Econometrica Supplementary Material

SUPPLEMENT TO "OVERIDENTIFICATION IN REGULAR MODELS"

(*Econometrica*, Vol. 86, No. 5, September 2018, 1771–1817)

XIAOHONG CHEN
Cowles Foundation for Research in Economics, Yale University

ANDRES SANTOS
Department of Economics, UCLA

THIS SUPPLEMENTARY APPENDIX contains additional results and proofs to support the main text.

Appendix D contains the proof of the limiting experiment results in Appendix A and additional lemmas. Appendix E presents the technical lemmas and their proofs that are used in the proofs of Appendix B and Appendix C. Appendix F contains the proofs of the results in Section 4. Appendix G provides sufficient conditions for verifying Assumption 3.1 in the general nonparametric conditional moment restriction models studied in Section 4. Appendix H provides additional discussion of the Examples in Section 4 as well as a final example.

APPENDIX D: PROOFS FOR APPENDIX A AND ADDITIONAL LEMMAS

In this appendix, we provide the proofs of Theorem A.1 and additional technical lemmas.

PROOF OF THEOREM A.1: To establish part (i), we first note that, for any $g \in L_0^2(P)$, it is possible to construct a path $t \mapsto P_{t,g}$ whose score is g; see Example 3.2.1 in Bickel, Klaassen, Ritov, and Wellner (1993) for a concrete construction. Further, any two paths $t \mapsto \tilde{P}_{t,g}$ and $t \mapsto P_{t,g}$ with the same score $g \in L_0^2(P)$ satisfy

$$\lim_{n\to\infty}\left|\int \phi_n \, d\tilde{P}^n_{1/\sqrt{n},g} - \int \phi_n \, dP^n_{1/\sqrt{n},g}\right| \leq \lim_{n\to\infty}\int \left|dP^n_{1/\sqrt{n},g} - d\tilde{P}^n_{1/\sqrt{n},g}\right| = 0 \tag{D.1}$$

for any $0 \leq \phi_n \leq 1$ by Lemma D.1 (below). Thus, for each $g \in L_0^2(P)$, we may select an arbitrary path $t \mapsto P_{t,g}$ whose score is indeed g, and for $\mathcal{B}$ the σ-algebra on $\mathbf{X}$, we consider the sequence of experiments

$$\mathcal{E}_n \equiv \left(\mathbf{X}^n, \mathcal{B}^n, P^n_{1/\sqrt{n},g} : g \in L_0^2(P)\right). \tag{D.2}$$

Next, since $\{\psi_k^T\}_{k=1}^{d_T} \cup \{\psi_k^{T^\perp}\}_{k=1}^{d_{T^\perp}}$ forms an orthonormal basis for $L_0^2(P)$, we obtain from Lemma D.3 (below) that $\mathcal{E}_n$ converges weakly to the experiment $\mathcal{E}$ given by

$$\mathcal{E} \equiv \left(\mathbf{R}^{d_T} \times \mathbf{R}^{d_{T^\perp}}, \mathcal{A}^{d_T} \times \mathcal{A}^{d_{T^\perp}}, Q_g : g \in L_0^2(P)\right), \tag{D.3}$$

where $\mathcal{A}$ denotes the Borel σ-algebra on $\mathbf{R}$ and we exploited that for $d_P \equiv \dim\{L_0^2(P)\}$ we have $\mathbf{R}^{d_T} \times \mathbf{R}^{d_{T^\perp}} = \mathbf{R}^{d_P}$ and $\mathcal{A}^{d_T} \times \mathcal{A}^{d_{T^\perp}} = \mathcal{A}^{d_P}$. The existence of a test function

Xiaohong Chen: xiaohong.chen@yale.edu
Andres Santos: andres@econ.ucla.edu

 https://doi.org/10.3982/ECTA13559

$\phi : (\mathbb{Y}^T, \mathbb{Y}^{T^\perp}) \to [0,1]$ satisfying $\pi(g_0) = \int \phi\, dQ_{g_0}$ for all $g_0 \in L_0^2(P)$ then follows from Theorem 7.1 in van der Vaart (1991). To establish part (i) of the theorem, it thus only remains to show that ϕ must control size in (A.2). To this end, note that $\Pi_{T^\perp}(g_0) = 0$ if and only if $g_0 \in \bar{T}(P)$. Fixing $\delta > 0$, then observe that for any $g_0 \in \bar{T}(P)$, there exists a $\tilde{g} \in T(P)$ such that $\|g_0 - \tilde{g}\|_{P,2} < \delta$. Moreover, since $\tilde{g} \in T(P)$, there exists a path $t \mapsto \tilde{P}_{t,\tilde{g}} \in \mathbf{P}$ with score $\tilde{g}$ and hence we can conclude that

$$\begin{aligned}\int \phi\, dQ_{g_0} &= \lim_{n\to\infty} \int \phi_n\, dP^n_{1/\sqrt{n}, g_0} \\ &\le \lim_{n\to\infty} \int \phi_n\, d\tilde{P}^n_{1/\sqrt{n},\tilde{g}} + \limsup_{n\to\infty} \int \left| dP^n_{1/\sqrt{n},g_0} - d\tilde{P}^n_{1/\sqrt{n},\tilde{g}} \right| \\ &\le \alpha + 2\left\{1 - \exp\left\{-\frac{\delta^2}{4}\right\}\right\}^{1/2},\end{aligned} \tag{D.4}$$

where the first inequality employed $0 \le \phi_n \le 1$, and the second inequality exploited Lemma D.1 and that ϕ_n is a local asymptotic level α test. Since $\delta > 0$ is arbitrary, we conclude from (11) and (D.4) that $\pi(g_0) = \int \phi\, dQ_{g_0} \le \alpha$ whenever $g_0 \in \bar{T}(P)$, and hence part (i) of the theorem follows.

For part (ii) of the theorem, we first note that since $T(P)$ is linear by Assumption 2.1(ii), and $\hat{\theta}_n$ is regular by hypothesis, Lemma D.4 (below) and Theorem 5.2.3 in Bickel et al. (1993) imply θ is pathwise differentiable at P; that is, there exists a bounded linear operator $\dot{\theta} : \bar{T}(P) \to \mathbf{B}$ such that, for any submodel $t \mapsto P_{t,g} \in \mathbf{P}$, it follows that

$$\lim_{t\downarrow 0} \left\| t^{-1}\{\theta(P_{t,g}) - \theta(P)\} - \dot{\theta}(g) \right\|_{\mathbf{B}} = 0. \tag{D.5}$$

Then note that for any $b^* \in \mathbf{B}^*$, $b^* \circ \dot{\theta} : \bar{T}(P) \to \mathbf{R}$ is a continuous linear functional. Hence, since $\bar{T}(P)$ is a Hilbert space under $\|\cdot\|_{P,2}$, the Riesz representation theorem implies there exists a $\dot{\theta}_{b^*} \in \bar{T}(P)$ such that, for all $g \in \bar{T}(P)$, we have

$$b^*\big(\dot{\theta}(g)\big) = \int \dot{\theta}_{b^*} g\, dP. \tag{D.6}$$

Moreover, since $\hat{\theta}_n$ is an asymptotically linear regular estimator of $\theta(P)$, it follows that $b^*(\hat{\theta}_n)$ is an asymptotically linear regular estimator of $b^*(\theta(P))$ with influence function $b^* \circ \nu$. Proposition 3.3.1 in Bickel et al. (1993) then implies that for all $g \in \bar{T}(P)$,

$$\int \big(\dot{\theta}_{b^*} - b^* \circ \nu\big) g\, dP = 0. \tag{D.7}$$

In particular, (D.7) implies that $\dot{\theta}_{b^*} = \Pi_T(b^* \circ \nu)$, and therefore, by asymptotic linearity,

$$\sqrt{n}\{b^*(\hat{\theta}_n) - b^*\big(\theta(P)\big)\} \xrightarrow{L} N\big(0, \|\dot{\theta}_{b^*}\|^2_{P,2} + \left\|\Pi_{T^\perp}\big(b^* \circ \nu\big)\right\|^2_{P,2}\big), \tag{D.8}$$

where we have exploited the central limit theorem and $b^* \circ \nu = \Pi_T(b^* \circ \nu) + \Pi_{T^\perp}(b^* \circ \nu)$. To conclude, we next define the maps $F^T(\mathbb{Y}^T)$ and $F^{T^\perp}(\mathbb{Y}^{T^\perp})$ to be given by

$$F^T(\mathbb{Y}^T) = \sum_{k=1}^{d_T} \mathbb{Y}_k^T \int \{\dot{\theta}_{b^*}\}\psi_k^T \, dP,$$
$$F^{T^\perp}(\mathbb{Y}^{T^\perp}) = \sum_{k=1}^{d_{T^\perp}} \mathbb{Y}_k^{T^\perp} \int \{\Pi_{T^\perp}(b^* \circ \nu)\}\psi_k^{T^\perp} \, dP. \tag{D.9}$$

We aim to show that if $(\mathbb{Y}^T, \mathbb{Y}^{T^\perp}) \sim Q_{g_0}$ with $g_0 = 0$, then $F^T(\mathbb{Y}^T) \sim N(0, \|\dot{\theta}_{b^*}\|_{P,2}^2)$, which is immediate if $d_T < \infty$, and thus we assume $d_T = \infty$. Defining the partial sums

$$\mathbb{V}_K \equiv \sum_{k=1}^{K} \mathbb{Y}_k^T \int \{\dot{\theta}_{b^*}\}\psi_k^T \, dP, \tag{D.10}$$

we then observe $\mathbb{V}_K \sim N(0, \sigma_K^2)$ where $\sigma_K^2 \equiv \sum_{k=1}^K \int \{\int \dot{\theta}_{b^*}\psi_k^T\}^2 \, dP$ and $\sigma_K^2 \uparrow \|\dot{\theta}_{b^*}\|_{P,2}^2$ by Parseval's identity. By the martingale convergence theorem (see, e.g., Theorem 12.1.1 in Williams (1991)), it follows that $\mathbb{V}_K$ converges almost surely and thus that $F^T(\mathbb{Y}^T)$ is well defined. Moreover, for any continuous bounded function $f : \mathbf{R} \to \mathbf{R}$, it follows that

$$\begin{aligned} E\big[f\big(F^T(\mathbb{Y}^T)\big)\big] &= \lim_{K\to\infty} E\big[f(\mathbb{V}_K)\big] = \lim_{K\to\infty} \frac{1}{\sqrt{2\pi}\sigma_K} \int f(z) \exp\left\{-\frac{z^2}{2\sigma_K^2}\right\} dz \\ &= \frac{1}{\sqrt{2\pi}\|\dot{\theta}_{b^*}\|_{P,2}} \int f(z) \exp\left\{-\frac{z^2}{2\|\dot{\theta}_{b^*}\|_{P,2}^2}\right\} dz \end{aligned} \tag{D.11}$$

due to $\sigma_K^2 \uparrow \|\dot{\theta}_{b^*}\|_{P,2}^2$. We conclude from (D.11) that $F^T(\mathbb{Y}^T) \sim N(0, \|\dot{\theta}_{b^*}\|_{P,2}^2)$ when $(\mathbb{Y}^T, \mathbb{Y}^{T^\perp}) \sim Q_{g_0}$ with $g_0 = 0$. Identical arguments imply $F^{T^\perp}(\mathbb{Y}^{T^\perp}) \sim N(0, \|\Pi_{T^\perp}(b^* \circ \nu)\|_{P,2}^2)$. Thus, part (ii) of the theorem follows from (D.8) and independence of $\mathbb{Y}^T$ and $\mathbb{Y}^{T^\perp}$. *Q.E.D.*

LEMMA D.1: *If $t \mapsto P_{t,g_1}$ and $t \mapsto P_{t,g_2}$ are arbitrary paths, then it follows that*

$$\limsup_{n\to\infty} \int \big|dP^n_{1/\sqrt{n},g_1} - dP^n_{1/\sqrt{n},g_2}\big| \le 2\left\{1 - \exp\left\{-\frac{1}{4}\|g_1 - g_2\|_{P,2}^2\right\}\right\}^{1/2}. \tag{D.12}$$

PROOF: Since $t \mapsto P_{t,g_1}$ and $t \mapsto P_{t,g_2}$ satisfy (1), we must have

$$\lim_{n\to\infty} n \int \big[dP^{1/2}_{1/\sqrt{n},g_1} - dP^{1/2}_{1/\sqrt{n},g_2}\big]^2 = \frac{1}{4} \int \big[g_1 \, dP^{1/2} - g_2 \, dP^{1/2}\big]^2 = \frac{1}{4}\|g_1 - g_2\|_{P,2}^2. \tag{D.13}$$

Moreover, by Theorem 13.1.2 in Lehmann and Romano (2005), we can also conclude

$$
\begin{aligned}
&\frac{1}{2}\int \left|dP^n_{1/\sqrt{n},g_1} - dP^n_{1/\sqrt{n},g_2}\right| \\
&\quad \le \left\{1 - \left[\int \left\{dP^n_{1/\sqrt{n},g_1}\right\}^{1/2}\left\{dP^n_{1/\sqrt{n},g_2}\right\}^{1/2}\right]^2\right\}^{1/2} \\
&\quad = \left\{1 - \left[\int dP^{1/2}_{1/\sqrt{n},g_1}\, dP^{1/2}_{1/\sqrt{n},g_2}\right]^{2n}\right\}^{1/2} \\
&\quad = \left\{1 - \left[1 - \frac{1}{2}\int \left[dP^{1/2}_{1/\sqrt{n},g_1} - dP^{1/2}_{1/\sqrt{n},g_2}\right]^2\right]^{2n}\right\}^{1/2},
\end{aligned} \tag{D.14}
$$

where in the first equality we exploited that $P^n_{1/\sqrt{n},g_1}$ and $P^n_{1/\sqrt{n},g_2}$ are product measures, while the second equality follows from direct calculation. Thus, by (D.13) and (D.14),

$$
\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{2}\int \left|dP^n_{1/\sqrt{n},g_1} - dP^n_{1/\sqrt{n},g_2}\right| \\
&\quad \le \limsup_{n\to\infty}\left\{1 - \left[1 - \frac{1}{2n}\int n\left[dP^{1/2}_{1/\sqrt{n},g_1} - dP^{1/2}_{1/\sqrt{n},g_2}\right]^2\right]^{2n}\right\}^{1/2} \\
&\quad = \left\{1 - \exp\left\{-\frac{1}{4}\|g_1 - g_2\|^2_{P,2}\right\}\right\}^{1/2},
\end{aligned} \tag{D.15}
$$

which establishes the claim of the lemma. *Q.E.D.*

LEMMA D.2: *Let $\{P_n\}$, $\{Q_n\}$, $\{V_n\}$ be probability measures defined on a common space. If $\{dQ_n/dP_n\}$ is asymptotically tight under P_n and $\int |dP_n - dV_n| = o(1)$, then*

$$
\left|\frac{dQ_n}{dP_n} - \frac{dQ_n}{dV_n}\right| \overset{P_n}{\to} 0. \tag{D.16}
$$

PROOF: Throughout, let $\mu_n = P_n + Q_n + V_n$, note μ_n dominates P_n, Q_n, and V_n, and set $p_n \equiv dP_n/d\mu_n$, $q_n \equiv dQ_n/d\mu_n$, and $v_n \equiv dV_n/d\mu_n$. We then obtain

$$
\begin{aligned}
\int \left|\frac{dP_n}{dV_n} - 1\right| dV_n &= \int \left|\frac{p_n}{v_n} - 1\right| v_n\, d\mu_n = \int_{v_n>0}\left|\frac{p_n}{v_n} - \frac{v_n}{v_n}\right| v_n\, d\mu_n \\
&\le \int |p_n - v_n|\, d\mu_n = \int |dP_n - dV_n| = o(1),
\end{aligned} \tag{D.17}
$$

where the second to last equality follows by definition, and the final equality by assumption. Hence, by (D.17) and Markov's inequality, we obtain $dP_n/dV_n \overset{V_n}{\to} 1$. Moreover, since $\int |dV_n - dP_n| = o(1)$ implies $\{P_n\}$ and $\{V_n\}$ are mutually contiguous, we conclude

$$
\frac{dP_n}{dV_n} \overset{P_n}{\to} 1. \tag{D.18}
$$

Next, observe that for any continuous and bounded function $f : \mathbf{R} \to \mathbf{R}$, we have that

$$\begin{aligned}\int f\left(\frac{dQ_n}{dP_n} - \frac{dQ_n}{dV_n}\right) dP_n &= \int f\left(\frac{q_n}{p_n} - \frac{q_n}{v_n}\right) p_n \, d\mu_n \\ &= \int_{p_n > 0} f\left(\frac{q_n}{p_n}\left(1 - \frac{p_n}{v_n}\right)\right) p_n \, d\mu_n \\ &= \int f\left(\frac{dQ_n}{dP_n}\left(1 - \frac{dP_n}{dV_n}\right)\right) dP_n \to f(0),\end{aligned} \tag{D.19}$$

where the final result follows from (D.18), dQ_n/dP_n being asymptotically tight under P_n, and continuity and boundedness of f. Since (D.19) holds for any continuous and bounded f, we conclude $dQ_n/dP_n - dQ_n/dV_n$ converges in law (under P_n) to zero, and hence also in P_n probability, which establishes (D.16). *Q.E.D.*

LEMMA D.3: *Let $H \subseteq L_0^2(P)$, assume for each $g \in H$ there is a path $t \mapsto P_{t,g}$ with score g, recall $\mathcal{B}$ is the σ-algebra on $\mathbf{X}$, let $\mathcal{A}$ be the Borel σ-algebra on $\mathbf{R}$, and set*

$$\mathcal{E}_n \equiv \left(\mathbf{X}^n, \mathcal{B}^n, P^n_{1/\sqrt{n},g} : g \in H\right). \tag{D.20}$$

If $0 \in H$, $\{\psi_k\}_{k=1}^{d_P}$ is an orthonormal basis for $L_0^2(P)$, and Φ denotes the standard normal measure on $\mathbf{R}$, then $\mathcal{E}_n$ converges weakly to the dominated experiment $\mathcal{E}$

$$\mathcal{E} \equiv \left(\mathbf{R}^{d_P}, \mathcal{A}^{d_P}, Q_g : g \in H\right), \tag{D.21}$$

where for each $g \in H$, $Q_g(\cdot) = Q_0(\cdot - T(g))$ for $T(g) \equiv \{\int g\psi_k \, dP\}_{k=1}^{d_P}$ and $Q_0 = \bigotimes_{k=1}^{d_P} \Phi$.

PROOF: The conclusion of the lemma is well known (see, e.g., Section 8.2 in van der Vaart (1991)), but we were unable to find a concrete reference and hence we include its proof for completeness. Since the lemma is straightforward when the dimension of $L_0^2(P)$ is finite ($d_P < \infty$), we focus on the case $d_P = \infty$. To analyze $\mathcal{E}$, let

$$\ell^2 \equiv \left\{\{c_k\}_{k=1}^{\infty} \in \mathbf{R}^{\infty} : \sum_{k=1}^{\infty} c_k^2 < \infty\right\}, \tag{D.22}$$

and note that by Example 2.3.5 in Bogachev (1998), ℓ^2 is the Cameron–Martin space of Q_0.[1] Hence, since for any $g \in L_0^2(P)$ we have $\{\int g\psi_k \, dP\}_{k=1}^{\infty} \in \ell^2$ due to $\{\psi_k\}_{k=1}^{\infty}$ being an orthonormal basis for $L_0^2(P)$, Theorem 2.4.5 in Bogachev (1998) implies

$$Q_g \equiv Q_0\left(\cdot - T(g)\right) \ll Q_0 \tag{D.23}$$

for all $g \in L_0^2(P)$, and thus $\mathcal{E}$ is dominated by Q_0. Denoting an element of $\mathbf{R}^{\infty}$ by $\omega = \{\omega_k\}_{k=1}^{\infty}$, we then obtain from $\{\int g\psi_k \, dP\}_{k=1}^{\infty} \in \ell^2$ and the Martingale convergence theorem

[1] See page 44 in Bogachev (1998) for a definition of a Cameron–Martin space.

(see, e.g., Theorem 12.1.1 in Williams (1991)) that

$$Q_0\left(\omega : \lim_{n\to\infty} \sum_{k=1}^{n} \omega_k \int \psi_k g\, dP \text{ exists}\right) = 1, \tag{D.24}$$

$$\lim_{n\to\infty} \int \left(\sum_{k=n+1}^{\infty} \omega_k \int g\psi_k\, dP\right)^2 dQ_0(\omega) = 0. \tag{D.25}$$

Therefore, Example 2.3.5 and Corollary 2.4.3 in Bogachev (1998) yield, for any $g \in L_0^2(P)$,

$$\begin{aligned}\log\left(\frac{dQ_g}{dQ_0}(\omega)\right) &= \sum_{k=1}^{\infty} \omega_k \int g\psi_k\, dP - \frac{1}{2}\int\left(\sum_{k=1}^{\infty} \omega_k \int g\psi_k\, dP\right)^2 dQ_0(\omega) \\ &= \sum_{k=1}^{\infty} \omega_k \int g\psi_k\, dP - \frac{1}{2}\int g^2\, dP,\end{aligned} \tag{D.26}$$

where the right-hand side of the first equality is well defined Q_0 almost surely by (D.24), while the second equality follows from (D.25) and $\sum_{k=1}^{\infty}(\int g\psi_k\, dP)^2 = \int g^2\, dP$ due to $\{\psi_k\}_{k=1}^{\infty}$ being an orthonormal basis for $L_0^2(P)$.

Next, select an arbitrary finite subset $\{g_j\}_{j=1}^{J} \equiv I \subseteq H$ and vector $(\lambda_1, \dots, \lambda_J)' \equiv \lambda \in \mathbf{R}^J$. From result (D.26), we then obtain Q_0 almost surely that

$$\sum_{j=1}^{J} \lambda_j \log\left(\frac{dQ_{g_j}}{dQ_0}(\omega)\right) = \sum_{k=1}^{\infty} \omega_k \int \left(\sum_{j=1}^{J} \lambda_j g_j\right)\psi_k\, dP - \sum_{j=1}^{J} \frac{\lambda_j}{2}\int g_j^2\, dP. \tag{D.27}$$

In particular, we can conclude from Example 2.10.2 and Proposition 2.10.3 in Bogachev (1998) together with (D.25) and $\sum_{j=1}^{J} \lambda_j g_j \in L_0^2(P)$ that, under Q_0, we have

$$\sum_{j=1}^{J} \lambda_j \log\left(\frac{dQ_{g_j}}{dQ_0}\right) \sim N\left(-\sum_{j=1}^{J} \frac{\lambda_j}{2}\int g_j^2\, dP, \int\left(\sum_{j=1}^{J} \lambda_j g_j\right)^2 dP\right). \tag{D.28}$$

Thus, for $\mu_I \equiv \frac{1}{2}(\int g_1^2\, dP, \dots, \int g_J^2\, dP)'$ and $\Sigma_I \equiv \int (g_1, \dots, g_J)'(g_1, \dots, g_J)\, dP$, we have

$$\left(\log\left(\frac{dQ_{g_1}}{dQ_0}\right), \dots, \log\left(\frac{dQ_{g_J}}{dQ_0}\right)\right)' \sim N(-\mu_I, \Sigma_I), \tag{D.29}$$

under Q_0 due to (D.28) holding for arbitrary $\lambda \in \mathbf{R}^J$.

To obtain an analogous result for the sequence of experiments $\mathcal{E}_n$, let $\{X_i\}_{i=1}^{n} \sim P^n$ where $P^n \equiv \bigotimes_{i=1}^{n} P$. From Lemma 25.14 in van der Vaart (1998), we obtain, under P^n,

$$\sum_{i=1}^{n} \log\left(\frac{dP_{1/\sqrt{n},g_j}}{dP}(X_i)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_j(X_i) - \frac{1}{2}\int g_j^2\, dP + o_p(1) \tag{D.30}$$

for any $1 \le j \le J$. Thus, defining $P^n_{1/\sqrt{n},g_j} \equiv \bigotimes_{i=1}^{n} P_{1/\sqrt{n},g_j}$, we can conclude that

$$\left(\log\left(\frac{dP^n_{1/\sqrt{n},g_1}}{dP^n}\right), \dots, \log\left(\frac{dP^n_{1/\sqrt{n},g_J}}{dP^n}\right)\right)' \xrightarrow{L} N(-\mu_I, \Sigma_I) \tag{D.31}$$

under P^n by (D.30), the central limit theorem, and the definitions of μ_I and Σ_I. Furthermore, also note Lemma D.1 implies $\int |dP^n - dP^n_{1/\sqrt{n},0}| = o(1)$ and hence

$$\left(\frac{dP^n_{1/\sqrt{n},g_1}}{dP^n}, \dots, \frac{dP^n_{1/\sqrt{n},g_J}}{dP^n}\right)' = \left(\frac{dP^n_{1/\sqrt{n},g_1}}{dP^n_{1/\sqrt{n},0}}, \dots, \frac{dP^n_{1/\sqrt{n},g_J}}{dP^n_{1/\sqrt{n},0}}\right)' + o_p(1) \tag{D.32}$$

under P^n by Lemma D.2 and result (D.31). Thus, by (D.31) and (D.32), we obtain

$$\left(\log\left(\frac{dP^n_{1/\sqrt{n},g_1}}{dP^n_{1/\sqrt{n},0}}\right), \dots, \log\left(\frac{dP^n_{1/\sqrt{n},g_J}}{dP^n_{1/\sqrt{n},0}}\right)\right)' \xrightarrow{L} N(-\mu_I, \Sigma_I), \tag{D.33}$$

under P^n, and since $\int |dP^n - dP^n_{1/\sqrt{n},0}| = o(1)$ also under $P^n_{1/\sqrt{n},0}$. Hence, the lemma follows from (i) (D.29), (ii) (D.33), and (iii) $\{P^n_{1/\sqrt{n},g}\}$ and $\{P^n_{1/\sqrt{n},0}\}$ being mutually contiguous for any $g \in H$ by (D.30) and Corollary 12.3.1 in Lehmann and Romano (2005), which together verify the conditions of Lemma 10.2.1 in LeCam (1986). *Q.E.D.*

LEMMA D.4: *Let Assumption* 2.1(i) *hold*, $\mathbf{B}$ *be a Banach space*, *and* $\hat{\theta}_n$ *be an asymptotically linear estimator for* $\theta(P) \in \mathbf{B}$ *such that* $\sqrt{n}\{\hat{\theta}_n - \theta(P)\} \xrightarrow{L} \mathbb{D}$ *under* P^n *on* $\mathbf{B}$ *for some tight Borel* $\mathbb{D}$. *Then*: *for any function* $h \in L^2_0(P)$, $(\sqrt{n}\{\hat{\theta}_n - \theta(P)\}, \frac{1}{\sqrt{n}}\sum_{i=1}^n h(X_i))$ *converges in distribution under* P^n *on* $\mathbf{B} \times \mathbf{R}$.

PROOF: For notational simplicity, let $\eta(P) \equiv (\theta(P), 0) \in \mathbf{B} \times \mathbf{R}$ and define $\hat{\eta}_n \equiv (\hat{\theta}_n, \frac{1}{n}\sum_{i=1}^n h(X_i)) \in \mathbf{B} \times \mathbf{R}$. Further let $(\mathbf{B} \times \mathbf{R})^*$ denote the dual space of $\mathbf{B} \times \mathbf{R}$ and note that, for any $d^* \in (\mathbf{B} \times \mathbf{R})^*$, there are $b^*_{d^*} \in \mathbf{B}^*$ and $r^*_{d^*} \in \mathbf{R}$ such that $d^*((b, r)) = b^*_{d^*}(b) + r^*_{d^*}(r)$ for all $(b, r) \in \mathbf{B} \times \mathbf{R}$. For ν the influence function of $\hat{\theta}_n$, then define $\zeta_{d^*}(X_i) \equiv \{b^*_{d^*}(\nu(X_i)) + r^*_{d^*}(h(X_i))\}$ to obtain that under $\bigotimes_{i=1}^n P$, we have

$$d^*\big(\sqrt{n}\{\hat{\eta}_n - \eta(P)\}\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \zeta_{d^*}(X_i) + o_p(1) \tag{D.34}$$

by asymptotic linearity of $\hat{\theta}_n$. Thus, for any finite set $\{d^*_k\}_{k=1}^K \subset (\mathbf{B} \times \mathbf{R})^*$, we have

$$\big(d^*_1\big(\sqrt{n}\{\hat{\eta}_n - \eta(P)\}\big), \dots, d^*_K\big(\sqrt{n}\{\hat{\eta}_n - \eta(P)\}\big)\big) \xrightarrow{L} (\mathbb{W}_{d^*_1}, \dots, \mathbb{W}_{d^*_K}) \tag{D.35}$$

for $(\mathbb{W}_{d^*_1}, \dots, \mathbb{W}_{d^*_K})$ a multivariate normal random variable satisfying $E[\mathbb{W}_{d^*_k}] = 0$ for all $1 \le k \le K$ and $E[\mathbb{W}_{d^*_j}\mathbb{W}_{d^*_k}] = E[\zeta_{d^*_j}(X_i)\zeta_{d^*_k}(X_i)]$ for all $1 \le j \le k \le K$.

Next, note that since $\sqrt{n}\{\hat{\theta}_n - \theta(P)\}$ is asymptotically measurable and asymptotically tight by Lemma 1.3.8 in van der Vaart and Wellner (1996), it follows that $\sqrt{n}\{\hat{\eta}_n - \eta(P)\}$ is asymptotically measurable and asymptotically tight on $\mathbf{B} \times \mathbf{R}$ by Lemmas 1.4.3 and 1.4.4 in van der Vaart and Wellner (1996). Hence, we conclude by Theorem 1.3.9 in van der Vaart and Wellner (1996) that any sequence $\{n_k\}$ has a subsequence $\{n_{k_j}\}$ with

$$\sqrt{n}_{k_j}\big\{\hat{\eta}_{n_{k_j}} - \eta(P)\big\} \xrightarrow{L} \mathbb{W} \tag{D.36}$$

under $\bigotimes_{i=1}^{n_{k_j}} P$ for $\mathbb{W}$ some tight Borel law on $\mathbf{B} \times \mathbf{R}$. However, letting $C_b(\mathbf{R}^K)$ denote the set of continuous and bounded functions on $\mathbf{R}^K$, we obtain from (D.35), (D.36), and the continuous mapping theorem that, for any $\{d_k^*\}_{k=1}^K \subset (\mathbf{B}\times\mathbf{R})^*$ and $f \in C_b(\mathbf{R}^K)$,

$$E\big[f\big(\big(d_1^*(\mathbb{W}),\dots,d_K^*(\mathbb{W})\big)\big)\big] = E\big[f\big((\mathbb{W}_{d_1^*},\dots,\mathbb{W}_{d_K^*})\big)\big]. \tag{D.37}$$

Since $\mathcal{G} \equiv \{f\circ(d_1^*,\dots,d_K^*) : f\in C_b(\mathbf{R}^K), \{d_k^*\}_{k=1}^K \subset (\mathbf{B}\times\mathbf{R})^*, 1\le K<\infty\}$ is a vector lattice that separates points in $\mathbf{B}\times\mathbf{R}$, Lemma 1.3.12 in van der Vaart and Wellner (1996) implies there is a unique tight Borel measure $\mathbb{W}$ on $\mathbf{B}\times\mathbf{R}$ satisfying (D.37). Thus, since the original sequence $\{n_k\}$ was arbitrary, we conclude all limit points of the law of $\sqrt{n}\{\hat{\eta}_n - \eta(P)\}$ coincide, and the lemma follows. *Q.E.D.*

APPENDIX E: TECHNICAL LEMMAS USED IN APPENDIX B AND APPENDIX C

In this appendix, we present technical lemmas that are used in Appendix B and Appendix C.

LEMMA E.1: *If $\mathbb{Z}_n \in \mathbf{B}$ is asymptotically tight and asymptotically measurable and satisfies $b^*(\mathbb{Z}_n) \xrightarrow{p} 0$ for any $b^*\in\mathbf{B}^*$, then it follows that $\mathbb{Z}_n = o_p(1)$ in $\mathbf{B}$.*

PROOF: For an arbitrary subsequence $\{n_j\}_{j=1}^\infty$, Theorem 1.3.9(ii) in van der Vaart and Wellner (1996) implies there exists a further subsequence $\{n_{j_k}\}_{k=1}^\infty$ along which $\mathbb{Z}_{n_{j_k}}$ converges in distribution to a tight limit $\mathbb{Z}$. Moreover, note that by the continuous mapping theorem, $b^*(\mathbb{Z}) = 0$ for all $b^*\in\mathbf{B}^*$. Therefore, letting $C_b(\mathbf{R}^K)$ denote the set of bounded and continuous functions on $\mathbf{R}^K$ and defining $\mathcal{G}\equiv\{f\circ(b_1^*,\dots,b_K^*) : f\in C_b(\mathbf{R}^K), \{b_k^*\}_{k=1}^K \subset \mathbf{B}^*, 1\le K<\infty\}$, we then obtain, for any $g\in\mathcal{G}$,

$$E\big[g(\mathbb{Z})\big] = g(0). \tag{E.1}$$

In particular, since $\mathcal{G}$ is a vector lattice that contains the constant functions and separates points in $\mathbf{B}$, Lemma 1.3.12 in van der Vaart and Wellner (1996) implies $\mathbb{Z}=0$ almost surely. We conclude that $\mathbb{Z}_{n_{j_k}}$ converges in probability to zero along $\{n_{j_k}\}_{k=1}^\infty$. Thus, since the original subsequence $\{n_j\}_{j=1}^\infty$ was arbitrary, it follows that $\mathbb{Z}_n = o_p(1)$. *Q.E.D.*

LEMMA E.2: *Let Assumptions* 2.1(i) *and* 3.1 *hold, and for any $g\in L_0^2(P)$ define $\Delta_g : \mathbf{T}\to\mathbf{R}$ to be given by $\Delta_g(\tau)\equiv\int s_\tau g\,dP$. It then follows that for any path $t\mapsto P_{t,g}\in\mathcal{M}$,*

$$\hat{\mathbb{G}}_n \xrightarrow{L_{n,g}} \mathbb{G}_0 + \Delta_g \quad \text{in } \ell^\infty(\mathbf{T}). \tag{E.2}$$

Further, under Assumption 2.1(ii), *$\Delta_g = 0$ whenever $\Pi_S(g)=0$ where $S(P)=\{s_\tau\in\bar{T}(P)^\perp : \tau\in\mathbf{T}\}$.*

PROOF: We first note Lemma 25.14 in van der Vaart (1998) implies

$$\sum_{i=1}^n \log\bigg(\frac{dP_{1/\sqrt{n},g}}{dP}(X_i)\bigg) = \frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i) - \frac{1}{2}\int g^2\,dP + o_p(1) \tag{E.3}$$

under P^n for any path $t\mapsto P_{t,g}\in\mathcal{M}$. Thus, by Example 3.10.6 in van der Vaart and Wellner (1996), P^n and $P^n_{1/\sqrt{n},g}$ are mutually contiguous, and hence applying Lemma D.4 and Lemma A.8.6 in Bickel et al. (1993) yields that, for any path $t\mapsto P_{t,g}\in\mathcal{M}$,

$$\hat{\mathbb{G}}_n \xrightarrow{L_{n,g}} \mathbb{G}_0 + \Delta_g \quad \text{in } \ell^\infty(\mathbf{T}), \tag{E.4}$$

which establishes (E.2). Moreover, if $t \mapsto P_{t,g} \in \mathbf{P}$, then by definition $g \in T(P)$ and hence $\int g s_\tau \, dP = 0$ for all $\tau \in \mathbf{T}$ due to $s_\tau \in \bar{T}(P)^\perp$ by Assumption 3.1(i). More generally, $\Delta_g = 0$ for any path $t \mapsto P_{t,g} \in \mathcal{M}$ with $\Pi_S(g) = 0$. *Q.E.D.*

LEMMA E.3: *Let Assumptions* 2.1(i) *and* 3.1 *hold, and for any* $g \in L_0^2(P)$ *define* $\Delta_g : \mathbf{T} \to \mathbf{R}$ *by* $\Delta_g(\tau) \equiv \int s_\tau g \, dP$. *It then follows that* Δ_g *is in the support of* $\mathbb{G}_0$.

PROOF: Define $\mathcal{S} \equiv S(P) = \{s_\tau : \tau \in \mathbf{T}\}$ and let the map $B : \mathbf{T} \to \mathcal{S}$ be given by $B(\tau) = s_\tau$ for any $\tau \in \mathbf{T}$. In addition, for any $s \in \mathcal{S}$, we define a selection $E : \mathcal{S} \to \mathbf{T}$ that assigns to each $s \in \mathcal{S}$ a unique element $E(s) \in B^{-1}(s)$. Our first goal is to show

$$P\Bigl(\sup_{s \in \mathcal{S}} \sup_{\tau \in B^{-1}(s)} \bigl|\mathbb{G}_0(\tau) - \mathbb{G}_0\bigl(E(s)\bigr)\bigr| = 0\Bigr) = 1, \tag{E.5}$$

and to this end we fix $\epsilon, \eta > 0$, and note that since $\mathbb{G}_0$ is tight by Assumption 3.1(ii), we obtain by Lemma 1.3.8 in van der Vaart and Wellner (1996) that $\hat{\mathbb{G}}_n$ is asymptotically tight. Thus, since $\mathbb{G}_0$ is Gaussian, Theorem 1.5.7 in van der Vaart and Wellner (1996) implies that, for any $\epsilon, \eta > 0$, there exists a $\delta(\epsilon, \eta) > 0$ such that

$$\limsup_{n \to \infty} P\Bigl(\sup_{\tau_1, \tau_2 : \|s_{\tau_1} - s_{\tau_2}\|_{P,2} < \delta(\epsilon, \eta)} \bigl|\hat{\mathbb{G}}_n(\tau_1) - \hat{\mathbb{G}}_n(\tau_2)\bigr| > \epsilon\Bigr) < \eta. \tag{E.6}$$

Moreover, since $\|s_{E(s)} - s_\tau\|_{P,2} = 0$ for any $\tau \in B^{-1}(s)$, by the Portmanteau Theorem (see, e.g., Theorem 1.3.4(ii) in van der Vaart and Wellner (1996)), we obtain

$$\begin{aligned} &P\Bigl(\sup_{s \in \mathcal{S}} \sup_{\tau \in B^{-1}(s)} \bigl|\mathbb{G}_0(\tau) - \mathbb{G}_0\bigl(E(s)\bigr)\bigr| > \epsilon\Bigr) \\ &\quad \le P\Bigl(\sup_{\tau_1, \tau_2 : \|s_{\tau_1} - s_{\tau_2}\|_{P,2} < \delta(\epsilon, \eta)} \bigl|\mathbb{G}_0(\tau_1) - \mathbb{G}_0(\tau_2)\bigr| > \epsilon\Bigr) < \eta. \end{aligned} \tag{E.7}$$

In particular, since $\epsilon, \eta > 0$ were arbitrary, we conclude that (E.5) holds by result (E.7) and the monotone convergence theorem.

Next, define a map $\Upsilon : \ell^\infty(\mathbf{T}) \to \ell^\infty(\mathcal{S})$ to be given by $\Upsilon(f)(s) = f(E(s))$ for any $f \in \ell^\infty(\mathbf{T})$, and note that Υ is linear and continuous. Thus, setting $\mathbb{S}_0 = \Upsilon(\mathbb{G}_0)$, we note $\mathbb{S}_0$ is a tight Gaussian process on $\ell^\infty(\mathcal{S})$, which by Assumption 3.1(i) satisfies

$$E\bigl[\mathbb{S}_0(s_1)\mathbb{S}_0(s_2)\bigr] = \int s_1 s_2 \, dP \tag{E.8}$$

for any $s_1, s_2 \in \mathcal{S}$. Similarly, let $S_g \in \ell^\infty(\mathcal{S})$ be given by $S_g \equiv \Upsilon(\Delta_g)$ and note that

$$S_g(s) = \Delta_g\bigl(E(s)\bigr) = \int g s_{E(s)} \, dP = \int g s \, dP \tag{E.9}$$

for any $s \in \mathcal{S}$. Further note that by Lemma 1.5.9 in van der Vaart and Wellner (1996) and Gaussianity of $\mathbb{G}_0$, $\mathbf{T}$ is totally bounded under the semimetric $d(\tau_1, \tau_2) \equiv \|s_{\tau_1} - s_{\tau_2}\|_{P,2}$ and the sample paths of $\mathbb{G}_0$ are almost surely uniformly continuous with respect to $d(\cdot, \cdot)$. It follows that $\mathbb{S}_0$ is almost surely uniformly continuous on $\mathcal{S}$ with respect to $\|\cdot\|_{P,2}$, which implies its sample paths almost surely admit a unique extension to $\bar{\mathcal{S}}$ for $\bar{\mathcal{S}}$ the closure of $\mathcal{S}$ under $\|\cdot\|_{P,2}$, and thus we may view $\mathbb{S}_0$ as an element of the space

$$C(\bar{\mathcal{S}}) \equiv \{S : \bar{\mathcal{S}} \to \mathbf{R} \text{ that are continuous under } \|\cdot\|_{P,2}\}. \tag{E.10}$$

Moreover, since $\mathbf{T}$ being totally bounded under $d(\cdot,\cdot)$ implies $\mathcal{S}$ is totally bounded under $\|\cdot\|_{P,2}$, it follows that $\bar{\mathcal{S}}$ is compact under $\|\cdot\|_{P,2}$. Since $\mathbb{S}_0$ is Radon by Theorem A.3.11 in Bogachev (1998), we can conclude from (E.8) and (E.9) and Lemma E.4 (below) that S_g belongs to the support of $\mathbb{S}_0$. In particular, we conclude, for any $\epsilon > 0$, that

$$0 < P(\|S_g - \mathbb{S}_0\|_\infty > \epsilon) = P(\|Y(\Delta_g) - Y(\mathbb{G}_0)\|_\infty > \epsilon) = P(\|\Delta_g - \mathbb{G}_0\|_\infty > \epsilon), \tag{E.11}$$

where the first equality follows by definition of $Y : \ell^\infty(\mathbf{T}) \to \ell^\infty(\mathcal{S})$, and the second equality is implied by (E.5) and $\Delta_g(\tau) = Y(\Delta_g)(s)$ for any $\tau \in B^{-1}(s)$. Thus, since $\epsilon > 0$ was arbitrary, we conclude from (E.11) that Δ_g is in the support of $\mathbb{G}_0$. *Q.E.D.*

LEMMA E.4: *Let Assumption* 2.1(i) *hold, $\mathcal{S} \subset L^2_0(P)$ be compact under $\|\cdot\|_{P,2}$, for any $g \in L^2_0(P)$ let $S_g : \mathcal{S} \to \mathbf{R}$ be given by $S_g(s) = \int sg\,dP$, and define*

$$C(\mathcal{S}) \equiv \{S : \mathcal{S} \to \mathbf{R} \textit{ is continuous under } \|\cdot\|_{P,2}\}, \tag{E.12}$$

which is endowed with the norm $\|S\|_\infty = \sup_{s\in\mathcal{S}} |S(s)|$. If $\mathbb{S}_0$ is a centered Radon Gaussian measure on $C(\mathcal{S})$ satisfying $E[\mathbb{S}_0(s_1)\mathbb{S}_0(s_2)] = \int s_1 s_2\,dP$ for any $s_1, s_2 \in \mathcal{S}$, then it follows that S_g belongs to the support of $\mathbb{S}_0$ for any $g \in L^2_0(P)$.

PROOF: Fix $g \in L^2_0(P)$ and note the Cauchy–Schwarz inequality yields

$$|S_g(s_1) - S_g(s_2)| \leq \int |g||s_1 - s_2|dP \leq \|g\|_{P,2}\|s_1 - s_2\|_{P,2} \tag{E.13}$$

for any $s_1, s_2 \in \mathcal{S}$, and therefore $S_g \in C(\mathcal{S})$. Let $\bar{V}$ denote the closure of the linear span of $\mathcal{S}$ in $L^2_0(P)$ and set $\Pi_{\bar{V}}(g)$ to equal the metric projection of g onto $\bar{V}$. For any $s \in \mathcal{S}$, then define $S_{\Pi_{\bar{V}}(g)} : \mathcal{S} \to \mathbf{R}$ by $S_{\Pi_{\bar{V}}(g)}(s) = \int \{\Pi_{\bar{V}}(g)\} s\,dP$ and note that

$$S_g(s) = \int gs\,dP = \int \{\Pi_{\bar{V}}(g)\} s\,dP = S_{\Pi_{\bar{V}}(g)}(s). \tag{E.14}$$

Moreover, since $\Pi_{\bar{V}}(g) \in \bar{V}$, it follows that there is a sequence $\{g_k\}_{k=1}^\infty$ such that

$$\lim_{k\to\infty} \|g_k - \Pi_{\bar{V}}(g)\|_{P,2} = 0, \tag{E.15}$$

where each g_k satisfies, for some $\{\alpha_{j,k}, s_{j,k}\}_{j=1}^k$ with $(\alpha_{j,k}, s_{j,k}) \in \mathbf{R} \times \mathcal{S}$, the relation

$$g_k = \sum_{j=1}^k \alpha_{j,k} s_{j,k}. \tag{E.16}$$

Defining $S_{g_k} : \mathcal{S} \to \mathbf{R}$ by $S_{g_k}(s) = \int g_k s\, dP$, we then conclude from results (E.14) and (E.15) together with the Cauchy–Schwarz inequality that

$$\begin{aligned}\lim_{k\to\infty} \|S_{g_k} - S_g\|_\infty &= \lim_{k\to\infty} \|S_{g_k} - S_{\Pi_{\bar{V}}(g)}\|_\infty \le \lim_{k\to\infty} \sup_{s\in\mathcal{S}} \int |s|\big|g_k - \Pi_{\bar{V}}(g)\big|\, dP \\ &\le \sup_{s\in\mathcal{S}} \|s\|_{P,2} \times \lim_{k\to\infty} \big\|g_k - \Pi_{\bar{V}}(g)\big\|_{P,2} = 0,\end{aligned} \tag{E.17}$$

where in the final equality we exploited that $\sup_{s\in\mathcal{S}} \|s\|_{P,2} < \infty$ since $\mathcal{S}$ is compact under $\|\cdot\|_{P,2}$ by hypothesis. In particular, since the topological support of $\mathbb{S}_0$ is a closed subset of $C(\mathcal{S})$, result (E.17) implies that to establish the lemma, it suffices to show S_{g_k} belongs to the support of $\mathbb{S}_0$ for all k. To this end, we let $\mathrm{ca}(\mathcal{S})$ denote the set of finite signed Borel (w.r.t. $\|\cdot\|_{P,2}$) measures on $\mathcal{S}$, and note that by Theorem 14.15 in Aliprantis and Border (2006), it follows $\mathrm{ca}(\mathcal{S})$ is the dual space of $C(\mathcal{S})$. Next, for any k, we define a measure $\nu_k \in \mathrm{ca}(\mathcal{S})$ by setting, for each Borel set $A \subseteq C(\mathcal{S})$,

$$\nu_k(A) = \sum_{j=1}^{k} \alpha_{j,k} 1\{s_{j,k} \in A\}. \tag{E.18}$$

Following the notation in Bogachev (1998), for any k we additionally introduce the linear map $R(\nu_k) : \mathrm{ca}(\mathcal{S}) \to \mathbf{R}$ which, for any $\mu \in \mathrm{ca}(\mathcal{S})$, is given by

$$R(\nu_k)(\mu) = E\bigg[\bigg\{\int \mathbb{S}_0(s)\nu_k(ds)\bigg\}\bigg\{\int \mathbb{S}_0(s)\mu(ds)\bigg\}\bigg]. \tag{E.19}$$

By results (E.18) and (E.19), Fubini's theorem (see, e.g., Corollary 3.4.2 in Bogachev (2007)), and $E[\mathbb{S}_0(s_1)\mathbb{S}_0(s_2)] = \int s_1 s_2\, dP$ for any $s_1, s_2 \in \mathcal{S}$, we then obtain

$$\begin{aligned}R(\nu_k)(\mu) &= E\Bigg[\Bigg\{\sum_{j=1}^{k} \alpha_{j,k}\mathbb{S}_0(s_{j,k})\Bigg\}\bigg\{\int \mathbb{S}_0(s)\mu(ds)\bigg\}\Bigg] \\ &= \int E\Bigg[\Bigg\{\sum_{j=1}^{k} \alpha_{j,k}\mathbb{S}_0(s_{j,k})\Bigg\}\mathbb{S}_0(s)\Bigg]\mu(ds) = \int S_{g_k}(s)\mu(ds),\end{aligned} \tag{E.20}$$

where the last equality follows from (E.16). Result (E.20) implies we may identify the linear map $R(\nu_k) : \mathrm{ca}(\mathcal{S}) \to \mathbf{R}$ with S_{g_k}, and therefore Theorem 3.2.3 in Bogachev (1998) implies S_{g_k} is in the Cameron–Martin space of $\mathbb{S}_0$. However, by Theorem 3.6.1 in Bogachev (1998), the Cameron–Martin space of $\mathbb{S}_0$ is a subset of its support, and hence we conclude S_{g_k} is in the support of $\mathbb{S}_0$. The lemma then follows from (E.17). *Q.E.D.*

LEMMA E.5: *Let $\mathbb{G}_0$ be a centered Gaussian measure on a separable Banach space $\mathbf{B}$ and $0 \neq \Delta \in \mathbf{B}$ belong to the support of $\mathbb{G}_0$. Further suppose $\Psi : \mathbf{B} \to \mathbf{R}_+$ is continuous, convex, and nonconstant, and satisfies $\Psi(0) = 0$, $\Psi(b) = \Psi(-b)$ for all $b \in \mathbf{B}$, and $\{b \in \mathbf{B} : \Psi(b) \le t\}$ is bounded for any $0 < t < \infty$. For any finite $t > 0$, it then follows that*

$$P\big(\Psi(\mathbb{G}_0 + \Delta) < t\big) < P\big(\Psi(\mathbb{G}_0) < t\big).$$

PROOF: Let $\|\cdot\|_{\mathbf{B}}$ denote the norm of $\mathbf{B}$, fix $t > 0$, and define

$$C \equiv \{b \in \mathbf{B} : \Psi(b) < t\}. \tag{E.21}$$

For $\mathbf{B}^*$ the dual space of $\mathbf{B}$, let $\|\cdot\|_{\mathbf{B}^*}$ denote its norm, and $\nu_C : \mathbf{B}^* \to \mathbf{R}$ be given by

$$\nu_C(b^*) = \sup_{b \in C} b^*(b), \tag{E.22}$$

which constitutes the support functional of C. Then note that, for any $b^* \in \mathbf{B}^*$, we have

$$\nu_C(-b^*) = \sup_{b \in C} -b^*(b) = \sup_{b \in C} b^*(-b) = \sup_{b \in -C} b^*(b) = \nu_C(b^*), \tag{E.23}$$

due to $C = -C$ since $\Psi(b) = \Psi(-b)$ for all $b \in \mathbf{B}$. Moreover, note that $0 \in C$ since $\Psi(0) = 0 < t$, and hence there exists a $M_0 > 0$ such that $\{b \in \mathbf{B} : \|b\|_{\mathbf{B}} \leq M_0\} \subseteq C$ by continuity of Ψ. Thus, by definition of $\|\cdot\|_{\mathbf{B}^*}$ we obtain, for any $b^* \in \mathbf{B}^*$, that

$$\nu_C(b^*) = \sup_{b \in C} b^*(b) \geq \sup_{\|b\|_{\mathbf{B}} \leq M_0} b^*(b) = M_0 \times \sup_{\|b\|_{\mathbf{B}} \leq 1} |b^*(b)| = M_0 \|b^*\|_{\mathbf{B}^*}. \tag{E.24}$$

Analogously, note that by assumption, $M_1 \equiv \sup_{b \in C} \|b\|_{\mathbf{B}} < \infty$, and thus for any $b^* \in \mathbf{B}^*$,

$$\nu_C(b^*) = \sup_{b \in C} b^*(b) \leq \|b^*\|_{\mathbf{B}^*} \times \sup_{b \in C} \|b\|_{\mathbf{B}} = M_1 \|b^*\|_{\mathbf{B}^*}. \tag{E.25}$$

We next aim to define a norm on $\mathbf{B}$ under which C is the open unit sphere. To this end, recall that the original norm $\|\cdot\|_{\mathbf{B}}$ on $\mathbf{B}$ may be written as

$$\|b\|_{\mathbf{B}} = \sup_{\|b^*\|_{\mathbf{B}^*} = 1} b^*(b); \tag{E.26}$$

see, for instance, Lemma 6.10 in Aliprantis and Border (2006). Similarly, instead define

$$\|b\|_{\mathbf{B},C} \equiv \sup_{\|b^*\|_{\mathbf{B}^*} = 1} \frac{b^*(b)}{\nu_C(b^*)}, \tag{E.27}$$

and note that: (i) $\|b_1 + b_2\|_{\mathbf{B},C} \leq \|b_1\|_{\mathbf{B},C} + \|b_2\|_{\mathbf{B},C}$ for any $b_1, b_2 \in \mathbf{B}$ by direct calculation, (ii) $\|\alpha b\|_{\mathbf{B},C} = |\alpha| \|b\|_{\mathbf{B},C}$ for any $\alpha \in \mathbf{R}$ and $b \in \mathbf{B}$ by (E.23) and (E.27), and (iii) results (E.24), (E.25), (E.26), and (E.27) imply that, for any $b \in \mathbf{B}$,

$$M_0 \|b\|_{\mathbf{B},C} \leq \|b\|_{\mathbf{B}} \leq M_1 \|b\|_{\mathbf{B},C}, \tag{E.28}$$

which establishes $\|b\|_{\mathbf{B},C} = 0$ if and only if $b = 0$, and hence we conclude $\|\cdot\|_{\mathbf{B},C}$ is indeed a norm on $\mathbf{B}$. In fact, (E.28) implies that the norms $\|\cdot\|_{\mathbf{B}}$ and $\|\cdot\|_{\mathbf{B},C}$ are equivalent, and hence $\mathbf{B}$ remains a separable Banach space and its Borel σ-algebra unchanged when endowed with $\|\cdot\|_{\mathbf{B},C}$ in place of $\|\cdot\|_{\mathbf{B}}$.

Next, note that the continuity of Ψ implies C is open, and thus, for any $b_0 \in C$, there is an $\epsilon > 0$ such that $\{b \in \mathbf{B} : \|b - b_0\|_{\mathbf{B}} \leq \epsilon\} \subset C$. We then obtain

$$\nu_C(b^*) \geq \sup_{\|b - b_0\|_{\mathbf{B}} \leq \epsilon} b^*(b) = \sup_{\|b\|_{\mathbf{B}} \leq 1} \{b^*(b_0) + \epsilon b^*(b)\} = b^*(b_0) + \epsilon \|b^*\|_{\mathbf{B}^*}, \tag{E.29}$$

where the final equality follows as in (E.24). Thus, from (E.25) and (E.29), we conclude $1-\epsilon/M_1 \geq b^*(b_0)/\nu_C(b^*)$ for all b^* with $\|b^*\|_{\mathbf{B}^*}=1$, and hence we conclude

$$C \subseteq \big\{b \in \mathbf{B} : \|b\|_{\mathbf{B},C} < 1\big\}. \tag{E.30}$$

Suppose, on the other hand, that $\|b_0\|_{\mathbf{B},C} < 1$, and note (E.27) implies, for some $\delta > 0$,

$$b^*(b_0) < \nu_C\big(b^*\big)(1-\delta) \tag{E.31}$$

for all $b^* \in \mathbf{B}^*$ with $\|b^*\|_{\mathbf{B}^*} = 1$. Setting $\eta \equiv \delta M_0$ and arguing as in (E.29) then yields

$$\begin{aligned} &\sup_{\|b^*\|_{\mathbf{B}^*}=1}\ \sup_{\|b-b_0\|_{\mathbf{B}}\leq\eta} \big\{b^*(b)-\nu_C\big(b^*\big)\big\} \\ &\quad= \sup_{\|b^*\|_{\mathbf{B}^*}=1} \big\{b^*(b_0)+\eta\big\|b^*\big\|_{\mathbf{B}^*}-\nu_C\big(b^*\big)\big\} \\ &\quad< \sup_{\|b^*\|_{\mathbf{B}^*}=1} \big\{\eta-\nu_C\big(b^*\big)\delta\big\} = \sup_{\|b^*\|_{\mathbf{B}^*}=1} \delta\big(M_0-\nu_C\big(b^*\big)\big) \leq 0, \end{aligned} \tag{E.32}$$

where the first inequality follows from (E.31), the second equality by definition of η, and the final inequality follows from (E.24). Since C is convex by hypothesis, (E.32) and Theorem 5.12.5 in Luenberger (1969) imply $\{b \in \mathbf{B} : \|b-b_0\|_{\mathbf{B}} \leq \eta\} \subseteq \bar{C}$. We conclude b_0 is in the interior of $\bar{C}$, and since C is convex and open, Lemma 5.28 in Aliprantis and Border (2006) yields that $b_0 \in C$. Thus, we can conclude that

$$\big\{b \in \mathbf{B} : \|b\|_{\mathbf{B},C} < 1\big\} \subseteq C, \tag{E.33}$$

which together with (E.30) yields $C = \{b \in \mathbf{B} : \|b\|_{\mathbf{B},C} < 1\}$. Therefore, $\mathbf{B}$ being separable under $\|\cdot\|_{\mathbf{B},C}$, $0 \neq \Delta$ being in the support of $\mathbb{G}_0$ by hypothesis, and Corollary 2 in Lewandowski, Ryznar, and Zak (1995) finally enable us to derive

$$P\big(\Psi(\mathbb{G}_0+\Delta) < t\big) = P(\mathbb{G}_0+\Delta \in C) < P(\mathbb{G}_0 \in C) = P\big(\Psi(\mathbb{G}_0) < t\big), \tag{E.34}$$

which establishes the claim of the lemma. *Q.E.D.*

LEMMA E.6: *Suppose Assumption* 5.1 *holds,* $f : \mathbf{X} \to \mathbf{R}$ *is bounded, and let* $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbf{R}$ *be an estimator of* $\int f\,dP$. *If* $\hat{\theta}_n$ *is such that, for any path* $t \mapsto P_{t,g} \in \mathbf{P}$,

$$\sqrt{n}\left\{\hat{\theta}_n - \int f\,dP_{1/\sqrt{n},g}\right\} \overset{L_{n,g}}{\to} \mathbb{Z}_g \tag{E.35}$$

for some tight law $\mathbb{Z}_g$, *then* (E.35) *holds for any path* $t \mapsto P_{t,g} \in \mathcal{M}$ *with* $g \in \bar{T}(P)$. *Moreover, if a sequence* $\{g_j\}_{j=1}^\infty \subseteq T(P)$ *satisfies* $\|g_j - g_0\|_{P,2} = o(1)$ *for some* $g_0 \in \bar{T}(P)$, *then it follows that* $\mathbb{Z}_{g_j} \to \mathbb{Z}_{g_0}$ *in the weak topology.*

PROOF: Fix a score $g_0 \in \bar{T}(P)$, select a sequence $\{g_j\}_{j=1}^\infty \subseteq T(P)$ with $\|g_j - g_0\|_{P,2} = o(1)$, and set Δ_j to equal

$$\Delta_j \equiv 2\left\{1-\exp\left\{-\frac{1}{4}\|g_0-g_j\|_{P,2}\right\}\right\}^{1/2}, \tag{E.36}$$

and note $\Delta_j = o(1)$. We further observe that since $f : \mathbf{X} \to \mathbf{R}$ is bounded, we obtain

$$\begin{aligned}
&\left|\sqrt{n}\left\{\int f\,dP_{1/\sqrt{n},g} - \int f\,dP\right\} - \int fg\,dP\right| \\
&\quad \leq \|f\|_\infty \int \left|\frac{g}{2}\,dP^{1/2}\left(dP^{1/2}_{1/\sqrt{n},g} - dP^{1/2}\right)\right| \\
&\qquad + \|f\|_\infty \int \left|\sqrt{n}\left\{dP^{1/2}_{1/\sqrt{n},g} - dP^{1/2}\right\} - \frac{g}{2}\,dP^{1/2}\right|\left(dP^{1/2}_{1/\sqrt{n},g} + dP^{1/2}\right)
\end{aligned} \tag{E.37}$$

for any path $t \mapsto P_{t,g}$. In particular, result (E.37) and the Cauchy–Schwarz inequality imply $t \mapsto \int f\,dP_{t,g}$ has pathwise derivative $\int fg\,dP$ at $t = 0$. Hence, since $g_j \in T(P)$ implies there exists a path $t \mapsto P_{t,g_j} \in \mathbf{P}$ such that (E.35) holds, we obtain

$$\sqrt{n}\left\{\hat{\theta}_n - \int f\,dP_{1/\sqrt{n},g_0}\right\} \overset{L_{n,g_j}}{\to} \mathbb{Z}_{g_j} + \int f(g_j - g_0)\,dP \tag{E.38}$$

for any j by the pathwise differentiability of $t \mapsto \int f\,dP_{t,g}$ and the continuous mapping theorem. For any continuous and bounded map $F : \mathbf{R} \to \mathbf{R}$, we then note

$$\begin{aligned}
&\limsup_{n\to\infty} \int F\left(\sqrt{n}\left\{\hat{\theta}_n - \int f\,dP_{1/\sqrt{n},g_0}\right\}\right) dP_{1/\sqrt{n},g_0} \\
&\quad \leq \limsup_{n\to\infty} \int F\left(\sqrt{n}\left\{\hat{\theta}_n - \int f\,dP_{1/\sqrt{n},g_0}\right\}\right) dP_{1/\sqrt{n},g_j} + \|F\|_\infty \Delta_j \\
&\quad = \liminf_{n\to\infty} \int F\left(\sqrt{n}\left\{\hat{\theta}_n - \int f\,dP_{1/\sqrt{n},g_0}\right\}\right) dP_{1/\sqrt{n},g_j} + \|F\|_\infty \Delta_j \\
&\quad \leq \liminf_{n\to\infty} \int F\left(\sqrt{n}\left\{\hat{\theta}_n - \int f\,dP_{1/\sqrt{n},g_0}\right\}\right) dP_{1/\sqrt{n},g_0} + 2\|F\|_\infty \Delta_j,
\end{aligned} \tag{E.39}$$

where the inequalities follow from (E.36) and Lemma D.1, and the equality from the limit existing by (E.38) and $F : \mathbf{R} \to \mathbf{R}$ being continuous and bounded. Since $\Delta_j = o(1)$, (E.39) implies the following limit exists for any continuous and bounded $F : \mathbf{R} \to \mathbf{R}$:

$$L(F) \equiv \lim_{n\to\infty} \int F\left(\sqrt{n}\left\{\hat{\theta}_n - \int f\,dP_{1/\sqrt{n},g_0}\right\}\right) dP_{1/\sqrt{n},g_0}. \tag{E.40}$$

In addition, for any $\epsilon > 0$, there exists a $j(\epsilon)$ such that $\Delta_{j(\epsilon)} < \epsilon/2$ and, since $\mathbb{Z}_{g_{j(\epsilon)}}$ is tight, a compact set K_ϵ such that $P(\mathbb{Z}_{g_{j(\epsilon)}} + \int f(g_{j(\epsilon)} - g_0)\,dP \in K_\epsilon) \geq 1 - \epsilon/2$. For any $\delta > 0$, let $K_\epsilon^\delta \equiv \{a \in \mathbf{R} : \inf_{b \in K_\epsilon} \|a - b\| < \delta\}$, and note Portmanteu's Theorem (see Theorem 1.3.4(ii) in van der Vaart and Wellner (1996)) (E.38), and Lemma D.1 yield

$$\begin{aligned}
&\liminf_{n\to\infty} P_{1/\sqrt{n},g_0}\left(\sqrt{n}\left\{\hat{\theta}_n - \int f\,dP_{1/\sqrt{n},g_0}\right\} \in K_\epsilon^\delta\right) \\
&\quad\geq \liminf_{n\to\infty} P_{1/\sqrt{n},g_{j(\epsilon)}}\left(\sqrt{n}\left\{\hat{\theta}_n - \int f\,dP_{1/\sqrt{n},g_0}\right\} \in K_\epsilon^\delta\right) - \Delta_{j(\epsilon)} \\
&\quad\geq P\left(\mathbb{Z}_{g_{j(\epsilon)}} + \int (g_{j(\epsilon)} - g_0)\,dP \in K_\epsilon\right) - \Delta_{j(\epsilon)} \\
&\quad\geq 1 - \epsilon.
\end{aligned} \tag{E.41}$$

Since ϵ was arbitrary, result (E.41) implies that the law of $\sqrt{n}\{\hat{\theta}_n - \int f\,dP_{1/\sqrt{n},g_0}\}$ under $P^n_{1/\sqrt{n},g_0}$ is asymptotically tight. Prohorov's theorem (see, e.g., Theorem 1.3.9 in van der Vaart and Wellner (1996)) then yields that every subsequence of $\sqrt{n}\{\hat{\theta}_n - \int f\,dP_{1/\sqrt{n},g_0}\}$ has a further subsequence that converges in distribution under $P^n_{1/\sqrt{n},g_0}$. However, in combination with result (E.40), these observations imply that $\sqrt{n}\{\hat{\theta}_n - \int f\,dP_{1/\sqrt{n},g_0}\}$ must itself converge in distribution under $P^n_{1/\sqrt{n},g_0}$, and we denote the limit law by $\mathbb{Z}_{g_0}$:

$$\sqrt{n}\left\{\hat{\theta}_n - \int f\,dP_{1/\sqrt{n},g_0}\right\} \stackrel{L_{n,g_0}}{\to} \mathbb{Z}_{g_0}. \tag{E.42}$$

Moreover, for any continuous and bounded $F : \mathbf{R} \to \mathbf{R}$, (E.38), (E.39), (E.42) imply

$$E\big[F(\mathbb{Z}_{g_0})\big] - \|F\|_\infty \Delta_j \leq E\left[F\left(\mathbb{Z}_{g_j} + \int f(g_j - g_0)\,dP\right)\right] \leq E\big[F(\mathbb{Z}_{g_0})\big] + \|F\|_\infty \Delta_j. \tag{E.43}$$

Since $\Delta_j = o(1)$, it therefore follows from (E.43) that $\mathbb{Z}_{g_j} + \int f(g_j - g_0)\,dP \to \mathbb{Z}_{g_0}$ in the weak topology. However, by the Cauchy–Schwarz inequality and $\|g_j - g_0\|_{P,2} = o(1)$, we can further conclude that $\int f(g_j - g_0)\,dP = o(1)$, and thus by the continuous mapping theorem we obtain that $\mathbb{Z}_{g_j} \to \mathbb{Z}_{g_0}$ in the weak topology. *Q.E.D.*

APPENDIX F: PROOFS OF MAIN RESULTS IN SECTION 4

In this appendix, we first provide the proofs for Theorem 4.1, Lemma 4.1, Lemma 4.2, and Corollary 4.1. We then establish a theorem (Theorem F.1) that includes Theorem 4.2 as a special case for models defined by sequential moment restrictions.

LEMMA F.1: *Let $\mathcal{F} \subset L_0^2(P)$ be such that $|f| \leq F$ for all $f \in \mathcal{F}$ and some $F \in L^2(P)$. Then, for any path $t \mapsto P_{t,g} \in \mathcal{M}$ satisfying $\int F^2\,dP_{t,g} = O(1)$, it follows that*

$$\limsup_{t\downarrow 0} \sup_{f\in\mathcal{F}} \left| \int \frac{f}{t}\{dP_{t,g} - dP\} - \int fg\,dP \right| = 0.$$

PROOF: Since $t \mapsto P_{t,g}$ is a path, the Cauchy–Schwarz inequality implies

$$\limsup_{t\downarrow 0} \sup_{f\in\mathcal{F}} \left| \int f\left[\frac{1}{t}\{dP_{t,g}^{1/2} - dP^{1/2}\} - \frac{g}{2}\,dP^{1/2}\right]\left(dP_{t,g}^{1/2} + dP^{1/2}\right) \right| = 0, \tag{F.1}$$

where we exploited that $F \in L^2(P)$ and $\int F^2\,dP_{t,g} = O(1)$ by hypothesis. Next, for any $M < \infty$, we obtain by the Cauchy–Schwarz and triangle inequalities that

$$\begin{aligned}\sup_{f\in\mathcal{F}}\left|\int fg\,dP^{1/2}\big(dP_{t,g}^{1/2} - dP^{1/2}\big)\right| &\leq M\left\{\int g^2\,dP\right\}^{1/2}\left\{\int \big(dP_{t,g}^{1/2} - dP^{1/2}\big)^2\right\}^{1/2} \\ &\quad + \left\{\int_{|F|>M} g^2\,dP\right\}^{1/2}\left\{\int F^2\big(dP_{t,g}^{1/2} - dP^{1/2}\big)^2\right\}^{1/2}.\end{aligned} \tag{F.2}$$

Therefore, result (F.2), $t \mapsto P_{t,g}$ being a path, $\int F^2\,dP_{t,g} = O(1)$ by hypothesis, $g \in L_0^2(P)$, and $P(|F(X)| > M)$ converging to zero as M diverges to infinity, yield

$$\begin{aligned}&\lim_{t\downarrow 0}\sup_{f\in\mathcal{F}}\left|\int fg\,dP^{1/2}\big(dP_{t,g}^{1/2} - dP^{1/2}\big)\right| \\ &\quad \leq \lim_{M\uparrow\infty}\left\{\int_{|F|>M} g^2\,dP\right\}^{1/2} \times \lim_{t\downarrow 0}\left\{\int F^2\big(dP_{t,g}^{1/2} - dP^{1/2}\big)^2\right\}^{1/2} = 0.\end{aligned} \tag{F.3}$$

Hence, results (F.1) and (F.3) and the triangle inequality together establish

$$\begin{aligned}&\lim_{t\downarrow 0}\sup_{f\in\mathcal{F}}\left|\frac{1}{t}\int f(dP_{t,g} - dP) - \int fg\,dP\right| \\ &\quad \leq \lim_{t\downarrow 0}\sup_{f\in\mathcal{F}}\left|\frac{1}{2}\int fg\,dP^{1/2}\big(dP_{t,g}^{1/2} - dP^{1/2}\big)\right| \\ &\qquad + \lim_{t\downarrow 0}\sup_{f\in\mathcal{F}}\left|\int f\left\{\frac{1}{t}\big(dP_{t,g}^{1/2} - dP^{1/2}\big) - \frac{1}{2}g\,dP^{1/2}\right\}\big(dP_{t,g}^{1/2} + dP^{1/2}\big)\right| = 0,\end{aligned} \tag{F.4}$$

and therefore the claim of the lemma follows. *Q.E.D.*

PROOF OF THEOREM 4.1: Consider any path $t \mapsto P_{t,g}$ satisfying Condition A, and an arbitrary bounded function $\psi_j \in L^2(W_j)$ for any $1 \leq j \leq J$. Then, from P satisfying (26) by Assumption 4.1(i) and $t \mapsto P_{t,g}$ satisfying Condition A(i), we obtain

$$\begin{aligned}0 &= \frac{1}{t}\left\{\int \rho_j(\cdot, h_t)\psi_j\,dP_{t,g} - \int \rho_j(\cdot, h_P)\psi_j\,dP\right\} \\ &= \frac{1}{t}\left\{\int \rho_j(\cdot, h_t)\psi_j(dP_{t,g} - dP) + \int \big(\rho_j(\cdot, h_t) - \rho_j(\cdot, h_P)\big)\psi_j\,dP\right\}.\end{aligned} \tag{F.5}$$

Furthermore, since the path $t \mapsto P_{t,g}$ satisfies Condition A(iii), we obtain, by Lemma F.1,

$$\lim_{t\downarrow 0}\frac{1}{t}\int \rho_j(\cdot, h_t)\psi_j(dP_{t,g} - dP) = \lim_{t\downarrow 0}\int \rho_j(\cdot, h_t)\psi_j g\,dP = \int \rho_j(\cdot, h_P)\psi_j g\,dP, \tag{F.6}$$

where the final equality follows by Assumption 4.1(iii), the Cauchy–Schwarz inequality, ψ_j being bounded, and $\|h_t - h_P\|_{\mathbf{H}} = o(1)$ by Condition A(ii). On the other hand, since $m_j(W_j, \cdot) : \mathbf{H} \to L^2(W_j)$ is Fréchet differentiable and $\|t^{-1}(h_t - h) - \Delta\|_{\mathbf{H}} = o(1)$ as $t \downarrow 0$

by Condition A(ii) for some $\Delta \in \mathbf{H}$, we can in addition conclude that

$$\lim_{t\downarrow 0} \frac{1}{t} E\big[\psi_j(W_j)\big\{m_j(W_j, h_t) - m_j(W_j, h_P)\big\}\big] = E\big[\psi_j(W_j)\nabla m_j(W_j, h_P)[\Delta]\big]. \tag{F.7}$$

Therefore, combining results (F.5), (F.6), and (F.7), we can obtain that any path $t \mapsto P_{t,g}$ satisfying Condition A must have a score $g \in L_0^2(P)$ satisfying the restriction

$$E\left[\left\{\sum_{j=1}^{J} \psi_j(W_j)\rho_j(Z, h_P)\right\} g(X)\right] = -E\left[\left\{\sum_{j=1}^{J} \psi_j(W_j)\nabla m_j(W_j, h_P)[\Delta]\right\}\right], \tag{F.8}$$

for any collection $(\psi_1, \dots, \psi_J) \in \bigotimes_{j=1}^J L^2(W_j)$ of bounded functions. However, since the set of bounded functions of W_j is dense in $L^2(W_j)$ for any $1 \le j \le J$, the Cauchy–Schwarz inequality and $E[\rho_j^2(Z, h_P)|W_j]$ being bounded almost surely by Assumption 4.2(i) imply that (F.8) actually holds for all $(\psi_1, \dots, \psi_J) \in \bigotimes_{j=1}^J L^2(W_j)$. In particular, we note that if we select $(\psi_1, \dots, \psi_J) \in \bar{\mathcal{R}}^\perp$, then the right-hand side of (F.8) is equal to zero, and therefore, result (F.8) implies the set inclusion

$$\left\{f \in L_0^2(P) : f = \sum_{j=1}^{J} \rho_j(Z, h_P)\psi_j(W_j) \text{ for some } (\psi_1, \dots, \psi_J) \in \bar{\mathcal{R}}^\perp\right\} \subseteq \bar{T}(P)^\perp. \tag{F.9}$$

In order to establish the theorem, we therefore only need to show the reverse inclusion in (F.9). As a preliminary result towards this goal, we first aim to establish that

$$\begin{aligned}&\left\{f \in L_0^2(P) : f = \sum_{j=1}^{J} \rho_j(Z, h_P)\psi_j(W_j) \text{ for some } (\psi_1, \dots, \psi_J) \in \bar{\mathcal{R}}^\perp\right\} \\ &\quad = \bar{\mathcal{V}} \cap \bar{T}(P)^\perp.\end{aligned} \tag{F.10}$$

To this end, note that by result (F.9) and the definition of $\bar{\mathcal{V}}$, we obtain the set inclusion

$$\begin{aligned}&\left\{f \in L_0^2(P) : f = \sum_{j=1}^{J} \rho_j(Z, h_P)\psi_j(W_j) \text{ for some } (\psi_1, \dots, \psi_J) \in \bar{\mathcal{R}}^\perp\right\} \\ &\quad \subseteq \bar{\mathcal{V}} \cap \bar{T}(P)^\perp.\end{aligned} \tag{F.11}$$

Next, we note that since $\bar{\mathcal{R}}$ is a closed linear subspace of $\bigotimes_{j=1}^J L^2(W_j)$, Theorem 3.4.1 in Luenberger (1969) implies we may decompose $\bigotimes_{j=1}^J L^2(W_j) = \bar{\mathcal{R}} \oplus \bar{\mathcal{R}}^\perp$. For any $(f_1, \dots, f_J) \in \bigotimes_{j=1}^J L^2(W_j)$, we in turn denote its projection onto $\bar{\mathcal{R}}$ and $\bar{\mathcal{R}}^\perp$ as $(\Pi_{\mathcal{R}} f_1, \dots, \Pi_{\mathcal{R}} f_J)$ and $(\Pi_{\mathcal{R}^\perp} f_1, \dots, \Pi_{\mathcal{R}^\perp} f_J)$, respectively. Selecting an arbitrary $f \in \mathcal{V} \cap \bar{T}(P)^\perp$, which by definition of $\mathcal{V}$ must be of the form $f = \sum_{j=1}^J \rho_j(\cdot, h_P)\psi_j^f$ for some $(\psi_1^f, \dots, \psi_J^f) \in \bigotimes_{j=1}^J L^2(W_j)$, we then observe that result (F.8) implies that, for any path $t \mapsto P_{t,g}$ satisfying Condition A, we must have the equality

$$E\left[\left\{\sum_{j=1}^{J} \psi_j^f(W_j)\rho_j(Z, h_P)\right\} g(X)\right] = -E\left[\sum_{j=1}^{J} \{\Pi_{\mathcal{R}} \psi_j^f(W_j)\}\nabla m_j(W_j, h_P)[\Delta]\right]. \tag{F.12}$$

However, by Assumption 4.1(iv), if $(\Pi_{\mathcal{R}}\psi_1^f, \dots, \Pi_{\mathcal{R}}\psi_J^f) \neq 0$ (in $\bigotimes_{j=1}^J L^2(W_j)$), then there is a path $t \mapsto P_{t,g}$ satisfying Condition A with $\|t^{-1}(h_t - h_P) - \Delta\|_{\mathbf{H}} = 0$ and

$$E\left[\sum_{j=1}^J \{\Pi_{\mathcal{R}}\psi_j^f(W_j)\}\nabla m_j(W, h_P)[\Delta]\right] \neq 0. \tag{F.13}$$

Therefore, if $f \in \mathcal{V}$ is such that $(\Pi_{\mathcal{R}}\psi_1^f, \dots, \Pi_{\mathcal{R}}\psi_J^f) \neq 0$, then results (F.12) and (F.13) establish that there exists a path $t \mapsto P_{t,g}$ satisfying Condition A and for which

$$E\left[\left\{\sum_{j=1}^J \psi_j^f(W_j)\rho_j(Z, h_P)\right\}g(X)\right] = -E\left[\sum_{j=1}^J \{\Pi_{\mathcal{R}}\psi_j^f(W_j)\}\nabla m_j(W, h_P)[\Delta]\right] \neq 0, \tag{F.14}$$

thus violating that $f \in \mathcal{V} \cap \bar{T}(P)^\perp$. In particular, it follows that any $f \in \mathcal{V} \cap \bar{T}(P)^\perp$ satisfies $(\Pi_{\mathcal{R}}\psi_1^f, \dots, \Pi_{\mathcal{R}}\psi_J^f) = 0$, and hence from $\bigotimes_{j=1}^J L^2(W_j) = \bar{\mathcal{R}} \oplus \bar{\mathcal{R}}^\perp$ we obtain

$$\begin{aligned}&\mathcal{V} \cap \bar{T}(P)^\perp \\ &\quad\subseteq \left\{f \in L_0^2(P) : f = \sum_{j=1}^J \rho_j(Z, h_P)\psi_j(W_j) \text{ for some } (\psi_1, \dots, \psi_J) \in \bar{\mathcal{R}}^\perp\right\}.\end{aligned} \tag{F.15}$$

Next, let $\bar{f} \in \bar{\mathcal{V}}$ be arbitrary, and note that $\|\bar{f} - \sum_{j=1}^J \rho_j(\cdot, h_P)\psi_j\|_{P,2}$ diverges to infinity as $\sum_{j=1}^J \|\psi_j\|_{P,2}$ diverges to infinity due to Assumption 4.2(ii). Thus, we obtain

$$\begin{aligned} 0 &= \inf_{(\psi_1,\dots,\psi_J)\in\bigotimes_{j=1}^J L^2(W_j)} \left\|\bar{f} - \sum_{j=1}^J \rho_j(\cdot, h_P)\psi_j\right\|_{P,2} \\ &= \min_{(\psi_1,\dots,\psi_J)\in\bigotimes_{j=1}^J L^2(W_j)} \left\|\bar{f} - \sum_{j=1}^J \rho_j(\cdot, h_P)\psi_j\right\|_{P,2}, \end{aligned} \tag{F.16}$$

where the first equality holds because $\bar{f} \in \bar{\mathcal{V}}$, and attainment in the second equality is implied by Proposition 38.14 in Zeidler (1984). However, attainment in (F.16) implies that $\bar{f} \in \mathcal{V}$, and hence, since $\bar{f} \in \bar{\mathcal{V}}$ was arbitrary, we can conclude $\bar{\mathcal{V}} = \mathcal{V}$. The claim in (F.10) then holds by result (F.11) and result (F.15).

In order to establish the theorem, we next aim to show that Assumption 4.1 implies

$$\bar{\mathcal{V}}^\perp \subseteq \bar{T}(P). \tag{F.17}$$

Selecting an arbitrary $g \in \bar{\mathcal{V}}^\perp \cap L^\infty(P)$, we define a path with density (w.r.t. P) equal to

$$\frac{dP_{t,g}}{dP} = 1 + tg, \tag{F.18}$$

which we note implies $P_{t,g}$ is indeed a probability measure for t small enough since $g \in L^\infty(P)$. The score of such a path is equal to g by direct calculation. Moreover, for any $\psi_j \in L^2(W_j)$ and $1 \leq j \leq J$, we have that $\rho_j(\cdot, h_P)\psi_j \in \mathcal{V}$ implies

$$\int \rho_j(\cdot, h_P)\psi_j \, dP_{t,g} = E\big[\rho_j(Z, h_P)\big(1 + tg(X)\big)\psi_j(W_j)\big] = 0, \tag{F.19}$$

where we exploited that $g \in \bar{\mathcal{V}}^{\perp}$ and $E[\rho_j(Z, h_P)|W_j] = 0$. Since $\psi_j \in L^2(W_j)$ was arbitrary, (F.19) in fact implies the path $t \mapsto P_{t,g}$ satisfies Condition A(i) with $h_t = h_P$ for all t, and hence also Conditions A(ii)–(iii). We conclude that $t \mapsto P_{t,g}$ satisfies Condition A, and as a result that $g \in \bar{T}(P)$. Since $g \in L^\infty(P) \cap \bar{\mathcal{V}}^{\perp}$ was arbitrary, it follows that $L^\infty(P) \cap \bar{\mathcal{V}}^{\perp} \subseteq \bar{T}(P)$, and by Assumption 4.1(v), that (F.17) indeed holds. Thus, we further obtain from result (F.17) that $\bar{T}(P)^{\perp} \subseteq (\bar{\mathcal{V}}^{\perp})^{\perp}$, and since $(\bar{\mathcal{V}}^{\perp})^{\perp} = \bar{\mathcal{V}}$ by Theorem 3.4.1 in Luenberger (1969), we can conclude that $\bar{T}(P)^{\perp} \subseteq \bar{\mathcal{V}}$. The theorem therefore follows from result (F.10). *Q.E.D.*

PROOF OF LEMMA 4.1: Recall that the map $\nabla m(W, h_P) : \mathbf{H} \to \bigotimes_{j=1}^J L^2(W_j)$ equals

$$\nabla m(W, h_P)[h] \equiv \big(\nabla m_1(W_1, h_P)[h], \dots, \nabla m_J(W_J, h_P)[h]\big)', \tag{F.20}$$

which is linear and continuous by the stated assumption that $m_j(W_j, \cdot) : \mathbf{H} \to L^2(W_j)$ is Fréchet differentiable at h_P for $1 \leq j \leq J$. Since $\bigotimes_{j=1}^J L^2(W_j)$ is its own dual, the adjoint $\nabla m(W, h_P)^*$ of the map $\nabla m(W, h_P)$ has domain $\bigotimes_{j=1}^J L^2(W_j)$. Moreover, because $\nabla m_j(W_j, h_P)^*$ is the adjoint of $\nabla m_j(W_j, h_P)$, it follows that

$$\nabla m(W, h_P)^*[f] = \sum_{j=1}^J \nabla m_j(W_j, h_P)^*[f_j] \tag{F.21}$$

for any $f = (f_1, \dots, f_J) \in \bigotimes_{j=1}^J L^2(W_j)$. Letting $\mathcal{N}(\nabla m(W, h_P)^*)$ denote the null space of $\nabla m(W, h_P)^* : \bigotimes_{j=1}^J L^2(W_j) \to \mathbf{H}^*$, and noting that $\mathcal{R}$ (as defined in (29)) equals the range of $\nabla m(W, h_P) : \mathbf{H} \to \bigotimes_{j=1}^J L^2(W_j)$, we obtain by Theorem 6.6.1 in Luenberger (1969) that, for $[\mathcal{R}]^{\perp}$ the orthocomplement of $\mathcal{R}$ in $\bigotimes_{j=1}^J L^2(W_j)$, we have

$$[\mathcal{R}]^{\perp} = \mathcal{N}\big(\nabla m(W, h_P)^*\big). \tag{F.22}$$

Furthermore, since $[\mathcal{R}]^{\perp} = \bar{\mathcal{R}}^{\perp}$ by continuity, and $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ if and only if $\bar{\mathcal{R}}^{\perp} = \{0\}$, Equation (F.22) yields $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ if and only if $\mathcal{N}(\nabla m(W, h_P)^*) = \{0\}$, which together with (F.21) establishes the lemma. *Q.E.D.*

PROOF OF LEMMA 4.2: Since each $\nabla m_j(W_j, h_P) : \mathbf{H} \to L^2(W_j)$ is linear and $\mathbf{H} = \bigotimes_{j=1}^J \mathbf{H}_j$, for each j there are linear maps $\nabla m_{j,k}(W_j, h_P) : \mathbf{H}_k \to L^2(W_j)$ such that

$$\nabla m_j(W_j, h_P)[h] = \sum_{k=1}^J \nabla m_{j,k}(W_j, h_P)[h_k] \tag{F.23}$$

for all $(h_1, \dots, h_J) = h \in \mathbf{H}$. Moreover, since $\nabla m_{j,k}(W_j, h_P)[h_k] = 0$ for any $h_k \in \mathbf{H}_k$ whenever $k > j$ by hypothesis, the decomposition in (F.23) implies that

$$\nabla m_j(W_j, h_P)[h] = \sum_{k=1}^j \nabla m_{j,k}(W_j, h_P)[h_k]. \tag{F.24}$$

We first suppose that $\bar{\mathcal{R}}_j = L^2(W_j)$ for all j and aim to show $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$. To this end, let $(f_1, \dots, f_J) \in \bigotimes_{j=1}^J L^2(W_j)$ and $\epsilon > 0$ be arbitrary. Then observe that

$$\left\| \nabla m_{1,1}(W_1, h_P)\left[h_1^*\right] - f_1 \right\|_{P,2} < \epsilon \tag{F.25}$$

for some $h_1^* \in \mathbf{H}_1$ since $\bar{\mathcal{R}}_1 = L^2(W_1)$. For $2 \le j \le J$, we may then exploit that $\bar{\mathcal{R}}_j = L^2(W_j)$ to inductively select $h_j^* \in \mathbf{H}_j$ to satisfy the inequality

$$\left\| \nabla m_{j,j}(W_j, h_P)\left[h_j^*\right] - \left(f_j - \sum_{k=1}^{j-1} \nabla m_{j,k}(W_j, h_P)\left[h_k^*\right] \right) \right\|_{P,2} < \epsilon. \tag{F.26}$$

Therefore, setting $(h_1^*, \dots, h_J^*) = h^* \in \mathbf{H}$ and employing (F.24) and (F.26), we obtain

$$\sum_{j=1}^J \left\| \nabla m_j(W_j, h_P)\left[h^*\right] - f_j \right\|_{P,2} < J\epsilon, \tag{F.27}$$

which, since $(f_1, \dots, f_J) \in \bigotimes_{j=1}^J L^2(W_j)$ and $\epsilon > 0$ were arbitrary, implies that $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$.

We next suppose $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ and aim to show $\bar{\mathcal{R}}_j = L^2(W_j)$ for all j. First note that by (F.24), it is immediate that $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ implies $\bar{\mathcal{R}}_1 = L^2(W_1)$. Thus, we focus on showing $\bar{\mathcal{R}}_j = L^2(W_j)$ for all $j > 1$. To this end, we select an arbitrary $1 < k^* \le J$ and $g^* \in L^2(W_{k^*})$, and define $(f_1^*, \dots, f_J^*)$ to satisfy $f_j^* = g^*$ if $j = k^*$ and $f_j^* = 0$ if $j \ne k^*$. Next, note that since $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ by hypothesis, there is a sequence $(h_{1n}, \dots, h_{Jn}) = h_n \in \mathbf{H}$ such that

$$\lim_{n\to\infty} \sum_{j=1}^J \left\| \nabla m_j(W_j, h_P)[h_n] - f_j^* \right\|_{P,2} = 0. \tag{F.28}$$

In particular, since $k^* > 1$, result (F.24) and h_n satisfying (F.28) together yield that

$$\lim_{n\to\infty} \left\| \nabla m_{1,1}(W_1, h_P)[h_{1n}] \right\|_{P,2} = 0. \tag{F.29}$$

Moreover, employing (F.24) and requirement (34), we obtain for any $2 \le j \le J$ that

$$\begin{aligned} &\left\| \nabla m_{j,j}(W_j, h_P)[h_{jn}] - f_j^* \right\|_{P,2} \\ &\quad \le \left\| \nabla m_j(W_j, h_P)[h_n] - f_j^* \right\|_{P,2} + C \sum_{k=1}^{j-1} \left\| \nabla m_{k,k}(W_k, h_P)[h_{kn}] \right\|_{P,2}. \end{aligned} \tag{F.30}$$

Evaluating (F.30) at any $j < k^*$ and proceeding inductively from (F.29) then implies

$$\lim_{n\to\infty} \left\| \nabla m_{j,j}(W_j, h_P)[h_{jn}] \right\|_{P,2} = 0 \tag{F.31}$$

since $f_j^* = 0$ for all $j < k^*$. Finally, evaluating (F.30) at $j = k^*$ and employing (F.31) implies $f_{k^*}^* = g^* \in \bar{\mathcal{R}}_{k^*}$. Since $1 < k^* \le J$ and $g^* \in L^2(W_{k^*})$ were arbitrary, it follows that $\bar{\mathcal{R}}_j =$

$L^2(W_j)$ for all j. Thus, we conclude $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ if and only if $\bar{\mathcal{R}}_j = L^2(W_j)$ for all $1 \leq j \leq J$. Finally, let $\nabla m_{j,j}(W_j, h_P)^* : L^2(W_j) \to \mathbf{H}_j^*$ denote the adjoint of $\nabla m_{j,j}(W_j, h_P) : \mathbf{H}_j \to L^2(W_j)$. Theorem 6.6.1 in Luenberger (1969) then implies that $\bar{\mathcal{R}}_j^\perp = \{f \in L^2(W_j) : \nabla m_{j,j}(W_j, h_P)^*[f] = 0\}$. Therefore, we further obtain that $\bar{\mathcal{R}}_j = L^2(W_j)$ for all $1 \leq j \leq J$ if and only if $\{f \in L^2(W_j) : \nabla m_{j,j}(W_j, h_P)^*[f] = 0\} = \{0\}$ for all $1 \leq j \leq J$. *Q.E.D.*

PROOF OF COROLLARY 4.1: Notice that the conditions of Lemma 4.2 are trivially satisfied. Therefore, Lemma 4.2 implies that $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ if and only if $\bar{\mathcal{R}}_j = L^2(W_j)$ for all $1 \leq j \leq J$, where $\bar{\mathcal{R}}_j$ denotes the closure of $\mathcal{R}_j$ in $L^2(W_j)$, and $\mathcal{R}_j$ is given by

$$\mathcal{R}_j = \{f \in L^2(W_j) : f = d_j(W_j)h_j(W_j) \text{ for some } h_j \in \mathbf{H}_j\}. \tag{F.32}$$

Hence, the claim of the corollary follows if, for all $1 \leq j \leq J$, $\bar{\mathcal{R}}_j = L^2(W_j)$ if and only if $\mathbf{H}_j$ is dense in $L^2(W_j)$ and $P(d_j(W_j) \neq 0) = 1$. The conditions of the corollary are equivalent for all $1 \leq j \leq J$, and therefore, without loss of generality, we focus on the case $j = 1$. To this end, we first suppose $\bar{\mathcal{R}}_1 = L^2(W_1)$ and define $f_1 \in L^2(W_1)$ by

$$f_1(W_1) \equiv 1\{d_1(W_1) = 0\}. \tag{F.33}$$

Next observe that since $\bar{\mathcal{R}}_1 = L^2(W_1)$ by hypothesis, it follows that $f_1 \in \bar{\mathcal{R}}_1$ and therefore

$$0 = \inf_{h_1 \in \mathbf{H}_1} E\big[\{d_1(W_1)h_1(W_1) - f_1(W_1)\}^2\big] \geq E\big[\{f_1(W_1)\}^2\big] = P(d_1(W_1) = 0), \tag{F.34}$$

where in the first equality we exploited (F.32), the inequality follows from definition (F.33) implying $d_1(W_1)f_1(W_1) = 0$ almost surely, and the final equality results from (F.33). Hence, we conclude that if $\bar{\mathcal{R}}_1 = L^2(W_1)$, then $P(d_1(W_1) \neq 0) = 1$. Moreover, for any $h_1 \in \mathbf{H}_1 \subseteq L^2(W_1)$, we have $d_1 h_1 \in L^2(W_1)$ since d_1 is bounded, and thus

$$\begin{aligned} 0 &= \inf_{h_1 \in \mathbf{H}_1} E\big[\{d_1(W_1)h_1(W_1) - d_1(W_1)f(W_1)\}^2\big] \\ &= \min_{h_1 \in \bar{\mathbf{H}}_1} E\big[\{d_1(W_1)\}^2\{h_1(W_1) - f(W_1)\}^2\big], \end{aligned} \tag{F.35}$$

for any $f \in L^2(W_1)$, and where the first equality follows from $\bar{\mathcal{R}}_1 = L^2(W_1)$, while the final equality holds for $\bar{\mathbf{H}}_1$ the closure of $\mathbf{H}_1$ in $L^2(W_1)$, and attainment of the infimum is guaranteed by the criterion being convex and diverging to infinity as $\|h_1\|_{P,2} \uparrow \infty$ and Proposition 38.15 in Zeidler (1984). Thus we conclude from (F.34) and (F.35) that, for any $f \in L^2(W_1)$, there exists a $h_1 \in \bar{\mathbf{H}}_1$ such that $P(f(W_1) = h_1(W_1)) = 1$. Since $\mathbf{H}_1 \subseteq L^2(W_1)$ by hypothesis, we conclude that in fact $\bar{\mathbf{H}}_1 = L^2(W_1)$.

We next suppose instead that $\bar{\mathbf{H}}_1 = L^2(W_1)$ and $P(d_1(W_1) \neq 0) = 1$ and aim to establish that $\bar{\mathcal{R}}_1 = L^2(W_1)$. First, since $\nabla m_1(W_1, h_P)[h] = d_1 h_1$ for any $(h_1, \dots, h_J) = h \in \mathbf{H}$ and d_1 is bounded, we may view $\nabla m_1(W_1, h_P)$ as a map from $\bar{\mathbf{H}}_1$ into $L^2(W_1)$ by, with some abuse of notation, setting $\nabla m_1(W_1, h_P)[h_1] = d_1 h_1$ for any $h_1 \in \bar{\mathbf{H}}_1 \setminus \mathbf{H}_1$ as well. Furthermore, since $\bar{\mathbf{H}}_1 = L^2(W_1)$ by hypothesis, direct calculation reveals that $\nabla m_1(W_1, h_P) : L^2(W_1) \to L^2(W_1)$ is self-adjoint. Thus, Theorem 6.6.3 in Luenberger (1969) implies $\bar{\mathcal{R}}_1 = L^2(W_1)$ if and only if $\nabla m_1(W_1, h_P) : L^2(W_1) \to L^2(W_1)$ is injective. However, injectivity of $\nabla m_1(W_1, h_P) : L^2(W_1) \to L^2(W_1)$ is equivalent to $P(d_1(W_1) \neq 0) = 1$, and therefore $\bar{\mathcal{R}}_1 = L^2(W_1)$. *Q.E.D.*

Previously, Ai and Chen (2012) derived the semiparametric efficiency bound for a general class of "smooth" functionals of P defined by nonparametric sequential moment restriction model (31). The next theorem, which is a restatement of Theorem 4.2, exploits their results and our Corollary 3.1 to obtain an alternative characterization of local just identification of P by model (31). In the following, we let Ω^*_f denote the semiparametric efficient variance bound for estimating population mean $\theta_f(P) \equiv \int f\,dP$ for f in any dense subset $\mathcal{D}$ of $L^2(P)$.

THEOREM F.1: *Let Assumption* 4.3 *hold*. *Then*: *There is a dense subset* $\mathcal{D}$ *of* $L^2(P)$ *such that* $\Omega^*_f = \mathrm{Var}\{f(X)\}$ *for all* $f \in \mathcal{D}$ *if and only if* $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$. *Hence*: P *is locally just identified by model* (31) *if and only if* $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$.

PROOF: We let $L^\infty(P) \equiv \{f : |f| \text{ is bounded } P\text{-a.s.}\}$, and $L^\infty(W_j)$ and $L^\infty(Z)$ be the subsets of $L^\infty(P)$ depending only on W_j and Z, respectively. We recall that $L^2(W_j)$ and $L^2(Z)$ are analogously defined. In addition, we note that Assumption 4.3(i) implies that if $j \le j'$, then it follows that

$$W_j = F(W_{j'}) \tag{F.36}$$

for some measurable function $F : \mathbf{W}_{j'} \to \mathbf{W}_j$; see, for example, Theorem 20.1 in Billingsley (2008). We define a subset $\mathcal{Q} \subseteq L^2(P)$ as

$$\mathcal{Q} \equiv \left\{ f : f(X) = \left\{ \sum_{j=1}^J \rho_j(Z, h_P) q_j(W_j) + C \right\} \text{ a.s. for some } q_j \in L^2(W_j), C \in \mathbf{R} \right\}, \tag{F.37}$$

and $\bar{\mathcal{Q}}$ as the closure of $\mathcal{Q}$ under $\|\cdot\|_{P,2}$. We set the desired subset $\mathcal{D}$ to equal $\mathcal{D} \equiv L^\infty(P) \setminus \bar{\mathcal{Q}}$ and note $\mathcal{D}$ is a subset of $L^2(P)$ since $\mathcal{D} \subseteq L^\infty(P) \subset L^2(P)$. To establish that $\mathcal{D}$ is dense in $L^2(P)$, we let j^* be the smallest j satisfying $1 \le j \le J$ and such that $L^2(W_{j^*})$ is infinite-dimensional—note existence of j^* is guaranteed by Assumption 4.3(vi). We next aim to show that

$$L^2(W_{j^*}) \cap \bar{\mathcal{Q}} \neq L^2(W_{j^*}), \tag{F.38}$$

and to this end, we note that since $L^2(W_{j^*})$ is infinite-dimensional and $L^2(W_j)$ is finite-dimensional for all $j < j^*$, it follows that there exists a $g \in L^2(W_{j^*})$ with $\|g\|_{P,2} > 0$ and

$$E\left[g(W_{j^*}) \left\{ \sum_{j=1}^{j^*-1} \rho_j(Z, h_P) q_j(W_j) + C \right\} \right] = 0 \tag{F.39}$$

for all $C \in \mathbf{R}$ and $q_j \in L^2(W_j)$—here, if $j^* = 1$, then (F.39) should be understood as just requiring $E[g(W_{j^*})] = 0$. On the other hand, Assumption 4.3(i) and the law of iterated expectations together imply that for any $q_j \in L^2(W_j)$, we have

$$E\left[g(W_{j^*}) \left\{ \sum_{j=j^*}^{J} \rho_j(Z, h_P) q_j(W_j) \right\} \right] = 0. \tag{F.40}$$

Thus, (F.39) and (F.40) imply $\|f - g\|_{P,2} = \|f\|_{P,2} + \|g\|_{P,2} > 0$ for any $f \in \bar{\mathcal{Q}}$, from which we conclude (F.38) holds. Since $L^\infty(W_{j^*})$ is dense in $L^2(W_{j^*})$, (F.38) further yields that there is a $\tilde{g} \in L^\infty(W_{j^*}) \setminus \bar{\mathcal{Q}}$ and for any $f \in L^\infty(P)$ and $\epsilon_n \downarrow 0$ we set

$$f_n = \begin{cases} f, & \text{if } f \notin \bar{\mathcal{Q}}, \\ f + \epsilon_n \tilde{g}, & \text{if } f \in \bar{\mathcal{Q}}, \end{cases} \tag{F.41}$$

and note $f_n \in \mathcal{D} \equiv L^\infty(P) \setminus \bar{\mathcal{Q}}$ and $\|f_n - f\|_{P,2} = o(1)$. We conclude that $\mathcal{D}$ is dense in $L^\infty(P)$ with respect to $\|\cdot\|_{P,2}$ and hence also in $L^2(P)$ since $L^\infty(P)$ is a dense subset of $L^2(P)$ under $\|\cdot\|_{P,2}$.

While we have so far avoided stating an explicit formulation for Ω_f^* for ease of exposition, it is now necessary to characterize it for all $f \in \mathcal{D}$. To this end, we follow Ai and Chen (2012) by setting $\varepsilon_J(Z,h) \equiv \rho_J(Z,h)$ and recursively defining

$$\varepsilon_s(Z,h) \equiv \rho_s(Z,h) - \sum_{j=s+1}^{J} \Gamma_{s,j}(W_j)\varepsilon_j(Z,h) \tag{F.42}$$

for $1 \le j \le J-1$, and where for any $1 \le s < j \le J$, the function $\Gamma_{s,j}(W_j)$ is given by

$$\Gamma_{s,j}(W_j) \equiv E\big[\rho_s(Z,h_P)\varepsilon_j(Z,h_P)|W_j\big]\big\{\Sigma_j(W_j)\big\}^{-1}, \tag{F.43}$$

$$\Sigma_j(W_j) \equiv E\big[\big\{\varepsilon_j(Z,h_P)\big\}^2|W_j\big], \tag{F.44}$$

and we note Assumptions 4.3(i), (iv), (v) and simple calculations together imply

$$P\big(\eta \le \Sigma_j(W_j) \le M\big) = 1 \tag{F.45}$$

for all $1 \le j \le J$ and some $\eta, M \in (0, +\infty)$. We further set $\Sigma_f \equiv \operatorname{Var}\{f(X)\}$ and define

$$\Sigma_0 \equiv \operatorname{Var}\left\{f(X) - \sum_{j=1}^{J} \Lambda_j(W_j)\varepsilon_j(Z,h_P)\right\}, \tag{F.46}$$

$$\Lambda_j(W_j) \equiv E\big[f(X)\varepsilon_j(Z,h_P)|W_j\big]\big\{\Sigma_j(W_j)\big\}^{-1}, \tag{F.47}$$

and note: (i) $\Lambda_j(W_j) \in L^2(W_j)$ by (F.44), (F.45), $f \in \mathcal{D} \subset L^\infty(P)$, and Jensen's inequality; (ii) $\Sigma_0 > 0$ since $f \notin \bar{\mathcal{Q}}$ and $\bar{\mathcal{Q}}$ is closed; and (iii) by direct calculation,

$$\Sigma_0 = \Sigma_f - \sum_{j=1}^{J} E\big[\big\{\Lambda_j(W_j)\big\}^2 \Sigma_j(W_j)\big]. \tag{F.48}$$

Next, we define the maps $a_j(W_j, \cdot) : \mathbf{H} \to L^2(W_j)$ for any $1 \le j \le J$ to be given by

$$a_j(W_j, h) \equiv E\big[\varepsilon_j(Z,h)|W_j\big]. \tag{F.49}$$

We further note that result (F.45) and Assumption 4.3(v) imply by arguing inductively that $\Gamma_{s,j}(W_j) \in L^\infty(W_j)$. Hence, it can be shown from definition (F.42) and Assumption 4.3(ii) that the maps $a_j(W_j, \cdot)$ are Fréchet differentiable at h_P and we denote their derivatives by $\nabla a_j(W_j, h_P) : \mathbf{H} \to L^2(W_j)$. Therefore, the Fisher norm of a $s \in \mathbf{H}$ is

$$\|s\|_w^2 \equiv \sum_{j=1}^{J} E\big[\big\{\Sigma_j(W_j)\big\}^{-1}\big\{\nabla a_j(W_j, h_P)[s]\big\}^2\big] + \{\Sigma_0\}^{-1}\left\{E\left[\sum_{j=1}^{J} \Lambda_j(W_j)\nabla a_j(W_j, h_P)[s]\right]\right\}^2 \tag{F.50}$$

(see eq. (4) in Ai and Chen (2012)), and we note $\|s\|_w < \infty$ for any $s \in \mathbf{H}$ since $\nabla a_j(W_j, h_P)[s] \in L^2(W_j)$, $\{\Sigma_j(W_j)\}^{-1} \in L^\infty(W_j)$ by (F.45), and as argued, $\Lambda_j(W_j) \in L^\infty(W_j)$. Letting $\mathcal{W}$ denote the completion of $\mathbf{H}$ under $\|\cdot\|_w$, we then obtain

$$\big\{\Omega_f^*\big\}^{-1} = \inf_{s\in\mathcal{W}}\left\{\{\Sigma_0\}^{-1}\left\{1 + \sum_{j=1}^{J} E\big[\Lambda_j(W_j)\nabla a_j(W_j, h_P)[s]\big]\right\}^2 + \sum_{j=1}^{J} E\big[\big\{\Sigma_j(W_j)\big\}^{-1}\big\{\nabla a_j(W_j, h_P)[s]\big\}^2\big]\right\} \tag{F.51}$$

by Theorem 2.1 in Ai and Chen (2012).

It is convenient for our purposes, however, to exploit the structure of our problem to further simplify the characterization in (F.51). To this end, note that (F.50) and the Cauchy–Schwarz inequality imply that the objective in (F.51) is continuous under $\|\cdot\|_w$. Hence, since $\mathcal{W}$ is the completion of $\mathbf{H}$ under $\|\cdot\|_w$, it follows from (F.51) that

$$\big\{\Omega_f^*\big\}^{-1} = \inf_{s\in\mathbf{H}}\left\{\{\Sigma_0\}^{-1}\left\{1 + \sum_{j=1}^{J} E\big[\Lambda_j(W_j)\nabla a_j(W_j, h_P)[s]\big]\right\}^2 + \sum_{j=1}^{J} E\big[\big\{\Sigma_j(W_j)\big\}^{-1}\big\{\nabla a_j(W_j, h_P)[s]\big\}^2\big]\right\}. \tag{F.52}$$

Next, note that we may view $(\nabla a_1(W_1, h_P), \ldots, \nabla a_J(W_J, h_P))$ as a map from $\mathbf{H}$ onto the product space $\bigotimes_{j=1}^{J} L^2(W_j)$, and we denote the range of this map by

$$\mathcal{A} \equiv \left\{\{r_j\}_{j=1}^{J} \in \bigotimes_{j=1}^{J} L^2(W_j) : \text{for some } s \in \mathbf{H}, r_j = \nabla a_j(W_j, h_P)[s] \text{ for all } 1 \le j \le J\right\}$$

and let $\bar{\mathcal{A}}$ denote the closure of $\mathcal{A}$ in the product topology. Result (F.52) then implies

$$\begin{aligned}\{\Omega_f^*\}^{-1} &= \inf_{\{r_j\}\in\bar{\mathcal{A}}}\left\{\{\Sigma_0\}^{-1}\left\{1+\sum_{j=1}^J E\big[\Lambda_j(W_j)r_j(W_j)\big]\right\}^2\right. \\ &\quad \left.+\sum_{j=1}^J E\big[\{\Sigma_j(W_j)\}^{-1}\{r_j(W_j)\}^2\big]\right\} \\ &= \min_{\{r_j\}\in\bar{\mathcal{A}}}\left\{\{\Sigma_0\}^{-1}\left\{1+\sum_{j=1}^J E\big[\Lambda_j(W_j)r_j(W_j)\big]\right\}^2\right. \\ &\quad \left.+\sum_{j=1}^J E\big[\{\Sigma_j(W_j)\}^{-1}\{r_j(W_j)\}^2\big]\right\},\end{aligned} \tag{F.53}$$

where attainment follows from $\bar{\mathcal{A}}$ being a vector space since $(\nabla a_1(W_1,h_P),\ldots,\nabla a_J(W_J,h_P))$ is linear and $\mathbf{H}$ is a vector space, the criterion in (F.53) being convex and diverges to infinity as $\sum_j \|r_j\|_{P,2}\uparrow\infty$, and Proposition 38.15 in Zeidler (1984). In particular, note that if $\{r_j^*\}\in\bar{\mathcal{A}}$ is the minimizer of (F.53), then for any $\{\delta_j\}\in\bar{\mathcal{A}}$,

$$\begin{aligned}&\sum_{j=1}^J E\left[\delta_j(W_j)\left\{\{\Sigma_j(W_j)\}^{-1}r_j^*(W_j)\right.\right. \\ &\quad \left.\left.+\Sigma_0^{-1}\Lambda_j(W_j)\left\{1+\sum_{s=1}^J E\big[\Lambda_s(W_s)r_s^*(W_s)\big]\right\}\right\}\right]=0.\end{aligned} \tag{F.54}$$

Next, we aim to solve the optimization in (F.54) under the hypothesis that $\bar{\mathcal{A}}=\bigotimes_{j=1}^J L^2(W_j)$. In that case, (F.54) must hold for all $\{\delta_j\}\in\bigotimes_{j=1}^J L^2(W_j)$, which implies

$$r_j^*(W_j)=-\Sigma_0^{-1}\left\{1+\sum_{s=1}^J E\big[\Lambda_s(W_s)r_s^*(W_s)\big]\right\}\Lambda_j(W_j)\Sigma_j(W_j). \tag{F.55}$$

It is evident from (F.55) that $r_j^*(W_j)=-\Lambda_j(W_j)\Sigma_j(W_j)C_0$ for some $C_0\in\mathbf{R}$ independent of j, and, plugging into (F.55), we solve for C_0 and exploit (F.48) to find

$$r_j^*(W_j)=-\{\Sigma_f\}^{-1}\Lambda_j(W_j)\Sigma_j(W_j). \tag{F.56}$$

Thus, combining (F.53) and (F.56), and repeatedly exploiting (F.48), we conclude

$$\begin{aligned}\{\Omega_f^*\}^{-1} &= \Sigma_0^{-1}\left\{1-\{\Sigma_f\}^{-1}\sum_{j=1}^J E\big[\Lambda_j^2(W_j)\Sigma_j(W_j)\big]\right\}^2+\{\Sigma_f\}^{-2}\sum_{j=1}^J E\big[\Lambda_j^2(W_j)\Sigma_j(W_j)\big] \\ &= \Sigma_0^{-1}\{1-\{\Sigma_f\}^{-1}\{\Sigma_f-\Sigma_0\}\}^2+\{\Sigma_f\}^{-2}\{\Sigma_f-\Sigma_0\}=\{\Sigma_f\}^{-1},\end{aligned} \tag{F.57}$$

or equivalently, $\Omega_f^* = \Sigma_f$. While (F.57) was derived while supposing $\bar{\mathcal{A}} = \bigotimes_{j=1}^J L^2(W_j)$, we note that since $\bar{\mathcal{A}} \subseteq \bigotimes_{j=1}^J L^2(W_j)$, the minimum in (F.53) is attained, and $r_j^*(W_j) = -\{\Sigma_f\}^{-1}\Lambda_j(W_j)\Sigma_j(W_j)$ is the unique minimizer on $\bigotimes_{j=1}^J L^2(W_j)$, we must have

$$\Omega_f^* = \Sigma_f \quad \text{if and only if} \quad \left\{-\Sigma_f^{-1}\Lambda_j(W_j)\Sigma_j(W_j)\right\}_{j=1}^J \in \bar{\mathcal{A}}. \tag{F.58}$$

Since result (F.58) holds for all $f \in \mathcal{D}$ and $\bar{\mathcal{A}}$ is a vector space, (F.47) implies

$$\Omega_f^* = \Sigma_f \quad \forall f \in \mathcal{D} \quad \text{if and only if} \quad \left\{E\left[f(X)\varepsilon_j(Z, h_P)|W_j\right]\right\}_{j=1}^J \in \bar{\mathcal{A}} \quad \forall f \in \mathcal{D}.$$

Also note that if $\|f_n - f\|_{P,2} = o(1)$, then by the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} &\lim_{n\to\infty} E\left[\left\{E\left[f_n(X)\varepsilon_j(Z, h_P)|W_j\right] - E\left[f(X)\varepsilon_j(Z, h_P)|W_j\right]\right\}^2\right] \\ &\quad \le \lim_{n\to\infty} E\left[\left\{f_n(X) - f(X)\right\}^2 \Sigma_j(W_j)\right] = 0, \end{aligned} \tag{F.59}$$

where the final equality follows from $\Sigma_j(W_j) \in L^\infty(W_j)$ by result (F.45). Therefore, since as argued, $\mathcal{D}$ is a dense subset of $L^2(P)$, in addition $\bar{\mathcal{A}}$ is closed under the product topology in $\bigotimes_{j=1}^J L^2(W_j)$, and result (F.59) holds for all $1 \le j \le J$, we conclude

$$\Omega_f^* = \Sigma_f \quad \forall f \in \mathcal{D} \quad \text{if and only if} \quad \left\{E\left[f(X)\varepsilon_j(Z, h_P)|W_j\right]\right\}_{j=1}^J \in \bar{\mathcal{A}} \quad \forall f \in L^2(P). \tag{F.60}$$

Next, fix an arbitrary $\{g_j\}_{j=1}^J \in \bigotimes_{j=1}^J L^\infty(W_j)$ and note that result (F.45) then yields

$$f_0(X) \equiv \sum_{j=1}^J g_j(W_j)\varepsilon_j(Z, h_P)\left\{\Sigma_j(W_j)\right\}^{-1}$$

belongs to $L^2(P)$ since $g_j \in L^\infty(W_j)$. Since $E[\{\varepsilon_j(Z, h_P)\}^2|W_j] = \Sigma_j(W_j)$, $E[\varepsilon_j(Z, h_P)\varepsilon_s(Z, h_P)|W_j] = 0$ whenever $s < j$, we obtain from result (F.36) that

$$\begin{aligned} &E\left[f_0(X)\varepsilon_j(Z, h_P)|W_j\right] \\ &\quad = E\left[\left(\sum_{s=1}^J g_s(W_s)\varepsilon_s(Z, h_P)\left\{\Sigma_s(W_s)\right\}^{-1}\right)\varepsilon_j(Z, h_P)\Big|W_j\right] \\ &\quad = \sum_{s=1}^J E\left[g_s(W_s)\left\{\Sigma_s(W_s)\right\}^{-1} E\left[\varepsilon_s(Z, h_P)\varepsilon_j(Z, h_P)|W_{s\vee j}\right]|W_j\right] = g_j(W_j). \end{aligned} \tag{F.61}$$

In particular, (F.61) holds for any $1 \le j \le J$, and since $\{g_j\}_{j=1}^J \in \bigotimes_{j=1}^J L^\infty(W_j)$ was arbitrary, it follows that if $\{E[f(X)\varepsilon_j(Z, h_P)|W_j]\}_{j=1}^J \in \bar{\mathcal{A}}$ for all $f \in L^2(P)$, then $\bigotimes_{j=1}^J L^\infty(W_j) \subseteq \bar{\mathcal{A}}$. However, since $\bar{\mathcal{A}}$ is closed in the product topology of $\bigotimes_{j=1}^J L^2(W_j)$, we have that if $\bigotimes_{j=1}^J L^\infty(W_j) \subseteq \bar{\mathcal{A}}$, then $\bigotimes_{j=1}^J L^2(W_j) = \bar{\mathcal{A}}$, and hence (F.60) yields

$$\Omega_f^* = \Sigma_f \quad \forall f \in \mathcal{D} \quad \text{if and only if} \quad \bigotimes_{j=1}^J L^2(W_j) = \bar{\mathcal{A}}. \tag{F.62}$$

To conclude, we note that since, as previously argued, $\Gamma_{s,j}(W_j) \in L^\infty(W_j)$ for all $1 \le s < j \le J$, definitions (F.43) and (F.49) and an inductive calculation imply that

$$\nabla m_j(Z, h_P)[s] = \nabla a_j(Z, h_P)[s] + \sum_{k=j+1}^{J} E\big[\Gamma_{j,k}(W_k)\nabla a_k(Z, h_P)[s] | W_j\big] \tag{F.63}$$

with $\nabla m_J(Z, h_P)[s] = \nabla a_J(Z, h_P)[s]$. Thus, from (F.63), we conclude $\bar{\mathcal{A}} = \bigotimes_{j=1}^J L^2(W_j)$ if and only if $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ and therefore the theorem follows from (F.62). *Q.E.D.*

APPENDIX G: SUFFICIENT CONDITIONS FOR ASSUMPTION 3.1

In this appendix, we illustrate how to construct a statistic $\hat{\mathbb{G}}_n$ satisfying Assumption 3.1 in the context of models defined by nonparametric conditional moment restrictions as studied in Section 4. Concretely, we let $\{X_i = (Z_i, W_i)\}_{i=1}^n$ be a random sample from the distribution P satisfying model (26), which is restated below for the purpose of easy reference:

$$E\big[\rho_j(Z_i, h_P) | W_{ij}\big] = 0 \quad \text{for all } 1 \le j \le J \text{ for some } h_P \in \mathbf{H}. \tag{G.1}$$

The parameter h_P can be estimated via the method of sieves by regularizing through either the choice of sieve, employing a penalization, or a combination of both approaches (Chen and Pouzo (2012)). Here, we assume $h_P \in \mathcal{H} \subseteq \mathbf{H}$, and consider a sequence of sieve spaces $\mathcal{H}_k \subseteq \mathcal{H}_{k+1} \subseteq \mathcal{H}$, with $\mathcal{H}_k$ growing suitably dense in $\mathcal{H}$ as k diverges to infinity. In turn, we estimate the unknown conditional expectation by series regression. Specifically, for $\{p_{jl}\}_{l=1}^\infty$ a sequence of approximating functions in $L^2(W_j)$, we let $p_j^{l_{jn}}(w_j) \equiv (p_{j1}(w_j), \dots, p_{jl_{jn}}(w_j))'$, set $P_{jn} \equiv (p_j^{l_{jn}}(W_{1j}), \dots, p_j^{l_{jn}}(W_{nj}))'$, and define

$$\hat{m}_j(w_j, h) \equiv \left\{\sum_{i=1}^n \rho_j(Z_i, h) p_j^{l_{jn}}(W_{ji})'\right\} \left(P_{jn}' P_{jn}\right)^- p_j^{l_{jn}}(w_j), \tag{G.2}$$

where $(P_{jn}' P_{jn})^-$ denotes the Moore–Penrose pseudoinverse of $P_{jn}' P_{jn}$. For a sequence k_n diverging to infinity with the sample size, the estimator $\hat{h}_n$ is then defined as

$$\hat{h}_n \in \arg\min_{h \in \mathcal{H}_{k_n}} \sum_{i=1}^n \sum_{j=1}^J \hat{m}_j^2(W_{ij}, h). \tag{G.3}$$

See Chen and Pouzo (2012) and references therein for sufficient conditions for the convergence rates of $\hat{h}_n$ to h_P.

For a set $\mathbf{T}$ and known function $\psi_j : \mathbf{W}_j \times \mathbf{T} \to \mathbf{R}$, we let $\psi(w, \tau) \equiv (\psi_1(w_1, \tau), \dots, \psi_J(w_J, \tau))'$ similarly define $\rho(z, h) \equiv (\rho_1(z, h), \dots, \rho_J(z, h))'$ and set

$$\hat{\mathbb{G}}_n(\tau) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \big\{\psi(W_i, \tau)\big\}' \rho(Z_i, \hat{h}_n). \tag{G.4}$$

Note that the resulting process $\hat{\mathbb{G}}_n$ may be viewed as an element of $\ell^\infty(\mathbf{T})$ provided that the functions $\psi_j(W_{ij}, \cdot)$ are bounded almost surely. To see why $\hat{\mathbb{G}}_n$ might satisfy Assumption 3.1, we observe that, for any $\tau \in \mathbf{T}$, $\hat{\mathbb{G}}_n(\tau)$ is an estimator of the parameter

$$\theta_P(\tau) \equiv E\big[\{\psi(W,\tau)\}'\rho(Z,h_P)\big]. \tag{G.5}$$

However, since h_P satisfies (G.1) by hypothesis, the model in fact dictates that $\theta_P(\tau) = 0$, and thus the efficient estimator for $\theta_P(\tau)$ is simply zero. As a result, $\hat{\mathbb{G}}_n(\tau)$ is an inefficient estimator of $\theta_P(\tau)$, and by Lemma 3.1(ii) it should satisfy Assumption 3.1 provided that it is regular and asymptotically linear. Similarly, $\hat{\mathbb{G}}_n$ could be constructed so that specification tests built on it aim their power at particular violations of the model by setting $\hat{\mathbb{G}}_n(\tau)$ to be the efficient estimator of $\theta_P(\tau)$ under the maintained alternative model; see Lemma 3.2(iii) and related discussion.

Denote $m(W,h) \equiv (m_1(W_1,h),\dots,m_J(W_J,h))'$. We assume that the maps $m_j(W_j,h) \equiv E[\rho_j(Z,h)|W_j]$, $j = 1,\dots,J$, are Fréchet differentiable at h_P with derivative $\nabla m_j(W_j,h_P) : \mathbf{H} \to L^2(W_j)$ (i.e., we impose Assumption 4.1(ii) as in Section 4). Recall that the linear map $\nabla m(W,h_P) : \mathbf{H} \to \bigotimes_{j=1}^J L^2(W_j)$ is given by

$$\nabla m(W,h_P)[h] \equiv \big(\nabla m_1(W_1,h_P)[h],\dots,\nabla m_J(W_J,h_P)[h]\big)', \tag{G.6}$$

and its range space equals

$$\mathcal{R} \equiv \left\{ f \in \bigotimes_{j=1}^J L^2(W_j) : f = \nabla m(W,h_P)[h] \text{ for some } h \in \mathbf{H} \right\}, \tag{G.7}$$

which is closed under addition, and its norm closure (in $\bigotimes_{j=1}^J L^2(W_j)$), denoted $\bar{\mathcal{R}}$, is a vector subspace of $\bigotimes_{j=1}^J L^2(W_j)$. With some abuse of notation, for any $(f_1,\dots,f_J) = f \in \bigotimes_{j=1}^J L^2(W_j)$, we let $\|f\|_{P,2}^2 = \sum_{j=1}^J \int f_j^2\,dP$ and we observe $\bigotimes_{j=1}^J L^2(W_j)$ is a Hilbert space under $\|\cdot\|_{P,2}$ and its corresponding inner product. Therefore, since $\bar{\mathcal{R}}$ is a closed subspace of $\bigotimes_{j=1}^J L^2(W_j)$, we obtain from Theorem 3.4.1 in Luenberger (1969) that

$$\bigotimes_{j=1}^J L^2(W_j) = \bar{\mathcal{R}} \oplus \bar{\mathcal{R}}^\perp. \tag{G.8}$$

For any $f \in \bigotimes_{j=1}^J L^2(W_j)$, we let $\Pi_{\bar{\mathcal{R}}} f$ and $\Pi_{\bar{\mathcal{R}}^\perp} f$ denote the projection of f under $\|\cdot\|_{P,2}$ onto $\bar{\mathcal{R}}$ and $\bar{\mathcal{R}}^\perp$, respectively. We emphasize that the projection of $(f_1,\dots,f_J) = f \in \bigotimes_{j=1}^J L^2(W_j)$ onto $\bar{\mathcal{R}}$ *need not equal* a coordinate by coordinate projection of f. Finally, recall that by Theorem 4.1, P is locally just identified if and only if $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$, or if and only if $\bar{\mathcal{R}}^\perp = \{0\}$.

We, in addition, impose the following assumptions to study the process $\hat{\mathbb{G}}_n$.

ASSUMPTION G.1: (i) $\mathcal{F} \equiv \{f = \{\{\Pi_{\bar{\mathcal{R}}^\perp}\psi(\cdot,\tau)\}'\rho(\cdot,h) : (\tau,h) \in \mathbf{T}\times\mathcal{H}\}$ *is* P-Donsker; (ii) $\|\Pi_{\bar{\mathcal{R}}^\perp}\psi(w,\tau)\|$ *is bounded on* $\bigotimes_{j=1}^J \mathbf{W}_j \times \mathbf{T}$; (iii) $\mathcal{H}_k \subseteq \mathcal{H}$ *for all* k.

ASSUMPTION G.2: (i) $\sum_{j=1}^J \|\rho_j(\cdot,\hat{h}_n) - \rho_j(\cdot,h_P)\|_{P,2} = o_p(1)$; (ii) $E[\|m(W_i,\hat{h}_n) - m(W_i,h_P) - \nabla m(W_i,h_P)[\hat{h}_n - h_P]\|] = o_p(n^{-1/2})$, (iii) $\frac{1}{n}\sum_{i=1}^n \{\Pi_{\bar{\mathcal{R}}}\psi(W_i,\tau)\}'\rho(Z_i,\hat{h}_n) = o_p(n^{-1/2})$ *uniformly in* $\tau \in \mathbf{T}$.

Assumption G.1(i) ensures that the empirical process indexed by $f \in \mathcal{F}$ converges in distribution in $\ell^\infty(\mathbf{T})$, Assumption G.1(ii) demands that the weights in the linear combinations of moments be bounded, and Assumption G.1(iii) implies that $\hat{h}_n \in \mathcal{H}$ with probability 1. Assumption G.2 imposes high-level conditions on $\hat{h}_n$ that are transparent in their role played in the proof, though they can be verified under lower-level requirements on the sieve bases, the sieve approximation errors, and the smoothness of the map $m(W, \cdot)$ near h_P. In particular, Assumption G.2(i) imposes that $\rho(\cdot, \hat{h}_n)$ be consistent for $\rho(\cdot, h_P)$ in $\bigotimes_{j=1}^J L^2(P)$. Assumption G.2(ii) demands the rate of convergence of $\hat{h}_n$ to be sufficiently fast to enable us to obtain a suitable expansion of $m(W_i, \hat{h}_n)$ around h_P. Both Assumption G.2(i) and G.2(ii) can be verified under lower-level conditions by employing the results in Chen and Pouzo (2012). Finally, Assumption G.2(iii) intuitively follows from $\hat{h}_n$ satisfying (G.3) and $\nabla\hat{m}(W_i, \hat{h}_n)$ approximating $\nabla m(W_i, h_P)$;[2] see Ai and Chen (2003) and Chen and Pouzo (2009) for related arguments.

We next establish the asymptotic behavior of $\hat{\mathbb{G}}_n$.

LEMMA G.1: *Let Assumptions* 4.1(i), (ii), *and* G.1 *and* G.2 *hold*. *Then*:

$$\hat{\mathbb{G}}_n(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\Pi_{\mathcal{R}^\perp}\psi(W_i, \tau)\}' \rho(Z_i, h_P) + o_p(1) \tag{G.9}$$

uniformly in $\tau \in \mathbf{T}$, *and* $\hat{\mathbb{G}}_n \xrightarrow{L} \mathbb{G}_0$ *in* $\ell^\infty(\mathbf{T})$ *for some tight Gaussian measure* $\mathbb{G}_0$.

Lemma G.1 establishes the asymptotic linearity of $\hat{\mathbb{G}}_n$ as a process in $\ell^\infty(\mathbf{T})$. Since the influence function of $\hat{\mathbb{G}}_n$ obeys a functional central limit theorem by Assumption G.1(i), the conclusion that $\hat{\mathbb{G}}_n$ converges to a tight Gaussian process is immediate from result (G.9). Therefore, given Lemma G.1, the main requirement remaining in verifying $\hat{\mathbb{G}}_n$ satisfies Assumption 3.1 is showing that the influence function of $\hat{\mathbb{G}}_n(\tau)$ is orthogonal to the scores of the model for any $\tau \in \mathbf{T}$. However, the latter claim is immediate from the characterization of $\bar{T}(P)^\perp$ derived in Theorem 4.1.

PROOF OF LEMMA G.1: We first note that Assumption G.2(iii) allows us to conclude

$$\hat{\mathbb{G}}_n(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\Pi_{\mathcal{R}^\perp}\psi(W_i, \tau)\}' \rho(Z_i, \hat{h}_n) + o_p(1) \tag{G.10}$$

uniformly in $\tau \in \mathbf{T}$ since $\psi(W_i, \tau) = \Pi_{\mathcal{R}}\psi(W_i, \tau) + \Pi_{\mathcal{R}^\perp}\psi(W_i, \tau)$. Moreover, by the Cauchy–Schwarz inequality, and Assumptions G.1(ii) and G.2(i), we obtain that

$$\begin{aligned} &E\big[\big(\Pi_{\mathcal{R}^\perp}\psi(W_i, \tau)'\{\rho(Z_i, \hat{h}_n) - \rho(Z_i, h_P)\}\big)^2\big] \\ &\quad \le \sup_{(w,\tau)} \big\|\psi(w, \tau)\big\|^2 \times \big\|\rho(\cdot, \hat{h}_n) - \rho(\cdot, h_P)\big\|_{P,2}^2 = o_p(1). \end{aligned} \tag{G.11}$$

[2]Recall $\Pi_{\mathcal{R}}\psi(W_i, \tau) = \nabla m(W_i, h_P)[v_n] + o(1)$ for some sequence $\{v_n\}_{n=1}^\infty \in \mathbf{H}$, while $\hat{h}_n$ solving (G.3) can be exploited to show $\frac{1}{n}\sum_{i=1}^n \{\nabla\hat{m}(W_i, \hat{h}_n)[v_n]\}'\rho(Z_i, \hat{h}_n) = o_p(n^{-1/2})$.

Thus, since $\mathcal{F} = \{f(x) = \{\Pi_{\mathcal{R}^\perp}\psi(w, \tau)\}'\rho(z, h) : (\tau, h) \in \mathbf{T} \times \mathcal{H}\}$ is P-Donsker by Assumption G.1(i) and $\hat{h}_n \in \mathcal{H}$ by Assumption G.1(iii), result (G.11) yields

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\Pi_{\mathcal{R}^\perp}\psi(W_i, \tau)\}'\rho(Z_i, \hat{h}_n) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\Pi_{\mathcal{R}^\perp}\psi(W_i, \tau)\}'\rho(Z_i, h_P)\\ &\quad = \sqrt{n}E\big[\{\Pi_{\mathcal{R}^\perp}\psi(W_i, \tau)\}'\{\rho(Z_i, \hat{h}_n) - \rho(Z_i, h_P)\}\big] + o_p(1)\end{aligned} \tag{G.12}$$

uniformly in $\tau \in \mathbf{T}$. Furthermore, the law of iterated expectations, the Cauchy–Schwarz inequality, and Assumptions G.1(ii) and G.2(ii) together yield uniformly in $\tau \in \mathbf{T}$ that

$$\begin{aligned}&\sqrt{n}E\big[\{\Pi_{\mathcal{R}^\perp}\psi(W_i, \tau)\}'\{\rho(Z_i, \hat{h}_n) - \rho(Z_i, h_P)\}\big]\\ &\quad = \sqrt{n}E\big[\{\Pi_{\mathcal{R}^\perp}\psi(W_i, \tau)\}'\nabla m(W_i, h_P)[\hat{h}_n - h_P]\big] + o_p(1) = o_p(1),\end{aligned} \tag{G.13}$$

where in the final equality we exploited that, by definition of $\mathcal{R}$ and $\mathcal{R}^\perp$, it follows that for any $h \in \mathbf{H}$, we have $E[\{\Pi_{\mathcal{R}^\perp}\psi(W_i, \tau)\}'\nabla m(W_i, h_P)[h]] = 0$. Hence, the lemma follows from results (G.10), (G.12), and (G.13), and the class $\mathcal{F} = \{f(x) = \{\Pi_{\mathcal{R}^\perp}\psi(w, \tau)\}'\rho(z, h) : (\tau, h) \in \mathbf{T} \times \mathcal{H}\}$ being P-Donsker by Assumption G.1(i). *Q.E.D.*

APPENDIX H: EXAMPLES FOR SECTION 4

In this appendix, we provide additional discussions on Examples 4.1, 4.2, and 4.3 to illustrate how to employ Theorem 4.1, Lemmas 4.1 and 4.2, and Corollary 4.1 to determine whether P is locally overidentified by the model $\mathbf{P}$ in specific applications. We also introduce a final example based on DiNardo, Fortin, and Lemieux (1996).

EXAMPLE 4.1: In this application, Z represents the distinct elements of $(V, Y_1, \ldots, Y_J)$ and there are J moment restrictions. For any $h = (h_1, \ldots, h_J) \in \mathbf{H} = \bigotimes_{j=1}^{J}\mathbf{H}_j$, each $\rho_j : \mathbf{Z} \times \mathbf{H} \to \mathbf{R}$ then equals

$$\rho_j(Z, h) = Y_j - h_j(V). \tag{H.1}$$

Therefore, for any $h = (h_1, \ldots, h_J) \in \mathbf{H}$, $m_j(W_j, h) = E[Y_j - h_j(V)|W_j]$ which is affine and continuous by Jensen's inequality and $\mathbf{H}_j \subseteq L^2(V)$. Hence, $m_j(W_j, h) : \mathbf{H} \to L^2(W_j)$ is Fréchet differentiable with

$$\nabla m_j(W_j, h_P)[h] = -E\big[h_j(V)|W_j\big] \tag{H.2}$$

for any $h = (h_1, \ldots, h_J) \in \mathbf{H}$. In particular, note that the conditions of Lemma 4.2 are trivially satisfied since $\nabla m_{j,k}(W_j, h_P)[h_k] = 0$ for all $k \neq j$ and $1 \leq j \leq J$. Hence, defining

$$\mathcal{R}_j \equiv \big\{f \in L^2(W_j) : f = E\big[h_j(V)|W_j\big] \text{ for some } h_j \in \mathbf{H}_j\big\}, \tag{H.3}$$

we conclude from Lemma 4.2 that P is locally just identified if and only if $\bar{\mathcal{R}}_j = L^2(W_j)$ for all $1 \leq j \leq J$—that is, we may study local overidentification by examining each moment condition separately. We gain insight into the condition $\bar{\mathcal{R}}_j = L^2(W_j)$ by considering two separate cases.

Case I: We first suppose $W_j = V$ (i.e., V is exogenous in the jth moment restriction). In this case, we may view $E[\cdot|W_j] : \mathbf{H}_j \to L^2(W_j)$ as the identity mapping, and hence $\bar{\mathcal{R}}_j =$

$L^2(W_j)$ whenever $\mathbf{H}_j = L^2(V)$. Notice in fact that by Corollary 4.1, $\bar{\mathcal{R}}_j$ continues to satisfy $\bar{\mathcal{R}}_j = L^2(W_j)$ if we set $\mathbf{H}_j$ to be any Banach space that is dense in $L^2(W_j)$, such as the set of bounded functions, continuous functions, or differentiable functions. On the other hand, Corollary 4.1 implies $\bar{\mathcal{R}}_j \neq L^2(W_j)$ whenever $\mathbf{H}_j$ is a strict closed subspace of $L^2(V)$, which occurs, for example, when we impose a partially linear or an additively separable specification for h_j (Robinson (1988), Stone (1985)).

Case II: We next consider the case where W_j is an instrument, so that $W_j \neq V$. We let $\mathbf{H}_j = L^2(V)$; the condition that the closure of the range of $E[\cdot|W_j] : L^2(V) \to L^2(W_j)$ be equal to $L^2(W_j)$ is most easily interpreted through Lemma 4.2. Note that the adjoint of $E[\cdot|W_j]$ is $E[\cdot|V] : L^2(W_j) \to L^2(V)$. Thus, Lemma 4.2 implies that $\bar{\mathcal{R}}_j = L^2(W_j)$ if and only if

$$\{0\} = \{f \in L^2(W_j) : E[f(W_j)|V] = 0\}. \tag{H.4}$$

The requirement in (H.4) is known as the distribution of (V, W_j) being L^2-complete with respect to W_j, which is an untestable property of the distribution of the data (Andrews (2017), Canay, Santos, and Shaikh (2013)). As in Case I, however, we may obtain $\bar{\mathcal{R}}_j \neq L^2(W_j)$ by restricting the parameter space for h_j. Suppose, for example, that $\mathbf{H}_j$ is a closed subspace of $L^2(V)$, such as in a partially linear or an additive separable specification. For any $f \in L^2(W_j)$, then let

$$\Pi_{\mathbf{H}_j} f \equiv \arg\min_{h \in \mathbf{H}_j} \|f - h\|_{P,2} \tag{H.5}$$

and note $\Pi_{\mathbf{H}_j} : L^2(W_j) \to \mathbf{H}_j$ is the adjoint of $E[\cdot|W_j] : \mathbf{H}_j \to L^2(W_j)$. Thus, applying Lemma 4.2, we obtain that $\bar{\mathcal{R}}_j = L^2(W_j)$ if and only if

$$\{0\} = \{f \in L^2(W_j) : \Pi_{\mathbf{H}_j} f = 0\}. \tag{H.6}$$

Condition (H.6) may be viewed as a generalization of (H.4), and can be violated even when $\mathbf{H}_j$ is infinite-dimensional yet a strict subspace of $L^2(V)$.

EXAMPLE 4.2: We will study a general nonparametric specification for the parameter space and aim to show P is nonetheless locally overidentified. To this end, let

$$C^1([0,1]) \equiv \{f : [0,1] \to \mathbf{R} : f \text{ is continuously differentiable on } [0,1]\}, \tag{H.7}$$

which is a Banach space when endowed with the norm $\|f\|_{C^1} \equiv \|f\|_\infty + \|f'\|_\infty$ for f' the derivative of f. We then set the parameter space $\mathbf{H}$ to be given by

$$\mathbf{H} = L^\infty((V, R)) \times L^2(V) \times L^2(V) \times C^1([0,1]) \times C^1([0,1]), \tag{H.8}$$

and assume $(s_P, g_{0,P}, g_{1,P}, \lambda_{0,P}, \lambda_{1,P}) = h_P \in \mathbf{H}$; that is, we require the $\lambda_{d,P}$ functions in (42) to be continuously differentiable. For $X = (Y, D, V, R)$ and $W_1 = (V, R)$, the moment restriction in (43) corresponds to setting, for any $(s, g_0, g_1, \lambda_0, \lambda_1) = h \in \mathbf{H}$,

$$\rho_1(Z, h) = D - s(V, R). \tag{H.9}$$

In turn, for the moment restriction in (44), we let $W_2 = (V, R, D)$ and define

$$\rho_2(Z, h) = D\{Y - g_1(V) - \lambda_1(s(V, R))\} + (1 - D)\{Y - g_0(V) - \lambda_0(s(V, R))\} \tag{H.10}$$

for any $(s, g_0, g_1, \lambda_0, \lambda_1) = h \in \mathbf{H}$.[3] We note that by (H.9), $m_1(W_1, \cdot) : \mathbf{H} \to L^2(W_1)$ is affine and continuous and therefore Fréchet differentiable with derivative

$$\nabla m_1(W_1, h_P)[h] = -s(V, R) \tag{H.11}$$

for any $(s, g_0, g_1, \lambda_0, \lambda_1) = h \in \mathbf{H}$. The second restriction is Fréchet differentiable as well, and for any $(s, g_0, g_1, \lambda_0, \lambda_1) \in \mathbf{H}$, $\nabla m_2(W_2, h_P) : \mathbf{H} \to L^2(W_2)$ is given by

$$\begin{aligned}
&\nabla m_2(W_2, h_P)[h] \\
&\quad = D\{-g_1(V) - \lambda_1(s_P(V,R)) - \lambda'_{1,P}(s_P(V,R))s(V,R)\} \\
&\qquad + (1-D)\{-g_0(V) - \lambda_0(s_P(V,R)) - \lambda'_{0,P}(s_P(V,R))s(V,R)\}.
\end{aligned} \tag{H.12}$$

To verify this claim, first note $\nabla m_2(W_{2,}, h_P) : \mathbf{H} \to L^2(W_2)$ is continuous when $\mathbf{H}$ is endowed with the product topology. Moreover, by the mean value theorem, we have

$$\begin{aligned}
&\{\lambda_{d,P} + \lambda_d\}(s_P(V,R) + s(V,R)) \\
&\qquad - \{\lambda_{d,P} + \lambda_d\}(s_P(V,R)) - \lambda'_{d,P}(s_P(V,R))s(V,R) \\
&\quad = (\lambda'_{d,P}(\bar{s}(V,R)) - \lambda'_{d,P}(s_P(V,R)))s(V,R) + \lambda'_d(\bar{s}(V,R))s(V,R)
\end{aligned} \tag{H.13}$$

for some $\bar{s}(V,R)$ a convex combination of $s_P(V,R)$ and $s_P(V,R) + s(V,R)$. Exploiting (H.10), (H.12), and (H.13), we can then obtain that

$$\begin{aligned}
&\|m_2(W_2, h_P + h) - m_2(W_2, h_P) - \nabla m_2(W_2, h_P)[h]\|_{P,2} \\
&\quad = o\left(\|s\|_\infty \left\{1 + \sum_{d=1}^{2} \|\lambda'_d\|_\infty\right\}\right)
\end{aligned} \tag{H.14}$$

since $\lambda'_{d,P}$ is uniformly continuous on $[0, 1]$ and $\|s_P - \bar{s}\|_\infty \leq \|s\|_\infty$. Thus, from (H.14), we conclude $\nabla m_2(W_2, h_P)$ is indeed the Fréchet derivative of $m_2(W_2, \cdot) : \mathbf{H} \to L^2(W_2)$. In order to show that P is locally overidentified, we note that the moment restrictions (H.9) and (H.10) possess a triangular structure. Hence, we aim to apply Lemma 4.2 with $s_P \in \mathbf{H}_1 = L^\infty((V,R))$ and $(g_{0,P}, g_{1,P}, \lambda_{0,P}, \lambda_{1,P}) = h_{P,2} \in \mathbf{H}_2 = L^2(V) \times L^2(V) \times C^1([0,1]) \times C^1([0,1])$, for which (34) then holds since $\|\lambda'_{d,P}\|_\infty < \infty$ for $d \in \{0, 1\}$. Moreover, Corollary 4.1 implies $\bar{\mathcal{R}}_1 = L^2(W_1)$ since $L^\infty(W_1)$ is dense in $L^2(W_1)$ under $\|\cdot\|_{P,2}$. Therefore, letting $S = s_P(V,R)$ and $h_2 = (g_0, g_1, \lambda_0, \lambda_1)$ for notational simplicity, we note (H.12) and Lemma 4.2 together imply that P is locally just identified by $\mathbf{P}$ if and only if

$$\begin{aligned}
\mathcal{R}_2 = \{f \in L^2(W_2) : f(W_2) = {}& D\{g_1(V) + \lambda_1(S)\} \\
&+ (1-D)\{g_0(V) + \lambda_0(S)\} \text{ for } h_2 \in \mathbf{H}_2\}
\end{aligned} \tag{H.15}$$

is dense in $L^2(W_2)$. However, if $S = s_P(V,R)$ is not a measurable function of V (i.e., the instrument R is relevant), then $\{f \in L^2((V,S)) : f(V,S) = g(V) + \lambda(S)\}$ is not dense in $L^2((V,S))$. Hence, from (H.15) we conclude that $\bar{\mathcal{R}}_2$ is not dense in $L^2(W_2)$ and thus by Lemma 4.2 and Theorem 4.1 that P is locally overidentified.

[3]Technically, $\lambda_j(s(V,R))$ may not be well defined if $s(V,R) \notin [0,1]$ since $\lambda_j \in C^1([0,1])$. However, note $s_P(V,R) \in [0,1]$ almost surely by (43) so for notational simplicity we ignore this issue.

EXAMPLE 4.3: We study a more general version of the model introduced in the main text. In particular, we still maintain that for some U_{it} mean independent of (K_{it}, L_{it}, I_{it}),

$$Y_{it} = g_P(K_{it}, L_{it}) + \omega_{it} + U_{it}. \tag{H.16}$$

However, we now let L_{it} be a possibly dynamic variable, in which case (48) becomes

$$\omega_{it} = \lambda_P(K_{it}, L_{it}, I_{it}). \tag{H.17}$$

Maintaining that ω_{it} follows an AR(1) process with coefficient π_P, and recalling that $W_i = (K_{i1}, L_{i1}, I_{i1})$, we then obtain the following two conditional moment restrictions:

$$E\big[Y_1 - \nu_P(W)|W\big] = 0, \tag{H.18}$$

$$E\big[Y_2 - g_P(K_2, L_2) - \pi_P\big(\nu_P(W) - g_P(K_1, L_1)\big)|W\big] = 0, \tag{H.19}$$

where $\nu_P(W) = g_P(K_1, L_1) + \lambda_P(K_1, L_1, I_1)$. Let $L^2((K_1, L_1)) = L^2((K_2, L_2))$, $h_P = (\nu_P, g_P, \pi_P)$, and the parameter space be $\mathbf{H} = L^2(W) \times L^2((K_1, L_1)) \times \mathbf{R}$. It is straightforward to verify that in this model, we have, for any $h = (\nu, g, \pi) \in \mathbf{H}$,

$$\nabla m_1(W, h_P)[h] = -\nu(W), \tag{H.20}$$

$$\begin{aligned} \nabla m_2(W, h_P)[h] = &-E\big[g(K_2, L_2)|W\big] - \pi\lambda_P(K_1, L_1, I_1) \\ &- \pi_P\big(\nu(W) - g(K_1, L_1)\big). \end{aligned} \tag{H.21}$$

Since the model defined by (H.18) and (H.19) has a triangular structure, we next apply Lemma 4.2 to establish that it is locally overidentified. Let $\nu_P \in \mathbf{H}_1 = L^2(W)$ and $(g_P, \pi_P) \in \mathbf{H}_2 = L^2((K_1, L_1)) \times \mathbf{R}$. Since $\pi_P < \infty$, condition (34) of Lemma 4.2 is satisfied by (H.20), (H.21), and direct calculation. Applying Corollary 4.1 to (H.20) and since $\mathbf{H}_1 = L^2(W)$, we trivially obtain that $\bar{\mathcal{R}}_1 = L^2(W)$. Therefore, by Theorem 4.1 and Lemma 4.2, we can conclude that P is locally overidentified if and only if $\bar{\mathcal{R}}_2 \neq L^2(W)$, where

$$\begin{aligned} \mathcal{R}_2 = \big\{&-E\big[g(K_2, L_2)|W\big] + \pi_P g(K_1, L_1) - \pi\lambda_P(K_1, L_1, I_1) : \\ &(g, \pi) \in L^2\big((K_1, L_1)\big) \times \mathbf{R}\big\}. \end{aligned} \tag{H.22}$$

Inspecting (H.22), a sufficient condition for P to be locally overidentified is therefore for the map $g \mapsto E[g(K_2, L_2)|W]$ to not be able to generate arbitrary functions of $W = (K_1, L_1, I_1)$. Formally, defining the spaces

$$\begin{aligned} \bar{F} &\equiv \mathrm{cl}\big\{E\big[g(K_2, L_2)|W\big] - E\big[g(K_2, L_2)|K_1, L_1\big] : g \in L^2\big((K_1, L_1)\big)\big\}, \\ L^2\big((K_1, L_1)\big)^{\perp} &\equiv \left\{f \in L^2(W) : \int fg\, dP = 0 \text{ for all } g \in L^2\big((K_1, L_1)\big)\right\}, \end{aligned} \tag{H.23}$$

we note $\bar{F} \subseteq L^2((K_1, L_1))^{\perp}$ by the law of iterated expectations, and therefore we may decompose $L^2((K_1, L_1))^{\perp} = \bar{F} \oplus \bar{F}^{\perp}$. A sufficient condition for P to be locally overidentified is then that the dimension of $\bar{F}^{\perp}$ is at least 2.

We conclude by discussing an additional example based on the nonparametric analysis of changes in the wage distribution by DiNardo, Fortin, and Lemieux (1996).

EXAMPLE H.1: Suppose we observe $\{H_i, D_i, V_i, T_i\}_{i=1}^n$, where, for each individual i, H_i denotes hourly wages, D_i is a dummy variable for union membership, V_i is a vector of covariates, and $T_i \in \{1, 2\}$ indicates the time period individual i was measured in. The parameter of interest θ_P is the counterfactual τth quantile of wages that would have held in period two if unionization rates had been constant between periods, which solves

$$E\left[\tau - 1\{H \leq \theta_P, D = 1\}\frac{\lambda_{1,P}(V)}{\lambda_{2,P}(V)} - 1\{H \leq \theta_P, D = 0\}\frac{1 - \lambda_{1,P}(V)}{1 - \lambda_{2,P}(V)}\Big| T = 2\right] = 0 \quad \text{(H.24)}$$

for $\lambda_{t,P}(V)$ the unionization rate conditional on V at time t.[4] Since $\lambda_{t,P}$ satisfies

$$E\big[D - \lambda_{t,P}(V)|V, T = t\big] = 0 \quad \text{for } t \in \{1, 2\}, \quad \text{(H.25)}$$

this setting fits model (26) with parameters $(\lambda_{1,P}, \lambda_{2,P}, \theta_P)$. Specifically, we suppose that $\lambda_{t,P} \in \mathbf{L}_t \subseteq L^\infty(V)$, let $\mathbf{H} = \mathbf{L}_1 \times \mathbf{L}_2 \times \mathbf{R}$, set $X = (Z, W) = (H, D, V, T)$ and define

$$\rho_1(Z, h) = 1\{T = 1\}\big(D - \lambda_1(V)\big) + 1\{T = 2\}\big(D - \lambda_2(V)\big),$$
$$\rho_2(Z, h) = 1\{T = 2\}\left(\tau - 1\{H \leq \theta, D = 1\}\frac{\lambda_1(V)}{\lambda_2(V)} - 1\{H \leq \theta, D = 0\}\frac{1 - \lambda_1(V)}{1 - \lambda_2(V)}\right),$$

for any $h = (\lambda_1, \lambda_2, \theta) \in \mathbf{H}$, and where $W_1 = (V, T)$, and since the second moment restriction is *unconditional*, we set $W_2 = \{1\}$. It is straightforward to verify that in this model,

$$\nabla m_1(W_1, h_P)[h] = -1\{T = 1\}\lambda_1(V) - 1\{T = 2\}\lambda_2(V) \quad \text{(H.26)}$$

for any $h = (\lambda_1, \lambda_2, \theta) \in \mathbf{H}$. For notational simplicity, we let $R = (V, T, D)$ and $G_{H|R}(h|r)$ and $g_{H|R}(h|r)$ respectively denote the cdf and density of H conditional on R. Then, by direct calculation, it follows that, for any $h = (\lambda_1, \lambda_2, \theta) \in \mathbf{H}$, we have

$$\begin{aligned} m_2(W_2, h) = E\Bigg[1\{T = 2\}\bigg(&\tau - G_{H|R}(\theta|R)1\{D = 1\}\frac{\lambda_1(V)}{\lambda_2(V)} \\ &- G_{H|R}(\theta|R)1\{D = 0\}\frac{1 - \lambda_1(V)}{1 - \lambda_2(V)}\bigg)\Bigg], \end{aligned}$$

and that sufficient conditions for $m_2(W_2, \cdot) : \mathbf{H} \to \mathbf{R}$ to be Fréchet differentiable are that: (i) $g_{H|R}(H|R)$ be continuously differentiable in H with almost surely bounded level and derivative in (H, R), and (ii) $P(1 - \epsilon \geq \lambda_{2,P}(V) \geq \epsilon) = 1$ for some $\epsilon > 0$. In addition, notice that this model possesses the triangular structure required in Lemma 4.2 with $\mathbf{H}_1 = \mathbf{L}_1 \times \mathbf{L}_2$ and $\mathbf{H}_2 = \mathbf{R}$, while requirement (34) holds under the additional assumption that $P(P(T = t|V) \geq \epsilon) = 1$ for $t \in \{1, 2\}$ and some $\epsilon > 0$. Employing the notation of Lemma 4.2, we obtain by direct calculation that

$$\mathcal{R}_1 = \left\{f \in L^2(W_1) : f(T, V) = \sum_{t=1}^{2} 1\{T = t\}\lambda_t(V) \text{ for } (\lambda_1, \lambda_2) \in \mathbf{L}_1 \times \mathbf{L}_2\right\},$$
$$\mathcal{R}_2 = \left\{\theta \times E\left[1\{T = 2\}g_{H|R}(\theta_P|R)\left(1\{D = 1\}\frac{\lambda_{1,P}(V)}{\lambda_{2,P}(V)} + 1\{D = 0\}\frac{1 - \lambda_{1,P}(V)}{1 - \lambda_{2,P}(V)}\right)\right] : \theta \in \mathbf{R}\right\}.$$

[4]That is, as in DiNardo, Fortin, and Lemieux (1996), we desire the τth quantile of $G(H|D, V, T = 2)G(D|V, T = 1)G(V|T = 2)$, where for any (A, B), $G(A|B)$ denotes the distribution of A conditional on B.

Notice, however, that the expectation defining $\mathcal{R}_2$ is necessarily positive, and thus $\mathcal{R}_2 = \mathbf{R}$—an equality that simply reflects that assuming existence of θ_P offers no additional information. Thus, Lemma 4.2 implies that P is locally just identified if and only if $\bar{\mathcal{R}}_1 = L^2(W_1)$. Equivalently, P is locally just identified iff $\mathbf{L}_t$ is dense in $L^2(V)$ for $t \in \{1, 2\}$. For example, P is locally overidentified if we restrict $\lambda_{t,P}$ through an additive separable specification. In that case, the choice of estimator for $\lambda_{t,P}$ can affect the asymptotic distribution of a plug-in estimator for θ_P.

REFERENCES

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1844. [29]

——— (2012): "The Semiparametric Efficiency Bound for Models of Sequential Moment Restrictions Containing Unknown Functions," *Journal of Econometrics*, 170, 442–457. [22-24]

ALIPRANTIS, C. D., AND K. C. BORDER (2006): *Infinite Dimensional Analysis—A Hitchhiker's Guide*. Berlin: Springer-Verlag. [11-13]

ANDREWS, D. W. K. (2017): "Examples of L2-Complete and Boundedly-Complete Distributions," *Journal of Econometrics*, 199, 213–220. [31]

BICKEL, P., C. KLAASSEN, Y. RITOV, AND J. WELLNER (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press. [1,2,8]

BILLINGSLEY, P. (2008): *Probability and Measure*. John Wiley & Sons. [22]

BOGACHEV, V. I. (1998): *Gaussian Measures*. Providence: American Mathematical Society. [5,6,10,11]

——— (2007): *Measure Theory: Volume I*. Berlin: Springer-Verlag. [11]

CANAY, I. A., A. SANTOS, AND A. M. SHAIKH (2013): "On the Testability of Identification in Some Nonparametric Models With Endogeneity," *Econometrica*, 81, 2535–2559. [31]

CHEN, X., AND D. POUZO (2009): "Efficient Estimation of Semiparametric Conditional Moment Models With Possibly Nonsmooth Residuals," *Journal of Econometrics*, 152, 46–60. [29]

——— (2012): "Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals," *Econometrica*, 80, 277–321. [27,29]

DINARDO, J., N. M. FORTIN, AND T. LEMIEUX (1996): "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1044. [30,33,34]

LECAM, L. (1986): *Asymptotic Methods in Statistical Decision Theory*. New York: Springer-Verlag. [7]

LEHMANN, E., AND J. ROMANO (2005): *Testing Statistical Hypotheses*. Springer Verlag. [4,7]

LEWANDOWSKI, M., M. RYZNAR, AND T. ZAK (1995): "Anderson Inequality Is Strict for Gaussian and Stable Measures," *Proceedings of the American Mathematical Society*, 123, 3875–3880. [13]

LUENBERGER, D. (1969): *Optimization by Vector Space Methods*. John Wiley & Sons. [13,17,19,21,28]

ROBINSON, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954. [31]

STONE, C. J. (1985): "Additive Regression and Other Nonparametric Models," *The Annals of Statistics*, 13, 689–705. [31]

VAN DER VAART, A. (1991): "An Asymptotic Representation Theorem," *International Statistical Review*, 59, 97–121. [2,5]

——— (1998): *Asymptotic Statistics*. Cambridge University Press. [6,8]

VAN DER VAART, A., AND J. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer Verlag. [7-9,15]

WILLIAMS, D. (1991): *Probability With Martingales*. New York: Cambridge University Press. [3,6]

ZEIDLER, E. (1984): *Nonlinear Functional Analysis and Its Applications III*. New York: Springer-Verlag. [18,21, 25]

Co-editor Lars Peter Hansen handled this manuscript.

Manuscript received 22 June, 2015; final version accepted 1 June, 2018; available online 1 June, 2018.