SUPPLEMENT TO "TWO-STEP SEMIPARAMETRIC EMPIRICAL LIKELIHOOD INFERENCE"

By Francesco Bravo [‡] , Juan Carlos Escanciano [*,§] and Ingrid Van Keilegom [†,¶]

University of York [‡], *Universidad Carlos III de Madrid* [§] *and KU Leuven* [¶]

This supplement contains four appendices, and is organized as follows. Appendix A gathers all the proofs for the examples. Appendix B proves the validity of a general numerical algorithm for estimating the pathwise derivative, Appendix C extends the main result of the paper to the case of over-identified models, and Appendix D shows an auxiliary result regarding Donsker and Glivenko-Cantelli classes.

A. Proofs of the Propositions.

Proof of Proposition E1. Assumption A(i) follows by standard empirical process arguments after noticing that the class $\mathcal{F} = \{g(\cdot,\theta_0,h) + \phi(\cdot,\theta_0,h) : h \in \mathcal{H}\}$ is the sum of two $\mathbb{P}-$Donsker classes. Define $\delta_0(x) = f_{X_2}(x_2)/f_X(x)$ and note

$$\begin{aligned}M(h) &= \mathbb{E}\left[\theta_0 - \int_0^M \eta(w, X_2)\, dw - (Y - \eta(X))\delta(X)\right] \\ &= \mathbb{E}\left[\theta_0 - \eta(X)\,\delta_0(X) - (Y - \eta(X))\delta(X)\right]\end{aligned}$$

where $\delta(x) = \int_0^M h_2(w, x_2)\, dw/h_2(x)$ and $h = (h_1, h_2)$. Simple algebra then shows

$$M(h) - M(h_0) = \mathbb{E}\left[(\eta_0(X) - \eta(X))(\delta_0(X) - \delta(X))\right].$$

Then, Assumption A(ii) follows from Assumption E1 and standard results in kernel estimation. Assumption A(iv) also holds since the class $\mathcal{M} := \{m(\cdot,\theta_0,h) : h \in \mathcal{H}^\delta\}$ is $\mathbb{P}$-Glivenko-Cantelli with an envelope function $\theta_0 + C$, where C is a positive constant. Therefore, the class $\mathcal{M}^2 := \{m(\cdot,\theta_0,h)^2 : h \in \mathcal{H}^\delta\}$ is $\mathbb{P}$-Glivenko-Cantelli.

[*]Juan Carlos Escanciano gratefully acknowledges support by the Ministerio Economia y Competitividad (Spain), ECO2017-86675-P & MDM 2014-0431, and by Comunidad de Madrid (Spain), MadEco-CM S2015/HUM-3444.

[†]Ingrid Van Keilegom acknowledges financial support from the European Research Council (2016-2021, Horizon 2020/ERC grant agreement No. 694409).

To verify A(v) first note that by the triangle inequality and for an arbitrarily small $\epsilon > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(MEL_n(\theta_0, \widehat{h}) = 0\right) \\
&= \mathbb{P}\left(m(Z_i, \theta_0, \widehat{h}) > 0 \text{ or } m(Z_i, \theta_0, \widehat{h}) < 0 \text{ for all } Z_i \in \mathcal{Z}\right) \\
&\leq \prod_{i=i}^{n} \mathbb{P}\Big(m(Z_i, \theta_0, h_0) > \epsilon\Big) \\
&\quad + \prod_{i=i}^{n} \mathbb{P}\Big(m\left(Z_i, \theta_0, h_0\right) < -\epsilon\Big) \\
&\quad + \mathbb{P}\left(\sup_{Z_i \in \mathcal{Z}} \left| m(Z_i, \theta_0, \widehat{h}) - m(Z_i, \theta_0, h_0)\right| > \epsilon \right) \to 0
\end{aligned}
$$

by the facts that $E\left(m\left(Z, \theta_0, h_0\right)\right) = 0$, $E\left(m\left(Z, \theta_0, h_0\right)^2\right) < \infty$ and

$$
\begin{aligned}
&\max_i \left| m\left(Z_i, \theta_0, \widehat{h}\right) - m\left(Z_i, \theta_0, h_0\right)\right| \\
&\leq \max_i \left| \int_0^M \left(\widehat{\eta} - \eta_0\right)(u, X_{2i}) du\right| + \max_i \left| \left(\widehat{\eta} - \eta_0\right)(X_i) \frac{\widehat{f}_W\left(X_{2i}\right)}{\widehat{f}_X\left(X_i\right)}\right| \\
&\quad + \max_i \left|Y_i - \eta_0(X_i)\right| \max_i \left| \frac{\widehat{f}_W\left(X_{2i}\right)}{\widehat{f}_X\left(X_i\right)} - \frac{f_W\left(X_{2i}\right)}{f_X\left(X_i\right)}\right| \\
&= o_{\mathbb{P}}(1),
\end{aligned}
$$

by a standard result on the uniform consistency of kernel estimators (see for example [46]). To show the second part of A(v) note that by the triangle inequality and Chebychev's inequality,

$$
\begin{aligned}
&\mathbb{P}\Big(\max_i \left| m\left(Z_i, \theta_0, \widehat{h}\right)\right| > 2\sqrt{n}\varepsilon\Big) \\
&\leq \mathbb{P}\Big(\max_i \left|m\left(Z_i, \theta_0, h_0\right)\right| > \sqrt{n}\varepsilon\Big) \\
&\quad + \sum_i \mathbb{P}\left(\left|\int_0^M \left(\widehat{\eta} - \eta_0\right)(u, X_{2i}) du - Y_i \left(\frac{\widehat{f}_{X_2}\left(X_{2i}\right)}{\widehat{f}_X\left(X_i\right)} - \frac{f_{X_2}\left(X_{2i}\right)}{f_X\left(X_i\right)}\right)\right.\right. \\
&\qquad \left.\left. + \left(\widehat{\eta} - \eta_0\right)(X_i) \frac{\widehat{f}_{X_2}\left(X_{2i}\right)}{\widehat{f}_X\left(X_i\right)} + \eta_0(X_i)\left(\frac{\widehat{f}_{X_2}\left(X_{2i}\right)}{\widehat{f}_X\left(X_i\right)} - \frac{f_{X_2}\left(X_{2i}\right)}{f_X\left(X_i\right)}\right)\right| > \sqrt{n}\varepsilon\right) \\
&\leq \mathbb{P}\Big(\max_i \left|m\left(Z_i, \theta_0, h_0\right)\right| > \sqrt{n}\varepsilon\Big) + \frac{4}{\varepsilon^2}\left[\mathbb{E}\left(\int_0^M \left(\widehat{\eta} - \eta_0\right)(u, X_2) du\right)^2\right.
\end{aligned}
$$

$$
+\mathbb{E}\left(\frac{(Y-\eta_0(X))}{\widehat{f}_X(X)\, f_X(X)}\left(\widehat{f}_{X_2}(X_2)\, f_X(X)-f_{X_2}(X_2)\,\widehat{f}_X(X)\right)\right)^2
$$

$$
\left. +\mathbb{E}\left((\widehat{\eta}-\eta_0)(X)\frac{\widehat{f}_{X_2}(X_2)}{\widehat{f}_X(X)}\right)^2\right]
$$

$$
=o(1),
$$

where the last equality follows from the Cauchy-Schwarz inequality combined with the L_2-consistency of kernel estimators ([4, 65]). This shows the part related to Theorem 3.1.

For Theorem 3.3 we need to verify Assumption C. Since m is linear in θ Assumptions C(i-ii) are trivially satisfied. Since m is bounded in a neighborhood of θ_0 and $\mathbb{E}\left[|f_{X_2}(X_2)/f_X(X)|^{2+\delta}\right]<\infty$ then Assumption C(iii) also holds. Finally, for Theorem 3.4 we proceed to verify Condition 1 in [2]. Note that the derivative of g with respect to θ_0 is one, and similarly the derivative of the conditional moment that defines the first-step is also non-singular. By

$$
\mathbb{E}\left[\theta_0-\int_0^M \eta(w,X_2)\,dw\right]=\mathbb{E}\left[\theta_0-\eta(X)\,\delta_0(X)\right], \tag{A.1}
$$

and since $\mathbb{E}\left[\delta_0^2(X)\right]<\infty$, the left hand side of (A.1) is a bounded linear functional of η. This verifies Condition 1 in [2] and the conclusions regarding Theorem 3.4 for this example. □

Proof of Proposition E2. We consider the classes of functions

$$
\mathcal{G}=\left\{z=(y,d,x')'\rightarrow l(z,\eta)=\frac{yd}{\eta(x)}-\frac{y(1-d)}{1-\eta(x)}:\eta\in\bar{\mathcal{C}}^q_{1,\varepsilon}(\mathcal{S}_X)\right\}
$$

and $\mathcal{H}=\bar{\mathcal{C}}^q_{1,\varepsilon}(\mathcal{S}_X)\times\mathcal{C}^q_M(\mathcal{S}_X)\times\mathcal{C}^q_M(\mathcal{S}_X)$, with $\|h\|_{\mathcal{H}}=\|\eta\|_\infty+\|\mu_0\|_\infty+\|\mu_1\|_\infty$, for $h=(\eta,\mu_0,\mu_1)\in\mathcal{H}$. We show first that $\mathcal{G}$ is $\mathbb{P}-$Donsker. Notice that, for $\varepsilon<\eta<1-\varepsilon$,

$$
\left|\frac{\partial l(z,\eta)}{\partial\eta}\right|=\left|\frac{-yd}{\eta^2}-\frac{y(1-d)}{(1-\eta)^2}\right|\leq\frac{2\,|y|}{\varepsilon^2}\equiv C(y).
$$

Then, for any η and η_1,

$$
|l(z,\eta)-l(z,\eta_1)|\leq C(y)\,\|\eta-\eta_1\|_\infty\,.
$$

We use brackets of the form $[l(z,\eta_j)-\delta C(y),l(z,\eta_j)+\delta C(y)]$, with η_j the center of balls of radius δ covering $\bar{\mathcal{C}}^q_{1,\varepsilon}(\mathcal{S}_X)$. These brackets have $||\cdot$

$||_2-$size of $2\delta||C(\cdot)||_2$. For any $\eta \in \bar{\mathcal{C}}^q_{1,\varepsilon}(\mathcal{S}_X)$ there exists $j \in \{1,...,N_\delta \equiv N(\delta, \bar{\mathcal{C}}^q_{1,\varepsilon}(\mathcal{S}_X), ||\cdot||_\infty)\}$ such that $||\eta - \eta_j||_\infty < \delta$. Moreover, by the previous display,

$$|l(z,\eta) - l(z,\eta_j)| \leq C(y)\delta.$$

Therefore,

$$N_{[\cdot]}(2\delta||C(\cdot)||_2, \mathcal{G}, \|\cdot\|_2) \leq N(\delta, \bar{\mathcal{C}}^q_{1,\varepsilon}(\mathcal{S}_X), \|\cdot\|_\infty),$$

and $\mathcal{G}$ is $\mathbb{P}-$Donsker provided $q > d_x/2$, see [72], p.154. Similarly, the class

$$\Phi := \left\{ z \to \phi(z,\theta_0,h) = (d - \eta(x)) \left(\frac{\mu_1(x)}{\eta(x)} + \frac{\mu_0(x)}{1-\eta(x)} \right) : h \in \mathcal{H} \right\}$$

is $\mathbb{P}-$Donsker. Then, Assumption A(i) holds.

To check for A(ii) we verify Assumption B. The moment is twice Hadamard differentiable with a bounded second derivative and a zero first derivative. The zero derivative condition follows from [51], p. 1361, and the fact that the derivative with respect to μ_1 and μ_0 of $h \to \mathbb{E}[\phi(Z,\theta_0,h)]$ at h_0 is zero by the conditional restriction

$$\mathbb{E}[D - \eta_0(X)|X] = 0 \ a.s.$$

To see that the second derivative is bounded, we use that the propensity score is bounded away from zero and one to conclude that

$$\left| \frac{\partial^2 l(z,\theta_0,\eta)}{\partial \eta^2} \right| = \left| \frac{-2yd}{\eta^3} + \frac{2y(1-d)}{(1-\eta)^3} \right| \leq \frac{4|y|}{\varepsilon^3} \equiv C_2(y),$$

and similarly for $\partial^2\phi(z,\theta_0,h)/\partial h^2$. The rate condition for Assumption B(ii) follows from Assumption E1 and standard results in kernel estimation. Finally, the verification of Assumption A(iv) follows from the class

$$\mathcal{F} := \{m(\cdot,\theta_0,h) : h \in \mathcal{H}\},$$

being $\mathbb{P}-$Donsker, and hence, $\mathbb{P}-$Glivenko-Cantelli, and the fact that $\mathcal{F}$ has a square integrable envelope function by E2(i). To check A(v) the same arguments as those used in the proof of Proposition E1 show first that $\mathbb{P}(MEL_n(\theta_0,\widehat{h}) = 0) \to 0$ since $\max_i |\widehat{\eta}_i - \eta_0| = o_{\mathbb{P}}(1)$ and $\max_i |\widehat{\iota}_i - \iota_0| = o_{\mathbb{P}}(1)$ where $\widehat{\eta}_i = \widehat{\eta}(X_i)$, $\widehat{\iota}_i = \widehat{\mu}_1(X_i)/\widehat{\eta}_i + \widehat{\mu}_0(X_i)/(1-\widehat{\eta}_i)$. To verify the second claim note that, for any $\delta > 0$,

$$\mathbb{P}\Big(\max_i \left| m\left(Z_i,\theta_0,\widehat{h}\right) \right| > 2\sqrt{n}\delta \Big)$$

$$\le \mathbb{P}\Big(\max_i |m(Z_i,\theta_0,h_0)| > \sqrt{n}\delta\Big)$$
$$+\frac{5}{\delta^2}\left[\mathbb{E}\left(\frac{YD}{\widehat{\eta}\eta_0}(\widehat{\eta}-\eta_0)\right)^2 + \mathbb{E}\left(\frac{Y(1-D)}{(1-\eta_0)(1-\widehat{\eta})}(\widehat{\eta}-\eta_0)\right)^2\right.$$
$$\left.+\mathbb{E}\left(D(\widehat{\iota}-\iota_0)\right)^2 + \mathbb{E}\left(\widehat{\eta}(\widehat{\iota}-\iota_0)\right)^2 + \mathbb{E}\left(\iota_0(\widehat{\eta}-\eta_0)\right)^2\right].$$

The rest follows as in the proof of Proposition E1. □

Proof of Proposition E3. By same arguments used in Proposition E2, it is shown that the class of functions $\mathcal{F} := \{m(\cdot,\theta_0,h) : h \in \mathcal{H}^\delta\}$ is $\mathbb{P}-$Donsker. Then, Assumption A(i) holds. To check A(ii) we verify Assumption B. The moment is twice pathwise differentiable at $h = h_0$ with uniformly bounded second derivatives, as $p_0 \in \mathcal{C}^q_{1,\varepsilon}(\mathcal{S}_X)$. The rate condition for Assumption B(ii) follows from Assumption E1 and standard results in kernel estimation. The zero derivative condition follows from [51], p. 1361. That is, the pathwise derivative with respect to q_0 is clearly zero, since

$$\mathbb{E}\left[\left(1-\frac{D}{p_0(X)}\right)\middle| X\right] = 0 \ a.s.$$

The derivative with respect to p_0 is more involved, but it is equally zero by the last display and the condition

$$\mathbb{E}\left[\frac{D}{p_0^2(X)}\left(s(X,W,\theta_0) - \frac{q_0(X,\theta_0)}{p_0(X)}\right)\middle| X\right] = 0 \ a.s.,$$

which holds by the conditional independence assumption. Finally, Assumption A(iv) follows because the class $\mathcal{F}$ is $\mathbb{P}-$Donsker. To check A(v) the first part follows as in the proof of Proposition E1 since $\max_i |\widehat{p}(X_i) - p_0(X_i)| = o_{\mathbb{P}}(1)$ and $\max_i |\widehat{q}(X_i,\theta_0) - q_0(X_i,\theta_0)| = o_{\mathbb{P}}(1)$. For the second part, note that for any $\delta > 0$,

$$\mathbb{P}\Big(\max_i \left|m\left(Z_i,\theta_0,\widehat{h}\right)\right| > 2\sqrt{n}\delta\Big)$$
$$\le \mathbb{P}\Big(\max_i |m(Z_i,\theta_0,h_0)| > \sqrt{n}\delta\Big)$$
$$+\frac{3}{\delta^2}\mathbb{E}\left[\frac{Ds(X,W,\theta_0)}{\widehat{p}(X)p_0(X)}(p_0(X)-\widehat{p}(X))\right]^2$$
$$+\frac{3}{\delta^2}\mathbb{E}\left[\left(1-\frac{D}{\widehat{p}(X)}\right)\frac{\widehat{q}(X,\theta_0)-q_0(X,\theta_0)}{\widehat{p}(X)}\right]^2$$
$$+\frac{3}{\delta^2}\mathbb{E}\left[\left(1-\frac{D}{\widehat{p}(X)}\right)\frac{q_0(X,\theta_0)}{\widehat{p}(X)p_0(X)}(\widehat{p}(X)-p_0(X))\right]^2 \to 0,$$

as in the proof of the second part of Proposition E1. □

Proof of Proposition E4. We start by verifying assumption A(i). Lemma 6.1 in [45] shows that the class $\mathcal{G}$ satisfies for any $\epsilon > 0$,

$$\log N_{[\cdot]}(\epsilon, \mathcal{G}, \|\cdot\|_{\mathcal{H}}) = O(\epsilon^{-2/(1+\delta)}),$$

and hence we also have that $\log N_{[\cdot]}(\epsilon, \mathcal{H}, \|\cdot\|_{\mathcal{H}}) = O(\epsilon^{-2/(1+\delta)})$. Define $\mathcal{M} = \{z \to m(z,\theta_0,h) : h \in \mathcal{H}\}$. Since the map $h \to m(z,\theta_0,h)$ is uniformly Lipschitz continuous (in the sense that $|m(z,\theta_0,h) - m(z,\theta_0,\tilde{h})| \le c(z)\|h - \tilde{h}\|_{\mathcal{H}}$, with $\mathbb{E}(c^2(Z)) < \infty$), it easily follows (as in the proof of Theorem 3 in [14]) that

$$\log N_{[\cdot]}(\epsilon, \mathcal{M}, \|\cdot\|_2) = O(\epsilon^{-2/(1+\delta)}),$$

and hence $\mathcal{M}$ is $\mathbb{P}$-Donsker, which implies A(i).

For Assumption A(ii), we use a direct approach. Using the fact that $\mathbb{E}(m(Z,\theta_0,h_0)) = 0$, straightforward calculations show that

$$\begin{aligned}
&\mathbb{E}(m(Z,\theta_0,\widehat{h}))\\
&= \mathbb{E}\Big[X\Big\{-\frac{1-\tau}{\eta_0(X'\theta_0|X)}\big(\widehat{\eta}(X'\theta_0|X) - \eta_0(X'\theta_0|X)\big)\\
&\quad - (1-F(X'\theta_0|X))\Big(\int_{-\infty}^{X'\theta_0}\frac{\widehat{H}(s|X) - H(s|X)}{(1-H(s|X))^2}dH_c(s|X)\\
&\qquad + \frac{\widehat{H}_c(X'\theta_0|X) - H_c(X'\theta_0|X)}{1-H(X'\theta_0|X)} - \int_{-\infty}^{X'\theta_0}\frac{\widehat{H}_c(s|X) - H_c(s|X)}{(1-H(s|X))^2}dH(s|X)\Big)\Big\}\Big].
\end{aligned}$$

Using the i.i.d. representation of $\widehat{\eta}(X'\theta_0|X) - \eta_0(X'\theta_0|X)$ given in (4.4), the latter can be written as

$$\begin{aligned}
&n^{-1}\sum_{i=1}^n \mathbb{E}\Big[X\frac{1-\tau}{f_X(X)}k_b(X_i - X)\xi(Y_i,\Delta_i,X'\theta_0|X)\Big]\\
&- (1-\tau)\sum_{i=1}^n \mathbb{E}\Big[XW_i(X,b_n)\Big(\int_{-\infty}^{X'\theta_0}\frac{I(Y_i \le s) - H(s|X)}{(1-H(s|X))^2}dH_c(s|X)\\
&\quad + \frac{I(Y_i \le X'\theta_0, \Delta_i = 1) - H_c(X'\theta_0|X)}{1-H(X'\theta_0|X)}\\
&\quad - \int_{-\infty}^{X'\theta_0}\frac{I(Y_i \le s, \Delta_i = 1) - H_c(s|X)}{(1-H(s|X))^2}dH(s|X)\Big)\Big]\\
&+ o_{\mathbb{P}}(n^{-1/2})\\
&= n^{-1}\sum_{i=1}^n \mathbb{E}\Big[X(1-\tau)k_b(X_i - X)\Big\{\frac{1}{f_X(X)} - \frac{1}{\widehat{f}_X(X)}\Big\}\xi(Y_i,\Delta_i,X'\theta_0|X)\Big]
\end{aligned}$$

$$
\begin{aligned}
&+ o_{\mathbb{P}}(n^{-1/2}) \\
&= o_{\mathbb{P}}(n^{-1/2}),
\end{aligned}
$$

with $\widehat{f}_X(x) = n^{-1}\sum_{i=1}^n k_b(X_i - x)$, since $\sup_x |\widehat{f}_X(x) - f_X(x)| = O_{\mathbb{P}}((nb_n)^{-1/2} (\log n)^{1/2} + b_n^r)$ and $\sup_x |n^{-1}\sum_{i=1}^n k_b(X_i - x)\xi(Y_i, \Delta_i, x'\theta_0|x)| = O_{\mathbb{P}}((nb_n)^{-1/2} (\log n)^{1/2} + b_n^r)$, which are $o_{\mathbb{P}}(n^{-1/4})$ provided $nb_n^{4r} \to 0$ and $nb_n^2(\log n)^{-2} \to \infty$. Hence, A(ii) follows.

We next consider A(iii). It follows from Proposition 1 and 2 in [3] and and Lemma 6.2 in [45] that $\mathbb{P}(\widehat{h} \in \mathcal{H}^\delta) \to 1$. Moreover, $\|\widehat{h} - h_0\|_{\mathcal{H}} = o_{\mathbb{P}}(1)$, because of the uniform consistency of the Nadaraya-Watson estimator and of the Beran estimator ([6]). Finally, A(iv) follows using similar arguments as for A(i), and A(v) can be shown in the same way as for the previous examples. □

B. Numerical computation of pathwise derivatives. It is useful to have a general and practical way to compute and estimate ϕ. [36] have recently proposed a simple method to compute the influence function of semiparametric estimators. This method when applied to the functional $\mu(F)$ in (2.10) can be used to compute ϕ. This approach is a smoothed version of Hampel's ([27, 28]) characterization of influence functions as Gateaux derivatives, i.e.

$$
\phi(z, \theta_0, \eta(F_0)) = \left.\frac{\partial \mu((1-t)F_0 + t\delta_z)}{\partial t}\right|_{t=0}, \tag{B.1}
$$

where δ_z is the Dirac measure at z. The problem with the calculation in (B.1) is that generally $(1-t)F_0 + t\delta_z$ will not be in the domain of $\mu(\cdot)$ for many examples of interest. [36] have addressed this limitation by proposing a smoothed version of the Dirac measure, i.e. using the paths of the form $F_{tz}^b = (1-t)F_0 + tG_{\tilde{z}}^b$, where G_z^b approaches a point-mass at z as $b \downarrow 0$. The leading example of G_z^b is a distribution with density function $g_z^b(\tilde{z}) = K_b(\tilde{z} - z)$, for $K_b(z) := b^{-d_z}\prod_{l=1}^{d_z} k(z_l/b)$, some univariate bounded kernel function $k(\cdot)$ with compact support, and a bandwidth parameter $b \downarrow 0$. The paths $\{F_{tz}^b\}$ will be in the domain of $\mu(\cdot)$ for many examples of interest. These paths have an associated score given by $s(\tilde{z}) = \left[g_z^b(\tilde{z})/f_0(\tilde{z}) - 1\right]$, where $f_0(z)$ is the Lebesgue density of the data. Then, by (2.11) and $\phi(\cdot, \theta_0, \eta(F_0)) \in L_2^0$,

$$
\begin{aligned}
\left.\frac{\partial \mu(F_{tz}^b)}{\partial t}\right|_{t=0} &= \int \phi(\tilde{z}, \theta_0, \eta(F_0)) \left[\frac{g_z^b(\tilde{z})}{f_0(\tilde{z})} - 1\right] f_0(\tilde{z}) d\tilde{z} \\
&= \int \phi(\tilde{z}, \theta_0, \eta(F_0)) g_z^b(\tilde{z}) d\tilde{z}, \qquad \text{(B.2)}
\end{aligned}
$$

which converges to $\phi(z, \theta_0, \eta(F_0))$ if this function is continuous in z, by dominated convergence. Thus, this discussion leads to the formula (see Theorem 1 in [36])

$$\phi(z, \theta_0, \eta(F_0)) = \lim_{b \to 0} \frac{\partial \mu(F^b_{tz})}{\partial t}\bigg|_{t=0}. \tag{B.3}$$

We now use the sample analog of (B.3) to construct an estimator for the pathwise derivative ϕ. We first replace F^b_{tz} by an estimated and uncentered path (a different centering will turn out to be convenient for estimation)

$$\hat{F}^b_{tz} = \hat{F} + tG^b_z,$$

where $\hat{F}$ is a nonparametric estimator of F_0, such as

$$\hat{F}(z) = \int_{(-\infty, z]} \hat{f}(u)du,$$

with $\hat{f}$ a kernel estimator of f_0. We also replace F_0 in the functional μ by the empirical distribution F_n to construct the estimator (cf. (B.3))

$$\hat{\delta}(z) := \frac{\partial \mu_n(\hat{F}^{b_n}_{tz})}{\partial t}\bigg|_{t=0}$$

if the derivative of g is analytical, or

$$\hat{\delta}(z) := \frac{\mu_n(\hat{F}^{b_n}_{t_n z}) - \mu_n(\hat{F})}{t_n},$$

if a numerical derivative is considered, where $\mu_n(F)$ is the empirical expectation operator

$$\mu_n(F) := \frac{1}{n}\sum_{j=1}^{n} g(Z_j, \theta_0, \eta(F)),$$

and $b \equiv b_n$ and $t \equiv t_n$ are positive sequences converging to zero at suitable rates (see Assumption C below). In any case, we estimate $\phi(Z_i, \theta_0, \eta(F_0))$ by

$$\hat{\phi}_i = \widehat{\delta}_i - \frac{1}{n}\sum_{j=1}^{n} \widehat{\delta}_j, \tag{B.4}$$

where, henceforth, for a generic measurable function $\delta(Z)$ of Z, $\delta_i = \delta(Z_i)$. This approach only uses the function g and the estimator $\hat{F}$. It does require knowledge on the specific structure of the problem. However, it involves smoothing on all components of Z in the estimation of ϕ, which might be suboptimal. For this reason, we consider a more specialized setting well suited for cases where the nuisance parameter η_0 is given by

$$\eta_0(x,\theta_0) = \mathbb{E}\left[V(Z,\theta_0)|X=x\right], \tag{B.5}$$

where $V \equiv V(Z,\theta_0)$ is a known vector of measurable functions of the data Z and the parameter θ_0. Many examples considered in applications have this structure.

For problems involving projections as above, the estimation of the influence function can be done using paths $\left\{F^b_{tz}\right\}$ such that $\eta^b_{tz} = \eta(F^b_{tz})$ is given by

$$\eta^b_{tz}(\bar{x}) = \frac{\eta^b_{2tz}(\bar{x})}{\eta^b_{1tz}(\bar{x})}, \tag{B.6}$$

where

$$\eta^b_{2tz}(\bar{x}) = \eta_{20}(\bar{x}) + tV(z,\theta_0)K_b\left(\bar{x}-x\right), \qquad \eta^b_{1tz}(\bar{x}) = f_X(\bar{x}) + tK_b\left(\bar{x}-x\right),$$

$\eta_{20}(x) = \eta_0(x,\theta_0)f_X\left(x\right)$ and $f_X\left(x\right)$ is the density of X. The sample analog of the path is given by

$$\widehat{\eta}^b_{ti}(\bar{x}) = \frac{\widehat{\eta}^b_{2ti}(\bar{x})}{\widehat{\eta}^b_{1ti}(\bar{x})},$$

where

$$\widehat{\eta}^b_{2ti}(\bar{x}) = \widehat{\eta}_2(\bar{x}) + tV_iK_b\left(\bar{x}-X_i\right), \qquad \widehat{\eta}^b_{1ti}(\bar{x}) = \widehat{f}_X(\bar{x}) + tK_b\left(\bar{x}-X_i\right),$$

$\widehat{\eta}_2(x) = n^{-1}\sum_{i=1}^n V_iK_b\left(X_i-x\right)$, $V_i \equiv V(Z_i,\theta_0)$, $\widehat{f}_X\left(x\right) = n^{-1}\sum_{i=1}^n K_b\left(X_i-x\right)$, and $\widehat{\eta} = \widehat{\eta}_2/\widehat{f}_X$.

Then, similarly as above define

$$\widehat{\delta}_i := \frac{1}{n}\sum_{j=1}^n \left.\frac{\partial g\left(Z_j,\theta_0,\widehat{\eta}^b_{ti}\right)}{\partial t}\right|_{t=0}, \tag{B.7}$$

if the derivative of g is analytical, or numerically

$$\widehat{\delta}_i := \frac{1}{nt_n}\sum_{j=1}^n \left[g\left(Z_j,\theta_0,\widehat{\eta}^{b_n}_{ti}\right) - g\left(Z_j,\theta_0,\widehat{\eta}\right)\right]. \tag{B.8}$$

Then, proceed as in (B.4). The estimator of the influence function using the analytical derivative in (B.7) and with V independent of θ_0 was proposed in [51] for the purpose of consistent asymptotic variance estimation. In this section, we build on [51] and obtain asymptotic pivotal EL ratios, for analytical and numerical derivatives and for moments that are non-smooth as functions of nuisance parameters.[1]

The automatic modification uses the estimated modified moments

$$\hat{m}_i := \hat{g}_i + \hat{\phi}_i,$$

where the dependence on θ_0 is dropped for simplicity, $\hat{g}_i = g\left(Z_i, \theta_0, \widehat{\eta}\right)$ and $\hat{\phi}_i$ is given by (B.4). Then, for the validity of our results we first need to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{m}_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_i + o_{\mathbb{P}}(1), \tag{B.9}$$

where $m_i = m\left(Z_i, \theta_0, h_0\right)$. In view of (2.5), a sufficient condition for (B.9) is

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\phi}_i \xrightarrow{\mathbb{P}} 0, \tag{B.10}$$

which is trivially satistied because $\hat{\phi}_i$ is centered. Then, our results will hold with the estimated influence function if we show

$$\hat{\Sigma} \equiv \frac{1}{n} \sum_{i=1}^{n} \hat{m}_i \hat{m}_i' \xrightarrow{\mathbb{P}} \Sigma. \tag{B.11}$$

For moments that are smoothed in nuisance parameters, transformations $V_i \equiv V(Z_i, \theta_0)$ that do not depend on θ_0 and analytical derivatives, ([51], Lemma 5.5) has found sufficient conditions for (B.11). We aim to find similar sufficient conditions for our more general setting with possibly non-smooth moments g, general transformations $V(Z_i, \theta_0)$ and numerical derivatives. Non-smooth moments appear in e.g. quantile regression applications. Transformations $V(Z_i, \theta_0)$ that depend on θ_0 appear in many applications, such as the general missing data problem considered in the main text.

The following result provides sufficient conditions for (B.11) in our general setting with nuisance parameters defined by (B.5) for non-smooth moments and numerical derivatives (the case of analytical derivatives is simpler). In this section, $h_0 = (\eta_0, f_X) \in \mathcal{H} := \mathcal{C}_1^q\left(\mathcal{S}_X\right) \times \mathcal{C}_{M,\varepsilon}^q\left(\mathcal{S}_X\right)$, $q \geq 1$, where

[1] Our results on influence function estimation are also complementary to [50, 1, 2, 32], who proposed analytical and numerical derivatives estimates, respectively, for series (sieves) first-steps.

henceforth we assume $\mathcal{S}_X$ is compact, convex and has non-empty interior. Define $\|h_0\|_{\mathcal{H}} = \|\eta_0\|_\infty + \|f_X\|_\infty$. A generic element of $\mathcal{H}$ is denoted by $h = (\eta, f)$. Theorem A3, which is used in the following assumption, is given in Section D.

Assumption D:

(i) $m_n(z, v, h) = t_n^{-1}\left\{g\left(z, \theta_0, \eta_{t,v}^b(h)\right) - g(z, \theta_0, h)\right\}$ satisfies the conditions of Theorem A3(i) with $\rho > 1$, $h = (\eta, f) \in \mathcal{H}$, and $v = (y, x) \in \mathcal{V} \equiv \mathcal{S}_Z$, some semi-metric $|\cdot|_{\mathcal{V}}$ and where

$$\eta_{t,v}^b(h)(\bar{x}) = \frac{\eta(\bar{x})f(\bar{x}) + tV(v, \theta_0)K_b(\bar{x} - x)}{f(\bar{x}) + tK_b(\bar{x} - x)}.$$

(ii) The map $h \to \mu(h) = \mathbb{E}[g(Z, \theta_0, h)]$ from $\mathcal{H}^\delta$ to $\mathbb{R}^p$ is directionally differentiable at $h = h_1 \in \mathcal{H}^{\delta_n}$ with derivative $V_h^\mu(h_1)[h - h_1]$ in all directions $[h - h_1] \in \mathcal{H}^\delta$; and for all $h \in \mathcal{H}^{\delta_n}$ with $\delta_n \downarrow 0$, it holds that

$$\left|\mu(h) - \mu(h_1) - V_h^\mu(h_1)[h - h_1]\right| \le c\|h - h_1\|_{\mathcal{H}}^2$$

for a constant $c \ge 0$. Furthermore, $V_h^\mu(h_0)[h] = \mathbb{E}[\psi(X)h(X)]$, where $\equiv \psi_{h_0}$ is Hölder continuous of exponent $\alpha_\psi > 0$, $\mathbb{E}\left[|\psi(X)|^2\right] < \infty$ and

$$\left|V_h^\mu(h_0)[h] - V_h^\mu(h_1)[h]\right| \le C\|h_0 - h_1\|_{\mathcal{H}}\|h\|_{\mathcal{H}}.$$

(iii) The kernel function $k(t) : \mathbb{R} \to [0, \infty)$ is bounded, symmetric, and integrates up to one. Furthermore, $\int |tk(t)|\,dt < \infty$.

(iv) The deterministic sequence of positive numbers $b \equiv b_n$ and $t \equiv t_n$ satisfy: (a) $b_n \to 0$ and $b_n^{2d_x}n/\log n \to \infty$; and (b) $t_n \to 0$, $ct_n^{-1}\|\widehat{h} - h_0\|_{\mathcal{H}}^2 = o_{\mathbb{P}}(1)$ and $t_nb_n^{-d_x} \to 0$.

(v) $\mathbb{E}\left[|\psi_{h_0}(X)(V(Z, \theta_0) - \eta_0(X))|^2\right] < \infty$, and $\mathbb{E}[|V(Z, \theta_0)|| X = x]$ and $\mathbb{E}\left[|V(Z, \theta_0)|^2 \middle| X = x\right]$ are continuous in $x \in \mathcal{S}_X$.

When $g(z, \theta_0, h)$ is smooth in h, one can easily find primitive conditions for Theorem A3(i) to hold, as required in Assumption D(i). Note that, more generally, Theorem A3(i) allows for non-smooth moment functions of first-steps. Assumption D(ii) is similar to B(i), but with an explicit representation for the linearization. From [50], the pathwise derivative is given by $\phi(z, \theta_0, h_0) = \psi_{h_0}(x)(V(z, \theta_0) - \eta_0(x))$. The assumptions on the kernel and bandwidth are standard.

THEOREM B.1. *Assumptions A and D are sufficient for (B.11).*

Note that although we have used kernel first-steps $\widehat{h} = (\widehat{\eta}, \widehat{f}_X)$ to estimate the pathwise derivative ϕ_i, other nonparametric smoothing techniques could be used as well. The proof of Theorem B.1 does not depend on the specific first-step used, but of course the precise regularity conditions (i.e. primitive conditions for Assumptions A and D) will in that case need to be adapted to the particular smoothing method at hand.

Proof of Theorem B.1. Assume $p = 1$ for notational simplicity. Write

$$\begin{aligned}
\hat{\Sigma} - \Sigma &= \frac{1}{n}\sum_{i=1}^{n}\left\{\left(\hat{g}_i + \widehat{\delta}_i - \frac{1}{n}\sum_{j=1}^{n}\widehat{\delta}_j\right)^2 - \left(\hat{g}_i + \widetilde{\delta}_i - \frac{1}{n}\sum_{j=1}^{n}\widetilde{\delta}_j\right)^2\right\} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{\left(\hat{g}_i + \widetilde{\delta}_i - \frac{1}{n}\sum_{j=1}^{n}\widetilde{\delta}_j\right)^2 - \left(\hat{g}_i + \widetilde{\delta}_i\right)^2\right\} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{\left(\hat{g}_i + \widetilde{\delta}_i\right)^2 - \left(\hat{g}_i + \delta_i\right)^2\right\} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{\left(\hat{g}_i + \delta_i\right)^2 - \left(g_i + \delta_i\right)^2\right\} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{\left(g_i + \delta_i\right)^2 - \Sigma\right\} \\
&\equiv T_1 + T_2 + T_3 + T_4 + T_5,
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{\delta}_i &= t_n^{-1}\mathbb{E}\left[g\left(Z, \theta_0, \widehat{\eta}_{ti}^{b_n}\right) - g\left(Z, \theta_0, \widehat{\eta}\right)\right]; \\
\delta_i &= t_n^{-1}\mathbb{E}\left[g\left(Z, \theta_0, \eta_{ti}^{b_n}\right) - g\left(Z, \theta_0, \eta_0\right)\right]; \\
\hat{g}_i &= g\left(Z_i, \theta_0, \widehat{\eta}\right); g_i = g\left(Z_i, \theta_0, \eta_0\right);
\end{aligned}$$

$$\eta_{t,z}^{b_n}(h)(\bar{x}) = \frac{\eta_{2t,z}^{b_n}}{\eta_{1t,z}^{b_n}}(\bar{x}),\ \ z = (y, x);\ \ \widehat{\eta}_{ti}^{b_n} = \eta_{t,Z_i}^{b_n}(\widehat{h});\ \ \eta_{ti}^{b_n} = \eta_{t,Z_i}^{b_n}(h_0);$$

$$\eta_{2t,z}^{b_n}(\bar{x}) = \eta(\bar{x})f(\bar{x}) + tV(z, \theta_0)K_b\left(\bar{x} - x\right); \eta_{1ti}^{b_n}(\bar{x}) = f(\bar{x}) + tK_b\left(\bar{x} - x\right),$$

and Z is a random vector with distribution F_0 and independent of the sample $\{Z_i\}_{i=1}^{n}$.

We show that each term $T_j = o_{\mathbb{P}}(1)$ for $j = 1, ..., 5$. Define the class of functions

$$\mathcal{T}_1 = \left\{z \to t_n^{-1}\left\{g\left(z, \theta_0, \eta_{t,v}^{b}(h)\right) - g\left(z, \theta_0, h\right)\right\} : h = (\eta, f) \in \mathcal{H}, v = (y, x) \in \mathcal{S}_Z\right\}.$$

By Assumption D(i) and Theorem A1(i), $\mathcal{T}_1$ is Glivenko-Cantelli. Thus,

$$\sup_{h\in\mathcal{H},v\in\mathcal{Z}}\left|\frac{1}{nt_n}\sum_{i=1}^{n}\left\{\left(g\left(Z_i,\theta_0,\eta_{t,v}^{b}(h)\right)-g\left(Z_i,\theta_0,h\right)\right)\right.\right.$$
$$\left.\left.-\mathbb{E}\left[g\left(Z,\theta_0,\eta_{t,v}^{b}(h)\right)-g\left(Z,\theta_0,h\right)\right]\right\}\right|=o_{\mathbb{P}}(1),$$

and hence, by Assumption A(iii)

$$\max_{1\leq i\leq n}\left|\widehat{\delta}_i-\widetilde{\delta}_i\right|=o_{\mathbb{P}}(1).$$

Then,

$$|T_1|\leq\left(\frac{1}{n}\sum_{i=1}^{n}\left|2\hat{g}_i+\widehat{\delta}_i+\widetilde{\delta}_i-\frac{1}{n}\sum_{j=1}^{n}\widehat{\delta}_j-\frac{1}{n}\sum_{j=1}^{n}\widetilde{\delta}_j\right|\right)\times 2\max_{1\leq i\leq n}\left|\widehat{\delta}_i-\widetilde{\delta}_i\right|$$
$$=O_{\mathbb{P}}(1)o_{\mathbb{P}}(1)=o_{\mathbb{P}}(1),$$

where the $O_{\mathbb{P}}(1)$ follows from our results below.

We now show that $\left|n^{-1}\sum_{j=1}^{n}\widetilde{\delta}_j\right|=o_{\mathbb{P}}(1)$. In turn, this will follow if we show that

$$\left|\frac{1}{n}\sum_{j=1}^{n}\widetilde{\delta}_j-\delta_j\right|=o_{\mathbb{P}}(1)\tag{B.12}$$

and

$$\left|\frac{1}{n}\sum_{j=1}^{n}\delta_j\right|=o_{\mathbb{P}}(1).\tag{B.13}$$

Use D(ii) and D(iv) to show that the left hand side of (B.12) is bounded by

$$\left|\frac{1}{n}\sum_{j=1}^{n}V_h^{\mu}\left(\eta_{ti}^{b_n}\right)\left[\frac{\widehat{\eta}_{ti}^{b_n}-\eta_{ti}^{b_n}}{t_n}\right]-V_h^{\mu}\left(\eta_0\right)\left[\frac{\widehat{\eta}-\eta_0}{t_n}\right]\right|+O_{\mathbb{P}}\left(ct_n^{-1}\|\widehat{h}-h_0\|_{\mathcal{H}}^2\right)$$
$$\leq\left|\frac{1}{n}\sum_{j=1}^{n}V_h^{\mu}\left(\eta_0\right)\left[\frac{\widehat{\eta}_{ti}^{b_n}-\widehat{\eta}}{t_n}-\frac{\eta_{ti}^{b_n}-\eta_0}{t_n}\right]\right|$$
$$+\left|\frac{1}{n}\sum_{j=1}^{n}V_h^{\mu}\left(\eta_{ti}^{b_n}\right)\left[\frac{\widehat{\eta}_{ti}^{b_n}-\eta_{ti}^{b_n}}{t_n}\right]-V_h^{\mu}\left(\eta_0\right)\left[\frac{\widehat{\eta}_{ti}^{b_n}-\eta_{ti}^{b_n}}{t_n}\right]\right|+o_{\mathbb{P}}\left(1\right)$$

(B.14)

$$\equiv A_1 + A_2 + o_{\mathbb{P}}(1).$$

Define the class of functions

$$\mathcal{T}_2 = \left\{ z \to V_h^{\mu}(\eta_0) \left[\frac{\eta_{t,z}^b(h) - \eta}{t_n} \right] : h \in \mathcal{H} \right\}.$$

We show that $\mathcal{T}_2$ is Glivenko-Cantelli. We shall apply Theorem A3(i) with $m_n(z, v, h) \equiv m_n(z, h) =: t_n^{-1} V_h^{\mu}(\eta_0) \left[\eta_{t,z}^b(h) - \eta\right], h \in \mathcal{H}$. We proceed to verify (D.1). Define $\Delta_{t,z}^b(h)(\bar{x}) = \left(\eta_{t,z}^b(h) - \eta\right)(\bar{x})/t$, and note that

$$\Delta_{t,z}^b(h)(\bar{x}) = \frac{1}{\eta_{1t,z}^b(\bar{x})} \left[V(z,\theta_0) - \eta(\bar{x})\right] K_b(\bar{x} - x).$$

By the representation of V_h^{μ} in D(ii), we can write, for each $z \in \mathcal{S}_Z$ and $h \in \mathcal{H}$,

$$\begin{aligned}
&V_h^{\mu}(h)\left[\Delta_{t,z}^b(h)\right] \\
&= \mathbb{E}\left[\psi(X)\frac{1}{\eta_{1t,z}^{b_n}(X)}\left[V(z,\theta_0) - \eta(X)\right] K_b(X - x)\right] \\
&= \mathbb{E}\left[\psi(X)\frac{1}{f_X(X)}\left[V(z,\theta_0) - \eta(X)\right] K_b(X - x)\right] \\
&\quad + \mathbb{E}\left[\psi(X)\frac{f_X(X) - \eta_{1t,z}^{b_n}(X)}{f_X(X)\eta_{1t,z}^{b_n}(X)}\left[V(z,\theta_0) - \eta(X)\right] K_b(X - x)\right]
\end{aligned}$$

(B.15)

$$=: \psi(x)\left[V(z,\theta_0) - \eta(x)\right] + \psi(x)\frac{f_X(x) - f(x)}{f(x)}\left[V(z,\theta_0) - \eta(x)\right] + R_t^b(x),$$

where

$$R_t^b(x) = O(b^{\alpha_\psi} + |\psi(x)|\, b + tb^{-d_x}\, |a(x)|),$$

as it follows from a change of variables argument applied to $a(\bar{x}) = \psi(\bar{x})[V(z,\theta_0) - \eta(\bar{x})]$, i.e.

$$\begin{aligned}
&\mathbb{E}\left[\psi(X)\frac{1}{f_X(X)}\left[V(z,\theta_0) - \eta(X)\right] K_b(X - x)\right] \\
&= \int \psi(\bar{x})\left[V(z,\theta_0) - \eta(\bar{x})\right] K_b(\bar{x} - x)\, d\bar{x} \\
&= \int a(x + ub) K(u)\, du
\end{aligned}$$

$$= a(x) + \int [a(x+ub) - a(x)] K(u) \, du$$
$$= a(x) + O(b^{\alpha_\psi} + |\psi(x)| \, b),$$

and where we have used that $|a(x+ub) - a(x)| \leq C\{|ub|^{\alpha_\psi} + |\psi(x)| \, |ub|\}$ uniformly in $h \in \mathcal{H}$ (recall $\mathcal{H} := \mathcal{C}_1^q(\mathcal{S}_X) \times \mathcal{C}_{M,\varepsilon}^q(\mathcal{S}_X)$ for $q \geq 1$) and

$$\left| \mathbb{E}\left[\psi(X) \frac{1}{f_X(X) \eta_{1t,z}^{b_n}(X)} [V(z,\theta_0) - \eta(X)] K_b^2(X - x) \right] \right|$$
$$\leq b_n^{-d_x} \int \frac{a(x+ub_n)}{\eta_{1t,z}^{b_n}(x+ub_n)} K^2(u) \, du$$
$$= O(b_n^{-d_x} |a(x)|).$$

Then, by (B.15), for a sufficiently large n,

$$\mathbb{E}\left[\sup_{h : |h_0 - h|_{\mathcal{H}} < \delta} |m_n(Z,h) - m_n(Z,h_0)|^\rho \right]$$
$$= \mathbb{E}\left[\sup_{h : |h_0 - h|_{\mathcal{H}} < \delta} \left| V_h^\mu(h) \left[\Delta_{t,Z}^b(h)(\bar{x}) \right] - V_h^\mu(h_0) \left[\Delta_{t,Z}^b(h_0)(\bar{x}) \right] \right|^\rho \right]$$
$$\leq \delta^\rho \mathbb{E}\left[|\psi(X)|^\rho \right] + o(\delta^\rho) = O(\delta^\rho).$$

Thus, by the Glivenko-Cantelli property

$$\sup_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^n \left\{ V_h^\mu(\eta_0) \left[\Delta_{t,Z_i}^b(h) \right] - \mathbb{E}\left[V_h^\mu(\eta_0) \left[\Delta_{t,Z}^b(h) \right] \right] \right\} \right| = o_{\mathbb{P}}(1).$$

The equality in (B.15) also shows that the limit is mean square continuous in $h \in \mathcal{H}$. Therefore, we conclude from (B.14) and the last display that $A_1 = o_{\mathbb{P}}(1)$.

We now show that $A_2 = o_{\mathbb{P}}(1)$. Note that by D(ii),

$$A_2 \leq \frac{C \left\| \widehat{h} - h_0 \right\|_{\mathcal{H}}}{n} \sum_{i=1}^n \| [V(Z_i,\theta_0) - \eta_0(\cdot)] K_b(\cdot - X_i) \|_{\mathcal{H}} = o_{\mathbb{P}}(1).$$

Hence, (B.12) holds.

Next, we show that (B.13) holds. To that end, write

$$\left| \frac{1}{n} \sum_{i=1}^n \delta_i \right| \leq \left| \frac{1}{n} \sum_{i=1}^n V_h^\mu(\eta_0) \left[\frac{\eta_{ti}^{b_n} - \eta_0}{t_n} \right] \right| + c \left| \frac{1}{n} \sum_{i=1}^n t_n^{-1} \| \eta_{ti}^{b_n} - \eta_0 \|_{\mathcal{H}}^2 \right|$$

$$\le \left| \frac{1}{n} \sum_{i=1}^{n} \psi(X_i) \left[V(Z_i, \theta_0) - \eta_0(X_i) \right] \right| + O_{\mathbb{P}} \left(ct_n b_n^{-d_x} \right)$$
$$= o_{\mathbb{P}}(1).$$

Then, we conclude by (B.12) and (B.13) that

$$|T_2| \le \left(\frac{1}{n} \sum_{i=1}^{n} \left| 2\hat{g}_i + 2\widetilde{\delta}_i - \frac{1}{n} \sum_{j=1}^{n} \widetilde{\delta}_j \right| \right) \times \left| \frac{1}{n} \sum_{j=1}^{n} \widetilde{\delta}_j \right|$$
$$= O_{\mathbb{P}}(1) o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

Similarly, using the same arguments as for (B.12), it follows that

$$|T_3| \le \left(\frac{1}{n} \sum_{i=1}^{n} \left| 2\hat{g}_i + \delta_i + \widetilde{\delta}_i \right| \right) \times \max_{1 \le i \le n} \left| \widetilde{\delta}_i - \delta_i \right| = O_{\mathbb{P}}(1) o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1).$$

To deal with T_4, we use D(ii) to conclude

$$\delta_z =: \psi(x) \left[V(z, \theta_0) - \eta_0(x) \right] + r_t^b(z), \tag{B.16}$$

where $r_t^b(z)$ satisfies

$$\frac{1}{n} \sum_{i=1}^{n} \left| r_t^b(Z_i) \right| \le \frac{ct}{n} \sum_{i=1}^{n} \| \frac{1}{\eta_{1t, Z_i}^{b_n}(\cdot)} \left[V(Z_i, \theta_0) - \eta(\cdot) \right] K_b\left(\cdot - X_i \right) \|_{\mathcal{H}}^2$$
$$= O_{\mathbb{P}}(ctb^{-d_x}) = o_{\mathbb{P}}(1).$$

Using this expansion and Assumption A(iv) and A(iii) we show that

$$T_4 \le \left| \frac{1}{n} \sum_{i=1}^{n} \hat{g}_i^2 - g_i \right| + \left| \frac{1}{n} \sum_{i=1}^{n} 2\psi(X_i) \left[V(Z_i, \theta_0) - \eta_0(X_i) \right] (\hat{g}_i - g_i) \right| + o_{\mathbb{P}}(1)$$
$$= o_{\mathbb{P}}(1).$$

Finally, by (B.16) and the law of large numbers

$$\frac{1}{n} \sum_{i=1}^{n} (g_i + \delta_i)^2 = \frac{1}{n} \sum_{i=1}^{n} \left(g_i + \psi(X_i) \left[V(Z_i, \theta_0) - \eta_0(X_i) \right] \right)^2 + o_{\mathbb{P}}(1)$$
$$= \Sigma + o_{\mathbb{P}}(1),$$

which shows that $T_5 = o_{\mathbb{P}}(1)$ and concludes the proof. □

C. Extension to the over-identified case. In this Appendix we extend the results of the paper to the over-identified case. Let Z be a random vector defined on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$ and with values on $\mathcal{S}_Z \subseteq \mathbb{R}^{d_z}$, and let $\{Z_i\}_{i=1}^n$ be independent copies of Z. Assume Z satisfies the estimating equations

$$\mathbb{E}\left[g\left(Z, \theta_0, \eta_0\right)\right] = 0, \tag{C.1}$$

where $g(\cdot) : \mathcal{S}_Z \times \Theta \times \mathcal{E} \to \mathbb{R}^l$ is a vector-valued measurable known function, $\theta_0 \in \Theta \subset \mathbb{R}^p$ $(p \leq l)$ denotes the finite-dimensional parameter of interest, Θ is a compact set, and $\eta_0 \in \mathcal{E}$ denotes the possibly infinite-dimensional nuisance parameter taking values in a semi-metric space $\mathcal{E}$. Let

$$m\left(Z, \theta_0, h_0\right) = g\left(Z, \theta_0, \eta_0\right) + \phi\left(Z, \theta_0, h_0\right)$$

denote the modified estimating equations, where $\phi\left(Z, \theta_0, h_0\right)$ is the pathwise derivative, and define

$$l_n\left(\lambda, \theta, h\right) = \sum_{i=1}^{n} \log\left(1 + \lambda' m\left(Z_i, \theta, h\right)\right),$$

where λ satisfies $0 = \sum_{i=1}^n m\left(Z_i, \theta, h\right) / \left(1 + \lambda' m\left(Z_i, \theta, h\right)\right)$. The main result is to show that under the regularity conditions listed in Assumption F below, the modified empirical likelihood ratio for the hypothesis $H_0 : \theta = \theta_0$ is asymptotically pivotal, that is

$$2\left(l_n\left(\widetilde{\lambda}, \theta_0, \widehat{h}\right) - l_n\left(\widehat{\lambda}, \widehat{\theta}, \widehat{h}\right)\right) \xrightarrow{d} \chi_p^2,$$

where $\widetilde{\lambda}$ and $\widehat{\lambda}$ satisfy

$$0 = \sum_{i=1}^{n} \frac{m\left(Z_i, \theta_0, \widehat{h}\right)}{\left(1 + \widetilde{\lambda}' m\left(Z_i, \theta_0, \widehat{h}\right)\right)}, 0 = \sum_{i=1}^{n} \frac{m\left(Z_i, \widehat{\theta}, \widehat{h}\right)}{\left(1 + \widehat{\lambda}' m\left(Z_i, \widehat{\theta}, \widehat{h}\right)\right)},$$

and $\widehat{\theta}$ is the two-step maximum empirical likelihood estimator defined as

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} l_n\left(\widehat{\lambda}, \theta, \widehat{h}\right) + o_{\mathbb{P}}\left(n^{-1}\right). \tag{C.2}$$

Assumption F: The measurable function $m(\cdot, \theta, h)$ is such that:

(i) $\theta_0 \in int(\Theta)$ satisfies $\mathbb{E}\left[m\left(\cdot, \theta_0, h_0\right)\right] = 0$,
(ii) $\left\|\widehat{\theta} - \theta_0\right\| = o_{\mathbb{P}}(1)$, $\left\|\widehat{\lambda}\right\| = o_{\mathbb{P}}(1)$, $\left\|h - h_0\right\|_{\mathcal{H}} = o_{\mathbb{P}}(1)$,

(iii) $\mathbb{E}\left[m\left(\cdot,\theta,h_0\right)\right]$ is differentiable at $\theta=\theta_0$ with derivative G_0 of full column rank p,

(iv) Uniform asymptotic no bias condition: for all positive sequences $\delta_n\downarrow 0$

$$\mathbb{P}\left[m\left(\cdot,\theta,\widehat{h}\right)-m\left(\cdot,\theta,h_0\right)\right]=o_{\mathbb{P}}(n^{-1/2})$$

for all $\|\theta-\theta_0\|\le\delta_n$,

(v) Stochastic equicontinuity in θ and h: for all positive sequences $\delta_n\downarrow 0$,

$$\sup_{\|\theta-\theta_0\|\le\delta_n,,\|h-h_0\|_{\mathcal{H}}\le\delta_n}\left|\mathbb{G}_n m\left(\cdot,\theta,h\right)-\mathbb{G}_n m\left(\cdot,\theta_0,h_0\right)\right|=o_{\mathbb{P}}\left(1\right),$$

(vi) Uniform consistency: for all positive sequences $\delta_n\downarrow 0$

$$\sup_{\|\theta-\theta_0\|\le\delta_n,\|\lambda\|\le\delta_n\|h-h_0\|_{\mathcal{H}}\le\delta_n}\left\|\mathbb{P}_n\left(\frac{\partial^2 l\left(\lambda,\theta,h\right)}{\partial\lambda\partial\lambda'}\right)-\mathbb{P}_n\left(\frac{\partial^2 l\left(0,\theta_0,h_0\right)}{\partial\lambda\partial\lambda'}\right)\right\|$$
$$=o_{\mathbb{P}}\left(1\right),$$

(vii) The matrix $\Sigma=\mathbb{E}\left[m\left(\cdot,\theta_0,h_0\right)m'\left(\cdot,\theta_0,h_0\right)\right]$ is positive definite and finite, $\mathbb{P}(l_n\left(\widetilde{\lambda},\theta_0,\widehat{h}\right)=0)\to 0$, $\mathbb{P}(l_n\left(\widehat{\lambda},\widehat{\theta},\widehat{h}\right)=0)\to 0$ and $\max_{1\le i\le n}\sup_{\theta\in\Theta}|m(Z_i,\theta,\widehat{h})|=o_{\mathbb{P}}(n^{1/2})$.

Note that Assumption F(ii) can be verified using the same arguments as those used in the proof of Theorem 3.1 of [53] , which rely on F(i) and the uniform consistency assumption $\sup_{\sup_{\|h-h_0\|_{\mathcal{H}}\le\delta_n}n}\left|\mathbb{P}_n m\left(\cdot,\theta,h\right)-\mathbb{P}_n m\left(\cdot,\theta_0,h_0\right)\right|$ $=o_{\mathbb{P}}\left(1\right)$, but does not require the differentiability of $m\left(\cdot,\theta,h\right)$ with respect to θ.

Theorem C.1. *Under Assumption F*

$$2\left(l_n\left(\widetilde{\lambda},\theta_0,\widehat{h}\right)-l_n\left(\widehat{\lambda},\widehat{\theta},\widehat{h}\right)\right)\xrightarrow{d}\chi^2_p.$$

Proof of Theorem C.1 We first establish the $n^{1/2}$ consistency of $\widehat{\theta}$ and $\widehat{\lambda}$. By Assumption F(iii) and the triangle inequality, there exists a constant $C>0$ such that

$$\begin{aligned}&\left\|\widehat{\theta}-\theta_0\right\|C\le\left\|\mathbb{E}\left[m\left(\cdot,\widehat{\theta},h_0\right)\right]\right\|\\ &\le\left\|\mathbb{E}\left[m\left(\cdot,\widehat{\theta},h_0\right)\right]-\mathbb{E}\left[m\left(\cdot,\widehat{\theta},\widehat{h}\right)\right]\right\|\\ &\quad+\left\|\mathbb{E}\left[m\left(\cdot,\widehat{\theta},\widehat{h}\right)\right]-\mathbb{P}_n\left[m\left(\cdot,\widehat{\theta},\widehat{h}\right)\right]+\mathbb{P}_n\left[m\left(\cdot,\theta_0,h_0\right)\right]\right\|\end{aligned}$$

$$+ \left\| \mathbb{P}_n \left[m \left(\cdot, \widehat{\theta}, \widehat{h} \right) \right] \right\| + \left\| \mathbb{P}_n \left[m \left(\cdot, \theta_0, h_0 \right) \right] \right\| .$$

Note that since $\left\| \widehat{\theta} - \theta_0 \right\| \leq \delta_n$ and $\|h - h_0\|_{\mathcal{H}} \leq \delta_n$ with probability approaching 1, Assumptions F(iv)-(v) imply that

$$\text{(C.3)} \quad \begin{aligned} &\left\| \mathbb{E} \left[m \left(\cdot, \widehat{\theta}, h_0 \right) \right] - \mathbb{E} \left[m \left(\cdot, \widehat{\theta}, \widehat{h} \right) \right] \right\| = o_{\mathbb{P}}(n^{-1/2}), \\ &\left\| \mathbb{E} \left[m \left(\cdot, \widehat{\theta}, \widehat{h} \right) \right] - \mathbb{P}_n \left[m \left(\cdot, \widehat{\theta}, \widehat{h} \right) \right] + \mathbb{P}_n \left[m \left(\cdot, \theta_0, h_0 \right) \right] \right\| = o_{\mathbb{P}}(n^{-1/2}). \end{aligned}$$

Since $\mathbb{P}_n \left[m \left(\cdot, \widehat{\theta}, \widehat{h} \right) \right] = \mathbb{P}_n \left[m \left(\cdot, \widehat{\theta}, h_0 \right) \right] + o_{\mathbb{P}}(n^{-1/2})$, the same arguments of Lemma A3 of [53] can be used to show that $\left\| \mathbb{P}_n \left[m \left(\cdot, \widehat{\theta}, h_0 \right) \right] \right\| = O_{\mathbb{P}}(n^{-1/2})$, hence $\left\| \widehat{\theta} - \theta_0 \right\| C \leq \left\| \mathbb{E} \left[m \left(\cdot, \widehat{\theta}, h_0 \right) \right] \right\| \leq O_{\mathbb{P}}(n^{-1/2})$. Lemma A2 of [53] can then be used to show that $\left\| \widehat{\lambda} \right\| = O_{\mathbb{P}}(n^{-1/2})$. Next we obtain the asymptotic representations of $n^{1/2} \left(\widehat{\theta} - \theta_0 \right)$ and $n^{1/2} \widehat{\lambda}$. We follow the same idea of [57] (see also [14]) and consider the approximating function

$$\mathcal{L}_n \left(\lambda, \theta, h \right) = G_0 \left(\widehat{\theta} - \theta_0 \right)' \lambda - \mathbb{P}_n \left[m \left(\cdot, \theta_0, h \right) \right] - \frac{1}{2} \lambda' \Sigma \lambda.$$

By a second order Taylor expansion about $\lambda = 0$

$$\mathbb{P}_n l \left(\widehat{\lambda}, \widehat{\theta}, \widehat{h} \right) = \widehat{\lambda}' \mathbb{P}_n \left[m \left(\cdot, \widehat{\theta}, \widehat{h} \right) \right] + \frac{1}{2} \widehat{\lambda}' \mathbb{P}_n \left(\frac{\partial^2 l \left(\overline{\lambda}, \widehat{\theta}, \widehat{h} \right)}{\partial \lambda \partial \lambda'} \right) \widehat{\lambda},$$

where $\overline{\lambda}$ is in the line segment between 0 and $\widehat{\lambda}$, hence by the triangle inequality and Assumptions F(iii)-(vi)

$$\text{(C.4)} \quad \begin{aligned} &|\mathbb{P}_n l(\widehat{\lambda}, \widehat{\theta}, \widehat{h}) - \mathcal{L}_n(\widehat{\lambda}, \widehat{\theta}, \widehat{h})| \\ &\leq \|\mathbb{P}_n[m(\cdot, \widehat{\theta}, \widehat{h}) - m(\cdot, \theta_0, \widehat{h})] - G_0(\widehat{\theta} - \theta_0)\| \|\widehat{\lambda}\| \\ &\quad + \frac{1}{2} \|[\mathbb{P}_n(\frac{\partial^2 l(\overline{\lambda}, \widehat{\theta}, \widehat{h})}{\partial \lambda \partial \lambda'}) - \Sigma]\| \|\widehat{\lambda}\|^2 \\ &\leq \{n^{-1/2} \|\mathbb{G}_n m(\cdot, \widehat{\theta}, \widehat{h}) - \mathbb{G}_n m(\cdot, \theta_0, h_0)\| \\ &\quad + \|\mathbb{E}[m(\cdot, \widehat{\theta}, \widehat{h})] - \mathbb{E}[m(\cdot, \widehat{\theta}, h_0)]\| + \mathbb{P}_n[m(\cdot, \theta_0, \widehat{h})] \\ &\quad + \|\mathbb{E}[m(\cdot, \widehat{\theta}, h_0)] - G_0(\widehat{\theta} - \theta_0)\|\} \|\widehat{\lambda}\| \\ &= (o_{\mathbb{P}}(n^{-1/2}) + o_{\mathbb{P}}(\|\widehat{\theta} - \theta_0\|)) \|\widehat{\lambda}\| + \frac{\|\widehat{\lambda}\|^2}{2} o_{\mathbb{P}}(1) = o_{\mathbb{P}}(n^{-1}), \end{aligned}$$

that is $|\mathbb{P}_n l(\widehat{\lambda}, \widehat{\theta}, \widehat{h}) - \mathcal{L}_n(\widehat{\lambda}, \widehat{\theta}, \widehat{h})| = o_{\mathbb{P}}(n^{-1})$. Consider the program $\min_{\theta\in\Theta}\max_{\lambda\in\mathbb{R}^l}\mathcal{L}_n(\lambda,\theta,\widehat{h})$. Since $\mathcal{L}_n(\cdot)$ is concave in λ and $\theta_0 \in int(\Theta)$ the first order conditions

$$G_0\left(\check{\theta}-\theta_0\right) - \mathbb{P}_n\left[m\left(\cdot,\theta_0,\widehat{h}\right)\right] - \Sigma\check{\lambda} = 0,\; G_0'\check{\lambda} = 0, \tag{C.5}$$

are satisfied with probability approaching 1. Rearranging, we have

$$\begin{aligned} n^{1/2}\begin{bmatrix} \check{\lambda} \\ \check{\theta}-\theta_0 \end{bmatrix} &= \begin{bmatrix} -\Sigma & G_0 \\ G_0' & 0 \end{bmatrix}^{-1} n^{1/2}\mathbb{P}_n\left[m\left(\cdot,\theta_0,\widehat{h}\right)\right] \\ &= \begin{bmatrix} P \\ \left(G_0'\Sigma^{-1}G_0\right)^{-1}G_0'\Sigma^{-1} \end{bmatrix} n^{1/2}\mathbb{P}_n\left[m\left(\cdot,\theta_0,h_0\right)\right] + o_{\mathbb{P}}(1), \end{aligned} \tag{C.6}$$

where $P = \Sigma^{-1} - \Sigma^{-1}G_0\left(G_0'\Sigma^{-1}G_0\right)^{-1}G_0'\Sigma^{-1}$. Finally we show that $n^{1/2}(\widehat{\theta}-\check{\theta})$ and $n^{1/2}(\widehat{\lambda}-\check{\lambda})$ are both $o_{\mathbb{P}}(1)$. By similar arguments to (C.3) and (C.4), it is possible to show that

$$\left|\mathbb{P}_n l\left(\widehat{\lambda},\check{\theta},\widehat{h}\right) - \mathcal{L}_n\left(\widehat{\lambda},\check{\theta},\widehat{h}\right)\right| = o_{\mathbb{P}}(n^{-1}),$$

hence by (C.2) and (C.4)

$$\begin{aligned} \mathcal{L}_n\left(\widehat{\lambda},\widehat{\theta},\widehat{h}\right) - o_{\mathbb{P}}(n^{-1}) = \mathbb{P}_n l\left(\widehat{\lambda},\widehat{\theta},\widehat{h}\right) &\le \mathbb{P}_n l\left(\widehat{\lambda},\check{\theta},\widehat{h}\right) + o_{\mathbb{P}}(n^{-1}) \\ &= \mathcal{L}_n\left(\widehat{\lambda},\check{\theta},\widehat{h}\right) + o_{\mathbb{P}}(n^{-1}), \end{aligned}$$

which implies that

$$\left|\mathcal{L}_n\left(\widehat{\lambda},\widehat{\theta},\widehat{h}\right) - \mathcal{L}_n\left(\widehat{\lambda},\check{\theta},\widehat{h}\right)\right| = \left\|G_0\left(\widehat{\theta}-\check{\theta}\right)\right\|\left\|\widehat{\lambda}\right\| \le o_{\mathbb{P}}(n^{-1}),$$

and thus $n^{1/2}\left(\widehat{\theta}-\check{\theta}\right) = o_{\mathbb{P}}(1)$ by $\left\|\widehat{\lambda}\right\| = O_{\mathbb{P}}\left(n^{-1/2}\right)$. It remains to show that $\left\|n^{1/2}\left(\widehat{\lambda}-\check{\lambda}\right)\right\| = o_{\mathbb{P}}(1)$. By definition $\mathcal{L}_n\left(\widehat{\lambda},\check{\theta},\widehat{h}\right) \le \mathcal{L}_n\left(\widehat{\lambda},\widehat{\theta},\widehat{h}\right) + o_{\mathbb{P}}(n^{-1})$, hence $\mathcal{L}_n\left(\widehat{\lambda},\check{\theta},\widehat{h}\right) = \mathcal{L}_n\left(\widehat{\lambda},\widehat{\theta},\widehat{h}\right) + o_{\mathbb{P}}(n^{-1})$ and from (C.5)

$$\left|\mathcal{L}_n\left(\widehat{\lambda},\check{\theta},\widehat{h}\right) - \mathcal{L}_n\left(\widehat{\lambda},\widehat{\theta},\widehat{h}\right)\right| \le \left\|\left(\check{\lambda}-\widehat{\lambda}\right)\right\|^2\|\Sigma\| = o_{\mathbb{P}}(n^{-1}),$$

which implies $\left\|n^{1/2}\left(\widehat{\lambda}-\check{\lambda}\right)\right\| = o_{\mathbb{P}}(1)$. Finally by a second order Taylor expansion, Assumptions F(vi)-(vii), (C.6) and the results of Theorem **3.1**

$$2\left(l_n\left(\widetilde{\lambda},\theta_0,\widehat{h}\right) - l_n\left(\widehat{\lambda},\widehat{\theta},\widehat{h}\right)\right)$$

$$= n\left(\widetilde{\lambda}'\Sigma\widetilde{\lambda} - \check{\lambda}\Sigma\check{\lambda}\right) + o_{\mathbb{P}}(1)$$
$$= n\mathbb{P}_n m\left(\cdot,\theta_0,\widehat{h}\right)\left(\Sigma^{-1} - P\right)\mathbb{P}_n m\left(\cdot,\theta_0,\widehat{h}\right) o_{\mathbb{P}}(1)$$
$$= n\mathbb{P}_n m\left(\cdot,\theta_0,\widehat{h}\right)\Sigma^{-1/2}\left(\Sigma^{-1/2}G_0\left(G_0'\Sigma^{-1}G_0\right)^{-1}G_0'\Sigma^{-1/2}\right)\Sigma^{-1/2}$$
$$\times\,\mathbb{P}_n m\left(\cdot,\theta_0,\widehat{h}\right) + o_{\mathbb{P}}(1),$$

and the conclusion follows by a standard result on quadratic forms with an idempotent matrix, see [61] for details. □

D. Auxiliary Result. The following result is instrumental in proving Donsker and Glivenko-Cantelli properties for semiparametric moments that are non-smooth in nuisance parameters. It is a slight variation of Theorem 3 in [14], in that it allows for classes of functions that change with n. Define the class $\mathcal{F}_n = \{z \to m_n(z,v,h) : (v,h) \in \mathcal{V}\times\mathcal{H}\}$. Define for a totally bounded semi-metric space $(T,\|\cdot\|)$ and $\delta > 0$,

$$J(\delta,T,\|\cdot\|) = \int_0^\delta \sqrt{\log N(\varepsilon,T,\|\cdot\|)}d\varepsilon.$$

THEOREM D.1. *Assume that for all $(v_0,h_0)\in V\times H$,*

(D.1)
$$\sup_{n\geq1}\left(\mathbb{E}\left[\sup_{v:|v_0-v|_{\mathcal{V}}<\delta,h:|h_0-h|_{\mathcal{H}}<\delta}|m_n(Z,v,h)-m_n(Z,v_0,h_0)|^\rho\right]\right)^{1/\rho}\leq C\delta^s$$

and

$$\sup_{n\geq1}\mathbb{E}\left[|m_n(Z,v_0,h_0)|^\rho\right]\leq C, \tag{D.2}$$

for all sufficiently small $\delta>0$, and constants$\rho\geq1$, $s>0$, and $C>0$.

- **(i)** *Assume that for each $\varepsilon>0$, $N(\varepsilon,\mathcal{V},|\cdot|_{\mathcal{V}})<\infty$ and $N(\varepsilon,\mathcal{H},|\cdot|_{\mathcal{H}})<\infty$. Then, if $\rho>1$ the class $\mathcal{F}_n$ is Glivenko-Cantelli. If $\mathcal{F}_n$ does not depend on n, this conclusion also holds with $\rho=1$.*
- **(ii)** *Assume that for each $\delta_n\downarrow0$, $J(\delta_n,V,|\cdot|_{\mathcal{V}})\downarrow0$ and $J(\delta_n,\mathcal{H},|\cdot|_{\mathcal{H}})\downarrow0$. Then, if $\rho>2$ the class $\mathcal{F}_n$ is Donsker. If $\mathcal{F}_n$ does not depend on n, this conclusion also holds with $\rho=2$.*

Proof of Theorem D.1 By the proof of Theorem 3(ii) in [14] we conclude

$$N_{[\cdot]}(\varepsilon,\mathcal{F}_n,\|\cdot\|_\rho)\leq N\left(\left(\frac{\varepsilon}{2C}\right)^{1/s},\mathcal{V},|\cdot|_{\mathcal{V}}\right)\times N\left(\left(\frac{\varepsilon}{2C}\right)^{1/s},\mathcal{H},|\cdot|_{\mathcal{H}}\right),$$

and there exists ε-brackets for $(\mathcal{F}_n, \|\cdot\|_\rho)$ of the form

$$[m_n(Z, v_k, h_j) - b_n(Z, v_k, h_j, \delta), m_n(Z, v_k, h_j) + b_n(Z, v_k, h_j, \delta)],$$

where

$$b_n(Z, v_k, h_j, \delta) = \sup_{v:|v-v_k|_{\mathcal{V}}<\delta, h:|h-h_j|_{\mathcal{H}}<\delta} |m_n(Z, v, h) - m_n(Z, v_k, h_j)|.$$

Then, the theorem follows by the classical proof of Glivenko-Cantelli Theorem, since each function in the brackets satisfies the conditions for law of large numbers and central limit theorems, respectively, for functions of iid variables that change with n (i.e. $\sup_{n\geq 1} \mathbb{E}[|m_n(Z, v_k, h_j)|^\rho] < \infty$ and $\sup_{n\geq 1} \mathbb{E}[|b_n(Z, v_k, h_j, \delta)|^\rho] < \infty$). □

References.

[1] Ackerberg, D., X. Chen, and J. Hahn, (2012). A Practical Asymptotic Variance Estimator for Two-step Semiparametric Estimators. *Rev. Econ. and Stat.* 94, 481-498.

[2] Ackerberg, D., X. Chen, J. Hahn and Z. Liao (2014). Asymptotic Efficiency of Semiparametric Two-step GMM. *Rev. of Econ. Stud.* 81, 919-943.

[3] Akritas, M.G. and Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist.*, **28**, 549-568.

[4] Andrews, D.W.K. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric. Theory*, **11**, 560-596.

[5] Banerjee, M. (2012). Current Status Data in the 21st Century: Some Interesting Developments. *Interval-Censored Time-to-Event Data: Methods and Applications*, 1-31, Chapman and Hall/CRC Biostatistics Series.

[6] Beran, R. (1981). *Nonparametric regression with randomly censored survival data.* Technical Report, Univ. California, Berkeley.

[7] Bertail P. (2006). Empirical likelihood in some semi-parametric models. *Bernoulli*, **12**, 299-331.

[8] Bickel, P.J., Klaassen, C.A., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins Univ. Press, Baltimore.

[9] Bickel,P.J., Y. Ritov, and T. M. Stoker (2006). Tailor-made tests for goodness of fit to semiparametric hypotheses. *Ann. Stat.,* **34**, 721-741.

[10] Bravo, F. (2004). Empirical likelihood based inference with applications to some econometric models. *Econometric. Theory,* **20**, 231-264.

[11] Bravo, F. (2018). *Second order asymptotics for nonparametric conditional moment restrictions.* Working paper.

[12] Bravo, F., Chu, B. and Jacho-Chavez, D.T. (2017). Semiparametric estimation of moment conditions models with weakly dependent data. *J. Nonpar. Statist.*, **29**, 108-136.

[13] Chen, X.H., Hong, H. and Tarozzi, A. (2008). Semiparametric efficiency in GMM models with auxiliary data. *Ann. Statist.*, **36**, 808-843.

[14] Chen, X., Linton, O. and Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, **71**, 1591-1608.

[15] Chen, S.X. and Cui, H. (2006). On the Bartlett correctability of empirical likelihood in the presence of nuisance parameters. *Biometrika*, **93**, 215-220.

[16] Chen, S.X. and Cui, H. (2007). On the second order properties of empirical likelihood with moment restrictions. *J. Econometrics*, **141**, 492-516.

[17] Chen, S.X. and Van Keilegom, I. (2009). A review on empirical likelihood methods for regression. *TEST*, **18**, 415-447.

[18] Chernozhukov, V., Escanciano, J.C., Newey, W.K., Ichimura, H., J. Robins (2017). Locally Robust Semiparametric Estimation. Working paper.

[19] Choi, S., Hall, W.J. and Schick, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.*, **24**, 841-861.

[20] DiCiccio, T. and Romano, (1989). On adjustments based on the signed root of the empirical likelihood ratio statistics. *Biometrika*, **76**, 447-456.

[21] DiCiccio, T., Hall, P. and Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.*, **19**, 1053-1061.

[22] Du, Y. and Akritas, M.G. (2002). I.i.d representations of the conditional Kaplan-Meier process for arbitrary distributions. *Math. Methods Statist.*, **11**, 152-182.

[23] Escanciano, J.C., D. T. Jacho-Chavez and A. Lewbel (2014). Uniform convergence of weighted sums of non- and semi-parametric residuals for estimation and testing.

J. Econometrics, **178**, 426-443.
[24] Fan, J. and Zhang, J. (2004). Sieve empirical likelihood tests for nonparametric functions. *Ann. Statist.*, **32**, 1858-1907.
[25] Goldstein, L. and Messer, K. (1992). Optimal plug-in estimators for nonparametric functional estimation. *Ann. Statist.*, **20,** 1306-1328.
[26] Groeneboom, P. and Wellner, J.A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhauser, Basel.
[27] Hampel, F. (1968). *Contributions to the Theory of Robust Estimation*, PhD thesis, University of California, Berkeley.
[28] Hampel, F. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.*, **69**, 383-393.
[29] He, S., Liang, W., Shen, J. and Yang G. (2016). Empirical likelihood for right censored lifetime data. *J. Amer. Stat. Assoc.,* **111**, 646-655.
[30] Hirano, K., Imbens, G.W. and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica,* **71**, 1161-1189.
[31] Hjort, N.L., McKeague, I.W. and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.*, **37**, 1079-1111.
[32] Hong, H. Mahajan, A. and D. Nekipelov (2015): "Extremum estimation and numerical derivatives," *J. Econometrics* 188, 250-263.
[33] Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.*, **24**, 540-568.
[34] Huang, J. and J. A. Wellner (1997). *Interval Censored Survival Data: A Review of Recent Progress*. Eds. D. Lin and T. Fleming. Springer-Verlag, New York.
[35] Ichimura, H. and Linton, O. (2005). Asymptotic expansions for some semiparametric program evaluation estimators. *Identification and Inference for Econometric Models*, edited by D. Andrews and J. Stock, Cambridge University Press, NY.
[36] Ichimura, H. and W. Newey (2017). The Influence Function of Semiparametric Estimators. CEMMAP working paper.
[37] Inglot, T. and Ledwina, T. (2006). Data-driven score tests for homoscedastic linear regression model : asymptotic results. *Probab. Math. Statist.*, **26**, 41-61.
[38] Jewell, N. P. and Van der Laan, M. (2003). Current status data: Review, recent developments and open problems. *Handbook of Statistics,* 5, 291-306.
[39] Koshevnik, Y. A., and Levit, B. Y. (1976). On a non-parametric analogue of the information matrix. *Theory Probab. Applic.*, **21**, 738-753.
[40] Leng, C. and Tong, X. (2013). A quantile regression estimator for censored data. *Bernoulli*, **19**, 344-361.
[41] Lepage, Y. (1973). A maximin test for means. *Stat. Neerlandica*, **27**, 1.
[42] Lewbel, A., O.B. Linton and D. McFadden (2011). Estimating features of a distribution from binomial data. *J. Econometrics.*, **162**, 170-188.
[43] Li. G., Lin, L. and Zhu, L. (2012). Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters. *J. Mult. Anal.,* **105**, 85-111.
[44] Li, Q., Racine, J. and Wooldridge, J. (2008). Estimating average treatment effects with continuous and discrete covariates: the case of Swan-Ganz catherization. *Amer. Econ. Review: Papers Proc.*, **98**, 357-362.
[45] Lopez, O. (2011). Nonparametric estimation of the multivariate distribution function in a censored regression model with applications. *Comm. Statist.: Theory Meth.*, **40**, 2639-2660.
[46] Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *J. Time Ser. Anal.*, **17**, 571-599.

[47] Matsushita, Y. and Otsu, T. (2016). Likelihood inference on semiparametric models with generated regressors. LSE working paper.

[48] Matsushita, Y. and Otsu, T. (2018). Likelihood inference on semiparametric models: average derivative and treatment effect. *Jap. Econ. Rev.*, **69**, 2.

[49] Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Mult.. Anal.*, **101**, 1067-1078.

[50] Newey, W.K. (1994a). Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, **10**, 233-253.

[51] Newey, W.K. (1994b). The asymptotic variance of semiparametric estimators. *Econometrica*, **62**, 1349-1382.

[52] Newey, W.K., F. Hsieh, and J. Robins (2004). Twicing kernels and a small bias property of semiparametric estimators. *Econometrica,* **72**, 947-962.

[53] Newey, W.K. and Smith R.J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, **72**, 219-255.

[54] Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237-249.

[55] Owen, A.B. (1990). Empirical likelihood ratio confidence Regions. *Ann. Statist.*, **18**, 90-120.

[56] Owen, A.B. (2001). *Empirical Likelihood*. Chapman and Hall, London.

[57] Pakes, A. and Pollard D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, **57**, 1027-1057.

[58] Pfanzagl, J. (1982). *Contributions to a general statistical theory.* Lecture Notes in Statistics. 13. New York: Springer-Verlag.

[59] Powell, J.L., Stock, J.H. and Stoker, T.M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, **57**, 1403-1430.

[60] Qin, J. and Lawless, J. (1994). Empirical Likelihood and General Estimating Equations. *Ann. Statist*, 22, 300-325.

[61] Rao, C.R. and S. K. Mitra (1971). textitGeneralized inverse of matrices and its applications. John Wiley and Sons, New York.

[62] Rosembaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41-55.

[63] Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psych.*, **66**, 688-701.

[64] Shao, J. and Sitter, R.R. (1996). Bootstrap for imputed survey data. *J. Amer. Statist. Assoc.*, **91**, 1278-1288.

[65] Stone, C.J. (1982). Optimal global rate of convergence for nonparametric regression. *Ann. Statist.*, **10**, 1040-1053.

[66] Strasser, H. (1985). *Mathematical Theory of Statistics.* De Gruyter, Berlin.

[67] Sun, J. (2006): *The Statistical Analysis of Interval-censored Failure Time Data.* Springer-Verlag, New York.

[68] Tang, X., Li, J. and Lian,H. (2013). Empirical likelihood for partially linear proportional hazards models with growing dimensions. *J. Mult. Anal.,* **121**, 22-32.

[69] Tang, C.Y. and Qin Y. (2012). An efficient empirical likelihood approach for estimating equations with missing data. *Biometrika*, **99**, 1001-1007.

[70] Van der Vaart, A.W. (1991). On differentiable functionals. *Ann. Stat.,* **19**, 178-204.

[71] Van der Vaart, A.W. (1998). *Asymptotics Statistics.* Cambridge University Press, Cambridge.

[72] Van der Vaart, A.W. and Wellner, J.A (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New-York.

[73] Wang, D. and Chen, S.X. (2009). Empirical likelihood for estimating equation with missing values. *Ann. Statist.*, **37**, 490-517.

[74] Wang, Q.-H. and Jing, B.-Y. (2003). Empirical likelihood for partial linear models. *Ann. Inst. Statist. Math.*, **55**, 585-595.

[75] Wang, Q., Linton, O. and Härdle, W. (2004). Semiparametric regression analysis with missing response at random. *J. Amer. Statist. Assoc.*, **99**, 334-345.

[76] Wang, Q.H. and Rao, J.N.K. (2001). Empirical likelihood for linear regression models under imputation for missing responses. *Canad. J. Statist.*, **29**, 597-608.

[77] Wang, Q.H. and Rao, J.N.K. (2002a). Empirical likelihood-based inference in linear models with missing data. *Scand. J. Statist.*, **29**, 563-576.

[78] Wang, Q.H. and Rao, J.N.K. (2002b). Empirical likelihood-based inference under imputation for missing response data. *Ann. Stat.,* **30**, 896-924.

[79] Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, **9**, 60-62.

[80] Wu, T.T., Li, G. and Tang, C. (2015). Empirical likelihood for censored linear regression and variable selection. *Scand. J. Stat.,* **42**, 798-812.

[81] Yang, G., Cui, X. and Hou, S. (2017). Empirical likelihood confidence regions in the single-index model with growing dimensions. *Commun. Stat.,* **46**, 7562-7579.

[82] Xue, L. (2009). Empirical likelihood confidence intervals for response mean with data missing at random. *Scand. J. Stat.,* **36**, 671-685.

[83] Xue, L. and Wang, Q. (2012). Empirical likelihood for single-index varying-coefficient models. *Bernoulli*, **18**, 836-856.

[84] Xue, L. and Xue, D. (2011). Empirical likelihood for semiparametric regression model with missing response data. *J. Multiv. Anal.,* **102**, 723-740.

[85] Xue, L.-G. and Zhu, L. (2006). Empirical likelihood for single-index models. *J. Multiv. Anal.*, **97**, 1295-1312.

[86] Xue, L.G. and Zhu, L. (2007). Empirical likelihood semiparametric regression analysis for longitudinal data. *Biometrika,* **94**, 921-937.

[87] Xue, L.G. and Zhu, L. (2012). Empirical likelihood in some nonparametric and semiparametric models. *Statist. Interf.*, **5**, 367-378.

[88] Zheng, M., Zhao, Z. and Yu, W. (2012). Empirical likelihood methods based on influence functions. *Statist. Interf.,* **5**, 355-366.

[89] Zhu, L., L. Lu, X. Cui and G. Li (2010). Bias-corrected empirical likelihood in a multi-link semiparametric model. *J. Mult. Anal.,* **101**, 850-868.

[90] Zhu, L. and Xue, L. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *J. Roy. Statist. Soc. - Ser. B*, **68**, 549-570.

Department of Economics
University of York
Heslington, York YO10 5DD
UK
E-mail: francesco.bravo@york.ac.uk

Department of Economics
Universidad Carlos III de Madrid
Calle Madrid 126, Getafe, 28902
Spain
E-mail: jescanci@eco.uc3m.es

ORSTAT
KU Leuven
Naamsestraat 69, 3000 Leuven
Belgium
E-mail: ingrid.vankeilegom@kuleuven.be