Hoesch, Lukas; Lee, Adam; Mesters, Geert

Article

Locally robust inference for non-Gaussian SVAR models

Quantitative Economics

Provided in Cooperation with:
The Econometric Society

Suggested Citation: Hoesch, Lukas; Lee, Adam; Mesters, Geert (2024) : Locally robust inference for non-Gaussian SVAR models, Quantitative Economics, ISSN 1759-7331, The Econometric Society, New Haven, CT, Vol. 15, Iss. 2, pp. 523-570, https://doi.org/10.3982/QE2274

Supplement to "Locally robust inference for non-Gaussian SVAR models"

(*Quantitative Economics*, Vol. 15, No. 2, May 2024, 523–570)

Lukas Hoesch
Department of Econometrics and Data Science, Vrije Universiteit Amsterdam and Tinbergen Institute

Adam Lee
Department of Data Science and Analytics, BI Norwegian Business School

Geert Mesters
Department of Economics and Business, Universitat Pompeu Fabra, Barcelona School of Economics, and CREI

In this Supplementary Material, we provide the following additional results:

SA: Choice for the parametrization

SB: Technical details for the main proofs

SC: Some technical tools

Keywords. Weak identification, semiparametric inference, hypothesis testing, impulse responses, independent component analysis.
JEL classification. C32, C39, C51.

Appendix SA: Parametrization of the semiparametric SVAR model

Under the main assumptions of the paper (i.e., Assumptions 1 and 2), the parameters of the SVAR are generally not locally identified. Even under the *additional* assumption that the errors $\epsilon_{k,t}$ follow non-Gaussian distributions, we have that $A(\alpha, \sigma)$ can only be identified up to permutation and sign changes of its rows (e.g., Comon (1994)).

Therefore, to ensure that we study economically interesting permutations we typically need to impose additional identifying restrictions, such as zero or sign restrictions. The choice for such restrictions interacts with the chosen parametrization for $A(\alpha, \sigma)$ for which we give a few examples.

Example S1 (Supply and demand). Following Baumeister and Hamilton (2015), when the SVAR defines a demand and a supply equation, we can set

$$A^{-1}(\alpha, \sigma) = \begin{pmatrix} -\alpha^d & 1 \\ -\alpha^s & 1 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}, \tag{S1}$$

Lukas Hoesch: l.hoesch@vu.nl
Adam Lee: adam.lee@bi.no
Geert Mesters: geert.mesters@upf.edu

Available at http://qeconomics.org. https://doi.org/10.3982/QE2274

where $\alpha = (\alpha^d, \alpha^s)'$ are the short run demand and supply elasticities, and $\sigma = (\sigma_1, \sigma_2)'$ scales the structural shocks. With independent non-Gaussian errors, A is identified up to permutation and sign changes of its rows. To pin down an economically interesting rotation, we can impose the sign restrictions $\alpha^d \leq 0$, $\alpha^s \geq 0$, and $\sigma_1, \sigma_2 > 0$.

EXAMPLE S2 (Rotation matrix). A canonical choice sets

$$A^{-1}(\alpha, \sigma) = \Sigma^{1/2}(\sigma) R(\alpha), \tag{S2}$$

where $\Sigma^{1/2}(\sigma)$ is a lower triangular matrix (with positive diagonal elements) defined by the vector σ and $R(\alpha)$ is a rotation matrix that is parametrized by the vector α. Different parametrizations for the rotation matrix are possible; see Magnus, Pijls, and Sentana (2021) for a detailed discussion. Similar to in Example S1, even with independent non-Gaussian errors $R(\alpha)$ is not uniquely identified and additional zero-, sign-, or long-run-restrictions are needed to pin down the desired rotation.

As the above examples make clear, several commonly used parametrizations can be adopted. Three general comments apply.

First, pinning down a specific permutation, as in the first example, is necessary for the economic interpretation of the results, but it is not necessary for the score testing methodology of the paper, which fixes α under the null.

Second, the robust non-Gaussian approach of this paper can be combined with any of the existing SVAR identification approaches to obtain an economically interesting specification. Besides zero and sign restrictions, one can also think of combining with external instruments or more general prior information as in Baumeister and Hamilton (2015) or Braun (2021).

Third, often multiple parametrizations are possible. We recommend jointly testing the possibly weakly identified parameters when they are of direct economic interest (e.g., Example 1). In contrast, when the interest is in more general functions, such as impulse responses or forecast error variances, we suggest to parameterize A such that α is as low-dimensional as possible, for example, via the rotation matrix specification as in Example 2. In this way, the Bonferroni procedure of Algorithm 2 can be executed over the smallest possible grid for α, which reduces the computational burden.

APPENDIX SB: TECHNICAL DETAILS FOR THE MAIN PROOFS

Here, we establish some technical details utilized in the proofs in Section A of the main text.

SB.1 *Markov structure*

Define $Z_t := (Y_t', Y_{t-1}', \ldots, Y_{t-p+1}')'$, $\mathsf{C}_\theta := (c_\theta', 0', \ldots, 0')'$,

$$\mathsf{B}_\theta := \begin{bmatrix} B_{\theta,1} & B_{\theta,2} & \cdots & B_{\theta,p-1} & B_{\theta,p} \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}, \qquad \mathsf{D}_\theta := \begin{bmatrix} A_\theta^{-1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and note that we can write

$$Z_t = \mathsf{C}_\theta + \mathsf{B}_\theta Z_{t-1} + \mathsf{D}_\theta \epsilon_t. \tag{S3}$$

This can be rewritten in demeaned form as

$$\tilde{Z}_t = \mathsf{B}_\theta \tilde{Z}_{t-1} + \mathsf{D}_\theta \epsilon_t \tag{S4}$$

with $\tilde{Z}_t := Z_t - m_\theta$, for $m_\theta := (\sum_{i=0}^{\infty} \mathsf{B}_\theta)\mathsf{C}_\theta = (I - \mathsf{B}_\theta)^{-1}\mathsf{C}_\theta$.

LEMMA S1. *Suppose that Assumption* 1 *holds. Define $U_{\theta,t}$ as the* (*unique, strictly*) *stationary solution to* (S3). *Then $U_{\theta,t}$ has the representation*

$$U_{\theta,t} = m_\theta + \sum_{j=0}^{\infty} \mathsf{B}_\theta^j \mathsf{D}_\theta \epsilon_{t-j}, \quad m_\theta := (I - \mathsf{B}_\theta)^{-1}\mathsf{C}_\theta, \sum_{j=0}^{\infty} \left\| \mathsf{B}_\theta^j \right\| < \infty.$$

If ρ_θ is the largest absolute eigenvalue of the companion matrix B_θ and $\upsilon > 0$ is such that $\rho_\theta + \upsilon < 1$, then

$$\mathbb{E}\,\|U_{\theta,t} - m_\theta\|^\rho \le \frac{\mathbb{E}\,\|\mathsf{D}_\theta \epsilon_t\|^\rho}{1 - (\rho_\theta + \upsilon)^\rho}, \quad \rho \in [1, 4 + \delta].$$

PROOF. Rewriting (S3) as (S4) and applying Theorem 11.3.1 in Brockwell and Davis (1991) yields the first part. For the second part,

$$\|U_{\theta,t} - m_\theta\| \le \sum_{j=0}^{\infty} \left\| \mathsf{B}_\theta^j \right\| \|\mathsf{D}_\theta \epsilon_{t-j}\| \le \sum_{j=0}^{\infty} \|\mathsf{B}_\theta\|^j \|\mathsf{D}_\theta \epsilon_{t-j}\| \le \sum_{j=0}^{\infty} (\rho_\theta + \nu)^j \|\mathsf{D}_\theta \epsilon_{t-j}\|.$$

Since $\mathbb{E}\,\|\mathsf{D}_\theta \epsilon_{t-j}\|^\rho = \mathbb{E}\,\|\mathsf{D}_\theta \epsilon_t\|^\rho < \infty$ for all $t \in \mathbb{N}$, all $j \ge 0$ and $\rho \in [1, 4 + \delta]$, it follows that

$$\mathbb{E}\,\|U_{\theta,t} - m_\theta\|^\rho \le \sum_{j=0}^{\infty} (\rho_\theta + \nu)^{j\rho}\, \mathbb{E}\,\|\mathsf{D}_\theta \epsilon_{t-j}\|^\rho = \frac{\mathbb{E}\,\|\mathsf{D}_\theta \epsilon_t\|^\rho}{1 - (\rho_\theta + \nu)^\rho}.$$

□

LEMMA S2. *Let $Q_{n,\theta}$ be the probability measure corresponding to $\bar{q}_{n,\theta} := \frac{1}{n}\sum_{t=1}^{n} q_{\theta,t}$, where $q_{\theta,t}$ is the density of X_t under P_θ^n $(1 \le t \le n)$.*[S1] *Then $Q_{n,\theta} \xrightarrow{\text{TV}} Q_\theta$, where Q_θ is the distribution of the* (*unique, strictly*) *stationary solution to* (1).

[S1] Here, and throughout the Supplementary Material, any reference to the density of X_t is to be understood as to the density of the nondeterministic parts of X_t.

PROOF. By Lemma S1, (S4) has a (unique, strictly) stationary solution with finite second moments. Applying Theorem 2 in Saikkonen (2007) gives that the Markov chain $(\tilde{Z}_t)$ is V-geometrically ergodic with $\mathsf{V}(x) = 1 + \|x\|^2$. That is, for an invariant probability measure $\tilde{\pi}_\theta$, some $r \in (1, \infty)$ and some $R < \infty$,

$$\sum_{n=1}^{\infty} r^n \left\| \tilde{P}_\theta^n(\cdot, \tilde{z}) - \tilde{\pi}_\theta \right\|_{\mathrm{TV}} \leq \sum_{n=1}^{\infty} r^n \left\| \tilde{P}_\theta^n(\cdot, \tilde{z}) - \tilde{\pi}_\theta \right\|_{\mathsf{V}} \leq R\mathsf{V}(\tilde{z}) = R\left(\|\tilde{z}\|^2 + 1\right) < \infty, \tag{S5}$$

where $\tilde{P}_\theta^n(\cdot, \tilde{z})$ is the n-step transition probability and $\tilde{z}$ is the initial condition.[S2] $\tilde{\pi}_\theta$ is the distribution of $U_{\theta,t} - m_\theta$ as defined in Lemma S1 (Kallenberg (2021, Theorem 11.11)).

Let $f_\theta : \mathbb{R}^{Kp} \to \mathbb{R}^K$ be defined as

$$f_\theta(x) := \begin{bmatrix} I_K & 0_{K \times K(p-1)} \end{bmatrix} (x + m_\theta),$$

that is, the function which adds m_θ to its argument and then returns the first K elements. The distribution of X_t under P_θ^n (given the initial condition $\tilde{z}$) is then $Q_\theta^{t-1}(\cdot, \tilde{z}) = \tilde{P}_\theta^{t-1}(\cdot, \tilde{z}) \circ f_\theta^{-1}$, that is, the pushforward of $\tilde{P}_\theta^{t-1}(\cdot, \tilde{z})$ under f_θ. Henceforth, we shall omit the $\tilde{z}$ in the notation. Similarly, let $Q_\theta = \tilde{\pi}_\theta \circ f_\theta^{-1}$, that is, the pushforward of $\tilde{\pi}_\theta$ under f. That Q_θ is the distribution of the (unique, strictly) stationary solution to (1) can be seen by noting that the first K elements of $U_{\theta,t}$ form a (strictly) stationary time series and satisfy the defining equation (1); by Theorem 11.3.1 in Brockwell and Davis (1991) it is therefore the unique solution. Then by (S5),

$$\begin{aligned} \left\| \frac{1}{n} \sum_{t=1}^{n} Q_\theta^t - Q_\theta \right\|_{\mathrm{TV}} &\leq \frac{1}{n} \sum_{t=1}^{n} \left\| Q_\theta^t - Q_\theta \right\|_{\mathrm{TV}} \\ &\leq \frac{1}{n} \sum_{t=1}^{n} \left\| \tilde{P}_\theta^{t-1} - \tilde{\pi}_\theta \right\|_{\mathrm{TV}} \\ &\leq \frac{1}{n} \sum_{t=1}^{n} \left\| \tilde{P}_\theta^{t} - \tilde{\pi}_\theta \right\|_{\mathrm{TV}} + o(1) \\ &\to 0. \end{aligned}$$

□

SB.2 *Moment bounds*

LEMMA S3. *Suppose that Assumption* 1 *holds. Then for any sequence* $\theta_n = (\gamma + g_n/\sqrt{n}, \eta)$ *with* $g_n \to g \in \mathbb{R}^L$, *for some* $\rho > 0$, *under* $P_{\theta_n}^n$:

(i) $\sup_{n \in \mathbb{N}} \mathbb{E}[\|\dot{\ell}_{\theta_n}\|^{2+\rho}] < \infty$;

(ii) $\sup_{n \in \mathbb{N}} \mathbb{E}[\|\tilde{\ell}_{\theta_n}\|^{2+\rho}] < \infty$.

[S2] The norm $\|\nu\|_{\mathsf{V}}$ is defined by $\|\nu\|_{\mathsf{V}} := \sup_{f \leq \mathsf{V}} |\int f \, \mathrm{d}\nu|$ where the supremum is taken over all measurable functions dominated by V for any probability measure ν.

PROOF. Since the deterministic terms in $\dot{\ell}_{\theta_n}$ and $\tilde{\ell}_{\theta_n}$ are either constants or continuous functions of γ (by Assumption 1(iii)), they are uniformly bounded, since $\{\gamma + g_n/\sqrt{n} : n \in \mathbb{N}\} \cup \{\gamma\}$ is compact. It is therefore sufficient to show that under $P^n_{\theta_n}$, each of

$$\sup_{n\in\mathbb{N},1\le t\le n} \mathbb{E}\big[\big|A(\theta_n)_{k\bullet}V_{\theta_n,t}\big|^{4+\delta}\big], \qquad \sup_{n\in\mathbb{N},1\le t\le n} \mathbb{E}\big[\big|\phi_k\big(A(\theta_n)_{k\bullet}V_{\theta_n,t}\big)\big|^{4+\delta}\big],$$
$$\sup_{n\in\mathbb{N},1\le t\le n} \mathbb{E}\big[\|X_t\|^{4+\delta}\big],$$

is finite. Since under $P^n_{\theta_n}$, each $A(\theta_n)_{k\bullet}V_{\theta_n,t} \sim \eta_k$, finiteness of the first two follow directly from Assumption 1(ii) For the third, recurse equation (S3) backwards under $\theta = \theta_n$, to obtain

$$Z_t = \sum_{j=0}^{t-1} \mathsf{B}^j_{\theta_n}\mathsf{C}_{\theta_n} + \sum_{j=0}^{t-1} \mathsf{B}^j_{\theta_n}\mathsf{D}_{\theta_n}\epsilon_{t-j} + \mathsf{B}^t_{\theta_n}Z_0.$$

Each of B_θ, C_θ, D_θ (depend on θ only through γ and) are continuous functions of γ, hence

$$\varrho := \sup_{n\in\mathbb{N}} \|\mathsf{B}_{\theta_n}\|_2 < 1, \qquad \sup_{n\in\mathbb{N}} \|\mathsf{C}_{\theta_n}\|_2 < C_1, \qquad \sup_{n\in\mathbb{N}} \|\mathsf{D}_{\theta_n}\|_2 < C_2,$$

where the first is due to Assumption 1(i). Since we condition on Z_0, by Assumption 1(ii),

$$\mathbb{E}\,\|Z_t\|^{4+\delta} \lesssim \left(\frac{C_1}{1-\varrho}\right)^{4+\delta} + \left(\frac{C_2}{1-\varrho}\right)^{4+\delta} \mathbb{E}\,|\epsilon_1|^{4+\delta} + \|Z_0\|^{4+\delta} < \infty. \tag{S6}$$

As the bound on the right-hand side is independent of t or n, the claim follows. □

LEMMA S4. *Let $W_{n,t}$ be as in the Proof of Proposition* 3 *and suppose the conditions of that Proposition hold. Then $P^n_\theta[|\sqrt{n}W_{n,t}|^{2+\rho}]$ is uniformly bounded for some $\rho > 0$. In consequence, under P^n_θ, $W_{n,t}$ satisfies*

$$\lim_{n\to\infty} \sum_{t=1}^{n} \mathbb{E}\big[W^2_{n,t}\mathbf{1}\{|\sqrt{n}W_{n,t}| > \varepsilon\sqrt{n}\}\big] = 0, \quad \textit{for any } \varepsilon > 0. \tag{S7}$$

PROOF. Uniform boundedness of $P^n_\theta[|\sqrt{n}W_{n,t}|^{2+\rho}]$ implies

$$\lim_{n\to\infty} \sum_{t=1}^{n} W^{2+\rho}_{n,t} = 0,$$

which in turns implies (S7) (cf. Billingsley (1995, p. 362)). For the uniform boundedness, as

$$2\sqrt{n}W_{n,t} = g'\dot{\ell}_\theta(Y_t, X_t) + \sum_{k=1}^{K} h_k\big(A_{k\bullet}(\alpha,\sigma)V_{\theta,t}\big),$$

and the h_k are bounded, it suffices to note that by Lemma S3 $\mathbb{E}[(g'\dot{\ell}_\theta(X_t, Y_t))^{2+\rho}] \le C$ under P^n_θ for some $\rho > 0$. □

SB.3 *Log-likelihood ratios*

LEMMA S5 (DQM). *Suppose that Assumption* 1 *holds. Then with $W_{n,t}$ and $U_{n,t}$ defined as in the proof of Proposition* 3,

$$\lim_{n\to\infty} \mathbb{E}\sum_{t=1}^{n}(W_{n,t} - U_{n,t})^2 = 0,$$

where the expectation is taken under P_θ^n.

PROOF. We argue similar to Lemma 7.6 in van der Vaart (1998). Let $V_{\theta,t} := Y_t - BX_t$ and $\varphi(v) = (g, \eta_1 h_1, \ldots, \eta_K h_K)$ for $v = (g, h)$ with $g \in \mathbb{R}^L$, $h \in \dot{\mathscr{H}}$. Let

$$p_\theta(Y_t, X_t) := \big|A(\theta)\big| \prod_{k=1}^{K} \eta_k\big(A_{k\bullet}(\theta)V_{\theta,t}\big),$$

$$s_{\theta,u}(Y_t, X_t) := g'\dot{\ell}_{\theta+u\varphi(v)}(Y_t, X_t) + \sum_{k=1}^{K} \frac{h_k\big(A_{k\bullet}\big(\theta+u\varphi(v)\big)V_{\theta+u\varphi(v),t}\big)}{1+uh_k\big(A_{k\bullet}\big(\theta+u\varphi(v)\big)V_{\theta+u\varphi(v),t}\big)}$$
$$+ \sum_{k=1}^{K} \frac{uh_k'\big(A_{k\bullet}\big(\theta+u\varphi(v)\big)V_{\theta+u\varphi(v),t}\big)[\mathsf{D}_{1,k,u}V_{\theta+u\varphi(v),t} + \mathsf{D}_{2,k,u}X_t]}{1+uh_k\big(A_{k\bullet}\big(\theta+u\varphi(v)\big)V_{\theta+u\varphi(v),t}\big)},$$

with

$$\mathsf{D}_{1,k,u} := e_k' \sum_{l=1}^{L_\alpha} g_{\alpha,l} D_{\alpha,l}\big(\theta+u\varphi(v)\big) + e_k' \sum_{l=1}^{L_\sigma} g_{\sigma,l} D_{\sigma,l}\big(\theta+u\varphi(v)\big),$$

$$\mathsf{D}_{2,k,u} := -A_{k\bullet}\big(\theta+u\varphi(v)\big) \sum_{l=1}^{L_b} D_{b,l}\big(\theta+u\varphi(v)\big).$$

By Assumption 1 and standard computations, the derivative of $u \mapsto \sqrt{p_{\theta+u\varphi(v)}}$ at $u = \mathsf{u}$ is $\frac{1}{2}s_{\theta,\mathsf{u}}\sqrt{p_{\theta+\mathsf{u}\varphi(v)}}$ (everywhere). Inspection reveals that this is continuous in u.

For $q_{\theta,t}$, the density of X_t under P_θ^n and $s_\theta := s_{\theta,0}$,

$$\mathbb{E}\sum_{t=1}^{n}(W_{n,t} - U_{n,t})^2 = \frac{1}{n}\sum_{t=1}^{n}\int\left(\sqrt{n}\left[\sqrt{\frac{p_{\theta_n}}{p_\theta}} - 1\right] - \frac{1}{2}s_\theta\right)^2 p_\theta q_{\theta,t}\,\mathrm{d}\lambda$$
$$= \int\left(\sqrt{n}[\sqrt{p_{\theta_n}} - \sqrt{p_\theta}] - \frac{1}{2}s_\theta\sqrt{p_\theta}\right)^2 \bar{q}_{n,\theta}\,\mathrm{d}\lambda,$$

with $\bar{q}_{n,\theta} := \frac{1}{n}\sum_{t=1}^{n} q_{\theta,t}$. The integrand converges to zero as $n \to \infty$ by the differentiability of $u \mapsto \sqrt{p_{\theta+u\varphi(v)}}$ at $u = 0$.[S3] Let

$$I_{\theta,u,n} := \int s_{\theta,u}^2 p_{\theta+u\varphi(v)}\bar{q}_{n,\theta}\,\mathrm{d}\lambda = \int s_{\theta,u}^2\,\mathrm{d}G_{\theta,u,n},$$

[S3] Note that $p_{\theta_n} = p_{\theta_n(g,h)} = p_{\theta+\varphi(v)/\sqrt{n}}$.

where $G_{\theta,u,n}$ is the distribution of (Y_t, X_t) corresponding to the density $p_{\theta+u\varphi(v)}\bar{q}_{n,\theta}$. By Lemma S22, $G_{\theta,u/\sqrt{n},n} \xrightarrow{\text{TV}} G_\theta$, defined by

$$G_\theta(A) := \int_A p_\theta \,\mathrm{d}\big(\lambda(y) \otimes Q_\theta(x)\big).$$

For any $(u_n) \subset [0,1]$, we have that $s^2_{\theta,u_n/\sqrt{n}} \to s^2_\theta$ (pointwise). By Lemma S6 and Corollary 2.9 in Feinberg, Kasyanov, and Zgurovsky (2016), $\lim_{n\to\infty} I_{\theta,u_n/\sqrt{n},n} = \int s^2_\theta \,\mathrm{d}G_\theta < \infty$, and hence

$$\left|\int_0^1 I_{\theta,u/\sqrt{n},n}\,\mathrm{d}u - \int_0^1 \int s^2_\theta \,\mathrm{d}G_\theta \,\mathrm{d}u\right| \le \sup_{u\in[0,1]} \left|I_{\theta,u/\sqrt{n},n} - \int s^2_\theta \,\mathrm{d}G_\theta\right| \to 0.$$

By absolute continuity, Jensen's inequality and the Fubini–Tonelli theorem,

$$\begin{aligned}\int \big(\sqrt{n}[\sqrt{p_{\theta_n}} - \sqrt{p_\theta}]\big)^2 \bar{q}_{n,\theta}\,\mathrm{d}\lambda &\le \frac{1}{4}\int\int_0^1 (s_{\theta,u/\sqrt{n}}\sqrt{p_{\theta+u\varphi(v)/\sqrt{n}}})^2 \bar{q}_{n,\theta}\,\mathrm{d}u\,\mathrm{d}\lambda \\ &\le \int_0^1 I_{\theta,u/\sqrt{n},n}\,\mathrm{d}u.\end{aligned}$$

Combine these observations with Proposition 2.29 in van der Vaart (1998). □

LEMMA S6. *Suppose that Assumption* 1 *holds. Let $s_{\theta,u}$ and $G_{\theta,u,n}$ be as in the proof of Lemma* S5. *Then for any $(u_n)_{n\in\mathbb{N}} \subset [0,1]$, $s^2_{\theta,u_n/\sqrt{n}}$ is asymptotically uniformly $G_{\theta,u_n/\sqrt{n},n}$-integrable and $s_\theta \in L_2(G_\theta)$.*

PROOF. That $s_\theta \in L_2(G_\theta)$ follows from the moment bounds in Assumption 1.2, the boundedness of the h_k, the form of $\dot{\ell}_\theta$ given in equations (7)–(9) and Lemma S1 given that Q_θ is the law of the stationary solution to (1).

For the uniform integrability, let $\vartheta_n := \theta + u_n\varphi(v)/\sqrt{n} \to \theta$ and

$$\begin{aligned}s_{\vartheta_n,1}(Y_t, X_t) &:= g'\dot{\ell}_{\vartheta_n}(Y_t, X_t), \\ s_{\vartheta_n,2}(Y_t, X_t) &:= \sum_{k=1}^K \frac{h_k\big(A_{k\bullet}(\vartheta_n)V_{\vartheta_n,t}\big)}{1 + u_n h_k\big(A_{k\bullet}(\vartheta_n)V_{\vartheta_n,t}\big)/\sqrt{n}}, \\ s_{\vartheta_n,3}(Y_t, X_t) &:= \sum_{k=1}^K \frac{u_n h_k'\big(A_{k\bullet}(\vartheta_n)V_{\vartheta_n,t}\big)[\mathsf{D}_{1,k,u_n/\sqrt{n}}V_{\vartheta_n,t} + \mathsf{D}_{2,k,u_n/\sqrt{n}}X_t]/\sqrt{n}}{1 + u_n h_k\big(A_{k\bullet}(\vartheta_n)V_{\vartheta_n,t}\big)/\sqrt{n}}.\end{aligned}$$

It suffices to show that under $G_{\theta,u_n/\sqrt{n},n}$ each $s_{\vartheta_n,i}$ $(i = 1,2,3)$ has uniformly bounded $2+\rho$ moments for some $\rho > 0$ for all sufficiently large n.

We start with $s_{\vartheta_n,2}$: since each h_k is bounded, for all large enough n, each numerator is uniformly bounded above and each denominator is uniformly bounded below, away from zero. Thus, there is a M such that $|s_{\vartheta_n,2}(Y_t, X_t)| \le M$ for all such n.

For $s_{\vartheta_n,3}$, by Assumption 1 part 3, each $\mathsf{D}_{1,k,u_n/\sqrt{n}}$ and $\mathsf{D}_{2,k,u_n/\sqrt{n}}$ are uniformly bounded for all large enough n; the same is true of $\|A(\vartheta_n)^{-1}\|_2$. Using this, the fact that

$V_{\vartheta_n,t} = A(\vartheta_n)^{-1}\epsilon_t$ and arguing similar to as in the preceding paragraph we have that for some M and all large enough n, $|s_{\vartheta_n,3}(Y_t, X_t)| \leq M[\|\epsilon_t\| + \|X_t\|]$. Thus, it is enough to verify that

$$\sup_{n\geq N, 1\leq t\leq n} G_{\theta,u_n/\sqrt{n},n}\|\epsilon_t\|^{4+\delta} < \infty, \qquad \sup_{n\geq N, 1\leq t\leq n} G_{\theta,u_n/\sqrt{n},n}\|X_t\|^{4+\delta} < \infty. \tag{S8}$$

Under $G_{\theta,u_n/\sqrt{n},n}$, the elements $\epsilon_{t,k}$ are (independently across k) distributed according to $\eta_k(1 + u_n h_k/\sqrt{n})$, so there are $c, C < \infty$ such that

$$\begin{aligned} G_{\theta,u_n/\sqrt{n},n}\|\epsilon_t\|^{4+\delta} &\leq G_{\theta,u_n/\sqrt{n},n}\left[\sum_{k=1}^{K}\epsilon_{t,k}^2\right]^{\frac{4+\delta}{2}} \\ &\leq c\sum_{k=1}^{K}\left[\left(1+\frac{\bar{h}_k}{\sqrt{n}}\right)\int |x_k|^{4+\delta}\eta_k(x_k)\,\mathrm{d}x_k\right] \leq C, \end{aligned}$$

where $|h_k(x)| \leq \bar{h}_k$. By arguing analogously to as in in Lemma S3, one has (cf. (S6))

$$G_{\theta,u_n/\sqrt{n},n}\|Z_t\|^{4+\delta} \lesssim \left(\frac{C_1}{1-\varrho}\right)^{4+\delta} + \left(\frac{C_2}{1-\varrho}\right)^{4+\delta} G_{\theta,u_n/\sqrt{n},n}|\epsilon_1|^{4+\delta} + \|Z_0\|^{4+\delta},$$

which is uniformly bounded given the penultimate display.

Finally, consider $s_{\vartheta_n,1}$. It suffices to show that each component of $\dot{\ell}_{\vartheta_n}$ has $4+\delta$ moment bounded uniformly for all $n \geq N$.[S4] By Assumption 1.3, by increasing N if necessary, $\sup_{\vartheta\in\mathsf{T}}|\zeta^x_{l,k,j}(\vartheta)| \leq M$ for all l, k, j and $x \in \alpha, \sigma$ and likewise $\sup_{\vartheta\in\mathsf{T}}\|A_{k\bullet}(\vartheta) \times D_{b_l}(\vartheta)\| \leq M$. Recall that $V_{\vartheta_n,t} = A(\vartheta_n)^{-1}\epsilon_t$. Given (S8) and the observations in footnote S4 to complete the proof, it suffices to note that (for $\phi_k = \frac{\mathrm{d}\log\eta_k(x)}{\mathrm{d}x}$) and some $C < \infty$,

$$G_{\theta,u_n/\sqrt{n},n}|\phi_k|^{4+\delta} \leq \left(1+\frac{\bar{h}_k}{\sqrt{n}}\right)\int |\phi(x)|^{4+\delta}\eta_k(x)\,\mathrm{d}x \leq C. \qquad \square$$

LEMMA S7. *Let $W_{n,t}$ be as in the proof of Proposition* 3 *and suppose the conditions of that Proposition hold. Let G_θ be defined as in the proof of Lemma* S5*. Then, under P^n_θ,*

$$\lim_{n\to\infty}\mathbb{E}\left|\sum_{t=1}^{n}W_{n,t}^2 - \frac{\tau^2}{4}\right| = 0, \quad \textit{with } \tau^2 := G_\theta\left(g'\dot{\ell}_\theta(Y,X) + \sum_{k=1}^{K}h_k\big(A_{k\bullet}(\theta)V_\theta\big)\right)^2.$$

[S4] The form each such component is that given in equations (7)–(9). Note here that each ϕ_k is (implicitly) a function of η_k, and thus when evaluating equations (7)–(9) at ϑ_n, the ϕ_k that appear are $\phi_{k,u_n,n}$, defined as

$$\phi_{k,u,n} := \frac{\mathrm{d}\big(\log\eta_k(x) + \log\big(1+uh_k(x)/\sqrt{n}\big)\big)}{\mathrm{d}x} = \phi_k + \frac{uh_k'/\sqrt{n}}{1+uh_k/\sqrt{n}}.$$

Since each h_k, and h_k' are bounded, increasing N if necessary, one has for $n \geq N$,

$$|\phi_{k,u_n,n}| \leq |\phi_k| + M.$$

PROOF. Define

$$r_\theta(X_t) := \mathbb{E}\big[s_\theta(Y_t, X_t)^2 | X_t\big], \qquad s_\theta(Y, X) := g' \dot{\ell}_\theta(Y, X) + \sum_{k=1}^{K} h_k\big(A_{k\bullet}(\theta) V_\theta\big),$$

where the conditional expectation is taken under P_θ^n. Since conditional expectations are L_1 contractions, by Lemma S4, we have that $P_\theta^n[|r_\theta(X_t)|^{1+\rho/2}] \lesssim C < \infty$, and hence $(|r_\theta(X_t)|^{1+\rho/2})_{t\in\mathbb{N}}$ is uniformly P_θ^n–integrable. Moreover, we have for $\mathscr{F}_t := \sigma(\epsilon_1, \ldots, \epsilon_t)$,

$$r_\theta(X_t) = \mathbb{E}\big[s_\theta(Y_t, X_t)^2 | X_t\big] = \mathbb{E}\big[s_\theta(Y_t, X_t)^2 | \mathscr{F}_{t-1}\big],$$

as is clear from the definition of s_θ.[S5] Hence, $(s_\theta(Y_t, X_t)^2 - r_\theta(X_t), \mathscr{F}_t)$ is a martingale difference squence and by Theorem 19.7 in Davidson (1994),

$$\lim_{n\to\infty} \mathbb{E}\left|\frac{1}{n}\sum_{t=1}^{n}\big[s_\theta(Y_t, X_t)^2 - r_\theta(X_t)\big]\right|^{1+\rho/2} = 0.$$

Now define $u_\theta(X_t) := r_\theta(X_t) - \mathbb{E}[r_\theta(X_t)]$, which satisfies $P_\theta^n[|u_\theta(X_t)|^{1+\rho/2}] \lesssim C < \infty$ and is evidently mean zero. By Theorem 3 in Saikkonen (2007), Z_t, and hence $u_\theta(X_t)$ (e.g., Davidson (1994, Theorem 14.1)) has geometrically decaying β-mixing coefficients. Therefore, by Theorem 14.2 in Davidson (1994), $(u_\theta(X_t)/n)_{n\in\mathbb{N}, 1\le t\le n}$ is an L_1-mixingale array with respect to the filtration formed by $\mathsf{F}_{n,t} := \sigma(X_1, \ldots, X_t)$ relative to the sequence of positive constants

$$n^{-1} \le c_{n,t} = \max\Big\{1/n, \big(P_\theta^n\big[\big|u_\theta(X_t)/n\big|^{1+\rho/2}\big]\big)^{1/(1+\rho/2)}\Big\} \le n^{-1}\max\{C, 1\}.$$

By Theorem 19.11 in Davidson (1994),

$$\lim_{n\to\infty} \mathbb{E}\left|\frac{1}{n}\sum_{t=1}^{n} u_\theta(Y_t, X_t)\right| = 0.$$

It remains to show that $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[r_\theta(X_t)] \to \tau^2$. Since $\mathbb{E}[r_\theta(X_t)] = \mathbb{E}[s_\theta(Y_t, X_t)]$,

$$\tau_n^2 := G_{\theta,0,n}\big[s_\theta(Y, X)^2\big] = \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\, s_\theta(Y_t, X_t)^2 = \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\big[r_\theta(X_t)\big],$$

where $G_{\theta,0,n}$ is as defined in the proof of Lemma S5. That $\mathbb{E}\,\frac{1}{n}\sum_{t=1}^{n} s_\theta(Y_t, X_t)^2 \lesssim C$ follows from Lemma S4. Therefore, by Lemma S6, $s_\theta(Y, X)^2$ is uniformly $G_{\theta,0,n}$–integrable and also $\tau^2 < \infty$. Then, by Corollary 2.9 in Feinberg, Kasyanov, and Zgurovsky (2016) and Lemma S22, $\tau_n^2 \to \tau$. □

LEMMA S8. *In the setting of Proposition* 4,

$$\log \frac{p^n_{\theta_n(g_n,h)}}{p^n_{\theta_n(g,h)}} = o_{P^n_{\theta_n(g,h)}}(1).$$

[S5] See, for example, Theorem 7.3.1 in Chow and Teicher (1997) for the (almost sure) equality of the conditional expectations.

PROOF. Since by Proposition 3 and Example 6.5 in van der Vaart (1998) $P^n_{\theta_n(g,h)} \triangleleft\triangleright P^n_\theta$ it suffices to show that the left-hand side is $o_{P^n_\theta}(1)$. We first show that

$$\log \frac{p^n_{\theta_n(g_n,0)}}{p^n_\theta} = \frac{1}{\sqrt{n}} \sum_{t=1}^n g' \dot{\ell}_\theta(Y_t, X_t) - \mathbb{E}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^n g' \dot{\ell}_\theta(Y_t, X_t) \right)^2 + o_{P^n_\theta}(1),$$

$$\log \frac{p^n_{\theta_n(g,0)}}{p^n_\theta} = \frac{1}{\sqrt{n}} \sum_{t=1}^n g' \dot{\ell}_\theta(Y_t, X_t) - \mathbb{E}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^n g' \dot{\ell}_\theta(Y_t, X_t) \right)^2 + o_{P^n_\theta}(1).$$

For these log-likelihood expansions, we may appeal to Lemma 1 in Swensen (1985). The required Conditions (1.3)–(1.7) and (iii) of his Theorem 1 are all established in the proof of Proposition 3 (take each $h_k = 0$). It remains to show condition (1.2) for each of the cases in the above display. In particular, set

$$W_{n,t} := \frac{1}{2\sqrt{n}} g' \dot{\ell}_\theta(Y_t, X_t)$$

and (cf. equations (37), (38))

$$U_{n,t} := \left[\left(\frac{|A(\theta_n(g_n, h))|}{|A(\theta)|} \right) \times \prod_{k=1}^K \frac{\eta_k\big(A_{k\bullet}(\theta_n(g_n, h)) V_{\theta_n(g_n,h),t}\big)}{\eta_k\big(A_{k\bullet}(\theta) V_{\theta,t}\big)} \right]^{1/2} - 1,$$

where we note that $A(\theta) = A(\theta_n(0, h))$ and $V_\theta = V_{\theta_n(0,h)}$. We verify (1.2), that is, that

$$\lim_{n\to\infty} \mathbb{E}\left[\sum_{t=1}^n (W_{n,t} - U_{n,t})^2 \right] = 0,$$

under P^n_θ.[S6] The argument now follows similar to that in Lemma S5. To simplify the notation, let $p_\gamma := p_{(\gamma,\eta)}$ and $\dot{\ell}_\gamma := \dot{\ell}_{(\gamma,\eta)}$ where $\eta = (\eta_1, \ldots, \eta_K)$ will remain fixed. By Assumption 1 and standard computations, the derivative of $\gamma \mapsto \sqrt{p_\gamma}$ is $\frac{1}{2}\dot{\ell}_\gamma \sqrt{p_\gamma}$ (everywhere). Inspection reveals that this is continuous in γ.

Let $\gamma_n := \gamma + g_n/\sqrt{n}$. For $q_{\theta,t}$, the density of X_t under P^n_θ,

$$\begin{aligned}
\mathbb{E} \sum_{t=1}^n (W_{n,t} - U_{n,t})^2 &= \frac{1}{n} \sum_{t=1}^n \int \left(\sqrt{n} \left[\sqrt{\frac{p_{\gamma_n}}{p_\gamma}} - 1 \right] - \frac{1}{2} g' \dot{\ell}_\gamma \right)^2 p_\gamma q_{\theta,t} \, \mathrm{d}\lambda \\
&= \int \left(\sqrt{n} [\sqrt{p_{\gamma_n}} - \sqrt{p_\gamma}] - \frac{1}{2} g' \dot{\ell}_\gamma \sqrt{p_\gamma} \right)^2 \bar{q}_{n,\theta} \, \mathrm{d}\lambda,
\end{aligned}$$

with $\bar{q}_{n,\theta} := \frac{1}{n} \sum_{t=1}^n q_{\theta,t}$. The term inside the parentheses converges to zero as $n \to \infty$ by the differentiability of $\gamma \mapsto \sqrt{p_\gamma}$ and that $(g_n - g)' \dot{\ell}_\gamma \sqrt{p_\gamma} \to 0$ pointwise. Let

$$I_{\theta,u,n} := \int \left(g' \dot{\ell}_{\gamma + u g_n} \right)^2 p_{\gamma + u g_n} \bar{q}_{n,\theta} \, \mathrm{d}\lambda = \int \left(g' \dot{\ell}_{\gamma + u g_n} \right)^2 \mathrm{d}G_{\theta,u,n},$$

[S6] This suffices as the second expansion is just the special case $g_n = g$ for each $n \in \mathbb{N}$.

where $G_{\theta,u,n}$ is the distribution of (Y_t, X_t) corresponding to the density $p_{\gamma+ug_n}\bar{q}_{n,\theta}$. By Lemma S22, $G_{\theta,u_n/\sqrt{n},n} \xrightarrow{\text{TV}} G_\theta$, defined as in the proof of Lemma S5. For any $(u_n) \subset [0,1]$, we have that $(g'\dot{\ell}_{\gamma+u_n g_n/\sqrt{n}})^2 \to (g'\dot{\ell}_\gamma)^2$ (pointwise). Each component of $\dot{\ell}_\gamma \in L_2(G_\theta)$ by Lemma S6 and, moreover, $\sup_{n\geq N} G_{\theta,u_n/\sqrt{n},n}\|\dot{\ell}_{\gamma+u_n g_n/\sqrt{n}}\|^{2+\rho} \leq C$ for some $\rho > 0$.[S7] Therefore, by Corollary 2.9 in Feinberg, Kasyanov, and Zgurovsky (2016), $\lim_{n\to\infty} I_{\theta,u_n/\sqrt{n},n} = \int (g'\dot{\ell}_\gamma)^2 \,\mathrm{d}G_\theta < \infty$, and hence

$$\left|\int_0^1 I_{\theta,u/\sqrt{n},n}\,\mathrm{d}u - \int_0^1 \int s_\theta^2 \,\mathrm{d}G_\theta \,\mathrm{d}u\right| \leq \sup_{u\in[0,1]}\left|I_{\theta,u/\sqrt{n},n} - \int \left(g'\dot{\ell}_\gamma\right)^2 \mathrm{d}G_\theta\right| \to 0.$$

By the continuous differentiability of $\sqrt{p_\gamma}$, Jensen's inequality and the Fubini–Tonelli theorem,

$$\begin{aligned}\int \left(\sqrt{n}[\sqrt{p_{\gamma_n}} - \sqrt{p_\gamma}]\right)^2 \bar{q}_{n,\theta}\,\mathrm{d}\lambda &\leq \frac{1}{4}\int\int_0^1 \left((g'\dot{\ell}_{\gamma+ug_n/\sqrt{n}})\sqrt{p_{\gamma+ug_n/\sqrt{n}}}\right)^2 \bar{q}_{n,\theta}\,\mathrm{d}u\,\mathrm{d}\lambda \\ &\leq \int_0^1 I_{\theta,u/\sqrt{n},n}\,\mathrm{d}u.\end{aligned}$$

Combining these observations with Proposition 2.29 in van der Vaart (1998) verifies (1.2), and hence the claimed log-likelihood expansions follow from Lemma 1 in Swensen (1985).

To complete the proof, set

$$\tilde{u}_{k,n,t} := A_{k\bullet}\big(\theta_n(g_n,h)\big)V_{\theta_n(g_n,h),t}, \qquad u_{k,n,t} := A_{k\bullet}\big(\theta_n(g,h)\big)V_{\theta_n(g,h),t},$$

and observe that

$$\begin{aligned}&\log\frac{p^n_{\theta_n(g_n,h)}}{p^n_{\theta_n(g,h)}} - \left[\log\frac{p^n_{\theta_n(g_n,0)}}{p^n_\theta} - \log\frac{p^n_{\theta_n(g,0)}}{p^n_\theta}\right] \\ &\quad= \sum_{k=1}^K\sum_{i=1}^n \log\left(1 + \frac{h_k(\tilde{u}_{k,n,t})}{\sqrt{n}}\right) - \log\left(1 + \frac{h_k(u_{k,n,t})}{\sqrt{n}}\right),\end{aligned}$$

where the bracketed term is $o_{P^n_\theta}(1)$ by the preceding argument. Hence, it suffices to show that an arbitrary kth element of the outer sum on the right-hand side is also $o_{P^n_\theta}(1)$. Let $\varepsilon \in (0,1)$ be fixed and define

$$\begin{aligned}E_n &:= \left\{\max_{1\leq i\leq n}\left|h_k(\tilde{u}_{k,n,t})\right|/\sqrt{n} \leq \varepsilon\right\}, \\ F_n &:= \left\{\max_{1\leq i\leq n}\left|h_k(u_{k,n,t})\right|/\sqrt{n} \leq \varepsilon\right\}.\end{aligned}$$

[S7] This follows from (a) the continuity requirements in Assumption 1.3, (b) under $G_{\theta,u_n/\sqrt{n},n}$ we have that $e_k' A(\theta_n(u_n g_n,0))^{-1}V_{\theta_n(u_n g_n,0)} = \epsilon_k \sim \eta_k$ and (c) $\sup_{n\geq N,1\leq t\leq n} G_{\theta,u_n/\sqrt{n},n}\|X_t\|^{4+\delta} < \infty$, which can be shown by an argument analogous to that which is established in the proof of Lemma S6.

Since h_k is bounded, $P_\theta^n(E_n \cap F_n) \to 1$. On this set, we may perform a two-term Taylor expansion of $\log(1+x)$ to obtain

$$\begin{aligned}
&\log\left(1+\frac{h_k(\tilde{u}_{k,n,t})}{\sqrt{n}}\right) - \log\left(1+\frac{h_k(u_{k,n,t})}{\sqrt{n}}\right) \\
&\quad = \frac{h_k(\tilde{u}_{k,n,t}) - h_k(u_{k,n,t})}{\sqrt{n}} - \frac{1}{2}\frac{h_k(\tilde{u}_{k,n,t})^2 - h_k(u_{k,n,t})^2}{n} \\
&\qquad + R\left(\frac{h_k(\tilde{u}_{k,n,t})}{\sqrt{n}}\right) - R\left(\frac{h_k(u_{k,n,t})}{\sqrt{n}}\right),
\end{aligned}$$

where $|R(x)| \le |x|^3$. For the remainder terms, one has for any u_i,

$$\sum_{i=1}^n \left|R\left(\frac{h_k(u_i)}{\sqrt{n}}\right)\right| \le \max_{1\le i\le n} \frac{h_k(u_i)}{\sqrt{n}} \frac{1}{n}\sum_{i=1}^n h_k(u_i)^2 \lesssim \frac{1}{\sqrt{n}},$$

since h_k is bounded. For the first term in Taylor expansion, note that the derivative (in θ, σ) of $A(\theta,\sigma)$ is bounded on a neighborhood of (θ,σ) (by Assumption 1). Combine this with the boundedness of h_k' and the mean value theorem to conclude that

$$\left|h_k(\tilde{u}_{k,n,t}) - h_k(u_{k,n,t})\right| \lesssim n^{-1/2}\|g_n - g\|\big[\|\epsilon_t\| + \|X_t\|\big].$$

Using this, since h_k is bounded,

$$\left|h_k(\tilde{u}_{k,n,t})^2 - h_k(u_{k,n,t})^2\right| \lesssim n^{-1/2}\|g_n - g\|\big[\|\epsilon_t\| + \|X_t\|\big].$$

Therefore, using (S6) and Assumption 1.2,

$$\begin{aligned}
&\sum_{i=1}^n \left|\frac{h_k(\tilde{u}_{k,n,t}) - h_k(u_{k,n,t})}{\sqrt{n}} - \frac{1}{2}\frac{h_k(\tilde{u}_{k,n,t})^2 - h_k(u_{k,n,t})^2}{n}\right| \\
&\quad \lesssim \|g_n - g\|\left(1+\frac{1}{\sqrt{n}}\right)\frac{1}{n}\sum_{i=1}^n\big[\|\epsilon_t\| + \|X_t\|\big] = o_{P_\gamma^n}(1).
\end{aligned}$$

□

LEMMA S9. *In the setting of Proposition* 4,

$$\log\frac{p^n_{\theta_n(g_n,h_n)}}{p^n_{\theta_n(g_n,h)}} = o_{P^n_{\theta_n(g_n,h)}}(1).$$

PROOF. For notational ease, set

$$u_{k,n,t} := e_k' A\big(\theta_n(g_n,h)\big)V_{\theta_n(g_n,h),t} = e_k' A\big(\theta_n(g_n,h_n)\big)V_{\theta_n(g_n,h_n),t}.$$

One has that

$$\log\frac{p^n_{\theta_n(g_n,h_n)}}{p^n_{\theta_n(g_n,h)}} = \sum_{k=1}^K\sum_{t=1}^n \log\big(1+h_{k,n}(u_{k,n,t})/\sqrt{n}\big) - \log\big(1+h_k(u_{k,n,t})/\sqrt{n}\big),$$

hence it suffices to show that each

$$l_{n,k} := \sum_{t=1}^{n} \log\big(1 + h_{k,n}(u_{k,n,t})/\sqrt{n}\big) - \log\big(1 + h_k(u_{k,n,t})/\sqrt{n}\big) \xrightarrow{P^n_{\theta_n(g_n,h)}} 0.$$

Let $\varepsilon \in (0, 1)$ be fixed and define

$$E_n := \Big\{ \max_{1\le t\le n} \big|h_{k,n}(u_{k,n,t})\big|/\sqrt{n} \le \varepsilon \Big\};$$

$$F_n := \Big\{ \max_{1\le t\le n} \big|h_{k}(u_{k,n,t})\big|/\sqrt{n} \le \varepsilon \Big\}.$$

Since h_k is bounded, $P^n_{\theta_n(g_n,h)} F_n \to 1$; $P^n_{\theta_n(g_n,h)} E_n \to 1$ follows from Lemma S11. Hence, $P^n_{\theta_n(g_n,h)} F_n \cap E_n \to 1$. On $E_n \cap F_n$, we can perform a two-term Taylor expansion of $\log(1 + x)$ to obtain

$$\begin{aligned}
&\log\big(1 + h_{k,n}(u_{k,n,t})/\sqrt{n}\big) - \log\big(1 + h_k(u_{k,n,t})/\sqrt{n}\big) \\
&\quad = \frac{h_{k,n}(u_{k,n,t})}{\sqrt{n}} - \frac{1}{2}\frac{h_{k,n}(u_{k,n,t})^2}{n} - \frac{h_k(u_{k,n,t})}{\sqrt{n}} + \frac{1}{2}\frac{h_k(u_{k,n,t})^2}{n} \\
&\qquad + R\bigg(\frac{h_{k,n}(u_{k,n,t})}{\sqrt{n}}\bigg) - R\bigg(\frac{h_k(u_{k,n,t})}{\sqrt{n}}\bigg),
\end{aligned}$$

where $|R(x)| \le |x|^3$. It follows that

$$\begin{aligned}
l_{n,k} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h_{k,n}(u_{k,n,t}) - h_k(u_{k,n,t}) - \frac{1}{2}\frac{1}{n}\sum_{t=1}^{n}\big[h_{k,n}(u_{k,n,t})^2 - h_k(u_{k,n,t})^2\big] \\
+ \sum_{t=1}^{n} R\bigg(\frac{h_{k,n}(u_{k,n,t})}{\sqrt{n}}\bigg) - R\bigg(\frac{h_k(u_{k,n,t})}{\sqrt{n}}\bigg).
\end{aligned}$$

We will show that the remainder terms vanish. In particular, one has

$$\begin{aligned}
\sum_{t=1}^{n}\bigg|R\bigg(\frac{h_{k,n}(u_{k,n,t})}{\sqrt{n}}\bigg)\bigg| &\le \sum_{t=1}^{n}\bigg|\frac{h_{k,n}(u_{k,n,t})}{\sqrt{n}}\bigg|\bigg|\frac{h_{k,n}(u_{k,n,t})^2}{n}\bigg| \\
&\le \max_{1\le t\le n}\frac{\big|h_{k,n}(u_{k,t,n})\big|}{\sqrt{n}}\frac{1}{n}\sum_{t=1}^{n} h_{k,n}(u_{k,n,t})^2.
\end{aligned}$$

By Markov's inequality with Lemmas S10 and S11, this converges to zero in $P^n_{\theta_n(g_n,h)}$ probability. The same evidently holds for the case where $h_{k,n} = h_k$ for each $n \in \mathbb{N}$. Thus,

$$l_{n,k} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h_{k,n}(u_{k,n,t}) - h_k(u_{k,n,t}) - \frac{1}{2}\frac{1}{n}\sum_{t=1}^{n}\big[h_{k,n}(u_{k,n,t})^2 - h_k(u_{k,n,t})^2\big] + o_{P^n_{\theta_n(g_n,h)}}(1),$$

and it remains to show that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} h_{k,n}(u_{k,n,t}) - h_k(u_{k,n,t})$ and $\frac{1}{n}\sum_{t=1}^{n}[h_{k,n}(u_{k,n,t})^2 - h_k(u_{k,n,t})^2]$ also converge to zero in probability. The second of these follows directly

from Lemma S10, Markov's inequality, and the reverse triangle inequality since

$$\begin{aligned}
&P^n_{\theta_n(g_n,h)}\left(\left|\frac{1}{n}\sum_{t=1}^{n}\left[h_{k,n}(u_{k,n,t})^2 - h_k(u_{k,n,t})^2\right]\right| > \varepsilon\right) \\
&\leq \varepsilon^{-1}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[h_{k,n}(u_{k,n,t})^2 - h_k(u_{k,n,t})^2\right] \\
&= \varepsilon^{-1}\,\mathbb{E}\left[h_{k,n}(u_{k,n,t})^2 - h_k(u_{k,n,t})^2\right] \\
&\to 0.
\end{aligned}$$

For the remaining term, we start by noting that

$$\mathbb{E}\left[h_{k,n}(u_{k,n,t}) - h_k(u_{k,n,t})\right] = \frac{\mathbb{E}\left[\left(h_{k,n}(\epsilon_k) - h_k(\epsilon_k)\right)h_k(\epsilon_k)\right]}{\sqrt{n}}$$

so

$$\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{E}\left[h_{k,n}(u_{k,n,t})\right] - \mathbb{E}\left[h_k(u_{k,n,t})\right]\right| \leq \frac{1}{n}\sum_{t=1}^{n}\|h_{k,n} - h_k\|_{L_2(P^n_\theta)}\|h_k\|_{L_2(P^n_\theta)} \to 0.$$

Then, if we define $\tilde{h}_{k,n}(u_{k,n,t}) := h_{k,n}(u_{k,n,t}) - \mathbb{E}[h_{k,n}(u_{k,n,t})]$ and similarly $\tilde{h}_k(u_{k,n,t}) := h_k(u_{k,n,t}) - \mathbb{E}[h_k(u_{k,n,t})]$, it will suffice to show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{h}_{k,n}(u_{k,n,t}) - \tilde{h}_k(u_{k,n,t}) \xrightarrow{P^n_{\theta_n(g_n,h)}} 0.$$

By the reverse triangle inequality and Lemma S10,

$$\mathbb{E}\left[\left(\tilde{h}_{k,n}(u_{k,n,t}) - \tilde{h}_k(u_{k,n,t})\right)^2\right] \to 0, \quad \text{uniformly in } t.$$

Using this, the independence of the $u_{k,t,n}$ and Markov's inequality,

$$\begin{aligned}
&P^n_{\theta_n(g_n,h)}\left(\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{h}_{k,n}(u_{k,n,t}) - \tilde{h}_k(u_{k,n,t})\right| > \varepsilon\right) \\
&\leq \frac{1}{\varepsilon^2}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\left(\tilde{h}_{k,n}(u_{k,n,t}) - \tilde{h}_k(u_{k,n,t})\right)^2\right] \to 0.
\end{aligned}$$

This establishes that $\sum_{k=1}^{K} l_{n,k} \xrightarrow{P^n_{\theta_n(g_n,h)}} 0$, as required. □

LEMMA S10. *In the setting of Proposition* 4, *let* $u_{k,n,t} := e_k' A_{\theta_n(g_n,h)} V_{\theta_n(g_n,h),t}$. *Under* $P^n_{\theta_n(g_n,h)}$,

$$\mathbb{E}\left[h_{k,n}(u_{k,n,t}) - h_k(u_{k,n,t})\right]^2 \leq \|h_{n,k} - h_k\|_{L_2(P^n_\theta)}\left(1 + \frac{\|h_k\|_{L_\infty(P^n_\theta)}}{\sqrt{n}}\right).$$

PROOF. Under $P^n_{\theta_n(g_n,h)}$, $u_{k,n,t} \sim \eta_k(1 + h_k/\sqrt{n})$, so for $\epsilon_k \sim \eta_k$, since h_k is bounded,

$$\begin{aligned}
&\mathbb{E}\big[h_{k,n}(u_{k,n,t}) - h_k(u_{k,n,t})\big]^2 \\
&= \int \big[h_{n,k}(x) - h_k(x)\big]^2 \eta_k(x)\big(1 + h_k(x)/\sqrt{n}\big)\,\mathrm{d}x \\
&\le \mathbb{E}\big[h_{k,n}(\epsilon_k) - h_k(\epsilon_k)\big]^2 + \frac{1}{\sqrt{n}}\,\mathbb{E}\big[h_{k,n}(\epsilon_k) - h_k(\epsilon_k)\big]^2 \|h_k\|_{L_\infty(P^n_\theta)} \\
&\le \|h_{n,k} - h_k\|_{L_2(P^n_\theta)} + \|h_{n,k} - h_k\|_{L_2(P^n_\theta)} \|h_k\|_{L_\infty(P^n_\theta)}/\sqrt{n}.
\end{aligned}$$

□

LEMMA S11. *In the setting of Proposition* 4, *let* $u_{k,n,t} := e_k' A_{\theta_n(g_n,h)} V_{\theta_n(g_n,h),t}$. *Then*

$$\max_{1\le t\le n} \frac{\big|h_{k,n}(u_{k,n,t})\big|}{\sqrt{n}} \xrightarrow{P^n_{\theta_n(g_n,h)}} 0.$$

PROOF. Under $P^n_{\theta_n(g_n,h)}$, $u_{k,n,t} \sim \eta_k(1+h_k/\sqrt{n})$. By Lemma S10, $h_{k,n}(u_{k,n,t})$ is uniformly square $P^n_{\theta_n(g_n,h)}$–integrable, and hence the Lindeberg condition holds for $h_{k,n}(u_{k,n,t})/\sqrt{n}$:

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{t=1}^n \mathbb{E}\bigg[\frac{h_{k,n}(u_{k,n,t})^2}{n} \mathbf{1}\big\{\big|h_{n,k}(u_{k,n,t})\big| > \delta\sqrt{n}\big\}\bigg] \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n \mathbb{E}\big[h_{k,n}(u_{k,n,t})^2 \mathbf{1}\big\{\big|h_{n,k}(u_{k,n,t})\big| > \delta\sqrt{n}\big\}\big] \\
&= \lim_{n\to\infty} \mathbb{E}\big[h_{k,n}(u_{k,n,t})^2 \mathbf{1}\big\{\big|h_{n,k}(u_{k,n,t})\big| > \delta\sqrt{n}\big\}\big] \\
&= 0,
\end{aligned}$$

for any $\delta > 0$. This implies the claimed uniform asymptotic negligability condition (e.g., Gut (2005, Remark 7.2.4)):

$$\max_{1\le t\le n} \frac{\big|h_{k,n}(u_{k,n,t})\big|}{\sqrt{n}} \xrightarrow{P^n_{\theta_n(g_n,h)}} 0.$$

□

SB.4 *Scores*

LEMMA S12. *Suppose Assumption* 1 *holds. Let* p_θ *and* $\bar{q}_{n,\theta}$ *be as in the proof of Proposition* S5 *and suppose that* $\theta_n = (\gamma_n, \eta) \to (\gamma, \eta) = \theta$. *Then*

$$\lim_{n\to\infty} \int \big\| \tilde{\ell}_{\theta_n} p_{\theta_n}^{1/2} \bar{q}_{n,\theta}^{1/2} - \tilde{\ell}_\theta p_\theta^{1/2} \bar{q}_{n,\theta}^{1/2} \big\|^2 \,\mathrm{d}\lambda = 0. \tag{S9}$$

PROOF. The integral in (S9) can be rewritten as

$$\sum_{l=1}^L \int \big(\tilde{\ell}_{\theta_n,l}(y,x) p_{\theta_n}(y,x)^{1/2} - \tilde{\ell}_{\theta,l}(y,x) p_\theta(y,x)^{1/2}\big)^2 \,\mathrm{d}\big(\lambda(y) \otimes Q_{n,\theta}(x)\big)$$

Inspection of the forms of $\tilde{\ell}_\vartheta$ and p_ϑ reveals that each integrand in the preceding display converges to zero as $n \to \infty$. If we show that

$$\limsup_{n\to\infty} \int \tilde{\ell}^2_{\theta_n,l} p_{\theta_n} \, \mathrm{d}(\lambda \otimes Q_{n,\theta}) \leq \int \tilde{\ell}^2_{\theta,l} p_\theta \, \mathrm{d}(\lambda \otimes Q_\theta) < \infty, \tag{S10}$$

the proof will be complete in view of Lemma S2, Proposition S1, and Remark 1.[S8] The preceding display is equivalent to

$$\limsup_{n\to\infty} \int \tilde{\ell}^2_{\theta_n,l} \, \mathrm{d}G_{\theta_n,\theta,n} \leq \int \tilde{\ell}^2_{\theta,l} \, \mathrm{d}G_\theta < \infty,$$

for $G_{\vartheta,\theta,n}$ the distribution of (Y, X) corresponding to the density $p_\vartheta \bar{q}_{n,\theta}$ and G_θ as defined in the proof of Lemma S5. That $\tilde{\ell}^2_{\theta_n,l} \to \tilde{\ell}^2_{\theta,l}$ pointwise is clear from its form, as given in Lemma 1. The finiteness of each of the integrals in the above display along with the fact that for some $N \in \mathbb{N}$ and some $\rho > 0$,

$$\sup_{n\geq N} \int \tilde{\ell}^{2+\rho}_{\theta_n,l} \, \mathrm{d}G_{\theta_n,\theta,n} < \infty$$

follows from the form of $\tilde{\ell}^2_{\vartheta,l}$ (as given in Lemma 1) along with Assumption 1.[S9] □

LEMMA S13 (Smoothness). *Suppose that Assumption* 1 *holds. Then for any sequence* $\theta_n = (\gamma + g_n/\sqrt{n}, \eta)$ *with* $g_n \to g \in \mathbb{R}^L$,

$$R_n := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \big[\tilde{\ell}_{\theta_n}(Y_t, X_t) - \tilde{\ell}_\theta(Y_t, X_t)\big] + \tilde{I}_{\theta,n} g_n \xrightarrow{P^n_\theta} 0.$$

PROOF. From (the proof of) Lemma S8, we have

$$\lim_{n\to\infty} \int \left[\sqrt{n}\big(p^{1/2}_{\theta_n} - p^{1/2}_\theta\big)\bar{q}^{1/2}_{n,\theta} - \frac{1}{2} g' \dot{\ell}_\theta p^{1/2}_\theta \bar{q}^{1/2}_{n,\theta}\right]^2 \mathrm{d}\lambda = 0, \tag{S11}$$

while by Lemma S12 we have

$$\lim_{n\to\infty} \int \big\| \tilde{\ell}_{\theta_n} p^{1/2}_{\theta_n} \bar{q}^{1/2}_{n,\theta} - \tilde{\ell}_\theta p^{1/2}_\theta \bar{q}^{1/2}_{n,\theta} \big\|^2 \mathrm{d}\lambda = 0. \tag{S12}$$

Define

$$c_n^{-1} := \int p^{1/2}_{\theta_n} p^{1/2}_\theta \bar{q}_{n,\theta} \, \mathrm{d}\lambda = 1 - \frac{1}{2} \int \big(p^{1/2}_\theta - p^{1/2}_{\theta_n}\big)^2 \bar{q}_{n,\theta} \, \mathrm{d}\lambda.$$

[S8] Note that the product structure of $\lambda \otimes Q_{n,\theta}$ and Lemma S2 ensure that $\lambda \otimes Q_{n,\theta} \to \lambda \otimes Q_\theta$ setwise.

[S9] Compare the proof of Lemma S3: arguing in essentially the same manner as there allows one to obtain uniform boundedness of the $4 + \delta$ moments of ϵ_k, $\phi_k(\epsilon_k)$, X_t (uniformly in t) and all the nonstochastic terms in $\tilde{\ell}^2_{\theta_n,l}$.

We have

$$-n\big(p_\theta^{1/2}-p_{\theta_n}^{1/2}\big)^2=-\bigg(\sqrt{n}\big[p_{\theta_n}^{1/2}-p_\theta^{1/2}\big]-\frac{1}{2}g'\dot\ell_\theta p_\theta^{1/2}\bigg)^2+\bigg(\frac{1}{2}g'\dot\ell_\theta p_\theta^{1/2}\bigg)^2 - g'\dot\ell_\theta p_\theta^{1/2}\sqrt{n}\big(p_{\theta_n}^{1/2}-p_\theta^{1/2}\big),$$

and so by (S11) and the continuity of the inner product

$$\begin{aligned}\int\big(p_\theta^{1/2}-p_{\theta_n}^{1/2}\big)^2\bar q_{n,\theta}\,\mathrm{d}\lambda&=\frac{1}{n}\int g'\dot\ell_\theta p_\theta^{1/2}\bar q_{n,\theta}^{1/2}\sqrt{n}\big(p_{\theta_n}^{1/2}-p_\theta^{1/2}\big)\bar q_{n,\theta}^{1/2}\,\mathrm{d}\lambda\\&\quad-\frac{1}{n}\int\bigg(\frac{1}{2}g'\dot\ell_\theta p_\theta^{1/2}\bigg)^2\bar q_{n,\theta}\,\mathrm{d}\lambda+o\big(n^{-1}\big)\\&=\frac{1}{4}\big(n^{-1/2}g\big)'\dot I_{n,\theta}\big(n^{-1/2}g\big)+o\big(n^{-1}\big),\end{aligned}$$

where $\dot I_{n,\theta}:=\int\dot\ell_\theta\dot\ell_\theta' p_\theta\bar q_{n,\theta}\,\mathrm{d}\lambda=O(1)$.[S10] It follows that $c_n^{-1}=1-a_n$ with $a_n\to 0$ and $na_n=\frac{1}{4}g'\dot I_\theta g+o(1)$.

R_n is equal to the sum of

$$R'_{1,n}:=\frac{1}{\sqrt{n}}\sum_{t=1}^n\bigg[\tilde\ell_{\theta_n}(Y_t,X_t)\bigg(1-\frac{p_{\theta_n}(Y_t,X_t)^{1/2}}{p_\theta(Y_t,X_t)^{1/2}}\bigg)\bigg]+\frac{1}{2}\tilde I_{n,\theta}g_n\,;$$

$$R'_{2,n}:=\frac{1}{\sqrt{n}}\sum_{t=1}^n\bigg[\tilde\ell_{\theta_n}(Y_t,X_t)\frac{p_{\theta_n}(Y_t,X_t)^{1/2}}{p_\theta(Y_t,X_t)^{1/2}}-\tilde\ell_\theta(Y_t,X_t)\bigg]+\frac{1}{2}\tilde I_{n,\theta}g_n.$$

Since $\tilde I_{n,\theta}$ is $O(1)$ by Lemma S3, it suffices to prove that these converge in probability to zero with g_n replaced by g; let the corresponding expressions be called $R_{i,n}$ for $i=1,2$.

For $R_{1,n}$, we note that (omitting the arguments of the functions)

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=1}^n\tilde\ell_{\theta_n}\bigg(1-\frac{p_{\theta_n}^{1/2}}{p_\theta^{1/2}}\bigg)+\frac{1}{2}\frac{1}{n}\sum_{t=1}^n\tilde\ell_{\theta_n}\dot\ell_\theta' g\\&=\frac{1}{n}\sum_{t=1}^n\tilde\ell_{\theta_n}\sqrt{n}\bigg(1-\frac{p_{\theta_n}^{1/2}}{p_\theta^{1/2}}+\frac{1}{2\sqrt{n}}\dot\ell_\theta' g\bigg)\\&\le\frac{1}{n}\sum_{t=1}^n\|\tilde\ell_{\theta_n}\|^2\times\frac{1}{n}\sum_{t=1}^n\bigg[\sqrt{n}\bigg(1-\frac{p_{\theta_n}^{1/2}}{p_\theta^{1/2}}+\frac{1}{2\sqrt{n}}\dot\ell_\theta' g\bigg)\bigg]^2.\end{aligned}$$

The first term on the second line is $O_{P_{\theta_n}^n}(1)$, hence $O_{P_\theta^n}(1)$ (by contiguity). The second has $L_1(P_\theta^n)$ norm

$$\mathbb{E}\bigg|\frac{1}{n}\sum_{t=1}^n\bigg[\sqrt{n}\bigg(1-\frac{p_{\theta_n}^{1/2}}{p_\theta^{1/2}}+\frac{1}{2\sqrt{n}}\dot\ell_\theta' g\bigg)\bigg]^2\bigg|\le\int\bigg[\sqrt{n}\bigg(p_\theta^{1/2}-p_{\theta_n}^{1/2}+\frac{1}{2\sqrt{n}}\dot\ell_\theta' g p_\theta^{1/2}\bigg)\bigg]^2\bar q_{n,\theta}\,\mathrm{d}\lambda\to 0,$$

[S10] This follows by noting that $\|\dot\ell_\theta\|^2$ is uniformly integrable under $p_\theta\bar q_{n,\theta}$, which is a consequence of Lemma S3.

where the convergence is by (S11). Therefore, it suffices to show that

$$\frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\dot{\ell}'_{\theta} - \tilde{I}_{n,\theta} \xrightarrow{P_\theta^n} 0. \tag{S13}$$

We may replace $\tilde{I}_{n,\theta}$ in (S13) with $\tilde{I}_\theta := \int \tilde{\ell}_\theta \dot{\ell}'_\theta \, \mathrm{d}G_\theta$ with G_θ as defined in the proof of Lemma S5. In particular, let $G_{\theta,n} := G_{\theta,0,n}$ as defined in the proof of Lemma S5. Then, since $\|\tilde{\ell}_\theta(Y_t, X_t)\dot{\ell}_\theta(Y_t, X_t)'\|^{1+\rho/2}$ is uniformly $L_1(P_\theta^n)$ bounded (Lemma S3), one has

$$\sup_{n\in\mathbb{N}} \int \left\|\tilde{\ell}_\theta \dot{\ell}'_\theta\right\|^{1+\rho/2} \mathrm{d}G_{n,\theta} < \infty,$$

and so $\|\tilde{\ell}_\theta\dot{\ell}'_\theta\|$ is uniformly $G_{\theta,n}$–integrable. By Lemma S22 and Theorem 2.8 of Serfozo (1982),

$$\tilde{I}_{n,\theta} = \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\big[\tilde{\ell}_\theta(Y_t, X_t)\dot{\ell}_\theta(Y_t, X_t)'\big] = \int \tilde{\ell}_\theta\dot{\ell}'_\theta \, \mathrm{d}G_{n,\theta} \to \int \tilde{\ell}_\theta\dot{\ell}'_\theta \, \mathrm{d}G_\theta = \tilde{I}_\theta. \tag{S14}$$

For any $M > 0$, one has the decompositions

$$\begin{aligned} E_{n,1}^M &:= \frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\dot{\ell}'_\theta - \frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| \le M\}\dot{\ell}'_\theta\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\} \\ &= \frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| > M\}\dot{\ell}'_\theta + \frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| \le M\}\dot{\ell}'_\theta\mathbf{1}\{\|\dot{\ell}_\theta\| > M\} \end{aligned}$$

and

$$\begin{aligned} E_2^M &:= \tilde{I}_\theta - \int \tilde{\ell}_\theta\dot{\ell}'_\theta\mathbf{1}\{\|\tilde{\ell}_\theta\| \le M\}\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\}\,\mathrm{d}G_\theta \\ &= \int \tilde{\ell}_\theta\dot{\ell}'_\theta\mathbf{1}\{\|\tilde{\ell}_\theta\| > M\}\,\mathrm{d}G + \int \tilde{\ell}_\theta\dot{\ell}'_\theta\mathbf{1}\{\|\tilde{\ell}_\theta\| > M\}\mathbf{1}\{\|\dot{\ell}_\theta\| > M\}\,\mathrm{d}G_\theta. \end{aligned}$$

Additionally, for $\mathbb{E}$ taken under P_θ^n, define

$$E_{n,3}^M := \frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| \le M\}\dot{\ell}'_\theta\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\} - \mathbb{E}\big[\tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| \le M\}\dot{\ell}'_\theta\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\}\big];$$

$$E_{n,4}^M := \mathbb{E}\,\frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| \le M\}\dot{\ell}'_\theta\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\} - \int \tilde{\ell}_\theta\dot{\ell}'_\theta\mathbf{1}\{\|\tilde{\ell}_\theta\| \le M\}\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\}\,\mathrm{d}G_\theta.$$

Since $\|\tilde{\ell}_\theta\dot{\ell}'_\theta\mathbf{1}\{\|\tilde{\ell}_\theta\| > M\}\| \le \|\tilde{\ell}_\theta\dot{\ell}'_\theta\|$, $\|\tilde{\ell}_\theta\dot{\ell}'_\theta\mathbf{1}\{\|\tilde{\ell}_\theta\| > M\}\mathbf{1}\{\|\dot{\ell}_\theta\| > M\}\| \le \|\tilde{\ell}_\theta\dot{\ell}'_\theta\|$, and $\|\tilde{\ell}_\theta\dot{\ell}'_\theta\|$ is G_θ-integrable by Lemma S3, by the dominated convergence theorem, for any $\delta > 0$ there is an M such that $E_2^{M'} < \delta$ for $M' \ge M$. For any $M > 0$, by Theorem 3 in Saikkonen (2007), Theorem 14.1 in Davidson (1994) and Theorem 2 in Kanaya (2017) one has (cf. Lemma S14 below)

$$E_{n,3}^M = O_{P_\theta^n}\big(M^2/\sqrt{n}\big).$$

For $E^M_{n,4}$, we introduce a new measure: define μ_n as

$$\mu_n(A) := \int_A c_n p_{\theta_n}(x,y)^{1/2} p_\theta(x,y)^{1/2} \,\mathrm{d}\big(\lambda(y)\otimes Q_n(x)\big).$$

By Lemma S22, one has that $\mu_n \to G$, as well as $G_{n,\theta} \to G$, in TV. Then, by Cauchy–Schwarz and Lemma S3,

$$\begin{aligned}
&c_n^{-1}\int \tilde\ell_{\theta_n}\mathbf{1}\{\|\tilde\ell_{\theta_n}\|\le M\}\dot\ell'_\theta\mathbf{1}\{\|\dot\ell_\theta\|\le M\}\,\mathrm{d}\mu_n - \int \tilde\ell_\theta\mathbf{1}\{\|\tilde\ell_\theta\|\le M\}\dot\ell'_\theta\mathbf{1}\{\|\dot\ell_\theta\|\le M\}\,\mathrm{d}G_{n,\theta}\\
&\quad= \int \big(\tilde\ell_{\theta_n}\mathbf{1}\{\|\tilde\ell_{\theta_n}\|\le M\}p_{\theta_n}^{1/2} - \tilde\ell_\theta\mathbf{1}\{\|\tilde\ell_\theta\|\le M\}p_\theta^{1/2}\big)\dot\ell'_\theta\mathbf{1}\{\|\dot\ell_\theta\|\le M\}p_\theta^{1/2}\,\mathrm{d}(\lambda\otimes Q_{\theta,n})\\
&\quad= \int \big(\tilde\ell_{\theta_n}\mathbf{1}\{\|\tilde\ell_{\theta_n}\|> M\}p_{\theta_n}^{1/2} - \tilde\ell_\theta\mathbf{1}\{\|\tilde\ell_\theta\|> M\}p_\theta^{1/2}\big)\dot\ell'_\theta\mathbf{1}\{\|\dot\ell_\theta\|\le M\}p_\theta^{1/2}\,\mathrm{d}(\lambda\otimes Q_{\theta,n})\\
&\qquad+ \int \big(\tilde\ell_{\theta_n}p_{\theta_n}^{1/2} - \tilde\ell_\theta p_\theta^{1/2}\big)\dot\ell'_\theta\mathbf{1}\{\|\dot\ell_\theta\|\le M\}p_\theta^{1/2}\,\mathrm{d}(\lambda\otimes Q_{\theta,n})\\
&\quad\lesssim o(1) + \sup_{n\in\mathbb{N}}\mathbb{E}_{\theta_n}\big[\|\tilde\ell_{\theta_n}\|^2\mathbf{1}\{\|\tilde\ell_{\theta_n}\|>M\}\big] + \sup_{n\in\mathbb{N}}\mathbb{E}_{\theta}\big[\|\tilde\ell_{\theta}\|^2\mathbf{1}\{\|\tilde\ell_{\theta}\|>M\}\big].
\end{aligned}$$

The last two right-hand side terms can be made arbitrarily small, uniformly in n, by taking M large enough; the $o(1)$ term follows from (S12) and is uniform in M. Now, by $G_{n,\theta}\xrightarrow{\text{TV}} G_\theta$,

$$\begin{aligned}
&\left|\int \tilde\ell_\theta\mathbf{1}\{\|\tilde\ell_\theta\|\le M\}\dot\ell'_\theta\mathbf{1}\{\|\dot\ell_\theta\|\le M\}\,\mathrm{d}G_{\theta,n} - \int \tilde\ell_\theta\mathbf{1}\{\|\tilde\ell_\theta\|\le M\}\dot\ell'_\theta\mathbf{1}\{\|\dot\ell_\theta\|\le M\}\,\mathrm{d}G_\theta\right|\\
&\quad\le M^2\|G_{n,\theta}-G_\theta\|_{\text{TV}}.
\end{aligned}$$

Since $\mu_n\to G_\theta$ and $G_{n,\theta}\to G_\theta$ in total variation, one has that $\|\mu_n - G_{n,\theta}\|_{\text{TV}}\to 0$. Since $\tilde\ell_{\theta_n}\mathbf{1}\{\|\tilde\ell_{\theta_n}\|\le M\}\dot\ell'_\theta\mathbf{1}\{\|\dot\ell_\theta\|\le M\}$ is uniformly bounded, one has that

$$\begin{aligned}
&\left|\int \tilde\ell_{\theta_n}\mathbf{1}\{\|\tilde\ell_{\theta_n}\|\le M\}\dot\ell'_\theta\mathbf{1}\{\|\dot\ell_\theta\|\le M\}\,\mathrm{d}\mu_n - \int \tilde\ell_{\theta_n}\mathbf{1}\{\|\tilde\ell_{\theta_n}\|\le M\}\dot\ell'_\theta\mathbf{1}\{\|\dot\ell_\theta\|\le M\}\,\mathrm{d}G_{n,\theta}\right|\\
&\quad\le M^2\|\mu_n - G_{n,\theta}\|_{\text{TV}}.
\end{aligned}$$

As $c_n^{-1}-1 = -a_n\to 0$, it follows that

$$E^M_{n,4}\le M^2\big[\|\mu_n - G_{n,\theta}\|_{\text{TV}} + \|G_{n,\theta}-G_\theta\|_{\text{TV}}\big] + e_n + M^2|a_n| + r(M),$$

where $0\le r(M) := \sup_{n\in\mathbb{N}}\mathbb{E}_{P^n_{\theta_n}}[\|\tilde\ell_{\theta_n}\|^2\mathbf{1}\{\|\tilde\ell_{\theta_n}\|>M\}] + \sup_{n\in\mathbb{N}}\mathbb{E}_{P^n_{\theta}}[\|\tilde\ell_{\theta}\|^2\mathbf{1}\{\|\tilde\ell_{\theta}\|>M\}]\to 0$ as $M\to\infty$ and r does not depend on n and $e_n = o(1)$. For $E^M_{n,1}$, note that since $\|\dot\ell_\theta\|^2$ is uniformly P^n_θ-integrable (Lemma S3), $\frac{1}{n}\sum_{t=1}^n\|\dot\ell_\theta\|^2 = O_{P^n_\theta}(1)$. By Markov's inequality, for any $\delta>0$,

$$\begin{aligned}
P^n_{\theta_n}\left(\left|\frac{1}{n}\sum_{t=1}^n\|\tilde\ell_{\theta_n}\|^2\mathbf{1}\{\|\tilde\ell_{\theta_n}\|>M\}\right|>\delta\right) &\le \delta^{-1}\,\mathbb{E}\left[\left|\frac{1}{n}\sum_{t=1}^n\|\tilde\ell_{\theta_n}\|^2\mathbf{1}\{\|\tilde\ell_{\theta_n}\|>M\}\right|\right]\\
&\le \delta^{-1}\sup_{n\in\mathbb{N}}\mathbb{E}\,\|\tilde\ell_{\theta_n}\|^2\mathbf{1}\{\|\tilde\ell_{\theta_n}\|>M\}\\
&\le \delta^{-1}r(M).
\end{aligned}$$

Thus by taking $M \to \infty$, the probability on the left-hand side of the preceding display vanishes. Therefore, the same is true of

$$P_\theta^n\left(\left|\frac{1}{n}\sum_{t=1}^n \|\tilde{\ell}_{\theta_n}\|^2 \mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| > M\}\right| > \delta\right),$$

by contiguity. That is, we can take a large enough M such that the probability in the display above is arbitrarily small (for all large enough $n \in \mathbb{N}$).

Now, fix $\varepsilon > 0$, $\delta > 0$. By Lemma S3, $\frac{1}{n}\sum_{t=1}^n \|\tilde{\ell}_\theta\|^2 = O_{P_\theta^n}(1)$ and also $\frac{1}{n}\sum_{t=1}^n \|\tilde{\ell}_{\theta_n}\|^2 = O_{P_{\theta_n}^n}(1)$. By this and contiguity, we can choose $R > 0$ be such that for all $n \geq N_1$,

$$P_\theta^n\left(\frac{1}{n}\sum_{t=1}^n \|\tilde{\ell}_\theta\|^2 > R\right) < \varepsilon/4, \qquad P_\theta^n\left(\frac{1}{n}\sum_{t=1}^n \|\tilde{\ell}_{\theta_n}\|^2 > R\right) < \varepsilon/4.$$

Take M large enough that $\|E_2^M\| < \delta$, $r(M) < \delta$, and for all $n \geq N_2$

$$P_\theta^n\left(\left|\frac{1}{n}\sum_{t=1}^n \|\tilde{\ell}_{\theta_n}\|^2 \mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| > M_n\}\right| > \delta/R\right) < \varepsilon/4,$$

$$P_\theta^n\left(\left|\frac{1}{n}\sum_{t=1}^n \|\dot{\ell}_\theta\|^2 \mathbf{1}\{\|\dot{\ell}_\theta\| > M_n\}\right| > \delta/R\right) < \varepsilon/4,$$

where $M_n \geq M$ and $M_n \to \infty$ slowly. This ensures that $\|E_2^{M_n}\| < \delta$, $P_\theta^n(\|E_{n,1}^{M_n}\| > 2\delta) < \varepsilon$ for all $n \geq \max\{N_1, N_2\}$. Then let N be large enough such that $N \geq \max\{N_1, N2\}$, and for all $n \geq N$, $P_\theta^n(\|E_{n,3}^{M_n}\| > \delta) < \varepsilon$ and $\|E_{n,4}^{M_n}\| \leq 3\delta$.[S11] Combining these ensures that for all such n,

$$P_\theta^n\left(\left\|\frac{1}{n}\sum_{t=1}^n \tilde{\ell}_{\theta_n}\dot{\ell}'_\theta - \tilde{I}_\theta\right\| > 7\delta\right) < 2\varepsilon.$$

In conjunction with (S14), this establishes (S13).

We next show that $R_{2,n}$ converges to zero in P_θ^n-probability. Define

$$Z_{n,t} := \tilde{\ell}_{\theta_n}(Y_t, X_t)\frac{p_{\theta_n}(Y_t, X_t)^{1/2}}{p_\theta(Y_t, X_t)^{1/2}},$$

$$m_n(X_t) := \int \tilde{\ell}_{\theta_n}(y, X_t)\, p_{\theta_n}(y, X_t)^{1/2} p_\theta(y, X_t)^{1/2}\, dy,$$

and note that $m_n(X_t) = \mathbb{E}[Z_{n,t}|X_t]$ (P_θ^n-a.s.). Since $\mathbb{E}[\tilde{\ell}_{\theta_n}(Y_t, X_t)|X_t] = 0$ under $P_{\theta_n}^n$ (which is clear from its form),

$$\begin{aligned} m_n(X_t) &= \int \tilde{\ell}_{\theta_n}(y, X_t)\, p_{\theta_n}(y, X_t)^{1/2} p_\theta(y, X_t)^{1/2}\, \mathrm{d}y \\ &= \int \tilde{\ell}_{\theta_n}(y, X_t)\, p_{\theta_n}(y, X_t)^{1/2}\big[p_\theta(y, X_t)^{1/2} - p_{\theta_n}(y, X_t)^{1/2}\big]\, \mathrm{d}y. \end{aligned} \tag{S15}$$

[S11] That is, n such that $M_n^2|a_n| < \delta$, $|e_n| < \delta$, $M_n^2[\|\mu_n - G_{n,\theta}\|_{\mathrm{TV}} + \|G_{n,\theta} - G_\theta\|_{\mathrm{TV}}] < \delta$. Here, one needs to take $M_n \to \infty$ slowly enough that these sequences still converge to zero and $M_n^2/\sqrt{n} \to 0$.

Using (S11), (S12), and Cauchy–Schwarz yield

$$\lim_{n\to\infty}\left|\langle\tilde{\ell}_{\theta_n}p_{\theta_n}^{1/2}\bar{q}_{\theta,n}^{1/2},\sqrt{n}(p_\theta^{1/2}-p_{\theta_n}^{1/2})\bar{q}_{n,\theta}^{1/2}\rangle_\lambda-\left\langle\tilde{\ell}_\theta p_\theta^{1/2}\bar{q}_{n,\theta}^{1/2},-\frac{1}{2}g'\dot{\ell}_\theta p_\theta^{1/2}\bar{q}_{n,\theta}^{1/2}\right\rangle_\lambda\right|=0,$$

which implies that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}m_n(X_t)+\frac{1}{2}\tilde{I}_{n,\theta}g\xrightarrow{P_\theta^n}0,$$

given the representation of m_n in (S15). In consequence, it remains to show that

$$R_{2,n}^*:=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}Z_{t,n}-m_n(X_t)-\tilde{\ell}_\theta(Y_t,X_t)\xrightarrow{P_\theta^n}0.$$

Put $\mathcal{F}_{n,t}=\sigma(Y_t,X_t)$. Then, as is straightforward to verify, $(Z_{t,n}-m_n(X_t)-\tilde{\ell}_\theta(Y_t,X_t),\mathcal{F}_{n,t})_{n\in\mathbb{N},1\le t\le n}$ forms a martingale difference array. Hence, it suffices to show that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left\|Z_{t,n}-m_n(X_t)-\tilde{\ell}_\theta(Y_t,X_t)\right\|^2\xrightarrow{P_\theta^n}0.$$

The left-hand side of this display can be written as

$$\int\left\|\tilde{\ell}_{\theta_n}\frac{p_{\theta_n}^{1/2}}{p_\theta^{1/2}}-m_n-\tilde{\ell}_\theta\right\|^2p_\theta\bar{q}_{n,\theta}\,\mathrm{d}\lambda\le2\int\left\|\tilde{\ell}_{\theta_n}p_{\theta_n}^{1/2}\bar{q}_{n,\theta}^{1/2}-\tilde{\ell}_\theta p_\theta^{1/2}\bar{q}_{n,\theta}^{1/2}\right\|^2\mathrm{d}\lambda+2\int\|m_n\|^2\,\mathrm{d}Q_{n,\theta},$$

and so, given (S12) it suffices to show that the second term on the right-hand side converges to zero. For this, note that by Fubini's theorem and the Cauchy–Schwarz inequality

$$\begin{aligned}\int\|m_n\|^2\,\mathrm{d}Q_{n,\theta}&\le\int\left\|\tilde{\ell}_{\theta_n}p_{\theta_n}^{1/2}\left[p_\theta^{1/2}-p_{\theta_n}^{1/2}\right]\right\|^2\bar{q}_{n,\theta}\,\mathrm{d}\lambda\\&\le\int\left\|\tilde{\ell}_{\theta_n}p_{\theta_n}^{1/2}\bar{q}_{n,\theta}^{1/2}\right\|^2\mathrm{d}\lambda\int\left[(p_{\theta_n}^{1/2}-p_\theta^{1/2})\bar{q}_{n,\theta}^{1/2}\right]^2\mathrm{d}\lambda.\end{aligned}$$

The first term on the right-hand side is $O(1)$ by equation (S10), while the second converges to zero by (S11) and the uniform $G_{\theta,0,n}$—integrability of $g'\dot{\ell}_\theta$ as established in Lemma S6. □

SB.4.1 *Estimation*

Lemma S14. *Suppose that Assumption* 1 *holds and* g_n *are* ϱ*—integrable functions for some* $\varrho>2$ *such that* $\max_{t=1,\dots,n}\|g_n(Y_t,X_t)\|_{L_\varrho}\le M_n$ *(all under* P_θ^n*). Then*

$$\frac{1}{n}\sum_{t=1}^{n}g_n(Y_t,X_t)-\mathbb{E}\left[g_n(Y_t,X_t)\right]=O_{P_\theta}(M_n/\sqrt{n}).$$

PROOF. Let $\alpha_n(m)$ be the α-mixing coefficients of the array $\{g_n(Y_t, X_t) - \mathbb{E}[g_n(Y_t, X_t)] : n \in \mathbb{N}, 1 \leq t \leq n\}$. By (the proof of) Theorem 14.1 in Davidson (1994), $\alpha_n(m) \leq \tilde{\alpha}(m - p)$ (for $m \geq p$) where $\tilde{\alpha}(m)$ are the mixing coefficients of $\{Y_t : t \in \mathbb{N}\}$. By Theorem 3 in Saikkonen (2007) and Proposition 1.1.1 in Doukhan (1994) $\tilde{\alpha}(m) = O(a^m)$ for some $a \in (0, 1)$. Condition A1 in Kanaya (2017) then holds (with $\Delta = 1$) with $\beta > \varrho/(\varrho - 2)$. To see this, note that for all $m \geq M_1$ we have $\tilde{\alpha}(m - p) \leq Ca^m$ while $Ca^m \leq Am^{-\beta}$ whenever

$$\beta \leq \frac{\log(A) - \log(C) + m\big|\log(a)\big|}{\log(m)}.$$

As the right-hand side diverges as $m \to \infty$, for all m larger than some $M \geq M_1$, the inequality will hold for some $\beta > \varrho/(\varrho - 2)$. Noting that the inequality above continues to hold if we increase A, we may then choose A such that each $\tilde{\alpha}(m) \leq Am^{-\beta}$ for all $1 \leq m \leq M$. The result then follows by Theorem 2 in Kanaya (2017). □

LEMMA S15. *Suppose that Assumptions* 1 *and* 2 *hold. Then*:

(i) *If* $Z_{n,1} := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{\ell}_\theta(Y_t, X_t)$ *and* $Z_{n,2} := \Lambda^n_{\theta_n(g,h)}(Y^n)$, *then under* P^n_θ,

$$Z_n \rightsquigarrow Z \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ -\frac{1}{2}\sigma^2_{g,h} \end{pmatrix}, \begin{pmatrix} \tilde{I}_\theta & \tilde{I}_\theta g \\ g'\tilde{I}_\theta & \sigma^2_{g,h} \end{pmatrix}\right).$$

Additionally, let $\theta_n := \theta_n(g_n, 0) = (\gamma + g_n/\sqrt{n}, \eta)$ *for* $g_n \to g \in \mathbb{R}^L$. *Then*:

(ii) *We have that*

$$\frac{1}{n}\sum_{t=1}^{n}\big(\hat{\ell}_{\theta_n}(Y_t, X_t) - \tilde{\ell}_{\theta_n}(Y_t, X_t)\big) = o_{P^n_{\theta_n}}\big(n^{-1/2}\big).$$

(iii) $\|\hat{I}_{n,\theta_n} - \tilde{I}_\theta\| = o_{P^n_{\theta_n}}(\nu_n^{1/2})$ *where* ν_n *is defined in Assumption* 2, *and* $\tilde{I}_\theta := G_\theta \tilde{\ell}_\theta \tilde{\ell}'_\theta$ *with* G_θ *as in the proof of Lemma* S5.

PROOF. For part (i), let z_t be

$$z_t := \left(\tilde{\ell}_\theta(Y_t, X_t)', g'\dot{\ell}_\theta(Y_t, X_t) + \sum_{k=1}^{K} h_k(A_{k\bullet}V_{\theta,t})\right)',$$

and $\mathcal{F}_t := \sigma(\epsilon_1, \ldots, \epsilon_t)$. Under P^n_θ, $\{z_t, \mathcal{F}_t : t \in \mathbb{N}\}$ is a martingale difference sequence such that

$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\big[z_t z_t'\big] = \begin{bmatrix} \tilde{I}_{n,\theta} & \tilde{I}_{\theta,\theta} g \\ g'\tilde{I}_{n,\theta} & \sigma^2_{g,h,n} \end{bmatrix} \to \begin{bmatrix} \tilde{I}_\theta & \tilde{I}_\theta g \\ g'\tilde{I}_\theta & \sigma^2_{g,h} \end{bmatrix},$$

noting Lemma 1 and Theorem 12.14 of Rudin (1991). That $\sigma^2_{g,h,n}$ converges to a $\sigma^2_{g,h}$ is part of the conclusion of Proposition 3. That $\tilde{I}_{\theta,n} \to \tilde{I}_\theta$ follows by combining Lemma S3, the fact that $G_{\theta,0,n}$ as defined in the proof of Lemma S5 converges in total variation to

G_θ (cf. Lemma S22), and Corollary 2.9 in Feinberg, Kasyanov, and Zgurovsky (2016). Lindeberg's condition is satisfied since $\{\|z_t\|^2 : t \in \mathbb{N}\}$ is uniformly P_θ^n-integrable (by Lemma S3 and the fact that each h_k is bounded) and the variance convergence in the preceding display. Part (i) then follows from Proposition 3 and the central limit theorem for martingale differences.

Define $A_n := A(\theta_n)$, $B_n := B(\theta_n)$, and $\zeta^x_{n,l,k,j} := \zeta^x_{l,k,j}(\theta_n)$ for each triple (l, j, k) of indicies and $x \in \{\alpha, \sigma\}$. Note that each $A_{n,k}(Y_t - B_n X_t) \approx \epsilon_{k,t} \sim \eta_k$ under $P^n_{\theta_n}$. Hence,

$$\tilde{\ell}_{\theta_n,\alpha_l}(Y_t, X_t) \approx \sum_{k=1}^K \sum_{j=1, j\neq k}^K \zeta^\alpha_{l,k,j}\phi_k(\epsilon_{k,t})\epsilon_{j,t} + \sum_{k=1}^K \zeta^\alpha_{n,l,k,k}\big[\tau_{k,1}\epsilon_{k,t} + \tau_{k,2}\kappa(\epsilon_{k,t})\big], \tag{S16}$$

$$\tilde{\ell}_{\theta_n,\sigma_l}(Y_t, X_t) \approx \sum_{k=1}^K \sum_{j=1, j\neq k}^K \zeta^\sigma_{n,l,k,j}\phi_k(\epsilon_{k,t})\epsilon_{j,t} + \sum_{k=1}^K \zeta^\sigma_{l,k,k}\big[\tau_{k,1}\epsilon_{k,t} + \tau_{k,2}\kappa(\epsilon_{k,t})\big], \tag{S17}$$

$$\tilde{\ell}_{\theta_n,b_l}(Y_t, X_t) \approx \sum_{k=1}^K -A_{n,k\bullet}D_{b,l}\big[\phi_k(\epsilon_{k,t})(X_t - \mathbb{E}\,X_t) - \mathbb{E}\,X_t\big(\varsigma_{k,1}\epsilon_{k,t} + \varsigma_{k,2}\kappa(\epsilon_{k,t})\big)\big]. \tag{S18}$$

By Assumption 1.3, $\zeta^x_{n,l,k,j} \to \zeta^\alpha_{\infty,l,k,j} := [D_{x_l}(\alpha, \sigma)]_{k\bullet} A(\alpha, \sigma)^{-1}_{\bullet j}$ for $x \in \{\alpha, \sigma\}$. Note that the entries of $D_{b,l}$ are all zero except for entry l (corresponding to b_l), which is equal to one.

We verify (ii) for each component of the efficient score (S16)–(S18). For components (S16) and (S17), we define for x either of α, σ,

$$\varphi_{1,n,t} := \sum_{k=1}^K \sum_{j=1, j\neq k}^K \zeta^x_{l,k,j,n}\phi_k(A_{n,k\bullet}V_{n,t})A_{n,j\bullet}V_{n,t},$$

and

$$\hat{\varphi}_{1,n,t} := \sum_{k=1}^K \sum_{j=1, j\neq k}^K \zeta^x_{l,k,j,n}\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})A_{n,j\bullet}V_{n,t},$$

with $V_{n,t} = Y_t - B_n X_t$, and let $\overline{\zeta}_n := \max_{l\in[L], j\in[K], k\in[K]} |\zeta^x_{l,j,k,n}|$, which converges to $\overline{\zeta} := \max_{l\in[L], j\in[K], k\in[K]} |\zeta^x_{l,j,k,\infty}| < \infty$. We have that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^n (\hat{\varphi}_{1,n,t} - \varphi_{1,n,t}) \\
&\quad \leq \sqrt{n}\sum_{k=1}^K \sum_{j=1, j\neq k}^K \overline{\zeta}_n \left| \frac{1}{n}\sum_{t=1}^n \hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})A_{n,j\bullet}V_{n,t} - \phi_k(A_{n,k\bullet}V_{n,t})A_{n,j\bullet}V_{n,t} \right|,
\end{aligned}$$

Each $|\frac{1}{n}\sum_{t=1}^n \hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})A_{n,j\bullet}V_{n,t} - \phi_k(A_{n,k\bullet}V_{n,t})A_{n,j\bullet}V_{n,t}| = o_{P_{\theta_n}}(n^{-1/2})$ by applying Lemma 2 with $W_{n,t} = A_{n,j\bullet}V_{n,t}$ (noting that $A_{n,k\bullet}V_{n,s} \simeq \epsilon_{k,s}$ and $A_{n,j\bullet}V_{n,t} \simeq \epsilon_{j,t}$ with are independent for any s, t with $\mathbb{E}_{\theta_n}(A_{n,j\bullet}V_{n,t})^2 = 1$ by Assumption 1.2), and the outside

summations are finite, it follows that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\hat{\varphi}_{1,n,t}-\varphi_{1,n,t})=o_{P^n_{\tilde{\theta}_n}}(1). \tag{S19}$$

That $\hat{\tau}_{k,n}\xrightarrow{P^n_{\theta_n}}\tau_k$ follows from Lemma S16. Now, consider $\varphi_{2,\tau,n,t}$ defined by

$$\varphi_{2,\tau,n,t}:=\sum_{k=1}^{K}\zeta^{z}_{n,l,k,k}\big[\tau_{k,1}A_{n,k\bullet}V_{n,t}+\tau_{k,2}\kappa(A_{n,k\bullet}V_{n,t})\big],$$

for x equal to either α or σ. Since sum is finite and each $|\zeta^{x}_{n,l,k,k}|\to|\zeta^{x}_{\infty,l,k,k}|<\infty$ it is sufficient to consider the convergence of the summands. In particular, we have that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\hat{\tau}_{k,n,1}-\tau_{k,1}]A_{n,k\bullet}V_{n,t}=[\hat{\tau}_{k,n,1}-\tau_{k,1}]\frac{1}{\sqrt{n}}\sum_{i=1}^{n}A_{n,k\bullet}V_{n,t}\to 0,$$
$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\hat{\tau}_{k,n,2}-\tau_{k,2}]\kappa(A_{n,k\bullet}V_{n,t})=[\hat{\tau}_{k,n,2}-\tau_{k,2}]\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\kappa(A_{n,k\bullet}V_{n,t})\to 0,$$

in probability, since $A_{n,k\bullet}V_{n,t}\approx\epsilon_{k,t}\sim\eta_k$ and $(\epsilon_{k,t})_{t\geq 1}$ and $(\kappa(\epsilon_{k,t}))_{t\geq 1}$ are i.i.d. mean-zero sequences with finite second moments such that the central limit theorem holds.

Together, these yield that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varphi_{2,\hat{\tau}_n,n,t}-\varphi_{2,\tau,n,t})\xrightarrow{P^n_{\theta_n}}0. \tag{S20}$$

Combination of (S19) and (S20) yields (ii) for components of the type (S16), (S17).

For components (S18), let $a_{n,k,l}:=-A_{n,k\bullet}D_{b_l}$, $\tilde{\varsigma}_{k,n}:=\hat{\varsigma}_{k,n}-\varsigma_k$, $c_{n,t}:=\mathbb{E}_{\theta_n}X_t$, and $\bar{c}_n:=\frac{1}{n}\sum_{t=1}^{n}c_{n,t}$.

Since $a_{n,k,l}\to a_{\infty,k,l}:=A(\alpha,\sigma)_{k\bullet}D_{b_l}(\alpha,\sigma)$, it suffices to show that:

1. $\frac{1}{n}\sum_{t=1}^{n}[\phi_k(A_{n,k\bullet}V_{n,t})-\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})](X_t-c_{n,t})=o_{P^n_{\theta_n}}(n^{-1/2})$;
2. $\frac{1}{n}\sum_{t=1}^{n}[\phi_k(A_{n,k\bullet}V_{n,t})-\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})](\bar{X}_n-\bar{c}_n)=o_{P^n_{\theta_n}}(n^{-1/2})$;
3. $\frac{1}{n}\sum_{t=1}^{n}[\phi_k(A_{n,k\bullet}V_{n,t})-\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})](\bar{c}_n-c_{n,t})=o_{P^n_{\theta_n}}(n^{-1/2})$;
4. $\frac{1}{n}\sum_{t=1}^{n}\phi_k(A_{n,k\bullet}V_{n,t})(\bar{X}_n-\bar{c}_n)=o_{P^n_{\theta_n}}(n^{-1/2})$;
5. $\frac{1}{n}\sum_{t=1}^{n}\phi_k(A_{n,k\bullet}V_{n,t})(\bar{c}_n-c_{n,t})=o_{P^n_{\theta_n}}(n^{-1/2})$;
6. $\frac{1}{n}\sum_{t=1}^{n}\bar{X}_n[\tilde{\varsigma}_{k,n,1}A_{n,k\bullet}V_{n,t}+\tilde{\varsigma}_{k,n,2}\kappa(A_{n,k\bullet}V_{n,t})]=o_{P^n_{\theta_n}}(n^{-1/2})$;
7. $\frac{1}{n}\sum_{t=1}^{n}(\bar{X}_n-\bar{c}_n)[\varsigma_{k,1}A_{n,k\bullet}V_{n,t}+\varsigma_{k,2}\kappa(A_{n,k\bullet}V_{n,t})]=o_{P^n_{\theta_n}}(n^{-1/2})$;
8. $\frac{1}{n}\sum_{t=1}^{n}(\bar{c}_n-c_{n,t})[\varsigma_{k,1}A_{n,k\bullet}V_{n,t}+\varsigma_{k,2}\kappa(A_{n,k\bullet}V_{n,t})]=o_{P^n_{\theta_n}}(n^{-1/2})$

1. follows by (the first part of) Lemma 2 applied with $W_{n,t} = X_t - c_{n,t}$. This is mean-zero, independent of all $A_{n,k\bullet}V_{n,s}$ with $s \geq t$ and has uniformly bounded second moments (cf. (S6)).

2. follows by Jensen's inequality, (the second part of) Lemma 2 applied with $W_{n,t} = 1$, (S6), Lemma S14 and Corollary 1.

3. follows by Cauchy–Schwarz, (the second part of) Lemma 2 applied with $W_{n,t} = 1$ and Lemma S17.

For 4., $\frac{1}{\sqrt{n}}\sum_{t=1}^n \phi_k(A_{n,k\bullet}V_{n,t}) = O_{P^n_{\theta_n}}(1)$ by the central limit theorem and $\bar{X}_n - \bar{c}_n = \frac{1}{n}\sum_{t=1}^n [X_t - c_{n,t}] \xrightarrow{P_{\theta_n}} 0$, which follows by (S6), Lemma S14 and Corollary 1.

5. follows by Cauchy–Schwarz, the fact that $\mathbb{E}\,\phi_k(A_{n,k\bullet}V_{n,t})^2 = \mathbb{E}\,\phi_k(\epsilon_{k,t})^2$ is uniformly bounded, hence $\frac{1}{n}\sum_{t=1}^n \phi_k(A_{n,k\bullet}V_{n,t})^2 = O_{P^n_{\theta_n}}(1)$ by Markov's inequality and Lemma S17.

For 6., $\bar{X}_n = O_{P_{\theta_n}}(1)$ by, for example, Markov's inequality and (S6). By the central limit theorem, also $\frac{1}{\sqrt{n}}\sum_{t=1}^n U_t = O_{P^n_{\theta_n}(1)}$ for U_t equal to either $A_{n,k\bullet}V_{n,t}$ or $\kappa(A_{n,k\bullet}V_{n,t})$. The result therefore follows from Lemma S16.

For 7., $\frac{1}{\sqrt{n}}\sum_{t=1}^n U_t = O_{P^n_{\theta_n}(1)}$ for U_t equal to either $A_{n,k\bullet}V_{n,t}$ or $\kappa(A_{n,k\bullet}V_{n,t})$ (as in 6.). Therefore, it suffices to note that $\bar{X}_n - \bar{c}_n \xrightarrow{P_{\theta_n}} 0$, as noted for 4. .

For 8., for U_t equal to either $\varsigma_{k,1}A_{n,k\bullet}V_{n,t}$ or $\varsigma_{k,2}\kappa(A_{n,k\bullet}V_{n,t})$, by Markov's inequality

$$P^n_{\theta_n}\left(\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^n(\bar{c}_n - c_{n,t})U_t\right\| > \varepsilon\right) \leq \varepsilon^{-2}\,\mathbb{E}\,U_t^2\frac{1}{n}\sum_{t=1}^n\|\bar{c}_n - c_{n,t}\|^2 \lesssim \frac{1}{n}\sum_{t=1}^n\|\bar{c}_n - c_{n,t}\|^2 \to 0,$$

by Lemma S17.

To verify (iii), we note that

$$\|\hat{I}_{n,\theta_n} - \tilde{I}_\theta\|_2 \leq \|\hat{I}_{n,\theta_n} - \check{I}_{n,\theta_n}\|_2 + \|\check{I}_{n,\theta_n} - \tilde{I}_{n,\theta_n}\|_2 + \|\tilde{I}_{n,\theta_n} - \tilde{I}_\theta\|_2, \tag{S21}$$

where $\tilde{I}_\theta := \mathbb{E}[\tilde{\ell}_\theta(Y_t, X_t)\tilde{\ell}_\theta(Y_t, X_t)'] = \frac{1}{n}\sum_{t=1}^n \mathbb{E}[\tilde{\ell}_\theta(Y_t, X_t)\tilde{\ell}_\theta(Y_t, X_t)']$ with the expectation taken under G_θ, $\hat{I}_{n,\theta} := \frac{1}{n}\sum_{t=1}^n \hat{\ell}_\theta(Y_t, X_t)\hat{\ell}_\theta(Y_t, X_t)'$ and $\check{I}_{n,\theta} := \frac{1}{n}\sum_{t=1}^n \tilde{\ell}_\theta(Y_t, X_t) \times \tilde{\ell}_\theta(Y_t, X_t)'$. We will show each right-hand side term is $o_{P^n_{\theta_n}}(\nu_n^{1/2})$.

For the first right-hand side term in (S21), let $r \in \{\alpha, \sigma, b\}$ and let l denote an index, we write $\hat{U}_{n,t,r_l} := \hat{\ell}_{\theta_n,r_l}(Y_t, X_t)$, $\tilde{U}_{t,r_l} := \tilde{\ell}_{\theta_n,r_l}(Y_t, X_t)$, and $D_{n,t,r_l} := \hat{\ell}_{\theta_n,r_l}(Y_t, X_t) - \tilde{\ell}_{\theta_n,r_l}(Y_t, X_t)$.

Since it is the absolute value of the $(r,l)-(s,m)$ component of $\hat{I}_{n,\theta_n} - \check{I}_{n,\theta_n}$, it is sufficient to show that $|\frac{1}{n}\sum_{t=1}^n \hat{U}_{n,t,r_l}D_{n,t,s_m} + \frac{1}{n}\sum_{t=1}^n D_{n,t,r_l}\tilde{U}_{t,s_m}| = o_{P^n_{\theta_n}}(\nu_n^{1/2})$ as $n \to \infty$ for any $r, s \in \{\alpha, \sigma, b\}$ and l, m. By Cauchy–Schwarz and Lemma S19,

$$\left|\frac{1}{n}\sum_{t=1}^n D_{n,t,r_l}\tilde{U}_{t,s_m}\right| \leq \left(\frac{1}{n}\sum_{t=1}^n \tilde{U}_{t,s_m}^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^n D_{n,t,r_l}^2\right)^{1/2} = o_{P^n_{\theta_n}}\left(\nu_n^{1/2}\right),$$

$$\left|\frac{1}{n}\sum_{t=1}^n \hat{U}_{n,t,r_l}D_{n,t,s_m}\right| \leq \left(\frac{1}{n}\sum_{t=1}^n \hat{U}_{n,t,r_l}^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^n D_{n,t,s_m}^2\right)^{1/2} = o_{P^n_{\theta_n}}\left(\nu_n^{1/2}\right),$$

for any $(r, l) - (s, m)$. It follows that

$$\left[\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{n,t,r_l}D_{n,t,s_m} + D_{n,t,r_l}\tilde{U}_{t,s_m}\right]^2 \leq 2\left[\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{n,t,r_l}D_{n,t,s_m}\right]^2 + 2\left[\frac{1}{n}\sum_{t=1}^{n}D_{n,t,r_l}\tilde{U}_{t,s_m}\right]^2$$
$$= o_{P^n_{\theta_n}}(\nu_n),$$

and hence $\|\hat{I}_{n,\theta_n} - \check{I}_{n,\theta_n}\|_2 \leq \|\hat{I}_{n,\theta_n} - \check{I}_{n,\theta_n}\|_F = o_{P^n_{\theta_n}}(\nu_n^{1/2})$

For the second right-hand side term in (S21), let $Q^{r,ls}_{l,m,t,n} = \tilde{\ell}_{\theta_n,r_l}(Y_t, X_t)\tilde{\ell}_{\theta_n,s_m}(Y_t, X_t)$, where $r, s \in \{\alpha, \sigma, b\}$ and l, m denote the indices of the components of the efficient scores. Fix any r, s and l, m and note that by the fact that $\tilde{\ell}_{\theta_n}$ has uniformly bounded $2 + \delta/2$ moments under $P^n_{\theta_n}$, Theorem 3 of Saikkonen (2007) and Theorem 1 of Kanaya (2017) together imply that (cf. Lemma S14)

$$\frac{1}{n}\sum_{t=1}^{n}Q^{r,s}_{l,m,t,n} - \mathbb{E}_{\theta_n}Q^{r,s}_{l,m,t,n} = O_{P^n_{\theta_n}}\left(n^{(1/p-1)/2}\right) = o_{P^n_{\theta_n}}\left(\nu_n^{1/2}\right), \quad p \in (1, 1 + \delta/4],$$

hence $\|\check{I}_{n,\theta_n} - \tilde{I}_{n,\theta_n}\|_2 = o_{P^n_{\theta_n}}(\nu_n^{1/2})$.

That the last right-hand side term in (S21) is $o(\nu_n^{1/2})$ follows from the assumed local Lipschitz continuity of the map defining the ζ's, that of each $\beta \mapsto A(\alpha, \sigma)_{k\bullet}$, Theorem 11.11 of Kallenberg (2021) and Lemma S18. □

Lemma S16. *If Assumption* 1 *holds, then* $\|\hat{\varrho}_{k,n} - \varrho_{k,n}\|_2 = o_{P^n_{\tilde{\theta}_n}}(\nu_{n,p}) = o_{P^n_{\theta_n}}(\nu_n^{1/2})$, *where* $\tilde{\theta}_n$ *is as in Lemma* S15 *and* $\varrho \in \{\tau, \varsigma\}$.

Proof. Under $P^n_{\theta_n}$, $A_{n,k\bullet}V_{n,t} \approx \epsilon_{k,t} \sim \eta_k$, for $V_{n,t} := Y_t - B_nX_t$ and $A_n := A(\theta_n)$. Let $w \in \{(0, -2)', (1, 0)'\}$. Since the map $M \mapsto M^{-1}$ is Lipschitz at a positive definite matrix M_0, then for large enough n, with probability approaching one

$$\|\hat{\varrho}_{k,n} - \varrho_{k,n}\|_2 = \left\|\left(\hat{M}^{-1}_{k,n} - M^{-1}_k\right)w\right\|_2 \leq 2\left\|\hat{M}^{-1}_{k,n} - M^{-1}_k\right\|_2 \leq 2C\|\hat{M}_{k,n} - M_k\|_2, \tag{S22}$$

for some positive constant C. By Theorem 2.5.11 in Durrett (2019),

$$\frac{1}{n}\sum_{t=1}^{n}\left[(A_{n,k\bullet}V_{n,t})^3 - \mathbb{E}(A_{n,k\bullet}V_{n,t})^3\right] = o_{P^n_{\theta_n}}\left(n^{\frac{1-p}{p}}\right),$$
$$\frac{1}{n}\sum_{t=1}^{n}\left[(A_{n,k\bullet}V_{n,t})^4 - \mathbb{E}(A_{n,k\bullet}V_{n,t})^4\right] = o_{P^n_{\theta_n}}\left(n^{\frac{1-p}{p}}\right).$$

These together imply that

$$\|\hat{M}_{k,n} - M_k\|_2 \leq \|\hat{M}_{k,n} - M_k\|_F = o_{P^n_{\theta_n}}\left(n^{\frac{1-p}{p}}\right) = o_{P^n_{\theta_n}}(\nu_{n,p}).$$

Combining these convergence rates with equation (S22) yields the result. □

LEMMA S17. *In the setting of Lemma S15, let* $c_{n,t} := \mathbb{E}_{\theta_n} X_t$ *and* $\bar{c}_n := \frac{1}{n}\sum_{t=1}^n c_{n,t}$. *Then*

$$\frac{1}{n}\sum_{t=1}^n \|\bar{c}_n - c_{n,t}\|^2 = O(n^{-1}).$$

PROOF. Since $X_t = (1, Z_{t-1}')'$, it suffices to show that $\frac{1}{n}\sum_{t=1}^n \|\tilde{c}_{n,t} - \frac{1}{n}\sum_{t=1}^n \tilde{c}_{n,t}\|^2 = O(n^{-1})$ for $\tilde{c}_{n,t} := \mathbb{E}_{\theta_n} Z_{t-1}$. Let $\tilde{c}_{n,\infty} := \sum_{j=0}^\infty \mathsf{B}_{\theta_n}^j \mathsf{C}_{\theta_n}$. This converges uniformly in n since under Assumption 1 parts (i) and (iii), the sets $\{\|\mathsf{B}_{\theta_n}\|_2 : n \in \mathbb{N}\} \cup \{\|\mathsf{B}_\theta\|_2\}$ and $\{\|\mathsf{C}_{\theta_n}\|_2 : n \in \mathbb{N}\} \cup \{\|\mathsf{C}_\theta\|_2\}$ are bounded above by $\rho_\star < 1$ and $C_\star < \infty$, respectively. By Jensen's inequality,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^n \left\|\tilde{c}_{n,t} - \frac{1}{n}\sum_{t=1}^n \tilde{c}_{n,t}\right\|^2 &\lesssim \frac{1}{n}\sum_{t=1}^n \|\tilde{c}_{n,t} - \tilde{c}_{n,\infty}\|^2 + \frac{1}{n}\sum_{t=1}^n \left\|\frac{1}{n}\sum_{t=1}^n [\tilde{c}_{n,\infty} - \tilde{c}_{n,t}]\right\|^2 \\ &\le \frac{2}{n}\sum_{t=1}^n \|\tilde{c}_{n,t} - \tilde{c}_{n,\infty}\|^2\end{aligned}$$

so it suffices to show that $n/2$ times the last term is uniformly bounded above. One has

$$\begin{aligned}\sum_{t=1}^n \|\tilde{c}_{n,t} - \tilde{c}_{n,\infty}\|^2 &= \sum_{t=1}^n \left\|\sum_{j=t-1}^\infty \mathsf{B}_{\theta_n}^j \mathsf{C}_{\theta_n} - \mathsf{B}_{\theta_n}^{t-1} Z_0\right\|^2 \\ &\lesssim \sum_{t=1}^n \left\|\sum_{j=t-1}^\infty \mathsf{B}_{\theta_n}^j \mathsf{C}_{\theta_n}\right\|^2 + \sum_{t=1}^n \left\|\mathsf{B}_{\theta_n}^{t-1} Z_0\right\|^2 \\ &\le \sum_{t=1}^n \left[\sum_{j=t-1}^\infty \|\mathsf{B}_{\theta_n}\|_2^j \|\mathsf{C}_{\theta_n}\|_2\right]^2 + \sum_{t=1}^n \|B_{\theta_n}\|_2^{2(t-1)} \|Z_0\|^2 \\ &\le C_\star^2 \sum_{t=1}^n \left[\frac{\rho_\star^{t-1}}{1-\rho_\star}\right]^2 + \|Z_0\|^2 \sum_{t=1}^n \rho_\star^{2(t-1)} \\ &\le \left[\frac{C_\star^2}{(1-\rho_\star)^2} + \|Z_0\|^2\right] \frac{1}{1-\rho_\star^2}.\end{aligned}$$

□

LEMMA S18. *In the setting of Lemma S15, let* $\tilde{X}_t = (1, \tilde{Y}_{t-1}', \ldots, \tilde{Y}_{t-p}')'$ *where* $\tilde{Y}_t$ *is a stationary solution to* (1). *Then*:

(i) $\frac{1}{n}\sum_{t=1}^n \mathbb{E}_{\theta_n} X_t - \mathbb{E}_\theta \tilde{X}_t = o(\nu_n^{1/2})$,

(ii) $\frac{1}{n}\sum_{t=1}^n [\mathbb{E}_{\theta_n} X_t][\mathbb{E}_{\theta_n} X_t]' - [\mathbb{E}_\theta \tilde{X}_t][\mathbb{E}_\theta \tilde{X}_t]' = o(\nu_n^{1/2})$.

(iii) $\frac{1}{n}\sum_{t=1}^n \mathbb{E}_{\theta_n}[X_t - \mathbb{E}_{\theta_n} X_t][X_t - \mathbb{E}_{\theta_n} X_t]' - \mathbb{E}_\theta[X_t - \mathbb{E}_\theta X_t][X_t - \mathbb{E}_\theta X_t]' = o(\nu_n^{1/2})$.

PROOF. Note that $\|\mathbb{E}_{\theta_n} X_t - \mathbb{E}_{\theta_n} \tilde{X}_t\|^2 \le \|\tilde{c}_{n,t} - \tilde{c}_{n,\infty}\|^2$ in the notation of (the proof of) Lemma S17, which shows that $\frac{1}{n}\sum_{t=1}^n \|\tilde{c}_{n,t} - \tilde{c}_{n,\infty}\|^2 = O(n^{-1})$. Hence, by Jensen's in-

equality,

$$\frac{1}{n}\sum_{t=1}^{n}\|\mathbb{E}_{\theta_n}X_t-\mathbb{E}_{\theta_n}\tilde{X}_t\|=O\big(n^{-1/2}\big)=o\big(\nu_n^{1/2}\big),$$

Since $\beta\mapsto\mathbb{E}_\theta\tilde{X}_t=\mathrm{vec}(\iota_K,(\iota_p\otimes(I_K-B_1-\cdots-B_p)^{-1}c))$ is locally Lipschitz,

$$\frac{1}{n}\sum_{t=1}^{n}\|\mathbb{E}_{\theta_n}\tilde{X}_t-\mathbb{E}_{\theta}\tilde{X}_t\|=\|\mathbb{E}_{\theta_n}\tilde{X}_t-\mathbb{E}_{\theta}\tilde{X}_t\|=O\big(n^{-1/2}\big)=o\big(\nu_n^{1/2}\big).$$

Combination of the above two displays yields that $\frac{1}{n}\sum_{t=1}^{n}\|\mathbb{E}_{\theta_n}X_t-\mathbb{E}_{\theta}\tilde{X}_t\|=O(n^{-1/2})=o(\nu_n^{1/2})$ which implies (i). Moreover, combined with the uniform moment bounds given in (S6) and Lemma S1 this yields

$$\frac{1}{n}\sum_{t=1}^{n}\big\|[\mathbb{E}_{\theta_n}X_t][\mathbb{E}_{\theta_n}X_t]'-[\mathbb{E}_{\theta}\tilde{X}_t][\mathbb{E}_{\theta}\tilde{X}_t]'\big\|\lesssim\frac{1}{n}\sum_{t=1}^{n}\|\mathbb{E}_{\theta_n}X_t-\mathbb{E}_{\theta}\tilde{X}_t\|=O\big(n^{-1/2}\big)=o\big(\nu_n^{1/2}\big),$$

which implies (ii).

For (iii), let $U_{\vartheta,t}:=X_t-\mathbb{E}_\vartheta X_t$ and $\tilde{U}_{\vartheta,t}:=\tilde{X}_t-\mathbb{E}_\vartheta\tilde{X}_t$. Note that as $U_{\vartheta,t}=\sum_{j=0}^{t-2}\mathsf{B}_\vartheta^j\mathsf{D}_\vartheta\epsilon_{t-j}$ and $\tilde{U}_{\vartheta,t}=\sum_{j=0}^{\infty}\mathsf{B}_\vartheta^j\mathsf{D}_\vartheta\epsilon_{t-j}$, $U_{\theta_n,t}-\tilde{U}_{\theta_n,t}$, and $U_{\theta_n,t}$ are independent. Additionally by Assumption 1 parts 1 and 3, the sets $\{\|\mathsf{B}_{\theta_n}\|_2:n\in\mathbb{N}\}$ and $\{\|\mathsf{D}_{\theta_n}\|_2:n\in\mathbb{N}\}$ are bounded above by $\rho_\star<1$ and $D_\star<\infty$, respectively. Hence,

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\big\|\mathbb{E}_{\theta_n}\big[U_{\theta_n,t}U_{\theta_n,t}'-\tilde{U}_{\theta_n,t}\tilde{U}_{\theta_n,t}'\big]\big\|\\
&\quad\le\frac{1}{n}\sum_{t=1}^{n}\big\|\mathbb{E}_{\theta_n}\big[(U_{\theta_n,t}-\tilde{U}_{\theta_n,t})U_{\theta_n,t}'\big]\big\|+\frac{1}{n}\sum_{t=1}^{n}\big\|\mathbb{E}_{\theta_n}\big[(U_{\theta_n,t}-\tilde{U}_{\theta_n,t})\tilde{U}_{\theta_n,t}'\big]\big\|\\
&\quad\le\frac{1}{n}\sum_{t=1}^{n}\Big\|\mathbb{E}_{\theta_n}\sum_{k=0}^{\infty}\sum_{j=t-1}^{\infty}\mathsf{B}_{\theta_n}^j\mathsf{D}_{\theta_n}\epsilon_{t-j}\epsilon_{t-k}'\mathsf{D}_{\theta_n}'\big(\mathsf{B}_{\theta_n}^j\big)'\Big\|\\
&\quad\le\frac{1}{n}\sum_{t=1}^{n}\sum_{j=t-1}^{\infty}\|\mathsf{B}_{\theta_n}\|_2^{2j}\|\mathsf{D}_{\theta_n}\|_2^2\\
&\quad\le D_\star^2\times\frac{1}{n}\sum_{t=1}^{n}\sum_{j=t-1}^{\infty}\rho_\star^{2j}\\
&\quad\le\frac{D_\star^2}{1-\rho_\star^2}\times\frac{1-\rho_\star^{2n}}{1-\rho_\star^2}\times\frac{1}{n}\\
&\quad=O\big(n^{-1}\big).
\end{aligned}$$

Additionally, we can write $\operatorname{vec}(\mathbb{E}_\vartheta\, \tilde{U}_{\vartheta,t}\tilde{U}'_{\vartheta,t}) = (I - \mathsf{B}_\vartheta \otimes \mathsf{B}_\vartheta)^{-1}\operatorname{vec}(\mathsf{D}_\vartheta \mathsf{D}'_\vartheta)$, which is locally Lipschitz in β at θ. This implies that

$$\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}_{\theta_n}\, \tilde{U}_{\theta_n,t}\tilde{U}'_{\theta_n,t} - \mathbb{E}_\theta\, \tilde{U}_{\theta,t}\tilde{U}'_{\theta,t} = O\big(n^{-1/2}\big) = o\big(\nu_n^{1/2}\big).$$

The previous two displays suffice for (iii). □

LEMMA S19. *In the setting of Lemma S15, for each $r \in \{\alpha, \sigma, b\}$ and l,*

$$\frac{1}{n}\sum_{t=1}^{n}\big(\hat{\ell}_{\tilde{\theta}_n,r_l}(Y_t, X_t) - \tilde{\ell}_{\tilde{\theta}_n,r_l}(Y_t, X_t)\big)^2 = o_{P^n_{\tilde{\theta}_n}}(\nu_n).$$

PROOF. We start by considering elements in $\frac{1}{n}\sum_{t=1}^{n}(\hat{\ell}_{\tilde{\theta}_n,\alpha_l}(Y_t, X_t) - \tilde{\ell}_{\tilde{\theta}_n,\alpha_l}(Y_t, X_t))^2$. Define $\tilde{\tau}_{k,n,q} := \hat{\tau}_{k,n,q} - \tau_{k,q}$ and $V_{n,t} = Y_t - B_n X_t$. Since each $|\zeta^{\alpha}_{n,l,k,j}| < \infty$ and the sums over k, j are finite, it is sufficient to demonstrate that for every k, j, m, $s \in [K]$, with $k \neq j$ and $s \neq m$,

$$\frac{1}{n}\sum_{t=1}^{n}\big[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\big]\big[\hat{\phi}_{s,n}(A_{n,s\bullet}V_{n,t}) - \phi_s(A_{n,s\bullet}V_{n,t})\big] \times A_{n,j\bullet}V_{t,n}A_{n,m\bullet}V_{n,t}, \tag{S23}$$

$$\frac{1}{n}\sum_{t=1}^{n}\big[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\big] \times A_{n,j\bullet}V_{n,t}\big[\tilde{\tau}_{s,n,1}A_{n,s\bullet}V_{n,t} + \tilde{\tau}_{s,n,2}\kappa(A_{n,s\bullet}V_{n,t})\big], \tag{S24}$$

$$\frac{1}{n}\sum_{t=1}^{n}\big[\tilde{\tau}_{s,n,1}A_{n,s\bullet}V_{n,t} + \tilde{\tau}_{s,n,2}\kappa(A_{n,s\bullet}V_{n,t})\big]\big[\tilde{\tau}_{k,n,1}A_{n,k\bullet}V_{n,t} + \tilde{\tau}_{k,n,2}\kappa(A_{n,k\bullet}V_{n,t})\big] \tag{S25}$$

are each $o_{P^n_{\tilde{\theta}_n}}(\nu_n)$.

For (S25), let $\xi_1(x) = x$ and $\xi_2(x) = \kappa(x)$. Then we can split the sum into 4 parts, each of which has the following form for some q, $w \in \{1, 2\}$:

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\tilde{\tau}_{s,n,q}\tilde{\tau}_{k,n,w}\xi_q(A_{n,s\bullet}V_{n,t})\xi_w(A_{n,k\bullet}V_{n,t}) \\
&\quad= \tilde{\tau}_{s,n,q}\tilde{\tau}_{k,n,w}\frac{1}{n}\sum_{t=1}^{n}\xi_q(A_{n,s\bullet}V_{n,t})\xi_w(A_{n,k\bullet}V_{n,t}) \\
&\quad= o_{P^n_{\tilde{\theta}_n}}(\nu_n),
\end{aligned}$$

since we have that each $\tilde{\tau}_{s,n,q}\tilde{\tau}_{k,n,w} = o_{P^n_{\tilde{\theta}_n}}(\nu_n)$ by Lemma S16.[S12] For (S24), we can argue similarly. Again, let $\xi_1(x) = x$ and $\xi_2(x) = \kappa(x)$. Then we can split the sum into 2 parts, each of which has the following form for some $q \in \{1, 2\}$:

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\big[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\big]A_{n,j\bullet}V_{n,t}\tilde{\tau}_{s,n,q}\xi_q(A_{n,s\bullet}V_{n,t}) \\
&\quad\le \tilde{\tau}_{s,n,q}\left(\frac{1}{n}\sum_{t=1}^{n}\big[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\big]^2(A_{n,j\bullet}V_{n,t})^2\right)^{1/2} \\
&\qquad\times\left(\frac{1}{n}\sum_{t=1}^{n}\xi_q(A_{n,s\bullet}V_{n,t})^2\right)^{1/2} \\
&\quad= o_{P^n_{\tilde{\theta}_n}}(\nu_n).
\end{aligned}$$

by Lemma 2 applied with $W_{n,t} = A_{n,j\bullet}V_{n,t}$ and $\tilde{\tau}_{s,n,q} = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2})$.[S13] For (S23), use Cauchy–Schwarz with Lemma 2,

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\big[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\big]\big[\hat{\phi}_{s,n}(A_{n,s\bullet}V_{n,t}) - \phi_s(A_{n,s\bullet}V_{n,t})\big] \\
&\qquad\times A_{n,j\bullet}V_{n,t}A_{n,m\bullet}V_{n,t} \\
&\quad\le\left(\frac{1}{n}\sum_{t=1}^{n}\big[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\big]^2(A_{n,j\bullet}V_{n,t})^2\right)^{1/2} \\
&\qquad\times\left(\frac{1}{n}\sum_{t=1}^{n}\big[\hat{\phi}_{s,n}(A_{n,s\bullet}V_{n,t}) - \phi_s(A_{n,s\bullet}V_{n,t})\big]^2(A_{n,m\bullet}V_{n,t})^2\right)^{1/2} \\
&\quad= o_{P^n_{\tilde{\theta}_n}}(\nu_n).
\end{aligned}$$

This completes the proof for the components corresponding to α_l. We note that the components corresponding to σ_l follow analogously.

Finally, we consider the elements in $\frac{1}{n}\sum_{t=1}^{n}(\hat{\ell}_{\theta_n,b_l}(Y_t, X_t) - \tilde{\ell}_{\theta_n,b_l}(Y_t, X_t))^2$. Let $a_{n,k,l} := -A_{n,k\bullet}D_{b_l}$, $\tilde{\mathfrak{s}}_{k,n} := \hat{\mathfrak{s}}_{k,n} - \mathfrak{s}_k$, $c_{n,t} := \mathbb{E}_{\theta_n}X_t$ and $\bar{c}_n := \frac{1}{n}\sum_{t=1}^{n}c_{n,t}$. Since $a_{n,k,l} \to a_{\infty,k,l} := A(\alpha,\sigma)_{k\bullet}D_{b_l}(\alpha,\sigma)$, it suffices to show that:

(i) $\frac{1}{n}\sum_{t=1}^{n}[\phi_k(A_{n,k\bullet}V_{n,t}) - \hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})]^2\|X_t - c_{n,t}\|^2 = o_{P^n_{\theta_n}}(\nu_n)$;

(ii) $\frac{1}{n}\sum_{t=1}^{n}[\phi_k(A_{n,k\bullet}V_{n,t}) - \hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})]^2\|\bar{X}_n - \bar{c}_n\|^2 = o_{P^n_{\theta_n}}(\nu_n)$;

[S12] The fact that $\frac{1}{n}\sum_{t=1}^{n}\xi_q(A_{n,s\bullet}V_{n,t})\xi_w(A_{n,k\bullet}V_{n,t}) = O_{P^n_{\tilde{\theta}_n}}(1)$ can be seem to hold using the moment and i.i.d. assumptions from Assumption 1 and Markov's inequality, noting once more that $A_{n,k\bullet}V_{n,t} \simeq \epsilon_{k,t}$ under $P^n_{\tilde{\theta}_n}$.

[S13] See footnote S12.

(iii) $\frac{1}{n}\sum_{t=1}^{n}[\phi_k(A_{n,k\bullet}V_{n,t}) - \hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})]^2\|\bar{c}_n - c_{n,t}\|^2 = o_{P^n_{\theta_n}}(\nu_n)$;

(iv) $\frac{1}{n}\sum_{t=1}^{n}\phi_k(A_{n,k\bullet}V_{n,t})^2\|\bar{X}_n - \bar{c}_n\|^2 = o_{P^n_{\theta_n}}(\nu_n)$;

(v) $\frac{1}{n}\sum_{t=1}^{n}\phi_k(A_{n,k\bullet}V_{n,t})^2\|\bar{c}_n - c_{n,t}\|^2 = o_{P^n_{\theta_n}}(\nu_n)$;

(vi) $\frac{1}{n}\sum_{t=1}^{n}\|\bar{X}_n\|^2[\tilde{\varsigma}_{k,n,1}A_{n,k\bullet}V_{n,t} + \tilde{\varsigma}_{k,n,2}\kappa(A_{n,k\bullet}V_{n,t})]^2 = o_{P^n_{\theta_n}}(\nu_n)$;

(vii) $\frac{1}{n}\sum_{t=1}^{n}\|\bar{X}_n - \bar{c}_n\|^2[\varsigma_{k,1}A_{n,k\bullet}V_{n,t} + \varsigma_{k,2}\kappa(A_{n,k\bullet}V_{n,t})]^2 = o_{P^n_{\theta_n}}(\nu_n)$;

(viii) $\frac{1}{n}\sum_{t=1}^{n}\|\bar{c}_n - c_{n,t}\|^2[\varsigma_{k,1}A_{n,k\bullet}V_{n,t} + \varsigma_{k,2}\kappa(A_{n,k\bullet}V_{n,t})]^2 = o_{P^n_{\theta_n}}(\nu_n)$.

(i) follows from repeated application of Lemma 2 with $W_{n,t} = e_j'(X_t - c_{n,t})$.

(ii) follows from application of Lemma 2 with $W_{n,t} = 1$ and $\bar{X}_n - \bar{c}_n = \frac{1}{n}\sum_{t=1}^{n}[X_t - c_{n,t}] \xrightarrow{P_{\theta_n}} 0$, which follows by (S6), Lemma S14, and Corollary 1.

(iii) follows by Lemma 2 applied repeatedly with $W_{n,t} = e_j'(\bar{c}_n - c_{n,t})$.[S14]

For (iv), $\frac{1}{n}\sum_{t=1}^{n}\phi_k(A_{n,k\bullet}V_{n,t})^2 = O_{P^n_{\theta_n}}(1)$ since $\phi_k(A_{n,k\bullet}V_{n,t})^2$ has uniformly bounded second moments and $\bar{X}_n - \bar{c}_n = O_{P^n_{\theta_n}}(n^{-1/2})$, by (S6), Lemma S14 and Corollary 1.

For (v), use Markov's inequality and Lemma S17 to conclude

$$P^n_{\theta_n}\left(\frac{1}{n}\sum_{t=1}^{n}\phi_k(A_{n,k\bullet}V_{n,t})^2\|\bar{c}_n - c_{n,t}\|^2 > \nu_n\varepsilon\right) \leq \nu_n^{-1}\varepsilon^{-1}\,\mathbb{E}\big[\phi_k(\epsilon_k)^2\big]\frac{1}{n}\sum_{t=1}^{n}\|\bar{c}_n - c_{n,t}\|^2 \to 0.$$

For (vi), $\bar{X}_n = O_{P_{\theta_n}}(1)$ by, for example, Markov's inequality and (S6). Similarly, $\frac{1}{n}\sum_{t=1}^{n}U_{t,i}U_{t,j} = O_{P^n_{\theta_n}(1)}$ for $i, j \in \{1, 2\}$ with $U_{t,1} = A_{n,k\bullet}V_{n,t}$ and $U_{t,2} = \kappa(A_{n,k\bullet}V_{n,t})$. The result then follows from Lemma S16.

For (vii), $\frac{1}{n}\sum_{t=1}^{n}U_{t,i}U_{t,j} = O_{P^n_{\theta_n}(1)}$ for $i, j \in \{1, 2\}$ with $U_{t,1}$ and $U_{t,2}$ as in the preceding paragraph. Therefore, it suffices to note that $\bar{X}_n - \bar{c}_n = O_{P_{\theta_n}}(n^{-1/2})$, as noted for (iv).

For (viii), for $U_{t,1}$ and $U_{t,2}$ as in the preceding paragraph and $i, j \in \{1, 2\}$,

$$\begin{aligned}
&P^n_{\theta_n}\left(\left|\frac{1}{n}\sum_{t=1}^{n}\|\bar{c}_n - c_{n,t}\|^2\varsigma_{k,i}U_{t,i}\varsigma_{k,j}U_{t,j}\right| > \nu_n\varepsilon\right)\\
&\leq \nu_n^{-1}\varepsilon^{-1}|\varsigma_{k,i}\varsigma_{k,j}|\big[\mathbb{E}\,U_{t,i}^2\big]^{1/2}\big[\mathbb{E}\,U_{t,j}^2\big]^{1/2}\frac{1}{n}\sum_{t=1}^{n}\|\bar{c}_n - c_{n,t}\|^2\\
&\lesssim \nu_n^{-1}\frac{1}{n}\sum_{t=1}^{n}\|\bar{c}_n - c_{n,t}\|^2 \to 0,
\end{aligned}$$

by Markov's inequality and Lemma S17. □

[S14] That this is uniformly bounded follows from (S6).

SB.5 *Assumption 2.1(ii)(b)*

We provide a sufficient condition under which Assumption 1 part (ii)(b) holds, given part (ii)(a). For convenience, recall that part (ii) reads as:

(ii) Conditional on the initial values $(Y'_{-p+1}, \ldots, Y'_0)'$, $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{K,t})'$ is independently and identically distributed across t, with independent components $\epsilon_{k,t}$. Each $\eta = (\eta_1, \ldots, \eta_K) \in \mathcal{H}$ is such that each η_k is nowhere vanishing, dominated by Lebesgue measure on $\mathbb{R}$, continuously differentiable with log density scores denoted by $\phi_k(z) := \partial \log \eta_k(z)/\partial z$, and for all $k = 1, \ldots, K$,

(a) $\mathbb{E}\epsilon_{k,t} = 0, \mathbb{E}\epsilon_{k,t}^2 = 1, \mathbb{E}\epsilon_{k,t}^{4+\delta} < \infty, \mathbb{E}(\epsilon_{k,t}^4) - 1 > \mathbb{E}(\epsilon_{k,t}^3)^2$, and $\mathbb{E}\phi_k^{4+\delta}(\epsilon_{k,t}) < \infty$ (for some $\delta > 0$);

(b) $\mathbb{E}\,\phi_k(\epsilon_{k,t}) = 0$, $\mathbb{E}\,\phi_k^2(\epsilon_{k,t}) > 0$, $\mathbb{E}\,\phi_k(\epsilon_{k,t})\epsilon_{k,t} = -1$, $\mathbb{E}\,\phi_k(\epsilon_{k,t})\epsilon_{k,t}^2 = 0$, and $\mathbb{E}\,\phi_k(\epsilon_{k,t})\epsilon_{k,t}^3 = -3$;

In this assumption part (a) is standard—only imposes that the shocks are mean zero with unit variance, and that certain $4 + \delta$ moments are finite. In contrast, part (b) may seem strong at first sight.

An important observation is that (b) should not be understood independently from (a). Indeed, the following lemma shows that given (a), condition (b) follows if the structural shocks have densities that decays to zero at a polynomial rate.

LEMMA S20. *Let* $a_k = \inf\{x \in \mathbb{R} \cup \{-\infty\} : \eta_k(x) > 0\}$ *and* $b_k = \sup\{x \in \mathbb{R} \cup \{\infty\} : \eta_k(x) > 0\}$. *Suppose that, for* $r = 0, 1, 2, 3$: (*i*) *if* $a_k = -\infty$, *then* $\eta_k(x) = o(x^{-3})$ *as* $x \to -\infty$, *else* $a_k^r \lim_{x \downarrow a_k} \eta_k(x) = 0$, *and* (*ii*) *if* $b_k = \infty$ *then* $\eta_k(x) = o(x^{-3})$ *as* $x \to \infty$, *else* $b_k^r \lim_{x \uparrow b_k} \eta_k(x) = 0$. *Then, if part* (*a*) *of Assumption* 1(*ii*) *holds, part* (*b*) *is also satisfied.*

PROOF. Let $r \in \{0, 1, 2, 3\}$, $b_k = \sup\{x \in \mathbb{R} : \eta_k(x) > 0\}$ and $a_k = \inf\{x \in \mathbb{R} : \eta_k(x) > 0\}$. We have, by integration by parts, with G_k denoting the measure on $\mathbb{R}$ corresponding to η_k,

$$\int \phi_k(z) z^r \, \mathrm{d}G_k = \int \frac{\eta_k'(z)}{\eta_k(z)} \eta_k(z) z^r \, \mathrm{d}z = \int \eta_k'(z) z^r \, \mathrm{d}z = \eta_k(z) z^r |_{a_k}^{b_k} - \int \eta_k(z) \frac{\mathrm{d}z^r}{\mathrm{d}z} \, \mathrm{d}z.$$

Our hypothesis ensures that $z^r \eta_k(z)|_{a_k}^{b_k} = 0$. Therefore, we have $G_k \phi_k(z) z^r = -G_k \frac{\mathrm{d}}{\mathrm{d}z} z^r$. For $r = 0$, this equals zero as $\frac{\mathrm{d}}{\mathrm{d}z} z^0 = \frac{\mathrm{d}}{\mathrm{d}z} 1 = 0$. For $r \in \{1, 2, 3\}$, we have $\frac{\mathrm{d}z^r}{\mathrm{d}z} = r z^{r-1}$, and hence $G_k \phi_k(z) z^r = -r G_k z^{r-1}$. Since $G_k 1 = 1$, $G_k z = 0$, and $G_k z^2 = 1$, the result follows. □

We now provide two examples. The first is a mixture of normals. We directly verify the moment conditions in (a) and (b) are satisfied.

The second example is a normalized χ_2^2 distribution. We show that this does satisfy the moment conditions in (a) but not those in (b) (nor the conditions of Lemma S20).[S15]

[S15] Additionally, the (normalized) χ_2^2 distribution does not have a nowhere vanishing Lebesgue density.

EXAMPLE S3 (Normal mixtures). Suppose that ϵ_k has the density function

$$\eta_k(z) = \sum_{m=1}^{M} p_m f_m\big(z, \mu_m, \sigma_m^2\big),$$

$$p_m \geq 0, \qquad \sum_{m=1}^{M} p_m = 1, \qquad \sum_{m=1}^{M} p_m \mu_m = 0, \qquad \sum_{m=1}^{M} p_m\big(\sigma_m^2 + \mu_m^2\big) = 1,$$

where $f_m(z, \mu_m, \sigma_m^2)$ is the density function of a $e_m \sim \mathcal{N}(\mu_m, \sigma_m^2)$.

ϵ_k has mean zero and unit variance. We first establish that each of the conditions in (a) are satisfied. In particular, we first note that $\mathbb{E}[|\epsilon_k|^r]$ is finite for any positive integer r as

$$\mathbb{E}\big[|\epsilon_k|^r\big] = \sum_{m=1}^{M} p_m \mathbb{E}\big[|e_m|^r\big] < \infty, \tag{S26}$$

since the Normal distribution has finite moments of all orders. To establish that $\mathbb{E}[\epsilon_k^3]^2 < \mathbb{E}[\epsilon_k^4] - 1$, note that this is equivalent to the linear independence in L_2 of 1, ϵ_k, ϵ_k^2 (e.g., Horn and Johnson (2013, Theorem 7.2.10)). This is equivalent to the condition that

$$a_1^2 + 2a_1a_3 + a_2^2 + a_3^2\,\mathbb{E}\big[\epsilon_k^4\big] = 0 \quad \Longrightarrow \quad a_1 = a_2 = a_3 = 0.$$

This holds since $\mathbb{E}[\epsilon_k^4] \geq 1 = \mathbb{E}[\epsilon_k^2]$ by the fact that L_p norms are increasing and so

$$a_1^2 + 2a_1a_3 + a_2^2 + a_3^2\,\mathbb{E}\big[\epsilon_k^4\big] \geq a_1^2 + 2a_1a_3 + a_3^2 = (a_1 + a_3)^2 \geq 0,$$

where equality is possible only if $a_1 = a_2 = a_3 = 0$. Next, note that

$$\phi_k(z) = -\frac{\sum_{m=1}^{M} p_m \sigma_m^{-2}(z - \mu_m) f_m\big(z, \mu_m, \sigma_m^2\big)}{\eta_k(z)}, \tag{S27}$$

and for any integer r and some $\mu \in \mathbb{R}$,

$$\begin{aligned}\big|\phi_k(z)\big|^r &\lesssim \big|\phi_k(z)\big|^{r-1}\left|\eta_k(z)^{-1}\big(|z| + |\mu|\big)\sum_{m=1}^{M} p_m f_m\big(z, \mu_m, \sigma_m^2\big)\right| \\ &= \big|\phi_k(z)\big|^{r-1}\big(|z| + |\mu|\big).\end{aligned}$$

Recursively using this inequality from $r = 0$, yields (for some constant $C_r \in (0, \infty)$)

$$\big|\phi_k(z)\big|^r \leq C_r\big(|z|^r + |\mu|^r\big).$$

That $\mathbb{E}\,|\phi(\epsilon_k)|^r < \infty$ for any integer r then follows from (S26).

For the conditions in (b), note that by (S27),

$$\begin{aligned}\mathbb{E}\big[\phi_k(\epsilon_k)\epsilon_k^r\big] &= -\sum_{m=1}^{M} p_m \int z^r \frac{\sigma_m^{-2}(z-\mu_m) f_m\big(\epsilon_k, \mu_m, \sigma_m^2\big)}{\eta_k(z)} \eta_k(z)\,\mathrm{d}z \\ &= -\sum_{m=1}^{M} p_m \sigma_m^{-2} \int z^r (z-\mu_m) f_m\big(\epsilon_k, \mu_m, \sigma_m^2\big)\,\mathrm{d}z \\ &= -\sum_{m=1}^{M} p_m \sigma_m^{-2} \big(\mathbb{E}\big[e_m^{r+1}\big] - \mathbb{E}\big[e_m^r\big]\mu_m\big).\end{aligned}$$

Taking $r = 0, 1, 2, 3$ in the right-hand expression respectively gives:

$$\begin{aligned}\mathbb{E}\big[\phi_k(\epsilon_k)\big] &= -\sum_{m=1}^{M} p_m \sigma_m^{-2}(\mu_m - \mu_m) = 0, \\ \mathbb{E}\big[\phi_k(\epsilon_k)\epsilon_k\big] &= -\sum_{m=1}^{M} p_m \sigma_m^{-2}\big(\sigma_m^2 + \mu_m^2 - \mu_m^2\big) = -1, \\ \mathbb{E}\big[\phi_k(\epsilon_k)\epsilon_k^2\big] &= -\sum_{m=1}^{M} p_m \sigma_m^{-2}\big(\mu_m^3 + 3\mu_m\sigma_m^2 - \big(\sigma_m^2 + \mu_m^2\big)\mu_m\big) = 0, \\ \mathbb{E}\big[\phi_k(\epsilon_k)\epsilon_k^3\big] &= -\sum_{m=1}^{M} p_m \sigma_m^{-2}\big(\mu_m^4 + 6\mu_m^2\sigma_m^2 + 3\sigma_m^4 - \mu_m^4 - 3\mu_m^2\sigma_m^2\big) = -3.\end{aligned}$$

Example S4 (The normalized χ_2^2 distribution). Suppose that $\tilde{\epsilon}_k \sim \chi_2^2$ and let $\epsilon_k = (\tilde{\epsilon}_k - 2)/2$. Then ϵ_k has mean zero, variance one, and density function $\eta_k(z) = \exp(-z-1)$ on its support $[-1, \infty)$ on which we also have that $\phi_k(z) = -1$. The χ_2^2 distribution has finite moments of all orders and has moment generating function (e.g., Johnson, Kotz, and Balakrishnan (1995, p. 420))

$$M_{\tilde{\epsilon}}(t) = (1-2t)^{-1}, \quad t < 1/2.$$

Hence, ϵ_k has finite moments of all orders. The same is evidently true of $\phi_k(\epsilon_k) = -1$. Using the above display, we have

$$M_{\epsilon}(t) = e^{-t}(1-t)^{-1}, \quad t < 1$$

and, therefore, may directly calculate $\mathbb{E}[\epsilon_k^3] = 2$ and $\mathbb{E}[\epsilon_k^4] = 9$, hence $\mathbb{E}[\epsilon_k^3]^2 < \mathbb{E}[\epsilon_k^4] - 1$ holds. The moment conditions in part (a) are therefore all satisfied.

However, $\mathbb{E}\,\phi_k(z) = -1 \neq 0$, hence part (b) does not hold. Note also that this example does not satisfy the requirements of Lemma S20: we have $a_k = -1$, $b_k = \infty$, and

$$\lim_{z \downarrow a_k} \eta_k(x) = \lim_{z \downarrow -1} \exp(-z-1) = 1 \neq 0,$$

and hence the required condition is violated for $r = 0$.

Appendix SC: Technical tools

This section records some technical tools used in the proofs for ease of reference.

Lemma S21 (Discretization). *Suppose that P_n is a sequence of probability measures and $f_n : \Gamma \to \mathbb{R}$, $\Gamma \subset \mathbb{R}^L$ is a sequence of functions, which satisfy*

$$f_n(\gamma_n) \xrightarrow{P_n} 0 \tag{S28}$$

for any $\gamma_n := \gamma + g_n/\sqrt{n}$, $g_n \to g \in \mathbb{R}^L$. Suppose that the estimator sequence $\bar{\gamma}_n$ satisfies $\sqrt{n}\|\bar{\gamma}_n - \gamma\| = O_{P_n}(1)$ and $\bar{\gamma}_n$ takes values in $\mathscr{S}_n := \{CZ/\sqrt{n} : Z \in \mathbb{R}^L\}$ for some $L \times L$ matrix C. Then

$$f_n(\bar{\gamma}_n) \xrightarrow{P_n} 0.$$

Proof. Since $\bar{\gamma}_n$ is $\sqrt{n}$-consistent, there is an $M > 0$ such that $P_n(\sqrt{n}\|\bar{\gamma}_n - \gamma\| > M) < \varepsilon$. If $\sqrt{n}\|\bar{\gamma}_n - \gamma\| \le M$, then $\bar{\gamma}$ is equal to one of the values in the finite set $\mathscr{S}_n^c = \{\gamma^* \in \mathscr{S}_n : \|\gamma^* - \gamma\| \le n^{-1/2}M\}$. For each M, this set has finite number of elements bounded independently of n, call this upper bound $\overline{B}$. For any $\upsilon > 0$,

$$\begin{aligned} P_n\big(|f_n(\bar{\gamma}_n)| > \upsilon\big) &\le \varepsilon + \sum_{\gamma_n \in \mathscr{S}_n^c} P_n\big(\{|f_n(\gamma_n)| > \upsilon\} \cap \{\bar{\gamma}_n = \gamma_n\}\big) \\ &\le \varepsilon + \sum_{\gamma_n \in \mathscr{S}_n^c} P_n\big(|f_n(\gamma_n)| > \upsilon\big) \\ &\le \varepsilon + \overline{B} P_n\big(|f_n(\gamma_n^\star)| > \upsilon\big), \end{aligned}$$

where $\gamma_n^\star \in \mathscr{S}_n^c$ maximizes $\gamma \mapsto P_n(|f_n(\gamma)| > \upsilon)$. As $\gamma_n^\star \in \mathscr{S}_n^c$, $\|\gamma^\star - \gamma\| \le n^{-1/2}M$. Hence, letting $g_n := \sqrt{n}(\gamma_n^\star - \gamma)$, $\|g_n\| \le M$. Arguing along subsequences if necessary, we may therefore assume that $g_n \to g \in \mathbb{R}^L$, and hence $f_n(\gamma_n^\star) \xrightarrow{P_n} 0$ by (S28). The proof is complete on combining this with the previously established bound on $P_n(|f_n(\bar{\gamma}_n)| > \upsilon)$. □

Lemma S22. *Let $(X, \mathcal{B}(X))$ be a measurable space, and Q_n a sequence of probability measures on $(X, \mathcal{B}(X))$, which converges to a probability measure Q in total variation. Let $(Y, \mathcal{B}(Y), \lambda)$ be a measure space and suppose that $p_n : X \times Y \to [0, \infty)$ is a sequence of functions and $p : X \times Y \to [0, \infty)$ a function such that (i) $\int p_n(x, y)\,\mathrm{d}\lambda(y) = 1 = \int p(x, y)\,\mathrm{d}\lambda(y)$ for each $n \in \mathbb{N}$ and each $x \in X$ and (ii) $p_n \to p$ pointwise. Then, if G_n and G_n are defined according to*

$$\begin{aligned} G_n(A) &:= \int_A p_n(x, y)\,\mathrm{d}\big(\lambda(y) \otimes Q_n(x)\big); \\ G(A) &:= \int_A p(x, y)\,\mathrm{d}\big(\lambda(y) \otimes Q(x)\big), \end{aligned}$$

it follows that $G_n \xrightarrow{\mathrm{TV}} G$.

PROOF. For any x, $p_n(x,\cdot) \to p(x,\cdot)$ pointwise and since each $p_n(\cdot,x)$, $p(\cdot,x)$ has integral one under λ, by Proposition 2.29 in van der Vaart (1998),

$$\mathscr{Q}_n(x) := \int \left| p_n(x,y) - p(x,y) \right| \mathrm{d}\lambda(y) \to 0,$$

pointwise. Let $(\psi_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions on $X \times Y$ with $\psi_n \in [0,1]$. Then

$$\left| \int \int \psi_n(x,y)\big(p_n(x,y) - p(x,y)\big) \mathrm{d}\lambda(y) \mathrm{d}Q_n(x) \right| \leq \int \mathscr{Q}_n(x) \mathrm{d}Q_n(x).$$

The inequality $\mathscr{Q}_n(x) \leq \int p_n(x,y)\mathrm{d}\lambda(y) + \int p(x,y)\mathrm{d}\lambda(y) = 2$ ensures that the $\mathscr{Q}_n(x)$ are uniformly Q_n-integrable and uniformly Q-integrable. By Theorem 2.8 of Serfozo (1982), $\int \mathscr{Q}_n(x)\mathrm{d}Q_n(x) \to 0$. □

LEMMA S23. *Suppose that P_n and Q_n are probability measures (each pair (P_n, Q_n) is defined on a common measurable space) with corresponding densities p_n and q_n (with respect to some σ-finite measure ν_n). Let $l_n = \log q_n/p_n$ be the log-likelihood ratio.*[S16] *If*

$$l_n = o_{P_n}(1),$$

then $d_{\mathrm{TV}}(P_n, Q_n) \to 0$.

PROOF. By the continuous mapping theorem,

$$\frac{q_n}{p_n} = \exp(l_n) \xrightarrow{P_n} 1.$$

Le Cam's first lemma (e.g., van der Vaart (1998, Lemma 6.4)) then implies that $Q_n \triangleleft P_n$. Let ϕ_n be arbitrary measurable functions valued in $[0,1]$. Since the ϕ_n are uniformly tight, Prohorov's theorem ensures that for any arbitrary subsequence $(n_j)_{j\in\mathbb{N}}$ there exists a further subsequence $(n_m)_{m\in\mathbb{N}}$ such that $\phi_{n_m} \rightsquigarrow \phi \in [0,1]$ under P_{n_m}. Therefore,

$$\big(\phi_{n_m}, \exp(l_{n_m})\big) \rightsquigarrow (\phi, 1) \quad \text{under } P_{n_m}.$$

By Le Cam's third lemma (e.g., van der Vaart (1998, Theorem 6.6)), under Q_{m_n} the law of ϕ_{n_m} converges weakly to the law of ϕ. Since each $\phi_n \in [0,1]$,

$$\lim_{m\to\infty} [Q_{n_m}\phi_{n_m} - P_{n_m}\phi_{n_m}] = 0.$$

As $(n_j)_{j\in\mathbb{N}}$ was arbitrary, the preceding display holds also along the original sequence. □

PROPOSITION S1 (Cf. Proposition 2.29 in van der Vaart, 1998). *Suppose that on a measureable space $(S,\mathcal{S})$, $(\mu_n)_{n\in\mathbb{N}}$ is a sequence of measures and μ a measure such that $\mu(A) \leq \liminf_{n\to\infty} \mu_n(A)$ for each $A \in \mathcal{S}$. If $(f_n)_{n\in\mathbb{N}}$ and f are (real-valued) measurable functions such that $f_n \to f$ in μ-measure and $\limsup_{n\to\infty} \int |f_n|^p \mathrm{d}\mu_n \leq \int |f|^p \mathrm{d}\mu < \infty$ for some $p \geq 1$, then $\int |f_n - f|^p \mathrm{d}\mu_n \to 0$.*

[S16] l_n may be defined arbitrarily when $p_n = 0$.

PROOF. $(a+b)^p \leq 2^p(a^p+b^p)$ for any $a, b \geq 0$, and hence, under our hypotheses,

$$0 \leq 2^p|f_n|^p + 2^p|f|^p - |f_n - f|^p \to 2^{p+1}|f|^p \quad \text{in } \mu\text{-measure.}$$

By Lemma 2.2 of Serfozo (1982) and $\limsup_{n\to\infty} \int |f_n|^p \, d\mu_n \leq \int |f|^p \, d\mu < \infty$,

$$\begin{aligned}\int 2^{p+1}|f|^p \, d\mu &\leq \liminf_{n\to\infty} \int 2^p|f_n|^p + 2^p|f|^p - |f_n - f|^p \, d\mu_n \\ &\leq 2^{p+1} \int |f|^p \, d\mu - \limsup_{n\to\infty} \int |f_n - f|^p \, d\mu_n.\end{aligned}$$

□

REMARK 1. The condition that $\mu(A) \leq \liminf_{n\to\infty} \mu_n(A)$ for each $A \in \mathcal{S}$ in Propositions S1 is clearly satisfied if $\mu_n \to \mu$ setwise or in total variation.

REFERENCES

Baumeister, Christiane and James D. Hamilton (2015), "Sign restrictions, structural vector autoregressions, and useful prior information." *Econometrica*, 83 (5), 1963–1999. [1, 2]

Billingsley, Patrick (1995), *Probability and Measure*. Wiley. [5]

Braun, Robin (2023), "The importance of supply and demand for oil prices: Evidence from non-Gaussianity." *Quantitative Economics*, 14 (4), 1163–1198. [2]

Brockwell, Peter J. and Richard A. Davis (1991), *Time Series: Theory and Methods*, second edition. Springer Series in Statistics. Springer. [3, 4]

Chow, Yuan Shih and Henry Teicher (1997), *Probability Theory*, third edition. Springer Texts in Statstics. Springer. [9]

Comon, Pierre (1994), "Independent component analysis, a new concept?" Signal Processing, 36. [1]

Davidson, James (1994), *Stochastic Limit Theory*. Oxford University Press. [9, 18, 22]

Doukhan, Paul (1994), *Mixing: Properties and Examples*. Springer. [22]

Durrett, Rick (2019), *Probability Theory and Examples*, fith edition. Cambridge University Press, Cambridge, UK. [26]

Feinberg, Eugene A., Pavlo O. Kasyanov, and Michael Z. Zgurovsky (2016), "Uniform Fatou's lemma." *Journal of Mathematical Analysis and Applications*, 444 (1), 550–567. [7, 9, 11, 23]

Gut, Allan (2005), *Probability: A Graduate Course*. Springer Texts in Statistics. Springer. [15]

Horn, Roger A. and Charles R. Johnson (2013), *Matrix Analysis*, second edition. Cambridge University Press. [33]

Johnson, Norman L., Samuel Kotz, and Narayanaswamy Balakrishnan (1995), *Continuous Univariate Distributions*. Wiley Series in Probability and Statistics, Vol. 2. Wiley. [34]

Kallenberg, Olav (2021), *Foundations of Modern Probability. Probability Theory and Stochastic Modelling*. Springer. [4, 26]

Kanaya, Shin (2017), "Convergence rates of sums of α-mixing triangular arrays: With an application to nonparametric drift function estimation of continuous-time processes." *Econometric Theory*, 33 (5), 1121–1153. [18, 22, 26]

Magnus, Jan R., Henk G. J. Pijls, and Enrique Sentana (2021), "The Jacobian of the exponential function." *Journal of Economic Dynamics and Control*, 127, 104122. [2]

Rudin, Walter (1991), *Functional Analysis*, second edition. McGraw Hill. [22]

Saikkonen, Pentti (2007), "Stability of mixtures of vector autoregressions with autoregressive conditional heteroskedasticity." *Statistica Sinica*, 17 (1), 221–239. [4, 9, 18, 22, 26]

Serfozo, Richard (1982), "Convergence of Lebesgue integrals with varying measures." *Sankhyā: The Indian Journal of Statistics, Series A (1961–2002)*, 44 (3), 380–402. [18, 36, 37]

Swensen, Anders Rygh (1985), "The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend." *Journal of Multivariate Analysis*, 16 (1), 54–70. [10, 11]

van der Vaart, Aad W. (1998), *Asymptotic Statistics*, first edition. Cambridge University Press, New York, NY, USA. [6, 7, 10, 11, 36]

Co-editor James Hamilton handled this manuscript.

Manuscript received 24 October, 2022; final version accepted 24 January, 2024; available online 1 February, 2024.