Econometric Supplement for the paper "Inference in Ordered Response Games with Complete Information".*

Andres Aradillas-Lopez†
Pennsylvania State University

Adam M. Rosen‡
Duke University and CeMMAP

October 12, 2021

Abstract

This supplement contains the steps of the proofs of the econometric theorems and lemmas in the paper. Throughout, we make reference to restrictions (assumptions) described in the paper.

1 Proof of Theorem 2

We prove parts 1 and 2 in the statement of the Theorem in separate steps. Let $\tilde{\beta}^*$ and λ^* denote the population values of $\tilde{\beta} = (\alpha, \beta')'$ and λ, and let $\tilde{b}$ denote any alternative value of $(\alpha, \beta')'$ such that $\tilde{b} \neq \tilde{\beta}^*$.

Step 1. Suppose that F is known and define the sets

$$S^+_{\tilde{b}} \equiv \left\{ z : z_1 \left(\tilde{b} - \tilde{\beta}^* \right) > 0 \wedge z_2 \left(\tilde{b} - \tilde{\beta}^* \right) \geq 0 \right\},$$

$$S^-_{\tilde{b}} \equiv \left\{ z : z_1 \left(\tilde{b} - \tilde{\beta}^* \right) < 0 \wedge z_2 \left(\tilde{b} - \tilde{\beta}^* \right) \leq 0 \right\}.$$

*This paper has benefited from feedback from participants at several conferences and seminars. We thank Eleni Aristodemou, Andrew Chesher, Jeremy Fox, Ariel Pakes, Aureo de Paula, Elie Tamer, and especially Francesca Molinari for helpful discussion and suggestions, and we thank the editor, an associate editor, and two anonymous referees for several comments and suggestions that have helped to improve the paper. Xinyue Bei provided excellent research assistance. Adam Rosen gratefully acknowledges financial support from the UK Economic and Social Research Council through grants (RES-589-28-0001, RES-589-28-0002 and ES/P008909/1) to the ESRC Centre for Microdata Methods and Practice (CeMMAP) and through the funding of the "Programme Evaulation for Policy Analysis" node of the UK National Centre for Research Methods, from the European Research Council (ERC) through grant ERC-2009-StG-240910-ROMETA, and from a British Academy Mid-Career Fellowship.

†Address: Department of Economics, Pennsylvania State University, University Park, PA 16802, United States. Email: aaradill@psu.edu.

‡Address: Department of Economics, Duke University, 213 Social Sciences Building, Box 90019, United States. Email: adam.rosen@duke.edu.

For any $z \in S_{\tilde{b}}^{+}$ we have that

$$F\left(z_1\tilde{b}, z_2\tilde{b}\right) > F\left(z_1\tilde{\beta}^*, z_2\tilde{\beta}^*\right) = P\left(Y = (0,0)|z\right),$$

and likewise for any $z \in S_{\tilde{b}}^{-}$,

$$F\left(z_1\tilde{b}, z_2\tilde{b}\right) < F\left(z_1\tilde{\beta}^*, z_2\tilde{\beta}^*\right) = P\left(Y = (0,0)|z\right).$$

Under conditions (i) and (ii) in the statement of the Theorem we'll show that at least one of $P\left(Z \in S_{\tilde{b}}^{+}\right)$ and $P\left(Z \in S_{\tilde{b}}^{-}\right)$ is positive, implying that $\tilde{b}$ and $\tilde{\beta}^*$ are observationally distinct. Since the choice of $\tilde{b}$ is arbitrary, this establishes that $\tilde{\beta}^*$ is point identified.

To see why at least one of $P\left(Z \in S_{\tilde{b}}^{+}\right)$ and $P\left(Z \in S_{\tilde{b}}^{-}\right)$ is positive, first note that at least one of

$$P\left(Z_1\left(\tilde{b} - \tilde{\beta}^*\right) > 0\right), \qquad P\left(Z_1\left(\tilde{b} - \tilde{\beta}^*\right) < 0\right) \tag{1}$$

is positive, since if this were not true this would imply $P\left(Z_1\left(\tilde{b} - \tilde{\beta}^*\right) = 0\right) = 1$, which is a violation of condition (i). If only $P\left(Z_1\left(\tilde{b} - \tilde{\beta}^*\right) > 0\right)$ of the two probabilities in (1) is positive, then

$$P\left(Z \in S_{\tilde{b}}^{+}\right) = P\left(Z_2\left(\tilde{b} - \tilde{\beta}^*\right) \geq 0 | Z_1\left(\tilde{b} - \tilde{\beta}^*\right) > 0\right) P\left(Z_1\left(\tilde{b} - \tilde{\beta}^*\right) > 0\right),$$

which must be positive by application of condition (ii) in the statement of Theorem with $c = \tilde{b} - \tilde{\beta}^*$. Similar reasoning establishes that $P\left(Z \in S_{\tilde{b}}^{-}\right)$ is positive if only $P\left(Z_1\left(\tilde{b} - \tilde{\beta}^*\right) < 0\right)$ of the two probabilities in (1) is positive. If both probabilities in (1) are positive, then at least one of $P\left(Z \in S_{\tilde{b}}^{+}\right) = P\left(Z_2\left(\tilde{b} - \tilde{\beta}^*\right) \geq 0 | Z_1\left(\tilde{b} - \tilde{\beta}^*\right) > 0\right) \cdot P\left(Z_1\left(\tilde{b} - \tilde{\beta}^*\right) > 0\right)$ and $P\left(Z \in S_{\tilde{b}}^{-}\right) = P\left\{Z_2\left(\tilde{b} - \tilde{\beta}^*\right) \leq 0 | Z_1\left(\tilde{b} - \tilde{\beta}^*\right) < 0\right\} \cdot P\left(Z_1\left(\tilde{b} - \tilde{\beta}^*\right) < 0\right)$ must be strictly positive by (ii). Thus with λ^* known $\tilde{b}$ is observationally distinct from $\tilde{\beta}^*$, since for each $z \in S_{\tilde{b}}$, $P\left(Y = (0,0)|z\right) \neq F\left(z_1\tilde{b}, z_2\tilde{b}\right)$.

Define now the expectation of n^{-1} times the log-likelihood

$$\mathcal{L}_0\left(\tilde{\beta}, \lambda\right) \equiv E\left[\ell\left(\left(\tilde{\beta}', \lambda\right)'; W\right)\right],$$

where as in the main text, for $\theta_1 \equiv \left(\tilde{\beta}, \lambda\right)$, $w \equiv (y, x)$ and $z_j \equiv (1, -x_j)$:

$$\ell\left(\theta_1; w\right) \equiv 1\left[y = (0,0)\right] \cdot \log F\left(z_1\tilde{\beta}, z_2\tilde{\beta}; \lambda\right) + \left(1\left[y \neq (0,0)\right]\right) \cdot \log\left(1 - F\left(z_1\tilde{\beta}, z_2\tilde{\beta}; \lambda\right)\right).$$

If F is only known to belong to some class of distribution functions $\{F_\lambda : \lambda \in \Gamma\}$, the above reasoning implies that for each $\lambda \in \Gamma$, $\mathcal{L}_0\left(\tilde{\beta}, \lambda\right)$ is uniquely maximized with respect to $\tilde{\beta}$. Then the conclusion of the first claim of the Theorem follows letting $\tilde{\beta}^*(\lambda)$ denote the maximizer of $\mathcal{L}_0\left(\tilde{\beta}, \lambda\right)$ with respect to $\tilde{\beta} \in B$ for any $\lambda \in \Gamma$.

Step 2. Here we will show that, under the assumptions of the theorem,

$$Pr\Big[F\left(Z_1\widetilde{\beta}, Z_2\widetilde{\beta}; \lambda\right) \neq F\left(Z_1\widetilde{\beta}', Z_2\widetilde{\beta}'; \lambda'\right)\Big] > 0 \quad \forall\, (\widetilde{\beta}, \lambda) \neq (\widetilde{\beta}', \lambda') \in \Theta. \tag{2}$$

And identification of $(\widetilde{\beta}^*, \lambda^*)$ follows. Take a pair of parameter values $(\widetilde{\beta}, \lambda)$ and $(\widetilde{\beta}', \lambda')$ in Θ and abbreviate $Z_1\widetilde{\beta} \equiv u_1$, $Z_1\widetilde{\beta}' \equiv u_1'$ $Z_2\widetilde{\beta} \equiv u_2$ and $Z_2\widetilde{\beta}' \equiv u_2'$. Denote $u \equiv (u_1, u_2)$ and $u' \equiv (u_1', u_2')$. Let $C(u, u')$, $D(u')$ and $D(u)$ be as defined in the assumptions of the theorem. Then, $(\widetilde{\beta}, \lambda)$ and $(\widetilde{\beta}', \lambda')$ are observationally equivalent if and only if

$$C(u, u') = \lambda' D(u') - \lambda D(u) \quad \text{w.p.1.} \tag{3}$$

That is, if and only if

$$Pr\Big[F\left(Z_1\widetilde{\beta}, Z_2\widetilde{\beta}; \lambda\right) = F\left(Z_1\widetilde{\beta}', Z_2\widetilde{\beta}'; \lambda'\right)\Big] = 1$$

Suppose $\widetilde{\beta} = \widetilde{\beta}'$. Then, $u = u'$ w.p.1, which implies $C(u, u') = 0$ and $D(u) = D(u')$ w.p.1. In this case, (3) holds if and only if $\lambda = \lambda'$. Therefore, $(\widetilde{\beta}, \lambda) \neq (\widetilde{\beta}', \lambda')$ are observationally equivalent only if $\widetilde{\beta} \neq \widetilde{\beta}'$. However, in this case (3) would violate the full-rank condition stated in the assumptions of the theorem. Therefore we conclude that there does not exist a pair of parameter values $(\widetilde{\beta}, \lambda) \neq (\widetilde{\beta}', \lambda')$ in Θ that are observationally equivalent. From here it follows that $\mathcal{L}_0\left(\tilde{\beta}^*, \lambda^*\right) \neq \mathcal{L}_0\left(\tilde{\beta}, \lambda\right)$ for any $(\tilde{\beta}, \lambda) \in \Theta$ such that $(\tilde{\beta}, \lambda) \neq (\tilde{\beta}^*, \lambda^*)$. Since $\mathcal{L}_0\left(\tilde{\beta}^*, \lambda^*\right) \geq \mathcal{L}_0\left(\tilde{\beta}, \lambda\right)$ for all $(\tilde{\beta}, \lambda) \in \Theta$ (by the information inequality, implied in turn by Jensen's inequality), equation (2) implies that $\theta_1^* \equiv (\tilde{\beta}^*, \lambda^*)$ is the unique maximizer of $\mathcal{L}_0$ and is therefore identified. A standard mean value theorem expansion for maximum likelihood estimation then gives

$$\widehat{\theta}_1 = \theta_1^* + \frac{1}{n}\sum_{i=1}^{n} \psi_M(w_i) + o_p\left(n^{-1/2}\right),$$

where

$$\psi_M(w_i) \equiv H_0^{-1} \frac{\partial \ell(\theta_1; w_i)}{\partial \theta_1}$$

is the maximum likelihood influence function satisfying

$$n^{-1/2}\sum_{i=1}^{n} \psi_M(w_i) \rightarrow \mathcal{N}\left(0, H_0^{-1}\right), \quad \text{with} \quad H_0 = E\left[\frac{\partial \ell\left(\theta_1^*; W\right)}{\partial \theta_1} \frac{\partial \ell\left(\theta_1^*; W\right)}{\partial \theta_1}'\right].$$

■

2 Theorem 3

In this section we prove Theorem 3, guiding the reader through all the steps leading to the final proof. The steps rely on empirical process results. Throughout, we refer to various restrictions stated in the paper.

2.1 Euclidean classes of functions involved

The first step will consist of verifying that the relevant classes of functions involved have the *Euclidean* property as defined in Nolan and Pollard (1987, Definition 8) and Pakes and Pollard (1989, Definition 2.7). Let

$$\begin{aligned} \mathscr{H} &= \{f : \mathcal{X} \to \mathbb{R} \colon f(x) = \omega(x) \cdot K(x - x'; h) \colon x' \in \mathcal{X},\, h > 0\}, \\ \mathscr{G}_{1,k} &= \Big\{f : \mathcal{Y} \to \mathbb{R} \colon f(y) = 1\left[\mathcal{R}_\theta(y,x) \subseteq \mathcal{U}_k(x,y';\theta)\right] \text{ where } (y',x) \in \mathcal{W},\, \theta \in \Theta\Big\}, \\ \mathscr{G}_{2,k} &= \Big\{f : \mathbb{R}^2 \to \mathbb{R} \colon f(u_1,u_2) = 1\left[(u_1,u_2) \in \mathcal{U}_k(x,y;\theta)\right] \text{ for some } (y,x) \in \mathcal{W},\, \theta \in \Theta\Big\}. \end{aligned}$$

Since K is a function of bounded variation (see Restriction I2), $\mathscr{H}$ is a Euclidean class of functions for constant envelope $\overline{\omega} \cdot \overline{K}$, where $\overline{\omega} \equiv \sup \omega$ and $\overline{K} \equiv \sup |K|$ (see Pakes and Pollard (1989, Example 2.10)). Since the class of sets

$$\mathscr{D}_{1,k} = \Big\{y \in \mathcal{Y} \colon R_\theta(y,x) \subseteq \mathcal{U}_k(x,v;\theta) \text{ for some } (v,x) \in \mathcal{W},\, \theta \in \Theta\Big\}$$

is assumed to be a VC class of sets (see Restriction I3), $\mathscr{G}_{1,k}$ is a Euclidean class of functions for constant envelope 1 (see Pakes and Pollard (1989, p. 1033)). The same holds for the class of functions $\mathscr{G}_{k,2}$ since the class of sets

$$\mathscr{D}_{2,k} = \Big\{\mathcal{U}_k(x,y;\theta) \text{ for some } (y,x) \in \mathcal{W},\, \theta \in \Theta\Big\}$$

is assumed to be a VC class of sets in Restriction I3. Next, we turn our attention to $P_U(\mathcal{U}_k(x,y;\theta);\theta)$. Let

$$g(u) = \frac{e^{-u}}{(1+e^{-u})^2}, \quad G(u) = \frac{e^u}{1+e^u}.$$

denote the logistic pdf and cdf, respectively. According to our copula parameterization, the joint density of (u_1,u_2) is given by

$$f(u_1,u_2;\lambda) = g(u_1)g(u_2) \cdot \Big[(1 + \lambda \cdot \Big(1 - 2G(u_1) - 2G(u_2) + 4G(u_1)G(u_2)\Big)\Big].$$

Notice that

$$P_U(\mathcal{U}_k(x,y;\theta);\lambda) = \int_{u_2}\int_{u_1} 1\left[(u_1,u_2) \in \mathcal{U}_k(x,y;\theta)\right] \cdot f(u_1,u_2;\lambda)\, du_1 du_2.$$

Define

$$g_2(u) = \frac{e^{-u/2}}{2\cdot\left(1+e^{-u/2}\right)^2}.$$

g_2 is the pdf for a re-scaled Logistic random variable with variance $4\cdot\frac{\pi^2}{3}$ (instead of $\frac{\pi^2}{3}$). We have

$$\sup_u \left(\frac{g(u)}{g_2(u)}\right) = 2, \quad \sup_u \left(\frac{g(u)\cdot G(u)}{g_2(u)}\right) < 1.23.$$

Let $f_2(u_1,u_2) = g_2(u_1)\cdot g_2(u_2)$. Note that f_2 is a well-defined pdf. We can re-express

$$\begin{aligned} P_U(\mathcal{U}_k(x,y;\theta);\theta) &= \int_{u_2}\int_{u_1} 1\left[(u_1,u_2)\in\mathcal{U}_k(x,y;\theta)\right]\cdot\left(\frac{f(u_1,u_2;\lambda)}{f_2(u_1,u_2)}\right)\cdot f_2(u_1,u_2)du_1du_2 \\ &\equiv \int_{u_2}\int_{u_1} 1\left[(u_1,u_2)\in\mathcal{U}_k(x,y;\theta)\right]\cdot h(u_1,u_2;\lambda)\cdot f_2(u_1,u_2)du_1du_2, \end{aligned}$$

where

$$h(u_1,u_2;\lambda) \equiv \frac{f(u_1,u_2;\lambda)}{f_2(u_1,u_2)} = \frac{g(u_1)\cdot g(u_2)}{g_2(u_1)\cdot g_2(u_2)}\cdot\left[(1+\lambda\cdot\left(1-2G(u_1)-2G(u_2)+4G(u_1)G(u_2)\right)\right]$$

Let

$$\mathscr{G}_3 = \left\{h:\mathbb{R}^2\to\mathbb{R}\colon h(u_1,u_2) = \frac{f(u_1,u_2;\lambda)}{f_2(u_1,u_2)} \text{ for some } \lambda\in[-1,1]\right\}.$$

Recall that $-1\le\lambda\le 1$. Therefore, $|h|\le 24\ \forall\ h\in\mathscr{G}_3$. Also note that for any $\lambda,\lambda'\in[-1,1]$,

$$\begin{aligned} \left|h(u_1,u_2;\lambda)-h(u_1,u_2;\lambda')\right| &\le \left|[1-2G(u_1)-2G(u_2)+4G(u_1)G(u_2)]\cdot\frac{g(u_1)\cdot g(u_2)}{g_2(u_1)\cdot g_2(u_2)}\right|\cdot\left|\lambda-\lambda'\right| \\ &< 17\cdot\left|\lambda-\lambda'\right| \end{aligned}$$

From here, by Pakes and Pollard (1989, Lemma 2.13) we deduce that $\mathscr{G}_3$ is a Euclidean class of functions for the constant envelope $\sup_{u_1,u_2}|h(u_1,u_2;0)| + 2\cdot 17 = 4+17\cdot 2\cdot\sup_{\lambda\in[-1,1]}|\lambda-0| = 4+2\cdot 17 = 38$. Now define

$$\mathscr{G}_{4,k} = \{g_1g_2\colon g_1\in\mathscr{G}_{2,k},\ g_2\in\mathscr{G}_3\}.$$

From our previous arguments and Pakes and Pollard (1989, Lemma 2.14), the class of functions $\mathscr{G}_{4,k}$ is Euclidean for the constant envelope $1\cdot 38$. Now let

$$\mathscr{G}_{5,k} = \left\{\int_{u_2}\int_{u_1} h(u_1,u_2)\cdot f_2(u_1,u_2)du_1du_2 \text{ for some } h\in\mathscr{G}_{4,k}\right\}.$$

Note that $P_U(\mathcal{U}_k(x,y;\theta);\lambda) \in \mathscr{G}_{5,k}$. By Nolan and Pollard (1987, Lemma 20) or Sherman (1994, Lemma 5), the class of functions $\mathscr{G}_{5,k}$ is Euclidean for the constant envelope 38. Finally, let

$$\mathscr{F}_k = \left\{(g_1 - g_2)\cdot g_3\colon\, g_1 \in \mathscr{G}_{1,k},\, g_2 \in \mathscr{G}_{5,k},\, g_3 \in \mathscr{H}\right\}.$$

By Pakes and Pollard (1989, Lemma 2.14), the class of functions $\mathscr{F}_k$ is Euclidean for the constant envelope $39\cdot\overline{K}\cdot\overline{\omega} \equiv \overline{F}_k$. Notice that

$$\mathscr{F}_k = \{g : \mathcal{W} \to \mathbb{R}\colon\, g(y,x) = m_k(y,v,u;\theta)\cdot\omega(x)\cdot K(x-u;h) \text{ for some } (v,u)\in\mathcal{W},\, \theta\in\Theta,\, h>0\}.$$

Next, recall from Restriction I3 that the following is a VC class of sets,

$$\mathscr{D}_{5,k} = \Big\{w\in\mathcal{W}\colon\, c_1 \le T_k(w;\theta) < c_2 \text{ for some } \theta\in\Theta,\, c_1<c_2 \text{ in } \mathbb{R}\Big\}.$$

Therefore, the following is a Euclidean class of functions

$$\mathscr{G}_{6,k} = \{f : \mathcal{W}\to\mathbb{R}\colon\, f(w) = \omega(x)\cdot 1\left[-b \le T_k(w;\theta) < 0\right] \text{ for some } \theta\in\Theta \text{ and } b\in\mathbb{R}\},$$

for the constant envelope $\overline{\omega}$. Finally, recall again from Restriction I3 that the following is a VC class of sets,

$$\mathscr{D}_{6,k} = \Big\{(y,x)\in\mathcal{W}\colon\, T_k(w;\theta)\ge c \text{ for some } \theta\in\Theta,\, c\in\mathbb{R}\Big\}.$$

For $w_1 \equiv (y_1,x_1)$ and $w_2 \equiv (y_2,x_2)$ denote

$$v_k(w_2,w_1;\theta,h) \equiv \omega(x_1)\cdot\omega(x_2)\cdot m_k(y_2,y_1,x_1;\theta)\cdot\mathbf{K}(x_1-x_2;h)\cdot 1\left[T_k(w_1,\theta)\ge 0\right].$$

And consider the class of functions

$$\mathscr{V}_k = \{f : \mathcal{W}\times\mathcal{W}\to\mathbb{R}\colon\, f(w_1,w_2) = v_k(w_2,w_1;\theta,h) \text{ for some } \theta\in\Theta \text{ and } h>0.\}$$

From the VC property of the class of sets $\mathscr{G}_{6,k}$, the Euclidean properties of the classes of functions reviewed above and Pakes and Pollard (1989, Lemma 2.14), the class of functions $\mathscr{V}_k$ is Euclidean for the constant envelope $\overline{\omega}^2\cdot\overline{K}$.

2.2 Auxiliary maximal inequality results

Recall that we denote $w\equiv(y,x)$ and that the supports of Y and X are denoted as $\mathcal{Y}$ and $\mathcal{X}$, respectively. Also recall that the joint support of $W\equiv(Y,X)$ is denoted as $\mathcal{W}$ and that $\mathcal{W}^*$ is the projection of $\mathcal{X}^*$ (our pre-specified inference range) onto $\mathcal{W}$. Using the Euclidean properties for the classes of functions described in Section 2.1 we will now use the maximal inequality results in Sherman (1994) to derive asymptotic properties for some key empirical processes in our problem.

Let us begin with the process $\nu_{1n}^k(\cdot)$ defined as

$$\begin{aligned}\nu_{1n}^k(y,x;\theta) &= \frac{1}{n}\sum_{i=1}^{n}\Big(m_k(Y_i,y,x;\theta)\cdot\omega(X_i)\cdot K(X_i-x;h_n) - E\left[m_k(Y,y,x;\theta)\cdot\omega(X)\cdot K(X-x;h_n)\right]\Big)\\ &= h_n^z\cdot\Big(\widehat{T}_k(y,x;\theta) - E[\widehat{T}_k(y,x;\theta)]\Big).\end{aligned}$$

Note that here and throughout this supplement when it is helpful for clarity we make use of the minor abuse of notation $\widehat{T}_k(y,x;\theta) = \widehat{T}_k((y,x);\theta)$, $T_k(y,x;\theta) = T_k((y,x);\theta)$, and so forth, replacing $w = (y,x)$ with y,x as arguments before the semicolon. Now, using Sherman (1994, Corollary 4), the Euclidean property of the class of functions $\mathscr{F}_k$ described in Section 2.1 yields

$$\sup_{\substack{w\in\mathcal{W}\\ \theta\in\Theta}}\left|\nu_{1n}^k(y,x;\theta)\right| = O_p\left(\frac{1}{n^{1/2}}\right) \quad\Longrightarrow\quad \sup_{\substack{w\in\mathcal{W}\\ \theta\in\Theta}}\left|\widehat{T}_k(y,x;\theta) - E[\widehat{T}_k(y,x;\theta)]\right| = O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right).$$

Next note that Restrictions I1 and I2 coupled with an Mth order approximation yield

$$\sup_{\substack{w\in\mathcal{W}^*\\ \theta\in\Theta}}\left|E[\widehat{T}_k(y,x;\theta)] - T_k(y,x;\theta)\right| \le C_T\cdot h_n^M, \tag{4}$$

for some $C_T < \infty$. Therefore,

$$\begin{aligned}\sup_{\substack{w\in\mathcal{W}^*\\ \theta\in\Theta}}\left|\widehat{T}_k(y,x;\theta) - T_k(y,x;\theta)\right| &= O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right) + O\Big(h_n^M\Big),\\ \sup_{\substack{w\in\mathcal{W}^*\\ \theta\in\Theta}}\left|\widehat{T}_k(y,x;\theta)\right| &= O_p(1).\end{aligned} \tag{5}$$

Take any integer q and any $0 < \alpha < 1$. By Sherman (1994, Main Corollary), there exists a universal constant Λ (described there) such that,

$$E\left[\sup_{\substack{w\in\mathcal{W}\\ \theta\in\Theta}}\left|n^{1/2}\cdot\nu_{1n}^k(y,x;\theta)\right|^q\right] \le \Lambda\cdot\Big(\overline{\omega}\cdot\overline{K}\Big)^{q-\alpha},$$

Therefore,

$$E\left[\sup_{\substack{w\in\mathcal{W}\\ \theta\in\Theta}}\left|\widehat{T}_k(y,x;\theta) - E\Big[\widehat{T}_k(y,x;\theta)\Big]\right|^q\right] = O\left(\frac{1}{\Big(n^{1/2}\cdot h_n^z\Big)^q}\right) \quad \text{for any integer } q. \tag{6}$$

Next, note from (4) that

$$P\left[\sup_{\substack{w\in\mathcal{W}^*\\ \theta\in\Theta}}\left|\widehat{T}_k(y,x;\theta)-T_k(y,x;\theta)\right|>b_n\right]\leq P\left[\sup_{\substack{w\in\mathcal{W}^*\\ \theta\in\Theta}}\left|\widehat{T}_k(y,x;\theta)-E\left[\widehat{T}_k(y,x;\theta)\right]\right|>b_n-C_T\cdot h_n^M\right]$$

From here, (6) and Chebyshev's inequality (for higher-order moments) yields

$$P\left[\sup_{\substack{w\in\mathcal{W}^*\\ \theta\in\Theta}}\left|\widehat{T}_k(y,x;\theta)-T_k(y,x;\theta)\right|>b_n\right]=O\left(\frac{1}{\left(n^{1/2}\cdot h_n^z\cdot(b_n-C_T\cdot h_n^M)\right)^q}\right)\quad\text{for any integer } q. \tag{7}$$

Note from Restriction I2 that $n^{1/2}\cdot h_n^z\cdot(b_n-C_T\cdot h_n^M)\to\infty$. Denote

$$\mathcal{D}_n=1\left[\sup_{\substack{w\in\mathcal{W}^*\\ \theta\in\Theta}}\left|\widehat{T}_k(y,x;\theta)-T_k(y,x;\theta)\right|>b_n\right]. \tag{8}$$

Note that

$$\begin{aligned}\mathcal{D}_n&=\left(\mathcal{D}_n-E\left[\mathcal{D}_n\right]\right)+E\left[\mathcal{D}_n\right]=O_p\left(\sqrt{Var\left[\mathcal{D}_n\right]}\right)+E\left[\mathcal{D}_n\right]=O_p\left(\sqrt{E\left[\mathcal{D}_n\right]\cdot(1-E\left[\mathcal{D}_n\right])}\right)+E\left[\mathcal{D}_n\right]\\&=O_p\left(\sqrt{E\left[\mathcal{D}_n\right]}\right)+E\left[\mathcal{D}_n\right]=O_p\left(\frac{1}{\left(n^{1/2}\cdot h_n^z\cdot(b_n-C_T\cdot h_n^M)\right)^{q/2}}\right)\times\left[1+O\left(\frac{1}{\left(n^{1/2}\cdot h_n^z\cdot(b_n-C_T\cdot h_n^M)\right)^{q/2}}\right)\right]\\&=O_p\left(\frac{1}{\left(n^{1/2}\cdot h_n^z\cdot(b_n-C_T\cdot h_n^M)\right)^{q/2}}\right)\times\left[1+o(1)\right]=O_p\left(\frac{1}{\left(n^{1/2}\cdot h_n^z\cdot(b_n-C_T\cdot h_n^M)\right)^{q/2}}\right)\end{aligned} \tag{9}$$

for any integer q. Now take any $\Delta>0$ and consider $n^{1/2+\Delta}\cdot\mathcal{D}_n$. Let $\epsilon>0$ be as described in Restriction I2 and take any integer $q>(1+2\Delta)/\epsilon$ (i.e, $1/q+2\Delta/q<\epsilon$). By (9),

$$\begin{aligned}n^{1/2+\Delta}\cdot\mathcal{D}_n&=O_p\left(\frac{n^{1/2+\Delta}}{\left(n^{1/2}\cdot h_n^z\cdot(b_n-C_T\cdot h_n^M)\right)^{q/2}}\right)\\&=O_p\left(\frac{1}{\left(n^{1/2-1/q-2\Delta/q}\cdot h_n^z\cdot b_n-C_T\cdot n^{1/2-1/q-2\Delta/q}\cdot h_n^M\right)^{q/2}}\right)\\&=o_p(1),\end{aligned}$$

where the last result follows from the bandwidth convergence conditions in Restriction I2. Therefore,

$$\mathcal{D}_n=o_p\left(n^{-1/2-\Delta}\right)\ \forall\ \Delta>0. \tag{10}$$

Now let us proceed to analyze the process $\nu_{2n}^k(\cdot)$ defined as

$$\nu_{2n}^k(\theta) = \frac{1}{n}\sum_{i=1}^n \Big(\omega(x_i)\cdot 1\big[-2b_n \le T_k(w_i;\theta) < 0\big] - E\big[\omega(x_i)\cdot 1\big[-2b_n \le T_k(w_i;\theta) < 0\big]\big]\Big).$$

Using Sherman (1994, Corollary 4), the Euclidean property of the class of functions $\mathscr{G}_{6k}$ described in Section 2.1 yields

$$\sup_{\theta\in\Theta}\left|\nu_{2n}^k(\theta)\right| = O_p\left(\frac{1}{n^{1/2}}\right). \tag{11}$$

2.3 Proof of Lemma 2 in the paper

Let

$$\widehat{R}_k(\theta) \equiv \frac{1}{n}\sum_{i=1}^n \omega(x_i)\cdot \widehat{T}_k(w_i;\theta)\cdot 1\left[\widehat{T}_k(w_i,\theta) \ge -b_n\right].$$

Note that

$$\widehat{R}(\theta) = \sum_{k=1}^K \widehat{R}_k(\theta).$$

Now let

$$\widetilde{R}_k(\theta) \equiv \frac{1}{n}\sum_{i=1}^n \omega(x_i)\cdot \widehat{T}_k(w_i;\theta)\cdot 1\left[T_k(w_i,\theta) \ge 0\right].$$

As defined in the paper, let

$$\widetilde{R}(\theta) \equiv \frac{1}{n}\sum_{i=1}^n \omega(x_i)\left(\sum_{k=1}^K \widehat{T}_k(w_i;\theta)\cdot 1\left\{T_k(w_i,\theta) \ge 0\right\}\right),$$

Note that

$$\widetilde{R}(\theta) = \sum_{k=1}^K \widetilde{R}_k(\theta).$$

Lemma 2 asserts that, under Restrictions I1-I4,

$$\sup_{\theta\in\Theta}\left|\widetilde{R}(\theta) - \widehat{R}(\theta)\right| = o_p\left(n^{-1/2-\epsilon}\right),$$

where $\epsilon > 0$ is as described in Restriction I2. We will prove the lemma by showing that

$$\sup_{\theta\in\Theta}\left|\widetilde{R}_k(\theta) - \widehat{R}_k(\theta)\right| = o_p\left(n^{-1/2-\epsilon}\right) \quad \text{for } k = 1,\dots,K.$$

Let

$$\xi_{k,n}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\omega(x_i)\cdot\left|\widehat{T}_k(w_i;\theta)\right|\cdot\left|1[\widehat{T}_k(w_i;\theta)\geq -b_n] - 1[T_k(w_i;\theta)\geq 0]\right|.$$

Note that $\xi_{k,n}(\cdot)\geq 0$ and

$$\left|\widetilde{R}_k(\theta) - \widehat{R}_k(\theta)\right| \leq \xi_{k,n}(\theta).$$

We have

$$\begin{aligned}
&\left|1[\widehat{T}_k(y_i,x_i;\theta)\geq -b_n] - 1[T_k(w_i;\theta)\geq 0]\right| \\
&\quad = 1\Big[\widehat{T}_k(w_i;\theta)\geq -b_n,\ -2b_n\leq T_k(w_i;\theta)<0\Big] + 1\Big[\widehat{T}_k(w_i;\theta)\geq -b_n,\ T_k(w_i;\theta)<-2b_n\Big] \\
&\quad + 1\Big[\widehat{T}_k(w_i;\theta)<-b_n,\ T_k(w_i;\theta)\geq 0\Big] \\
&\quad \leq 1\Big[-2b_n\leq T_k(w_i;\theta)<0\Big] + 1\Big[\left|\widehat{T}_k(w_i;\theta)-T_k(w_i;\theta)\right|\geq b_n\Big].
\end{aligned}$$

From the above arguments, we have

$$\begin{aligned}
\xi_{k,n}(\theta) &\leq \frac{1}{n}\sum_{i=1}^{n}\omega(x_i)\cdot\left|T_k(w_i;\theta)\right|\cdot 1\Big[-2b_n\leq T_k(w_i;\theta)<0\Big] \\
&+ \frac{1}{n}\sum_{i=1}^{n}\omega(x_i)\cdot\left|\widehat{T}_k(w_i;\theta)-T_k(w_i;\theta)\right|\cdot 1\Big[-2b_n\leq T_k(w_i;\theta)<0\Big] \\
&+ \frac{1}{n}\sum_{i=1}^{n}\omega(x_i)\cdot\left|\widehat{T}_k(w_i;\theta)\right|\cdot 1\Big[\left|\widehat{T}_k(w_i;\theta)-T_k(w_i;\theta)\right|\geq b_n\Big]
\end{aligned}$$

Then,

$$\begin{aligned}
\xi_{k,n}(\theta) \leq &\left(2b_n + \sup_{\substack{w\in\mathcal{W}^*\\ \theta\in\Theta}}\left|\widehat{T}_k(w;\theta)-T_k(w;\theta)\right|\right)\times\frac{1}{n}\sum_{i=1}^{n}\omega(x_i)\cdot 1\Big[-2b_n\leq T_k(w_i;\theta)<0\Big] \\
&+ 1\left[\sup_{\substack{w\in\mathcal{W}^*\\ \theta\in\Theta}}\left|\widehat{T}_k(w;\theta)-T_k(w;\theta)\right|\geq b_n\right]\times\sup_{\substack{w\in\mathcal{W}^*\\ \theta\in\Theta}}\left|\widehat{T}_k(w;\theta)\right|\cdot\overline{\omega}
\end{aligned}$$

Using the results in equations (5) and (10),

$$\begin{aligned}
&\xi_{k,n}(\theta) \\
&\leq \left(2b_n + O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right) + O\Big(h_n^M\Big)\right)\times\frac{1}{n}\sum_{i=1}^{n}\omega(x_i)\cdot 1\Big[-2b_n\leq T_k(w_i;\theta)<0\Big] + o_p\Big(n^{-1/2-\Delta}\Big)\ \forall\ \Delta>0 \\
&= \left(2b_n + O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right) + O\Big(h_n^M\Big)\right)\times\Big(\nu_{2k}(\theta) + E\Big[\omega(X)\cdot 1\Big[-2b_n\leq T_k(Y,X;\theta)<0\Big]\Big]\Big) \\
&+ o_p\Big(n^{-1/2-\Delta}\Big)\ \forall\ \Delta>0
\end{aligned}$$

Let n_0 denote the smallest n such that $2b_n \leq \overline{b}$, where $\overline{b}$ is as described in Restriction I4. By the conditions described there, we have

$$\sup_{\theta\in\Theta} E\Big[\omega(X)\cdot 1\Big[-2b_n \leq T_k(Y,X;\theta) < 0\Big] \leq 2\overline{A}\cdot b_n \quad \forall\ n \geq n_0.$$

Combining this with the result in (11),

$$\begin{aligned}
\sup_{\theta\in\Theta} \xi_{k,n}(\theta) &= \left(2b_n + O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right) + O\Big(h_n^M\Big)\right)\times\left(O_p\left(\frac{1}{n^{1/2}}\right) + O(b_n)\right) + o_p\Big(n^{-1/2-\Delta}\Big)\ \forall\ \Delta > 0 \\
&= \left(2b_n + O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right) + O\Big(h_n^M\Big)\right)\times b_n \times\left(O_p\left(\frac{1}{n^{1/2}\cdot b_n}\right) + O(1)\right) + o_p\Big(n^{-1/2-\Delta}\Big)\ \forall\ \Delta > 0 \\
&= \left(2b_n + O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right) + O\Big(h_n^M\Big)\right)\times b_n \times\Big(o_p(1) + O(1)\Big) + o_p\Big(n^{-1/2-\Delta}\Big)\ \forall\ \Delta > 0 \\
&= O_p\left(b_n^2\right) + O_p\left(\frac{b_n}{n^{1/2}\cdot h_n^z}\right) + O\left(h_n^M\cdot b_n\right) + o_p\Big(n^{-1/2-\Delta}\Big)\ \forall\ \Delta > 0 \\
&= o_p\left(n^{-1/2-\epsilon}\right),
\end{aligned}$$

where $\epsilon > 0$ is as described in Restriction I2. Therefore, under Restrictions I1-I4,

$$\sup_{\theta\in\Theta}\left|\widetilde{R}_k(\theta) - \widehat{R}_k(\theta)\right| = o_p(n^{-1/2-\epsilon}) \quad \text{for } k = 1,\ldots,K.$$

From here,

$$\sup_{\theta\in\Theta}\left|\widetilde{R}(\theta) - \widehat{R}(\theta)\right| \leq \sum_{k=1}^{K}\left[\sup_{\theta\in\Theta}\left|\widetilde{R}_k(\theta) - \widehat{R}_k(\theta)\right|\right] = o_p(n^{-1/2-\epsilon}),$$

where $\epsilon > 0$ is as described in Restriction I2. This proves Lemma 2 in the paper. ■

2.4 Using Lemma 2 to setup the result in Theorem 3

For $w_1 \equiv (y_1, x_1)$ and $w_2 \equiv (y_2, x_2)$, let

$$v_k(w_2, w_1; \theta, h) \equiv \omega(x_1)\cdot\omega(x_2)\cdot m_k(y_2, y_1, x_1;\theta)\cdot \mathbf{K}(x_1 - x_2; h)\cdot 1\left[T_k(w_1,\theta) \geq 0\right]$$

From Lemma 2,

$$v_k(w_\ell, w_i; \theta, h) \equiv \omega(x_i)\cdot\omega(x_\ell)\cdot m_k(y_\ell, y_i, x_i;\theta)\cdot \mathbf{K}(x_i - x_\ell; h)\cdot 1\left[T_k(w_i,\theta) \geq 0\right]. \qquad (12)$$

From Lemma 2,

$$
\begin{aligned}
\widehat{R}(\theta) &= \sum_{k=1}^{K}\left(\frac{1}{n}\sum_{i=1}^{n}\omega(x_i)\cdot\widehat{T}_k(w_i;\theta)\cdot 1\left[T_k(w_i,\theta)\geq 0\right]\right)+\xi_n(\theta) \\
&= \sum_{k=1}^{K}\left(\frac{1}{n^2\cdot h_n^z}\sum_{i=1}^{n}\sum_{\ell=1}^{n}v_k(w_\ell,w_i;\theta,h_n)\right)+\xi_n(\theta),
\end{aligned}
\tag{13}
$$

where $\sup_{\theta\in\Theta}|\xi_n(\theta)| = o_p(n^{-1/2-\epsilon})$. Let

$$
\widetilde{v}_k(w_1,w_2;\theta,h) \equiv v_k(w_1,w_2;\theta,h)+v_k(w_2,w_1;\theta,h) \tag{14}
$$

Consider the U-process $U_{k,n}$ of order-2 described as

$$
U_{k,n}(\theta) = \binom{n}{2}^{-1}\sum_{i<j}\widetilde{v}_k(w_i,w_j;\theta,h_n) \tag{15}
$$

where $\sum_{i<j}$ denotes the sum over all $\binom{n}{2}$ combinations (i,j) out of n. Then,

$$
\widehat{R}(\theta) = \left(\frac{n-1}{n}\right)\times\frac{1}{2\cdot h_n^z}\cdot\sum_{k=1}^{K}\left(U_{k,n}(\theta)+\frac{1}{n}\times\left[\frac{1}{n}\sum_{i=1}^{n}v_k(w_i,w_i;\theta,h_n)\right]\right)+\xi_n(\theta)
$$

Note that

$$
\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}v_k(w_i,w_i;\theta,h_n)\right| \leq \overline{\omega}^2\cdot\overline{K}.
$$

Therefore,

$$
\frac{1}{n\cdot h_n^z}\times\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}v_k(w_i,w_i;\theta,h_n)\right| \leq \frac{\overline{\omega}^2\cdot\overline{K}}{n\cdot h_n^z} = o\left(n^{-1/2-\epsilon}\right),
$$

where $\epsilon > 0$ is as described in Restriction I2. Thus,

$$
\widehat{R}(\theta) = \left(\frac{n-1}{n}\right)\times\frac{1}{2\cdot h_n^z}\cdot\sum_{k=1}^{K}U_{k,n}(\theta)+\vartheta_n(\theta), \quad \text{where} \quad \sup_{\theta\in\Theta}\left|\vartheta_n(\theta)\right| = o_p\left(n^{-1/2-\epsilon}\right) \tag{16}
$$

Next we study the asymptotic properties of the U-processes $U_{k,n}(\cdot)$.

2.4.1 Asymptotic properties of the U-process $U_{k,n}(\cdot)$

Note that $v_k(w_2, w_1; \theta, h)$ belongs to the class of functions

$$\mathscr{V}_k = \{f : \mathcal{W} \times \mathcal{W} \to \mathbb{R}\colon f(w_1, w_2) = v_k(w_2, w_1; \theta, h) \ \text{ for some } \theta \in \Theta \text{ and } h > 0.\}.$$

As we pointed out in Section 2.1, this class of functions is Euclidean for the constant envelope $\overline{\omega}^2 \cdot \overline{K}$. Therefore $\widetilde{v}_k(w_1, w_2; \theta, h)$ belongs in a Euclidean class of functions with envelope $2 \cdot \overline{\omega}^2 \cdot \overline{K}$. Consider the U-process $U_{k,n}(\cdot)$ defined in (15). Let

$$\begin{aligned}
\mu_k(\theta, h) &\equiv E\left[\widetilde{v}_k(W_1, W_2; \theta, h)\right], \\
\widetilde{g}_k(w; \theta, h) &\equiv E\left[\widetilde{v}_k(w, W; \theta, h)\right] - \mu_k(\theta, h), \\
\widetilde{r}_k(w_1, w_2; \theta, h) &\equiv \widetilde{v}_k(w_1, w_2; \theta, h) - \mu_k(\theta, h) - \widetilde{g}_k(w_1; \theta, h) - \widetilde{g}_k(w_2; \theta, h).
\end{aligned}$$

The Hoeffding decomposition (projection) of $U_{k,n}(\theta)$ is given by (see Serfling (1980, pages 177-178) or Sherman (1994, equations (6)-(7))),

$$U_{k,n}(\theta) = \mu_k(\theta, h_n) + \frac{2}{n} \sum_{i=1}^{n} \widetilde{g}_k(w_i; \theta, h_n) + \binom{n}{2}^{-1} \sum_{i<j} \widetilde{r}_k(w_i, w_j; \theta, h_n). \tag{17}$$

The last term is a degenerate U-process of order 2. By the Euclidean property of the class of functions $\mathscr{V}_k$, Sherman (1994, Corollary 4) yields

$$\sup_{\substack{\theta \in \Theta \\ h>0}} \left| \binom{n}{2}^{-1} \sum_{i<j} \widetilde{r}_k(w_i, w_j; \theta, h) \right| = O_p\left(\frac{1}{n}\right) \quad \Longrightarrow \quad \sup_{\theta \in \Theta} \left| \binom{n}{2}^{-1} \sum_{i<j} \widetilde{r}_k(w_i, w_j; \theta, h_n) \right| = O_p\left(\frac{1}{n}\right)$$

As defined in Restriction I5, let

$$\begin{aligned}
\lambda_{2k}(y, x; \theta) &= E\left[m_k(Y, y, x; \theta) \cdot 1\left[T_k(Y, X; \theta) \geq 0\right] \, | X = x\right], \\
g_k(y, x; \theta) &= \omega(x) \cdot \left(T_k(w; \theta)\right)_+ + \omega(x)^2 \cdot \lambda_{2k}(y, x; \theta) \cdot f_X(x).
\end{aligned}$$

Under Restrictions I1, I2 and I5, an M^{th}-order approximation and dominated convergence arguments yield[1]

$$E\left[\widetilde{v}_k(w, W; \theta, h_n)\right] = h_n^z \times \left(g_k(y, x; \theta) + \mathcal{B}^a_{k,n}(w, \theta)\right), \quad \text{where} \quad \sup_{\substack{w \in \mathcal{W} \\ \theta \in \Theta}} \left|\mathcal{B}^a_{k,n}(w, \theta)\right| = O(h_n^M). \tag{18}$$

[1]Notice the usefulness of having the term $\omega(x_1) \cdot \omega(x_2)$ in $v_k(w_2, w_1; \theta, h)$, as it means that both w_1 and w_2 are constrained to belong to $\mathcal{W}^*$.

From here, a dominated convergence argument yields,

$$\mu_k(\theta,h_n) \equiv E\left[\widetilde{v}_k(W_1,W_2;\theta,h_n)\right] = h_n^z\cdot\left(E\left[g_k(Y,X;\theta)\right] + \mathcal{B}^b_{k,n}(\theta)\right) \quad \text{where} \quad \sup_{\theta\in\Theta}\left|\mathcal{B}^b_{k,n}(\theta)\right| = O(h_n^M)$$

Let

$$R_k(\theta) \equiv E\left[\omega(X)\cdot\left(T_k(Y,X;\theta)\right)_+\right].$$

Note that $R(\theta) = \sum_{k=1}^K R_k(\theta)$. By iterated expectations, we have $E\left[g_k(Y,X;\theta)\right] = 2\cdot R_k(\theta)$, and the result above yields

$$\mu_k(\theta,h_n) \equiv E\left[\widetilde{v}_k(W_1,W_2;\theta,h_n)\right] = h_n^z\times\left(2\cdot R_k(\theta) + \mathcal{B}^b_{k,n}(\theta)\right) \quad \text{where} \quad \sup_{\theta\in\Theta}\left|\mathcal{B}^b_{k,n}(\theta)\right| = O(h_n^M),$$

Let

$$\psi_R^k(y_i,x_i;\theta) \equiv g_k(y_i,x_i;\theta) - E\left[g_k(Y,X;\theta)\right].$$

Plugging the previous results back into the Hoeffding decomposition (17), we obtain

$$\begin{aligned}
U_{k,n}(\theta) &= h_n^z\times\left(E\left[g_k(Y,X;\theta)\right] + \frac{2}{n}\sum_{i=1}^n \psi_R^k(y_i,x_i;\theta) + \mathcal{B}_{k,n}(\theta)\right) + \varsigma_{k,n}(\theta) \\
&= h_n^z\times\left(2\cdot R_k(\theta) + \frac{2}{n}\sum_{i=1}^n \psi_R^k(y_i,x_i;\theta) + \mathcal{B}_{k,n}(\theta)\right) + \varsigma_{k,n}(\theta), \quad \text{where} \\
&\qquad \sup_{\theta\in\Theta}\left|\mathcal{B}_{k,n}(\theta)\right| = O(h_n^M), \quad \sup_{\theta\in\Theta}\left|\varsigma_{k,n}(\theta)\right| = O_p\left(\frac{1}{n}\right).
\end{aligned} \tag{19}$$

2.5 Proof of Theorem 3

Define

$$\psi_R(y_i,x_i;\theta) \equiv \sum_{k=1}^K \psi_R^k(y_i,x_i;\theta) = \sum_{k=1}^K\left(g_k(y_i,x_i;\theta) - E\left[g_k(Y,X;\theta)\right]\right).$$

Note that $R(\theta) = \sum_{k=1}^K R_k(\theta)$. From (19),

$$\begin{aligned}
\frac{1}{2\cdot h_n^z}\cdot\sum_{k=1}^K U_{k,n}(\theta) &= R(\theta) + \frac{1}{n}\sum_{i=1}^n \psi_R(y_i,x_i;\theta) + \varrho_n(\theta), \quad \text{where} \\
&\sup_{\theta\in\Theta}\left|\varrho_n(\theta)\right| = O\left(h_n^M\right) + O_p\left(\frac{1}{n\cdot h_n^z}\right) = o_p\left(n^{-1/2-\epsilon}\right),
\end{aligned}$$

where $\epsilon > 0$ is described in Restriction I2. Plugging this back into equation (16),

$$\begin{aligned}
\widehat{R}(\theta) &= \left(\frac{n-1}{n}\right) \times \frac{1}{2 \cdot h_n^z} \cdot \sum_{k=1}^{K} U_{k,n}(\theta) + \vartheta_n(\theta), \quad \text{where} \quad \sup_{\theta \in \Theta} \left|\vartheta_n(\theta)\right| = o_p\left(n^{-1/2-\epsilon}\right) \\
&= \left(\frac{n-1}{n}\right) \times \left[R(\theta) + \frac{1}{n}\sum_{i=1}^{n} \psi_R(y_i, x_i; \theta) + \rho_n(\theta)\right] + \vartheta_n(\theta) \\
&= R(\theta) + \frac{1}{n}\sum_{i=1}^{n} \psi_R(y_i, x_i; \theta) + \rho_n(\theta) + \vartheta_n(\theta) - \frac{1}{n} \times \left[R(\theta) + \frac{1}{n}\sum_{i=1}^{n} \psi_R(y_i, x_i; \theta) + \rho_n(\theta)\right]
\end{aligned}$$

where $\sup_{\theta \in \Theta} \left|\rho_n(\theta) + \vartheta_n(\theta)\right| = o_p\left(n^{-1/2-\epsilon}\right)$, with $\epsilon > 0$ is described in Restriction I2. Next, by Sherman (1994, Lemma 5 and Corollary 4), the Euclidean properties of the classes of functions involved yield the result

$$\sup_{\theta \in \Theta} \left|\frac{1}{n}\sum_{i=1}^{n} \psi_R(y_i, x_i; \theta)\right| = O_p(n^{-1/2})$$

From here, since $\sup_{\theta \in \Theta} |R(\theta)| = O(1)$, we have

$$\sup_{\theta \in \Theta} \left|\frac{1}{n} \times \left[R(\theta) + \frac{1}{n}\sum_{i=1}^{n} \psi_R(y_i, x_i; \theta) + \rho_n(\theta)\right]\right| = o_p\left(n^{-1/2-\epsilon}\right).$$

Combining all our results, we have

$$\widehat{R}(\theta) = R(\theta) + \frac{1}{n}\sum_{i=1}^{n} \psi_R(y_i, x_i; \theta) + \varepsilon_n(\theta), \ \text{where} \ \sup_{\theta \in \Theta} |\varepsilon_n(\theta)| = o_p\left(n^{-1/2-\epsilon}\right),$$

where $\epsilon > 0$ is described in Restriction I2. This is the statement of Theorem 3. ■

3 Theorems 4 and 5

3.1 Asymptotic properties of our variance-covariance estimators

The paper outlined all the arguments leading to the statements in Theorems 4 and 5. A key result we invoked there was the following,

$$\begin{aligned} \left\| \widehat{H}_0^{-1} - H_0^{-1} \right\| &= O_p(n^{-1/2}), \\ \sup_{\theta\in\Theta} \left\| \widehat{\Sigma}_{MR}(\theta) - \Sigma_{MR}(\theta) \right\| &= O_p\left(\frac{1}{n^{1/2}\cdot h_n^z} \right), \\ \sup_{\theta\in\Theta} \left| \widehat{\sigma}_R^2(\theta) - \sigma_R^2(\theta) \right| &= O_p\left(\frac{1}{n^{1/2}\cdot h_n^z} \right). \end{aligned} \tag{20}$$

We present a step-by-step proof of the result in (20) next. As before, we will refer to restrictions stated in the paper. We denoted

$$-E\left[\frac{\partial^2 \ell(\theta_1^*, W)}{\partial\theta_1^* \partial\theta_1'} \right] \equiv H_0,$$

and

$$\widehat{H}_0 = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \ell(\widehat{\theta}_1, w_i)}{\partial\theta_1\partial\theta_1'}.$$

We will show first that, under the conditions of Theorem 3 and Restriction I6, we have

$$\left\| \widehat{H}_0^{-1} - H_0^{-1} \right\| = O_p(n^{-1/2}).$$

As we did in the paper, define

$$\phi_1(t_1,t_2;\lambda) \equiv \tfrac{\partial F(t_1,t_2;\lambda)}{\partial t_1}\cdot H(t_1,t_2;\lambda)^{-1}, \quad \phi_2(t_1,t_2;\lambda) \equiv \tfrac{\partial F(t_1,t_2;\lambda)}{\partial t_2}\cdot H(t_1,t_2;\lambda)^{-1},$$
$$\phi_3(t_1,t_2;\lambda) \equiv \tfrac{\partial F(t_1,t_2;\lambda)}{\partial \lambda}\cdot H(t_1,t_2;\lambda)^{-1}.$$

As in Restriction I6, for $(\ell,m)\in\{1,2\}\times\{1,2\}$, define the following real-valued functions,

$$\begin{aligned} \delta_{\ell,m}(y,t_1,t_2;\lambda) &\equiv \frac{\partial\phi_\ell(t_1,t_2;\lambda)}{\partial t_m}\cdot\Big(1[y=(0,0)]-F(t_1,t_2;\lambda)\Big) - \phi_\ell(t_1,t_2;\lambda)\frac{\partial F(t_1,t_2;\lambda)}{\partial t_m}, \\ \eta_\ell(y,t_1,t_2;\lambda) &\equiv \frac{\partial\phi_\ell(t_1,t_2;\lambda)}{\partial \lambda}\cdot\Big(1[y=(0,0)]-F(t_1,t_2;\lambda)\Big) - \phi_\ell(t_1,t_2;\lambda)\frac{\partial F(t_1,t_2;\lambda)}{\partial \lambda}, \\ \Upsilon(y,t_1,t_2;\lambda) &\equiv \frac{\partial\phi_3(t_1,t_2;\lambda)}{\partial \lambda}\cdot\Big(1[y=(0,0)]-F(t_1,t_2;\lambda)\Big) - \phi_3(t_1,t_2;\lambda)\frac{\partial F(t_1,t_2;\lambda)}{\partial \lambda}. \end{aligned}$$

We have

$$E\left[\frac{\partial^2\ell(\theta_1,W)}{\partial\theta_1\partial\theta_1'} \right] = \begin{pmatrix} A_{11}(\theta_1) & A_{12}(\theta_1) \\ A_{12}(\theta_1)' & A_{22}(\theta_1) \end{pmatrix},$$

where

$$A_{11}(\theta_1) = E\left[\sum_{\ell=1}^{2}\sum_{m=1}^{2} Z_\ell Z_m' \delta_{\ell,m}(Y, Z_1\widetilde{\beta}, Z_2\widetilde{\beta}; \lambda)\right],$$

$$A_{12}(\theta_1) = E\left[\sum_{\ell=1}^{2} Z_\ell \eta_\ell(Y, Z_1\widetilde{\beta}, Z_2\widetilde{\beta}; \lambda)\right],$$

$$A_{22}(\theta_1) = E\left[\Upsilon(Y, Z_1\widetilde{\beta}, Z_2\widetilde{\beta}; \lambda)\right].$$

And,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \ell(\theta_1, w_i)}{\partial\theta_1 \partial\theta_1'} = \begin{pmatrix} \widehat{A}_{11}(\theta_1) & \widehat{A}_{12}(\theta_1) \\ \widehat{A}_{12}(\theta_1)' & \widehat{A}_{22}(\theta_1) \end{pmatrix},$$

where

$$\widehat{A}_{11}(\theta_1) = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{\ell=1}^{2}\sum_{m=1}^{2} z_{\ell i} z_{mi}' \delta_{\ell,m}(y_i, z_{1i}\widetilde{\beta}, z_{2i}\widetilde{\beta}; \lambda)\right),$$

$$\widehat{A}_{12}(\theta_1) = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{\ell=1}^{2} z_{\ell i} \eta_\ell(y_i, z_{1i}\widetilde{\beta}, z_{2i}\widetilde{\beta}; \lambda)\right),$$

$$\widehat{A}_{22}(\theta_1) = \frac{1}{n}\sum_{i=1}^{n} \Upsilon(y_i, z_{1i}\widetilde{\beta}, z_{2i}\widetilde{\beta}; \lambda).$$

Take observation $i \in \{1, \ldots, n\}$. A first-order approximation yields

$$\delta_{\ell,m}(y_i, z_{1i}\widehat{\widetilde{\beta}}, z_{2i}\widehat{\widetilde{\beta}}; \widehat{\lambda}) = \delta_{\ell,m}(y_i, z_{1i}\widetilde{\beta}^*, z_{2i}\widetilde{\beta}^*; \lambda^*) + \sum_{j=1}^{2} \frac{\partial \delta_{\ell,m}(y_i, z_{1i}\overline{\widetilde{\beta}}_i, z_{2i}\overline{\widetilde{\beta}}_i; \overline{\lambda}_i)}{\partial t_j} \cdot z_{ji}'\left(\widehat{\widetilde{\beta}} - \widetilde{\beta}^*\right)$$
$$+ \frac{\partial \delta_{\ell,m}(y_i, z_{1i}\overline{\widetilde{\beta}}_i, z_{2i}\overline{\widetilde{\beta}}_i; \overline{\lambda}_i)}{\partial\lambda} \cdot \left(\widehat{\lambda} - \lambda^*\right),$$

where $\overline{\theta}_{1i} \equiv \left(\overline{\widetilde{\beta}}_i', \overline{\lambda}_i\right)'$ belongs in the line segment connecting $\widehat{\theta}_1$ and θ_1^*. From Theorem 2, we have that, wp$\rightarrow 1$, our ML estimator $\widehat{\theta}_1$ belongs in the neighborhood $\mathcal{N}$ which contains θ_1^* and whose properties are described in Restriction I6. Therefore $\overline{\theta}_{1i}$ also belongs in $\mathcal{N}$ for all i. Therefore, from Theorem 2 and Restriction I6, wp$\rightarrow 1$, we have

$$\left|\delta_{\ell,m}(y_i, z_{1i}\widehat{\widetilde{\beta}}, z_{2i}\widehat{\widetilde{\beta}}; \widehat{\lambda}) - \delta_{\ell,m}(y_i, z_{1i}\widetilde{\beta}^*, z_{2i}\widetilde{\beta}^*; \lambda^*)\right| \leq D(w_i) \cdot \left(\sum_{j=1}^{2} \left\|z_{ji}\right\| \cdot \left\|\widehat{\widetilde{\beta}} - \widetilde{\beta}^*\right\| + \left|\widehat{\lambda} - \lambda^*\right|\right), \tag{21}$$

for $(\ell, m) \in \{1,2\} \times \{1,2\}$. And from here we obtain,

$$\begin{aligned}\left\|\widehat{A}_{11}(\widehat{\theta}_1) - \widehat{A}_{11}(\theta_1^*)\right\| &\leq \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{\ell=1}^{2}\sum_{m=1}^{2}\|z_{\ell i}\| \cdot \|z_{mi}\| \cdot \left(\sum_{j=1}^{2}\left\|z_{ji}\right\| \cdot D(w_i) \cdot \left\|\widehat{\widetilde{\beta}} - \widetilde{\beta}^*\right\| + D(w_i) \cdot \left|\widehat{\lambda} - \lambda^*\right|\right)\right) \\ &= \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{\ell=1}^{2}\sum_{m=1}^{2}\sum_{j=1}^{2}\|z_{\ell i}\| \cdot \|z_{mi}\| \cdot \left\|z_{ji}\right\| \cdot D(w_i)\right)\right)}_{O_p(1) \text{ by the conditions in Restriction I6}} \cdot \underbrace{\left\|\widehat{\widetilde{\beta}} - \widetilde{\beta}^*\right\|}_{O_p(n^{-1/2})} \\ &\quad + \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{\ell=1}^{2}\sum_{m=1}^{2}\|z_{\ell i}\| \cdot \|z_{mi}\| \cdot D(w_i)\right)\right)}_{O_p(1) \text{ by the conditions in Restriction I6}} \cdot \underbrace{\left|\widehat{\lambda} - \lambda^*\right|}_{O_p(n^{-1/2})} \\ &= O_p(n^{-1/2}).\end{aligned}$$

Theorem 2 and Restriction I6 also imply that the same inequality in (21) holds for η_ℓ, and wp$\to 1$, we have

$$\left|\eta_\ell(y_i, z_{1i}\widehat{\widetilde{\beta}}, z_{2i}\widehat{\widetilde{\beta}}; \widehat{\lambda}) - \eta_\ell(y_i, z_{1i}\widetilde{\beta}^*, z_{2i}\widetilde{\beta}^*; \lambda^*)\right| \leq D(w_i) \cdot \left(\sum_{j=1}^{2}\left\|z_{ji}\right\| \cdot \left\|\widehat{\widetilde{\beta}} - \widetilde{\beta}^*\right\| + \left|\widehat{\lambda} - \lambda^*\right|\right), \tag{21'}$$

for $\ell \in \{1,2\}$. Therefore, wp$\to 1$,

$$\begin{aligned}\left\|\widehat{A}_{12}(\widehat{\theta}_1) - \widehat{A}_{12}(\theta_1^*)\right\| &\leq \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{2}\sum_{j=1}^{2}\|z_{\ell i}\| \cdot \left\|z_{ji}\right\| \cdot D(w_i)\right)}_{O_p(1) \text{ by Restriction I6}} \cdot \underbrace{\left\|\widehat{\widetilde{\beta}} - \widetilde{\beta}^*\right\|}_{O_p(n^{-1/2})} + \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{2}\|z_{\ell i}\| \cdot D(w_i)\right)}_{O_p(1) \text{ by Restriction I6}} \cdot \underbrace{\left|\widehat{\lambda} - \lambda^*\right|}_{O_p(n^{-1/2})} \\ &= O_p(n^{-1/2}).\end{aligned}$$

Finally, the inequality in (21) also holds for Υ by Theorem 2 and Restriction I6, and wp$\to 1$,

$$\left|\Upsilon(y_i, z_{1i}\widehat{\widetilde{\beta}}, z_{2i}\widehat{\widetilde{\beta}}; \widehat{\lambda}) - \Upsilon(y_i, z_{1i}\widetilde{\beta}^*, z_{2i}\widetilde{\beta}^*; \lambda^*)\right| \leq D(w_i) \cdot \left(\sum_{j=1}^{2}\left\|z_{ji}\right\| \cdot \left\|\widehat{\widetilde{\beta}} - \widetilde{\beta}^*\right\| + \left|\widehat{\lambda} - \lambda^*\right|\right), \tag{21''}$$

and from here, wp$\to 1$ we have

$$\left|\widehat{A}_{22}(\widehat{\theta}_1)-\widehat{A}_{22}(\theta_1^*)\right| \leq \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2}D(w_i)\cdot\left\|z_{ji}\right\|\right)}_{O_p(1) \text{ by Restriction I6}}\cdot\underbrace{\left\|\widehat{\widetilde{\beta}}-\widetilde{\beta}^*\right\|}_{O_p(n^{-1/2})}+\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}D(w_i)\right)}_{O_p(1) \text{ by Restriction I6}}\cdot\underbrace{\left|\widehat{\lambda}-\lambda^*\right|}_{O_p(n^{-1/2})}$$
$$= O_p(n^{-1/2})$$

Next, the finite-variance conditions described in Restriction I6, and a Central Limit Theorem yield

$$\left\|\widehat{A}_{11}(\theta_1^*)-A_{11}(\theta_1^*)\right\| = O_p(n^{-1/2}),$$
$$\left\|\widehat{A}_{12}(\theta_1^*)-A_{12}(\theta_1^*)\right\| = O_p(n^{-1/2}),$$
$$\left|\widehat{A}_{22}(\theta_1^*)-A_{22}(\theta_1^*)\right| = O_p(n^{-1/2}).$$

From here, the invertibility condition in Restriction I6(i) finally yields

$$\left\|\widehat{H}_0^{-1}-H_0^{-1}\right\| = O_p(n^{-1/2}) \tag{22}$$

Let us continue analyzing the asymptotic properties of the remaining elements of our regularized variance matrix $\widehat{\Sigma}(\theta)$. In our construction we denoted $\sigma_R^2(\theta) \equiv E\left[\psi_R(W;\theta)^2\right]$ and $\Sigma_{MR}(\theta) \equiv E\left[\psi_M(W)\psi_R(W;\theta)\right]$. As a reminder (from Theorem 3), the influence function ψ_R is given by

$$\psi_R(y_i,x_i;\theta) \equiv \sum_{k=1}^{K}\Big(g_k(w_i;\theta)-E\left[g_k(W;\theta)\right]\Big),$$

where

$$g_k(w_i;\theta) \equiv \omega(x_i)\cdot(T_k(w_i;\theta))_+ + \omega(x_i)^2\cdot\lambda_{2k}(w_i;\theta)\cdot f_X(x_i).$$

In our construction, we proposed to use as estimators,

$$\widehat{\sigma}_R^2(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n}\widehat{\psi}_R(w_i;\theta)^2, \quad \text{and} \quad \widehat{\Sigma}_{MR}(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n}\widehat{\psi}_M(w_i)\widehat{\psi}_R(w_i;\theta).$$

Where

$$\widehat{\psi}_M(w_i) \equiv \widehat{H}_0^{-1}\cdot\frac{\partial\ell(\widehat{\theta}_1,w_i)}{\partial\theta_1}, \quad \text{and} \quad \widehat{\psi}_R(w_i;\theta) \equiv \sum_{k=1}^{K}\Big(\widehat{g}_k(w_i;\theta)-\widehat{E}[g_k(W;\theta)]\Big),$$

with

$$\widehat{g}_k(y_1,x_1;\theta) = \frac{1}{n}\sum_{j=1}^{n}\frac{\widetilde{v}_k(w_1,w_j;\theta,h_n)}{h_n^z}, \quad \text{and} \quad \widehat{E}[g_k(W;\theta)] = \binom{n}{2}^{-1}\sum_{i<j}\frac{\widetilde{v}_k(w_i,w_j;\theta,h_n)}{h_n^z},$$

where $\widetilde{v}_k(w_1, w_2;\theta, h)$ is as described in equation (14) (above). Notice that $\widehat{E}[g_k(W;\theta)] = \frac{1}{h_n^z}\cdot U_{k,n}(\theta)$, where $U_{k,n}$ is the U-process described in equation (15) whose asymptotic properties were studied in Section 2.4.1. Under the conditions of Theorem 3, the empirical process arguments used in Section 2.2 yield

$$\sup_{\substack{w\in\mathcal{W}\\ \theta\in\Theta}}\left|\frac{1}{n}\sum_{j=1}^{n}\Big(\widetilde{v}_k(w, w_j;\theta, h_n) - E\left[\widetilde{v}_k(w, W;\theta, h_n)\right]\Big)\right| = O_p(n^{-1/2}).$$

From here we obtain,

$$\widehat{g}_k(y, x;\theta) = E\left[\frac{1}{h_n^z}\widetilde{v}_k(w, W;\theta, h_n)\right] + \xi_n^g(w,\theta), \quad \text{where} \quad \sup_{\substack{w\in\mathcal{W}\\ \theta\in\Theta}}\left|\xi_n^g(w,\theta)\right| = O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right).$$

Under Restrictions I1, I2 and I5, we have (see equation (18)),

$$E\left[\frac{1}{h_n^z}\widetilde{v}_k(w, W;\theta, h_n)\right] = g_k(y, x;\theta) + \mathcal{B}_n(w;\theta), \quad \text{where} \quad \sup_{\substack{w\in\mathcal{W}\\ \theta\in\Theta}}|\mathcal{B}_n(w_1;\theta)| = O(h_n^M).$$

Since $n^{1/2}\cdot h_n^M \longrightarrow 0$ by Restriction I2, these results combined yield,

$$\widehat{g}_k(y, x;\theta) = g_k(y, x;\theta) + \zeta_n^g(w,\theta), \quad \text{where} \quad \sup_{\substack{w\in\mathcal{W}\\ \theta\in\Theta}}\left|\zeta_n^g(w,\theta)\right| = O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right).$$

Next, equation (19) states that,

$$\widehat{E}\left[g_k(Y, X;\theta)\right] = E\left[g_k(Y, X;\theta)\right] + \rho_n^g(\theta), \quad \text{where} \quad \sup_{\theta\in\Theta}\left|\rho_n^g(\theta)\right| = O_p(n^{-1/2}).$$

Finally, the last two results combined for $k = 1,\ldots,K$ yield the following asymptotic property for our estimated influence function $\widehat{\psi}_R$,

$$\sup_{\substack{w\in\mathcal{W}\\ \theta\in\Theta}}\left|\widehat{\psi}_R(y, x;\theta) - \psi_R(y, x;\theta)\right| = O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right). \tag{23}$$

Next, note by inspection that

$$\sup_{\substack{w\in\mathcal{W}\\ \theta\in\Theta}}\left|\psi_R(y, x;\theta)\right| \le 4\overline{K}\,\overline{\omega}^2\overline{f} \equiv \overline{\psi}_R. \tag{24}$$

where $\overline{f}$ and $\overline{K}$ are described in Restrictions I1 and I2 respectively and $\overline{\omega}$ is the upper bound for the trimming function $\omega(\cdot)$. We have

$$\widehat{\sigma}_R^2(\theta) - \sigma_R^2(\theta) = \frac{1}{n}\sum_{i=1}^{n}\Big(\psi_R(y_i, x_i;\theta)^2 - \sigma_R^2(\theta)\Big) + \frac{2}{n}\sum_{i=1}^{n}\psi_R(y_i, x_i;\theta)\cdot\Big(\widehat{\psi}_R(w_i;\theta) - \psi_R(y_i, x_i;\theta)\Big)$$
$$+\frac{1}{n}\sum_{i=1}^{n}\Big(\widehat{\psi}_R(w_i;\theta) - \psi_R(y_i, x_i;\theta)\Big)^2$$

Under the conditions of Theorem 3, the empirical process arguments used in Section 2.2 can be used to show that

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\Big(\psi_R(y_i, x_i;\theta)^2 - \sigma_R^2(\theta)\Big)\right| = O_p(n^{-1/2}).$$

From equations (23) and (24), we obtain

$$\sup_{\theta\in\Theta}\left|\frac{2}{n}\sum_{i=1}^{n}\psi_R(y_i, x_i;\theta)\cdot\Big(\widehat{\psi}_R(w_i;\theta) - \psi_R(y_i, x_i;\theta)\Big)\right| = O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right),$$
$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\Big(\widehat{\psi}_R(w_i;\theta) - \psi_R(y_i, x_i;\theta)\Big)^2\right| = O_p\left(\frac{1}{n\cdot h_n^{2z}}\right).$$

Together, these results yield,

$$\sup_{\theta\in\Theta}\left|\widehat{\sigma}_R^2(\theta) - \sigma_R^2(\theta)\right| = O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right). \tag{25}$$

Next, we have

$$\begin{aligned}
&\widehat{\Sigma}_{MR}(\theta) - \Sigma_{MR}(\theta) = \\
&\frac{1}{n}\sum_{i=1}^{n}\Big(\psi_M(w_i)\psi_R(y_i, x_i;\theta) - \Sigma_{MR}(\theta)\Big) + \frac{1}{n}\sum_{i=1}^{n}\psi_M(w_i)\Big(\widehat{\psi}_R(w_i;\theta) - \psi_R(y_i, x_i;\theta)\Big) \\
&+\frac{1}{n}\sum_{i=1}^{n}\Big(\widehat{\psi}_M(w_i) - \psi_M(w_i)\Big)\psi_R(y_i, x_i;\theta) + \frac{1}{n}\sum_{i=1}^{n}\Big(\widehat{\psi}_M(w_i) - \psi_M(w_i)\Big)\Big(\widehat{\psi}_R(w_i;\theta) - \psi_R(y_i, x_i;\theta)\Big).
\end{aligned} \tag{26}$$

Under the conditions of Theorem 3, the empirical process arguments used in Section 2.2 can be used to show that

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\Big(\psi_M(w_i)\psi_R(y_i, x_i;\theta) - \Sigma_{MR}(\theta)\Big)\right\| = O_p(n^{-1/2}).$$

Next, from (23) and the conditions of Theorem 3,

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\psi_M(w_i)\big(\widehat{\psi}_R(w_i;\theta)-\psi_R(y_i,x_i;\theta)\big)\right\| \leq \sup_{\substack{w\in\mathcal{W}\\ \theta\in\Theta}}\left|\widehat{\psi}_R(y,x;\theta)-\psi_R(y,x;\theta)\right|\times\left\|\frac{1}{n}\sum_{i=1}^{n}\psi_M(w_i)\right\|$$

$$= O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right)\times O_p(1) = O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right)$$

For the next term, note from (24) that

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\big(\widehat{\psi}_M(w_i)-\psi_M(w_i)\big)\psi_R(y_i,x_i;\theta)\right\| \leq \overline{\psi}_R\times\frac{1}{n}\sum_{i=1}^{n}\big\|\widehat{\psi}_M(w_i)-\psi_M(w_i)\big\|.$$

We analyze the term $\frac{1}{n}\sum_{i=1}^{n}\big\|\widehat{\psi}_M(w_i)-\psi_M(w_i)\big\|$ next. Note that

$$\widehat{\psi}_M(w_i)-\psi_M(w_i) = \big(\widehat{H}_0^{-1}+H_0^{-1}\big)\cdot\frac{\partial\ell(\theta_1^*,w_i)}{\partial\theta_1}+\widehat{H}_0^{-1}\cdot\left(\frac{\partial\ell(\widehat{\theta}_1,w_i)}{\partial\theta_1}-\frac{\partial\ell(\theta_1^*,w_i)}{\partial\theta_1}\right).$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\big\|\widehat{\psi}_M(w_i)-\psi_M(w_i)\big\| \leq \big\|\widehat{H}_0^{-1}-H_0^{-1}\big\|\times\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\partial\ell(\theta_1^*,w_i)}{\partial\theta_1}\right\|$$

$$+\big\|\widehat{H}_0^{-1}\big\|\times\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\partial\ell(\widehat{\theta}_1,w_i)}{\partial\theta_1}-\frac{\partial\ell(\theta_1^*,w_i)}{\partial\theta_1}\right\|$$

From equation (22), Theorem 3 and the conditions in Restriction I6,

$$\big\|\widehat{H}_0^{-1}-H_0^{-1}\big\|\times\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\partial\ell(\theta_1^*,w_i)}{\partial\theta_1}\right\| = O_p(n^{-1/2})\times O_p(1) = O_p(n^{-1/2}).$$

Next we analyze the term $\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\partial\ell(\widehat{\theta}_1,w_i)}{\partial\theta_1}-\frac{\partial\ell(\theta_1^*,w_i)}{\partial\theta_1}\right\|$. Note that

$$\frac{\partial\ell(\theta_1,w_i)}{\partial\theta_1} = \begin{pmatrix}\big(z_{1i}\cdot\phi_1(z_{1i}\widetilde{\beta},z_{2i}\widetilde{\beta};\lambda)+z_{2i}\cdot\phi_2(z_{1i}\widetilde{\beta},z_{2i}\widetilde{\beta};\lambda)\big)\cdot\big(1[y_i=(0,0)]-F(z_{1i}\widetilde{\beta},z_{2i}\widetilde{\beta};\lambda)\big)\\ \phi_3(z_{1i}\widetilde{\beta},z_{2i}\widetilde{\beta};\lambda)\cdot\big(1[y_i=(0,0)]-F(z_{1i}\widetilde{\beta},z_{2i}\widetilde{\beta};\lambda)\big)\end{pmatrix}$$

A first-order approximation yields,

$$\left(\frac{\partial\ell(\widehat{\theta}_1,w_i)}{\partial\theta_1}-\frac{\partial\ell(\theta_1^*,w_i)}{\partial\theta_1}\right)=$$

$$\begin{pmatrix}\Big(\sum_{\ell=1}^{2}\sum_{m=1}^{2}z_{\ell i}z_{mi}'\delta_{\ell,m}(y_i,z_{1i}\overline{\overline{\beta}}_i,z_{2i}\overline{\overline{\beta}}_i;\overline{\lambda}_i)\Big)\big(\widehat{\widetilde{\beta}}-\widetilde{\beta}^*\big)+\Big(\sum_{\ell=1}^{2}z_{\ell i}\eta_\ell(y_i,z_{1i}\overline{\overline{\beta}}_i,z_{2i}\overline{\overline{\beta}}_i;\overline{\lambda}_i)\Big)\big(\widehat{\lambda}-\lambda^*\big)\\ \Big(\sum_{\ell=1}^{2}\eta_\ell(y_i,z_{1i}\overline{\overline{\beta}}_i,z_{2i}\overline{\overline{\beta}}_i;\overline{\lambda}_i)z_{\ell i}'\Big)\big(\widehat{\widetilde{\beta}}-\widetilde{\beta}^*\big)+\Upsilon(y_i,z_{1i}\overline{\overline{\beta}}_i,z_{2i}\overline{\overline{\beta}}_i;\overline{\lambda}_i)\big(\widehat{\lambda}-\lambda^*\big)\end{pmatrix}$$

where $\overline{\theta}_{1i} \equiv \left(\overline{\overline{\beta}}_i', \overline{\lambda}_i\right)'$ belongs in the line segment connecting $\widehat{\theta}_1$ and θ_1^*. From Theorem 2, wp$\rightarrow$ 1, $\widehat{\theta}_1$ belongs in the neighborhood $\mathcal{N}$ which contains θ_1^* and whose properties are described in Restriction I6. Therefore $\overline{\theta}_{1i}$ also belongs in $\mathcal{N}$ for all i. Therefore, the conditions in Restriction I6 hold wp$\rightarrow$ 1 and we have,

$$\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{2}\sum_{m=1}^{2}\|z_{\ell i}\|\cdot\|z_{mi}\|\cdot\left|\delta_{\ell,m}(y_i, z_{1i}\overline{\overline{\beta}}_i, z_{2i}\overline{\overline{\beta}}_i;\overline{\lambda}_i)\right|\right)\cdot\left\|\widehat{\widetilde{\beta}} - \widetilde{\beta}^*\right\|$$

$$\leq \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{2}\sum_{m=1}^{2}\|z_{\ell i}\|\cdot\|z_{mi}\|\cdot D(w_i)\right)}_{O_p(1) \text{ by Restriction I6}}\cdot\underbrace{\left\|\widehat{\widetilde{\beta}} - \widetilde{\beta}^*\right\|}_{O_p(n^{-1/2})} = O_p(n^{-1/2}),$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{2}\|z_{\ell i}\|\cdot\left|\eta_{\ell}(y_i, z_{1i}\overline{\overline{\beta}}_i, z_{2i}\overline{\overline{\beta}}_i;\overline{\lambda}_i)\right|\right)\cdot\left|\widehat{\lambda} - \lambda^*\right| \leq \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{2}\|z_{\ell i}\|\cdot D(w_i)\right)}_{O_p(1) \text{ by Restriction I6}}\cdot\underbrace{\left|\widehat{\lambda} - \lambda^*\right|}_{O_p(n^{-1/2})} = O_p(n^{-1/2}),$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{2}\|z_{\ell i}\|\cdot\left|\eta_{\ell}(y_i, z_{1i}\overline{\overline{\beta}}_i, z_{2i}\overline{\overline{\beta}}_i;\overline{\lambda}_i)\right|\right)\cdot\left\|\widehat{\widetilde{\beta}} - \widetilde{\beta}^*\right\| \leq \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{2}\|z_{\ell i}\|\cdot D(w_i)\right)}_{O_p(1) \text{ by Restriction I6}}\cdot\underbrace{\left\|\widehat{\widetilde{\beta}} - \widetilde{\beta}^*\right\|}_{O_p(n^{-1/2})} = O_p(n^{-1/2}),$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{2}\left|\Upsilon(y_i, z_{1i}\overline{\overline{\beta}}_i, z_{2i}\overline{\overline{\beta}}_i;\overline{\lambda}_i)\right|\right)\cdot\left|\widehat{\lambda} - \lambda^*\right| \leq \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{2}D(w_i)\right)}_{O_p(1) \text{ by Restriction I6}}\cdot\underbrace{\left|\widehat{\lambda} - \lambda^*\right|}_{O_p(n^{-1/2})} = O_p(n^{-1/2}).$$

Combined with the fact that $\left\|\widehat{H}_0^{-1}\right\| = O(1)$, these results yield,

$$\left\|\widehat{H}_0^{-1}\right\| \times \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\partial\ell(\widehat{\theta}_1, w_i)}{\partial\theta_1} - \frac{\partial\ell(\theta_1^*, w_i)}{\partial\theta_1}\right\| = O_p(n^{-1/2}).$$

And therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{\psi}_M(w_i) - \psi_M(w_i)\right\| \leq$$

$$\underbrace{\left\|\widehat{H}_0^{-1} - H_0^{-1}\right\| \times \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\partial\ell(\theta_1^*, w_i)}{\partial\theta_1}\right\|}_{=O_p(n^{-1/2})} + \underbrace{\left\|\widehat{H}_0^{-1}\right\| \times \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\partial\ell(\widehat{\theta}_1, w_i)}{\partial\theta_1} - \frac{\partial\ell(\theta_1^*, w_i)}{\partial\theta_1}\right\|}_{=O_p(n^{-1/2})} = O_p(n^{-1/2}).$$

And,

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\Big(\widehat{\psi}_M(w_i)-\psi_M(w_i)\Big)\psi_R(y_i,x_i;\theta)\right\| \leq \overline{\psi}_R\times\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{\psi}_M(w_i)-\psi_M(w_i)\right\| = O_p(n^{-1/2})$$

Finally, combining our previous results, we have

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\Big(\widehat{\psi}_M(w_i)-\psi_M(w_i)\Big)\Big(\widehat{\psi}_R(w_i;\theta)-\psi_R(y_i,x_i;\theta)\Big)\right\| \\
&\leq \sup_{\substack{w\in\mathcal{W}\\ \theta\in\Theta}}\left|\widehat{\psi}_R(y,x;\theta)-\psi_R(y,x;\theta)\right|\times\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{\psi}_M(w_i)-\psi_M(w_i)\right\| = O_p\left(\frac{1}{n^{-1/2}\cdot h_n^z}\right)\times O_p\left(\frac{1}{n^{1/2}}\right) \\
&= O_p\left(\frac{1}{n\cdot h_n^z}\right).
\end{aligned}$$

Plugging these results back into (26), we obtain

$$\begin{aligned}
\sup_{\theta\in\Theta}\left\|\widehat{\Sigma}_{MR}(\theta)-\Sigma_{MR}(\theta)\right\| &\leq O_p\left(\frac{1}{n^{1/2}}\right)+O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right)+O_p\left(\frac{1}{n^{1/2}}\right)+O_p\left(\frac{1}{n\cdot h_n^z}\right) \\
&= O_p\left(\frac{1}{n^{1/2}\cdot h_n^z}\right)
\end{aligned} \tag{27}$$

Together, equations (22), (25) and (27) prove the result in (20). ■

3.2 Some intermediate steps leading to Theorems 4 and 5

Next we include for clarity some intermediate steps that lead to the asymptotic representation of our statistic $\widehat{Q}(\theta)$ described in the paper.

3.2.1 A useful decomposition of the variance-covariance matrix $\Sigma(\theta)$ over $\Theta\setminus\Theta_0^I$

As stated in the paper, let M be an invertible matrix satisfying $H_0^{-1}=MM'$ (and therefore $H_0=(M')^{-1}M^{-1}$). Let $\sigma_{MR}(\theta)$ be as defined in Restriction I7 and for all $\theta\in\Theta\setminus\Theta_0^I$ define

$$\begin{aligned}
&C(\theta)\equiv\begin{pmatrix} M & 0\\ \Sigma_{MR}(\theta)'(M')^{-1} & \sigma_{MR}(\theta)\end{pmatrix} \quad\Longrightarrow\quad C(\theta)^{-1}=\begin{pmatrix} M^{-1} & 0\\ -\frac{\Sigma_{MR}(\theta)'H_0}{\sigma_{MR}(\theta)} & \frac{1}{\sigma_{MR}(\theta)}\end{pmatrix}, \\
&\text{and note that}\quad \Sigma(\theta)=C(\theta)C(\theta)' \quad\text{and}\quad \Sigma(\theta)^{-1}=\Big(C(\theta)^{-1}\Big)'C(\theta)^{-1}.
\end{aligned} \tag{28}$$

As defined in the paper, let

$$\overline{\psi}(W;\theta) \equiv C(\theta)^{-1} \cdot \begin{pmatrix} \psi(W) \\ \psi_R(W;\theta) \end{pmatrix} = \begin{pmatrix} M^{-1}\psi_M(W) \\ \frac{-\Sigma_{MR}(\theta)' H_0 \psi_M(W) + \psi_R(W;\theta)}{\sigma_R^2(\theta) - \Sigma_{MR}(\theta)' H_0 \Sigma_{MR}(\theta)} \end{pmatrix} \equiv \begin{pmatrix} \overline{\psi}_M(W) \\ \frac{\psi_{MR}(W;\theta)}{\sigma_{MR}(\theta)} \end{pmatrix} \equiv \begin{pmatrix} \overline{\psi}_M(W) \\ \overline{\psi}_{MR}(W;\theta) \end{pmatrix}$$
$$T_n(\theta) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \overline{\psi}(w_i;\theta) \tag{29}$$

Note that

$$E\left[\overline{\psi}(W;\theta)\right] = 0, \quad E\left[\overline{\psi}(W;\theta)\overline{\psi}(W;\theta)'\right] = I_{r+1}, \quad Var\left(T_n(\theta)\right) = I_{r+1},$$
$$T_n(\theta)' T_n(\theta) = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_M(w_i) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_R(w_i;\theta) \end{pmatrix}' \Sigma^{-1}(\theta) \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_M(w_i) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_R(w_i;\theta) \end{pmatrix}.$$

As stated in the paper, under the results in Theorems 2 and 3, and Restriction I7, we then have,

$$\widehat{Q}(\theta) = n\mu(\theta)'\left(\Sigma(\theta)^{-1} + \vartheta_n(\theta)\right)\mu(\theta) + T_n(\theta)'T_n(\theta) + \xi_n^Q(\theta) \ \forall\ \theta \in \Theta \setminus \Theta_0^I,$$
$$\Longrightarrow \quad \widehat{Q}(\theta) = T_n(\theta)'T_n(\theta) + \xi_n^Q(\theta) \ \forall\ \theta \in \overline{\Theta}^*,$$

where $\sup\limits_{\theta\in\Theta\setminus\Theta_0^I} \left|\xi_n^Q(\theta)\right| = o_p(1)$ and $\sup\limits_{\theta\in\Theta\setminus\Theta_0^I} |\vartheta_n(\theta)| = o_p(1)$ (recall that $\overline{\Theta}^* \subseteq \Theta \setminus \Theta_0^I$ with $\mu(\theta) = 0$ for all $\theta \in \overline{\Theta}^*$).

3.2.2 Using the decomposition of the variance-covariance matrix $\Sigma(\theta)$ under Restriction I7'

As in the paper, define

$$\sigma_{MR,\kappa}^2(\theta) \equiv \max\{\sigma_R^2(\theta), \kappa\} - \Sigma_{MR}(\theta)' H_0 \Sigma_{MR}(\theta) = \sigma_{MR}^2(\theta) + \left(\max\{\sigma_R^2(\theta), \kappa\} - \sigma_R^2(\theta)\right),$$

and

$$\Sigma_\kappa(\theta) \equiv \begin{pmatrix} H_0^{-1} & \Sigma_{MR}(\theta) \\ \Sigma_{MR}(\theta)' & \max\{\sigma_R^2(\theta)\,,\,\kappa\} \end{pmatrix} \quad \Longrightarrow \quad \Sigma_\kappa(\theta)^{-1} = \begin{pmatrix} H_0 + \frac{H_0 \Sigma_{MR}(\theta)\Sigma_{MR}(\theta)' H_0}{\sigma_{MR,\kappa}^2(\theta)} & -\frac{H_0 \Sigma_{MR}(\theta)}{\sigma_{MR,\kappa}^2(\theta)} \\ -\frac{\Sigma_{MR}(\theta)' H_0}{\sigma_{MR,\kappa}^2(\theta)} & \frac{1}{\sigma_{MR,\kappa}^2(\theta)} \end{pmatrix}$$

Note from Restriction I7' that $\sigma_{MR}^2(\theta) \geq \sigma_R^2(\theta) \cdot (1 - \overline{d}) \ \forall\ \theta \in \Theta \setminus \Theta_0^I$. Take any $\theta \in \Theta \setminus \Theta_0^I$. If $\sigma_R^2(\theta) \geq \kappa$, we have $\sigma_{MR,\kappa}^2(\theta) = \sigma_{MR}^2(\theta) \geq \sigma_R^2(\theta) \cdot (1 - \overline{d}) \geq \kappa \cdot (1 - \overline{d})$. And if $\sigma_R^2(\theta) < \kappa$, we have $\sigma_{MR,\kappa}^2(\theta) = \sigma_{MR}^2(\theta) + \kappa - \sigma_R^2(\theta) \geq \sigma_R^2(\theta) \cdot (1 - \overline{d}) + \kappa - \sigma_R^2(\theta) = \kappa - \overline{d} \cdot \sigma_R^2(\theta) > \kappa - \overline{d} \cdot \kappa = \kappa \cdot (1 - \overline{d})$. Thus, Restriction I7' implies

$$\sigma_{MR,\kappa}^2(\theta) \geq \kappa \cdot (1 - \overline{d}) \ \forall\ \theta \in \Theta \setminus \Theta_0^I, \tag{30}$$

Next note from the result in (20) that

$$\sup_{\theta\in\Theta}\left|\max\left\{\widehat{\sigma}_R^2(\theta),\,\kappa\right\}-\max\left\{\sigma_R^2(\theta),\,\kappa\right\}\right|\le\sup_{\theta\in\Theta}\left|\widehat{\sigma}_R^2(\theta)-\sigma_R^2(\theta)\right|=o_p(1).$$

And from here, (20) once again yields

$$\sup_{\theta\in\Theta}\left|\left(\max\left\{\widehat{\sigma}_R^2(\theta),\,\kappa\right\}-\widehat{\Sigma}_{MR}(\theta)'\widehat{H}_0\widehat{\Sigma}_{MR}(\theta)\right)-\sigma_{MR,\kappa}^2(\theta)\right|=o_p(1).$$

From here, the bound in (30) yields

$$\sup_{\theta\in\Theta}\left|\left(\frac{\max\left\{\widehat{\sigma}_R^2(\theta),\,\kappa\right\}-\widehat{\Sigma}_{MR}(\theta)'\widehat{H}_0\widehat{\Sigma}_{MR}(\theta)}{\sigma_{MR,\kappa}^2(\theta)}\right)-1\right|=o_p(1),$$

and therefore, $\sup_{\theta\in\Theta\setminus\Theta_0^I}\left\|\widehat{\Sigma}(\theta)^{-1}-\Sigma_\kappa(\theta)^{-1}\right\|=o_p(1)$. Next let $C(\theta)$ be the matrix described in the factorization of $\Sigma(\theta)$ given in (28) and, as we defined in (29), let $\overline{\psi}(W;\theta)\equiv\begin{pmatrix}\overline{\psi}_M(W)\\ \overline{\psi}_{MR}(W;\theta)\end{pmatrix}$. We have

$$\begin{aligned}
&\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_M(w_i)\\ \frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_R(w_i;\theta)\end{pmatrix}'\Sigma_\kappa^{-1}(\theta)\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_M(w_i)\\ \frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_R(w_i;\theta)\end{pmatrix}\\
&=\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_M(w_i)\\ \frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_R(w_i;\theta)\end{pmatrix}'\left(C(\theta)^{-1}\right)'C(\theta)'\Sigma_\kappa^{-1}(\theta)C(\theta)C(\theta)^{-1}\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_M(w_i)\\ \frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_R(w_i;\theta)\end{pmatrix}\\
&=\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\overline{\psi}(w_i;\theta)\right)'\begin{pmatrix}I & 0\\ 0 & \left(\frac{\sigma_{MR}^2(\theta)}{\sigma_{MR,\kappa}^2(\theta)}\right)\end{pmatrix}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\overline{\psi}(w_i;\theta)\right)
\end{aligned}$$

As stated in the paper, under the results in Theorems 2 and 3, and Restriction I7', we then have,

$$\widehat{Q}(\theta)=n\mu(\theta)'\left(\Sigma_\kappa(\theta)^{-1}+\vartheta_n^\kappa(\theta)\right)\mu(\theta)+\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\overline{\psi}(w_i;\theta)\right)'\begin{pmatrix}I & 0\\ 0 & \left(\frac{\sigma_{MR}^2(\theta)}{\sigma_{MR,\kappa}^2(\theta)}\right)\end{pmatrix}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\overline{\psi}(w_i;\theta)\right)+\xi_n^{Q_\kappa}(\theta)$$

$\forall\ \theta\in\Theta\setminus\Theta_0^I$, and

$$\widehat{Q}(\theta)=\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\overline{\psi}(w_i;\theta)\right)'\begin{pmatrix}I & 0\\ 0 & \left(\frac{\sigma_{MR}^2(\theta)}{\sigma_{MR,\kappa}^2(\theta)}\right)\end{pmatrix}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\overline{\psi}(w_i;\theta)\right)+\xi_n^{Q_\kappa}(\theta)$$

$\forall\ \theta \in \Theta \setminus \overline{\Theta}^{*}$, where $\sup_{\theta\in\Theta\setminus\Theta_0^I} \left|\xi_n^{Q_\kappa}(\theta)\right| = o_p(1)$ and $\sup_{\theta\in\Theta\setminus\Theta_0^I} |\vartheta_n^{\kappa}(\theta)| = o_p(1)$. Alternatively, we can perform a decomposition directly on $\Sigma_\kappa(\theta)^{-1}$. Let

$$D_\kappa(\theta) \equiv \begin{pmatrix} M^{-1} & 0 \\ -\frac{\Sigma_{MR}(\theta)' H_0}{\sigma_{MR,\kappa}(\theta)} & \frac{1}{\sigma_{MR,\kappa}(\theta)} \end{pmatrix}$$

and note that $\Sigma_\kappa(\theta)^{-1} = D_\kappa(\theta)' D_\kappa(\theta)$ (note that this decomposition immediately shows that $\Sigma_\kappa(\theta)^{-1}$ is positive-definite). We have

$$\begin{aligned}
&\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_M(w_i) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_R(w_i;\theta) \end{pmatrix}' \Sigma_\kappa^{-1}(\theta) \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_M(w_i) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_R(w_i;\theta) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_M(w_i) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_R(w_i;\theta) \end{pmatrix}' D_\kappa(\theta)' D_\kappa(\theta) \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_M(w_i) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_R(w_i;\theta) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n M^{-1}\psi_M(w_i) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{-\Sigma_{MR}(\theta)' H_0 \psi_M(w_i) + \psi_R(w_i;\theta)}{\sigma_{MR,\kappa}(\theta)}\right) \end{pmatrix}' \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n M^{-1}\psi_M(w_i) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{-\Sigma_{MR}(\theta)' H_0 \psi_M(w_i) + \psi_R(w_i;\theta)}{\sigma_{MR,\kappa}(\theta)}\right) \end{pmatrix} \\
&= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \overline{\psi}(w_i;\theta)\right)' \begin{pmatrix} I & 0 \\ 0 & \left(\frac{\sigma_{MR}^2(\theta)}{\sigma_{MR,\kappa}^2(\theta)}\right) \end{pmatrix} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \overline{\psi}(w_i;\theta)\right)
\end{aligned}$$

this is the same result we showed above. ■

References

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.

Serfling, R. J. (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.