NONPARAMETRIC MODEL CHECKS
OF SINGLE-INDEX ASSUMPTIONS

Samuel Maistre[1,2,3] and Valentin Patilea[1]

[1]*CREST (Ensai)* [2]*Université de Lyon* [3]*Université de Strasbourg*

Supplementary Material

This supplementary material contains additional proofs, details and technical lemmas.

S1 Additional proofs and details

S1.1 Proof of Lemma A.1

Lemma A.1. *Let* (U_1, Z_1, W_1) *and* (U_2, Z_2, W_2) *be two independent draws of* (U, Z, W)*. Let* $K(\cdot)$ *be a bounded, even, integrable function with positive, integrable Fourier transform. Assume* $\mathbb{E}(\|Uw(Z)\|_{\mathcal{H}}^2) < \infty$*, Then for any* $h > 0$*,*

$$\mathbb{E}\left[U \mid Z, W\right] = 0 \ \textit{a.s.} \Leftrightarrow I(h) = 0.$$

Moreover, if $\mathbb{P}\left(\mathbb{E}\left[U \mid Z, W\right] = 0\right) < 1$, *then*

$$\inf_{h\in(0,1]} I(h) > 0.$$

Proof of Lemma A.1. Using the Inverse Fourier Transform and the independence of $(U_1,\, Z_1,\, W_1)$ and $(U_2,\, Z_2,\, W_2)$,

$$\begin{aligned}
&I(h) = \mathbb{E}\left[\langle U_1, U_2\rangle_{\mathcal{H}}\; w(Z_1)w(Z_2)h^{-q}K((Z_1-Z_2)/h)\phi(W_1-W_2)\right]\\
&= \mathbb{E}\left[\langle U_1, U_2\rangle_{\mathcal{H}}\; w(Z_1)w(Z_2)\int_{\mathbb{R}^q} e^{2\pi i v'(Z_1-Z_2)}\mathcal{F}[K](vh)dv \int_{\mathbb{R}^r} e^{2\pi i s'(W_1-W_2)}\mathcal{F}[\phi](s)ds\right]\\
&= \int_{\mathbb{R}^q}\int_{\mathbb{R}^r}\left\|\mathbb{E}\left[\mathbb{E}[U \mid Z, W]w(Z)e^{-2\pi i\{v'Z+s'W\}}\right]\right\|_{\mathcal{H}}^2 \mathcal{F}[K](vh)\mathcal{F}[\phi](s)dvds.
\end{aligned}$$

Clearly, for any $h > 0$, $I(h) = 0$ whenever $\mathbb{E}[U \mid Z, W] = 0$ a.s. For the reverse implication, since $\mathcal{F}[\phi], \mathcal{F}[K] > 0$, and $w(\cdot) > 0$, for any $h > 0$, one can deduce

$$\mathbb{E}\left[\mathbb{E}[U \mid Z, W]w(Z)e^{-2\pi i\{v'Z+s'W\}}\right] = 0, \qquad \forall v \in \mathbb{R}^q,\ \forall s \in \mathbb{R}^r.$$

Then necessarily $\mathbb{E}[U \mid Z, W]w(Z) = 0$ a.s., and thus $\mathbb{E}[U \mid Z, W] = 0$ a.s.

In the case $\mathbb{P}\left(\mathbb{E}\left[U \mid W, X\right] = 0\right) < 1$, by the Lebesgue Dominated Convergence Theorem, the map $h \mapsto I(h)$ is continuous on $(0,1]$ and

$$\lim_{h\to 0} I(h) = \mathcal{F}\left[K\right](0)\int_{\mathbb{R}^q}\int_{\mathbb{R}^r}\left\|\mathbb{E}\left[\mathbb{E}[U \mid Z, W]w(Z)e^{-2\pi i\{v'Z+s'W\}}\right]\right\|_{\mathcal{H}}^2 \mathcal{F}\left[\phi\right](s)\, dvds.$$

Since the nonnegative valued map

$$(v,s) \mapsto \left\|\mathbb{E}\left[\mathbb{E}[U \mid Z, W]w(Z)e^{-2\pi i\{v'Z+s'W\}}\right]\right\|_{\mathcal{H}}^2$$

is continuous, and non identically equal to 0 whenever $\mathbb{E}\left[U \mid Z, W\right] \neq 0$, and $\mathcal{F}\left[K\right](0), \mathcal{F}\left[\psi\right](\cdot) > 0$, $\lim_{h\to 0} I(h)$ is necessarily positive and $I(h)$ is bounded away from zero on the interval $(0,1]$. □

S1.2 Some details on equation (4.4)

Consider a sequence of alternatives

$$Y = m(Z(\beta_0)) + r_n\delta(Z(\beta_0), W(\beta_0)) + \varepsilon, \quad n \geq 1,$$

with $\mathbb{E}(\varepsilon \mid X) = 0$ a.s., and $\delta(\cdot)$ satisfying the conditions (4.3). In the following, we show that for each n, β_0 is solution of the equation

$$\frac{\partial}{\partial\beta}\mathbb{E}\left[\{Y - r_\beta(Z(\beta))\}^2\right] = 0.$$

Under suitable conditions on the second order derivative with respect to β, this justifies the equation (4.4). For this purpose, we want to differentiate $r_\beta(Z(\beta))$. Let

$$Y^0 = m(Z(\beta_0)) + \varepsilon \quad \text{and define} \quad r^0_\beta(Z(\beta)) = \mathbb{E}\left[Y^0 \mid Z(\beta)\right] = \mathbb{E}\left[m(Z(\beta_0)) \mid Z(\beta)\right].$$

Moreover, let

$$\delta_\beta(Z(\beta)) = \mathbb{E}\left[\delta(Z(\beta_0), W(\beta_0)) \mid Z(\beta)\right],$$

and notice

$$r_\beta(Z(\beta)) = r^0_\beta(Z(\beta)) + r_n \delta_\beta(Z(\beta)),$$

and that, by the first condition in equation (4.3), $\delta_{\beta_0}(Z(\beta_0)) = 0$. Next, by standard results from single-index regression models applied to the response Y^0 (see, for instance, Horowitz (2009) chapter 2), one has

$$\left.\frac{\partial}{\partial \beta} r^0_\beta(Z(\beta))\right|_{\beta=\beta_0} = m'(Z(\beta_0))\{X - \mathbb{E}[X \mid Z(\beta_0)]\}.$$

On the other hand, by the standard variance decomposition formula,

$$\begin{aligned} \mathbb{E}\left[\{\delta(Z(\beta_0), W(\beta_0)) - \delta_\beta(Z(\beta))\}^2\right] &\leq \mathbb{E}[\delta^2(Z(\beta_0), W(\beta_0))] \\ &= \mathbb{E}\left[\{\delta(Z(\beta_0), W(\beta_0)) - \delta_{\beta_0}(Z(\beta_0))\}^2\right], \end{aligned}$$

and thus one can deduce that

$$\mathbb{E}\left[\delta(Z(\beta_0),W(\beta_0))\left.\frac{\partial}{\partial\beta}\delta_\beta(Z(\beta))\right|_{\beta=\beta_0}\right]=0. \tag{S1.1}$$

Using this identity and the second condition in equation (4.3), under suitable technical conditions, for any n, one can deduce

$$\begin{aligned}
\left.\frac{\partial}{\partial\beta}\mathbb{E}\left[\{Y-r_\beta(Z(\beta))\}^2\right]\right|_{\beta=\beta_0} &= -2\mathbb{E}\left[\{Y-r_{\beta_0}(Z(\beta_0))\}\left.\frac{\partial}{\partial\beta}r_\beta(Z(\beta))\right|_{\beta=\beta_0}\right]\\
= -2\mathbb{E}\left[\{r_n\delta(Z(\beta_0),W(\beta_0))+\varepsilon\}\left\{\left.\frac{\partial}{\partial\beta}r^0_\beta(Z(\beta))\right|_{\beta=\beta_0}\right.\right.&\left.\left.+\left.\frac{\partial}{\partial\beta}r_n\delta_\beta(Z(\beta))\right|_{\beta=\beta_0}\right\}\right]\\
&= -2r_n\mathbb{E}\left[\delta(Z(\beta_0),W(\beta_0))m'(Z(\beta_0))\{X-\mathbb{E}[X\mid Z(\beta_0)\}\right]\\
&\quad -2r_n^2\mathbb{E}\left[\delta(Z(\beta_0),W(\beta_0))\left.\frac{\partial}{\partial\beta}\delta_\beta(Z(\beta))\right|_{\beta=\beta_0}\right]\\
&= 0.
\end{aligned}$$

Let us end this part with a comment on the condition (4.3). In general, in the case of sequence of alternatives, by standard tools for deriving asymptotic results, one can prove $\hat{\beta}-\beta^*=O_{\mathbb{P}}(n^{-1/2})$ for some $\beta^*\in\mathcal{B}$ that could depend on n. The value β^* tends to β_0 at a rate depending on the rate the alternative hypotheses approaches the null hypothesis. Following a common choice in the literature, herein we want to simplify the presentation and focus on the case where β^* does not depend on n. For this purpose, we impose some orthogonality conditions on the function $\delta(\cdot)$ such that $\beta^* = \beta_0$ when the index is estimated by semiparametric least-squares. See, for instance, the second condition of equation (3.11) in Guerre and Lavergne (2005) for a similar condition.

S1.3 Some details on the proof of Proposition 1

Herein, we provide detailed justifications of the claim that the norm of any column of $\mathbf{A}\left(\beta\right) - \mathbf{A}\left(\bar{\beta}\right)$ is bounded by $c\|\beta - \bar{\beta}\|$.

Firstly, recall that

$$\mathcal{B} \subset \{1\} \times \mathbb{R}^{p-1} \quad \text{or} \quad \mathcal{B} \subset \{\|\gamma\|^{-1}\gamma : \gamma = (\gamma_1, \ldots, \gamma_p)^\top \in \mathbb{R}^p, \gamma_1 > 0\}.$$

Then the norm of every β from the parameter space $\mathcal{B}$ is larger or equal to 1. Moreover, if $\overline{\beta} \in \mathcal{B}$ and $B(\overline{\beta}, r) \subset \mathbb{R}^d$ is the ball centered at $\overline{\beta}$ of radius $0 < r < 1/2$, then

$$\inf_{\beta \in B(\overline{\beta}, r) \cap \mathcal{B}} \frac{\langle \beta, \overline{\beta} \rangle}{\|\beta\| \|\overline{\beta}\|} \geq \frac{1}{\|\overline{\beta}\|} > 0. \tag{S1.2}$$

Indeed, for any $\beta \in B(\overline{\beta}, r)$,

$$\big|\|\beta\| - \|\overline{\beta}\|\big| \leq r \qquad \text{and thus} \qquad \|\beta\| \geq \|\overline{\beta}\| - r \geq 1 - r.$$

Then, since for any $\beta \in B(\overline{\beta}, r)$, $\|\beta - \overline{\beta}\|^2 \leq r^2$, we have

$$\langle \beta, \overline{\beta} \rangle = \frac{1}{2} \left\{ \|\overline{\beta}\|^2 + \|\beta\|^2 - \|\beta - \overline{\beta}\|^2 \right\} \geq \frac{\|\overline{\beta}\|^2}{2}.$$

Finally, divide by $\|\beta\| \|\overline{\beta}\|$ to derive the inequality (S1.2).

Now, consider $\{v_1, \ldots, v_{p-1}\}$, a basis in the orthogonal subspace $\{\overline{\beta}\}^\perp$. Then for any $\beta \in B(\overline{\beta}, r)$, the set of vectors $\{v_1, \ldots, v_{p-1}\} \cup \{\beta\}$ is a basis in $\mathbb{R}^p$. Indeed, equation (S1.2) implies that none of $\beta \in B(\overline{\beta}, r)$ could be spanned by $\{v_1, \ldots, v_{p-1}\}$.

Finally, if $\mathbf{A}(\beta)$ is built by the Gram-Schmidt process applied to $\{v_1, \ldots, v_{p-1}\} \cup \{\beta\}$, as described at the beginning of the proof of Proposition 1, and the vectors $\{v_1, \ldots, v_{p-1}\}$ are orthogonal, then the norm of any column of $\mathbf{A}(\beta) - \mathbf{A}\left(\bar{\beta}\right)$ is bounded by $c\|\beta - \bar{\beta}\|$ for some c depending only on the initial $p-1$ independent vectors.

The fact that $\{v_1, \ldots, v_{p-1}\}$ are orthogonal it is not a real constraint since, starting from an arbitrary basis in $\{\overline{\beta}\}^\perp$, one could first apply an orthogonalization process with that basis.

We consider the case $p = 3$, the case of larger p could be derived similarly. Let $\beta \in B(\overline{\beta}, 1/2)$ and consider v, w two linearly independent vectors from the space $\{\overline{\beta}\}^\perp$. The Gram-Schmidt process transforms the basis $\{\beta, v, w\}$ as follows:

$$\begin{aligned} u_1 &= \beta, & e_1 &= e_1(\beta) = \frac{u_1}{\|u_1\|}, \\ u_2 &= v - \langle v, e_1\rangle e_1, & e_2 &= e_2(\beta) = \frac{u_2}{\|u_2\|}, \\ u_3 &= w - \langle w, e_1\rangle e_1 - \langle w, e_2\rangle e_2, & e_3 &= e_3(\beta) = \frac{u_3}{\|u_3\|}. \end{aligned}$$

Since the matrix $\mathbf{A}(\beta)$ is build with the columns $e_2(\beta)$ and $e_3(\beta)$, it remains to check that Lipschitz condition for $e_2(\beta)$ and $e_3(\beta)$ as functions of β. First, note that

$$\|e_1(\beta) - e_1(\overline{\beta})\| = \left\| \frac{\|\overline{\beta}\|\beta - \|\beta\|\overline{\beta}}{\|\overline{\beta}\|\|\beta\|} \right\| \leq \frac{1}{\|\overline{\beta}\|} \big|\|\overline{\beta}\| - \|\beta\|\big| + \frac{1}{\|\overline{\beta}\|}\|\overline{\beta} - \beta\| \leq c_1 \|\overline{\beta} - \beta\|,$$

with $c_1 = 2/\|\overline{\beta}\|$. In particular,

$$\big|\|e_1(\beta)\| - \|e_1(\overline{\beta})\|\big| \leq c_1 \|\overline{\beta} - \beta\|.$$

Next, since $v \in \{\overline{\beta}\}^\perp = \{e_1(\overline{\beta})\}^\perp$,

$$\langle v, e_1(\beta)\rangle = \langle v, e_1(\overline{\beta})\rangle + \langle v, e_1(\beta) - e_1(\overline{\beta})\rangle = \langle v, e_1(\beta) - e_1(\overline{\beta})\rangle,$$

and thus

$$|\langle v, e_1(\beta)\rangle| \leq \|v\|\|e_1(\beta) - e_1(\overline{\beta})\| \leq c_1\|v\|\|\beta - \overline{\beta}\| \leq c_1 r\|v\|.$$

Moreover, $\forall \beta \in B(\overline{\beta}, r)$

$$\|u_2(\beta)\|^2 = \|v\|^2 - \langle v, e_1(\beta)\rangle^2 \geq (1 - c_1^2 r^2)\|v\|^2$$

and

$$\begin{aligned}
\big|\|u_2(\overline{\beta})\| - \|u_2(\beta)\|\big| &= \frac{\big|\|u_2(\overline{\beta})\|^2 - \|u_2(\beta)\|^2\big|}{\|u_2(\overline{\beta})\| + \|u_2(\beta)\|} \\
&\leq \frac{\big|\langle v, e_1(\overline{\beta})\rangle - \langle v, e_1(\beta)\rangle\big|\,\big|\langle v, e_1(\overline{\beta})\rangle + \langle v, e_1(\beta)\rangle\big|}{2(1 - c_1^2 r^2)^{1/2}\|v\|} \\
&\leq \frac{\|v\|\|e_1(\overline{\beta}) - e_1(\beta)\| \times 2\|v\|}{2(1 - c_1^2 r^2)^{1/2}\|v\|} \\
&\leq \frac{c_1\|v\|}{(1 - c_1^2 r^2)^{1/2}}\|\overline{\beta} - \beta\|.
\end{aligned}$$

Then

$$\begin{aligned}
\|e_2(\beta)-e_2(\overline{\beta})\| &= \left\| \frac{\|u_2(\overline{\beta})\|\{v-\langle v,e_1(\beta)\rangle e_1(\beta)\} - \|u_2(\beta)\|\{v-\langle v,e_1(\overline{\beta})\rangle e_1(\overline{\beta})\}}{\|u_2(\beta)\|\|u_2(\overline{\beta})\|} \right\| \\
&\leq \|v\| \frac{\left|\|u_2(\overline{\beta})\| - \|u_2(\beta)\|\right|}{\|u_2(\overline{\beta})\|\|u_2(\beta)\|} \\
&+ \frac{1}{\|u_2(\beta)\|} \left\|\langle v,e_1(\beta)\rangle e_1(\beta) - \langle v,e_1(\overline{\beta})\rangle e_1(\overline{\beta})\right\| \\
&+ \frac{\left\|\langle v,e_1(\overline{\beta})\rangle e_1(\overline{\beta})\right\|}{\|u_2(\overline{\beta})\|} \left|\|u_2(\overline{\beta})\| - \|u_2(\beta)\|\right| \\
&\leq c_2\|\overline{\beta}-\beta\|,
\end{aligned}$$

where c_2 is a positive constant that depends only on r, $\|\overline{\beta}\|$ and $\|v\|$. Smaller r is fixed, larger the constant c_2 could be taken. Finally, to get the Lipschitz condition for the map $\beta \mapsto e_3(\beta)$, we first need to bound from below $\|u_3(\beta)\|$. Using the orthogonality between $e_1(\beta)$ and $e_2(\beta)$, we get

$$\|u_3(\beta)\|^2 = \|w\|^2 - \langle w,e_1(\beta)\rangle^2 - \langle w,e_2(\beta)\rangle^2.$$

Since $w \in \{\overline{\beta}\}^\perp = \{e_1(\overline{\beta})\}^\perp$ and v and w are orthogonal,

$$|\langle w,e_1(\beta)\rangle| \leq \|w\|\|e_1(\beta)-e_1(\overline{\beta})\| \leq c_1\|w\|\|\beta-\overline{\beta}\|$$

and

$$|\langle w, e_2(\beta)\rangle| \leq \|w\|\|\langle v, e_1(\overline{\beta})\rangle e_1(\overline{\beta}) - \langle v, e_1(\beta)\rangle e_1(\beta)\| \leq c_2\|w\|\|\beta - \overline{\beta}\|.$$

Deduce that $\forall\beta \in B(\overline{\beta}, r)$,

$$\|u_3(\beta)\|^2 \geq (1 - c_1^2 r^2 - c_2^2 r^2)\|w\|^2.$$

The Lipschitz condition for the map $\beta \mapsto e_3(\beta)$ follows after repeatedly applying the triangle inequality.

S1.4 Proof of Proposition 2

Proposition 2. *Suppose the conditions in Assumption 1 in the Appendix are met and the null hypothesis (2.2) holds true. Consider* β_n *such that* $\beta_n - \beta_0 = O_{\mathbb{P}}(n^{-1/2})$. *Then* $nh^{1/2} I_n^{\{l\}}(\beta_n) / \hat{\omega}_n^{\{l\}}(\beta_n) \to \mathcal{N}(0,1)$ *in law under* H_0, *and*

$$\left[\hat{\omega}_n^{\{l\}}(\beta_0)\right]^2 \to \left[\omega^{\{l\}}(\beta_0)\right]^2 = 2\int K^2(u)\,du \times \int\int \Gamma^2(s,t)\,ds\,dt$$
$$\times \mathbb{E}\left[\int f_{\beta_0}^4(z)\,\phi^2(W_1(\beta_0) - W_2(\beta_0))\,\pi_{\beta_0}(z \mid W_1(\beta_0))\,\pi_{\beta_0}(z \mid W_2(\beta_0))\,dz\right],$$

in probability, where $\pi_{\beta_0}(\cdot \mid w)$ *is the conditional density of* $Z(\beta_0)$ *knowing that* $W(\beta_0) = w$, *and for* $t, s \in [0,1]$,

$$\Gamma(s,t) = \mathbb{E}\left[\epsilon(s)\,\epsilon(t)\right], \qquad \epsilon(t) = \mathbf{1}\{\Phi(Y) \leq t\} - \mathbb{P}[\Phi(Y) \leq t \mid X'\beta_0].$$

Proof of Proposition 2. Let us consider the simplified notation from equation (7.3) and further simplify in the case $\beta = \beta_0$ and write

$$L_{ij} = L_{ij}(\beta_0, g), \quad K_{ij} = K_{ij}(\beta_0, h), \quad \text{and} \quad \phi_{ij} = \phi(W_i(\beta_0) - W_j(\beta_0)). \tag{S1.3}$$

Notice that

$$\begin{aligned} I_n^{\{l\}}(\beta_0) &= \frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} \Big\{ \left\langle (r_i - \tilde{r}_i)(\cdot;\, \beta_0),\, (r_j - \tilde{r}_j)(\cdot;\, \beta_0) \right\rangle_{L^2} \\ &\qquad + \left\langle \epsilon_i(\cdot),\, \epsilon_j(\cdot) \right\rangle_{L^2} \\ &\qquad + \left\langle \tilde{\epsilon}_i(\cdot),\, \tilde{\epsilon}_j(\cdot) \right\rangle_{L^2} \\ &\qquad + 2\left\langle \epsilon_i(\cdot),\, (r_j - \tilde{r}_j)(\cdot;\, \beta_0) \right\rangle_{L^2} \\ &\qquad - 2\left\langle \tilde{\epsilon}_i(\cdot),\, (r_j - \tilde{r}_j)(\cdot;\, \beta_0) \right\rangle_{L^2} \\ &\qquad - 2\left\langle \epsilon_i(\cdot),\, \tilde{\epsilon}_j(\cdot) \right\rangle_{L^2} \Big\} \hat{f}_{\beta_0,i}\hat{f}_{\beta_0,j} K_{ij}\phi_{ij} \\ &= I_1(\beta_0) + I_2(\beta_0) + I_3(\beta_0) + 2I_4(\beta_0) - 2I_5(\beta_0) - 2I_6(\beta_0) \end{aligned}$$

with

$$\hat{f}_{\beta,i} = \frac{1}{(n-1)\,g} \sum_{k \neq i} L_{ik}(\beta), \quad r_i(t;\,\beta) = \mathbb{P}\left[Y_i \leq \Phi^{-1}(t) \mid X_i'\beta\right],$$

$$\tilde{r}_i(t;\,\beta) = \frac{1}{(n-1)\,g\hat{f}_{\beta,i}} \sum_{k \neq i} r_k(t;\,\beta)\, L_{ik}(\beta)$$

and $\tilde{\epsilon}_i(\cdot)$ is defined as $\tilde{r}_i(t;\,\beta)$ by replacing $r_i(t;\,\beta)$ by $\epsilon_i(\cdot)$. This decomposition of $I_n^{\{l\}}(\beta_0)$ is given by the identity

$$\widehat{U_i\omega(Z_i)}(\cdot;\,\beta_0) = \left[r_i(\cdot;\,\beta_0) - \tilde{r}_i(\cdot;\,\beta_0) + \epsilon_i(\cdot) - \tilde{\epsilon}_i(\cdot)\right] \hat{f}_{\beta_0,i}.$$

The terms $I_1(\beta_0)$ and $I_3(\beta_0)$ are treated in Lemmas 6 and 7 in Section S2. For $I_2(\beta_0)$, let us introduce

$$\omega_n^2(\beta) = \frac{2}{n(n-1)\,h} \sum_{i=1}^{n} \sum_{j \neq i} \int\!\!\int \Gamma^2(s,t)\, ds\, dt\, \hat{f}_{\beta,i}^2 \hat{f}_{\beta,j}^2 K_{ij}^2(\beta)\, \phi_{ij}^2(\beta).$$

Proposition 3 below ensures that $nh^{1/2}\omega_n^{-1}(\beta_0)\, I_2(\beta_0) \to \mathcal{N}(0,1)$ in law. The terms $I_4(\beta_0)$, $I_5(\beta_0)$ and $I_6(\beta_0)$ can be shown to be negligible in a similar way as $I_1(\beta_0)$ and $I_3(\beta_0)$. Lemma 9 shows that $\omega_n^2(\beta_0) \to \left[\omega^{\{l\}}(\beta_0)\right]^2$, in probability, with $\omega^{\{l\}}(\beta_0) > 0$ and thus $I_j(\beta_0)/\omega_n(\beta_0)$ is of the same order as $I_j(\beta_0)$ for $j \in \{1, 3, 4, 5, 6\}$. Finally, it is easy to check that $\omega_n(\beta_0) - \hat{\omega}_n^{\{l\}}(\beta_0) = o_{\mathbb{P}}(1)$. By Proposition 1, one can replace β_0 by β_n, an estimator of β_0 that converges in probability with the rate $O_{\mathbb{P}}(n^{-1/2})$, and have

$$nh^{1/2} I_n^{\{l\}}(\beta_n) / \hat{\omega}_n^{\{l\}}(\beta_n) - nh^{1/2} I_n^{\{l\}}(\beta_0) / \hat{\omega}_n^{\{l\}}(\beta_0) = o_{\mathbb{P}}(1).$$

Then the result of the Proposition 2 follows. □

Proposition 3. *Under the conditions of Proposition 2,*

$$nh^{1/2}\omega_n^{-1}(\beta_0) I_2(\beta_0) \to \mathcal{N}(0,1) \quad \textit{in law.}$$

Proof. $\{S_{n,m},\, \mathcal{F}_{n,m},\, 1 \le m \le n,\, n \ge 1\}$ is a martingale array with $S_{n,1} = 0$ and

$$S_{n,m}(\beta_0) = \sum_{i=1}^{m} G_{n,i}(\beta_0)$$

with

$$G_{n,i}(\beta_0) = \frac{2h^{p/2}}{\omega_n (n-1) h} \left\langle \epsilon_i(\cdot) \hat{f}_{\beta_0,i}, \sum_{j=1}^{i-1} \epsilon_j(\cdot) \hat{f}_{\beta_0,j} K_{ij} \phi_{ij} \right\rangle_{L^2}$$

and $\mathcal{F}_{n,m}$ is the σ-field generated by $\{X_1, \ldots, X_n, Y_1, \ldots, Y_m\}$. Thus

$$nh^{1/2}\omega_n^{-1}(\beta_0) I_2(\beta_0) = S_{n,n}(\beta_0).$$

From Lemma 8 and the nesting of the σ-fields $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}$ for $1 \le i \le n$, $n \ge 1$, we have that the martingale array satisfies Corollary 3.1 of Hall and Heyde (1980) and the result follows. $\square$

S2 Technical lemmas

In the following results the kernels L and K are supposed to satisfy the conditions of Assumption 1-(f).

Lemma 1. *Assume that* $\mathbb{E}[\exp(a\|X\|)] < \infty$ *for some* $a > 0$. *Consider that* $g \to 0$ *and* $ng^{4/3}/\log n \to \infty$. *For any* $t \in [0,1]$ *let* $Y_k(t)$, $1 \le k \le n$, *be an i.i.d. random variables like in the proof of Proposition 1 such that* $\mathbb{E}[\sup_t |Y_k(t)|^a] < \infty$ *for some* $a > 8$. *Moreover, assume that the maps* $v \mapsto \mathbb{E}[|Y_k(t)| \mid X'\bar{\beta} = v] f_{\bar{\beta}}(v)$, $v \in \mathbb{R}$, $t \in [0,1]$, *are uniformly Lipschitz (the Lipschitz constant does not depend on* t*). Then*

$$\max_{1\le i\le n} \sup_{t\in[0,1]} \sup_{\beta\in\mathcal{B}_n} \left| \frac{1}{n-1} \sum_{k\neq i} Y_k(t) \frac{1}{g} \left[L_{ik}(\beta) - L_{ik}(\bar{\beta}) \right] \right| = O_{\mathbb{P}}(n^{-1/2} g^{-1/2} \log^{1/2} n + b_n).$$

Moreover,

$$\max_{1\le i\le n} \sup_{t\in[0,1]} \sup_{\beta\in\mathcal{B}_n} \left| \frac{1}{n-1} \sum_{k\neq i} \{Y_k(t) - \mathbb{E}[Y_k(t) \mid X_k'\bar{\beta}]\} \frac{1}{g} L_{ik}(\bar{\beta}) \right| = O_{\mathbb{P}}(n^{-1/2} g^{-1/2} \log^{1/2} n).$$

Proof of Lemma 1. Recall that $Y_i(t) \equiv Y_i$ (in the case of SIM for mean regression) or $Y_i(t) = \mathbf{1}\{Y_i \leq \Phi^{-1}(t)\}$ (for the case of single-index assumption on the conditional law), and $r_i(t;\bar\beta) = \mathbb{E}[Y_i(t) \mid Z(\bar\beta)]$, $t \in [0,1]$. For any $t \in [0,1]$ we decompose

$$\begin{aligned}
\frac{1}{ng}\sum_{k\neq i} Y_k(t) L_{ik}(\beta) &= \frac{1}{ng}\sum_{k=1}^{n}\{Y_k(t)L\left((X_i - X_k)'\beta/g\right) - \mathbb{E}\left[Y(t)L\left((X_i - X)'\beta/g\right) \mid X_i\right]\} \\
&\quad + \mathbb{E}\left[Y(t)g^{-1}L\left((X_i - X)'\beta/g\right) \mid X_i\right] - n^{-1}g^{-1}L(0)Y_i(t) \\
&= \Sigma_{1ni}(\beta,t) + \Sigma_{2ni}(\beta,t) - n^{-1}g^{-1}L(0)Y_i(t).
\end{aligned}$$

The moment condition on Y guarantees that $\max_{1\leq i\leq n}\sup_t |Y_i(t)| = o_{\mathbb{P}}(n^b)$ for some $0 < b < 1/8$. This and the fact that $ng^{4/3}/\log n \to \infty$ make that $\max_{1\leq i\leq n}\sup_t n^{-1}g^{-1}|Y_i(t)| = o_{\mathbb{P}}(n^{-1/2}g^{-1/2}\log^{1/2} n)$. On the other hand, by Lemma 5,

$$\max_{1\leq i\leq n}\sup_{t\in[0,1]}\sup_{\beta\in\mathcal{B}_n}\left|\Sigma_{2ni}(\beta,t) - \Sigma_{2ni}(\bar\beta,t)\right| = O_{\mathbb{P}}(b_n).$$

It remains to uniformly bound $\Sigma_{1ni}(\beta,t)$ and for this purpose we use empirical process tools. Let us introduce some notation. Let $\mathcal{G}$ be a class of functions of the observations with envelope function G and let

$$J(\delta,\mathcal{G},L^2) = \sup_Q \int_0^\delta \sqrt{1+\log N(\varepsilon\|G\|_2,\mathcal{G},L^2(Q))}\,d\varepsilon, \qquad 0<\delta\leq 1,$$

denote the uniform entropy integral, where the supremum is taken over all finitely discrete probability distributions Q on the space of the observations, and $\|G\|_2$ denotes the norm of G in $L^2(Q)$. Let $Z_1, \cdots, Z_n$ be a sample of independent observations and let

$$\mathbb{G}_n g = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma(Z_i), \qquad \gamma \in \mathcal{G},$$

be the empirical process indexed by $\mathcal{G}$. If the covering number $N(\varepsilon, \mathcal{G}, L^2(Q))$ is of polynomial order in $1/\varepsilon$, there exists a constant $c > 0$ such that $J(\delta, \mathcal{G}, L^2) \leq c\delta\sqrt{\log(1/\delta)}$ for $0 < \delta < 1/2$. Now if $\mathbb{E}\gamma^2 < \delta^2 \mathbb{E}G^2$ for every γ and some $0 < \delta < 1$, and $\mathbb{E}G^{(4\upsilon-2)/(\upsilon-1)} < \infty$ for some $\upsilon > 1$, under mild additional measurability conditions that are satisfied in our context, Theorem 3.1 of van der Vaart and Wellner (2011) implies

$$\sup_{\mathcal{G}} |\mathbb{G}_n \gamma| = J(\delta, \mathcal{G}, L^2) \left(1 + \frac{J(\delta^{1/\upsilon}, \mathcal{G}, L^2)}{\delta^2 \sqrt{n}} \frac{\|G\|_{(4\upsilon-2)/(\upsilon-1)}^{2-1/\upsilon}}{\|G\|_2^{2-1/\upsilon}} \right)^{\upsilon/(2\upsilon-1)} \|G\|_2 O_{\mathbb{P}}(1), \tag{S2.4}$$

where $\|G\|_2^2 = \mathbb{E}G^2$ and the $O_{\mathbb{P}}(1)$ term is independent of n. Note that the family $\mathcal{G}$ could change with n, as soon as the envelope is the same for all n. We apply this result to the family of functions $\mathcal{G} = \{\gamma(\cdot; \beta, w, t) - \gamma(\cdot; \bar{\beta}, w, t) : t \in [0,1], \beta \in \mathcal{B}, w \in \mathbb{R}\}$ where

$$\gamma(Y, X; \beta, w, t) = Y(t) L((X'\beta - w) g^{-1}))$$

for a sequence g that converges to zero and the envelope

$$G(Y,X) = \sup_{t\in[0,1]} |Y(t)| \sup_{w\in\mathbb{R}} L(w).$$

Its entropy number is of polynomial order in $1/\varepsilon$, independently of n, as $L(\cdot)$ is of bounded variation and the families of indicator functions have polynomial complexity, see for instance van der Vaart (1998). Now for any $\gamma\in\mathcal{G}$, $\mathbb{E}\gamma^2 \leq Cg\mathbb{E}G^2$, for some constant C. Let $\delta = g^{1/2}$, so that $\mathbb{E}\gamma^2 \leq C'\delta^2\mathbb{E}G^2$, for some constant C' and $\upsilon = 3/2$, which corresponds to $\mathbb{E}G^8 < \infty$ that is guaranteed by our assumptions. Thus the bound in (S2.4) yields

$$\sup_{\mathcal{G}} \left| \frac{1}{g\sqrt{n}}\, \mathbb{G}_n\gamma \right| = \frac{\log^{1/2}(n)}{\sqrt{ng}} \left[1 + n^{-1/2}g^{-2/3}\log^{1/2}(n) \right]^{3/4} O_{\mathbb{P}}(1),$$

where the $O_{\mathbb{P}}(1)$ term is independent of n. Since $ng^{4/3}/\log n \to \infty$,

$$\max_{1\leq i\leq n} \sup_{t\in[0,1]} \sup_{\beta\in\mathcal{B}_n} |\Sigma_{1ni}(\beta,t) - \Sigma_{1ni}(\bar{\beta},t)| = O_{\mathbb{P}}(n^{-1/2}g^{-1/2}\log^{1/2} n).$$

The second part of the statement is now obvious. $\square$

Lemma 2. *Assume that the density $f_{\bar\beta}(\cdot)$ is Lipschitz. Then*

$$\max_{1\le i\le n}\left|\frac{1}{n-1}\sum_{k\neq i}\frac{1}{g}L_{ik}(\bar\beta)-f_{\bar\beta}(X_i'\bar\beta)\right| = O_{\mathbb{P}}(n^{-1/2}g^{-1/2}\log^{1/2} n+g).$$

Proof of Lemma 2. We can write

$$\begin{aligned}\frac{1}{n-1}\sum_{k\neq i}\frac{1}{g}L_{ik}(\bar\beta)-f_{\bar\beta}(X_i'\bar\beta) \quad &= \quad \frac{1}{n}\sum_{k=1}^{n}\left\{g^{-1}L_{ik}(\bar\beta)-\mathbb{E}[g^{-1}L_{ik}(\bar\beta)\mid X_i]\right\}\\ &\quad +\mathbb{E}[g^{-1}L_{ik}(\bar\beta)\mid X_i]-f_{\bar\beta}(X_i'\bar\beta)+O(n^{-1}g^{-1}).\end{aligned}$$

By the empirical process arguments used in Lemma 1, the sum on the right-hand side of the display is of rate $O_{\mathbb{P}}(n^{-1/2}g^{-1/2}\log^{1/2} n)$ uniformly with respect to i. The Lipschitz property of $f_{\bar\beta}$ and the fact that $\int|vL(v)|dv<\infty$ guarantee that

$$\max_{1\le i\le n}|\mathbb{E}[g^{-1}L_{ik}(\bar\beta)\mid X_i]-f_{\bar\beta}(X_i'\bar\beta)|\le Cg$$

for some constant C. □

Lemma 3. *For any $t\in[0,1]$ let $Y_k(t)$, $1\le k\le n$, be an independent sample from a random variable $Y(t)$ defined like in the proof of Proposition 1. Let $r(v;t,\bar\beta)=\mathbb{E}[Y(t)\mid X'\bar\beta=v]$, $v\in\mathbb{R}$, and assume that $r(\cdot;t,\bar\beta)$*

is twice differentiable and the second derivative is bounded by a constant independent of t. If $r'(v;t,\bar\beta)$ is the first derivative of $r(\cdot;t,\bar\beta)$, then, for any $t\in[0,1]$,

$$\frac{1}{n-1}\sum_{k\neq i}\{r(X_i'\bar\beta;t,\bar\beta)-r(X_k'\bar\beta;t,\bar\beta)\}\frac{1}{g}L_{ik}(\bar\beta)=r'(X_i'\bar\beta;t,\bar\beta)gD_{1,ni}+g^2D_{1,ni}(t),$$

where $\max_{1\leq i\leq n}|D_{1,ni}|=n^{-1/2}g^{-1/2}\log^{1/2}n$ *and* $\max_{1\leq i\leq n}\sup_{t\in[0,1]}|D_{1,ni}(t)|=O_{\mathbb{P}}(1)$.

Proof of Lemma 3. By Taylor expansion

$$\begin{aligned}\frac{1}{n-1}\sum_{k\neq i}\{r(X_i'\bar\beta;t,\bar\beta)-r(X_k'\bar\beta;t,\bar\beta)\}\frac{1}{g}L_{ik}(\bar\beta)&=r'(X_i'\bar\beta;t,\bar\beta)\frac{1}{n}\sum_{k=1}^{n}(X_i-X_k)'\bar\beta\frac{1}{g}L_{ik}(\bar\beta)\\&+\frac{1}{n}\sum_{k=1}^{n}r''(x_{ik}(t);t,\bar\beta)[(X_i-X_k)'\bar\beta]^2\frac{1}{g}L_{ik}(\bar\beta),\end{aligned}$$

where r'' stands for the second derivative with respect to v and $x_{ik}(t)$ is a point between $X_i'\bar\beta$ and $X_k'\bar\beta$. Since $L(\cdot)$ is symmetric, by the empirical process arguments as in Lemma 1

$$\max_{1\leq i\leq n}\left|\frac{1}{n}\sum_{k=1}^{n}\frac{(X_i-X_k)'\bar\beta}{g}\frac{1}{g}L_{ik}(\bar\beta)\right|=O_{\mathbb{P}}(n^{-1/2}g^{-1/2}\log^{1/2}n).$$

The result follows taking absolute values in the last sum in the last display, using the boudedness of r'' and the fact that

$$\max_{1\leq i\leq n}\left|\frac{1}{n}\sum_{k=1}^{n}\frac{[(X_i-X_k)'\bar{\beta}]^2}{g^2}\frac{1}{g}L_{ik}(\bar{\beta})-f_{\bar{\beta}}(X_i'\bar{\beta})\int_{\mathbb{R}}v^2|L(v)|dv\right|=o_{\mathbb{P}}(1).$$

□

Lemma 4. *Assume that* $\mathbb{E}[\exp(a\|X\|)]<\infty$ *for some* $a>0$. *Moreover the kernels* K *and* L *are of bounded variation, differentiable except at most a finite set of points, and* $\int_{\mathbb{R}}|uK(u)|du<\infty$. *Let* $\mathcal{B}_n$ *be a subset in the parameter space such that the event defined in equation (7.2) with* $b_n\to 0$ *and* $b_n n^{1/2}/\log n\to\infty$ *has probability tending to 1. Let*

$$K_{12}(\beta)=K((X_1-X_2)'\beta/h),\quad L_{12}(\beta)=L((X_1-X_2)'\beta/g)$$

and $\phi(\beta)=\phi((X_1-X_2)'\mathbf{A}(\beta))$. *If the density* $f_{\bar{\beta}}$ *is Lipschitz with constant* $C_{1,\bar{\beta}}$, *then there exists a constant* C *depending only on* K, L, $\|f_{\bar{\beta}}\|_\infty$ *and* $C_{1,\bar{\beta}}$ *such that*

$$\mathbb{P}\left\{\mathbb{E}\left[\sup_{b\in\mathcal{B}_n}\left|K_{12}(\beta)\phi_{12}(\beta)-K_{12}(\bar{\beta})\phi_{12}(\bar{\beta})\right|\mid X_1\right]\leq Cb_nh^{1/2}\right\}\to 1,\quad \text{(S2.5)}$$

$$\mathbb{E}\left[\sup_{b\in\mathcal{B}_n}\left|K_{12}(\beta)\phi_{12}(\beta)-K_{12}(\bar\beta)\phi_{12}(\bar\beta)\right|\right]\leq Cb_nh^{1/2}, \tag{S2.6}$$

$$\mathbb{P}\left\{\mathbb{E}\left[\sup_{b\in\mathcal{B}_n}\left|L_{12}(\beta)-L_{12}(\bar\beta)\right|^2\mid X_1\right]\leq Cb_ng^{-1}\right\}\to 1 \tag{S2.7}$$

$$\mathbb{P}\left\{\mathbb{E}\left[\sup_{b\in\mathcal{B}_n}\left|L_{13}(\beta)-L_{13}(\bar\beta)\right|^2|K_{12}(\bar\beta)|^2\mid X_2,X_3\right]\leq Chb_ng^{-1}\right\}\to 1, \tag{S2.8}$$

and

$$\mathbb{E}\left[\sup_{b\in\mathcal{B}_n}\left|L_{13}(\beta)-L_{13}(\bar\beta)\right|^2|K_{12}(\bar\beta)|^2\phi_{12}^2(\bar\beta)\right]\leq Chb_ng^{-1}, \tag{S2.9}$$

In Lemma 4 we provide different bounds for $L(\cdot)$ and $K(\cdot)$ because the bandwidths g and h have to satisfy the condition $h/g^2\to 0$. Hence we need less restrictive conditions on the range of h if we want to allow for a larger domain for the pair (g,h).

Proof of Lemma 4. Since the univariate kernel K is of bounded kernels, let K_1 and K_2 non decreasing bounded functions such that $K=K_1-K_2$ and denote $K_{1h}=K_1(\cdot/h)$. Clearly, it is sufficient to prove the result with K_1. Similar arguments apply for K_2 and hence we get the results for K. For simpler writings we assume that K is differentiable and let $K_1(x) = \int_{-\infty}^{x} [K\,'(t)]^+ dt$ and $K_2(x) = \int_{-\infty}^{x} [K\,'(t)]^- dt$, $x \in \mathbb{R}$. Here $[K\,']^+$ (resp. $[K\,']^-$) denotes the positive (resp. negative) part of $K\,'$. The general case where a finite set of nondifferentiability is allowed can be handled with obvious modifications. Let $K_{1h}(t) = K_1(t/h)$ and recall that $Z_i(\beta) = X_i'\beta$. Note that $|\exp(-t^2) - \exp(-s^2)| \leq \sqrt{2}|t-s|$. For any $\beta \in \mathcal{B}_n$ and an elementary event in the set $\mathcal{C}_n = \{\max_{1\leq i\leq n} \|X_i\| \leq c \log n\} \subset \mathcal{E}_n$ for some large constant c,

$$
\begin{aligned}
&\left|K_{1h}\left(Z_1(\beta) - Z_2(\beta)\right)\phi_{12}(\beta) - K_{1h}\left(Z_1(\bar\beta) - Z_2(\bar\beta)\right)\phi_{12}(\bar\beta)\right| \\
&\qquad\qquad \leq \sqrt{2}\, b_n K_{1h}(Z_1(\bar\beta) - Z_2(\bar\beta) + 2b_n) \\
&\qquad + [K_{1h}(Z_1(\bar\beta) - Z_2(\bar\beta) + 2b_n) - K_{1h}\left(Z_1(\bar\beta) - Z_2(\bar\beta) - 2b_n\right)]\phi_{12}(\bar\beta).
\end{aligned}
$$

The upper bound on the left-hand side is uniform with respect to β. By a suitable change of variable and since the density $f_{\overline{\beta}}$ is bounded, it is easy to check that

$$
\mathbb{E}\left[K_{1h}\left(Z_1(\bar\beta) - Z_2(\bar\beta) + 2b_n\right) \mid Z_1(\bar\beta)\right]
$$

is bounded by a constant times h. Next, note that since $nh \to \infty$, there exists a constant C' independent of n such that on the set $\mathcal{C}_n$ we have

$|Z_1(\bar\beta) - Z_2(\bar\beta) \pm 2b_n|/h \leq C' h^{-1/2}$. Then, applying twice a change of variables and using the Lipschitz property of $f_{\bar\beta}$, on the set $\mathcal{C}_n$,

$$
\begin{aligned}
&\mathbb{E}\left[\left|K_{1h}\left(Z_1(\bar\beta) - Z_2(\bar\beta) + 2b_n\right) - K_{1h}\left(Z_1(\bar\beta) - Z_2(\bar\beta) - 2b_n\right)\right| \mathbf{1}\{\mathcal{C}_n\} \mid Z_1(\bar\beta)\right] \\
&\leq h \int_{[-C'/h^{1/2},\, C'/h^{1/2}]} K_1(u) \left| f_{\bar\beta}(2b_n + Z_1(\bar\beta) - uh) - f_{\bar\beta}(-2b_n + Z_1(\bar\beta) - uh)\right| du \\
&\qquad \leq h \times \sup_{t\in\mathbb{R}} \left| f_{\bar\beta}(2b_n + t) - f_{\bar\beta}(-2b_n + t)\right| \int_{[-C'/h^{1/2},\, C'/h^{1/2}]} K_1(u) du \\
&\qquad\qquad \leq C h^{1/2} b_n,
\end{aligned}
$$

for some constant $C > 0$. Since by a suitable choice of c the probability of $\mathbf{1}\{\mathcal{C}_n\}$ given $Z_1(\bar\beta)$ could be made smaller than any fixed negative power of n, and the probability of the event $\{|Z_1(\bar\beta)| \leq c \log n\}$ could be also made very small, the bound in the last display implies the statement (S2.5). For the statement (S2.6) it suffices to take expectation.

For the bound in equation (S2.7), recall that $L(t) = L(|t|)$ for any $t \in \mathbb{R}$ so that we can consider only nonnegative t. Moreover, without loss of generality we can consider L nonnegative and decreasing on $[0, \infty)$, otherwise, since L is of bounded variation, it could be written an the difference of two nonnegative decreasing functions on $[0, \infty)$. Moreover, let $Z_{13}(\beta) = |Z_1(\beta) - Z_3(\beta)|$ and $L_{g,13}(\beta) = L(Z_{13}(\beta)/g)$. We split the problem in two cases: $Z_{13}(\beta) \leq Z_{13}(\bar\beta)$ and $Z_{13}(\beta) > Z_{13}(\bar\beta)$. Then, for $\beta \in \mathcal{B}_n$

and on the set $\mathcal{C}_n$ we have

$$
\begin{aligned}
&\left|L_{g,13}(\beta) - L_{g,13}(\bar{\beta})\right| \mathbf{1}\{Z_{13}(\beta) \leq Z_{13}(\bar{\beta})\} \\
&\leq [L(0) - L_{g,13}(\bar{\beta})]\mathbf{1}\{Z_{13}(\beta) \leq Z_{13}(\bar{\beta}), Z_{13}(\bar{\beta}) \leq 2b_n\} \\
&+ [L_{g,13}(\beta) - L_{g,13}(\bar{\beta})]\mathbf{1}\{Z_{13}(\beta) \leq Z_{13}(\bar{\beta}), Z_{13}(\bar{\beta}) \geq 2b_n\} \\
&\leq Cb_n^2 g^{-2}\mathbf{1}\{Z_{13}(\bar{\beta}) \leq 2b_n\} \\
&+ \left[L((Z_{13}(\bar{\beta}) - 2b_n)/g) - L(Z_{13}(\bar{\beta})/g)\right] \mathbf{1}\{Z_{13}(\beta) \leq Z_{13}(\bar{\beta}), Z_{13}(\bar{\beta}) \geq 2b_n\} \\
&= Cb_n^2 g^{-2}\mathbf{1}\{Z_{13}(\bar{\beta}) \leq 2b_n\} + A_n
\end{aligned}
$$

and

$$
\begin{aligned}
&\left|L_{g,13}(\beta) - L_{g,13}(\bar{\beta})\right| \mathbf{1}\{Z_{13}(\beta) > Z_{13}(\bar{\beta})\} \\
&\leq \left[L(Z_{13}(\bar{\beta})/g) - L((Z_{13}(\bar{\beta}) + 2b_n)/g)\right] \mathbf{1}\{Z_{13}(\beta) > Z_{13}(\bar{\beta})\} = B_n,
\end{aligned}
$$

for some constant C. Let us notice that

$$\begin{aligned}
A_n + B_n &\leq [L([Z_{13}(\bar{\beta}) - 2b_n]/g) - L([Z_{13}(\bar{\beta}) + 2b_n]/g)]\mathbf{1}\{Z_{13}(\bar{\beta}) \geq 2b_n\} \\
&\quad + [L([Z_{13}(\bar{\beta})]/g) - L([Z_{13}(\bar{\beta}) + 2b_n]/g)]\mathbf{1}\{0 \leq Z_{13}(\bar{\beta}) \leq 2b_n\} \\
&\leq [L([Z_{13}(\bar{\beta}) - 2b_n]/g) - L([Z_{13}(\bar{\beta}) + 2b_n]/g)]\mathbf{1}\{Z_{13}(\bar{\beta}) \geq 2b_n\} \\
&\quad + Cb_n^2 g^{-2}\mathbf{1}\{Z_{13}(\bar{\beta}) \leq 2b_n\} \\
&\leq [L([Z_{13}(\bar{\beta}) - 2b_n]/g) - L([Z_{13}(\bar{\beta}) + 2b_n]/g)] \\
&\quad + 2Cb_n^2 g^{-2}\mathbf{1}\{Z_{13}(\bar{\beta}) \leq 2b_n\} \\
&= D_n + 2Cb_n^2 g^{-2}\mathbf{1}\{Z_{13}(\bar{\beta}) \leq 2b_n\}.
\end{aligned}$$

On the other hand, $0 \leq D_n \leq 4b_n g^{-1}|L'(\widetilde{Z})|$ where $\widetilde{Z}$ is some value such that $|\widetilde{Z} - Z_{13}(\bar{\beta})| \leq 2b_n g^{-1}$. Since, for some constant c, $|L'(v)| \leq c|v|$ in a neighborhood of the origin,

$$D_n \leq 4b_n g^{-1}|L'(Z_{13}(\bar{\beta})| + C' b_n^2 g^{-2},$$

for some constant C'. Since L' is bounded, deduce that $|L_{g,13}(\beta) - L_{g,13}(\bar{\beta})|^2$ is bounded by $Cb_n^2 g^{-1}|g^{-1}L'(Z_{13}(\bar{\beta})| + o(b_n^2 g^{-1})$ for some constant C. Take conditional expectation given X_1, that is the same with the conditional expectation given $Z_1(\beta)$, and deduce the bound in equation (S2.7).

On the set of events $\mathcal{C}_n$,

$$\sup_{\beta\in\mathcal{B}_n}\left|L_{13}(\beta)-L_{13}(\bar\beta)\right||K_{12}(\bar\beta)|\le\{D_n+3Cb_n^2g^{-2}\mathbf{1}\{Z_{13}(\bar\beta)\le 2b_n\}\}|K_{12}(\bar\beta)|.$$

Take conditional expectation and use standard change of variables to derive the bound in equation (S2.8). Take expectation and remember that $\phi_{12}(\bar\beta)$ is bounded to derive the moment bound in equation (S2.9). □

Lemma 5. *Under the conditions of Lemma 1*

$$\sup_{t\in[0,1]}\sup_{\beta\in\mathcal{B}_n}\max_{1\le i\le n}\left|\Sigma_{2ni}(\beta,t)-\Sigma_{2ni}(\bar\beta,t)\right|=O_{\mathbb{P}}(b_n).$$

Proof of Lemma 5. We can write

$$\begin{aligned}\left|\Sigma_{2ni}(\beta,t)-\Sigma_{2ni}(\bar\beta,t)\right|&\le\mathbb{E}\left[|Y(t)|\left|g^{-1}L\left((X_i-X)'\beta/g\right)-g^{-1}L\left((X_i-X)'\bar\beta/g\right)\right|\mid X_i\right]\\&=\mathbb{E}\left[\mathbb{E}\left\{|Y(t)|X\right\}g^{-1}\left|L\left((X_i-X)'\beta/g\right)-L\left((X_i-X)'\bar\beta/g\right)\right|\mid X_i\right].\end{aligned}$$

Now, we can apply the monotonicity argument we used in Lemma 4 and deduce the bound. □

Lemma 6. *Under the conditions of Proposition 2,* $I_1(\beta_0)=o_{\mathbb{P}}\left(n^{-1}h^{-1/2}\right)$.

Proof of Lemma 6. With the notation defined in equation (S1.3) we have

$$I_1(\beta_0) = \frac{1}{n(n-1)^3 g^2 h} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq j} \langle (r_i - r_k)(\cdot;\, \beta_0),\, (r_j - r_l)(\cdot;\, \beta_0) \rangle_{L^2} L_{ik} L_{jl} K_{ij} \phi_{ij}$$

and if we denote by $I_{1,1}(\beta_0)$ the term where i, j, k and l are all different, then

$$\mathbb{E}[I_{1,1}(\beta_0)] = \frac{(n-2)(n-3)}{(n-1)^2 g^2 h} \mathbb{E}\left[\langle \mathbb{E}[(r_i - r_k)(\cdot;\, \beta_0) L_{ik} \mid Z_i(\beta_0)],\right.$$
$$\left. \mathbb{E}[(r_j - r_l)(\cdot;\, \beta_0) L_{jl} \mid Z_j(\beta_0)] \rangle_{L^2} K_{ij} \phi_{ij}\right] = O\left(g^4\right)$$

as soon as $g^{-1}\mathbb{E}[(r_i - r_k)(t;\, \beta_0) L_{ik}(\beta_0) \,|\, Z_i(\beta_0)] = O(g^2) D(t; Z_i(\beta_0))$ with $D(\cdot)$ bounded, which is guaranteed by Assumption 1-(c). When i, j, k and l take no more than 3 different values, the number of terms is reduced by a factor n, and thus we have that $\mathbb{E}[I_{1,2}(\beta_0)] = O(n^{-1} g^{-1}) = o\left(n^{-1} h^{-1/2}\right)$. Similar reasoning can be applied to prove that $\mathbb{E}[I_1^2(\beta_0)] = o(n^{-2} h^{-1})$. See also Proposition A.1. in Fan and Li (1996). □

Lemma 7. *Under the conditions of Proposition 2,* $I_3(\beta_0) = o_{\mathbb{P}}\left(n^{-1} h^{-1/2}\right)$

.

Proof of Lemma 7. Write

$$\begin{aligned}
I_3(\beta_0) &= \frac{1}{n(n-1)^3 g^2 h}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i}\sum_{l\neq j}\left\langle \epsilon_k(\cdot),\, \epsilon_l(\cdot)\right\rangle_{L^2} L_{ik}L_{jl}K_{ij}\phi_{ij} \\
&= \frac{1}{n(n-1)^3 g^2 h}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i}\sum_{l\neq j,k}\left\langle \epsilon_k(\cdot),\, \epsilon_l(\cdot)\right\rangle_{L^2} L_{ik}L_{jl}K_{ij}\phi_{ij} \\
&\quad + \frac{1}{n(n-1)^3 g^2 h}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}\left\|\epsilon_k(\cdot)\right\|_{L^2}^2 L_{ik}L_{ji}K_{ij}\phi_{ij} \\
&\quad + \frac{1}{n(n-1)^3 g^2 h}\sum_{i=1}^{n}\sum_{j\neq i}\left\|\epsilon_j(\cdot)\right\|_{L^2}^2 L_{ij}L_{ji}K_{ij}\phi_{ij} \\
&= I_{3,1}(\beta_0) + I_{3,2}(\beta_0) + I_{3,3}(\beta_0).
\end{aligned}$$

Then

$$\begin{aligned}
\mathbb{E}\left[I_{3,1}(\beta_0)\right] &= \frac{1}{(n-1)^2 g^2 h}\mathbb{E}\left[\left\langle \epsilon_1(\cdot),\, \epsilon_2(\cdot)\right\rangle_{L^2} L_{12}^2 K_{12}\phi_{12}\right] \\
&= O\left(n^{-2}g^{-2}\right)\mathbb{E}\left[\left|\left\langle \epsilon_1(\cdot),\, \epsilon_2(\cdot)\right\rangle_{L^2} h^{-1}K_{12}\right|\right] \\
&= O\left(n^{-2}g^{-2}\right),
\end{aligned}$$

$\mathbb{E}\left[I_{3,2}(\beta_0)\right] = O\left(n^{-1}g^{-1}\right)$ and $\mathbb{E}\left[I_{3,3}(\beta_0)\right] = O\left(n^{-2}g^{-2}\right)$, thus $\mathbb{E}\left[I_3(\beta_0)\right] = o\left(n^{-1}h^{-1/2}\right)$. By quite straightforward but tedious calculations, it can be proved that $\mathbb{E}\left[I_3^2(\beta_0)\right] = o\left(n^{-2}h^{-1}\right)$ and the rate of $I_3(\beta_0)$ follows. □

Lemma 8. *Under the conditions of Proposition 2,*

$$V_n^2(\beta_0) = \sum_{i=2}^{n} \mathbb{E}\left[G_{n,i}^2(\beta_0) \mid \mathcal{F}_{n,i-1}\right] \to 1, \quad \textit{in probability,}$$

and the martingale difference array $\{G_{n,i},\, \mathcal{F}_{n,i},\ 1 \le i \le n\}$ *satisfies the conditional Lindeberg condition*

$$\forall \varepsilon > 0, \quad \sum_{i=2}^{n} \mathbb{E}\left[G_{n,i}^2 I\left(|G_{n,i}| > \varepsilon\right) \mid \mathcal{F}_{n,i-1}\right] \to 0, \quad \textit{in probability.}$$

Proof of Lemma 8. First, decompose

$$\begin{aligned} V_n^2(\beta_0) &= \frac{4}{\omega_n^2 (n-1)^2 h} \sum_{i=2}^{n} \int\int \Gamma(s,t) \hat{f}_{\beta_0,i}^2 \left(\sum_{j=1}^{i-1} \epsilon_j(s) \hat{f}_{\beta_0,j} K_{ij} \phi_{ij} \right) \\ &= \frac{4}{\omega_n^2 (n-1)^2 h} \sum_{i=2}^{n} \sum_{j=1}^{i-1} \int\int \Gamma(s,t) \hat{f}_{\beta_0,i}^2 \epsilon_j(s) \epsilon_j(t) \hat{f}_{\beta_0,j}^2 K_{ij}^2 \phi_{ij}^2 ds\, dt \\ &+ \frac{8}{\omega_n^2 (n-1)^2 h} \sum_{i=3}^{n} \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \int\int \Gamma(s,t) \hat{f}_{\beta_0,i}^2 \epsilon_j(s) \epsilon_k(t) \hat{f}_{\beta_0,j} \hat{f}_{\beta_0,k} K_{ij}^2 \phi_{ij}^2 ds\, dt \\ &= A_n(\beta_0) + B_n(\beta_0). \end{aligned} \tag{S2.10}$$

We have

$$
\begin{aligned}
&\mathbb{E}\left[A_n\left(\beta_0\right)\right]=\mathbb{E}\left[\mathbb{E}\left[A_n\left(\beta_0\right) \mid X_1, \ldots, X_n\right]\right] \\
&\quad=\mathbb{E}\left[\frac{2 n}{\omega_n^2\left(\beta_0\right)(n-1) h} \int \int \Gamma(s, t) \hat{f}_{\beta_0, i}^2 \mathbb{E}\left[\epsilon_j(s) \epsilon_j(t)\right] \hat{f}_{\beta_0, j}^2 K_{i j}^2 \phi_{i j}^2 d s\, d t\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad=\frac{n}{n-1} \xrightarrow{n \rightarrow \infty} 1 .
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
\operatorname{Var}\left(A_n\left(\beta_0\right)\right) \leq\; & \frac{64\|\phi\|_{\infty}^4}{(n-1)^4 h^2} \sum_{i=3}^n \sum_{j=2}^{i-1} \sum_{j^{\prime}=1}^{j-1} \mathbb{E}\left[\omega_n^{-2}\left(\beta_0\right) \hat{f}_{\beta_0, i}^4 \hat{f}_{\beta_0, j}^2 \hat{f}_{\beta_0, j^{\prime}}^2 K_{i j}^2 K_{i j^{\prime}}^2\right] \\
& \times \int \int \int \int \Gamma^2(s, t) \Gamma^2(u, v)\, d s d t d u d v \\
& +\frac{32\|\phi\|_{\infty}^4}{(n-1)^4 h^2} \sum_{i=3}^n \sum_{i^{\prime}=2}^{i-1} \sum_{j=1}^{i^{\prime}-1} \mathbb{E}\left[\omega_n^{-2}\left(\beta_0\right) \hat{f}_{\beta_0, i}^2 \hat{f}_{\beta_0, i^{\prime}}^2 \hat{f}_{\beta_0, j}^4 K_{i j}^2 K_{i^{\prime} j}^2\right] \\
& \times \int \int \int \int \Gamma(s, t) \Gamma(u, v) \mathfrak{G}(s, t, u, v)\, d s d t d u d v \\
& +\frac{16\|\phi\|_{\infty}^4}{(n-1)^4 h^2} \sum_{i=3}^n \sum_{i^{\prime}=2}^{i-1} \sum_{j=1}^{i^{\prime}-1} \mathbb{E}\left[\omega_n^{-2}\left(\beta_0\right) \hat{f}_{\beta_0, i}^4 \hat{f}_{\beta_0, j}^4 K_{i j}^4\right] \\
& \times \int \int \int \int \Gamma(s, t) \Gamma(u, v) \mathfrak{G}(s, t, u, v)\, d s d t d u d v \\
=\; & o\left(n^{-1} h^{-1 / 2}\right),
\end{aligned}
$$

where $\mathfrak{G}(s, t, u, v)=\mathbb{E}[\epsilon(s) \epsilon(t) \epsilon(u) \epsilon(v)]$. The decomposition of $\mathbb{E}\left[B_n^2\right]$ involves the same type of terms and is therefore also of rate $o\left(n^{-1} h^{-1 / 2}\right)$, so that the convergence of $V_n^2\left(\beta_0\right)$ is met.

For the conditional Lindeberg condition, we have $\forall \varepsilon > 0$, $\forall n \geq 1$ and $1 < i \leq n$

$$\mathbb{E}\left[G_{n,i}^2 I\left(\left|G_{n,i}\right| > \varepsilon\right) \mid \mathcal{F}_{n,i-1}\right] \leq \frac{\mathbb{E}\left[G_{n,i}^4 \mid \mathcal{F}_{n,i-1}\right]}{\varepsilon^2}.$$

Then

$$\begin{aligned}
&\sum_{i=2}^{n} \mathbb{E}\left[G_{n,i}^2 I\left(\left|G_{n,i}\right| > \varepsilon\right) \mid \mathcal{F}_{n,i-1}\right] \\
&\qquad \leq \frac{1}{\varepsilon^2} \sum_{i=2}^{n} \mathbb{E}\left[G_{n,i}^4 \mid \mathcal{F}_{n,i-1}\right] \\
&\qquad \leq \frac{1}{\varepsilon^2} \frac{16}{(n-1)^4 h^2} \sum_{i=2}^{n} \int\int\int\int \mathfrak{G}\left(s_1, s_2, s_3, s_4\right) \hat{f}_{\beta_0,i}^4 \\
&\qquad\qquad\qquad\qquad\qquad \times \prod_{k=1}^{4} \sum_{j_k=1}^{i-1} \epsilon_{j_k}\left(s_k\right) \hat{f}_{\beta_0,j_k} K_{ij_k} \phi_{ij_k} ds_k.
\end{aligned}$$

The expectation of the last majorant is of rate

$$\begin{aligned}
O\left(n^{-1}\right) &\int\int\int\int \mathfrak{G}\left(s_1, s_2, s_3, s_4\right) \Gamma\left(s_1, s_2\right) \Gamma\left(s_3, s_4\right) ds_1 ds_2 ds_3 ds_4 \\
&\times \mathbb{E}\left[\hat{f}_{\beta_0,i}^4 \hat{f}_{\beta_0,j}^2 \hat{f}_{\beta_0,j'}^2 h^{-1} K_{ij}^2 h^{-1} K_{ij'}^2 \phi_{ij}^2 \phi_{ij'}^2\right] \\
&+ O\left(n^{-2} h^{-1}\right) \sum_{i=2}^{n} \int\int\int\int \mathfrak{G}^2\left(s_1, s_2, s_3, s_4\right) ds_1 ds_2 ds_3 ds_4 \\
&\times \mathbb{E}\left[\hat{f}_{\beta_0,i}^4 \hat{f}_{\beta_0,j}^4 h^{-1} K_{ij}^4 \phi_{ij}^4\right] \\
=& o\left(n^{-1} h^{-1/2}\right).
\end{aligned}$$

□

Lemma 9. *Under the conditions of Proposition 2,* $\omega_n^2(\beta_0) \to \omega^2(\beta_0) > 0$, *in probability.*

Proof of Lemma 9. We have

$$\mathbb{E}\left[\omega_n^2(\beta_0)\right] = 2\mathbb{E}\left[\hat{f}_{\beta,i}\hat{f}_{\beta,j}h^{-1}K_{ij}^2(\beta)\phi_{ij}^2(\beta)\right] \times \int\int \Gamma^2(s,t)\,ds\,dt.$$

On the other hand,

$$\begin{aligned}
&\mathbb{E}\left[\hat{f}_{\beta,i}\hat{f}_{\beta,j}h^{-1}K_{ij}^2\phi_{ij}^2\right] \\
=&\frac{1}{g^2h}\mathbb{E}\left[\sum_{k\neq i}\sum_{l\neq j}\sum_{k'\neq i}\sum_{l'\neq j} L_{ik}L_{jl}L_{ik'}L_{jl'}h^{-1}K_{ij}^2\phi_{ij}^2\right] \\
=&\frac{1}{g^2h(n-1)^2}\mathbb{E}\left[\sum_{k\neq i}\sum_{l\neq j}\sum_{k'\neq i}\sum_{l'\neq j} L_{ik}L_{jl}L_{ik'}L_{jl'}h^{-1}K_{ij}^2\phi_{ij}^2\right] \\
=&\frac{1}{g^2h(n-1)^4}\mathbb{E}\left[\sum_{k\neq i}\sum_{l\neq j}\sum_{k'\neq i}\sum_{l'\neq j} L_{ik}L_{jl}L_{ik'}L_{jl'}h^{-1}K_{ij}^2\phi_{ij}^2\right] \\
&+o\left(n^{-1}h^{-1/2}\right) \\
=&\frac{(n-1)^3}{(n-2)(n-3)(n-4)}\tilde{\omega}_n^2(\beta_0) + o\left(n^{-1}h^{-1/2}\right)
\end{aligned}$$

where

$$
\begin{aligned}
\tilde{\omega}_n^2\left(\beta_0\right) = \mathbb{E}\Bigg[\int\int\int\int & \frac{1}{g}L\left(\frac{z_i - z_k}{g}\right)\frac{1}{g}L\left(\frac{z_j - z_l}{g}\right)\frac{1}{g}L\left(\frac{z_i - z_{k'}}{g}\right)\frac{1}{g}L\left(\frac{z_j - z_{l'}}{g}\right) \\
& \times \frac{1}{h}K^2\left(\frac{z_i - z_j}{h}\right)\phi_{ij} \\
& \times f_{\beta_0}\left(z_k\right) f_{\beta_0}\left(z_l\right) f_{\beta_0}\left(z_{k'}\right) f_{\beta_0}\left(z_{l'}\right) \\
& \times \pi_{\beta_0}\left(z_i \mid W_i\left(\beta_0\right)\right)\pi_{\beta_0}\left(z_j \mid W_j\left(\beta_0\right)\right) dz_i dz_j dz_k dz_l dz_{k'} dz_{l'}\Bigg] \\
= \mathbb{E}\Bigg[\int\int\int\int & f_{\beta_0}\left(z_i - gs_1\right) f_{\beta_0}\left(z_i - gs_2\right) f_{\beta_0}\left(z_j - gt_1\right) f_{\beta_0}\left(z_j - gt_2\right) \\
& \times \pi_{\beta_0}\left(z_i \mid W_i\left(\beta_0\right)\right)\pi_{\beta_0}\left(z_j \mid W_j\left(\beta_0\right)\right)\phi_{ij} \\
& \times L\left(s_1\right) L\left(t_1\right) L\left(s_2\right) L\left(t_2\right)\frac{1}{h}K^2\left(\frac{z_i - z_j}{h}\right) dz_i dz_j ds_1 dt_1 ds_2 dt_2\Bigg] \\
= \mathbb{E}\Bigg[\int\int\int\int & f_{\beta_0}\left(z_i - gs_1\right) f_{\beta_0}\left(z_i - gs_2\right) f_{\beta_0}\left(z_i - gu - gt_1\right) f_{\beta_0}\left(z_i - gu - gt_2\right) \\
& \times \pi_{\beta_0}\left(z_i \mid W_i\left(\beta_0\right)\right)\pi_{\beta_0}\left(z_i - gu \mid W_j\left(\beta_0\right)\right)\phi_{ij} \\
& \times L\left(s_1\right) L\left(t_1\right) L\left(s_2\right) L\left(t_2\right) K^2\left(u\right) dz_i ds_1 dt_1 ds_2 dt_2 du\Bigg] \\
\to \mathbb{E}\Bigg[\int & f_{\beta_0}^4\left(z\right)\pi_{\beta_0}\left(z \mid W_i\left(\beta_0\right)\right)\pi_{\beta_0}\left(z \mid W_j\left(\beta_0\right)\right)\phi_{ij} dz\Bigg] \times \int K^2\left(u\right) du
\end{aligned}
$$

where the limit is obtained by standard arguments, using uniform continuity of $f_{\beta_0}\left(\cdot\right)$ and $\pi_{\beta_0}\left(\cdot \mid w\right)$. $\square$

References

Fan, Y. and Li, Q. (1996). Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms. *Econometrica* **64**, 865-90.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application*, New York: Academic Press Inc. [Harcourt Brace Jovanovich Publishers], probab. Math. Statist.

van der Vaart, A. W. (1998). *Asymptotic statistics*, vol. 3 of Camb. Ser. Stat. Probab. Math., Cambridge: Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (2011). A local maximal inequality under uniform entropy. *Electron. J. Stat.* **5**, 192-203.