillas-Lopez, A. (2008): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information," Working Paper, Princeton University.

Collomb, G., and W. Hardle (1986): "Strong Uniform Convergence Rates in Robust Nonparametric Time Series Analysis and Prediction : Kernel Regression from Dependent Observations," *Stochastic Processes and their Applications*, 23, 77–89.

Hall, P. (1984): "Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators," *Journal of Multivariate Analysis*, 14, 1–16.

Serfling, R. (1980): *Approximation Theorems of Mathematical Statistics*. Wiley, New York, NY.

Silverman, B. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.