Supplement to "Specification testing for conditional moment restrictions under local identification failure"

(*Quantitative Economics*, Vol. 15, No. 3, July 2024, 849–891)

Prosper Dovonon
Department of Economics, Concordia University, CIREQ, and CIRANO

Nikolay Gospodinov
Research Department, Federal Reserve Bank of Atlanta

Proofs of Lemmas B.1, B.2, B.3, and B.4 in Appendix B

The proofs of Lemma B.2(i), Proposition S.2, and Theorem 5.1 rely on the following result.

Lemma S.1 (Theorem 7.4 of Roussas and Ioannides (1987)). *Suppose that F_n^m are the σ-algebras generated by $(x_n, x_{n+1}, \dots, x_m)$, where $\{x_i : i \in \mathbb{Z}\}$ is a stationary α-mixing process with mixing coefficient $\alpha(i)$. For some positive integer ℓ, let $\eta_i \in F_{s_i}^{t_i}$, where $s_1 < t_1 < s_2 < t_2 < \cdots < t_\ell$ and $s_{i+1} - t_i \geq \tau$ for all i. In addition, let $\|\eta\|_p = [E|\eta|^p]^{1/p}$ for $1 < p < \infty$ and $\|\eta\|_\infty = esssup|\eta|$, respectively, and assume further that*

$$\|\eta\|_{p_i} < \infty \quad \text{for some } 1 < p_i \leq \infty \text{ with } q = \sum_{i=1}^{\ell} \frac{1}{p_i} < 1.$$

Then

$$\left| \mathbb{E}\left(\prod_{i=1}^{\ell} \eta_i\right) - \prod_{i=1}^{\ell} \mathbb{E}(\eta_i) \right| \leq 10(\ell - 1)\alpha(\tau)^{1-q} \prod_{i=1}^{\ell} \|\eta_i\|_{p_i}.$$

Proof of Lemma B.1.

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{t=1}^{n} g^{(k)}(x_t) U_t(\bar{\theta})' - \mathbb{E}\big(g^{(k)}(x_t) U_t(\theta_0)'\big) \right\|_2 \\
&\quad \leq O_P\left(\sqrt{\frac{k}{n}}\right) + \left\| \mathbb{E}\big(g^{(k)}(x_t) U_t(\bar{\theta})'\big) - \mathbb{E}\big(g^{(k)}(x_t) U_t(\theta_0)'\big) \right\|_2 \\
&\quad \leq O_P\left(\sqrt{\frac{k}{n}}\right) + \left(\sum_{l=1}^{k} \sum_{r=1}^{m} \big[\mathbb{E}\big(g_l(x_t) U_{t,r}(\bar{\theta})\big) - \mathbb{E}\big(g_l(x_t) U_{t,r}(\theta_0)\big)\big]^2 \right)^{1/2}
\end{aligned}$$

Prosper Dovonon: prosper.dovonon@concordia.ca
Nikolay Gospodinov: nikolay.gospodinov@atl.frb.org

 https://doi.org/10.3982/QE2242

$$\leq O_P\left(\sqrt{\frac{k}{n}}\right) + \sqrt{km}\cdot c\cdot \|\bar\theta - \theta_0\|_2 = O_P\big(\sqrt{k}\cdot\big(n^{-1/2}\vee \|\bar\theta-\theta_0\|_2\big)\big). \qquad \square$$

Proof of Lemma B.2. (i) Recall that

$$\|\bar V_k - V_k\|_2 \leq k \max_{1\leq h,l\leq k}\big|[\bar V_k - V_k]_{hl}\big|.$$

Letting $s_{t,hl} = v_t(\theta_0) - E(v_t(\theta_0))$, with $v_t(\theta) = g_h(x_t)g_l(x_t)u_t(\theta)^2$, we have

$$\begin{aligned}
[\bar V_k - V_k]_{hl} &= \frac{1}{n}\sum_{t=1}^n v_t(\bar\theta) - \mathbb{E}\big[v_t(\theta_0)\big] \\
&= \left(\frac{1}{n}\sum_{t=1}^n s_{t,hl}\right) + \frac{2}{n}\sum_{t=1}^n g_h(x_t)g_l(x_t)u_t(\bar\theta)\nabla_\theta u_t(\tilde\theta)(\bar\theta-\theta_0) \\
&:= \big(a'\big) + \big(b'\big). \qquad \text{(S.1)}
\end{aligned}$$

The first equality follows by definition and the second one follows from a mean-value expansion of $(1/n)\sum_{t=1}^n v_t(\bar\theta)$ around θ_0 with $\tilde\theta \in (\bar\theta,\,\theta_0)$. We need to determine the order of magnitude of (a') and (b'). Pick $0<\xi<1/2$ and let $\alpha = (1-2\xi)/2$. We have

$$\mathbb{E}\big(n^\xi a'\big)^2 = n^{2\xi-2}\sum_{t,j=1}^n \mathbb{E}(s_{t,hl}s_{j,hl}).$$

From Assumption A.1(ii), we have $\max_{h,l} E|s_{t,hl}|^\delta < \infty$ with $\delta > 2$. Also, using Assumption 1, we can apply Lemma S.1, and claim (with $q = 2/\delta < 1$) that: for all t, j,

$$\begin{aligned}
\big|\mathbb{E}(s_{t,hl}s_{j,hl}) - \mathbb{E}(s_{t,hl})\mathbb{E}(s_{j,hl})\big| &\leq 10\rho^{(1-q)|t-j|}\big(\mathbb{E}|s_{t,hl}|^\delta\big)^{2/\delta} \leq 10\rho^{(1-q)|t-j|}\Big(\max_{h,l}\mathbb{E}|s_{t,hl}|^\delta\Big)^{2/\delta} \\
&:= c\rho^{(1-q)|t-j|}.
\end{aligned}$$

Note that for $|t-j| > n^\alpha$,

$$\big|\mathbb{E}(s_{t,hl}s_{j,hl})\big| = \big|\mathbb{E}(s_{t,hl}s_{j,hl}) - \mathbb{E}(s_{t,hl})\mathbb{E}(s_{j,hl})\big| \leq c\rho^{(1-q)n^\alpha}$$

and for $|t-j|\leq n^\alpha$,

$$\big|\mathbb{E}\big((s_{t,hl}s_{j,hl})\big)\big| \leq \big(\mathbb{E}\big(s_{t,hl}^2\big)\mathbb{E}\big(s_{j,hl}^2\big)\big)^{1/2} \leq c.$$

It follows that

$$\begin{aligned}
\sum_{t,j=1}^n \mathbb{E}(s_{t,hl}s_{j,hl}) &= \sum_{|t-j|\leq n^\alpha}\mathbb{E}(s_{t,hl}s_{j,hl}) + \sum_{|t-j|>n^\alpha}\mathbb{E}(s_{t,hl}s_{j,hl}) \\
&\leq c\cdot n\cdot n^\alpha + 2c\cdot n\cdot \sum_{j=n^\alpha+1}^n \rho^{(1-q)j}
\end{aligned}$$

$$\leq c \cdot n \cdot n^{\alpha} + 2c \cdot n\rho^{(1-q)(n^{\alpha}+1)} \frac{1-\rho^{(1-q)(n-n^{\alpha})}}{1-\rho^{1-q}}$$

$$\leq c \cdot n \cdot n^{\alpha} + \frac{4c}{1-\rho^{1-q}} \cdot n.$$

As a result, for any $n \geq 1$,

$$n^{2\xi-2} \sum_{t,j} \mathbb{E}(s_{t,hl}s_{j,hl}) \leq c \cdot n^{-\alpha} + c' \cdot n^{2\xi-1} = c \cdot n^{-\alpha} + c' \cdot n^{-2\alpha}$$

uniformly in h, l. Letting $\xi \to 1/2$, $\alpha \to 0$, and hence $n^{-1} \sum_{t,j} E(s_{t,hl}s_{j,hl}) \leq c + c'$. We can therefore claim that

$$\max_{h,l} \left| \frac{1}{n} \sum_{t=1}^{n} s_{t,hl} \right| = O_P\big(n^{-1/2}\big).$$

Besides,

$$\big|(b')\big| \leq 2\left(\frac{1}{n}\sum_{t=1}^{n}\big|g_h(x_t)\big|\big|g_l(x_t)\big| \sup_{\theta\in\mathcal{N}}\big|u_t(\theta)\big|\big\|\nabla_\theta u_t(\theta)\big\|\right)\|\bar{\theta}-\theta_0\|$$

$$\leq 2\left(\frac{1}{n}\sum_{t=1}^{n} g_l(x_t)^2 \sup_{\theta\in\mathcal{N}}\big|u_t(\theta)\big|^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n} g_h(x_t)^2 \sup_{\theta\in\mathcal{N}}\big\|\nabla_\theta u_t(\theta)\big\|^2\right)^{1/2}\|\bar{\theta}-\theta_0\|$$

$$= O_P\big(\|\bar{\theta}-\theta_0\|\big).$$

We deduce from (S.1) that

$$\max_{1\leq h,l\leq k}\big|[\hat{V}_k - V_k]_{hl}\big| = O_P\big(n^{-1/2}\big) + O_P\big(\|\bar{\theta}-\theta_0\|\big) = O_P\big(n^{-1/2} \vee \|\bar{\theta}-\theta_0\|\big)$$

and the result follows.

(ii) and (iii): Let c be the unit vector of $\mathbb{R}^k$ such that $c'V_kc = \lambda_k$ if $\lambda_k \leq \lambda_{\min}(\bar{V}_k)$. (Choose c such that $c'\bar{V}_kc = \lambda_{\min}(\bar{V}_k)$ otherwise.) We then have

$$\big|\lambda_k - \lambda_{\min}(\bar{V}_k)\big| \leq \big|c'(V_k - \bar{V}_k)c\big| \leq \sum_{h,l=1}^{k} |c_h||c_l|\big|[V_k - \bar{V}_k]_{hl}\big|$$

$$\leq \max_{1\leq h,l\leq k}\big|[V_k - \bar{V}_k]_{hl}\big|\left(\sum_{l=1}^{k}|c_l|\right)^2$$

$$\leq k \cdot \max_{1\leq h,l\leq k}\big|[V_k - \bar{V}_k]_{hl}\big|$$

and (ii) follows from the proof of part (i). Part (iii) is obtained by dividing each side of (ii) by λ_k.

We establish (iv) by recalling that $\bar{V}_k^{-1} - V_k^{-1} = -\bar{V}_k^{-1}(\bar{V}_k - V_k)V_k^{-1}$. Thus, by the Cauchy–Schwarz inequality, we have

$$\|\bar{V}_k^{-1} - V_k^{-1}\|_2 \leq \lambda_{\max}(\bar{V}_k^{-1})\|\hat{V}_k - V_k\|_2\lambda_{\max}(V_k^{-1}) = \frac{1}{\lambda_{\min}(\bar{V}_k)\lambda_k}\|\bar{V}_k - V_k\|_2$$
$$= \frac{\lambda_k}{\lambda_{\min}(\bar{V}_k)}\lambda_k^{-2}\|\bar{V}_k - V_k\|_2.$$

This yields the stated order by using (i) and (iii). The proof of (v) follows the same arguments as those in the proof of (iii). □

PROOF OF LEMMA B.3. (i) We have $\|\bar{D}_2'\hat{W}_k\bar{D}_2\|_2 = \lambda_{\max}(\bar{D}_2'\hat{W}_k\bar{D}_2) \leq [\lambda_{\max}(\hat{W}_k)/\bar{\lambda}_k]\bar{\lambda}_k \times \lambda_{\max}(\bar{D}_2'\bar{D}_2) = O_P(\bar{\lambda}_k k)$.

(ii) Same as (i) by noting that $\|\hat{W}_k^{1/2}\bar{f}_k\|_2 = \sqrt{\lambda_{\max}(\bar{f}_k'\hat{W}_k\bar{f}_k)} = \sqrt{\bar{f}_k'\hat{W}_k\bar{f}_k}$.

(iii) We have $\|\hat{W}_k^{1/2}\bar{H}\|_2 \leq \|\hat{W}_k^{1/2}(\bar{H} - H)\|_2 + \|\hat{W}_k^{1/2}H\|_2 \leq \|\hat{W}_k^{1/2}\|_2\|\bar{H} - H\|_2 + O_P(\sqrt{\bar{\lambda}_k k})$. Note that $\|\hat{W}_k^{1/2}\|_2\|\bar{H} - H\|_2 \leq \sqrt{\lambda_{\max}(\hat{W}_k)}O_P(\sqrt{k}[n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2]) = o_P(\sqrt{\bar{\lambda}_k k})$ and the result follows.

(iv) We have $\|(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\|_2 = [\lambda_{\min}(\bar{D}_1'\hat{W}_k\bar{D}_1)]^{-1}$. But $\lambda_{\min}(\bar{D}_1'\hat{W}_k\bar{D}_1) \geq \lambda_{\min}(\hat{W}_k) \times \lambda_{\min}(\bar{D}_1'\bar{D}_1)$. Note that, proceeding as in the proof of Lemma B.2(iii), we obtain

$$|\lambda_{\min}(\bar{D}_1'\bar{D}_1)/\lambda_{\min}(D_1'D_1) - 1| = O_P(\|\bar{D}_1'\bar{D}_1 - D_1'D_1\|_2/\lambda_{\min}(D_1'D_1)).$$

Using Lemma B.1, it is not difficult to see that $\|\bar{D}_1'\bar{D}_1 - D_1'D_1\|_2 = O_P(k[n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2]) = o_P(k)$. As a result, $\lambda_{\min}(\bar{D}_1'\bar{D}_1)/\lambda_{\min}(D_1'D_1) - 1 = o_P(1)$. We can therefore claim that

$$\|(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\|_2 \leq \frac{1}{\lambda_{\min}(\hat{W}_k)\lambda_{\min}(\bar{D}_1'\bar{D}_1)} = O_P([\lambda_{\min}(W_k)\lambda_{\min}(D_1'D_1)]^{-1}) = O_P(\bar{\lambda}_k^{-1}k^{-1}).$$

(v) We have

$$\begin{aligned}
&\bar{M}^{(k)} - M^{(k)} \\
&\quad = (\hat{W}_k^{1/2} - W_k^{1/2})\bar{D}_1(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\bar{D}_1'\hat{W}_k^{1/2} + W_k^{1/2}(\bar{D}_1 - D_1)'(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\bar{D}_1'\hat{W}_k^{1/2} \\
&\qquad + W_k^{1/2}D_1[(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1} - (D_1'W_kD_1)^{-1}]\bar{D}_1'\hat{W}_k^{1/2} \\
&\qquad + W_k^{1/2}D_1(D_1'W_kD_1)^{-1}(\bar{D}_1 - D_1)\hat{W}_k^{1/2} \\
&\qquad + W_k^{1/2}D_1(D_1'W_kD_1)^{-1}D_1(\hat{W}_k^{1/2} - W_k^{1/2}) := (1) + (2) + (3) + (4) + (5).
\end{aligned}$$

Note that

$$(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1} - (D_1'W_kD_1)^{-1} = -(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}(\bar{D}_1'\hat{W}_k\bar{D}_1 - D_1'W_kD_1)(D_1'W_kD_1)^{-1}.$$

From (iv), $\|(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\|_2 = O_P(\bar{\lambda}_k^{-1}k^{-1})$ and $\|D_1'W_kD_1\|_2 = O_P(\bar{\lambda}_k^{-1}k^{-1})$. Also,

$$\bar{D}_1'\hat{W}_k\bar{D}_1 - D_1'W_kD_1 = (\bar{D}_1 - D_1)'\hat{W}_k\bar{D}_1 + D_1'(\hat{W}_k - W_k)\bar{D}_1 + D_1'W_k(\bar{D}_1 - D_1).$$

Hence,

$$\big\|\bar{D}_1'\hat{W}_k\bar{D}_1 - D_1'W_kD_1\big\|_2 = O_P\big(\bar{\lambda}_k k\big[n^{-1/2} \vee \|\bar{\theta}-\theta_0\|_2\big]\big) + O_P\big(k\|\hat{W}_k - W_k\|_2\big).$$

Thus,

$$\begin{aligned}&\big\|\big(\bar{D}_1'\hat{W}_k\bar{D}_1\big)^{-1} - \big(D_1'W_kD_1\big)^{-1}\big\|_2\\&\quad= O_P\big(\bar{\lambda}_k^{-1}k^{-1}\big[n^{-1/2} \vee \|\bar{\theta}-\theta_0\|_2\big]\big) + O_P\big(\bar{\lambda}_k^{-2}k^{-1}\|\hat{W}_k - W_k\|_2\big).\end{aligned}$$

Also, using the Ando–van Hemmen inequality (see Ando and Leo van Hemmen (1980)), we have

$$\big\|\hat{W}_k^{1/2} - W_k^{1/2}\big\|_2 \le \frac{1}{\lambda_{\min}(\hat{W}_k)^{1/2} + \lambda_{\min}(W_k)^{1/2}}\|\hat{W}_k - W_k\|_2 \le \lambda_{\min}(W_k)^{-1/2}\|\hat{W}_k - W_k\|_2.$$

Then, going back to the expression for $\bar{M}^{(k)} - M^{(k)}$, we have

$$\begin{aligned}\big\|(1)\big\|_2 &\le \lambda_{\min}(W_k)^{-1/2}\|\hat{W}_k - W_k\|_2 O_P(\sqrt{k})O_P\big(\bar{\lambda}_k^{-1}k^{-1}\big)O_P(\sqrt{k})O_P(\sqrt{\bar{\lambda}_k})\\&= O_P\big(\bar{\lambda}_k^{-1}\|\hat{W}_k - W_k\|_2\big).\end{aligned}$$

Similarly, we can verify that $\|(5)\|_2 = O_P(\bar{\lambda}_k^{-1}\|\hat{W}_k - W_k\|_2)$ and

$$\begin{aligned}\big\|(2)\big\|_2 &\le \bar{\lambda}_k^{1/2}O_P\big(\sqrt{k}\big[n^{-1/2} \vee \|\bar{\theta}-\theta_0\|_2\big]\big)O_P\big(\bar{\lambda}_k^{-1}k^{-1}\big)O_P(\sqrt{k})O_P\big(\bar{\lambda}_k^{1/2}\big)\\&= O_P\big(n^{-1/2} \vee \|\bar{\theta}-\theta_0\|_2\big).\end{aligned}$$

The same holds for $\|(4)\|_2$. Finally,

$$\begin{aligned}\big\|(3)\big\|_2 &\le O_P(\bar{\lambda}_k)O_P(k)\big[O_P\big(\bar{\lambda}_k^{-1}k^{-1}\big[n^{-1/2} \vee \|\bar{\theta}-\theta_0\|_2\big]\big) + O_P\big(\bar{\lambda}_k^{-2}k^{-1}\|\hat{W}_k - W_k\|_2\big)\big]\\&= O_P\big(\big[n^{-1/2} \vee \|\bar{\theta}-\theta_0\|_2\big]\big) + O_P\big(\bar{\lambda}_k^{-1}\|\hat{W}_k - W_k\|_2\big)\end{aligned}$$

and the result follows.

(vi) Note that

$$\begin{aligned}\|\Delta_{1n}\|_2 &= \big\|\bar{H}'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{H} - H'W_k^{1/2}M^{(k)}W_k^{1/2}H\big\|_2\\&\le \|\bar{H}-H\|_2\|\hat{W}_k\|_2\|\bar{H}\|_2 + \|H\|_2\big\|\hat{W}_k^{1/2} - W_k^{1/2}\big\|_2\big\|\hat{W}_k^{1/2}\big\|_2\|\bar{H}\|_2\\&\quad+ \|H\|_2\big\|W_k^{1/2}\big\|_2\big\|\bar{M}^{(k)} - M^{(k)}\big\|_2\|\bar{H}\|_2\big\|\hat{W}_k^{1/2}\big\|_2\\&\quad+ \|H\|_2\big\|W_k^{1/2}\big\|_2\big\|\hat{W}_k^{1/2} - W_k^{1/2}\big\|_2\|\bar{H}\|_2 + \|H\|_2\|W_k\|_2\|\bar{H}-H\|_2.\end{aligned}$$

The result follows by the same steps as above.

(vii) Similarly,

$$\begin{aligned}\|\Delta_{2n}\|_2 &= \big\|\bar{H}'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{f}_k - H'W_k^{1/2}M^{(k)}W_k^{1/2}\bar{f}_k\big\|_2\\&\le \|\bar{H}-H\|_2\|\hat{W}_k\|_2\|\bar{f}_k\|_2 + \|H\|_2\big\|\hat{W}_k^{1/2} - W_k^{1/2}\big\|_2\big\|\hat{W}_k^{1/2}\big\|_2\|\bar{f}_k\|_2\end{aligned}$$

$$+ \|H\|_2 \big\| W_k^{1/2} \big\|_2 \big\| \bar{M}^{(k)} - M^{(k)} \big\|_2 \big\| \hat{W}_k^{1/2} \big\|_2 \|\bar{f}_k\|_2$$
$$+ \|H\|_2 \big\| W_k^{1/2} \big\|_2 \big\| \hat{W}_k^{1/2} - W_k^{1/2} \big\|_2 \|\bar{f}_k\|_2$$

and the result follows along similar lines as above.

(viii) We have

$$\begin{aligned}\big\| \hat{M}^{(k)} - \bar{M}^{(k)} \big\|_2 \le \|\hat{W}_k\|_2 \big(\|\hat{D}_1 - \bar{D}_1\|_2 \big\| \big(\hat{D}_1' \hat{W}_k \hat{D}_1\big)^{-1} \big\|_2 \|\hat{D}_1\|_2 \\ + \|\hat{D}_1\|_2 \|\bar{D}_1\|_2 \big\| \big(\hat{D}_1' \hat{W}_k \hat{D}_1\big)^{-1} - \big(\bar{D}_1' \hat{W}_k \bar{D}_1\big)^{-1} \big\|_2 \\ + \|\bar{D}_1\|_2 \big\| \big(\bar{D}_1' \hat{W}_k \bar{D}_1\big)^{-1} \big\|_2 \|\hat{D} - \bar{D}_1\|_2 \big) := (1) + (2) + (3).\end{aligned}$$

From (iv), $\|(\bar{D}_1' \hat{W}_k \bar{D})_1^{-1}\|_2 = O_P(\bar{\lambda}_k^{-1} k^{-1})$ and $\|(\hat{D}_1' \hat{W}_k \hat{D}_1)^{-1}\|_2 = O_P(\bar{\lambda}_k^{-1} k^{-1})$. Note that

$$\begin{aligned}\|\hat{D}_1 - \bar{D}_1\|_2 &\le \|\hat{D}_1 - D_1\|_2 + \|\bar{D}_1 - D_1\|_2 \\ &\le O_P\big(\sqrt{k}\big[n^{-1/2} \vee \|\hat{\theta} - \theta_0\|_2\big]\big) + O_P\big(\sqrt{k}\big[n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2\big]\big).\end{aligned}$$

Since $\bar{\theta} \in (\theta_0, \hat{\theta})$, we can write $\bar{\theta} = t\theta_0 + (1-t)\hat{\theta}$ for $t \in (0,1)$. Therefore, we can claim that $\|\hat{D}_1 - \bar{D}_1\|_2 = O_P(\sqrt{k}[n^{-1/2} \vee \|\hat{\theta} - \theta_0\|_2])$. Using similar arguments as above, we can also show that

$$\big\| \big(\hat{D}_1' \hat{W}_k \hat{D}_1\big)^{-1} - \big(\bar{D}_1' \hat{W}_k \bar{D}_1\big)^{-1} \big\|_2 = O_P\big(\bar{\lambda}_k^{-1} k^{-1} \big[n^{-1/2} \vee \|\hat{\theta} - \theta_0\|_2\big]\big).$$

By combining this with the fact that $\|\hat{W}_k\|_2 = O_P(\bar{\lambda}_k)$, $\|\bar{D}_1\|_2 = O_P(\sqrt{k})$ and $\|\hat{D}_1\|_2 = O_P(\sqrt{k})$, the result follows readily. □

Proof of Lemma B.4. Note that conditions of the theorem ensure that the maps $\theta \mapsto \mathbb{E}(g_l(x)u(\theta))$ are continuous on Θ for each l. By Lemma 1 of de Jong and Bierens (1994), we can claim that, for each $\theta_i \in \Theta$, there exists l_i such that $|\mathbb{E}(g_{l_i}(x)u(\theta_i))| = \delta_i > 0$. By continuity of the map $\theta \mapsto \mathbb{E}(g_{l_i}(x)u(\theta))$, there exists an open neighborhood V_i of θ_i such that $|\mathbb{E}(g_{l_i}(x)u(\theta))| > \delta_i/2$ for all $\theta \in V_i$. Clearly, $\Theta \subset \bigcup_{\theta_i \in \Theta} V_i$ and, by compactness of Θ, we can extract a finite number of elements from the sequence of V_i to cover Θ. That is, $\Theta \subset \bigcup_{j=1}^{m_0} V_{i_j}$, with m_0 finite. Let $k_0 = \max_{1 \le j \le m_0} l_{i_j}$. k_0 is a finite integer and, by construction, for any $\theta \in \Theta$, $\mathbb{E}(g^{(k_0)}(x)u(\theta)) \neq 0$. Also, taking $\delta_0 = \min_{1 \le j \le m_0}(\delta_{i_j}/2)^2$, we obviously have $\|\mathbb{E}(g^{(k_0)}(x)u(\theta))\|_2^2 \ge \delta_0 > 0$ and this concludes the proof. □

Proof of equation (B.6) in Appendix B

We have

$$\begin{aligned}&\bar{f}_k(\hat{\theta})' \hat{W}_k^{1/2} \bar{P}^{(k)} \hat{W}_k^{1/2} \bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)' \hat{W}_k^{1/2} \bar{P}^{(k)} \hat{W}_k^{1/2} \bar{f}_k(\theta_0) \\ &\quad = 2\bar{f}_k(\hat{\theta})' \hat{W}_k^{1/2} \bar{P}^{(k)} \hat{W}_k^{1/2} \bar{f}_k(\hat{\theta}) \\ &\qquad - 2\bar{f}_k(\theta_0)' \hat{W}_k^{1/2} \bar{P}^{(k)} \hat{W}_k^{1/2} \bar{f}_k(\hat{\theta}) - \big(\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)\big)' \hat{W}_k^{1/2} \bar{P}^{(k)} \hat{W}_k^{1/2} \big(\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)\big).\end{aligned}$$

From (B.3) and (B.4), $\|\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)\|_2 = O_P(\sqrt{k}\|\hat{\eta} - \eta_0\|_2) = O_P(\sqrt{k}\|\hat{\theta} - \theta_0\|_2)$. Hence,

$$\begin{aligned}\big|\big(\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)\big)'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\big(\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)\big)\big| &\le \lambda_{\max}(\hat{W}_k)\big\|\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)\big\|_2^2 \\ &= O_P\big(\lambda_k k\|\hat{\theta} - \theta_0\|_2^2)\big).\end{aligned}$$

Let $\hat{D}_1 := \bar{D}_1(\hat{\theta})$. By the first-order necessary optimality condition, $\hat{D}_1'\hat{W}_k\bar{f}_k(\hat{\theta}) = 0$.

Let $\hat{P}^{(k)} := \hat{W}_k^{1/2}\hat{D}_1(\hat{D}_1'\hat{W}_k\hat{D}_1)^{-1}\hat{D}_1'\hat{W}_k^{1/2}$. Obviously, $\hat{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) = 0$. Thus,

$$\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) = \big(\bar{P}^{(k)} - \hat{P}^{(k)}\big)\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) \quad \text{and}$$

$$\big\|\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta})\big\|_2 \le \big\|\bar{P}^{(k)} - \hat{P}^{(k)}\big\|_2 \cdot O_P(\sqrt{\bar{\lambda}_k k}).$$

From Lemma B.3, $\|\bar{P}^{(k)} - \hat{P}^{(k)}\|_2 = O_P(n^{-1/2} \vee \|\hat{\theta} - \theta_0\|_2)$ and it follows that

$$\big\|\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta})\big\|_2 = O_P\big(\sqrt{k\bar{\lambda}_k}\big[n^{-1/2} \vee \|\hat{\theta} - \theta_0\|_2\big]\big).$$

Thus,

$$\bar{f}_k(\hat{\theta})'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) = O_P\big(\bar{\lambda}_k k\big[n^{-1} \vee \|\hat{\theta} - \theta_0\|_2^2\big]\big)$$

and

$$\bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) = O_P\big(\bar{\lambda}_k k\big[n^{-1/2} \vee \|\hat{\theta} - \theta_0\|_2\big]\big).$$

As a result,

$$\begin{aligned}&\big|\bar{f}_k(\hat{\theta})'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\theta_0)\big| \\ &\quad = O_P\big(\bar{\lambda}_k k n^{-1/2}\big) + O_P\big(\bar{\lambda}_k k\|\hat{\theta} - \theta_0\|_2\big).\end{aligned}$$

Proof of Theorem 5.1 with some preliminary results

The asymptotic theory for degenerate U-statistics has been extensively studied in the literature. Existing results cover cases where x_t is assumed to be i.i.d. or time dependent as well as cases where the kernel function is sample-size dependent—as above—or fixed. A CLT for i.i.d. data and sample-size dependent kernel has been developed by Hall (1984) and played a key role in the main results of de Jong and Bierens (1994). An extension of Hall's (1984) results to β-mixing processes has been provided by Fan and Li (1999). More recent contributions to this literature include Leucht (2012), who proposes a CLT for τ-dependent processes[1] and fixed kernels,[2] and Gao (2007) and Gao and Hong (2008) who consider α-mixing processes and sample-size dependent kernels. Kim, Luo, and Kim (2011) further extend these results by establishing the CLT for U-statistics under quite general conditions. Our kernel is more consistent with the formulation in Kim,

[1]See Dedecker and Prieur (2005) for a definition. Note that i.i.d. $\Rightarrow$ β-mixing $\Rightarrow$ α-mixing $\Rightarrow$ τ-dependence.

[2]CLT for degenerate U-statistics with fixed kernels gives rise to nonstandard asymptotic distribution taking the form of a quadratic function of an infinite number of independent Gaussian variables, while sample-size dependent kernels are typically associated with the standard normal distribution.

Luo, and Kim (2011) but the special factorization that they require is not well aligned with our framework, which features an inner product with an increasing dimension.

The following proposition is used in the proof of Theorem 5.1 below.

PROPOSITION S.2. *Under Assumptions*-CLT 1 *and* 2, $\mathrm{Var}(U_n) = 2 + o(1)$.

PROOF OF PROPOSITION S.2. We shall assume, without loss of generality, that $V_k = I_k$. This amounts to taking the scaled version of $f_k(\theta_0)$ by $V_k^{-1/2}$. In the following expressions, we let $f_i = f_k(x_i)$ and $f_j = f_k(x_j)$. We have

$$\begin{aligned}
\mathrm{Var}(U_n) &= \frac{1}{n^2 k}\mathbb{E}\left(\sum_{i=1}^{n}\sum_{j\neq i} f_i' f_j\right)^2 = \frac{1}{n^2 k}\sum_{i_1\neq j_1}\sum_{i_2\neq j_2}\mathbb{E}(f_{i_1}' f_{j_1}\cdot f_{i_2}' f_{j_2}) \\
&= \frac{2}{n^2 k}\sum_{i\neq j}\mathbb{E}\big[(f_i' f_j)^2\big] + \frac{1}{n^2 k}\sum_{\substack{i_1\neq j_1, i_2\neq j_2 \\ \text{3 diff. indices}}}\mathbb{E}\big[(f_{i_1}' f_{j_1})\cdot(f_{i_2}' f_{j_2})\big] \\
&\quad + \frac{1}{n^2 k}\sum_{\substack{i_1\neq j_1, i_2\neq j_2 \\ \text{4 diff. indices}}}\mathbb{E}\big[(f_{i_1}' f_{j_1})\cdot(f_{i_2}' f_{j_2})\big].
\end{aligned} \tag{S.2}$$

Because of the martingale difference dynamics of f_i, all the terms in the last summation are 0. We next show that the second expression is $o(1)$ and the first one is equal to $2 + o(1)$.

(a) Consider

$$A_n = \frac{1}{n^2 k}\sum_{\substack{i\neq j_1, i\neq j_2 \\ j_1\neq j_2}}\mathbb{E}\big[f_i' f_{j_1}\cdot f_i' f_{j_2}\big].$$

Let $\pi_n = n^a$, $a \in (0, 1)$. In the following, we will consider two indices i, j to be connected iff $|i - j| \leq \pi_n$ and a set of three indices to be connected iff any one of them is connected to at least another one of them.

In the summation above, we have three configurations that stand out:

(i) The three indices i, j_1, j_2 are connected. Denote S_1 to be the collection of such indices.

(ii) Only two of the indices, say $\{i, j_1\}$, $\{i, j_2\}$, or $\{j_1, j_2\}$ are connected. Denote S_2, S_3, and S_4 the subset of indices (i, j_1, j_2) satisfying these descriptions, respectively.

(iii) All the three indices are isolated from one another. Denote S_5 to be the collection of such indices.

We first deal with (i). We have

$$\big|\mathbb{E}\big[f_i' f_{j_1}\cdot f_i' f_{j_2}\big]\big| = \left|\sum_{h,h'=1}^{k}\mathbb{E}[f_{ih} f_{j_1 h} f_{ih'} f_{j_2 h'}]\right| \leq \sum_{h,h'=1}^{k}\mathbb{E}|f_{ih} f_{j_1 h} f_{ih'} f_{j_2 h'}|$$

$$\leq \sum_{h,h'=1}^{k} \left[\mathbb{E}(f_{ih}^4)\mathbb{E}(f_{j_1h}^4)\mathbb{E}(f_{ih'}^4)\mathbb{E}(f_{j_2h'}^4)\right]^{1/4}$$
$$\leq \sum_{h,h'=1}^{k} \left(\mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{2/4+\epsilon}\left(\mathbb{E}|f_{ih'}|^{4+\epsilon}\right)^{2/4+\epsilon},$$

where the first inequality is the triangle inequality, the second one follows from the Cauchy–Schwarz inequality, the third one follows from the monotonicity (in p) of L_p-norms and the stationarity assumption. Hence,

$$\left|\mathbb{E}\left[f_i' f_{j_1} \cdot f_i' f_{j_2}\right]\right| \leq k^2\left(\frac{1}{k}\sum_{h=1}^{k}\left(\mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{2/4+\epsilon}\right)^2 \leq k^2\left(\frac{1}{k}\sum_{h=1}^{k}\left(\mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{4/4+\epsilon}\right)$$
$$\leq k^2\left(\frac{1}{k}\sum_{h=1}^{k}\mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{4/4+\epsilon} \leq k^2\Delta_\epsilon,$$

where $\Delta_\epsilon = \Delta^{4/4+\epsilon}$ is an absolute constant. The last inequality holds by assumption and the previous two follow from the Jensen's inequality. Thus,

$$A_{1n} \equiv \left|\frac{1}{n^2 k}\sum_{(i,j_1,j_2)\in S_1}\mathbb{E}\left[f_i' f_{j_1}\cdot f_i' f_{j_2}\right]\right| \leq N(S_1)\frac{k^2\Delta_\epsilon}{n^2 k},$$

where $N(S)$ denotes the cardinality of S. It is not hard to see that $N(S_1) \leq n\pi_n^2$. Then

$$A_{1n} \leq \Delta_\epsilon k\pi_n^2/n.$$

Choosing $\pi_n = o(\sqrt{n/k})$ is sufficient to claim that $A_{1n} = o(1)$.

We next deal with (ii). Assume that (i, j_1) are connected and j_2 is isolated from both so that $(i, j_1, j_2) \in S_2$. Take $p_1 = \frac{4+\epsilon}{3}$ and $p_2 = 4+\epsilon$. Let $q = \frac{1}{p_1} + \frac{1}{p_2} = \frac{4}{4+\epsilon}$. We have $q < 1$ and by Lemma S.1, it follows that

$$\mathbb{E}(f_{ih}f_{ih'}f_{j_1h}f_{j_2h'}) = \mathbb{E}(f_{ih}f_{ih'}f_{j_1h})\mathbb{E}(f_{j_2h'}) + O\left(\rho^{(1-q)\pi_n}\|f_{ih}f_{ih'}f_{j_1h}\|_{p_1}\|f_{j_2h'}\|_{p_2}\right)$$
$$= O\left(\rho^{(1-q)\pi_n}\|f_{ih}f_{ih'}f_{j_1h}\|_{p_1}\|f_{j_2h'}\|_{p_2}\right).$$

By the Cauchy–Schwarz inequality, we have

$$\|f_{ih}f_{ih'}f_{j_1h}\|_{p_1}\|f_{j_2h'}\|_{p_2} \leq \left(\mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{\frac{1}{4+\epsilon}}\left(\mathbb{E}|f_{ih'}|^{4+\epsilon}\right)^{\frac{1}{4+\epsilon}}\left(\mathbb{E}|f_{j_1h}|^{4+\epsilon}\right)^{\frac{1}{4+\epsilon}}$$
$$= \left(\mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{\frac{2}{4+\epsilon}}\left(\mathbb{E}|f_{ih'}|^{4+\epsilon}\right)^{\frac{1}{4+\epsilon}},$$

where the last equality holds by stationarity. Hence,

$$\mathbb{E}(f_{ih}f_{ih'}f_{j_1h}f_{j_2h'}) = O\left(\rho^{(1-q)\pi_n}\|f_{ih}\|_{4+\epsilon}^2\|f_{ih'}\|_{4+\epsilon}^2\right).$$

This yields

$$\sum_{h,h'=1}^{k} \mathbb{E}(f_{ih} f_{ih'} f_{j_1 h} f_{j_2 h'}) = O\left(\rho^{(1-q)\pi_n} k^2 \left(\frac{1}{k}\sum_{h=1}^{k} \|f_{ih}\|_{4+\epsilon}^2\right)^2\right).$$

But, by the Jensen's inequality,

$$\begin{aligned}\left(\frac{1}{k}\sum_{h=1}^{k} \|f_{ih}\|_{4+\epsilon}^2\right)^2 &\leq \frac{1}{k}\sum_{h=1}^{k} \|f_{ih}\|_{4+\epsilon}^4 = \frac{1}{k}\sum_{h=1}^{k} \left(\mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{\frac{4}{4+\epsilon}} \\ &\leq \left(\frac{1}{k}\sum_{h=1}^{k} \mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{\frac{4}{4+\epsilon}} \leq \left(\sup_k \frac{1}{k}\sum_{h=1}^{k} \mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{\frac{4}{4+\epsilon}} = \Delta_\epsilon < \infty.\end{aligned}$$

Thus,

$$\sum_{h,h'=1}^{k} \mathbb{E}(f_{ih} f_{ih'} f_{j_1 h} f_{j_2 h'}) = O\left(k^2 \rho^{(1-q)\pi_n}\right).$$

Hence,

$$A_{2n} \equiv \left|\frac{1}{n^2 k} \sum_{(i,j_1,j_2)\in S_2} \mathbb{E}\left[f_i' f_{j_1} \cdot f_i' f_{j_2}\right]\right| = O\left(\frac{N(S_2)}{n^2 k} k^2\right).$$

Clearly, $N(S_2) \leq n^2 \pi_n$. Thus, choosing $\pi_n = n^a$ for some $a > 0$ ensures that $A_{2n} = o(1)$. Likewise, summation over $(i, j_1, j_2) \in S_3$ leads to a negligible quantity.

Now, consider S_4, where j_1 and j_2 are connected while i is isolated. Applying Lemma S.1, we have

$$\begin{aligned}\mathbb{E}(f_{ih} f_{ih'} f_{j_1 h} f_{j_2 h'}) &= \mathbb{E}(f_{ih} f_{ih'})\mathbb{E}(f_{j_1 h} f_{j_2 h'}) + O\left(\rho^{(1-q)\pi_n} \|f_{ih} f_{ih'}\|_{p_1} \|f_{j_1 h} f_{j_2 h'}\|_{p_2}\right) \\ &= O\left(\rho^{(1-q)\pi_n} \|f_{ih} f_{ih'} f_{j_1 h}\|_{p_1} \|f_{j_2 h'}\|_{p_2}\right),\end{aligned}$$

with $p_1 = p_2 = \frac{4+\epsilon}{2}$, and then $q = \frac{4}{4+\epsilon}$. From the Cauchy–Schwarz inequality and the stationarity assumption, we have

$$\|f_{ih} f_{ih'} f_{j_1 h}\|_{p_1} \|f_{j_2 h'}\|_{p_2} \leq \left(|f_{ih}|^{4+\epsilon}\right)^{\frac{2}{4+\epsilon}} \left(|f_{ih'}|^{4+\epsilon}\right)^{\frac{2}{4+\epsilon}}.$$

Similar derivations as above lead to

$$A_{3n} \equiv \left|\frac{1}{n^2 k} \sum_{(i,j_1,j_2)\in S_4} \mathbb{E}\left[f_i' f_{j_1} \cdot f_i' f_{j_2}\right]\right| = O\left(\frac{N(S_4)}{n^2 k} k^2\right).$$

Again, $N(S_4) \leq n^2 \pi_n$. Thus, choosing $\pi_n = n^a$ for some $a > 0$ ensures that $A_{3n} = o(1)$.

We next consider (iii), where $\{i, j_1, j_2\}$ has no pairs connected. Again, by Lemma S.1, we have

$$\begin{aligned}\mathbb{E}(f_{ih} f_{ih'} f_{j_1 h} f_{j_2 h'}) &= \mathbb{E}(f_{ih} f_{ih'})\mathbb{E}(f_{j_1 h})\mathbb{E}(f_{j_2 h'}) + O\left(\rho^{(1-q)\pi_n} \|f_{ih} f_{ih'}\|_{p_1} \|f_{j_1 h}\|_{p_2} \|f_{j_2 h'}\|_{p_3}\right) \\ &= O\left(\rho^{(1-q)\pi_n} \|f_{ih} f_{ih'}\|_{p_1} \|f_{j_1 h}\|_{p_2} \|f_{j_2 h'}\|_{p_3}\right),\end{aligned}$$

with $p_1 = \frac{4+\epsilon}{2}$, and $p_2 = p_3 = 4+\epsilon$. By the Cauchy–Schwarz inequality and stationarity, we have, as before,

$$\|f_{ih}f_{ih'}\|_{p_1}\|f_{j_1h}\|_{p_2}\|f_{j_2h'}\|_{p_3} \le (\mathbb{E}\big(|f_{ih}|^{4+\epsilon}\big)^{\frac{2}{4+\epsilon}}(\mathbb{E}\big(|f_{ih'}|^{4+\epsilon}\big)^{\frac{2}{4+\epsilon}}.$$

Hence,

$$\sum_{h,h'=1}^{k} \mathbb{E}(f_{ih}f_{ih'}f_{j_1h}f_{j_1h'}) = O\big(\rho^{(1-q)\pi_n}k^2\big).$$

Thus,

$$A_{4n} \equiv \left|\frac{1}{n^2k}\sum_{(i,j_1,j_2)\in S_5}\mathbb{E}\big[f_i'f_{j_1}\cdot f_i'f_{j_2}\big]\right| = O\left(\frac{1}{n^2k}N(S_5)\rho^{(1-q)\pi_n}k^2\right).$$

Note that $N(S_5)\le n^3$. Thus, $A_{4n} = O(n^2\rho^{(1-q)\pi_n})$. Taking $\pi_n = n^a$ for some $a>0$ ensures that $A_{4n} = o(1)$.

Therefore, since $k = n^\alpha$, with $\alpha \in (0,1)$, we can always find $\pi_n = n^a$ with $a>0$ small enough so that all A_{1n}, A_{2n}, A_{3n}, and A_{4n} are all $o(1)$. This shows that $A_n = o(1)$. Hence, the second term in the expression of the variance of U_n in (S.2) is $o(1)$, that is,

$$\frac{1}{n^2k}\sum_{\substack{i_1\ne j_1, i_2\ne j_2\\ \text{3 diff. indices}}}\mathbb{E}\big[\big(f_{i_1}'f_{j_1}\big)\cdot\big(f_{i_2}'f_{j_2}\big)\big] = o(1).$$

(b) Consider the first term in the expression of the variance of U_n in (S.2). We have

$$\frac{1}{n^2k}\sum_{i\ne j}\mathbb{E}\big[\big(f_i'f_j\big)^2\big] = \frac{1}{n^2k}\sum_{i\ne j}\sum_{h,h'=1}^{k}\mathbb{E}(f_{ih}f_{jh}f_{ih'}f_{jh'}).$$

As previously, we can show that

$$\big|\mathbb{E}(f_{ih}f_{jh}f_{ih'}f_{jh'})\big| \le \big(\mathbb{E}|f_{ih}|^{4+\epsilon}\big)^{\frac{2}{4+\epsilon}}\big(\mathbb{E}|f_{jh}|^{4+\epsilon}\big)^{\frac{2}{4+\epsilon}}$$

and

$$\sum_{h,h'=1}^{k}\mathbb{E}(f_{ih}f_{jh}f_{ih'}f_{jh'}) \le k^2\cdot\Delta_\epsilon.$$

If i, j are connected,

$$A_{5n} \equiv \left|\frac{1}{n^2k}\sum_{\substack{i\ne j\\ |i-j|\le\pi_n}}\sum_{h,h'=1}^{k}\mathbb{E}(f_{ih}f_{jh}f_{ih'}f_{jh'})\right| \le \frac{Nk^2\Delta_\epsilon}{n^2k} = O\left(\frac{k\pi_n}{n}\right),$$

where $N\le 2n\pi_n$ is the number of connected pairs (i,j). Choosing $\pi_n = o(n/k)$ is enough to claim that $A_{5n} = o(1)$.

If i, j are not connected, $|i-j|\ge\pi_n$, we have

$$\mathbb{E}(f_{ih}f_{jh}f_{ih'}f_{jh'}) = \mathbb{E}(f_{ih}f_{ih'})\mathbb{E}(f_{jh}f_{jh'}) + O\big(\rho^{(1-q)\pi_n}\|f_{ih}f_{ih'}\|_{p_1}\|f_{jh}f_{jh'}\|_{p_2}\big),$$

with $p_1 = p_2 = \frac{4+\epsilon}{2}$ and $q = \frac{4}{4+\epsilon}$. As in the previous lines, we can claim that

$$\frac{1}{n^2k}\sum_{\substack{i\neq j\\ |i-j|\geq \pi_n}}\sum_{h,h'=1}^{k}\mathbb{E}(f_{ih}f_{jh}f_{ih'}f_{jh'}) = \frac{1}{n^2k}\sum_{\substack{i\neq j\\ |i-j|\geq \pi_n}}\mathbb{E}^*\big((f_i'f_j)^2\big) + O\big(k\rho^{(1-q)\pi_n}\big),$$

where E^* denote expectation under independence of f_i and f_j. By taking $\pi_n = n^a$ for some $a > 0$ small enough, the second term in the right-hand side is $o(1)$. Note that the number N of connected pairs (i, j) satisfies

$$n(n-2\pi_n) \leq N \leq n^2.$$

Also,

$$\mathbb{E}^*\big(f_i'f_j\big)^2 = \mathbb{E}^*\big(f_i'f_jf_j'f_i\big) = \mathrm{trace}\big(\mathbb{E}^*[f_jf_j'f_if_i']\big) = k.$$

It follows that $\frac{1}{n^2k}\sum_{i\neq j,|i-j|\geq\pi_n}E^*((f_i'f_j)^2) = 1 + o(1)$ and as a result,

$$\frac{1}{n^2k}\sum_{i\neq j}\mathbb{E}\big[(f_i'f_j)^2\big] = 1 + o(1).$$

This completes the proof. □

Proof of Theorem 5.1. As previously, we assume, without loss of generality, that $V_k = I_k$. Again, let $f_i := f_k(x_i)$ and set $z_i := \max_{1\leq i\leq n}\|f_i\|/\sqrt{k}$ and $z_{ij} := \max_{1\leq i\neq j\leq n}|f_i'f_j|/\sqrt{k}$. Let $M > 0$. Then we have

$$U_n = U_n1_{\{(z_i\leq M\log^\beta n)\cap(z_{ij}\leq M\log^\beta n)\}} + U_n1_{\{(z_i> M\log^\beta n)\cup(z_{ij}> M\log^\beta n)\}}. \tag{S.3}$$

For the second term in this expression, we have

$$\mathbb{E}\big[|U_n|1_{\{z_i>M\log^\beta n\cup z_{ij}>M\log^\beta n\}}\big] \leq \mathbb{E}\big[|U_n|1_{\{z_i>M\log^\beta n\}}\big] + \mathbb{E}\big[|U_n|1_{\{z_{ij}>M\log^\beta n\}}\big].$$

By the Cauchy–Schwarz and the Markov inequalities,

$$\mathbb{E}\big[|U_n|1_{\{z_i>M\log^\beta n\}}\big] \leq \big(\mathbb{E}\big(U_n^2\big)\Pr\big(z_i > M\log^\beta n\big)\big)^{1/2} \leq \frac{1}{\sqrt{M\log^\beta n}}\big(\mathbb{E}\big(U_n^2\big)\mathbb{E}(z_i)\big)^{1/2}.$$

Thus, by Proposition S.2 and Assumption-CLT 3, there exists a constant $c > 0$ such that

$$\sup_n \mathbb{E}\big[|U_n|1_{\{z_i>M\log^\beta n\}}\big] \leq \frac{c}{\sqrt{M}}.$$

The same claim can be made about $\sup_n E[|U_n|1_{\{z_{ij}>M\log^\beta n\}}]$ and it then follows that

$$\lim_{M\to\infty}\sup_n \mathbb{E}\big[|U_n|1_{\{(z_i>M\log^\beta n)\cup(z_{ij}>M\log^\beta n)\}}\big] = 0.$$

Hence, $U_n 1_{\{(z_i > M\log^\beta n)\cup(z_{ij} > M\log^\beta n)\}}$ can be made stochastically small by taking M large and we can only focus on the first term in (S.3). We shall therefore consider throughout that

$$\max_{1\le i\le n}\frac{\|f_i\|}{\sqrt{k}\log^\beta n}\le M \quad\text{and}\quad \max_{1\le i\ne j\le n}\frac{|f_i' f_j|}{\sqrt{k}\log^\beta n}\le M \tag{S.4}$$

for a fixed constant $M > 0$.

Next, we show that the moments of $U_n/\sqrt{2}$ converge to those of the standard normal distribution. Let $r \in N$. We have

$$\begin{aligned} U_n^{2r+1} &= \frac{1}{(n\sqrt{k})^{2r+1}}\sum_{\substack{i_1,j_1,\dots,i_{2r+1},j_{2r+1}=1\\ i_a\ne j_a,\,a=1,\dots,2r+1}}^{n}\prod_{s=1}^{2r+1} f_{i_s}' f_{j_s}\\ &= \frac{1}{(n\sqrt{k})^{2r+1}}\sum_{\substack{i_1,j_1,\dots,i_{2r+1},j_{2r+1}=1\\ i_a\ne j_a,\,a=1,\dots,2r+1}}^{n}\ \sum_{h_1,\dots,h_{2r+1}=1}^{k}\prod_{s=1}^{2r+1} f_{i_s,h_s} f_{j_s,h_s}. \end{aligned} \tag{S.5}$$

To derive this moment, we shall rely, following Kim, Luo, and Kim (2011), on some notions from graph theory. To make the discussion self-contained, we introduce some of these notions below. More details may be found in Kim, Luo, and Kim (2011). Each term in the right-hand side of (S.5) can be associated with an undirected graph with vertices $i_1, j_1, i_2, j_2, \dots, i_{2r+1}, j_{2r+1}$. Let $\pi_n(\le n)$ be an increasing sequence of n. We say that a_1, a_2 in the graph $\{i_1, j_1, i_2, j_2, \dots, i_{2r+1}, j_{2r+1}\}$ are connected if $|a_1 - a_2| \le \pi_n$ or there exist $\{b_1, \dots, b_l\} \subset \{i_1, j_1, i_2, j_2, \dots, i_{2r+1}, j_{2r+1}\}$, such that

$$|a_1 - b_1| \le \pi_n,\ |b_1 - b_2| \le \pi_n,\ \dots,\ |b_{l-1} - b_l| \le \pi_n,\ |b_l - a_2| \le \pi_n.$$

Note that in a graph, many i_s and/or j_s may take the same value. A component of the graph is a subset $\mathcal{I}$ of $\{i_1, j_1, i_2, j_2, \dots, i_{2r+1}, j_{2r+1}\}$ such that every vertex in $\mathcal{I}$ is connected to at least another one. A graph can be partitioned into m components: $\mathcal{I}_1, \dots, \mathcal{I}_m$. We may assume, without loss of generality, that the components are arranged in the increasing order of vertices so that, for $u < v$,

$$i < j \quad \text{for all } i \in \mathcal{I}_u \text{ and } j \in \mathcal{I}_v.$$

The distance between two successive components $\mathcal{I}_u$ and $\mathcal{I}_{u+1}$ is defined as

$$d_u := d(\mathcal{I}_u, \mathcal{I}_{u+1}) = \inf_{i\in\mathcal{I}_u,\, j\in\mathcal{I}_{u+1}} (j - i).$$

Note that $d_u \ge \pi_n$ for any u. For given m components $\mathcal{I}_1, \dots, \mathcal{I}_m$, let $d_{(u)}$ denote the uth smallest distance among $d_1, \dots, d_{m-1}$. The size of a component is the number of vertices (accounting for multiplicity) it contains.

Suppose that the graph $\mathcal{G} = \{i_1, j_1, i_2, j_2, \dots, i_{2r+1}, j_{2r+1}\}$ comprises m components.

Since $\|f_i\|/\sqrt{k}\log^\beta n \le M$ for all i, we also have $|f_{i,h}|/\sqrt{k}\log^\beta n \le M$ for all i and $h = 1,\dots,k$. We can then apply Lemma S.1 with $p_1 = \cdots = p_k = \infty$ and claim that

$$\mathbb{E}\left(\prod_{s=1}^{2r+1} f_{i_s,h_s} f_{j_s,h_s}\right) = \prod_{\ell=1}^{m} \mathbb{E}\left(\prod_{s:i_s\in\mathcal{I}_\ell} f_{i_s,h_s} \prod_{s:j_s\in\mathcal{I}_\ell} f_{j_s,h_s}\right) + O\big(\rho^{d_{(1)}} k^{2r+1} \log^{\beta(4r+1)} n\big).$$

We shall use the type of decomposition above routinely throughout our subsequent derivations.

Let $\mathcal{G}_m$ be the set of graphs having exactly m components. We have

$$\begin{aligned}\mathbb{E}\big(U_n^{2r+1}\big) &= \frac{1}{(n\sqrt{k})^{2r+1}} \sum_{m=1}^{4r+2} \sum_{\substack{i_1,j_1,\dots,i_{2r+1},j_{2r+1}\in\mathcal{G}_m\\ i_a\neq j_a,\, a=1,\dots,2r+1}} \mathbb{E}\left(\prod_{s=1}^{2r+1} f_{i_s}' f_{j_s}\right)\\ &= \frac{1}{(n\sqrt{k})^{2r+1}} \sum_{m=1}^{4r+2} \sum_{\substack{i_1,j_1,\dots,i_{2r+1},j_{2r+1}\in\mathcal{G}_m\\ i_a\neq j_a,\, a=1,\dots,2r+1}} \sum_{h_1,\dots,h_{2r+1}=1}^{n} \mathbb{E}\left(\prod_{s=1}^{2r+1} f_{i_s,h_s} f_{j_s,h_s}\right). \qquad \text{(S.6)}\end{aligned}$$

(a) Consider a graph with $m \ge 2r+2$ components. Then there exists at least one component, $\mathcal{I}_{\ell_0} = \{i_0\}$, with size 1 and

$$\mathbb{E}\left(\prod_{s:i_s\in\mathcal{I}_{\ell_0}} f_{i_s,h_s} \prod_{s:j_s\in\mathcal{I}_{\ell_0}} f_{j_s,h_s}\right) = \mathbb{E}(f_{i_0,h_s}) = 0.$$

Thus, for such graphs,

$$\mathbb{E}\left(\prod_{s=1}^{2r+1} f_{i_s,h_s} f_{j_s,h_s}\right) = O\big(\rho^{d_{(1)}} k^{2r+1} \log^{\beta(4r+2)} n\big) = O\big(\rho^{\pi_n} k^{2r+1} \log^{\beta(4r+2)} n\big).$$

Clearly, the total number of graphs with at least $2r+2$ components is less than or equal to $(n(n-1))^{2r+1}$, the total number of graphs in the expansion in (S.6). As a result,

$$\begin{aligned}&\frac{1}{(n\sqrt{k})^{2r+1}} \sum_{m=2r+2}^{4r+2} \sum_{\substack{i_1,j_1,\dots,i_{2r+1},j_{2r+1}\in\mathcal{G}_m\\ i_a\neq j_a,\, a=1,\dots,2r+1}} \sum_{h_1,\dots,h_{2r+1}=1}^{n} \mathbb{E}\left(\prod_{s=1}^{2r+1} f_{i_s,h_s} f_{j_s,h_s}\right)\\ &\quad= O\big(n^{2r+1} k^{3(r+1/2)} \rho^{\pi_n} \log^{\beta(4r+2)} n\big) = o(1),\end{aligned}$$

where the final order of magnitude is obtained by setting $\pi_n = n^a$ for some $a \in (0,1)$.

(b) Consider graphs with $m \le 2r$ components. Recalling (S.4), we have

$$A_n := \left|\frac{1}{(n\sqrt{k})^{2r+1}} \sum_{m=1}^{2r} \sum_{\substack{i_1,j_1,\dots,i_{2r+1},j_{2r+1}\in\mathcal{G}_m\\ i_a\neq j_a,\, a=1,\dots,2r+1}} \mathbb{E}\left(\prod_{s=1}^{2r+1} f_{i_s}' f_{j_s}\right)\right| \le \frac{c_{M,r}\big(\sqrt{k}\log^\beta n\big)^{2r+1}}{(n\sqrt{k})^{2r+1}} \sum_{m=1}^{2r} N_m,$$

where N_m is the number of graphs with exactly m components and $c_{M,r}$ is a constant depending only on M and r. We next find an upper bound for N_m. Define the degree of a component as the number of different vertices it contains and let δ_ℓ denote the degree of the component $\mathcal{I}_\ell$ for $\ell = 1, \dots, m$. Let $N_m(\delta_1, \dots, \delta_m)$ be the number of graphs with m components with δ_ℓ's being the respective components. Recall that $m \le \sum_{u=1}^{m} \delta_u \le 4r+2$. Hence,

$$N_m(\delta_1, \dots, \delta_m) = n\pi_n^{\delta_1 - 1} n\pi_n^{\delta_2 - 1} \cdots n\pi_n^{\delta_m - 1} = n^m \pi_n^{\delta_1 + \cdots + \delta_m - m} \le n^m \pi_n^{4r+2-m}.$$

Thus,

$$N_m \le (4r+2)^m n^m \pi_n^{4r+2-m}.$$

For $1 \le m \le 2r$, we have

$$N_m \le (4r+2)^{2r} n^{2r} \pi_n^{4r+1} = O\big(n^{2r}\pi_n^{4r+1}\big) \quad \text{and}$$
$$\sum_{m=1}^{2r} N_m \le 2r(4r+2)^{2r} n^{2r} \pi_n^{4r+1} = O\big(n^{2r}\pi_n^{4r+1}\big).$$

Taking $\pi_n = n^{\frac{1}{4r+1} - \epsilon}$ for some $\epsilon \in (0, \frac{1}{4r+1})$ ensures that $A_n = o(1)$.

(c) We are left with the graphs with $m = 2r+1$ components. Note that if there is one component of such graphs that has more than two vertices (accounting for multiplicities), then there is at least one component with exactly one vertex. Similar to part (a) of the proof, the contribution of such graphs to the moment in (S.6) is negligible. We are left with the case where each component has exactly 2 vertices. Again, if any of these components, say $\mathcal{I}_{\ell_0}$, has size 2, then by the martingale difference assumption, $\mathbb{E}(f_{i,h} f_{j,h'}) = 0$ for $i \neq j \in \mathcal{I}_{\ell_0}$ and $h, h' = 1, \dots, k$. Graphs containing such components also have negligible contribution. There only remain graphs with components each having two equal vertices. Let G'_{2r+1} be the set of all graphs with $2r+1$ components, each of size 1. Note that each graph in G'_{2r+1} contains $2r+1$ different vertices, each featuring exactly twice. We can check along the same lines as in (a) that

$$\frac{1}{(n\sqrt{k})^{2r+1}} \sum_{\substack{i_1, j_1, \dots, i_{2r+1}, j_{2r+1} \in \mathcal{G}'_{2r+1} \\ i_a \neq j_a, a=1,\dots,2r+1}} \mathbb{E}\left(\prod_{s=1}^{2r+1} f'_{i_s} f_{j_s}\right)$$
$$= \frac{1}{(n\sqrt{k})^{2r+1}} \sum_{\substack{i_1, j_1, \dots, i_{2r+1}, j_{2r+1} \in \mathcal{G}'_{2r+1} \\ i_a \neq j_a, a=1,\dots,2r+1}} \mathbb{E}^*\left(\prod_{s=1}^{2r+1} f'_{i_s} f_{j_s}\right) + o(1),$$

where $\mathbb{E}^*$ is the expectation under pairwise independence of $(2r+1)$ distinct $f_\bullet$'s appearing in the expectation.

Take $E := \prod_{s=1}^{2r+1} f'_{i_s} f_{j_s}$ with $i_1, j_1, \dots, i_{2r+1}, j_{2r+1} \in \mathcal{G}'_{2r+1}$. Since each f_{i_s} (or f_{j_s}) appears exactly twice in the product, it either appears in a square inner product: $(f'_{i_s} f_{j_s})^2$ and nowhere else, or it appears in two different inner products: $(f'_{i_s} f_{j_s})(f'_{i_{s'}} f_{j_{s'}})$ (with $i_s = i_{s'}$) and nowhere else. Because we are in the presence of odd number $(2r+1)$ of inner product terms in E, there is at least one i_{s_1} such that $f_{i_{s_1}}$ appears in two different inner products, that is, factors such as $(f'_{i_{s_1}} f_{j_{s_1}})$ and $(f'_{i_{s_2}} f_{j_{s_2}})$ with $i_{s_1} = i_{s_2}$ and $j_{s_1} \neq j_{s_2}$ appear in E.

Let

$$(f'_{i_{s_1}} f_{j_{s_1}}) \cdot (f'_{i_{s_2}} f_{j_{s_2}}) \cdots (f'_{i_{s_u}} f_{j_{s_u}})$$

be a product of factors from E featuring pairwise different inner products but with each vertex appearing exactly twice. Clearly, $u \geq 3$. From the discussion above, such a product can be found for any graph in $\mathcal{G}'_{2r+1}$.

Now, consider $\mathbb{E}^*(\prod_{s=1}^{2r+1} f'_{i_s} f_{j_s})$. By (S.4), we have

$$\begin{aligned}
\mathbb{E}^*\left(\prod_{s=1}^{2r+1} f'_{i_s} f_{j_s}\right) &\leq c_{M,r}\big(\sqrt{k}\log^\beta n\big)^{2r+1-u}\mathbb{E}^*\big((f'_{i_{s_1}} f_{j_{s_1}}) \cdot (f'_{i_{s_2}} f_{j_{s_2}}) \cdots (f'_{i_{s_u}} f_{j_{s_u}})\big) \\
&\leq c_{M,r}\big(\sqrt{k}\log^\beta n\big)^{2r-2}\mathbb{E}^*\big((f'_{i_{s_1}} f_{j_{s_1}}) \cdot (f'_{i_{s_2}} f_{j_{s_2}}) \cdots (f'_{i_{s_u}} f_{j_{s_u}})\big) \\
&= c_{M,r}\big(\sqrt{k}\log^\beta n\big)^{2r-2} k,
\end{aligned}$$

where $c_{M,r}$ is a constant depending only on M and r. Since the number of graphs in $\mathcal{G}'_{2r+1}$ is less than n^{2r+1} (a graph in $\mathcal{G}'_{2r+1}$ corresponds in particular to a subset of $2r+1$ elements from a set of n elements), we have

$$\begin{aligned}
\frac{1}{(n\sqrt{k})^{2r+1}} \sum_{\substack{i_1, j_1, \ldots, i_{2r+1}, j_{2r+1} \in \mathcal{G}'_{2r+1} \\ i_a \neq j_a, a=1,\ldots,2r+1}} \mathbb{E}^*\left(\prod_{s=1}^{2r+1} f'_{i_s} f_{j_s}\right) &\leq \frac{c_{M,r}\big(\sqrt{k}\log^\beta n\big)^{2r-2} k n^{2r+1}}{(n\sqrt{k})^{2r+1}} \\
&= \frac{c_{M,r}\log^{\beta(2r-2)} n}{\sqrt{k}} = o(1).
\end{aligned}$$

Parts (a), (b), and (c) allow us to conclude that $\mathbb{E}(U_n^{2r+1}) = o(1)$.

(d) Let us now obtain the limit of $\mathbb{E}(U_n/\sqrt{2})^{2r}$. Similar to (S.6), we can claim that

$$\mathbb{E}\big((U_n/\sqrt{2})^{2r}\big) = \frac{1}{(n\sqrt{2k})^{2r}} \sum_{m=1}^{4r} \sum_{\substack{i_1, j_1, \ldots, i_{2r}, j_{2r} \in \mathcal{G}_m \\ i_a \neq j_a, a=1,\ldots,2r}} \mathbb{E}\left(\prod_{s=1}^{2r} f'_{i_s} f_{j_s}\right). \tag{S.7}$$

Here, we shall distinguish the cases: $m \geq 2r+1$, $m \leq 2r-1$, and $m = 2r$. Similar to the arguments in (a) and (b) above, we can claim that those graphs have negligible contribution to $\mathbb{E}((U_n/\sqrt{2})^{2r})$. The discussion in (c) is also valid here and only graphs in $\mathcal{G}'_{2r}$ can contribute to this expectation. Also, graphs with vertices of equal value appearing in different inner products do not contribute significantly. We are left with graphs featuring only square of inner products. Such graphs exist since the number of inner product in each term of the summation in (S.7) is even ($2r$). We can write

$$\mathbb{E}\big((U_n/\sqrt{2})^{2r}\big) = \frac{1}{(n\sqrt{2k})^{2r}} \sum_{\substack{i_1, j_1, \ldots, i_r, j_r \in \mathcal{G}'_{2r} \\ i_a \neq j_a, a=1,\ldots,r}} \mathbb{E}^*\left(\prod_{s=1}^{r} \big(f'_{i_s} f_{j_s}\big)^2\right) + o(1).$$

(In this expression, we represent graphs in $\mathcal{G}'_{2r}$ that determine $\prod_{s=1}^{r}(f'_{i_s}f_{j_s})^2$ by $i_1, j_1, \dots, i_r, j_r$ without the need to repeat each vertex.) Recall that $U_n = (2/n)\sum_{i<j}(f'_i f_j/\sqrt{k})$ so that

$$\mathbb{E}\left(\frac{U_n}{\sqrt{2}}\right)^{2r} = 2^{2r}\frac{1}{(n\sqrt{2k})^{2r}}\sum_{\substack{i_1,j_1,\dots,i_r,j_r\in\mathcal{G}'_{2r}\\ i_a<j_a,a=1,\dots,r}}\mathbb{E}^*\left(\prod_{s=1}^{r}\left(f'_{i_s}f_{j_s}\right)^2\right)+o(1). \tag{S.8}$$

We have

$$\mathbb{E}^*\left(\prod_{s=1}^{r}\left(f'_{i_s}f_{j_s}\right)^2\right) = \prod_{s=1}^{r}\mathbb{E}^*\left(\left(f'_{i_s}f_{j_s}\right)^2\right) = \prod_{s=1}^{r}\mathbb{E}^*\left(f'_{i_s}f_{j_s}f'_{j_s}f_{i_s}\right) = \prod_{s=1}^{r}\mathrm{trace}\mathbb{E}^*\left(f_{i_s}f'_{i_s}\right) = k^r.$$

The number of times that a specific term $\prod_{s=1}^{r}(f'_{i_s}f_{j_s})^2$ appears in the expansion of $\mathbb{E}(U_n/\sqrt{2})^{2r}$ is given by the multinomial formula as $(2r)!/2^r$. We are left to determine the cardinality number of graphs in $\mathcal{G}'_{2r}$ such that $i_s < j_s$ for $s = 1, \dots, r$. That is the cardinality of the set

$$S = \left\{\left\{(i_1, j_1), \dots, (i_r, j_r)\right\} \subset \Delta_n : \forall s, s' = 1, \dots, r, |a_s - a_{s'}| \geq \pi_n \text{ with } a_\bullet = i_\bullet \text{ or } j_\bullet\right\},$$

with $\Delta_n := \{(i, j) : i < j, \text{ and } i, j = 1, \dots, n\}$.

Define

$$S_r = \left\{\left\{(i_1, j_1), \dots, (i_r, j_r)\right\} \subset \Delta_n\right\},$$

and

$$S_2 = \left\{\left\{(i_1, j_1), \dots, (i_r, j_r)\right\} \subset \Delta_n : \exists s, s' \in \{1, \dots, r\} : |a_s - a_{s'}| \leq \pi_n\right\}.$$

We have $S = S_r \setminus S_2$ so that $\mathrm{Card}(S) = \mathrm{Card}(S_r) - \mathrm{Card}(S_2)$. Note that

$$\mathrm{Card}(S_1) = \begin{pmatrix}\frac{n(n-1)}{2}\\ r\end{pmatrix} = \frac{1}{r!}\frac{n(n-1)}{2}\left(\frac{n(n-1)}{2}-1\right)\cdots\left(\frac{n(n-1)}{2}-r+1\right)$$

and

$$\mathrm{Card}(S_2) \leq \sum_{j=1}^{2r-1}\binom{2r-1}{j}\pi_n^j n^{2r-j} = o\left(n^{2r}\right),$$

where in this summation j represents the number of increments smaller than π_n as we sort the $2r$ vertices i_s's and j_s's in increasing order. Hence,

$$\mathrm{Card}(S) = \frac{n^{2r}}{r!2^r}\left(1+o(1)\right).$$

The result follows by noting that (S.8) can be rewritten as

$$\mathbb{E}\left(\frac{U_n}{\sqrt{2}}\right)^{2r} = 2^{2r}\frac{1}{(n\sqrt{2k})^{2r}}\frac{n^{2r}}{r!2^r}\left(1+o(1)\right)\frac{(2r)!}{2^r}k^r = \frac{(2r)!}{2^r r!}+o(1). \qquad \square$$

Additional simulation evidence

This additional set of simulations is based on the following design. Let the sample $\{(y_t, x_t) : t = 1, \dots, T\}$ be such that $y_t \sim iid(\theta_0, 1)$ and y_t^2 is unrelated to x_t in the sense that $\mathbb{E}((y_t - \theta_0)^2|x_t) = 1$ while the relation between y_t and x_t is left unrestricted. We can verify that the conditional moment restriction

$$\mathbb{E}\big((y_t - \theta)^2 - 1|x_t\big) = 0$$

point identifies θ_0. Letting $h(x_t)$ be any suitable function of x_t, the Jacobian of $\theta \mapsto \mathbb{E}(h(x_t) \cdot ((y_t - \theta)^2 - 1))$ at θ_0 is given by

$$-2\mathbb{E}\big(h(x_t) \cdot (y_t - \theta_0)\big) = \mu_h.$$

If x_t is independent of y_t, $\mu_h = 0$ for any choice of h and we have first-order local identification failure. Otherwise, μ_h may be nonzero for some h and the model in this case is first-order locally identified.

For the purpose of illustrating the properties of our test in the presence of first-order local identification failure, we specialize the data generating process further. We assume that $y_t \sim iidN(0, 1)$ and $x_t \sim iidN(0, 1)$ for $t = 1, \dots, n$, with y being independent of x. The moment condition tested is

$$H_0 : \Pr\big\{\mathbb{E}\big(u_t(\theta)|x_t\big) = 0\big\} = 1,$$

where $u_t(\theta) = (y_t - \theta)^2 - 1$. The null H_0 is correct and identifies the true value $\theta_0 = 0$ of the parameter θ. This design is characterized with first-order local identification failure even though global identification is ensured. Local identification is obtained at second order according to Assumption 2(iii). Here, $k_0 = 1$ since with the instrument $z_{t,1} = 1$, we have $\mathbb{E}(\nabla_{\theta\theta}(z_{t,1}u_i(\theta_0))) = -2 \neq 0$.

We present results for three specification tests: the traditional J test, the conditional moment restriction test by Smith (2007) and Tripathi and Kitamura (2003), denoted by S-TK, and the test $\hat{Z}$ based on the $N(0, 1)$ asymptotic approximation. The J test is performed on the unconditional moment conditions $\mathbb{E}(z_t u_t(\theta)) = 0$ with $z_t = (1, x_t^2)'$. We present results for the J test under the standard chi-squared asymptotics with $k_z - p$ (in this case, one) degrees-of-freedom, the conservative chi-squared asymptotics with k_z (in this case, two) degrees-of-freedom, as well as the mixture of chi-squared distribution $\frac{1}{2}\chi^2(1) + \frac{1}{2}\chi^2(2)$ proposed in Dovonon and Renault (2013).

As argued above, the Jacobian in this example is zero. Lee and Liao (2018) propose to add this as a moment condition, that is, $\mathbb{E}(z_t e_t(\theta)) = 0$ with $e_t(\theta) = y_t - \theta$, which will restore the first-order local identification and standard inference. For the chosen instruments $z_t = (1, x_t^2)'$, we use this augmented set of four moment conditions to obtain the two-step GMM estimator and compare the resulting test for overidentifying restrictions against the $\chi^2(3)$ critical values. This method is denoted by "augmented" J test.

The conditional moment restriction tests, S-TK and $\hat{Z}$, involve a choice of tuning parameters. The tuning parameters for the S-TK test are the biweight kernel $k(x) = \frac{15}{16}(1 - x^2)^2 1_{\{|x| \le 1\}}$ for a kernel function, $K^{**} = \frac{1,168,780}{2,263,261}$, kernel bandwidth $b_n = n^{-1/4}$,

TABLE S.1. Empirical rejection rates of specification tests under the null (size).

	J Test											
	$\chi^2(1)$		$\chi^2(2)$		Mixture		Augmented		S-TK Test		$\hat{Z}$ Test	
n	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
100	10.5	4.3	2.4	0.9	4.4	1.5	18.4	11.7	2.9	1.5	8.3	4.5
200	11.0	5.0	3.3	1.2	5.1	2.1	15.6	9.8	3.4	1.9	8.8	4.9
500	13.1	6.4	3.9	1.8	6.6	2.9	13.1	7.4	4.4	2.1	8.1	4.0
1000	13.6	6.9	4.2	1.9	7.1	3.1	12.5	7.0	6.0	2.8	9.1	4.5

Note: The table presents the empirical size of the tests at 5% and 10% nominal level. 'Mixture" stands for $\frac{1}{2}\chi^2(1)+\frac{1}{2}\chi^2(2)$ and "augmented" denotes the method by Lee and Liao (2018) that uses critical values from a $\chi^2(3)$ distribution. The J test is performed on the unconditional moment conditions $\mathbb{E}(z_t u_t(\theta))=0$ or $\mathbb{E}(z_t\binom{u_t(\theta)}{e_t(\theta)})=0$ (for the "augmented" method) with $z_t=(1,x_t^2)'$, S-TK test is the conditional moment restriction test by Smith (2007) and Tripathi and Kitamura (2003), and $\hat{Z}$ test is the conditional moment restriction test proposed in this paper. The results are based on 10,000 Monte Carlo replications.

and trimming function $1_{\{|x|\le S_*\}}$ with $S_*=1.96$. The test is computed for the continuously updated GMM. For our $\hat{Z}$ test, the tuning parameters are $\Psi(\cdot):\mathbb{R}\to[-\pi,\pi]$, $x\mapsto 2\arctan(x)$, series of bounded functions $g_l(\cdot):[-\pi,\pi]\to[-1,+1]$, $x\mapsto\cos(lx)$ for $l=1,\ldots k$, and $k=5n^{1/6}$ (see footnote 10 in the paper for further discussion on the choice of k).[3]

The results for the empirical size of the tests at 5% and 10% nominal level are reported in Table S.1 as the sample size n increases. The three standard J tests behave as expected: the $\chi^2(1)$ approximation is leading to overrejections, the $\chi^2(2)$ approximation is being conservative, and the mixture of chi-squares is approaching the nominal level as the sample size increases. The augmentation method by Lee and Liao (2018) results in overrejections in small samples as the empirical size is converging to the nominal level with the sample size. The S-TK test tends to be undersized although its properties appear to improve in large samples. The $\hat{Z}$ test controls size for all sample sizes with empirical rejections close to the nominal level.

To visualize the quality of the $N(0,1)$ approximation for the $\hat{Z}$ test in this setting, the left panel of Figure S.1 plots the kernel density obtained from 10,000 replications and $n=10{,}000$ against the $N(0,1)$ density. The simulated density appears to be very close to the limiting distribution. The middle plot in Figure S.1 presents the simulated kernel density for the GMM estimator of θ while the right plot provides the histogram for the squared GMM estimator of θ. The histogram for the squared GMM estimator nicely illustrates the half probability mass at zero suggested by theory (Theorem 4.2). In addition, the simulated kernel density for $\hat{\theta}$ reveals some features in the shape of the distribution that are not immediately evident in the histogram for $\hat{\theta}^2$.

[3] In less structured problems, the modern machine learning literature provides a wider and more flexible choice set of basis functions. By contrast, our setup has much more structure and the use of a preselected set of basis functions seems appropriate. The inclusion of an increasing $n^{1/6}$ number of basis functions, selected (in a deterministic fashion) from a predetermined set, yields a test that is asymptotically correct and consistent. It is possible to optimize some properties of the test by relying on the advances in machine learning in choosing a more suitable set of basis functions but this is beyond the scope of the paper.

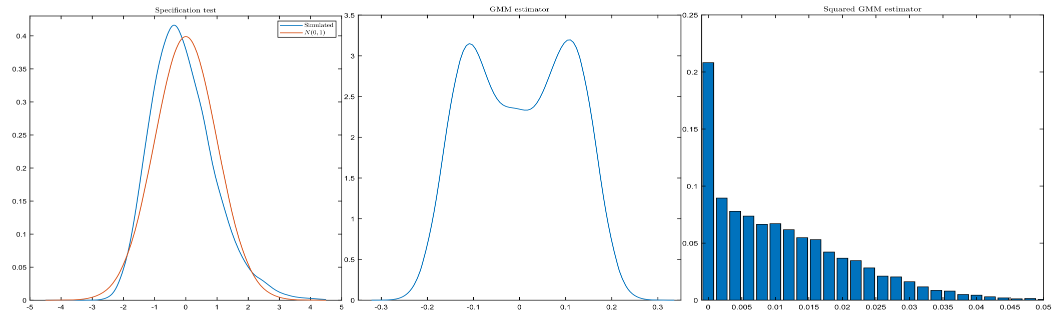

FIGURE S.1. The left graph presents the Monte Carlo kernel density and the limit $N(0,1)$ density for the specification test. The middle and right graphs plot the Monte Carlo kernel density and histogram for the GMM estimator and the squared GMM estimator, respectively. The sample size is $n = 10{,}000$ and the number of Monte Carlo replications is 10,000.

BOUNDS OF EIGENVALUES FOR COSINE WEIGHT FUNCTION

Let $g^{(k)}(x) = (\cos(2l\Psi(x)))_{l=1,\dots,k}$, where $\Psi(x) = \arctan(x)$, be the vector of instruments used for inference based on the moment condition model

$$\mathbb{E}\big(u(y_t, \theta_0)|x_t\big) = 0 \quad \text{a.s.}$$

$$V_k = \mathbb{E}\big(u(y_t, \theta_0)^2 g^{(k)}(x_t) g^{(k)\prime}(x_t)\big).$$

Let $\sigma^2(x)$ be the conditional expectation of $u(y, \theta_0)^2$ given x and $p(x)$ be the common density function of x_t with respect to the Lebesgue measure. We show the following result.

PROPOSITION S.3. (a) *If $\sigma^2(x)p(x)$ is bounded away from 0, that is, there exists $\delta_1 > 0$ such that $\inf_{x\in\mathbb{R}} \sigma^2(x)p(x) \geq \delta_1$, then*

$$\inf_k \lambda_{\min}(V_k) \geq \frac{\delta_1 \pi}{2} > 0.$$

(b) *If $\sigma^2(x)(1+x^2)p(x)$ is bounded from above, that is, there exists $\delta_2 > 0$ such that $\sup_{x\in\mathbb{R}} \sigma^2(x)(1+x^2)p(x) \leq \delta_2$, then*

$$\sup_k \lambda_{\max}(V_k) \leq \frac{\delta_2 \pi}{2} < \infty.$$

The condition in part (a) is stated by de Jong and Bierens ((1994), Example 1, p.75) in establishing that the eigenvalues of V_k are bounded away from 0 for a specific choice of weight function $g^{(k)}(x)$. This condition holds if $p(x) > 0$ and, as x grows, $p(x)$ does not vanish faster than $1/\sigma^2(x)$. It is worth mentioning that this condition may be restrictive as it excludes, for instance, the case where $p(x)$ is the standard normal density function and $\sigma^2(x)$ is linear. However, it has to be understood that this is a sufficient but not a necessary condition for the eigenvalues to be bounded away from 0. We investigate by simulation that the boundedness of eigenvalues claimed by this proposition holds in the case of normal distribution and Student distribution with one degree-of-freedom.

The condition in part (b) is less restrictive. It holds, for instance, if $\sigma^2(x)$ and $p(x)$ are continuous and x has a bounded support. It also holds if x is Gaussian (or more generally, $p(x)$ is continuous and $p(x)e^{-ax} \to 0$ as $x \to \infty$ for some $a > 0$) and $\sigma^2(x)$ is bounded by a polynomial function of x. The proof of Proposition S.3 is provided below.

Proof. (a) Let $c \in \mathbb{R}^k$ such that $c'c = 1$. We have

$$\begin{aligned}
c'V_k c &= \sum_{i,j=1}^{k} c_i c_j \mathbb{E}\big(u(y_t, \theta_0)^2 g_i^{(k)}(x_t) g_j^{(k)}(x_t)\big) \\
&= \sum_{i,j=1}^{k} c_i c_j \mathbb{E}\big(\sigma^2(x_t)\cos\big(2i\arctan(x_t)\big)\cos\big(2j\arctan(x_t)\big)\big) \\
&= \int \sigma^2(x) \sum_{i,j=1}^{k} c_i c_j \cos\big(2i\arctan(x)\big)\cos\big(2j\arctan(x)\big)p(x)\,dx \\
&= \int \sigma^2(x)p(x)\left(\sum_{j=1}^{k} c_j \cos\big(2j\arctan(x)\big)\right)^2 dx \\
&\geq \delta_1 \int_{-\pi/2}^{\pi/2} \left(\sum_{j=1}^{k} c_j \cos(2jx)\right)^2 \big(1+\tan^2(x)\big)\,dx \\
&\geq \delta_1 \int_{-\pi/2}^{\pi/2} \left(\sum_{j=1}^{k} c_j \cos(2jx)\right)^2 dx \\
&= \delta_1 \sum_{i=1}^{k} \int_{-\pi/2}^{\pi/2} c_i^2 \cos^2(2ix)\,dx + \delta_1 \sum_{i,j=1, i\neq j}^{k} c_i c_j \int_{-\pi/2}^{\pi/2} \cos(2ix)\cos(2jx)\,dx.
\end{aligned}$$

We obtain

$$\int_{-\pi/2}^{\pi/2} \cos^2(2ix)\,dx = \frac{1}{2}\left[\frac{1}{4i}\sin(4ix) + x\right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = \frac{\pi}{2}$$

and, using integration by parts, we have, for $i \neq j$,

$$\int_{-\pi/2}^{\pi/2} \cos(2ix)\cos(2jx)\,dx = \frac{1}{2(i-j)}\left[\sin(2ix)\cos(2jx) - \frac{j}{i+j}\sin\big(2(i+j)x\big)\right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = 0.$$

Thus, $c'V_k c \geq \frac{\delta_1 \pi}{2}$.

(b) We have

$$c'V_k c = \int \sigma^2(x)p(x)\left(\sum_{j=1}^{k} c_j \cos\big(2j\arctan(x)\big)\right)^2 dx$$

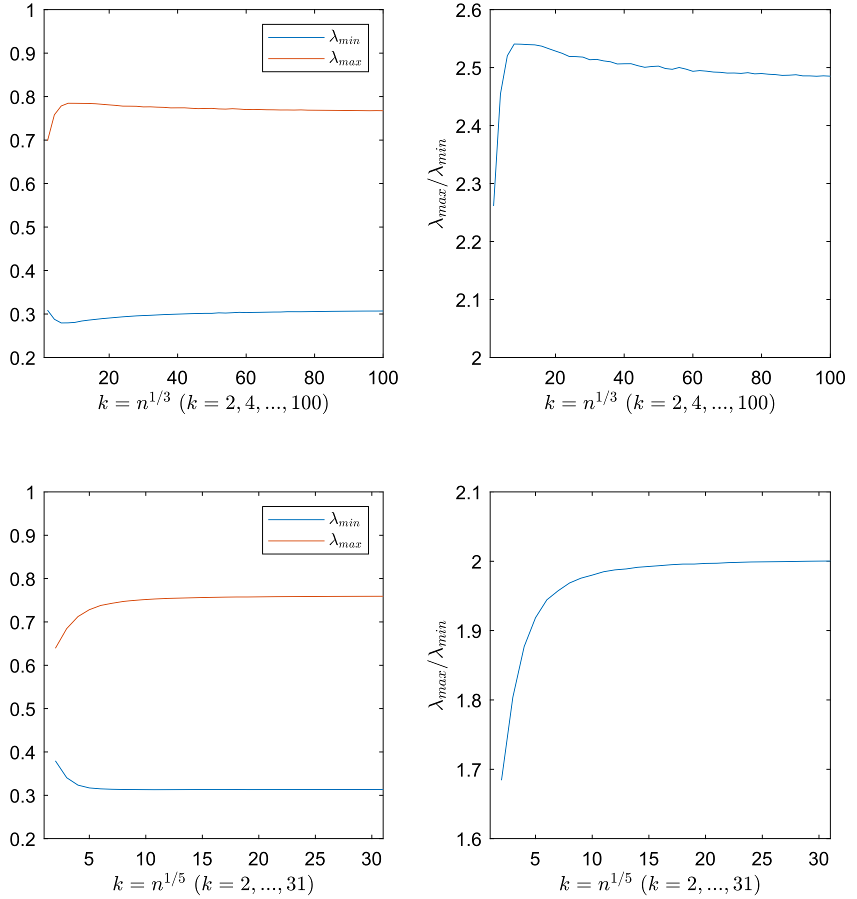

FIGURE S.2. Simulated smallest and largest eigenvalues and their ratios for $x \sim \text{NID}(0, 1)$. The top row is for $k = n^{1/3}$ ($n = 8, 64, \ldots,$ and 1,000,000) and the bottom row is for $k = n^{1/5}$($n = 32, 243, \ldots,$ and 28,629, 151).

$$= \int_{-\pi/2}^{\pi/2} \left(\sum_{j=1}^{k} c_j \cos(2jx) \right)^2 \sigma^2\big(\tan(x)\big)\big(1 + \tan^2(x)\big)\, p\big(\tan(x)\big)\, dx$$

$$\leq \delta_2 \int_{-\pi/2}^{\pi/2} \left(\sum_{j=1}^{k} c_j \cos(2jx) \right)^2 dx = \frac{\delta_2 \pi}{2}. \qquad \square$$

As mentioned above, deriving analytically a meaningful lower bound for the eigenvalues of V_k when x has a certain type of distribution is difficult. Instead, we proceed by simulations to obtain those bounds when x follows a standard normal distribution and a t distribution with one degree-of-freedom.

We plot the simulated the smallest ($\lambda_{\min}$) and largest eigenvalue ($\lambda_{\max}$) of $\mathbb{E}[g^{(k)}(x) \times g^{(k)\prime}(x)]$ along with their ratio. Note that, under homoskedasticity ($\sigma^2(x) = \text{constant}$), boundedness of V_k—both away from 0 and from above—is equivalent to that of

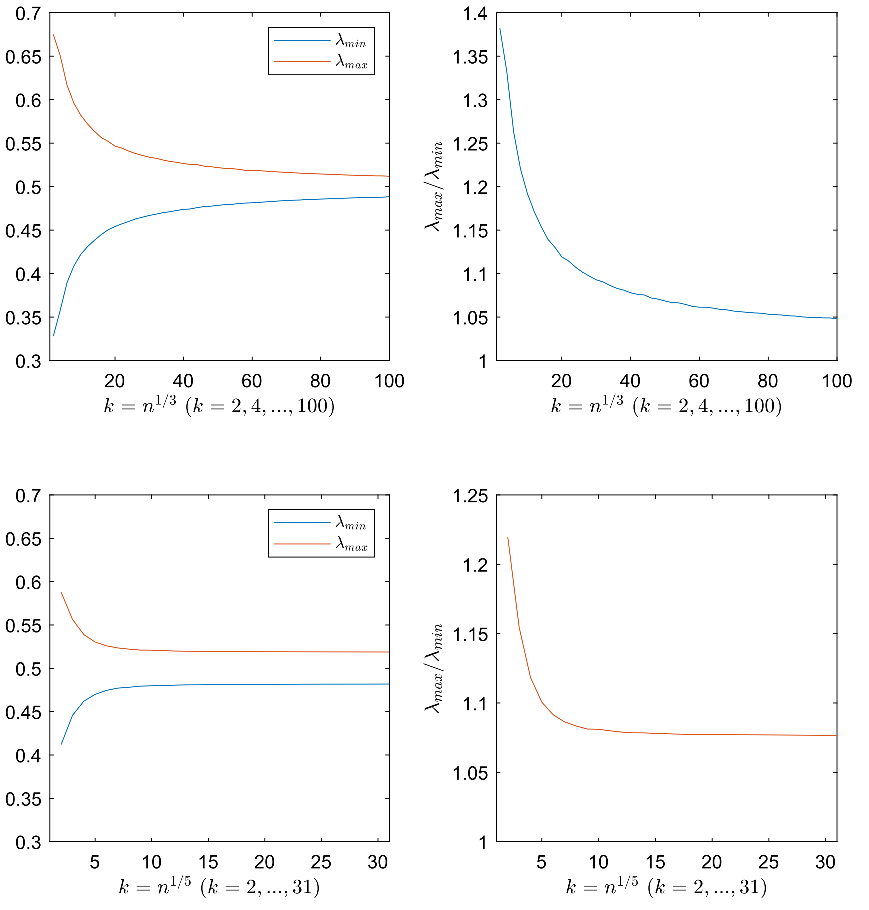

FIGURE S.3. Simulated smallest and largest eigenvalues and their ratios for $x \sim t(1)$. The top row is for $k = n^{1/3}(n = 8, 64, \ldots,$ and $1{,}000{,}000)$ and the bottom row is for $k = n^{1/5}$ $(n = 32{,}243, \ldots,$ and $28{,}629, 151)$.

$\mathbb{E}[g^{(k)}(x)g^{(k)'}(x)]$. We consider $g_j^{(k)}(x) = \cos(4j \arctan(x))$, $j = 1, \ldots, k$, with two choices of k: $k = n^{1/3}$ and $k = n^{1/5}$. For $k = n^{1/3}$, we consider $k = 2, 4, \ldots,$ and 100, that is, $n = 8, 64, \ldots,$ and 1,000,000. For $k = n^{1/5}$, we consider $k = 2, 3, \ldots,$ and 31, that is, $n = 32{,}243, \ldots,$ and $28{,}629, 151$. For each simulated sample, we calculate the eigenvalues of the sample mean of $g^{(k)}(x)g^{(k)'}(x)$, and for each sample size, simulated smallest and largest eigenvalues are obtained as an average of the smallest and largest eigenvalues over R replications. For $k = n^{1/3}$, $R = 1000$ for $k \leq 40$, and $R = 200$ for $k > 40$. For $k = n^{1/5}$, $R = 1000$ for $k \leq 17$, and $R = 200$ for $k > 17$.

Figures S.2 and S.3 plot $\lambda_{\min}$ and $\lambda_{\max}$ and the ratio: $\lambda_{\max}/\lambda_{\min}$. Specifically, in Figure S.2, $x \sim \text{NID}(0, 1)$ with $k = n^{1/3}$ (top) and $k = n^{1/5}$ (bottom). In Figure S.3, $x \sim t(1)$ with $k = n^{1/3}$ (top) and $k = n^{1/5}$ (bottom). These figures, along with the theoretical results in Proposition S.3, provide support for our assumption that the eigenvalues are bounded away from 0 and are bounded from above. In addition, as the sample size grows, the eigenvalues seem very stable and they appear to converge to fixed values.

References

Ando, Tsuyoshi and J. Leo van Hemmen (1980), "An inequality for trace ideals." *Communications in Mathematical Physics*, 76, 143–148. [5]

de Jong, Robert and Herman Bierens (1994), "On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity." *Econometric Theory*, 10, 70–90. [6, 7, 20]

Dedecker, Jérôme and Clémentine Prieur (2005), "New dependence coefficients. Examples and applications to statistics." *Probability Theory and Related Fields*, 132, 203–236. [7]

Dovonon, Prosper and Eric Renault (2013), "Testing for common conditionally heteroskedastic factors." *Econometrica*, 81, 2561–2586. [18]

Fan, Yanqin and Qi Li (1999), "Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing." *Journal of Nonparametric Statistics*, 10, 245–271. [7]

Gao, Jiti (2007), *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. CRC Press, Florida USA. [7]

Gao, Jiti and Yongmiao Hong (2008), "Central limit theorems for generalized U-statistics with applications in nonparametric specification." *Journal of Nonparametric Statistics*, 20, 61–76. [7]

Hall, Peter (1984), "Central limit theorem for integrated square error of multivariate nonparametric density estimators." *Journal of Multivariate Analysis*, 14, 1–16. [7]

Kim, Tae, Zhi-Ming Luo, and Chiho Kim (2011), "The central limit theorem for degenerate variable U-statistics under dependence." *Journal of Nonparametric Statistics*, 23, 683–699. [7, 8, 13]

Lee, Ji and Zhipeng Liao (2018), "On standard inference for GMM with local identification failure of known forms." *Econometric Theory*, 34, 790–814. [18, 19]

Leucht, Anne (2012), "Degenerate U- and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency." *Bernoulli*, 18, 552–585. [7]

Roussas, George and Dimitrios Ioannides (1987), "Moment inequalities for mixing sequences of random variables." *Stochastic Analysis and Applications*, 5, 61–120. [1]

Smith, Richard (2007), "Efficient information theoretic inference for conditional moment restrictions." *Journal of Econometrics*, 138, 430–460. [18, 19]

Tripathi, Gautam and Yuichi Kitamura (2003), "Testing conditional moment restrictions." *Annals of Statistics*, 31, 2059–2095. [18, 19]

Co-editor Stéphane Bonhomme handled this manuscript.

Manuscript received 30 August, 2022; final version accepted 6 May, 2024; available online 6 May, 2024.