Supplement to

" Robust Inference for Non-Gaussian Linear Simultaneous Equations Models"*

Adam Lee[a] *and* *Geert Mesters*[a,b]

[a] Universitat Pompeu Fabra and Barcelona GSE

[b] Vrije Universiteit Amsterdam

July 6, 2022

Abstract

In this supplementary material we provide the following additional results.

S1: Proof of Lemmas 1-3 from the main text.

S2: Auxiliary results

S3: Additional simulation results.

*Email: adam.lee@upf.edu, geert.mesters@upf.edu. Address: Jaume 1, Ramon Trias Fargas 25-27, 08005, Barcelona, Spain.

We thank numerous seminar participants for helpful comments. Mesters acknowledge support from the Spanish Ministry of Economy and Competitiveness through the Ramon y Cajal fellowship (RYC2019-028287-I), the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S), and the Netherlands Organization for Scientific Research (NWO) through the VENI research grant (016.Veni.195.036).

Throughout this document, references to lemmas, equations etc. which start with a "S" are references to this document. Those which consist of just a number refer to the main text.

S1 Proofs for Lemmas 1-3

In this section we provide the proofs for lemmas 1-3. The proofs of these lemmas depend on a number of supporting results which can be found in section S1.1. Many of these results are standard but are nevertheless included for convenience.

Proof of Lemma 1. The log density for the semiparametric LSEM is given by

$$\ell_\theta(y,\tilde{x}) \coloneqq \log p_\theta(y,\tilde{x}) = \log|A| + \sum_{k=1}^{K} \log \eta_k(A_{k\bullet}(y - Bx)) + \log \eta_0(\tilde{x}) \ .$$

For convenience let $v = v_\theta \coloneqq y - Bx$ with $x = (1, \tilde{x})$. We define $\dot{\ell}_\theta(y,\tilde{x}) \coloneqq \nabla_\gamma \ell_\theta(y,\tilde{x})$, where we recall that γ partitions as $\gamma = (\alpha, \beta)$, with $\beta = (\sigma, b)$, and some derivations show that the components of $\dot{\ell}_\theta(y,\tilde{x})$ can be written as

$$\begin{aligned}
\dot{\ell}_{\theta,\alpha_l}(y,\tilde{x}) &= \operatorname{tr}(A^{-1}D_{\alpha,l}(\alpha,\sigma)) + \sum_{k=1}^{K} \phi_k(A_{k\bullet}v) \times [D_{\alpha,l}(\alpha,\sigma)]_{k\bullet} v \\
&= \operatorname{tr}(D_{\alpha,l}(\alpha,\sigma)A^{-1}) + \sum_{k=1}^{K}\sum_{j=1}^{K} \phi_k(A_{k\bullet}v) \times \left([D_{\alpha,l}(\alpha,\sigma)]_{k\bullet}A^{-1}_{\bullet j}\right) A_{j\bullet} v \\
&= \sum_{k=1}^{K}\sum_{j=1, j\neq k}^{K} \zeta^{\alpha}_{l,k,j}\phi_k(A_{k\bullet}v)A_{j\bullet}v + \sum_{k=1}^{K} \zeta^{\alpha}_{l,k,k}\left(\phi_k(A_{k\bullet}v)A_{k\bullet}v + 1\right), \qquad \text{(S1)} \\
\dot{\ell}_{\theta,\sigma_l}(y,\tilde{x}) &= \sum_{k=1}^{K}\sum_{j=1, j\neq k}^{K} \zeta^{\sigma}_{l,k,j}\phi_k(A_{k\bullet}v)A_{j\bullet}v + \sum_{k=1}^{K} \zeta^{\sigma}_{l,k,k}\left(\phi_k(A_{k\bullet}v)A_{k\bullet}v + 1\right), \\
\dot{\ell}_{\theta,b_l}(y,\tilde{x}) &= \sum_{k=1}^{K} \phi_k(A_{k\bullet}v) \times [-A_{k\bullet}D_{b,l}x] \ ,
\end{aligned}$$

where $\zeta^{\alpha}_{l,k,j} \coloneqq [D_{\alpha,l}]_{k\bullet}A^{-1}_{\bullet j}$, $\zeta^{\sigma}_{l,k,j} \coloneqq [D_{\sigma,l}]_{k\bullet}A^{-1}_{\bullet j}$, $D_{\alpha,l} = \partial A(\alpha,\sigma)/\partial\alpha_l$, $D_{\sigma,l} = \partial A(\alpha,\sigma)/\partial\sigma_l$ and $D_{b,l} = \partial B/\partial b_l$. Paths of the form $t \to P_{\gamma+tg,\eta}$ have an associated tangent space given by

$$\mathcal{T}^{\gamma|\eta}_{P_\theta,\mathbb{R}^L} = \{g'\dot{\ell}_\theta(y,\tilde{x}) : g \in \mathbb{R}^L\} \ . \qquad \text{(S2)}$$

To constructing the tangent space of the non-parametric part we consider submodels of the following form. Let

$$\eta_{k,t}^{h_k}(\cdot) = \eta_k(\cdot)(1 + th_k(\cdot)) \qquad k = 0, \ldots, K \ ,$$

which for $t = 0$ recover η_k. For $k = 1, \ldots, K$, h_k is some function such that $h_k \in H_k$ with

$$H_k \coloneqq \left\{ h_k \in \mathcal{C}_b^1(\lambda) : \mathbb{E}h_k(\epsilon_k) = 0, \mathbb{E}\epsilon_k h_k(\epsilon_k) = 0, \mathbb{E}\kappa(\epsilon_k)h_k(\epsilon_k) = 0 \right\} \ , \tag{S3}$$

where $\mathcal{C}_b^1(\lambda)$ denotes the space of functions from $\mathbb{R} \to \mathbb{R}$ which are bounded and continuously differentiable with bounded derivatives λ-a.e.. Letting G_k be the law on $\mathbb{R}$ corresponding to η_k for $k = 1, \ldots, K$, it is clear that H_k is a linear subspace of $L_2(G_k)$. The additional restrictions on h_k ensure that for t small enough $\eta_{k,t} \in \mathscr{H}$. For $k = 0$, define

$$H_0 \coloneqq \left\{ h_0 \in \mathcal{C}_b(\lambda, \mathbb{R}^{d-1}) : \mathbb{E}h_0(\tilde{X}) = 0 \right\} \ , \tag{S4}$$

where $\mathcal{C}_b(\lambda, \mathbb{R}^{d-1})$ denotes the space of bounded λ-a.e. continuous functions from $\mathbb{R}^{d-1} \to \mathbb{R}$.[S1] Letting G_0 be the law on $\mathbb{R}^{d-1}$ corresponding to η_0, it is clear that H_0 is a linear subspace of $L_2(G_0)$. The additional restrictions on h_0 ensure that for t small enough $\eta_{0,t} \in \mathscr{Z}$. Now let $H \coloneqq \prod_{k=0}^K H_k$. For any $h = (h_0, h_1, \ldots, h_K) \in H$ and any $\theta \in \Theta$ we can define a path $\eta_t(\theta, h) \coloneqq (\eta_{0,t}^{h_0}, \eta_{1,t}^{h_1}, \ldots, \eta_{K,t}^{h_K})$. Given the preceding discussion, for each $h \in H$ there is a $\delta > 0$ small enough such that $\eta_{0,t}^{h_0} \in \mathscr{Z}$ and $\eta_{k,t}^{h_k} \in \mathscr{H}$ for each $k = 1, \ldots, K$ when $t \in (-\delta, \delta)$. Now, we use this to define a path $\theta_t(\theta, h) \coloneqq (\gamma, \eta_t(\theta, h))$. Then, $p_{\theta_t(\gamma,h)}$ defines a path towards p_θ according to:

$$p_{\theta_t(\theta,h)}(y, \tilde{x}) = |\det A| \times \prod_{k=1}^K \eta_{k,t}^{h_k}(A_{k\bullet}v) \times \eta_{0,t}^{h_0}(\tilde{x}) \ . \tag{S5}$$

Given the discussion above, for $t \in (-\delta, \delta)$, the submodel $\{P_{\theta_t(\theta,h)} : t \in (-\delta, \delta)\} \subset \mathcal{P}_\Theta$. Let $s : \mathbb{R}^K \to \mathbb{R}$ be given by

$$\begin{aligned} s(y, \tilde{x}) \coloneqq \left. \frac{\partial \log p_{\theta_t(\theta,h)}(y, \tilde{x})}{\partial t} \right|_{t=0} &= \left. \frac{h_0(\tilde{x})}{1 + th_0(\tilde{x})} \right|_{t=0} + \sum_{k=1}^K \left. \frac{h_k(A_{k\bullet}v)}{1 + th_k(A_{k\bullet}v)} \right|_{t=0} \\ &= h_0(\tilde{x}) + \sum_{k=1}^K h_k(A_{k\bullet}v) \ . \end{aligned} \tag{S6}$$

s is a score function associated to the differentiable path $t \mapsto P_{\theta_t(\theta,h)}$ from $[0, \delta) \to \mathcal{P}_\Theta$ and

[S1] We make no notational distinction between the Lebesgue measure on $\mathbb{R}$ and that on $\mathbb{R}^{d-1}$; which is meant can be inferred from context.

the associated tangent space for η is given by

$$\mathcal{T}^{\eta|\gamma}_{P_\theta,H} \coloneqq \left\{ y \mapsto h_0(\tilde{x}) + \sum_{k=1}^{K} h_k(A_{k\bullet}v) : h = (h_0, h_1, \ldots, h_K) \in H \right\}. \tag{S7}$$

These calculations establish the form of the score functions for the parameteric part and non-parametric part of the model separately. To verify assumption 3 we rather need to consider the (joint) paths given by $\theta_t(\theta, g, h) = (\gamma + tg, \eta_t(\theta, h))$.

By the definitions of $\mathcal{T}^{\gamma|\eta}_{P_\theta,\mathbb{R}^L}$ and $\mathcal{T}^{\eta|\gamma}_{P_\theta,H}$ given in (S2) and (S7) respectively and the fact that both $\mathbb{R}^L$ and H are linear spaces, it follows that $\mathcal{T}^{\gamma|\eta}_{P_\theta,\mathbb{R}^L}$ and $\mathcal{T}^{\eta|\gamma}_{P_\theta,H}$ are linear spaces, implying that the same is true of their sum. Therefore, provided we show that $\mathcal{T}_{P_\theta,\mathcal{J}}$ is a tangent *set* to the model at P_θ and that it is the sum of $\mathcal{T}^{\gamma|\eta}_{P_\theta,\mathbb{R}^L}$ and $\mathcal{T}^{\eta|\gamma}_{P_\theta,H}$, we immediately obtain that it is a tangent *space*. That the second equality in the display in the statement of the lemma holds is clear by the definition of a sum of linear subspaces and the form of the elements on the right hand side given in equations (S2) and (S7). So it remains to prove the first equality. That is, for any $g \in \mathbb{R}^L$ and $h \in H$ there is a small enough $\delta > 0$ such that the path $t \mapsto P_{\theta_t(\theta,g,h)}$ from $[0, \delta)$ to (a subset of) $\mathcal{P}_\Theta$ is a differentiable path with score function $y \mapsto g'\dot{\ell}_\theta(y) + h_0(\tilde{x}) + \sum_{k=1}^K h_k(A_{k\bullet}v)$. Fix $g \in \mathbb{R}^L, h \in H$ and $\theta \in \Theta$ and let θ_t abbreviate $\theta_t(\theta, g, h)$. Recall that γ partitions as $\gamma = ((\alpha, \sigma), b)$ and let $g = (g_1, g_2)$ be the conforming partition for any $g \in \mathbb{R}^L$. Further, let G_2 be such that $g_2 = \operatorname{vec}(G_2)$. Additionally throughout the proof we will let $M_k = M_{k\bullet}$ for any matrix M and to save on notation, we define $\tilde{A}(t) \coloneqq A((\alpha', \sigma')' + tg_1)$, $\tilde{B}(t) = B + G_2 t$, $\tilde{v}(t) \coloneqq y - \tilde{B}(t)x$ and $\tilde{D}_k(t) \coloneqq \frac{\mathrm{d}[\tilde{A}(a)]_k\tilde{v}(a)}{\mathrm{d}a}(t)$.

We will now compute the (pointwise) derivative of $t \mapsto \ell_{\theta_t}(y, \tilde{x}) \coloneqq \log p_{\theta_t}(y, \tilde{x})$ on $(-\delta, \delta)$. We have that

$$\begin{aligned}
\ell_{\theta_t}(y, \tilde{x}) = &\log|\det \tilde{A}(t)| + \log \eta_0(\tilde{x}) + \sum_{k=1}^{K} \log \eta_k\left([\tilde{A}(t)]_k \tilde{v}(t)\right) \\
&+ \log\left(1 + th_0(\tilde{x})\right) + \sum_{k=1}^{K} \log\left(1 + th_k\left([\tilde{A}(t)]_k\tilde{v}(t)\right)\right).
\end{aligned}$$

For sufficiently small t (i.e. there is some neighbourhood $(-\delta, \delta)$ on which) the arguments of the logarithms on the second line are positive. We proceed by repeatedly applying the chain rule to conclude that

$$\acute{\ell}_{\theta_t}(y,\tilde{x}) \coloneqq \frac{\partial \ell_{\theta_t}(y,\tilde{x})}{\partial t} = \operatorname{tr}\left(\left[\tilde{A}(t)\right]^{-1}\frac{\mathrm{d}\tilde{A}(t)}{\mathrm{d}t}\right) + \frac{h_0(\tilde{x})}{1+th_0(\tilde{x})} + \sum_{k=1}^{K}\left[\phi_k\left([\tilde{A}(t)]_k\tilde{v}(t)\right)\times\tilde{D}_k(t)\right] + \sum_{k=1}^{K}\frac{h_k([\tilde{A}(t)]_k\tilde{v}(t)) + th_k'([\tilde{A}(t)]_k\tilde{v}(t))\times\tilde{D}_k(t)}{1+th_k([\tilde{A}(t)]_k\tilde{v}(t))},$$

for all $y,\tilde{x}$ such that $p_{\theta_t}(y,\tilde{x})>0$ and define it as 0 elsewhere. Use (S1) to evaluate the preceding display at $t=0$ and obtain (for y such that $p_{\theta_t}(y,\tilde{x})>0$ and set it to 0 otherwise):

$$s(y,\tilde{x}) \coloneqq \frac{\partial \ell_{\theta_t}(y,\tilde{x})}{\partial t}_{|t=0} = \operatorname{tr}\left(\left[\tilde{A}(t)\right]^{-1}\frac{\mathrm{d}\tilde{A}(t)}{\mathrm{d}t}_{|t=0}\right) + h_0(\tilde{x}) + \sum_{k=1}^{K}\left[\phi_k\left(A_k v\right)\times\tilde{D}_k(0)v\right] + h_k\left(A_k v\right)$$
$$= g'\dot{\ell}_\theta + h_0(\tilde{x}) + \sum_{k=1}^{K} h_k\left(A_k v\right).$$

We will demonstrate that the conditions in Lemma 7.6 of van der Vaart (1998) (alternatively Lemma 1.8 of van der Vaart (2002)) are satisfied for the map $t\mapsto P_{\theta_t}$ from $(-\delta,\delta)$ to $\mathcal{P}_\Theta$, from which we will be able to conclude that this is a differentiable path with score function as in the preceding display.[S2]

Firstly, by the imposed continuous differentiability conditions we have that $t\mapsto\sqrt{p_{\theta_t}}$ is continuously differentiable λ-a.e..

It remains to show that $\int\left(\frac{\dot{p}_{\theta_t}}{p_{\theta_t}}\right)^2 \mathrm{d}P_{\theta_t}$ is finite and continuous in t. For this, note that when it exists we have $\acute{\ell}_{\theta_t} = \frac{\dot{p}_{\theta_t}}{p_{\theta_t}}$. Therefore, we can bound our integral by

$$\int\left(\acute{\ell}_{\theta_t}(y,\tilde{x})\right)^2\mathrm{d}P_{\theta_t} \lesssim \operatorname{tr}\left(\left[\tilde{A}(t)\right]^{-1}\frac{\mathrm{d}\tilde{A}(t)}{\mathrm{d}t}\right)^2 + \int\left(\frac{h_0(\tilde{x})}{1+th_0(\tilde{x})}\right)^2\mathrm{d}P_{\theta_t}$$
$$+\sum_{k=1}^{K}\int\left[\phi_k\left([\tilde{A}(t)]_k\tilde{v}(t)\right)\times\tilde{D}_k(t)\right]^2\mathrm{d}P_{\theta_t}$$
$$+\sum_{k=1}^{K}\int\left(\frac{h_k([\tilde{A}(t)]_k\tilde{v}(t)) + th_k'([\tilde{A}(t)]_k\tilde{v}(t))\times\tilde{D}_k(t)}{1+th_k([\tilde{A}(t)]_k\tilde{v}(t))}\right)^2\mathrm{d}P_{\theta_t}.$$

The first rhs term can be ensured finite by choosing δ small enough since $[\tilde{A}(t)]^{-1}\frac{\mathrm{d}\tilde{A}(t)}{\mathrm{d}t}$ is

[S2] Strictly speaking, applying lemma 7.6 as stated in van der Vaart (1998) would require continuous differentiability for every y. Nevertheless, with appropriate modifications, the same proof demonstrates the claim remains valid with continuous differentiability holding "only" λ-a.e.. See also proposition 2.1.1 of Bickel et al. (1998).

continuous in t.[S3] The same is true of the second term, since h_0 is bounded λ-a.s., hence G_0-a.s., and

$$\int \left(\frac{h_0(\tilde{x})}{1+th_0(\tilde{x})}\right)^2 \mathrm{d}P_{\theta_t} = \int \left(\frac{h_0(\tilde{x})}{1+th_0(\tilde{x})}\right)^2 \eta_0(\tilde{x})(1+th_0(\tilde{x}))\,\mathrm{d}\lambda = \int \frac{h_0(\tilde{x})^2}{1+th_0(\tilde{x})}\,\mathrm{d}G_0(\tilde{x}).$$

For the third term it suffices to consider the integral for an arbitrary $k \in [K]$, which by Cauchy-Schwarz is bounded by

$$\int \left[\phi_k\left([\tilde{A}(t)]_k\tilde{v}(t)\right) \times \tilde{D}_k(t)\right]^2 \mathrm{d}P_{\theta_t} \leq \left\|\phi_k\left([\tilde{A}(t)]_k\tilde{v}(t)\right)^2\right\|_{P_{\theta_t},2} \left\|\left[\tilde{D}_k(t)\right]^2\right\|_{P_{\theta_t},2} < \infty$$

For the first term observe that if $Y, \tilde{X}$ has law P_{θ_t}, then $[\tilde{A}(t)]_k\tilde{v}(t)$ is distributed according to the density $\eta_k(1+th_k) \in \mathscr{H}$ (for small enough δ), and thus the integral is finite by the definition of $\mathscr{H}$, i.e. assumption 1-part 1. For the second term write

$$\tilde{D}_k(t) = \frac{\mathrm{d}[\tilde{A}(a)]_k}{\mathrm{d}a}(t)\left(z - \tilde{B}(t)x\right) - [\tilde{A}(t)]_k\left(\frac{\mathrm{d}\tilde{B}(a)}{\mathrm{d}a}(t)x\right),$$

and note that for small enough δ, $P_{\theta_t} \in \mathcal{P}_\Theta$ and so for some small enough $\nu > 0$, each $P_{\theta_t}|Y_k|^{4+\nu} < \infty$ and $P_{\theta_t}|X_l|^{4+\nu} < \infty$ (by assumption 1), hence $\left\|\left[\tilde{D}_k(t)\right]^2\right\|_{P_{\theta_t},2} = \sqrt{\int[\tilde{D}_k(t)]^4\,\mathrm{d}P_{\theta_t}} < \infty$ since $\int\|\tilde{D}_k(t)\|_2^{4+\nu}\,\mathrm{d}P_{\theta_t} < \infty$.

For the final term on the rhs, it is again sufficient to consider the integral for any arbitrary $k \in [K]$. Here, let $c > 0$ be a bound away from zero for $1 + th_k$ on $(-\delta, \delta)$ and let $M > 0$ bound both h_k and h_k' on the same interval, which we know to be possible by their definition. Then this integral can be bounded by

$$\int \left(\frac{h_k([\tilde{A}(t)]_k\tilde{v}(t)) + th_k'([\tilde{A}(t)]_k\tilde{v}(t)) \times \tilde{D}_k(t)}{1+th_k([\tilde{A}(t)]_k\tilde{v}(t))}\right)^2 \mathrm{d}P_{\theta_t} \leq \int \left(\frac{M + tM\tilde{D}_k(t)}{c}\right)^2 \mathrm{d}P_{\theta_t},$$

where the right hand side can be seen to be finite by the fact that $\int[\tilde{D}_k(t)]^2\,\mathrm{d}P_{\theta_t} < \infty$ as implied by the corresponding finite 4th moment obtained above.

To show continuity, let $t_n \to t$ be an arbitrary convergent sequence in $[0, \delta)$ with δ chosen such that if $0 \leq t \leq \delta$ then each $h_k, h_k', h_0 \leq M$ and $1 + th_k, 1 + th_0 \geq c > 0$. Suppose that $Z_n = (Y_n, \tilde{X}_n)$ and $Z = (Y, \tilde{X})$ have laws $P_{\theta_{t_n}}$ and P_{θ_t} respectively and let

[S3] By our assumptions that $(\alpha, \beta_1) \mapsto A(\alpha, \beta_1)$ is continuously differentiable and $A(\alpha, \beta_1)$ is invertible.

$\tilde{v}(t,Z) \coloneqq Y - \tilde{B}(t)X$. We have

$$b_n \coloneqq \operatorname{tr}\left(\left[\tilde{A}(t_n)\right]^{-1} \frac{\mathrm{d}\tilde{A}(t)}{\mathrm{d}t}(t_n)\right) \to b \coloneqq \operatorname{tr}\left(\left[\tilde{A}(t)\right]^{-1} \frac{\mathrm{d}\tilde{A}(t)}{\mathrm{d}t}(t)\right),$$

which converges by the continuity of all its constituent functions. Define for $k = 1, \dots, K$

$$\begin{aligned}
U_{k,n} &\coloneqq \phi_k\left([\tilde{A}(t_n)]_k \tilde{v}(t_n, Z_n)\right)\\
W_{k,n} &\coloneqq \tilde{D}_k(t_n)\\
V_{k,n} &\coloneqq \frac{h_k([\tilde{A}(t_n)]_k \tilde{v}(t_n, Z_n))}{1 + t_n h_k([\tilde{A}(t_n)]_k \tilde{v}(t_n, Z_n))}\\
Q_{k,n} &\coloneqq \frac{t_n h_k'([\tilde{A}(t_n)]_k \tilde{v}(t_n, Z_n))}{1 + t_n h_k([\tilde{A}(t_n)]_k \tilde{v}(t_n, Z_n))}\\
E_n &\coloneqq \frac{h_0(\tilde{X}_n)}{1 + t_n h_0(\tilde{X}_n)},
\end{aligned}$$

and analogously U_k, V_k, W_k, Q_k, E where the t_n are replaced by t and the Z_n by Z respectively. Since $p_{\theta_{t_n}} \to p_{\theta_t}$ we have that $\tilde{Z}_n \rightsquigarrow \tilde{Z}$ by Scheffé's theorem. Hence, by the continuous mapping theorem

$$\begin{aligned}
&(U_{1,n}, V_{1,n}, W_{1,n}, Q_{1,n}, \dots, U_{K,n}, V_{K,n}, W_{K,n}, Q_{K,n}, E_n)\\
&\qquad \rightsquigarrow (U_1, V_1, W_1, Q_1, \dots, U_K, V_K, W_K, Q_K, E).
\end{aligned}$$

Moreover, $V_{k,n}$, $Q_{k,n}$ and E_n are bounded above. We have that $(U_{k,n}^4)_{n\geq 1}$ and $(W_{k,n}^4)_{n\geq 1}$ are uniformly integrable for each $k \in [K]$. For the former, note that each $[\tilde{A}(t_n)]_k \tilde{v}(t_n, Z_n)$ is distributed according to the density $\eta_k(1 + t_n h_k)$. Hence we have for small enough but positive ν

$$\sup_{n\in\mathbb{N}} \mathbb{E}|U_{k,n}|^{4+\nu} = \sup_{n\in\mathbb{N}} \int |\phi_k(z)|^{4+\nu} \eta_k(z)(1 + t_n h_k(z))\,\mathrm{d}z \lesssim \int |\phi_k(z)|^{4+\nu} \eta_k(z)\,\mathrm{d}z < \infty.$$

Similarly, using Cauchy-Schwarz, for small enough but positive ν

$$\begin{aligned}\sup_{n\in\mathbb{N}}\mathbb{E}|W_{k,n}|^{4+\nu} &= \sup_{n\in\mathbb{N}}\int |\tilde{D}_k(t_n)|^{4+\nu}\,\mathrm{d}P_{\theta_{t_n}} \\ &\lesssim \sup_{n\in\mathbb{N}}\int \|\epsilon_n\|_2^{4+\nu}\,\mathrm{d}P_{\theta_{t_n}} + \sup_{n\in\mathbb{N}}\int \|X_n\|_2^{4+\nu}\,\mathrm{d}P_{\theta_{t_n}} \\ &\lesssim \sup_{n\in\mathbb{N}}\sum_{k=1}^{K}\int |e_k|^{4+\nu}\eta_k(e_k)(1+t_nh_k(e_k))\,\mathrm{d}e_k \\ &\quad + \sup_{n\in\mathbb{N}}\int \|(1,\tilde{x}')'\|_2^{4+v}\eta_0(\tilde{x})(1+t_n(h_0(\tilde{x})))\,\mathrm{d}\tilde{x} \\ &\lesssim \sum_{k=1}^{K}\int |\epsilon_k|^{4+\nu}\,\mathrm{d}G_k + \int \|(1,\tilde{X}')'\|_2^{4+v}\,\mathrm{d}G_0 \\ &< \infty.\end{aligned}$$

With this in hand, using continuous mapping theorem and noting that each of the relevant sequences is $P_{\theta_{t_n}}$-UI given the preceding discussion we have, as $n\to\infty$

$$P_{\theta_{t_n}}\left[b_n + E_n + \sum_{k=1}^{K}U_{k,n}W_{k,n} + \sum_{k=1}^{K}V_{k,n} + Q_{k,n}\right]^2 \to P_{\theta_t}\left[b + E + \sum_{k=1}^{K}U_kW_k + \sum_{k=1}^{K}V_k + Q_k\right]^2,$$

yielding the required continuity.

□

Proof of Lemma 2. To construct the efficient score function for γ, we need to project the elements of $\dot{\ell}_\theta(y)$, as given in (S1), onto the orthogonal complement of $\mathcal{T}^{\eta|\gamma}_{P_\theta,H}$ (equation (S7)), that is: $\breve{\ell}_{\theta,l} = \Pi\left(\dot{\ell}_{\theta,l}\middle|\left[\mathcal{T}^{\eta|\gamma}_{P_\theta,H}\right]^\perp\right)$.[S4] The efficient score then follows as $\tilde{\ell}_{\theta,l} := \dot{\ell}_{\theta,l} - \Pi_\theta\dot{\ell}_{\theta,l} = \dot{\ell}_{\theta,l} - \Pi\left(\dot{\ell}_{\theta,l}\middle|\left[\mathcal{T}^{\eta|\gamma}_{P_\theta,H}\right]^\perp\right)$.

We first provide some results that simplify the exposition. Lemma S2 proves that the closure of H_k is given by

$$\operatorname{cl} H_k = \{h_k \in L_2(G_k) : \mathbb{E}h_k(\epsilon_k) = 0, \mathbb{E}\epsilon_k h_k = 0, \mathbb{E}\kappa(\epsilon_k)h_k(\epsilon_k) = 0\},$$

and similarly

$$\operatorname{cl} H_0 = \{h_0 \in L_2(G_0) : \mathbb{E}h_0(\tilde{X}) = 0\}.$$

Now, let $\tilde{H}^\gamma_k := \{y \mapsto h_k(A_{k\bullet}v) : h_k \in H_k\}$ for $k = 1,\ldots,K$, $\tilde{H}^\gamma_0 := \{y \mapsto h_0(\tilde{x}) : h_0 \in$

[S4]See e.g. Section 2.2 of van der Vaart (2002).

$H_0\}$ and note that $\mathcal{T}^{\eta|\gamma}_{P_\theta,H}$ can be written as

$$\mathcal{T}^{\eta|\gamma}_{P_\theta,H} = \tilde{H}^\gamma_0 + \tilde{H}^\gamma_1 + \cdots + \tilde{H}^\gamma_K. \tag{S8}$$

It follows that for $k = 1, \ldots, K$

$$\operatorname{cl} \tilde{H}^\gamma_0 = \{y \mapsto h_0(\tilde{x}) : h_0 \in \operatorname{cl} H_0\}, \quad \operatorname{cl} \tilde{H}^\gamma_k = \{y \mapsto h_k(A_{k\bullet}v) : h_k \in \operatorname{cl} H_k\}, \tag{S9}$$

which are (closed) subspaces of $L_2(P_\theta)$.[S5]

Define $\mathscr{T} \coloneqq \operatorname{cl} \tilde{H}^\gamma_1 + \cdots + \operatorname{cl} \tilde{H}^\gamma_K$ and the following finite dimensional subset of $L_2(P_\theta)$

$$\mathscr{L}_0 \coloneqq \mathscr{L}_1 \cup \mathscr{L}_2 \coloneqq \{y \mapsto A_{k\bullet}v, y \mapsto \kappa(A_{k\bullet}v) : k \in [K]\} \cup \{y \mapsto \phi_k(A_{k\bullet}v)A_{j\bullet}v : j, k \in [K],\ j \neq k\}, \tag{S10}$$

where $\kappa(w) \coloneqq w^2 - 1$ and $\mathscr{L} \coloneqq \operatorname{lin} \mathscr{L}_0$. Lemma S4 proves that $\mathscr{L} \subset \mathscr{T}^\perp$.

Since orthogonal projections are linear we have that for $\kappa = \alpha, \sigma$

$$\begin{aligned}
\Pi\left(\dot{\ell}_{\theta,\kappa_l}|\mathscr{T}^\perp\right) &= \sum_{k=1}^K \sum_{j=1, j\neq k}^K \zeta^\kappa_{l,k,j} \Pi\left(\phi_k(A_{k\bullet}v)A_{j\bullet}v|\mathscr{T}^\perp\right) \\
&\quad + \sum_{k=1}^K \zeta^\kappa_{l,k,k} \Pi\left(\phi_k(A_{k\bullet}v)A_{k\bullet}v + 1|\mathscr{T}^\perp\right) \\
&= \sum_{k=1}^K \sum_{j=1, j\neq k}^K \zeta^\kappa_{l,k,j} \phi_k(A_{k\bullet}v)A_{j\bullet}v + \sum_{k=1}^K \zeta^\kappa_{l,k,k} \Pi\left(\phi_k(A_{k\bullet}v)A_{k\bullet}v + 1|\mathscr{T}^\perp\right)
\end{aligned}$$

where the second equality follows from $y \mapsto \phi_k(A_{k\bullet}v)A_{j\bullet}v \in \mathscr{L} \subset \mathscr{T}^\perp$, for $j \neq k$.

What remains is $\Pi\left(\phi_k(A_{k\bullet}v)A_{k\bullet}v + 1|\mathscr{T}^\perp\right)$. For this we specialise to the case for $\theta = (\gamma, \eta)$ such that $\eta \in \mathcal{H}_0$, for which we can establish an explicit expression.

In particular, we will show that for each $k \in [K]$, there are τ_i for $i = 1, 2$ such that $y \mapsto w(A_k v) \in \operatorname{cl} \tilde{H}^\gamma_k$ where $w(A_k v) \coloneqq \phi_k(A_k v)A_k v + 1 - r(A_k v)$ and $r(A_k v) \coloneqq \tau_1 A_k v + \tau_2 \kappa(A_k v)$. This would imply that we can write $\phi_k(A_k v)A_k v + 1 = w(A_k v) + r(A_k v)$ where the first summand on the right hand side is in $\mathscr{T}$ and the latter is in $\mathscr{L} \subset \mathscr{T}^\perp$.[S6] Since orthog-

[S5]To see this let $y \mapsto h_k(A_k v) \in \{y \mapsto h_k(A_k v) : h_k \in \operatorname{cl} H_k\}$. There are $h_{n,k} \in H_k$ such that $h_{n,k} \to h_k$ in $L_2(G_k)$. Hence, recalling that $A_k v$ is distributed according to η_k under P_θ, it follows immediately that $\int [h_{n,k}(A_k v) - h_k(A_k v)]^2 \, \mathrm{d}P_\theta \to 0$ as $n \to \infty$. Hence $y \mapsto h_k(A_k v) \in \operatorname{cl} \tilde{H}^\gamma_k$. For the reverse inclusion, let $y \mapsto h_k(A_k v) \in \operatorname{cl} \tilde{H}^\gamma_k$. So there are $y \mapsto h_{n,k}(A_k v)$ in $\tilde{H}^\gamma_k$ such that $\int [h_{n,k}(A_k v) - h_k(A_k v)]^2 \, \mathrm{d}P_\theta \to 0$ as $n \to \infty$. Again noting that $A_k v$ is distributed according to η_k under P_θ, this immediately implies that $h_{n,k} \to h_k$ in $L_2(G_k)$. That $\operatorname{cl} \tilde{H}^\gamma_k$ is a subspace of $L_2(P_\theta)$ follows directly from the fact that $\operatorname{cl} H_k$ is a subspace of $L_2(G_k)$ once more noting $A_k v$ is distributed according to η_k under P_θ. The argument for $\tilde{H}^\gamma_0$ is analogous.

[S6]Take $h_k = w$ and $h_j = 0$ for all $j \neq k$ to see that $y \mapsto w(A_k v) \in \operatorname{cl} \tilde{H}^\gamma_k$ implies $y \mapsto w(A_k v) \in \mathscr{T}$.

onal decompositions are unique this would further imply that $\Pi\left(\phi_k(A_kv)A_kv+1|\mathscr{T}^\perp\right) = \Pi\left(\phi_k(A_kv)A_kv+1|\mathscr{L}\right) = r(A_kv)$.[S7]

To show that $y \mapsto w(A_kv) \in \mathrm{cl}\,\tilde{H}_k^\gamma$ let $h_k(z) := \phi_k(z)z + 1 - \tau_{k,1}z - \tau_{k,2}\kappa(z)$. We first note that $h_k \in L_2(G_k)$, which can be easily seen by the triangle inequality along with the fact that all of ϵ_k, $\kappa(\epsilon_k)$, 1 and $\phi_k(\epsilon_k)\epsilon_k$ are in $L_2(G_k)$. Next, $\int \phi_k(z)z\,\mathrm{d}G_k + 1 - \tau_{k,1}\int z\,\mathrm{d}G_k - \tau_{k,2}\int \kappa(z)\,\mathrm{d}G_k = 1 + \int \phi_k(z)z\,\mathrm{d}G_k$, and so as $\eta \in \mathcal{H}_0$,

$$\int h_k(z)\,\mathrm{d}G_k = 1 - 1 = 0.$$

Next, we will demonstrate that $\tau_{k,1}$ and $\tau_{k,2}$ can be chosen such that $\int h_k(z)z\,\mathrm{d}G_k = \int h_k(z)\kappa(z)\,\mathrm{d}G_k = 0$. As $\eta \in \mathcal{H}_0$ we have that

$$\begin{aligned}
\int h_k(z)z\,\mathrm{d}G_k &= \int \phi_k(z)z^2\,\mathrm{d}G_k + \int z\,\mathrm{d}G_k - \tau_{k,1}\int z^2\,\mathrm{d}G_k - \tau_{k,2}\int \kappa(z)z\,\mathrm{d}G_k \\
&= -\tau_{k,1}\int z^2\,\mathrm{d}G_k - \tau_{k,2}\int z^3\,\mathrm{d}G_k + \tau_{k,2}\int z\,\mathrm{d}G_k \\
&= -\tau_{k,1}\mathbb{E}\epsilon_k^2 - \tau_{k,2}\mathbb{E}\epsilon_k^3 \\
&= -\tau_{k,1}1 - \tau_{k,2}\mathbb{E}\epsilon_k^3,
\end{aligned}$$

where we note that $\mathbb{E}\epsilon_k^2 = 1$. Similarly,

$$\begin{aligned}
\int h_k(z)\kappa(z)\,\mathrm{d}G_k &= \int \phi_k(z)(z^3 - z)\,\mathrm{d}G_k + \int \kappa(z)\,\mathrm{d}G_k - \tau_{k,1}\int z(z^2-1)\,\mathrm{d}G_k - \tau_{k,2}\int (z^2-1)^2\,\mathrm{d}G_k \\
&= -2 - \tau_{k,1}\left[\int z^3\,\mathrm{d}G_k - \int z\,\mathrm{d}G_k\right] - \tau_{k,2}\left[\int z^4\,\mathrm{d}G_k - 2\int z^2\,\mathrm{d}G_k + 1\right] \\
&= -2 - \tau_{k,1}\int z^3\,\mathrm{d}G_k - \tau_{k,2}\left[\int z^4\,\mathrm{d}G_k - 2\int z^2\,\mathrm{d}G_k + 1\right] \\
&= -2 - \tau_{k,1}\mathbb{E}\epsilon_k^3 - \tau_{k,2}[\mathbb{E}\epsilon_k^4 - 1].
\end{aligned}$$

Hence we need to choose $\tau_{k,1}$ and $\tau_{k,2}$ such that:

$$\begin{bmatrix} 1 & \mathbb{E}\epsilon_k^3 \\ \mathbb{E}\epsilon_k^3 & \mathbb{E}\epsilon_k^4 - 1 \end{bmatrix}\begin{bmatrix} \tau_{k,1} \\ \tau_{k,2} \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}.$$

The matrix $M_k := \left[\begin{smallmatrix} 1 & \mathbb{E}\epsilon_k^3 \\ \mathbb{E}\epsilon_k^3 & \mathbb{E}\epsilon_k^4 - 1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} \mathbb{E}\epsilon_k^2 & \mathbb{E}\epsilon_k^3 \\ \mathbb{E}\epsilon_k^3 & \mathbb{E}\epsilon_k^4 - 1 \end{smallmatrix}\right]$ is nonsingular by assumption 1; see footnote 5. Hence we can take $(\tau_{k,1}, \tau_{k,2})' = M_k^{-1}(0, -2)'$, which is non zero by the nonsingularity of

[S7] See e.g. Theorem 4.11 in Rudin (1987).

M_k^{-1}. We conclude that

$$\Pi\left(\dot{\ell}_{\theta,\alpha_l}|\mathscr{T}^{\perp}\right) = \sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K} \zeta^{\alpha}_{l,k,j}\phi_k(A_{k\bullet}v)A_{j\bullet}v + \sum_{k=1}^{K}\zeta^{\alpha}_{l,k,k}\left[\tau_{k,1}A_{k\bullet}v + \tau_{k,2}\kappa(A_{k\bullet}v)\right] \ ,$$
$$\Pi\left(\dot{\ell}_{\theta,\sigma_l}|\mathscr{T}^{\perp}\right) = \sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K} \zeta^{\sigma}_{l,k,j}\phi_k(A_{k\bullet}v)A_{j\bullet}v + \sum_{k=1}^{K}\zeta^{\sigma}_{l,k,k}\left[\tau_{k,1}A_{k\bullet}v + \tau_{k,2}\kappa(A_{k\bullet}v)\right] \ ,$$

Moreover, by independence, for any $h_0 \in \operatorname{cl}\tilde{H}_0^{\gamma}$

$$P_\theta\left[\Pi\left(\dot{\ell}_{\theta,\alpha_l}|\mathscr{T}^{\perp}\right)h_0\right] = P_\theta\left[\Pi\left(\dot{\ell}_{\theta,\alpha_l}|\mathscr{T}^{\perp}\right)\right]P_\theta h_0 = 0,$$
$$P_\theta\left[\Pi\left(\dot{\ell}_{\theta,\sigma_l}|\mathscr{T}^{\perp}\right)h_0\right] = P_\theta\left[\Pi\left(\dot{\ell}_{\theta,\sigma_l}|\mathscr{T}^{\perp}\right)\right]P_\theta h_0 = 0,$$

and so by lemma S3 we can conclude that (see e.g. Bickel et al., 1998, Proposition A.2.3.B)

$$\Pi\left(\dot{\ell}_{\theta,\alpha_l}|\mathscr{T}^{\perp}\right) = \Pi\left(\dot{\ell}_{\theta,\alpha_l}|\left[\mathcal{T}^{\eta|\gamma}_{P_\theta,H}\right]^{\perp}\right) \qquad \text{and} \qquad \Pi\left(\dot{\ell}_{\theta,\sigma_l}|\mathscr{T}^{\perp}\right) = \Pi\left(\dot{\ell}_{\theta,\sigma_l}|\left[\mathcal{T}^{\eta|\gamma}_{P_\theta,H}\right]^{\perp}\right).$$

For the remaining part corresponding to b, let $\varsigma_k := M_k^{-1}(1,0)'$ and define $q(y,\tilde{x}) := \phi_k(A_{k\bullet}v) + \varsigma_{k,1}A_{k\bullet}v + \varsigma_{k,2}\kappa(A_{k\bullet}v)$. Then we have that for any $a \in \mathbb{R}^d$

$$y \mapsto q(y,\tilde{X}) \times a'\mathbb{E}X \in \operatorname{cl}\tilde{H}_k^{\gamma} \subset \operatorname{cl}\mathcal{T}^{\eta|\gamma}_{P_\theta,H},$$

since letting $\tilde{a} := a'\mathbb{E}X$ we have $P_\theta(\tilde{a}q(Y,\tilde{X}))^2 < \infty$ by the triangle inequality & $P_\theta\tilde{a}q(Y,\tilde{X}) = 0$,

$$\begin{aligned}
P_\theta\tilde{a}q(Y,\tilde{X})A_{k\bullet v} &= \tilde{a}\left[\int \phi_k(\epsilon_k)\epsilon_k \,\mathrm{d}G_k + \varsigma_{k,1}\int \epsilon_k^2 \,\mathrm{d}G_k + \varsigma_{k,2}\int \epsilon_k^3 - \epsilon_k \,\mathrm{d}G_k\right] \\
&= \tilde{a}\left[-1 + \varsigma_{k,1} + \varsigma_{k,2}\mathbb{E}\epsilon_k^3\right] \\
&= 0
\end{aligned}$$

and

$$\begin{aligned}
P_\theta\tilde{a}q(Y,\tilde{X})\kappa(A_{k\bullet v}) &= \tilde{a}\left[\int \phi_k(\epsilon_k)(\epsilon_k^2-1) \,\mathrm{d}G_k + \varsigma_{k,1}\int \epsilon_k^3 - \epsilon_k \,\mathrm{d}G_k + \varsigma_{k,2}\int \epsilon_k^4 - 2\epsilon_k^2 + 1 \,\mathrm{d}G_k\right] \\
&= \tilde{a}\left[\varsigma_{k,1}\mathbb{E}\epsilon_k^3 + \varsigma_{k,2}(\mathbb{E}\epsilon_k^4 - 1)\right] \\
&= 0
\end{aligned}$$

by the choice of ς_k. Moreover, since for any $h \in \mathcal{T}^{\eta|\gamma}_{P_\theta,H}$ we have

$$
\begin{aligned}
&P_\theta\left(\left[a'X\phi_k(A_{k\bullet}v) - a'\mathbb{E}X\left(\phi_k(A_{k\bullet}v) + \varsigma_{k,1}A_{k\bullet}v + \varsigma_{k,2}\kappa(A_{k\bullet}v)\right)\right]h(Y,\tilde{X})\right) \\
&\quad = P_\theta\left(\left[a'(X-\mathbb{E}X)\phi_k(A_{k\bullet}v) - a'\mathbb{E}X\left(\varsigma_{k,1}A_{k\bullet}v + \varsigma_{k,2}\kappa(A_{k\bullet}v)\right)\right]\left[h_0(\tilde{X}) + \sum_{j=1}^{K} h_j(A_{j\bullet}v)\right]\right) \\
&\quad = 0,
\end{aligned}
$$

it follows that

$$
\Pi\left(\dot{\ell}_{\theta,b,l}\middle|\left[\mathcal{T}^{\eta|\gamma}_{P_\theta,H}\right]^{\perp}\right) = \sum_{k=1}^{K}[-A_{k\bullet}D_{b,l}]\left[(x-\mathbb{E}x)\phi_k(A_{k\bullet}v) - \mathbb{E}x\left(\varsigma_{k,1}A_{k\bullet}v + \varsigma_{k,2}\kappa(A_{k\bullet}v)\right)\right].
$$

□

Proof of Lemma 3. We start by showing that $\hat{\phi}_k$ satisfies equation (29). Under P_{θ_n}, we have that $A_{n,k\bullet}(Y_i - B_nX_i) \simeq \epsilon_{i,k} \sim \eta_k$, where $A_{n,k\bullet}$ denotes the kth row of $A_n \equiv A(\alpha_0, \sigma_n)$. Additionally, we can write

$$
\begin{aligned}
\left|\frac{1}{n}\sum_{i=1}^{n}\hat{\phi}_k(\epsilon_{i,k})W_{i,n} - \phi_k(\epsilon_{i,k})W_{i,n}\right| \le{} & \left|\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\phi}_k(\epsilon_{i,k}) - \tilde{\phi}_k(\epsilon_{i,k})\right]W_{i,n}\right| \\
&+ \left|\frac{1}{n}\sum_{i=1}^{n}\left[\tilde{\phi}_k(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})\right]W_{i,n}\right| \qquad \text{(S11)} \\
&+ \left|\frac{1}{n}\sum_{i=1}^{n}\left[\phi_{k,n}(\epsilon_{i,k}) - \phi_k(\epsilon_{i,k})\right]W_{i,n}\right|,
\end{aligned}
$$

where $\hat{\phi}_k(z) = \hat{\gamma}_k' b_k(z)$ as defined in equation (7), $\tilde{\phi}_k(z) := \gamma_k' b_k(z)$, where

$$
\gamma_k = -G_k[b_k b_k']^{-1}G_k c_k \ ,
$$

with G_k being the law corresponding to η_k. Finally, $\phi_{k,n} := \phi_k \mathbf{1}_{[\Xi^L_{k,n},\Xi^U_{k,n}]}$ as in Assumption (2) and ϕ_k is the true log density score. We will show that each of these three terms on the right hand side are $o_G(n^{-1/2})$, where G is the product of G_k and G_w, which implies that

$$
\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\phi}_{k,n}(A_{n,k}Y_i)W_{i,n} - \phi_k(A_{n,k}Y_i)W_{i,n}\right| \xrightarrow{P_{\theta_n}} 0.
$$

For the last term in (S11), by assumption $G_k\{\epsilon_{i,k} \notin [\Xi^L_{k,n}, \Xi^U_{k,n}]\} \downarrow 0$ and hence by independence and Cauchy-Schwarz

$$\begin{aligned} G\left([\phi_{k,n}(\epsilon_{i,k}) - \phi_k(\epsilon_{i,k})]^2 W_{i,n}^2\right) &= G_k\left[\phi_k(\epsilon_{i,k})^2 \mathbf{1}\{\epsilon_{i,k} \notin [\Xi_{k,n}^L, \Xi_{k,n}^U]\}\right] G_w W_{i,n}^2 \\ &\leq \left[G_k \phi_k(\epsilon_{i,k})^4\right]^{1/2} \left[G_k \mathbf{1}\{\epsilon_{i,k} \notin [\Xi_{k,n}^L, \Xi_{k,n}^U]\}\right]^{1/2} G_w W_{i,n}^2 \\ &\to 0. \end{aligned} \tag{S12}$$

By Markov's inequality it follows that for any $\upsilon > 0$,

$$G\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n [\phi_{k,n}(\epsilon_{i,k}) - \phi_k(\epsilon_{i,k})] W_{i,n}\right| > \upsilon\right) \leq \frac{nG\left([\phi_{k,n}(\epsilon_{i,k}) - \phi_k(\epsilon_{i,k})]^2 W_{i,n}^2\right)}{n\upsilon} \to 0.$$

For the second term, we note that by our hypotheses and lemma S6 we have

$$\begin{aligned} G\left([\tilde{\phi}_k(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})]^2 W_{i,n}^2\right) &= G_k\left([\tilde{\phi}_k(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})]^2\right) G_w W_{i,n}^2 \\ &\leq C^2 \delta_{k,n}^6 \|\phi_k^{(3)}\|_\infty^2 G_w W_{i,n}^2 \to 0 \end{aligned}, \tag{S13}$$

as $n \to \infty$, and hence again by Markov's inequality for any $\upsilon > 0$,

$$G\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n [\tilde{\phi}_k(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})] W_{i,n}\right| > \upsilon\right) \leq \frac{nG\left([\tilde{\phi}_k(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})]^2 W_{i,n}^2\right)}{n\upsilon} \to 0.$$

For the first term, by Cauchy-Schwarz

$$\left|\frac{1}{n}\sum_{i=1}^n \left[\hat{\phi}_k(\epsilon_{i,k}) - \tilde{\phi}_k(\epsilon_{i,k})\right] W_{i,n}\right| \leq \|\hat{\gamma}_k - \gamma_k\|_2 \left\|\frac{1}{n}\sum_{i=1}^n b_k(\epsilon_{i,k}) W_{i,n}\right\|_2 = o_G(n^{-1/2}),$$

by lemmas S7 and S8.

Next, we show that $\hat{\phi}_k$ satisfies equation (30). We write:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n \left(\left[\hat{\phi}_k(\epsilon_{i,k}) - \phi_k(\epsilon_{i,k})\right] W_{i,n}\right)^2 \leq \frac{4}{n}\sum_{i=1}^n &\left[\hat{\phi}_k(\epsilon_{i,k}) - \tilde{\phi}_k(\epsilon_{i,k})\right]^2 W_{i,n}^2 \\ &+ \frac{4}{n}\sum_{i=1}^n \left[\tilde{\phi}_k(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})\right]^2 W_{i,n}^2 \\ &+ \frac{4}{n}\sum_{i=1}^n \left[\phi_{k,n}(\epsilon_{i,k}) - \phi_k(\epsilon_{i,k})\right]^2 W_{i,n}^2. \end{aligned} \tag{S14}$$

We will show that (1/4 of) each of the right hand side terms is $o_G(\nu_n)$ under our assumptions, which is sufficient for equation (30) since $A_{k,n}(Y_i - B_n X_i) \simeq \epsilon_{i,k} \sim \eta_k$ under P_{θ_n}. For the last term, for any $\upsilon > 0$, by Markov's inequality, independence and Cauchy-Schwarz we have

$$G\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left[\phi_{k,n}(\epsilon_{i,k}) - \phi_k(\epsilon_{i,k})\right]^2 W_{i,n}^2\right| > \upsilon\nu_n\right) \lesssim \frac{G_k\mathbf{1}\{\epsilon_{i,k} \notin [\Xi_{k,n}^L, \Xi_{k,n}^U]\}G_w W_{i,n}^2}{\upsilon\nu_n} = o(1).$$

For the second term, for any $\upsilon > 0$, by Markov's inequality, independence and lemma S6:

$$\begin{aligned}G\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left[\tilde{\phi}_k(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})\right]^2 W_{i,n}^2\right| > \upsilon\nu_n\right) &\leq \frac{G_k\left([\tilde{\phi}_k(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})]^2\right)G_w W_{i,n}^2}{\upsilon\nu_n}\\ &\leq \frac{C\delta_{k,n}^6\|\phi_k^{(3)}\|_\infty^2 G_w W_{i,n}^2}{\upsilon\nu_n}\\ &= o(1).\end{aligned}$$

Finally, for the first term in the decomposition, by lemma S8 and Assumption 2-part (ii) we have

$$\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\phi}_k(\epsilon_{i,k}) - \tilde{\phi}_k(\epsilon_{i,k})\right]^2 W_{i,n}^2 \leq \|\hat{\gamma}_k - \gamma_k\|_2^2\left[\frac{1}{n}\sum_{i=1}^{n}\|b_k(\epsilon_{i,k})\|_2^2 W_{i,n}^2\right] = o_G(\nu_n).$$

□

S1.1 Supporting results

Definition S1. *Let $\mathcal{C}^k$ denote the space of real functions which have a continuous derivative of order k. Let $\mathcal{C}^\infty := \bigcap_{k\geq 1}\mathcal{C}^k$. Let $\mathcal{C}_c^\infty$ be the subset of $\mathcal{C}^\infty$ consisting of functions $f \in \mathcal{C}^\infty$ such that* $\operatorname{supp}(f)$ *is compact.*[S8]

Lemma S1. *Let μ be a probability measure on $\mathbb{R}$. Then, $\mathcal{C}_c^\infty$ is dense in $L_2(\mu)$.*

Proof. Let $\mathcal{C}_c$ denote the set of compactly supported real functions on $\mathbb{R}$. By theorem 1.1 of Billingsley (1999) and proposition 7.9 of Folland (1999), we have that $\mathcal{C}_c$ is dense in $L_2(\mu)$ and hence it suffices to show that $\mathcal{C}_c^\infty$ is dense in $\mathcal{C}_c$ with respect to the $L_2(\mu)$ norm.[S9] Now, let $g \in \mathcal{C}_c$ and choose $R > 0$ such that $\operatorname{supp}(g) \subset (-R, R) \subset \mathbb{R}$. By the $\mathcal{C}^\infty$ Urysohn lemma (8.18 in Folland, 1999), there is a $h \in \mathcal{C}_c^\infty$ such that $h \in [0,1]$, $h = 1$ on $\operatorname{supp}(g)$ and $\operatorname{supp}(h) \subset (-R, R)$. By the Weierstrass approximation theorem (see e.g. p. 247 of Royden

[S8] The *support* of f is $\operatorname{supp}(f) := \operatorname{cl}\{x : f(x) \neq 0\}$.

[S9] Suppose we have shown this. Then since for each $g \in \mathcal{C}_c$ we have $g \in \operatorname{cl}\mathcal{C}_c^\infty$ and hence $\mathcal{C}_c \subset \operatorname{cl}\mathcal{C}_c^\infty$. Noting that $\mathcal{C}_c$ is dense in $L_2(\mu)$ we obtain the chain of inclusions $L_2(\mu) \subset \operatorname{cl}\mathcal{C}_c \subset \operatorname{cl}\operatorname{cl}\mathcal{C}_c^\infty = \operatorname{cl}\mathcal{C}_c^\infty \subset L_2(\mu)$ where the last inclusion is evident from the fact that any function in $\mathcal{C}_c^\infty$ is bounded and hence in $L_2(\mu)$, which is itself closed.

and Fitzpatrick, 2010) there is a sequence of polynomials $(p_n)_{n\geq 1}$ such that $p_n \to g$ uniformly in $[-R, R]$. Note that the product $p_n h \in \mathcal{C}_c^\infty$. We have that $p_n h \to gh = g$ uniformly on $\operatorname{supp}(h)$. It follows that $\|p_n h - g\|_{\mu,2} \to 0$.[S10] □

Lemma S2. *Let H_k be defined as in (S3). We have that*

$$\operatorname{cl} H_k = \{h_k \in L_2(G_k) : \mathbb{E}h_k(\epsilon_k) = 0, \mathbb{E}\epsilon_k h_k = 0, \mathbb{E}\kappa(\epsilon_k)h_k(\epsilon_k) = 0\},$$

where G_k is the law on $\mathbb{R}$ corresponding to η_k and ϵ_k is distributed according to G_k.

Let H_0 be defined as in (S4). We have that

$$\operatorname{cl} H_0 = \{h_0 \in L_2(G_0) : \mathbb{E}h_0(\tilde{X}) = 0\},$$

where G_0 is the law on $\mathbb{R}^{d-1}$ corresponding to η_0 and $\tilde{X}$ is distributed according to G_0.

Proof. Let $h_k \in \operatorname{cl} H_k$. Then, there are $h_{n,k} \in H_k \subset L_2(G_k)$ with $\|h_{n,k} - h_k\|_{G_k,2} \to 0$. Hence, $h_k \in L_2(G_k)$. Since the inner product is continuous we have

$$\mathbb{E}h_k(\epsilon_k)\xi(\epsilon_k) = \langle h_k(\epsilon_k),\ \xi(\epsilon_k)\rangle_{G_k} = \lim_{n\to\infty} \langle h_{n,k}(\epsilon_k),\ \xi(\epsilon_k)\rangle_{G_k} = \lim_{n\to\infty} 0 = 0,$$

for each $\xi \in \{\xi_0, \xi_1, \kappa\}$, where $\xi_0(x) := 1$, $\xi_1(x) := x$. Hence, $h_k \in \{h_k \in L_2(G_k) : \mathbb{E}h_k(\epsilon_k) = 0, \mathbb{E}\epsilon_k h_k = 0, \mathbb{E}\kappa(\epsilon_k)h_k(\epsilon_k) = 0\}$ and thus we have that $\operatorname{cl} H_k \subset \{h_k \in L_2(G_k) : \mathbb{E}h_k(\epsilon_k) = 0, \mathbb{E}\epsilon_k h_k = 0, \mathbb{E}\kappa(\epsilon_k)h_k(\epsilon_k) = 0\}$.

For the other inclusion, h_k be in $L_2(G_k)$ and orthogonal to each $\xi \in \{\xi_0, \xi_1, \kappa\}$. We want to approximate (in the $L_2(G_k)$ norm) h_k by functions in H_k. First ignore the orthogonality constraints: the space $\mathcal{C}_c^\infty$ (see definition S1) (of which $\mathcal{C}_b^1(\lambda) \subset L_2(G_k)$ is a superset) is dense in $L_2(G_k)$ by lemma S1. Hence there is a sequence $(h_{n,k})_{n\geq 1}$ in $\mathcal{C}_b^1(\lambda)$ such that $\|h_{n,k} - h_k\|_{G_k,2} \to 0$. Introduce the function

$$\tilde{h}_{n,k}(z) := h_{n,k}(z) + \upsilon_n + \nu_n v(z) + \omega_n w(z),$$

where each of υ_n, ν_n and ω_n are in $\mathbb{R}$ and $v, w \in \mathcal{C}_b^1(\lambda)$ are such that

$$\mathbb{E}v(\epsilon_k) = \mathbb{E}w(\epsilon_k) = 0, \quad \mathbb{E}\epsilon_k w(\epsilon_k) = \mathbb{E}\kappa(\epsilon_k)v(\epsilon_k) = 0, \quad \mathbb{E}\epsilon_k v(\epsilon_k) = \mathbb{E}\kappa(\epsilon_k)w(\epsilon_k) = 1,$$

and the existence of such functions is guaranteed by lemma S5. It is clear from its definition

[S10] Fix $\epsilon > 0$. Take N large enough that for all $n \geq N$ we have $|p_n h - g| < \epsilon$ on $\operatorname{supp}(h)$. Then, $\int (p_n h - g)^2 \,\mathrm{d}\mu = \int_{\operatorname{supp}(h)} (p_n h - g)^2 \,\mathrm{d}\mu + \int_{\mathbb{R}\setminus \operatorname{supp}(h)} (p_n h - g)^2 = \int_{\operatorname{supp}(h)} (p_n h - g)^2 \,\mathrm{d}\mu < \epsilon^2$ since $p_n h - g = 0$ outside of $\operatorname{supp}(h)$.

that $\tilde{h}_{n,k} \in \mathcal{C}_b^1(\lambda)$. Now, put

$$\upsilon_n := -\mathbb{E}h_{n,k}(\epsilon_k), \quad \nu_n := -\mathbb{E}[h_{n,k}(\epsilon_k)\epsilon_k], \quad \omega_n := -\mathbb{E}[h_{n,k}\kappa(\epsilon_k)].$$

Then, we clearly have that

$$\langle \tilde{h}_{n,k},\ \xi_0\rangle_{G_k} = \mathbb{E}\left[h_{n,k}(\epsilon_k) + \upsilon_n\right] = \mathbb{E}h_{n,k}(\epsilon_k) - \mathbb{E}h_{n,k}(\epsilon_k) = 0,$$

$$\langle \tilde{h}_{n,k},\ \xi_1\rangle_{G_k} = \mathbb{E}\left[h_{n,k}(\epsilon_k)\epsilon_k + \nu_n\mathbb{E}[v(\epsilon_k)\epsilon_k]\right] = \mathbb{E}\left[h_{n,k}(\epsilon_k)\epsilon_k\right] - \mathbb{E}[h_{n,k}(\epsilon_k)\epsilon_k] = 0$$

$$\langle \tilde{h}_{n,k},\ \kappa\rangle_{G_k} = \mathbb{E}\left[h_{n,k}\kappa(\epsilon_k) + \omega_n\mathbb{E}[w(\epsilon_k)\kappa(\epsilon_k)]\right] = \mathbb{E}[h_{n,k}\kappa(\epsilon_k)] - \mathbb{E}[h_{n,k}\kappa(\epsilon_k)] = 0.$$

Moreover, since $h_{n,k} \xrightarrow{L_2(G_k)} h_k$ we have that $(\upsilon_n, \nu_n, \omega_n) \to 0$ as $n \to \infty$. Therefore,

$$\begin{aligned}\|\tilde{h}_{n,k} - h_k\|_{G_k,2} &\leq \|h_{n,k} - h_k\|_{G_k,2} + \|\upsilon_n + \nu_n v + \omega_n w\|_{G_k,2} \\ &\leq \|h_{n,k} - h_k\|_{G_k,2} + |\upsilon_n| + |\nu_n|\|v\|_{G_k,2} + |\omega_n|\|w\|_{G_k,2} \\ &\to 0,\end{aligned}$$

as $n \to \infty$ where we note that $\|v\|_{G_k,2} < \infty$ and $\|w\|_{G_k,2} < \infty$ since the functions are bounded λ-a.e. (and hence G_k-a.s.). Thus $(\tilde{h}_{n,k})_{n\geq 1}$ is a sequence in H_k such that $\|\tilde{h}_{n,k} - h_k\|_{G_k,2} \to 0$ and we conclude that $\{h_k \in L_2(G_k) : \mathbb{E}h_k(\epsilon_k) = 0, \mathbb{E}\epsilon_k h_k = 0, \mathbb{E}\kappa(\epsilon_k)h_k(\epsilon_k) = 0\} \subset \operatorname{cl} H_k$.

For H_0 let $h_0 \in \operatorname{cl} H_0$. There are $(h_{n,0} \in H_0 \subset L_2(G_0)$ with $\|h_{n,0} - h_0\|_{G_0,2} \to 0$. Hence $h_0 \in L_2(G_0)$ and $\int h_0 \,\mathrm{d}G_0 = \lim_{n\to\infty} \int h_{n,0} \,\mathrm{d}G_0 = 0$. Conversely, suppose that $h_0 \in L_2^0(G_0)$. Since $C_b(\lambda, \mathbb{R}^{d-1}) \subset L_2(G_0)$ is a superset of the compactly supported continuous functions on $\mathbb{R}^{d-1}$ (when considered as elements of $L_2(G_0)$) it is dense in $L_2(G_0)$ by e.g. Theorem 3.14 in Rudin (1987). Hence there exists a sequence $(h_{n,0})_{n\geq 1} \subset C_b(\lambda, \mathbb{R}^{d-1})$ with $\|h_{n,0} - h_0\|_{G_0,2} \to 0$. This implies that $0 = \int h_0 \,\mathrm{d}G_0 = \lim_{n\to\infty} \int h_{n,0} \,\mathrm{d}G_0$ and so also $\|\tilde{h}_{n,0} - h_0\|_{G_0,2} \to 0\|_{G_0,2} \to 0$ where $\tilde{h}_{n,0} := h_{n,0} - \int h_{n,0} \,\mathrm{d}G_0 \in H_0$, implying that $h_0 \in \operatorname{cl} H_0$. □

Lemma S3. *Let $\tilde{H}_k^\gamma$ be defined as in the proof of Lemma 2. We have that*

$$\mathscr{T} = \operatorname{cl}\left(\tilde{H}_1^\gamma + \cdots + \tilde{H}_K^\gamma\right),$$

and

$$\operatorname{cl} \mathcal{T}_{P_\theta,H}^{\eta|\gamma} = \operatorname{cl} \tilde{H}_0^\gamma + \operatorname{cl} \tilde{H}_1^\gamma + \cdots + \operatorname{cl} \tilde{H}_K^\gamma = \operatorname{cl} \tilde{H}_0^\gamma + \mathscr{T}.$$

Proof. For the first display, the sets in the sum on the right hand side are pairwise orthogonal. Note that we have for any $k, j \in [K]$ and any $(h_j, h_k) \in H_j \times H_k$,

$$\langle h_j(A_j v),\ h_k(A_k v)\rangle_{P_\theta} = P_\theta h_j(A_j v)h_k(A_k v) = \mathbb{E}h_j(\epsilon_j)h_k(\epsilon_k) = \mathbb{E}h_j(\epsilon_j)\mathbb{E}h_k(\epsilon_k) = 0,$$

due to the independence of the elements of ϵ. So $y \mapsto h_j(A_j v) \in [\tilde{H}_k^\gamma]^\perp = [\mathrm{cl}\, \tilde{H}_k^\gamma]^\perp$.[S11] Recalling that the sum of closed pairwise orthogonal subspaces is closed,[S12] we conclude that $\mathrm{cl}\left(\tilde{H}_1^\gamma + \cdots + \tilde{H}_K^\gamma\right) \subset \mathrm{cl}\, \tilde{H}_1^\gamma + \cdots + \mathrm{cl}\, \tilde{H}_K^\gamma = \mathscr{T}$ since the closure of a set is the smallest closed set containing that set. For the opposite inclusion, let $g = \sum_{k=1}^K g_k \in \mathscr{T}$ and note there are $g_{i,n}(y) = h_{i,n}(A_i v) \in \tilde{H}_i^\gamma$ such that each $g_{i,n} \to g_i$ in $L_2(P_\theta)$. Let $g_n = \sum_{k=1}^K g_{k,n}$. Clearly this is in $\tilde{H}_1^\gamma + \cdots + \tilde{H}_K^\gamma$ and hence its limit g is in $\mathrm{cl}\left(\tilde{H}_1^\gamma + \cdots + \tilde{H}_K^\gamma\right)$. Thus $\mathscr{T} \subset \mathrm{cl}\left(\tilde{H}_1^\gamma + \cdots + \tilde{H}_K^\gamma\right)$. The second display is analogous, noting the independence between $\tilde{X}$ and ϵ. □

Lemma S4. *We have*

$$\mathscr{L} \subset \mathscr{T}^\perp,$$

where both are as defined in the proof of Lemma 2.

Proof. Suppose that $y \mapsto f(y)$ is in $\mathscr{L}_0$ and let $y \mapsto \sum_{k=1}^K h_k(A_k v) \in \tilde{H}_1^\gamma + \cdots + \tilde{H}_K^\gamma$.[S13] We have

$$\left\langle f(Y),\ \sum_{k=1}^K h_k(A_k V) \right\rangle_{P_\theta} = \sum_{k=1}^K \langle f(Y),\ h_k(A_k V)\rangle_{P_\theta},$$

where $V = Z - BX$ so it suffices to show that $\langle f(Y),\ h_k(A_k V)\rangle_{P_\theta} = 0$ for any $k \in [K]$ and any $h_k \in H_k$. First suppose that $f(Y) \in \{A_k V, \kappa(A_k V)\}$. Then, by the definition of H_k we have

$$\langle f(Y),\ h_k(A_k V)\rangle_{P_\theta} = \int f(y) h_k(A_k v)\, \mathrm{d}P_\theta = P_\theta[f(Y) h_k(A_k V)] = 0.$$

Second suppose that $f(Y) \in \{A_l V, \kappa(A_l V)\}$ for some $l \neq k$. Then we have that

$$\langle f(Y),\ h_k(A_k V)\rangle_{P_\theta} = \int f(y) h_k(A_k v)\, \mathrm{d}P_\theta = P_\theta[f(Y) h_k(A_k V)] = P_\theta f(Y) P_\theta h_k(A_k V) = 0,$$

by the independence of $A_k V = \epsilon_k$ and $A_l V = \epsilon_l$ and the fact that by the definition of H_k we have $P_\theta[h_k(A_k V)] = 0$. Now, let $i \neq j$ with both in $[K]$ and suppose that $f(Y) = \phi_i(A_i V) A_j V$. If $k = i \neq j$ we have

$$\langle f(Y),\ h_k(A_k V)\rangle_{P_\theta} = P_\theta[\phi_i(A_i V) A_j V h_k(A_k V)] = P_\theta[\phi_k(A_k V) h_k(A_k V)] P_\theta[A_j V] = 0,$$

by independence of $A_k V$ and $A_j V$ and that $P_\theta[A_j V] = 0$. If $k = j \neq i$,

$$\langle f(Y),\ h_k(A_k V)\rangle_{P_\theta} = P_\theta[\phi_i(A_i V) A_j V h_k(A_k V)] = P_\theta[h_k(A_k V) A_k V] P_\theta[\phi_i(A_i V)] = 0,$$

[S11] Note that for any Hilbert space V and a linear subspace U of V, $U^\perp = [\mathrm{cl}\, U]^\perp$.
[S12] See e.g. II.3.4 in Conway (1985).
[S13] See Lemma S3.

by independence of $A_k V$ and $A_i V$ and the definition of h_k. Lastly, if $k \neq j \neq i$ then

$$\langle f(Y),\ h_k(A_k V)\rangle_{P_\theta} = P_\theta[\phi_i(A_i V) A_j V h_k(A_k V)] = P_\theta[h_k(A_k V)] P_\theta[A_j V] P_\theta[\phi_i(A_i V)] = 0,$$

by independence of $A_k V, A_j V, A_i V$ and $P_\theta[A_j V] = 0$. □

Lemma S5. *Let $\kappa(x) := x^2 - 1$ and let $L_2(G_k)$ denote the space of functions from $\mathbb{R} \to \mathbb{R}$ square-integrable with respect to the probability measure G_k, which is absolutely continuous with respect to Lebesgue measure, λ. Let $\mathcal{C}_b^1(\lambda) \subset L_2(G_k)$ denote the subspace of functions which are bounded and continuously differentiable with bounded derivatives λ-a.e. Suppose that $\kappa \in L_2(G_k)$, $\int z \, \mathrm{d}G_k = \int \kappa(z) \, \mathrm{d}G_k = 0$ and $\int \kappa(z)^2 \, \mathrm{d}G_k > 0$. Then, there are functions $v, w \in \mathcal{C}_b^1(\lambda)$ such that*

$$\int v(z) \, \mathrm{d}G_k = \int w(z) \, \mathrm{d}G_k = 0,$$

$$\int z w(z) \, \mathrm{d}G_k = \int \kappa(z) v(z) \, \mathrm{d}G_k = 0$$

and

$$\int z v(z) = \int \kappa(z) w(z) \, \mathrm{d}G_k = 1.$$

Proof. We first note that the requirement that v, w be mean zero is easily met, once we have $\tilde{v}, \tilde{w}$ satisfying the other required properties. Suppose that is the case, then put $v := \tilde{v} - \int \tilde{v}(z) \, \mathrm{d}G_k$ and likewise for w. Clearly these are zero mean. Moreover, they are bounded and continuously differentiable with bounded derivative λ-a.e. and the inner product conditions also continue to hold in view of the assumption that $\int z \, \mathrm{d}G_k = \int \kappa(z) \, \mathrm{d}G_k = 0$. Therefore, we now construct $\tilde{v}, \tilde{w}$ ignoring the zero-mean requirement.

We start with $\tilde{v}$. Let $a < b < c$ and define

$$M := \begin{pmatrix} \int_a^b z \, \mathrm{d}G_k & \int_b^c z \, \mathrm{d}G_k \\ \int_a^b (z^2 - 1) \, \mathrm{d}G_k & \int_b^c (z^2 - 1) \, \mathrm{d}G_k \end{pmatrix}.$$

Provided M^{-1} exists there must exist a $v^* = (v_1^*, v_2^*)'$ such that $M v^* = (1, 0)'$. Then, we can define

$$\tilde{v}(z) := \begin{cases} v_1^* & \text{if } z \in [a, b) \\ v_2^* & \text{if } z \in [b, c) \\ 0 & \text{otherwise} \end{cases},$$

to yield

$$\begin{pmatrix} \int z\tilde{v}(z)\,\mathrm{d}G_k \\ \int (z^2-1)\tilde{v}(z)\,\mathrm{d}G_k \end{pmatrix} = \begin{pmatrix} v_1^* \int_a^b z\,\mathrm{d}G_k + v_2^* \int_b^c z\,\mathrm{d}G_k \\ v_1^* \int_a^b (z^2-1)\,\mathrm{d}G_k + v_2^* \int_b^c (z^2-1)\,\mathrm{d}G_k \end{pmatrix} = Mv^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

as required. It remains to demonstrate that there are a, b, c such that M^{-1} exists. To see that this is always possible, note first that since $\int z\,\mathrm{d}G_k = 0$ and $\int z^2\,\mathrm{d}G_k = 1$, G_k must place mass both on the negative and positive parts of the real line. Since also $\int (z^2-1)\,\mathrm{d}G_k = 0$ and $\int (z^2-1)^2\,\mathrm{d}G_k > 0$ at least one of $G_k([-1,0)) > 0$ or $G_k([0,1)) > 0$ must hold. Without loss of generality assume the latter.[S14] Take $a < 0$ such that $G_k((a,0)) > 0$. Take $b = 0$ and $c < 1$ such that $G_k([0,c)) > 0$ and $G_k([c,1)) > 0$. Note that this ensures that $\int_a^b z\,\mathrm{d}G_k < 0$ and $\int_b^c (z^2-1)\,\mathrm{d}G_k < 0$, so neither of the rows are 0. Now, either M is non-singular and we are done or there is a $\tau \neq 0$ such that $\int_a^b z\,\mathrm{d}G_k = \tau \int_a^b (z^2-1)\,\mathrm{d}G_k$ and $\int_b^c z\,\mathrm{d}G_k = \tau \int_b^c (z^2-1)\,\mathrm{d}G_k$. If $\tau > 0$, adjust c upwards to $c^* \in (c,1)$ such that $G_k([c,c^*)) > 0$. We have

$$\int_b^{c^*} z\,\mathrm{d}G_k > \int_b^c z\,\mathrm{d}G_k = \tau \int_b^c (z^2-1)\,\mathrm{d}G_k > \tau \int_b^{c^*} (z^2-1)\,\mathrm{d}G_k.$$

If $\tau < 0$, adjust c downwards to $c' > 0$ with $c' < c$ such that $G_k([c',c)) > 0$. We have

$$\int_b^{c'} z\,\mathrm{d}G_k < \int_b^c z\,\mathrm{d}G_k = \tau \int_b^c (z^2-1)\,\mathrm{d}G_k < \tau \int_b^{c'} (z^2-1)\,\mathrm{d}G_k.$$

Since $\int_a^b z\,\mathrm{d}G_k = \tau \int_a^b (z^2-1)\,\mathrm{d}G_k$ continues to hold, the two rows are now linearly independent and hence M is invertible.

We have constructed a $\tilde{v} \in \mathcal{C}_b^1(\lambda)$ satisfying the required conditions. The construction for $\tilde{w}$ can be perfomed analogously, taking $w^* := M^{-1}(0,1)'$. □

Lemma S6 (Cf. Lemma A.5, Chen and Bickel, 2006)**.** *Let $\tilde{\phi}_k(z) = \gamma_k' b_k$, with $\gamma_k = -G_k[b_k b_k']^{-1} G_k c_k$ and $\phi_{k,n}$ is defined as in Assumption 2. If part (iv) of Assumption 2 holds, we have*

$$G_k\left(\tilde{\phi}_k(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})\right)^2 \le C^2 \delta_{k,n}^6 \|\phi_{k,n}^{(3)}\|_\infty^2.$$

Proof. By the definition of $\tilde{\phi}_k$ and lemma S10 we have

$$G_k\left(\tilde{\phi}_k(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})\right)^2 = \inf_{g \in \mathscr{G}_k(\xi_{k,n})} G_k\left(g(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})\right)^2 \le C^2 \delta_{k,n}^6 \|\phi_{k,n}^{(3)}\|_\infty^2.$$

[S14] If instead $G_k([0,1)) = 0$, an analogous argument can be made, interchanging the roles of a and c.

The first inequality comes from the fact that we can equivalently see $\gamma_k = -G_k[b_k b_k']^{-1} G_k c_k$ as the solution to minimizing

$$\begin{aligned}\int (\phi_k(z) - \gamma_k' b_k(z))^2 \eta_k(z)\,\mathrm{d}z &= \int \phi_k^2\,\mathrm{d}G_k + \int (\gamma_k' b_k)^2\,\mathrm{d}G_k + 2\int \gamma_k' c_k(z)\eta_k(z)\,\mathrm{d}z \\ &= G_k\phi_k^2 + \gamma_k' G_k[b_k b_k']\gamma_k + 2\gamma_k' G_k c_k.\end{aligned} \tag{S15}$$

where we only integrate over the support of $\phi_{k,n}$ since this is also the support of b_k and c_k. □

Lemma S7 (Cf. Lemma A.3, Chen and Bickel, 2006)**.** *Under assumptions 1 and 2, and that $W_{i,n}$ is independent of $\epsilon_{i,k}$ we have*

$$\left\| \frac{1}{n}\sum_{i=1}^n b_k(\epsilon_{i,k}) W_{i,n} \right\|_2 = O_G(n^{-1/2}).$$

Proof. By the fact that $\sum_{m=1}^{B_k} b_{m,k}(x)^2 \leq 1$ (see e.g. (36) on p. 96 of de Boor, 2001) and the given assumptions we have that

$$G\left(\left\| \frac{1}{n}\sum_{i=1}^n b_k(\epsilon_{i,k}) W_{i,n} \right\|_2^2 \right) = \frac{1}{n} G_k \left(\sum_{m=1}^{B_k} b_{m,k}(\epsilon_{i,k})^2 \right) G_w W_{i,n}^2 \leq \frac{G_w W_{i,n}^2}{n}$$

Fix $\epsilon > 0$ and take $M > 0$ large enough such that $G_w W_{i,n}^2 / M^2 < \epsilon$. Markov's inequality yields

$$G\left(\sqrt{n} \left\| \frac{1}{n}\sum_{i=1}^n b_k(\epsilon_{i,k}) W_{i,n} \right\|_2 > M \right) \leq \frac{G\left(n \left\| \frac{1}{n}\sum_{i=1}^n b_k(\epsilon_{i,k}) W_{i,n} \right\|_2^2 \right)}{M^2} \leq \frac{G_w W_{i,n}^2}{M^2} < \epsilon.$$

□

Lemma S8 (Cf. Lemma A.2, Chen and Bickel, 2006)**.** *Let $\hat{\gamma}_k$ be as defined in equation (7) and $\gamma_k = -G_k[b_k b_k']^{-1} G_k c_k$. Suppose that Assumptions 1 and 2 hold. Then, if we define*

$$\hat{\Gamma}_{k,n} := \frac{1}{n}\sum_{i=1}^n b_k(\epsilon_{i,k}) b_k(\epsilon_{i,k})', \quad \Gamma_{k,n} := G_k b_k b_k',$$

and

$$\hat{C}_{k,n} := \frac{1}{n}\sum_{i=1}^n c_{k,n}(\epsilon_{i,k}), \quad C_{k,n} := G_k c_k,$$

we have that

(i) $\|C_{k,n}\|_2 = O(\delta_{k,n} B_k^{1/2})$,

(ii) $\|\hat{C}_{k,n} - C_{k,n}\|_2 = O_G\left(\sqrt{\frac{B_k \log B_k}{n\delta_{k,n}^2}}\right)$,

(iii) $\|\hat{\Gamma}_{k,n} - \Gamma_{k,n}\|_2 = O_G\left(\sqrt{\frac{B_k \log B_k}{n}}\right)$,

(iv) $\|\Gamma_{k,n}\|_2 = O(\delta_{n,k})$

(v) $\|\Gamma_{k,n}^{-1}\|_2 = O(\delta_{k,n}^{-2})$.

In particular, $\|\hat{\gamma}_k - \gamma_k\|_2 = O_G(n^{-1/2}\Delta_{k,n}\delta_{k,n}^{-4}(\Delta_{k,n}\delta_{k,n}^{-1})^{\iota}) = o_G(1)$ *and* $\|\hat{\Gamma}_{k,n}\|_2 = o_G(1)$.

Proof. The proof follows the relevant parts of the proof of lemma A.2 in Chen and Bickel (2006). Firstly, from the representation of the derivative of the cubic spline (e.g. de Boor, 2001) we can write $c_{k,i} = \left(b_{k,i}^{(3)} - b_{k,i+1}^{(3)}\right)/\delta_{k,n}$. We have, for large enough $n \in \mathbb{N}$,

$$\begin{aligned}
|C_{k,n,i}| = |G_k c_{k,i}| &= \delta_{k,n}^{-1}\left|\int b_{k,i}^{(3)}(t)\eta_k(t)\,\mathrm{d}t - \int b_{k,i+1}^{(3)}(t)\eta_k(t)\,\mathrm{d}t\right| \\
&= \delta_{k,n}^{-1}\left|\int b_{k,i}^{(3)}(t)\eta_k(t)\,\mathrm{d}t - \int b_{k,i}^{(3)}(t)\eta_k(t+\delta_{k,n})\,\mathrm{d}t\right| \\
&\leq \left|\int b_{k,i}^{(3)}(t)\frac{\eta_k(t+\delta_{k,n}) - \eta_k(t)}{\delta_{k,n}}\,\mathrm{d}t\right| \\
&\leq 2\|\eta_k'\|_\infty \int b_{k,i}^{(3)}(t)\,\mathrm{d}t \\
&\leq 6\|\eta_k'\|_\infty \delta_{k,n},
\end{aligned}$$

where the last inequality is due to (20) on p. 91 in de Boor (2001) and the fact that splines (of any order) take values in $[0,1]$.[S15] It follows immediately that for large enough $n \in \mathbb{N}$,

$$\sum_{i=1}^{B_k} C_{k,n,i}^2 \leq \sum_{i=1}^{B_k} 6^2\|\eta_k'\|_\infty^2 \delta_{k,n}^2 = B_k 6^2 \|\eta_k'\|_\infty^2 \delta_{k,n}^2,$$

from which (i) follows.

We have that $c_{k,i} = \left(b_{k,i}^{(3)} - b_{k,i+1}^{(3)}\right)/\delta_{k,n}$ and since splines (of any order) take values in $[0,1]$ (both as noted above), we have that $c_{k,i} \in [-\delta_{k,n}^{-1}, \delta_{k,n}^{-1}]$. Hence, by Hoeffdings's inequality for $t \geq 0$ we have

$$G\left(\left|\frac{1}{n}\sum_{i=1}^{n} c_{k,m}(\epsilon_{i,k}) - G_k c_{k,m}\right| \geq t\right) \leq 2\exp\left(\frac{-n^2t^2}{2n\delta_{k,n}^{-2}}\right) = 2\exp(-nt^2\delta_{k,n}^2/2).$$

[S15] This is evident from their definition. See also property (36) (p. 96) of de Boor (2001).

Therefore,

$$G\left(\|\hat{C}_{k,n} - C_{k,n}\|_2 \geq t\right) \leq \sum_{m=1}^{B_k} G\left(\left|\frac{1}{n}\sum_{i=1}^{n} c_{k,m}(\epsilon_{i,k}) - G_k c_{k,m}\right| \geq \frac{t}{\sqrt{B_k}}\right)$$
$$\leq 2B_k \exp(-nt^2 B_k^{-1}\delta_{k,n}^2/2),$$

and so for any fixed $\epsilon > 0$ we can take $t = \sqrt{\frac{4B_k \log B_k}{n\delta_{k,n}^2}}$ to obtain

$$G\left(\|\hat{C}_{k,n} - C_{k,n}\|_2 \geq t\right) \leq 2B_k^{-1} \to 0,$$

yielding (ii).

Since for any $m, s \in [B_k]$ we have $b_{k,m}b_{k,s} \in [0,1]$ by Hoeffding's inequality it follows that for any $t \geq 0$

$$G\left(\left|\frac{1}{n}\sum_{i=1}^{n} b_{k,m}(\epsilon_{i,k})b_{k,s}(\epsilon_{i,k}) - G_k b_{k,m}b_{k,s}\right| \geq t\right) \leq 2\exp\left(\frac{-2n^2t^2}{n}\right) = 2\exp(-2nt^2).$$

Therefore, since $\|\hat{\Gamma}_{k,n} - \Gamma_{k,n}\|_2 \leq \|\hat{\Gamma}_{k,n} - \Gamma_{k,n}\|_F$ and both $\hat{\Gamma}_{k,n}$ and $\Gamma_{k,n}$ are zero for all (m,s) entries where $|m-s| > 3$ (de Boor, 2001, (20), p. 91) we have that

$$\begin{aligned}
&G\left(\|\hat{\Gamma}_{k,n} - \Gamma_{k,n}\|_2 \geq t\right)\\
&\quad\leq G\left(\|\hat{\Gamma}_{k,n} - \Gamma_{k,n}\|_F \geq t\right)\\
&\quad\leq \sum_{m=1}^{B_k}\sum_{s=\max(m-3,1)}^{\min(B_k,m+3)} G\left(\left|\frac{1}{n}\sum_{i=1}^{n} b_{k,m}(\epsilon_{i,k})b_{k,s}(\epsilon_{i,k}) - G_k b_{k,m}b_{k,s}\right| \geq \frac{t}{\sqrt{7B_{k,n}}}\right)\\
&\quad\leq 14B_k \exp\left(\frac{-2nt^2}{7B_k}\right).
\end{aligned}$$

Putting $t = \sqrt{\frac{7B_k \log B_k}{n}}$ we obtain

$$G\left(\|\hat{\Gamma}_{k,n} - \Gamma_{k,n}\|_2 \geq t\right) \leq 14B_k^{-1} \to 0,$$

yielding (iii).

Since $\Gamma_{k,n}$ is symmetric and positive (semi-)definite we have that $\|\Gamma_{k,n}\|_2 \leq \|\Gamma_{k,n}\|_\infty =$

$\max_{m=1,\ldots,B_k}\sum_{s=1}^{B_k} G_k b_{k,m} b_{k,s}$.[S16] Then, since for any $z\in\mathbb{R}$, each row of $b_k(z)b_k(z)'$ has at most 7 non-zero entries,[S17] all of which are bounded above by 1 we have

$$
\begin{aligned}
\|\Gamma_{k,n}\|_2 &\leq \max_{m=1,\ldots,B_k}\sum_{s=1}^{B_k} G_k b_{k,m} b_{k,s} \\
&= \max_{m=1,\ldots,B_k}\sum_{s=1}^{B_k}\int_{\xi_{k,n,m}}^{\xi_{k,n,m+4}} b_{k,m}(z) b_{k,s}(z)\eta_k(z)\,\mathrm{d}z \\
&\leq \max_{m=1,\ldots,B_{k,n}} 7\|\eta_k\|_\infty 4\delta_{k,n} \\
&= 28\|\eta_k\|_\infty \delta_{k,n},
\end{aligned}
$$

which yields (iv) in conjunction with requirement (iii) of Assumption 2.

By Assumption 2 part (v), on $[\Xi^L_{k,n},\Xi^U_{k,n}]$ we have $\eta(x)\geq c\delta_{k,n}$. Hence $\eta(x)-c\delta_{k,n}\geq 0$ and so $\int b_k b_k'(\eta-c\delta_{k,n})\lambda = \int (b_k\sqrt{\eta-c\delta_{k,n}})(b_k\sqrt{\eta-c\delta_{k,n}})'\lambda$. Note that the functions $b_{k,i}\sqrt{\eta-c\delta_{k,n}}$ satisfy $\int (b_{k,i}\sqrt{\eta-c\delta_{k,n}})^2\,\mathrm{d}\lambda<\infty$ and hence belong to $L_2(\lambda)$. It follows that the matrix $\int b_k b_k'(\eta-c\delta_k)\lambda$ is a Gram matrix and hence positive semi-definite. This implies that $\Gamma_{k,n}\succeq c\delta_{k,n}\tilde{\Gamma}_{k,n}$ where $\tilde{\Gamma}_{k,n}$ is defined as in lemma S9. Hence, by the Rayleigh quotient theorem (see e.g. Theorem 4.2.2 in Horn and Johnson, 2013) and lemma S9

$$
\lambda_{\min}(\Gamma_{k,n})\geq\lambda_{\min}(c\delta_{k,n}\tilde{\Gamma}_{k,n}) = c\delta_{k,n}\lambda_{\min}(\tilde{\Gamma}_{k,n})\geq cv\delta^2_{k,n},
$$

for a $v>0$, from which we may conclude that

$$
\|\Gamma^{-1}_{k,n}\|_2 = \frac{1}{\lambda_{\min}(\Gamma_{k,n})}\leq (cv)^{-1}\delta^{-2}_{k,n},
$$

which yields (v).

To demonstrate the last claim, note that with the results just derived, under our assumptions we have,

$$
\|\hat{C}_{k,n}\|_2\leq\|\hat{C}_{k,n}-C_{k,n}\|_2+\|C_{k,n}\|_2 = O_G\left(\sqrt{\frac{B_k\log B_k}{n\delta^2_{k,n}}}\right)+O\left(\delta_{k,n}\sqrt{B_k}\right) = O_G\left(\delta_{k,n}\sqrt{B_k}\right),
$$

[S16] See e.g. Theorem 5.6.9 in Horn and Johnson (2013).

[S17] $b_{k,m}(z)=0$ outside $[\xi_{k,n,m},\xi_{k,n,m+4})$. See (20) on p. 91 in de Boor (2001).

and, using inequality (5.8.2) from Horn and Johnson (2013),

$$
\begin{aligned}
\|\hat{\Gamma}_{k,n}^{-1}\|_2 &\leq \|\Gamma_{k,n}^{-1}(I + [\hat{\Gamma}_{k,n} - \Gamma_{k,n}]\Gamma_{k,n}^{-1})^{-1}\|_2 \\
&\leq \|\Gamma_{k,n}^{-1}\|_2 \|(I + [\hat{\Gamma}_{k,n} - \Gamma_{k,n}]\Gamma_{k,n}^{-1})^{-1}\|_2 \\
&\leq \|\Gamma_{k,n}^{-1}\|_2 \left(1 - \|[\hat{\Gamma}_{k,n} - \Gamma_{k,n}]\Gamma_{k,n}^{-1}\|_2\right)^{-1} \\
&\leq \|\Gamma_{k,n}^{-1}\|_2 \left(1 - \|\hat{\Gamma}_{k,n} - \Gamma_{k,n}\|_2 \|\Gamma_{k,n}^{-1}\|_2\right)^{-1} \\
&= O_G(\delta_{k,n}^{-2}).
\end{aligned} \tag{S16}
$$

Using these intermediate results along with (ii) - (v) and our hypotheses we obtain that

$$
\begin{aligned}
\|\hat{\gamma}_k - \gamma_k\|_2 &= \|\hat{\Gamma}_{k,n}^{-1}\hat{C}_{k,n} - \Gamma_{k,n}^{-1}C_{k,n}\|_2 \\
&\leq \|(\hat{\Gamma}_{k,n}^{-1} - \Gamma_{k,n}^{-1})\hat{C}_{k,n}\|_2 + \|\Gamma_{k,n}^{-1}(\hat{C}_{k,n} - C_{k,n})\|_2 \\
&\leq \|\Gamma_{k,n}^{-1}\|_2 \|\Gamma_{k,n} - \hat{\Gamma}_{k,n}\|_2 \|\hat{\Gamma}_{k,n}^{-1}\|_2 \|\hat{C}_{k,n}\|_2 + \|\Gamma_{k,n}^{-1}\|_2 \|\hat{C}_{k,n} - C_{k,n}\|_2 \\
&= O_G\left(\sqrt{\frac{B_k^2 \log B_k}{\delta_{k,n}^6 n}}\right) + O_G\left(\sqrt{\frac{B_k \log B_k}{\delta_{k,n}^6 n}}\right) \\
&= o_G(1),
\end{aligned}
$$

by Assumption 2 part (ii), since we have $B_k \leq \Delta_{k,n}\delta_{k,n}^{-1}$ and hence the dominant term above vanishes since for all large enough n,

$$
\sqrt{\frac{B_k^2 \log B_k}{\delta_{k,n}^6 n}} \leq n^{-1/2}\Delta_{k,n}\delta_{k,n}^{-4}\log(\Delta_{k,n}\delta_{k,n}^{-1}) \leq n^{-1/2}\Delta_{k,n}\delta_{k,n}^{-4}(\Delta_{k,n}\delta_{k,n}^{-1})^{\iota} = o(1).
$$

Finally, by (iii) and (iv) and Assumption 2 part (ii) we have

$$
\|\hat{\Gamma}_{k,n}\|_2 \leq \|\hat{\Gamma}_{k,n} - \Gamma_{k,n}\|_2 + \|\Gamma_{k,n}\|_2 = O_G\left(\sqrt{\frac{B_{k,n}\log B_k}{n}}\right) + O(\delta_{k,n}) = o_G(1),
$$

since $\delta_{k,n} \to 0$ and for large enough n,

$$
\sqrt{\frac{B_k \log B_k}{n}} \leq n^{-1/2}\Delta_{k,n}\delta_{k,n}^{-1}\log(\Delta_{k,n}\delta_{k,n}^{-1}) \leq \delta_{k,n}^3 n^{-1/2}\Delta_{k,n}\delta_{k,n}^{-4}(\Delta_{k,n}\delta_{k,n}^{-1})^{\iota} = o(1).
$$

□

Lemma S9. *The smallest eigenvalue of the $B_k \times B_k$ Gram matrix $\tilde{\Gamma}_{k,n} := \int b_k b_k' \,\mathrm{d}\lambda$ satisfies*

$$
\lambda_{\min}(\tilde{\Gamma}_{k,n}) \geq \upsilon\delta_{k,n} > 0,
$$

for a $\upsilon > 0$.

Proof. Since $b_{k,m}(x)b_{k,s}(x)$ is non-zero only for $|m-s| \leq 3$ and each $b_{k,m}$ is non-zero only on $[\xi_{k,n,m}, \xi_{k,n,m+4})]$ (e.g. (20) p. 91 of de Boor, 2001), $\tilde{\Gamma}_{k,n}$ is a symmetric banded Toeplitz matrix.[S18] Its entries can be computed by direct integration:

$$[\tilde{\Gamma}_{k,n}]_{m,s} = \delta_{k,n} \times \begin{cases} \frac{151}{315} & \text{if } m = s \\ \frac{397}{1680} & \text{if } |m-s| = 1 \\ \frac{1}{42} & \text{if } |m-s| = 2 \\ \frac{1}{5040} & \text{if } |m-s| = 3 \\ 0 & \text{if } |m-s| > 3 \end{cases}.$$

For simplicity of notation let $f_0 := \frac{151}{315}$, $f_1 := f_{-1} := \frac{397}{1680}$, $f_2 := f_{-2} := \frac{1}{42}$ and $f_3 := f_{-3} := \frac{1}{5040}$ and let $f_s := 0$ for $|s| > 3$. Now, let $f(\theta) := \sum_{s=-3}^{3} f_s e^{i(s\theta)}$. Then, $\tilde{\Gamma}_{k,n}/\delta_{k,n}$ is then the matrix generated by f in the sense that $\tilde{\Gamma}_{k,n}/\delta_{k,n} = \mathscr{T}_n(f) := \sum_{s=-\min(B_k-1,3)}^{\min(B_k-1,3)} f_k J_n^s$ where each J_n^s is the $B_k \times B_k$ matrix which is zero everywhere except for the (i,j)-th entries where $i - j = s$, where it has a value of 1.[S19] Since $f \in L_1([-\pi,\pi])$ and is real on $[-\pi,\pi]$ by Theorem 6.1 in Garoni and Serra-Capizzano (2017) we have that $\lambda_{\min}(\tilde{\Gamma}_{k,n}) = \delta_{k,n}\lambda_{\min}(\tilde{\Gamma}_{k,n}/\delta_{k,n}) \geq \delta_{k,n} \inf_{\theta\in[-\pi,\pi]} f(\theta) = \delta_{k,n}\upsilon$, where $\upsilon := \inf_{\theta\in[-\pi,\pi]} f(\theta) > 0$. □

Lemma S10. *Suppose* $\xi \in \mathbb{R}^{N+1}$ *such that* $a = \xi_0 < \xi_1 < \cdots < \xi_N = b$, $h := \max_{i\in[N]} \xi_i - \xi_{i-1}$, *and let* $\mathscr{G}_k(\xi)$ *be the linear space formed by degree* k *splines with knots* ξ. *Then, if* $f \in C^{k-1}[a,b]$ *we have that*

$$\inf_{g\in\mathscr{G}_k(\xi)} \|g - f\|_\infty \leq \frac{(k+1)!}{2^k} h^{k-1} \|f^{(k-1)}\|_\infty = c_k h^{k-1} \|f^{(k-1)}\|_\infty,$$

where c_k *depends only on* k.

Proof. This follows as a special case of Theorem 20.3 in Powell (1981). □

S2 Additional auxillary results

We present a few additional results that explicitly prove some claims made in the main text. First, we show that if two errors $\epsilon_{i,k}$ and $\epsilon_{i,j}$ are Gaussian the efficient information matrix

[S18] As can be easily verified, unlike in the case of linear ($\kappa = 2$) or quadratic splines ($\kappa = 3$), this matrix is *not* diagonally dominant. In the case of $\kappa \in \{2,3\}$ this argument could be completed in a simpler fashion by using the Gershgorin circle theorem.

[S19] See section 6.1 in Garoni and Serra-Capizzano (2017), noting that it is clear that $f \in L_1([-\pi,\pi])$.

becomes singular. Second, we provide an explicit example of a density which satisfies the first part of the Assumption 1 but not the second. Third we prove that if Assumption 1 part 1 holds then a sufficient condition for part 2 is that η_k has tails that decay to zero at a polynomial rate.

Lemma S11. *Consider the LSEM model (3) with $B = 0$ (for ease of exposition only) and Assumption 1 parts 1-3 hold. Define the vector-valued function $Q : \mathbb{R}^K \to \mathbb{R}^{K^2}$ according to*

$$Q(y) = (Q_1(y)', \ldots, Q_K(y)')',$$

where each $Q_k : \mathbb{R}^K \to \mathbb{R}^K$ and the j-th element of Q_k for $j \in [K]$ is given by

$$Q_{k,j}(y) = \begin{cases} \phi_k(A_k y) A_j y & \text{if } k \neq j \\ \tau_{k,1} A_k y + \tau_{k,2} \kappa(A_k y) & \text{if } k = j \end{cases}.$$

Next define the $K^2 \times L$ matrix ζ according to $\zeta = (\mathrm{vec}([D_1(\alpha)A^{-1}]'), \ldots, \mathrm{vec}([D_L(\alpha)A^{-1}]'))$, where in the definition of both Q and ζ we have $A = A(\gamma)$. Equipped with these definitions, we can write the efficient score function as defined in lemma 2 as

$$\tilde{\ell}_\theta(y) = \zeta' Q(y). \tag{S17}$$

Then,

(i) $\mathbb{E}_\theta QQ'$ is non-singular if and only if for each pair (k, j) with $k \neq j$ and each $k, j \in [K]$ we have that $[\mathbb{E}_\theta \phi_k^2(A_k Y)][\mathbb{E}_\theta \phi_j^2(A_j Y)] \neq 1$.

(ii) $\tilde{I}_\theta$ is non-singular if $\mathrm{rank}(\zeta) = L$ and $\mathbb{E}_\theta QQ'$ is non-singular.

(iii) If $\mathrm{rank}(\zeta) < L$ then $\tilde{I}_\theta$ is singular.

(iv) If $L = K^2$ and $\mathbb{E}_\theta QQ'$ is singular then $\tilde{I}_\theta$ is singular.

(v) If $\mathbb{E}_\theta QQ'$ is singular, $\tilde{I}_\theta$ may be singular when $\mathrm{rank}(\zeta) = L < K^2$.

In particular, if both ϵ_k and ϵ_j ($k \neq j$) have a Gaussian distribution and $L = K^2$, $\tilde{I}_\theta$ is singular.

Proof. For (i), let j, k, m, i all be in $[K]$. We will consider the entries of the matrix $\mathbb{E}_\theta QQ'$, which are of the form $\langle Q_{k,j},\ Q_{m,i} \rangle_{P_\theta}$. In particular, the s, t-th element of the matrix is given by the form $\langle Q_{k,j},\ Q_{m,i} \rangle_{P_\theta}$ where $(k-1)K + j = s$ and $(m-1)K + i = t$. If $k = j = m = i$ we have $s = t$ and $\langle Q_{k,j},\ Q_{m,i} \rangle_{P_\theta} = \mathbb{E}_\theta[\tau_{k,1} A_k Y + \tau_{k,2} \kappa(A_k Y)]^2$. The other diagonal entries occur when $k = m \neq j = i$, and have the form $\langle Q_{k,j},\ Q_{m,i}\rangle_{P_\theta} = \mathbb{E}_\theta[\phi_k^2(A_k Y)]$. Inspection of the other possible cases reveals that the only other case with non-zero entries is $k = i \neq m = j$ which has value $\langle Q_{k,j},\ Q_{m,i}\rangle_{P_\theta} = \mathbb{E}_\theta[\phi_k(A_k Y)A_k Y]\mathbb{E}_\theta[\phi_k(A_m Y)A_m Y] = 1$ by assumption 1.

Therefore for any $k, j \in [K]$, column $(k-1)K+j$ has non-zero entries in row $(k-1)K+j$ only if $k = j$ and otherwise in rows $(k-1)K+j$ and $(j-1)K+k$, with values $\mathbb{E}_\theta\phi_k^2(A_k Y)$ and 1 respectively. There are therefore no columns that can be linearly related to column $(k-1)K+j$ if $k = j$. If $k \neq j$, then column $(k-1)K+j$ has zeros everywhere except row $(k-1)K+j$ where it has $\mathbb{E}_\theta\phi_k^2(A_k Y)$ and row $(j-1)+k$ where it has 1. Column $(j-1)K+k$ has zeros everywhere except row $(j-1)K+k$ where it has $\mathbb{E}_\theta\phi_j^2(A_j Y)$ and row $(k-1)K+j$ where it has 1. Since no other columns have entries in these rows, it follows that column $(k-1)K+j$ is linearly independent of all the other columns if and only if it is linearly independent of column $(j-1)K+k$, which occurs if and only if $[\mathbb{E}_\theta\phi_k^2(A_k Y)][\mathbb{E}_\theta\phi_j^2(A_j Y)] \neq 1$.

For (ii), suppose that $\operatorname{rank}(\zeta) = L$ and $\mathbb{E}_\theta QQ'$ is non-singular. Then there is a (unique) positive definite $[\mathbb{E}_\theta QQ']^{1/2}$ and we have $\tilde{I}_\theta = \left([\mathbb{E}_\theta QQ']^{1/2}\zeta\right)'\left([\mathbb{E}_\theta QQ']^{1/2}\zeta\right)$ which has full rank, since $\left([\mathbb{E}_\theta QQ']^{1/2}\zeta\right)$ has full column rank.

For the remaining parts note first that

$$\tilde{I}_\theta = \mathbb{E}_\theta\tilde{\ell}_\theta\tilde{\ell}_\theta' = \zeta'\left[\mathbb{E}_\theta QQ'\right]\zeta,$$

and so $\operatorname{rank}(\tilde{I}_\theta) \leq \min\{\operatorname{rank}(\zeta'\mathbb{E}_\theta QQ'), \operatorname{rank}(\zeta)\}$. Hence if $\operatorname{rank}(\zeta) < L$, $\operatorname{rank}(\tilde{I}_\theta) < L$ implying (iii).

For (iv), suppose that $\operatorname{rank}(\mathbb{E}_\theta QQ') < K^2 = L$. Then, there is a non-zero $x \in \mathbb{R}^L$ such that $\mathbb{E}_\theta QQ'x = 0$ and hence $\zeta'\mathbb{E}_\theta QQ'x = 0$. Hence $\dim(N(\zeta'\mathbb{E}_\theta QQ')) \geq 1$. It follows that $\operatorname{rank}(\zeta'\mathbb{E}_\theta QQ') \leq L-1 < L$ and hence $\operatorname{rank}(\tilde{I}_\theta) \leq \min\{\operatorname{rank}(\zeta'\mathbb{E}_\theta QQ'), \operatorname{rank}(\zeta)\} < L$.

For (v) suppose that $K = 2$, ϵ_1 and ϵ_2 are both Gaussian and $A(\gamma) = \left[\begin{smallmatrix}\cos(\gamma) & -\sin(\gamma)\\ \sin(\gamma) & \cos(\gamma)\end{smallmatrix}\right]$. We have for $l \in \{1, 2\}$, $\phi_l(z) = -z$, hence $\phi_l^2(z) = z^2$ and so $\mathbb{E}_\theta\phi_l^2(\epsilon_l) = \mathbb{E}_\theta\phi_l^2(A_l Y) = 1$. $D_1(\gamma) = \left[\begin{smallmatrix}-\sin(\gamma) & -\cos(\gamma)\\ \cos(\gamma) & -\sin(\gamma)\end{smallmatrix}\right]$ and hence

$$D_1(\gamma)A(\gamma)^{-1} = D_1(\gamma)A(\gamma)' = \begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix},$$

which implies $\zeta = (0, -1, 1, 0)'$ and hence $\operatorname{rank}(\zeta) = 1 = L < K^2 = 4$. Explicit calculation reveals that

$$E_\theta QQ' = \begin{bmatrix}8/9 & 0 & 0 & 0\\ 0 & 1 & 1 & 0\\ 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 8/9\end{bmatrix},$$

which is clearly singular with rank 3. We have

$$\tilde{I}_\theta = \zeta' \left[\mathbb{E}_\theta QQ'\right] \zeta = \zeta' \begin{bmatrix} 8/9 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 8/9 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} = \zeta' \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 0.$$

For the last part, suppose that $k \neq j$ and ϵ_k and ϵ_j are both Gaussian. Since both have zero mean and unit variance, we have for $l \in \{k, j\}$, $\phi_l(z) = -z$, hence $\phi_l^2(z) = z^2$ and so $\mathbb{E}_\theta \phi_l^2(\epsilon_l) = \mathbb{E}_\theta \phi_l^2(A_l Y) = 1$. $E_\theta QQ'$ is singular by (i) and hence by (iv) $\tilde{I}_\theta$ is singular. □

Example S1 (Necessity of part 2 of assumption 1)**.** *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 $\phi_k(z) = -1$. Explicit calculation reveals that part 1 of assumption 1 is satisfied. However, $\mathbb{E}\phi_k(z) = -1 \neq 0$ as would be required by part 2 of assumption 1.*

Note also that this example does not satisfy the requirements of lemma S12: 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 $r = 0$.

Lemma S12. *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 1 of assumption 1 holds, part 2 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,

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

Our hypothesis ensures that $z^r \eta_k(z)\big|_{a_k}^{b_k} = 0$. 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. □

S3 Additional simulation results

In this section we provide a number of additional simulation results.

S3.1 Additional power results for the baseline model

Figure 4 in the main text compared the power of different tests for the baseline model $Y_i = R'\epsilon_i$ for the case where $n = 1000$. Here we show the results for $n = 200$ and $n = 500$. Specifically, Figures S1 and S2 show the results.

Overall, the patterns that we find are similar as in the main text. One thing that stands out is that the $\mathrm{S}^{\mathrm{gmm}}$ test is not correctly sized for these smaller sample sizes, essentially confirming the results in Table 3. It is possible that a more careful selection of the relevant higher order moments will improve this finding.

Besides this our two main findings from the main text hold. First, the standard LM test is the preferred approach whenever the true density is known, but the semi-parametric score test comes close in terms of power. Second, for all other densities the semi-parametric score test shows the highest power.

S3.2 Additional power results for the LSEM

Figure 5 in the main text compared the power of different tests for the LSEM model for the case where $n = 1000$. Here we show the results for $n = 200$ and $n = 500$. Specifically, Figures S3 and S4 show the results.

We find that for $n = 200$ the power of tests is generally quite low, indicating that for small sample sizes little can be learned by exploiting deviations from the Gaussian density. This holds most notably for the Student's t densities, the skewed unimodal density and the bimodal density. Intuitively, given a small sample these densities are hard to distinguish from the normal density and little can be learned about the parameter α. A reassuring finding is that the size of the test remains well controlled. These findings persist, to a lesser extent, when we increase to $n = 500$.

Overall, the implementing the test with one-step efficient estimates leads to higher power, but the size of the test is controlled less well. Therefore we recommend using OLS estimates for β when the sample size is small.

References

Bickel, P. J., C. A. J. Klaasen, Y. Ritov, and J. A. Wellner. 1998. *Efficient and Adaptive Estimation for Semiparametric Models.* New York, NY, USA:Springer.

Billingsley, P. 1999. *Convergence of Probability Measures.* Wiley.

Chen, A., and P. J. Bickel. 2006. "Efficient Independent Component Analysis." *Annals of Statistics*, 34(6).

Conway, J. B. 1985. *A course in functional analysis.* New York, NY, USA:Springer.

de Boor, C. 2001. *A Practical Guide to Splines.* New York, NY, USA:Springer.

Folland, G. B. 1999. *Real Analysis: Modern Techniques and Their Applications.* New York, NY, USA:John Wiley & Sons, Inc.

Garoni, C., and S. Serra-Capizzano. 2017. *Generalized Locally Toeplitz Sequences: Theory and Applications.* Vol. 1, Cham, Switzerland:Springer.

Horn, R. A., and C. R. Johnson. 2013. *Matrix Analysis.* . 2 ed., Cambridge University Press.

Powell, M. J. D. 1981. *Approximation Theory and Methods.* Cambridge, UK:Cambridge University Press.

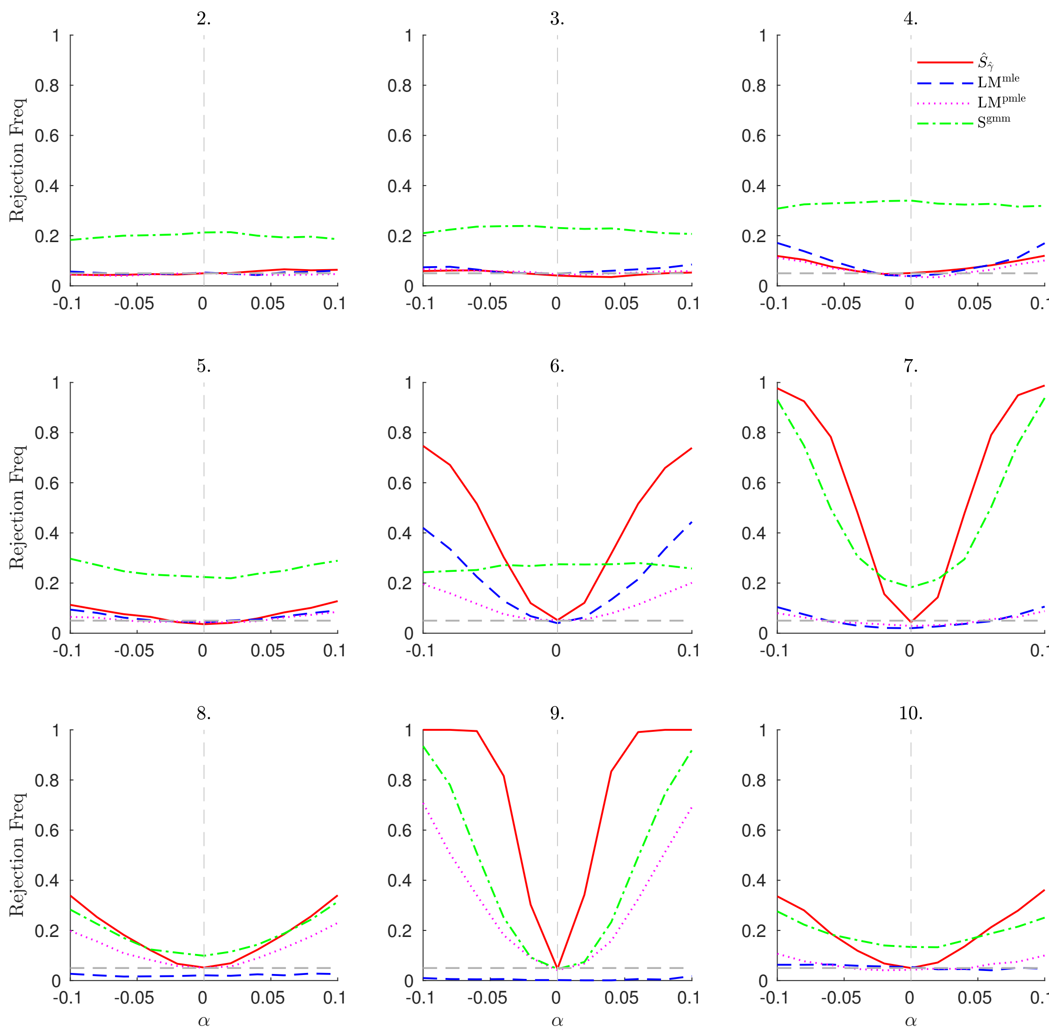
Royden, H.L., and P.M. Fitzpatrick. 2010. *Real Analysis.* Pearson Prentice Hall.

Rudin, W. 1987. *Real & Complex Analysis.* McGraw Hill.

van der Vaart, A. W. 1998. *Asymptotic Statistics.* . 1st ed., New York, NY, USA:Cambridge University Press.

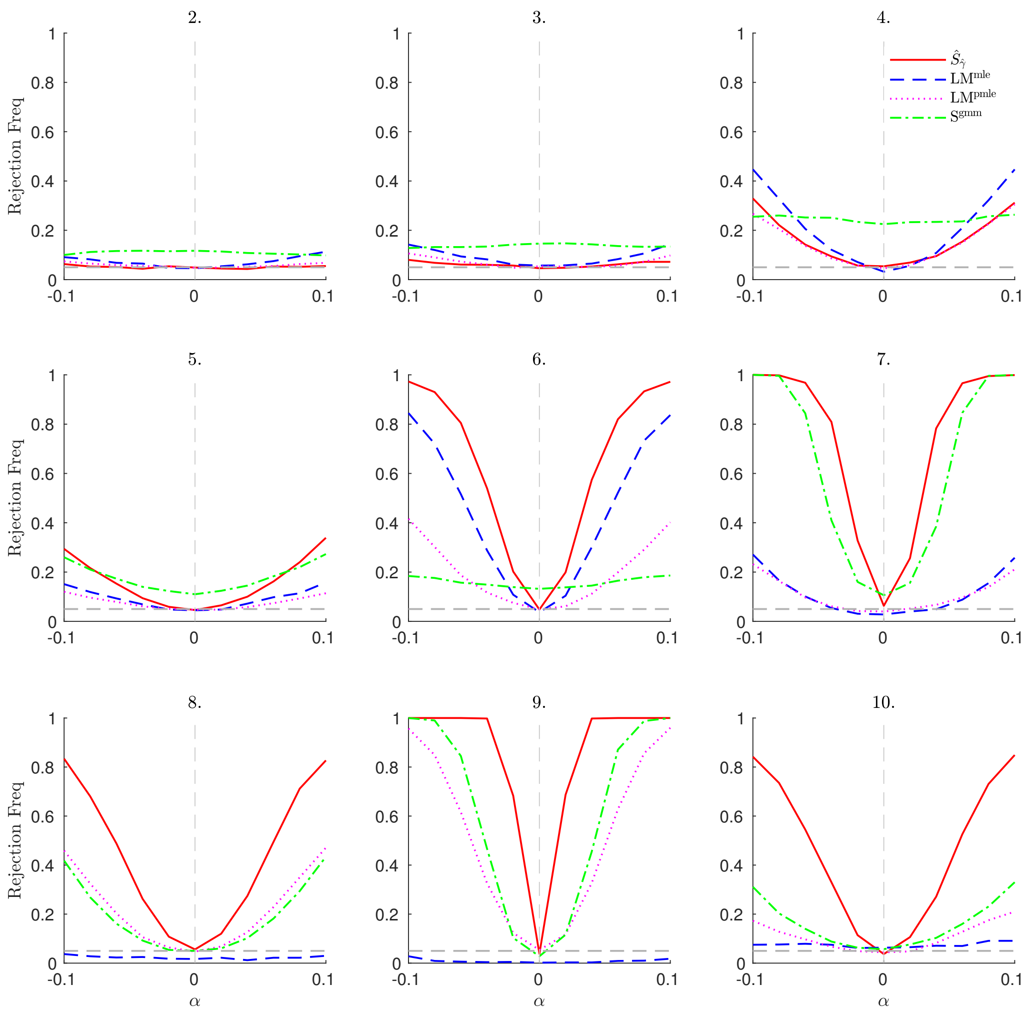
van der Vaart, A. W. 2002. "Semiparametric Statistics." In *Lectures on Probability Theory and Statistics: Ecole d'Eté de Probabilités de Saint-Flour XXIX - 1999.* , ed. P. Bernard. Berlin, Germany:Springer.

Figure S1: Power Comparison Baseline model $n = 200$



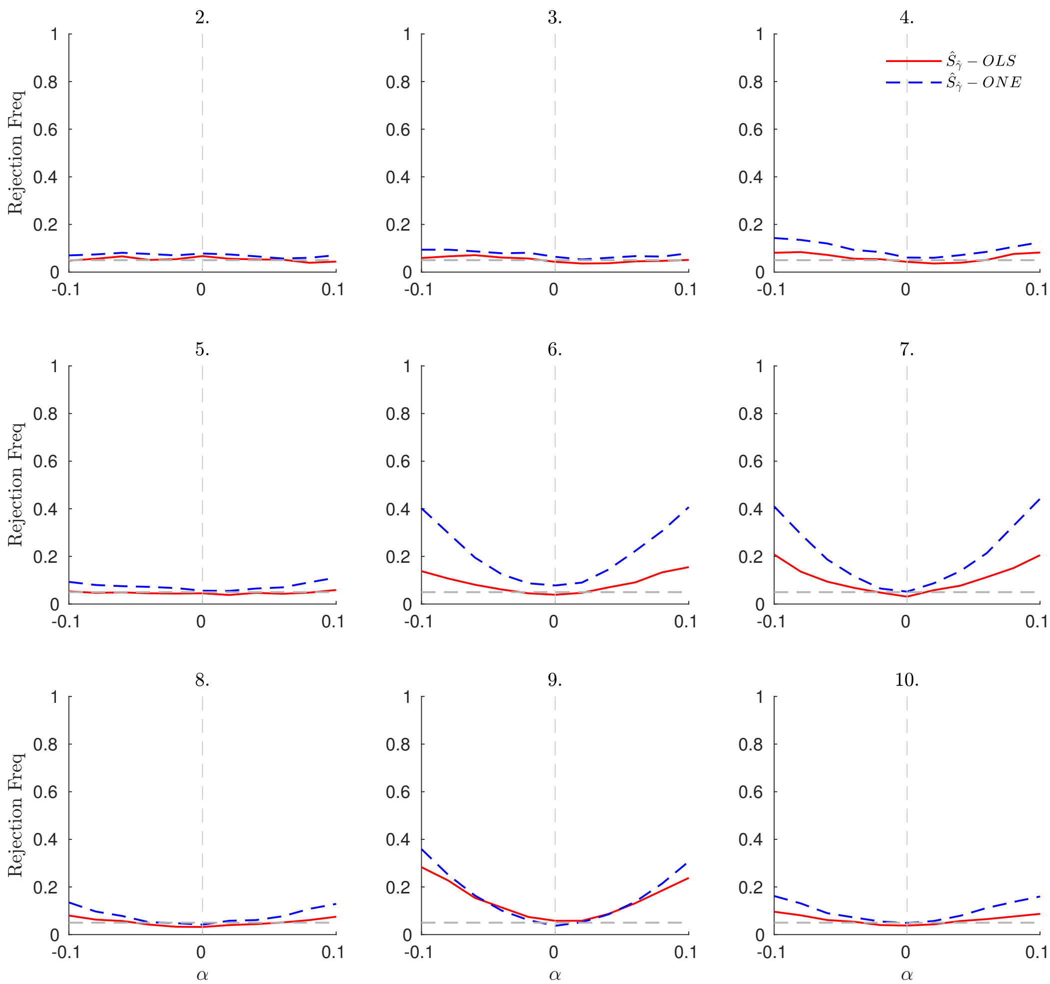
Notes: Empirical power curves for the baseline model with $k = 2$ and $n = 200$. Each plot corresponds to the choice for densities ϵ_k, for $k \geq 2$, where the numbers correspond to the different densities listed in Figure 3. The solid red line corresponds to $S_{\hat{\gamma}}$, the dashed blue line to LM$^{\text{mle}}$, the dotted pink line to LM$^{\text{pmle}}$ and the dot-dashed green line to S$^{\text{gmm}}$.

Figure S2: Power Comparison Baseline model $n = 500$



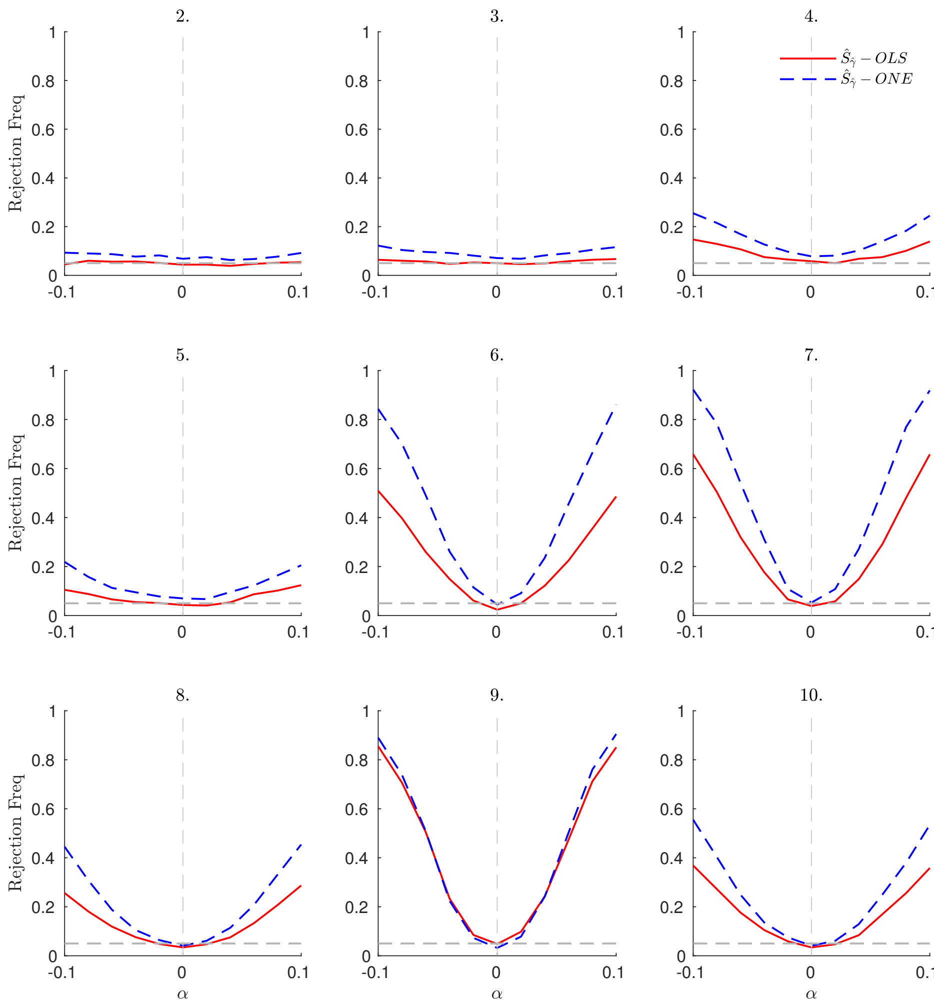
Notes: Empirical power curves for the baseline model with $k = 2$ and $n = 500$. Each plot corresponds to the choice for densities ϵ_k, for $k \geq 2$, where the numbers correspond to the different densities listed in Figure 3. The solid red line corresponds to $S_{\hat{\gamma}}$, the dashed blue line to LM^{mle}, the dotted pink line to LM^{pmle} and the dot-dashed green line to S^{gmm}.

Figure S3: Power LSEM $n = 200$



Notes: Empirical power curves for the LSEM model with $k = 2$, $d = 2$ and $n = 200$. Each plot corresponds to the choice for densities $\epsilon_{i,k}$, for $k \geq 2$, where the numbers correspond to the different densities shown in Figure 3. The solid red line corresponds to the empirical rejection frequency of the $\hat{S}_{\hat{\gamma}}$ test where $\hat{\gamma} = (\alpha_0, \hat{\beta})$, with $\hat{\beta}$ the OLS estimator. The dashed blue line corresponds to the empirical rejection frequency of the $\hat{S}_{\hat{\gamma}}$ test where $\hat{\gamma} = (\alpha_0, \hat{\beta})$, with $\hat{\beta}$ the one-step efficient MLE estimator.

Figure S4: Power LSEM $n = 500$



Notes: Empirical power curves for the LSEM model with $k = 2$, $d = 2$ and $n = 500$. Each plot corresponds to the choice for densities $\epsilon_{i,k}$, for $k \geq 2$, where the numbers correspond to the different densities shown in Figure 3. The solid red line corresponds to the empirical rejection frequency of the $\hat{S}_{\hat{\gamma}}$ test where $\hat{\gamma} = (\alpha_0, \hat{\beta})$, with $\hat{\beta}$ the OLS estimator. The dashed blue line corresponds to the empirical rejection frequency of the $\hat{S}_{\hat{\gamma}}$ test where $\hat{\gamma} = (\alpha_0, \hat{\beta})$, with $\hat{\beta}$ the one-step efficient MLE estimator.