Supplement to "Nonparametric estimation of triangular simultaneous equations models under weak identification"

(*Quantitative Economics*, Vol. 11, No. 1, January 2020, 161–202)

Sukjin Han
Department of Economics, University of Texas at Austin

Appendix B: Supplemental Appendix

This Supplemental Appendix contains the proof of identification (Theorem 2.2) in Section 2, the remaining proofs of the lemmas introduced in Appendix A of the main paper, the proof of the sufficiency of the assumptions for the technical assumptions, and the proof of Theorem 6.1 in Section 6.

B.1 *Proofs in identification analysis (Section 2)*

Throughout the Appendices, we suppress the subscript "0" for the true functions and, when no confusion arises, the subscript "n" of the true reduced-form function in Assumption L.

In this section, we prove Lemma 2.1 and Theorem 2.2. For Lemma 2.1, we first introduce a preliminary lemma. For nonempty sets A and B, define the following set:

$$A+B=\{a+b:(a,b)\in A\times B\}. \tag{B.1}$$

Then, for nonempty sets A, B, and C,

$$A+B=B+A \text{ (commutative)}, \tag{Rule 1}$$

$$A+(B\cup C)=(A+B)\cup(A+C) \text{ (distributive 1)}, \tag{Rule 2}$$

$$A+(B\cap C)=(A+B)\cap(A+C) \text{ (distributive 2)}, \tag{Rule 3}$$

$$(A+B)^c\subset A+B^c, \tag{Rule 4}$$

where the last rule is less obvious than the others but can be shown to hold. Let λ_{Leb} denote a Lebesgue measure on $\mathbb{R}^{d_x}$, and $\partial\mathcal{V}$ and $\text{int}(\mathcal{V})$ denote the boundary and interior of $\mathcal{V}$, respectively.

Lemma B.1. *Suppose Assumptions* ID1 *and* ID2$'$(a)(i) *and* (ii) *hold. Suppose* $\mathcal{Z}^r\neq\phi$ *and* $\mathcal{Z}^0\neq\phi$. *Then* (a) $\{\Pi(z)+v:z\in\mathcal{Z}^0,v\in\text{int}(\mathcal{V})\}\subset\mathcal{X}^r$ *and* (b) $\lambda_{\text{Leb}}(\Pi(\mathcal{Z}^0))=0$ *and* $\partial\mathcal{V}$ *is countable.*

Sukjin Han: sukjin.han@austin.utexas.edu

 https://doi.org/10.3982/QE975

We prove the main lemma first.

Proof of Lemma 2.1. When $\mathcal{Z}^r = \phi$ or $\mathcal{Z}^r = \mathcal{Z}$, we trivially have $\mathcal{X}^r = \mathcal{X}$. Suppose $\mathcal{Z}^r \neq \phi$ and $\mathcal{Z}^0 \neq \phi$. First, under Assumption ID2′(b) that $\mathcal{V} = \mathbb{R}^{d_x}$, we have the conclusion by

$$\mathcal{X}^r = \left\{\Pi(z) + v : z \in \mathcal{Z}^r, v \in \mathbb{R}^{d_x}\right\} = \mathbb{R}^{d_x} = \left\{\Pi(z) + v : x \in \mathcal{Z}, v \in \mathbb{R}^{d_x}\right\} = \mathcal{X}.$$

Now suppose Assumption ID2′(a) holds. By Assumption ID2′(a)(iii), for $z \in \mathcal{Z}^0$, the joint support of (z, v) is $\mathcal{Z}^0 \times \mathcal{V}$. Hence

$$\left\{\Pi(z) + v : z \in \mathcal{Z}^0, v \in \operatorname{int}(\mathcal{V})\right\} = \left\{\Pi(z) + v : (z, v) \in \mathcal{Z}^0 \times \operatorname{int}(\mathcal{V})\right\} = \Pi\big(\mathcal{Z}^0\big) + \operatorname{int}(\mathcal{V}).$$

But by Lemma B.1(a), $\Pi(\mathcal{Z}^0) + \operatorname{int}(\mathcal{V}) \subset \mathcal{X}^r$ or contrapositively, $\mathcal{X} \backslash \mathcal{X}^r \subset (\Pi(\mathcal{Z}^0) + \operatorname{int}(\mathcal{V}))^c$. Also, by (Rule 4), $(\Pi(\mathcal{Z}^0) + \operatorname{int}(\mathcal{V}))^c \subset \Pi(\mathcal{Z}^0) + \partial\mathcal{V}$. Therefore,

$$\mathcal{X} \backslash \mathcal{X}^r \subset \Pi\big(\mathcal{Z}^0\big) + \partial\mathcal{V}. \tag{B.2}$$

Let $\partial\mathcal{V} = \{\nu_1, \nu_2, \ldots, \nu_k, \ldots\} = \bigcup_{k=1}^{\infty}\{\nu_k\}$ by Lemma B.1(b). Then

$$\begin{aligned}\lambda_{\text{Leb}}\big(\Pi\big(\mathcal{Z}^0\big) + \partial\mathcal{V}\big) &= \lambda_{\text{Leb}}\left(\Pi\big(\mathcal{Z}^0\big) + \left(\bigcup_{k=1}^{\infty}\{\nu_k\}\right)\right) = \lambda_{\text{Leb}}\left(\bigcup_{k=1}^{\infty}\big(\Pi\big(\mathcal{Z}^0\big) + \{\nu_k\}\big)\right) \\ &\leq \sum_{k=1}^{\infty}\lambda_{\text{Leb}}\big(\Pi\big(\mathcal{Z}^0\big) + \{\nu_k\}\big) = \sum_{k=1}^{\infty}\lambda_{\text{Leb}}\big(\Pi\big(\mathcal{Z}^0\big)\big) = 0,\end{aligned}$$

where the second equality is from (Rule 2) and the third equality by the property of Lebesgue measure. The last equality is by Lemma B.1(b) that $\lambda_{\text{Leb}}(\Pi(\mathcal{Z}^0)) = 0$. Since x is continuously distributed, by (B.2), $\Pr[x \in \mathcal{X} \backslash \mathcal{X}^r] \leq \Pr[x \in \Pi(\mathcal{Z}^0) + \partial\mathcal{V}] = 0$. □

In the following proofs, we explicitly distinguish the r.v.'s with their realization. Let ξ, ζ, and ν denote the realizations of x, z, and v, respectively. We now prove Lemma B.1.

Proof of Lemma B.1(a). First, we claim that for any $\pi \in \Pi(\mathcal{Z}^0)$ there exists $\bigcup_{n=1}^{\infty}\{\pi_n\} \subset \Pi(\mathcal{Z}^r)$ such that $\lim_{n\to\infty}\pi_n = \pi$. By Proposition 4.21(a) of Lee (2011, p. 92), for any space $\mathcal{S}$, the path components of $\mathcal{S}$ form a partition of $\mathcal{S}$. Note that a path component of $\mathcal{S}$ is a maximal nonempty path connected subset of $\mathcal{S}$. Then for $\mathcal{Z}^0 \subset \mathbb{R}^{d_z}$, we have $\mathcal{Z}^0 = \bigcup_{\iota\in I}\mathcal{Z}^0_\iota$ where partitions $\mathcal{Z}^0_\iota$ are path components. Note that, since $\mathcal{Z}^0_\iota$ is path connected, for any ζ and $\tilde{\zeta}$ in $\mathcal{Z}^0_\iota$, there exists a path in $\mathcal{Z}^0_\iota$, namely, a piecewise continuously differentiable function $\gamma : [0,1] \to \mathcal{Z}^0_\iota$ such that $\gamma(0) = \zeta$ and $\gamma(1) = \tilde{\zeta}$. Note that $\{\gamma(t) : t \in [0,1]\} \subset \mathcal{Z}^0_\iota$. Consider a composite function $\Pi \circ \gamma : [0,1] \to \Pi(\mathcal{Z}^0_\iota) \subset \mathbb{R}^{d_x}$. Then $\Pi(\gamma(\cdot))$ is differentiable, and by the mean value theorem, there exists $t^* \in [0,1]$ such that

$$\Pi\big(\gamma(1)\big) - \Pi\big(\gamma(0)\big) = \frac{\partial\Pi\big(\gamma(t^*)\big)}{\partial t}(1-0) = \frac{\partial\Pi\big(\gamma(t^*)\big)}{\partial\zeta'}\frac{\partial\gamma(t^*)}{\partial t}.$$

Note that $\partial\Pi(\gamma(t^*))/\partial\zeta' = \mathbf{0}_{d_x\times d_x}$ since $\gamma(t^*) \in \mathcal{Z}^0_\iota \subset \mathcal{Z}^0$ and $d_x = 1$. This implies that $\Pi(\gamma(1)) = \Pi(\gamma(0))$ or $\Pi(\zeta) = \Pi(\tilde{\zeta})$. Therefore, for any $\zeta \in \mathcal{Z}^0_\iota$, $\Pi(\zeta) = c_\iota$ for some constant c_ι.

Consider any $\bigcup_{n=1}^{\infty}\{\zeta_n\} \subset \mathcal{Z}^0 \subset \mathcal{Z}$ such that $\lim \zeta_n = \bar{\zeta} \in \mathcal{Z}$. Then for each n, $\partial \Pi(\zeta_n)/\partial\zeta' = (0, 0, \dots, 0) = \mathbf{0}$ by the definition of $\mathcal{Z}^0$, and $\partial\Pi(\bar{\zeta})/\partial\zeta' = \partial\Pi(\lim_{n\to\infty}\zeta_n)/\partial\zeta' = \lim_{n\to\infty}\partial\Pi(\zeta_n)/\partial\zeta' = \mathbf{0}$ where the second equality is by continuity of $\partial\Pi(\cdot)/\partial\zeta'$. Therefore, $\bar{\zeta} \in \mathcal{Z}^0$, and hence $\mathcal{Z}^0$ is closed, which implies that $\mathcal{Z}^0_\iota$ is closed for each ι. That is, $\mathcal{Z}^0$ is partitioned to a closed disjoint union of $\mathcal{Z}^0_\iota$'s. But Assumption ID2′(a)(ii) says $\mathcal{Z}$ is a connected set in Euclidean space (i.e., $\mathbb{R}^{d_z}$). Therefore, for each $\iota \in I$, $\mathcal{Z}^0_\iota$ must contain accumulation points of $\mathcal{Z}^r$ (Taylor (1965, p. 76)). Now, for any $\pi = \Pi(\zeta) \in \Pi(\mathcal{Z}^0)$, it satisfies that $\zeta \in \mathcal{Z}^0_\iota$ for some $\iota \in I$. Let $\zeta_c \in \mathcal{Z}^0_\iota$ be an accumulation point of $\mathcal{Z}^r$, that is, there exists $\bigcup_{n=1}^{\infty}\{\zeta_n\} \subset \mathcal{Z}^r$ such that $\lim_{n\to\infty}\zeta_n = \zeta_c$. Then it follows that $\pi = \Pi(\zeta) = c_\iota = \Pi(\zeta_c) = \Pi(\lim_{n\to\infty}\zeta_n) = \lim_{n\to\infty}\Pi(\zeta_n)$, where the second and third equalities are from $\Pi(\zeta) = c_\iota$ for $\zeta \in \mathcal{Z}^0_\iota$ and the fourth by continuity of $\Pi(\cdot)$. Let $\pi_n = \Pi(\zeta_n)$, then $\pi_n \in \Pi(\mathcal{Z}^r)$ for every $n \geq 1$. Therefore, we can conclude that for any $\pi \in \Pi(\mathcal{Z}^0)$, there exists $\bigcup_{n=1}^{\infty}\{\pi_n\} \subset \Pi(\mathcal{Z}^r)$ such that $\lim_{n\to\infty}\pi_n = \pi$.

Next, we prove that if $\xi \in \{\Pi(z) + v : z \in \mathcal{Z}^0, v \in \operatorname{int}(\mathcal{V})\}$ then $\xi \in \mathcal{X}^r$. Suppose $\xi \in \{\Pi(z) + v : z \in \mathcal{Z}^0, v \in \operatorname{int}(\mathcal{V})\}$, that is, $\xi = \pi + \nu$ for $\pi \in \Pi(\mathcal{Z}^0)$ and $\nu \in \operatorname{int}(\mathcal{V})$. Then, by the result above, there exists $\bigcup_{n=1}^{\infty}\{\pi_n\} \subset \Pi(\mathcal{Z}^r)$ such that $\lim_{n\to\infty}\pi_n = \pi$. Define a sequence $\nu_n = \xi - \pi_n$ for $n \geq 1$. Notice that ν_n is not necessarily in $\mathcal{V}$. But, $\nu_n = (\pi + \nu) - \pi_n = \nu + (\pi - \pi_n)$, hence $\lim_{n\to\infty}\nu_n = \nu$. Since $\nu \in \operatorname{int}(\mathcal{V})$, there exists an open neighborhood $B_\varepsilon(\nu)$ of ν for some ε such that $B_\varepsilon(\nu) \subset \operatorname{int}(\mathcal{V})$. Also, by the fact that $\lim_n \nu_n = \nu$, there exists N_ε such that for all $n \geq N_\varepsilon$, $\nu_n \in B_\varepsilon(\nu)$. Therefore, by conveniently taking $n = N_\varepsilon$, ξ satisfies that $\xi = \pi_{N_\varepsilon} + \nu_{N_\varepsilon}$ where $\pi_{N_\varepsilon} \in \Pi(\mathcal{Z}^r)$ and $\nu_{N_\varepsilon} \in B_\varepsilon(\nu) \in \operatorname{int}(\mathcal{V}) \subset \mathcal{V}$. That is, $\xi \in \mathcal{X}^r$. □

Proof of Lemma B.1(b). Recall $d_x = 1$. Note that $\mathcal{V} \subset \mathbb{R}$ can be expressed by a union of disjoint intervals. Since we are able to choose a rational number in each interval, the union is a countable union. Since each interval has at most two end points which are the boundary of it, $\partial\mathcal{V}$ is countable. To prove that $\lambda_{\text{Leb}}(\Pi(\mathcal{Z}^0)) = 0$, note that $\mathcal{Z}^0$ is the support where $\partial\Pi(z)/\partial z_k = 0$ for $k \leq d_z$. Therefore, its bilateral (directional) derivative $D_\alpha\Pi(z)$ in the direction $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{d_z})'$ satisfies $D_\alpha\Pi(z) = \sum_{k=1}^{d_z}\alpha_k \cdot \partial\Pi(z)/\partial z_k = 0$. Since the bilateral derivative is zero, each unilateral derivative is also zero; see, for example, Giorgi, Guerraggio, and Thierfelder (2004, p. 94) for the definitions of various derivatives. Then, by Corollary 6.1.3 in Garg (1998), $\lambda_{\text{Leb}}(\Pi(\mathcal{Z}^0)) = 0$. □

Proof of Theorem 2.2. Consider equation (2.2) with $z = z_2$,

$$E[y|x, z] = E[y|v, z] = g\big(\Pi(z) + v\big) + \lambda(v), \tag{2.2}$$

where the conditional expectations and $\Pi(\cdot)$ are consistently estimable, and v can also be estimated. By differentiating both sides of (2.2) with respect to z, we have

$$\frac{\partial E[y|v, z]}{\partial z'} = \frac{\partial g(x)}{\partial x'} \cdot \frac{\partial\Pi(z)}{\partial z'}. \tag{B.3}$$

Now, suppose $\Pr[z \in \mathcal{Z}^r] > 0$. For any fixed value $\bar{z}$ such that $\bar{z} \in \mathcal{Z}^r$, we have $\operatorname{rank}(\partial\Pi(\bar{z})/\partial z') = d_x$ by definition, hence the system of equations (B.3) has a unique solution $\partial g(x)/\partial x'$ for x in the conditional support $\mathcal{X}_{\bar{z}}$. That is, $\partial g(x)/\partial x'$ is locally identified for

$x \in \mathcal{X}_{\tilde{z}}$. Now, since the above argument is true for any $z \in \mathcal{Z}^r$, we have that $\partial g(x)/\partial x'$ is identified on $x \in \mathcal{X}^r$. Now by Assumption ID2, the difference between $\mathcal{X}^r$ and $\mathcal{X}$ has probability zero, thus $\partial g(x)/\partial x'$ is identified. Once $\partial g(x)/\partial x'$ is identified, we can identify $\partial \lambda(v)/\partial v'$ by differentiating (2.2) with respect to v:

$$\frac{\partial E[y|v,z]}{\partial v'} = \frac{\partial g(x)}{\partial x'} + \frac{\partial \lambda(v)}{\partial v}.$$

Next, we prove the necessity part of the theorem. Suppose $\Pr[z \in \mathcal{Z}^r] = 0$. This implies $\Pr[z \in \mathcal{Z}^0] = 1$, but since $\mathcal{Z}^0$ is closed $\mathcal{Z}^0 = \mathcal{Z}$. Therefore, for any $z \in \mathcal{Z}$, the system of equations (B.3) either has multiple solutions or no solution, and hence $g(\Pi(z)+v)$ is not identified. □

B.2 *Technical proofs*

B.2.1 *Proofs of sufficiency of Assumptions B, C, D, and L*

Proof that B and L imply B†. For simplicity, assume $\Pr[z \in \mathcal{Z}^r(\tilde{\Pi})] = 1$ and we prove after replacing Q^{r*} with Q^* in Assumption B†(i). Note that $\tilde{\Pi}(\cdot)$ is piecewise one-to-one. Here, we prove the case where $\tilde{\Pi}(\cdot)$ is one-to-one. The general cases where $\tilde{\Pi}(\cdot)$ is piecewise one-to-one or where $0 < \Pr[z \in \mathcal{Z}^r(\tilde{\Pi})] < 1$ can be followed by conditioning on z in appropriate subset of $\mathcal{Z}$.

Consider the change of variables of $u = (z, v)$ into $\tilde{u} = (\tilde{z}, \tilde{v})$ where $\tilde{z} = \tilde{\Pi}(z)$ and $\tilde{v} = v$. Then it follows that $p^{*K}(u_i) = [1 \vdots \tilde{z}_i \partial p^\kappa(v_i)' \vdots p^\kappa(v_i)']' = p^K(\tilde{u}_i)$ where $p^K(\tilde{u}_i)$ is one particular form of a vector of approximating functions as specified in NPV (pp. 572–573). Moreover, the joint density of $\tilde{u}$ is

$$f_{\tilde{u}}(\tilde{z},\tilde{v}) = f_u\big(\tilde{\Pi}^{-1}(\tilde{z}),\tilde{v}\big) \cdot \begin{vmatrix} \frac{\partial \tilde{\Pi}^{-1}(\tilde{z})}{\partial \tilde{z}} & 0 \\ 0 & 1 \end{vmatrix} = f_u\big(\tilde{\Pi}^{-1}(\tilde{z}),\tilde{v}\big) \cdot \frac{\partial \tilde{\Pi}^{-1}(\tilde{z})}{\partial \tilde{z}}.$$

Since $\partial \tilde{\Pi}^{-1}(\tilde{z})/\partial \tilde{z} \neq 0$ by $\tilde{\Pi} \in \mathcal{C}_1(\mathcal{Z})$ (bounded derivative) and f_u is bounded away from zero and the support of u is compact by Assumption B, the support of $\tilde{u}$ is also compact and $f_{\tilde{u}}$ is also bounded away from zero. Then, by the proof of Theorem 4 in Newey (1997, p. 167), $\lambda_{\min}(Ep^K(\tilde{u}_i)p^K(\tilde{u}_i)')$ is bounded away from zero. Therefore, $\lambda_{\min}(Q^*)$ is bounded away from zero for all κ.

As for Q_1 that does not depend on the effect of weak instruments, the density of z being bounded away from zero implies that $\lambda_{\min}(Q_1)$ is bounded away from zero for all L by Newey (1997, p. 167). The maximum eigenvalues of Q^*, Q, and Q_1 are bounded by fixed constants by the fact that the polynomials are defined on bounded sets and by Assumption L that $\tilde{\Pi}(\cdot) \in \mathcal{C}_1(\mathcal{Z})$. Lastly, note that q_{jj} is either $E[p_j(x)^2]$ or $E[p_j(v)^2]$. But by Assumption B, the density of v is bounded below, and hence $E[p_j(v)^2]$ is bounded below. Similarly, $E[p_j(x)^2]$ is also bounded below since eventually x converges to v by Assumption L. □

Proof that C implies C†. The results in (i) follow by Daubechies (1992), and the results in (ii) follow by Chen (2007). □

PROOF THAT D IMPLIES D†. It follows from Chen (2007) that with wavelest and B-splines, $\zeta_r^v(K) = K^{\frac{1}{2}+r}$, and similarly for the restriction on L. The same results holds for $\xi_r(L)$. □

B.2.2 *Matrix algebra* The following are mathematical lemmas that are useful in proving Lemma A.1 and other results.

LEMMA B.2. *For symmetric $k \times k$ matrices A and B, let $\lambda_j(A)$ and $\lambda_j(B)$ denote their jth eigenvalues such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$. Then the following inequality holds: For $1 \leq j \leq k$,*

$$\left|\lambda_j(A) - \lambda_j(B)\right| \leq \left|\lambda_1(A - B)\right| \leq \|A - B\|.$$

LEMMA B.3. *If $K(n) \times K(n)$ symmetric random sequence of matrices, A_n satisfies $\lambda_{\max}(A_n) = O_p(a_n)$, then $\|B_n A_n\| \leq \|B_n\| O_p(a_n)$ for a given sequence of matrices B_n.*

PROOF OF LEMMA B.2. First, by Weyl (1912), for symmetric $k \times k$ matrices C and D

$$\lambda_{i+j-1}(C + D) \leq \lambda_i(C) + \lambda_j(D), \tag{B.4}$$

where $i + j - 1 \leq k$. Also, for any $k \times 1$ vector a such that $\|a\| = 1$, $(a'Da)^2 = \text{tr}(a'Daa' \times Da) = \text{tr}(DDaa') \leq \text{tr}(DD)\,\text{tr}(aa') = \text{tr}(DD)$. Since $\lambda_j(D) = a'Da$ for some a with $\|a\| = 1$, we have

$$\left|\lambda_j(D)\right| \leq \|D\|, \tag{B.5}$$

for $1 \leq j \leq k$. Now, in (B.4) and (B.5), take $j = 1$, $C = B$, and $D = A - B$, and we have the desired results. □

PROOF OF LEMMA B.3. Let A_n have eigenvalue decomposition $A_n = UDU^{-1}$. Then $\|B_n A_n\|^2 = \text{tr}(B_n U D^2 U^{-1} B_n') \leq \text{tr}(B_n U U^{-1} B_n') \cdot \lambda_{\max}(A_n)^2 = \|B_n\|^2 O_p(n^\delta)^2$. □

B.3 *Proof of asymptotic normality (Section 6)*

Assumption G in Section 6 implies the following technical assumption. Define $\tilde{Q} = P'P/n$ where $P = (p^K(w_1), \ldots, p^K(w_n))'$, $Q_\tau = Q + \tau_n I$, and

$$\Delta_{\hat{Q}} = \zeta_1^v(\kappa)^2 \Delta_\pi^2 + \kappa^{1/2} \zeta_1^v(\kappa) \Delta_\pi, \qquad \Delta_{\tilde{Q}} = \zeta_0^v(\kappa)\sqrt{\log(\kappa)/n}, \qquad \Delta_Q = \Delta_{\hat{Q}} + \Delta_{\tilde{Q}},$$

$$\Delta_{Q_1} = \xi(L)\sqrt{\log(L)/n}, \qquad \Delta_\pi = \sqrt{L/n} + L^{-s_\pi/d_z},$$

$$\Delta_h = R_n\left(\sqrt{K/n} + K^{-s/d_x} + \tau_n R_n \lambda_n + \Delta_\pi\right),$$

$$\Delta_H = L^{1/2} \zeta_1^v(\kappa) \Delta_\pi + K^{1/2} \xi(L)/\sqrt{n}.$$

ASSUMPTION G†. *The following quantities converge to zero as $n \to \infty$: $\sqrt{n}K^{-s/d_x}$, $\sqrt{n}L^{-s_\pi/d_z}$, $L^{1/2}\Delta_{Q_1}$, $R_n L^{1/2} \Delta_H$, $R_n^3 K^{1/2} \Delta_Q$, $L^{1/2} \zeta_0^v(K) \zeta_1^v(K) \Delta_h$, $R_n^2(\zeta_0^v(K)\sqrt{K} + \xi(L) \times \sqrt{L})/\sqrt{n}$. Also, $\tau_n R_n \lambda_n \leq C K^{-s/d_x}$ for some $C > 0$.*

PROOF THAT ASSUMPTION G IMPLIES ASSUMPTION G†. First, by $\sqrt{n}K^{-s/d_x} \to 0$ and $\sqrt{n}L^{-s_\pi/d_z} \to 0$, we have $\Delta_\pi = \sqrt{L/n}(1 + L^{-1/2}\sqrt{n}L^{-s_\pi/d_z}) = O(\sqrt{L/n})$ and $\Delta_h = O(R_n(\sqrt{K/n} + \sqrt{L/n}))$ (since $\tau_n^{1/2}\lambda_n \le CK^{-s/d_x}$). Therefore,

$$\begin{aligned}
L^{1/2}\Delta_{Q_1} &= O\big(L^{1/2}\xi(L)\sqrt{\log(L)/n}\big),\\
R_n L^{1/2}\Delta_H &= O\big(R_n\big(\zeta_1^v(K)L^{3/2} + K^{1/2}L^{1/2}\xi(L)\big)/\sqrt{n}\big),\\
R_n^3 K^{1/2}\Delta_Q &= O\big(R_n^3 K^{1/2}\big\{\zeta_1^v(K)^2 L/n + K^{1/2}\zeta_1^v(K)\sqrt{L/n} + \zeta_0^v(K)\sqrt{\log(K)/n}\big\}\big),\\
L^{1/2}\zeta_0^v(K)\zeta_1^v(K)\Delta_h &= O\big(R_n\zeta_0^v(K)\zeta_1^v(K)\big(K^{1/2}L^{1/2} + L\big)/\sqrt{n}\big).
\end{aligned}$$

Then by plugging in $\zeta_r^v(K) = K^{\frac{1}{2}+r}$ and $\xi(L) = L$, it is readily seen that Assumption G† is followed by Assumption G. □

PROOF OF THEOREM 6.1. The proof of Theorem 6.1 is a modification of the proof of Theorem 5.1 in NPV (with their trimming function being an identity function) in the setting of weak instruments and penalization. We use the components established in the proof of the convergence rate, which are distinct from NPV. The rest of the notation are the same as those of NPV. Let “MVE” abbreviate mean value expansion.

Under Assumptions A, B†(ii), and E and given our choice of basis functions, Lemma 2.1 in Chen and Christensen (2015) or Theorem 4.6 in Belloni et al. (2015) (which improves over Newey (1997)) yields $\|\tilde{Q} - Q\| = O_p(\Delta_{\tilde{Q}})$ and $\|\hat{Q}_1 - I\| = O_p(\Delta_{Q_1})$ by letting $Q_1 = I$. Also similar to the proof of Lemma A1 of NPV, but using $\|\hat{Q} - \tilde{Q}\| = \|\frac{1}{n}\sum_{i=1}^n(\hat{p}_i\hat{p}_i' - p_i p_i')\|$ instead of their (A.5), we have $\|\hat{Q} - \tilde{Q}\| = O_p(\Delta_{\hat{Q}})$. Also,

$$\|\hat{Q}_\tau - Q_\tau\| \le \|\hat{Q} - Q\| \le \|\hat{Q} - \tilde{Q}\| + \|\tilde{Q} - Q\| = O_p(\Delta_Q).$$

Let $\tilde{x}$ is a vector of variables that includes x and z, and $\omega(\tilde{x}, \pi)$ a vector of functions of $\tilde{x}$ and π and, trivially, $\omega(\tilde{x}, \pi) = (x, x - \Pi_n(z))$. Let $r_i = r^L(z_i)$. Define

$$\begin{aligned}
V_\tau &= AQ_\tau^{-1}\big(\Sigma + HQ_1^{-1}\Sigma_1 Q_1^{-1}H'\big)Q_\tau^{-1}A',\\
\Sigma &= E\big[p_i p_i' \operatorname{var}(y_i|\tilde{x}_i)\big], \qquad H = E\big[p_i\big\{\big[\partial h(w_i)/\partial w\big]'\partial\omega\big(X_i, \Pi_n(z_i)\big)/\partial\pi\big\}r_i'\big].
\end{aligned}$$

Note that H is a channel through which the first-stage estimation error affects into the variance of the estimator of $h(\cdot)$. We first prove

$$\sqrt{n}V_\tau^{-1/2}(\hat{\theta}_\tau - \theta_0) \to_d N(0, 1).$$

For notational simplicity, let $F = V_\tau^{-1/2}$. Let $h = (h(w_1), \ldots, h(w_n))'$ and $\tilde{h} = (h(\hat{w}_1), \ldots, h(\hat{w}_n))'$. Also let $\eta_i = y_i - h_0(w_i)$ and $\eta = (\eta_1, \ldots, \eta_n)'$. Let $\Pi_n = (\Pi_n(z_1), \ldots, \Pi_n(z_n))'$, $v_i = x_i - \Pi_n(z_i)$, and $U = (v_1, \ldots, v_n)'$. Similar to NPV, once we prove that (i) $\|FAQ_\tau^{-1}\| = O(R_n)$, (ii) $\sqrt{n}F[a(p^{K\prime}\tilde{\beta}) - a(h_0)] = o_p(1)$, (iii) $\sqrt{n}FA(\hat{P}'\hat{P} + n\tau_n I)^{-1}\hat{P}'(\tilde{h} - \hat{P}_{\#}\beta) = o_p(1)$, (iv) $FA\hat{Q}_\tau^{-1}\hat{P}'(h - \tilde{h})/\sqrt{n} = FAQ_\tau^{-1}HR'U/\sqrt{n} + o_p(1)$, and (v) $FA\hat{Q}_\tau^{-1}\hat{P}'\eta/\sqrt{n} = FAQ_\tau^{-1}P'\eta/\sqrt{n} + o_p(1)$ below, then we will have

$$\sqrt{n}V_\tau^{-1/2}(\hat{\theta}_\tau - \theta_0) = \sqrt{n}F\big(a\big(p^{K\prime}\hat{\beta}_\tau\big) - a\big(p^{K\prime}\tilde{\beta}\big) + a\big(p^{K\prime}\tilde{\beta}\big) - a(h_0)\big)$$

$$
\begin{aligned}
&= \sqrt{n} FA\hat{\beta}_\tau - \sqrt{n} FA\tilde{\beta} + o_p(1) \\
&= \sqrt{n} FA\big(\hat{P}'\hat{P} + n\tau_n I\big)^{-1}\hat{P}'(h+\eta) - \sqrt{n} FA\big(\hat{P}'\hat{P} + n\tau_n I\big)^{-1}\hat{P}'\tilde{h} \\
&\quad + \sqrt{n} FA\big(\hat{P}'\hat{P} + n\tau_n I\big)^{-1}\hat{P}'(\tilde{h} - \hat{P}_{\#}\tilde{\beta}) + o_p(1) \\
&= FA\hat{Q}_\tau^{-1}\hat{P}'\eta/\sqrt{n} - FA\hat{Q}_\tau^{-1}\hat{P}'(h-\tilde{h})/\sqrt{n} + o_p(1) \\
&= FAQ_\tau^{-1}\big(\hat{P}'\eta/\sqrt{n} + HR'U/\sqrt{n}\big) + o_p(1). \qquad \text{(B.6)}
\end{aligned}
$$

Then, for any vector ϕ with $\|\phi\| = 1$, let $Z_{in} = Z_{1,in}\eta_i + Z_{2,in}u_i$ with $Z_{1,in} = \phi' FAQ_\tau^{-1} p_i/\sqrt{n}$ and $Z_{2,in} = \phi' FAQ_\tau^{-1} H r_i/\sqrt{n}$. Note Z_{in} is i.i.d. for each n. Also $EZ_{in} = 0$, $\mathrm{var}(Z_{in}) = 1/n$. Furthermore, using $\|FAQ_\tau^{-1}\| = O(R_n)$ and $\|FAQ_\tau^{-1}H\| \le C\|FAQ_\tau^{-1}\| = O(R_n)$ by $CI - HH'$ being p.s.d., we can verify the Lindbergh condition that, for any $\varepsilon > 0$,

$$
\begin{aligned}
&nE\big[1\{|Z_{in}| > \varepsilon\}Z_{in}^2\big] \\
&\quad = n\varepsilon^2 E\big[1\{|Z_{in}/\varepsilon| > 1\}(Z_{in}/\varepsilon)^2\big] \le n\varepsilon^2 E\big[(Z_{in}/\varepsilon)^4\big] \\
&\quad \le \frac{n\varepsilon^2}{n^2\varepsilon^4}\|\phi\|^4 \|FAQ_\tau^{-1}\|^4 \{\|p_i\|^2 E\big[\|p_i\|^2 E\big[\eta_i^4|\tilde{x}_i\big]\big] + \|r_i\|^2 E\big[\|r_i\|^2 E\big[u_i^4|z_i\big]\big]\} \\
&\quad \le CO\big(R_n^4\big)\{\zeta_0^v(K)^2 E\|p_i\|^2 + \xi(L)^2 E\|r_i\|^2\}/n \\
&\quad \le CO\big(R_n^4\big)\{\zeta_0^v(K)^2 \,\mathrm{tr}(Q) + \xi(L)^2 \,\mathrm{tr}(Q_1)\}/n \\
&\quad \le O\big(R_n^4\{\zeta_0^v(K)^2 K + \xi(L)^2 L\}/n\big) = o(1)
\end{aligned}
$$

by G†. Then $\sqrt{n}F(\hat{\theta}_\tau - \theta_0) \to_d N(0,1)$ by the Lindbergh–Feller theorem and (B.6).

Now, we proceed with detailed proofs of (i)–(v). For simplicity, the remainder of the proof will be given for the scalar $\Pi(z)$ case and under Assumption F; A similar proof under Assumption H can be derived by analogously modifying the proof of Lemma A2 in NPV. Under Assumption F and by CS, for $h_K(w) = p^K(w)'\beta_K$, $|a(h_K)| = |A\beta_K| \le \|A\|\|\beta_K\| = \|A\|(Eh_K(w)^2)^{1/2}$ so $\|A\| \to \infty$. But since $\lambda_{\min}(Q_\tau^{-1})$ is bounded away from zero, $Q_\tau^{-1} \ge \lambda_{\min}(Q_\tau^{-1})I = CI$. Also since $\sigma^2(\tilde{x}) = \mathrm{var}(y|\tilde{x})$ is bounded away from zero by Assumption E and $\tau_n \to 0$, we have $\Sigma \ge CQ_\tau$. Hence

$$
V_\tau \ge AQ_\tau^{-1}\Sigma Q_\tau^{-1}A' \ge CAQ_\tau^{-1}Q_\tau Q_\tau^{-1}A' = CAQ_\tau^{-1}A' \ge \tilde{C}\|A\|^2. \qquad \text{(B.7)}
$$

Therefore, F is bounded. Now, by (B.7),

$$
\|FAQ_\tau^{-1/2}\|^2 = \mathrm{tr}\big(FAQ_\tau^{-1}A'F\big) \le \mathrm{tr}(CFV_\tau F) = C.
$$

Also $\lambda_{\max}(Q_\tau^{-1}) = O(R_n^2)$, which can readily be shown analogous to (A.26). Using these results,

$$
\|FAQ_\tau^{-1}\| = \|FAQ_\tau^{-1/2}Q_\tau^{-1/2}\| \le \lambda_{\max}\big(Q_\tau^{-1}\big)^{1/2}\|FAQ_\tau^{-1/2}\| = O(R_n).
$$

Then, combining with (A.26), (i) follows by

$$
\|FA\hat{Q}_\tau^{-1}\| \le \|FAQ_\tau^{-1}\| + \|FAQ_\tau^{-1}(\hat{Q}_\tau - Q_\tau)\hat{Q}_\tau^{-1}\|
$$

$$\leq O(R_n) + O(R_n)O_p\big(R_n^2\big)\|\hat{Q}_\tau - Q_\tau\|$$
$$= O(R_n) + O_p\big(R_n^3\Delta_Q\big) = O_p(R_n),$$

where $R_n^3\Delta_Q = o_p(1)$ is by G†. Also

$$\begin{aligned}\|FA\hat{Q}_\tau^{-1/2}\|^2 &\leq \|FAQ_\tau^{-1/2}\|^2 + \mathrm{tr}\big(FAQ_\tau^{-1}(\hat{Q}_\tau - Q_\tau)\hat{Q}_\tau^{-1}A'F\big)\\ &\leq C + \|FA'Q_\tau^{-1}\|\|\hat{Q}_\tau - Q_\tau\|\|FA'\hat{Q}_\tau^{-1}\|\\ &\leq C + O(R_n)O_p(\Delta_Q)O(R_n) = O_p(1)\end{aligned}$$

by G†. To prove (ii), by C† and G†,

$$\begin{aligned}\|\sqrt{n}F\big[a\big(p^{K\prime}\tilde{\beta}\big) - a(h_0)\big]\| = \|\sqrt{n}F\big[a\big(p^{K\prime}\tilde{\beta} - h_0\big)\big]\| &\leq \sqrt{n}|F|\sup_w\big|p^K(w)'\tilde{\beta} - h_0(w)\big|\\ &\leq C\sqrt{n}K^{-s/d_x} = o_p(1).\end{aligned}$$

For (iii),

$$\begin{aligned}\|FA\hat{Q}_\tau^{-1}\hat{P}'(\tilde{h} - \hat{P}_\#\tilde{\beta})/\sqrt{n}\| &\leq \|FA\hat{Q}_\tau^{-1}\hat{P}'/\sqrt{n}\|\sqrt{n}\sup_w\big|p^K(w)'\tilde{\beta} - h_0(w)\big|\\ &\quad + \|n\tau_n FA\hat{Q}_\tau^{-1}D_{K^*}\tilde{\beta}/\sqrt{n}\|\\ &\leq \|FA\hat{Q}_\tau^{-1/2}\|\sqrt{n}O\big(K^{-s/d_x}\big) + \sqrt{n}\tau_n\|FA\hat{Q}_\tau^{-1}\|\|D_{K^*}\tilde{\beta}\|\\ &\leq O_p(1)O\big(\sqrt{n}K^{-s/d_x}\big) + O_p\big(\sqrt{n}\tau_n R_n K^{*-s/d_x-1/2}\big) = o_p(1)\end{aligned}$$

by G†. To prove (iv), let $\gamma = (\gamma_1, \ldots, \gamma_L)'$, $d_i = d(\tilde{x}_i) = [\partial h(w_i)/\partial w]'\partial\omega(\tilde{x}_i, \Pi_0(z_i))/\partial\pi$ and $\bar{H} = \sum \hat{p}_i d(\tilde{x}_i)r_i'/n$. By a second-order MVE of each $h(\hat{w}_i)$ around w_i

$$\begin{aligned}FA\hat{Q}_\tau^{-1}\hat{P}'(h - \tilde{h})/\sqrt{n} &= FA\hat{Q}_\tau^{-1}\sum_i\hat{p}_i d_i\big[\hat{\Pi}(z_i) - \Pi_n(z_i)\big]/\sqrt{n} + \hat{\rho}\\ &= FA\hat{Q}_\tau^{-1}\bar{H}\hat{Q}_1^{-1}R'U/\sqrt{n} + FA\hat{Q}_\tau^{-1}\bar{H}\hat{Q}_1^{-1}R'\big(\Pi_n - R'\gamma\big)/\sqrt{n}\\ &\quad + FA\hat{Q}_\tau^{-1}\sum_i\hat{p}_i d_i\big[r_i'\gamma - \Pi_n(z_i)\big]/\sqrt{n} + \hat{\rho}.\end{aligned} \tag{B.8}$$

But $\|\hat{\rho}\| \leq C\sqrt{n}\|FA\hat{Q}_\tau^{-1/2}\|\zeta_0^v(K)\sum_i\|\hat{\Pi}(z_i) - \Pi_n(z_i)\|^2/n = O_p(\sqrt{n}\zeta_0^v(K)\Delta_\pi^2) = o_p(1)$ by G†. Also, by d_i being bounded and $n\bar{H}\hat{Q}_1^{-1}\bar{H}'$ being equal to the matrix sum of squares from the multivariate regression of $\hat{p}_i d_i$ on r_i, $\bar{H}\hat{Q}_1^{-1}\bar{H}' \leq \sum_i\hat{p}_i\hat{p}_i' d_i^2/n \leq C\hat{Q} \leq C\hat{Q}_\tau$. Therefore, the second term in (B.8) becomes

$$\begin{aligned}&\|FA\hat{Q}_\tau^{-1}\bar{H}\hat{Q}_1^{-1}R'\big(\Pi_n - R'\gamma\big)/\sqrt{n}\|\\ &\leq \|FA\hat{Q}_\tau^{-1}\bar{H}\hat{Q}_1^{-1}R'/\sqrt{n}\|\sqrt{n}\sup_{\mathcal{Z}}\big|\Pi_n(z) - r^L(z)'\gamma\big|\\ &\leq \big\{\mathrm{tr}\big(FA\hat{Q}_\tau^{-1}\bar{H}\hat{Q}_1^{-1}\hat{Q}_1\hat{Q}_1^{-1}\bar{H}'\hat{Q}_\tau^{-1}A'F'\big)\big\}^{1/2}O\big(\sqrt{n}L^{-s_\pi/d_z}\big)\\ &\leq C\big\{\mathrm{tr}\big(FA\hat{Q}_\tau^{-1}\hat{Q}_\tau\hat{Q}_\tau^{-1}A'F'\big)\big\}^{1/2}O\big(\sqrt{n}L^{-s_\pi/d_z}\big)\end{aligned}$$

$$\leq C\|FA\hat{Q}_\tau^{-1/2}\|O(\sqrt{n}L^{-s_\pi/d_z}) = O_p(\sqrt{n}L^{-s_\pi/d_z}) = o_p(1)$$

by G†. Similarly, the third term is

$$\left\|FA\hat{Q}_\tau^{-1}\sum_i \hat{p}_i d_i[r_i'\gamma - \Pi_n(z_i)]/\sqrt{n}\right\| \leq C\|FA\hat{Q}_\tau^{-1/2}\|O(\sqrt{n}L^{-s_\pi/d_z})$$
$$= O_p(\sqrt{n}L^{-s_\pi/d_z}) = o_p(1).$$

Next, we consider the first term $FA\hat{Q}_\tau^{-1}\bar{H}\hat{Q}_1^{-1}R'U/\sqrt{n}$ in (B.8). Note that $E\|R'U/\sqrt{n}\|^2 = \operatorname{tr}(\Sigma_1) \leq C\operatorname{tr}(I_L) \leq L$ by $E[u^2|z]$ bounded, so by MR, $\|R'U/\sqrt{n}\| = O_p(L^{1/2})$. Also, note that

$$\|FA\hat{Q}_\tau^{-1}\bar{H}\hat{Q}_1^{-1}\| \leq O_p(1)\|FA\hat{Q}_\tau^{-1/2}\| = O_p(1).$$

Therefore,

$$\|FA\hat{Q}_\tau^{-1}\bar{H}(\hat{Q}_1^{-1} - I)R'U/\sqrt{n}\| \leq \|FA\hat{Q}_\tau^{-1}\bar{H}\hat{Q}_1^{-1}\|\|\hat{Q}_1 - I\|\|R'U/\sqrt{n}\|$$
$$= O_p(1)O_p(\Delta_{Q_1})O_p(L^{1/2}) = o_p(1)$$

by G†. Similarly,

$$\|FA\hat{Q}_\tau^{-1}(\bar{H} - H)R'U/\sqrt{n}\| \leq \|FA\hat{Q}_\tau^{-1}\|\|\bar{H} - H\|\|R'U/\sqrt{n}\|$$
$$= O_p(R_n)O_p(\Delta_H)O_p(L^{1/2}) = o_p(1),$$

where $\|\bar{H} - H\| = O_p(\Delta_H)$ instead of (A.12) in NPV, and the last equation by G†. Note that HH' is the population matrix mean-square of the regression of $p_i d_i$ on r_i so that $HH' \leq C$, it follows that $E\|HR'U/\sqrt{n}\|^2 = \operatorname{tr}(H\Sigma_1 H') \leq CK$ and, therefore, $\|HR'U/\sqrt{n}\| = O_p(K^{1/2})$. Then

$$\|FA(\hat{Q}_\tau^{-1} - Q_\tau^{-1})HR'U/\sqrt{n}\| \leq \lambda_{\max}(\hat{Q}_\tau^{-1})\|FAQ_\tau^{-1}\|\|Q_\tau - \hat{Q}_\tau\|\|HR'U/\sqrt{n}\|$$
$$= O(R_n^2)O(R_n)O_p(\Delta_Q)O_p(K^{1/2}) = o_p(1).$$

Combining the results above and by TR, $FA\hat{Q}_\tau^{-1}\bar{H}\hat{Q}_1^{-1}R'U/\sqrt{n} = FAQ^{-1}HR'U/\sqrt{n} + o_p(1)$, and thus we have the result of (iv).

Lastly, to prove (v), similar to (A.19),

$$\|\hat{Q}_\tau^{-1/2}(P - \hat{P})'\eta/\sqrt{n}\| = O_p(R_n\zeta_1^v(K)^2\Delta_\pi^2) = o_p(1)$$

by G† (and by (A.6) of NPV), which implies

$$\|FA\hat{Q}_\tau^{-1}(\hat{P} - P)'\eta/\sqrt{n}\| \leq \|FA\hat{Q}_\tau^{-1/2}\|\|\hat{Q}_\tau^{-1/2}(P - \hat{P})'\eta/\sqrt{n}\| = O_p(1)o_p(1) = o_p(1).$$

Also, by $E[\eta|\tilde{x}] = 0$,

$$E[\|FA(\hat{Q}_\tau^{-1} - Q_\tau^{-1})P'\eta/\sqrt{n}\|^2|\tilde{x}_i]$$
$$\leq \operatorname{tr}\Big(FA(\hat{Q}_\tau^{-1} - Q_\tau^{-1})\Big[\sum p_i p_i' \operatorname{var}(y_i|\tilde{x}_i)/n\Big](\hat{Q}_\tau^{-1} - Q_\tau^{-1})A'F\Big)$$

$$\leq C\,\mathrm{tr}\big(FA(\hat{Q}_\tau^{-1}-Q_\tau^{-1})\hat{Q}_\tau(\hat{Q}_\tau^{-1}-Q^{-1}{}_\tau)A'F\big)$$
$$= C\,\mathrm{tr}\big(FAQ_\tau^{-1}(\hat{Q}_\tau-Q_\tau)\hat{Q}_\tau^{-1}(\hat{Q}_\tau-Q_\tau)Q_\tau^{-1}A'F\big)$$
$$\leq O_p\big(R_n^2\big)\big\|FAQ_\tau^{-1}\big\|^2\|\hat{Q}_\tau-Q_\tau\|^2 \leq O_p\big(R_n^2\Delta_Q\big)^2 = o_p(1)$$

by G†. Combining all of the previous results and by TR, we have the result of (v).

Now we can prove

$$\sqrt{n}\hat{V}_\tau^{-1/2}(\hat{\theta}_\tau-\theta_0)\to_d N(0,1)$$

by showing $|F\hat{V}_\tau F-1|\to_p 0$. Then $V_\tau^{-1}\hat{V}_\tau\to_p 1$, so that $\sqrt{n}\hat{V}_\tau^{-1/2}(\hat{\theta}_\tau-\theta_0)=\sqrt{n}V_\tau^{-1/2}(\hat{\theta}-\theta_0)/(V_\tau^{-1}\hat{V}_\tau)^{1/2}\to_d N(0,1)$. The rest part of the proof can analogously be followed by the relevant part of the proof of NPV (pp. 600–601), using $Q\neq I$ because of weak instruments and $F=V_\tau^{-1/2}$. Therefore, the following replace the corresponding parts in the proof: For any matrix B, we have $\|B\Sigma\|\leq C\|BQ_\tau\|$ by $\Sigma\leq CQ_\tau$. Therefore,

$$\begin{aligned}
&\big\|FA\big(\hat{Q}_\tau^{-1}\Sigma\hat{Q}_\tau^{-1}-Q_\tau^{-1}\Sigma Q_\tau^{-1}\big)A'F'\big\| \\
&\quad\leq \big\|FA\big(\hat{Q}_\tau^{-1}-Q_\tau^{-1}\big)\Sigma\hat{Q}_\tau^{-1}A'F'\big\| + \big\|FAQ_\tau^{-1}\Sigma\big(\hat{Q}_\tau^{-1}-Q_\tau^{-1}\big)A'F'\big\| \\
&\quad\leq \big\|FA\hat{Q}_\tau^{-1}(Q_\tau-\hat{Q}_\tau)Q_\tau^{-1}\Sigma\hat{Q}_\tau^{-1}A'F'\big\| + \big\|FAQ_\tau^{-1}\Sigma Q_\tau^{-1}(Q_\tau-\hat{Q}_\tau)\hat{Q}_\tau^{-1}A'F'\big\| \\
&\quad\leq \big\|FA\hat{Q}_\tau^{-1}\big\|^2\big\|(Q_\tau-\hat{Q}_\tau)Q_\tau^{-1}\Sigma\big\| + \big\|FAQ_\tau^{-1}\big\|\big\|\Sigma Q_\tau^{-1}(Q_\tau-\hat{Q}_\tau)\big\|\big\|FA\hat{Q}_\tau^{-1}\big\| \\
&\quad\leq C\big\|FA\hat{Q}_\tau^{-1}\big\|^2\big\|(Q_\tau-\hat{Q}_\tau)Q_\tau^{-1}Q_\tau\big\| + C\big\|FAQ_\tau^{-1}\big\|\big\|Q_\tau Q_\tau^{-1}(Q_\tau-\hat{Q}_\tau)\big\|\big\|FA\hat{Q}_\tau^{-1}\big\| \\
&\quad\leq O_p\big(R_n^2\big)O_p(\Delta_Q)+O_p\big(R_n^2\big)O_p(\Delta_Q)=o_p(1)
\end{aligned}$$

by Assumption G†. Also note that in our proof, Q_τ is introduced by penalization but the treatment is the same as above. Also, recall $\zeta_r(K)\leq\zeta_r^v(K)$ and Δ_h and Δ_Q are redefined in this paper. Under G†, we can prove the following:

$$\big\|FA\hat{Q}_\tau^{-1}(\hat{\Sigma}_\tau-\tilde{\Sigma})\hat{Q}_\tau^{-1}A'F'\big\| \leq C\,\mathrm{tr}(\hat{D})\max_{i\leq n}|\hat{h}_{\tau,i}-h_i| \leq O_p(1)O_p\big(\zeta_0^v(K)\Delta_h\big)=o_p(1),$$

$$\begin{aligned}
\big\|FA\hat{Q}_\tau^{-1}(\tilde{\Sigma}-\Sigma)\hat{Q}_\tau^{-1}A'F'\big\| &\leq \big\|FA\hat{Q}_\tau^{-1}\big\|^2\|\tilde{\Sigma}-\Sigma\| \leq O_p\big(R_n^2\big)O_p\big(\Delta_Q+\zeta_0^v(K)\sqrt{\log(K)/n}\big) \\
&= o_p(1),
\end{aligned}$$

$$\begin{aligned}
\|\hat{H}_\tau-\bar{H}\| &\leq C\left(\sum_{i=1}^n\|\hat{p}_i\|^2\|r_i\|^2/n\right)^{1/2}\left(\sum_{i=1}^n|\hat{d}_{\tau,i}-d_i|^2/n\right)^{1/2} \\
&= O_p\big(\zeta_0^v(K)L^{1/2}\big)O_p\big(\zeta_1^v(K)\Delta_h\big)=o_p(1),
\end{aligned}$$

where $\hat{d}_{\tau,i}=[\partial\hat{h}_\tau(w_i)/\partial w]'\partial w(\tilde{x}_i,\Pi_0(z_i))/\partial\pi$. The second term on the r.h.s. of $\|\tilde{\Sigma}-\Sigma\|\leq O_p(\Delta_Q+\zeta_0^v(K)\sqrt{\log(K)/n})$ is derived by showing $\|\sum_{i=1}^n p_ip_i'\eta_i^2/n-\Sigma\|=\zeta_0^v(K)\times\sqrt{\log(K)/n})$, which improves the rate in NPV by applying Theorem 4.1 in Chen and Christensen (2015) or Theorem 4.6 in Belloni et al. (2015). The rest of the proof thus follows. □

References

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015), "Some new asymptotic theory for least squares series: Pointwise and uniform results." *Journal of Econometrics*, 186, 345–366. [6, 10]

Chen, X. (2007), "Large sample sieve estimation of semi-nonparametric models." In *Handbook of Econometrics*, Vol. 6, 5549–5632. [4, 5]

Chen, X. and T. M. Christensen (2015), "Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions." *Journal of Econometrics*, 188, 447–465. [6, 10]

Daubechies, I. (1992), *Ten Lectures on Wavelets*, Vol. 61. SIAM. [4]

Garg, K. M. (1998), *Theory of Differentiation*. Wiley. [3]

Giorgi, G., A. Guerraggio, and J. Thierfelder (2004), *Mathematics of Optimization: Smooth and Nonsmooth Case*. Elsevier. [3]

Lee, J. M. (2011), "Topological spaces." In *Introduction to Topological Manifolds*, 19–48, Springer. [2]

Newey, W. K. (1997), "Convergence rates and asymptotic normality for series estimators." *Journal of Econometrics*, 79, 147–168. [4, 6]

Taylor, A. E. (1965), *General Theory of Functions and Integration*. Dover Publications. [3]

Weyl, H. (1912), "Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)." *Mathematische Annalen*, 71, 441–479. [5]

Co-editor Andres Santos handled this manuscript.

Manuscript received 10 September, 2017; final version accepted 11 June, 2019; available online 1 July, 2019.