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

Adam Lee[1] *and* *Geert Mesters*[2]

[1]BI Norwegian Business School

[2]Universitat Pompeu Fabra, Barcelona School of Economics and CREI

December 17, 2023

Abstract

In this supplementary material we provide the following additional results.

*Email: adam.lee@bi.no, geert.mesters@upf.edu.

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 Main proofs

In this section we provide our main proofs. Regarding notation: $x \coloneqq y$ means that x is defined to be y. The Lebesgue measure on $\mathbb{R}^K$ is denoted by λ_K with $\lambda \coloneqq \lambda_1$ and the standard basis vectors in $\mathbb{R}^K$ are $e_1, \ldots, e_K$. We will make use of the empirical process notation: $Pf \coloneqq \int f \, \mathrm{d}P$, $\mathbb{P}_n f \coloneqq \frac{1}{n}\sum_{i=1}^n f(Y_i)$ and $\mathbb{G}_n f \coloneqq \sqrt{n}(\mathbb{P}_n - P)f$. For any two sequence of probability measures $(Q_n)_{n\in\mathbb{N}}$ and $(P_n)_{n\in\mathbb{N}}$ (where Q_n and P_n are defined on a common measurable space for each $n \in \mathbb{N}$), $Q_n \triangleleft P_n$ indicates that $(Q_n)_{n\in\mathbb{N}}$ is contiguous with respect to $(P_n)_{n\in\mathbb{N}}$. $Q_n \triangleleft\triangleright P_n$ indicates that both $Q_n \triangleleft P_n$ and $P_n \triangleleft Q_n$ hold, see van der Vaart (1998, Section 6.2) for formal definitions. $X \perp\!\!\!\perp Y$ indicates that random vectors X and Y are independent; $X \simeq Y$ indicates that they have the same distribution. $a \lesssim b$ means that a is bounded above by Cb for some constant $C \in (0, \infty)$; the constant C may change from line to line. $\operatorname{cl} X$ means the closure of X. vec^{-1} is the inverse vec operator, i.e. if $b = \operatorname{vec}(B)$ then $B = \operatorname{vec}^{-1}(b)$. If S is a subset of an inner product space $(V, \langle\cdot,\ \cdot\rangle)$, $S^\perp$ is its orthogonal complement, i.e. $S^\perp = \{x \in V : \langle x,\ s\rangle = 0 \text{ for all } s \in S\}$. If $S \subset V$ is complete (hence a Hilbert space) the orthogonal projection of $x \in V$ onto S is $\Pi(x|S)$.

In this document we use notation which explicitly records the dependency of objects on $\theta = (\gamma, \eta)$, including in cases where this was left implicit in the main text to prevent the notation from becoming overly cumbersome. For instance, instead of $A_{k\bullet}$, in the appendices we write $A(\alpha, \sigma)_{k\bullet}$ or $e_k' A(\alpha, \sigma)$.

S1.1 Score functions and local asymptotic normality

We first review a number of definitions and establish the semiparametric framework underlying the robust testing approach outlined in this paper.

Formally, the considered model (3) is the collection

$$\mathcal{P}_\Theta = \{P_\theta : \theta \in \Theta\} , \tag{S1}$$

where each P_θ is the law of the data $W_i = (Y_i, \tilde{X}_i)$ which lies in $\mathcal{W} \subset \mathbb{R}^{K+d-1}$. The parameter space Θ has the form $\Theta = \mathcal{A} \times \mathcal{B} \times \mathcal{H}$, where $\mathcal{A} \subset \mathbb{R}^{L_\alpha}$, $\mathcal{B} \subset \mathbb{R}^{L_\beta}$. $\mathcal{H}$ has the form $\mathscr{Z} \times \prod_{k=1}^K \mathscr{H}$, where $\mathscr{Z}$ is the space of density functions η_0 and $\mathscr{H}$ is the space of density functions η_k such that if $\tilde{X} \sim \eta_0$ and $\epsilon_k \sim \eta_k$ then Assumption 2 parts 1, 3, 4 and 5 hold.[S1]

[S1] Part 2 of Assumption 2 serves to simplify the form of the effective score function derived in Lemma S3

We write a typical element of Θ as $\theta = (\alpha, \beta, \eta)$, where $\beta = (b', \sigma')'$ and it is understood that $\alpha \in \mathcal{A}$, $\beta \in \mathcal{B}$ and $\eta \in \mathcal{H}$. In what follows we will let $V_{\theta,i} := Y_i - BX_i$ be the reduced form error so that $A(\alpha, \sigma)V_{\theta,i} = \epsilon_i$. Each P_θ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^{K+d-1}$, with (Lebesgue) density given by

$$p_\theta(W_i) = |\det A(\alpha, \sigma)| \prod_{k=1}^{K} \eta_k(e_k' A(\alpha, \sigma)V_{\theta,i}) \times \eta_0(\tilde{X}_i) \ , \tag{S2}$$

and hence log-density

$$\ell_\theta(W_i) = \log|\det A(\alpha, \sigma)| + \sum_{k=1}^{K} \log \eta_k(e_k' A(\alpha, \sigma)V_{\theta,i}) + \log \eta_0(\tilde{X}_i) \ . \tag{S3}$$

We now define the scores of model (S1) following the definition in van der Vaart (2002).

Definition S1 (Cf. Definition 1.6 in van der Vaart, 2002)**.** *A* differentiable path *is a map $t \mapsto P_t$ from a neighborhood of $0 \in [0, \infty)$ to $\mathcal{P}_\Theta$ such that for some measurable function $s : \mathcal{W} \to \mathbb{R}$,*

$$\int \left[\frac{\sqrt{p_t} - \sqrt{p}}{t} - \frac{1}{2} s \sqrt{p} \right]^2 \mathrm{d}\mu \to 0 \ , \tag{S4}$$

as $t \to 0$, where p_t and p respectively denote the densities of P_t and P relative to a σ-finite measure μ. The map $t \to \sqrt{p_t}$ is the root density path and s is the ***score function*** *of the submodel $\{P_t : t \geq 0\}$ at $t = 0$.*

In words, a differentiable path is a one-dimensional parametric submodel $\{P_t : t \geq 0\}$ that is differentiable in quadratic mean at $t = 0$ with score function s. If we let $t \mapsto P_t$ range over a collection of submodels, indexed by $\mathcal{V}$, we will obtain a collection of score functions, say s_j for $j \in \mathcal{V}$.

The differentiable paths we consider have the following form. Let P_t be the measure corresponding to the density with form as in (S2) evaluated at $\theta_t := (\gamma + tg, \eta_t)$ where the k-th coordinate of η_t is $\eta_{k,t}^{h_k} := \eta_k(1 + th_k)$ $(k = 0, \ldots, K)$, and $(g, h) \in \mathbb{R}^L \times H$, where $H = \prod_{k=0}^{K} H_k$ and each H_k is defined following (6).

That such $t \mapsto P_t$ paths are indeed differentiable paths as in Definition S1 is established in the following lemma.

Lemma S1. *Suppose Assumptions 1 and 2 hold and that (α, β) is an interior point of $\mathcal{A} \times \mathcal{B}$. For each $(g, h) \in \mathbb{R}^L \times H := \mathcal{V}$, the map $t \mapsto P_{\theta_t}$ is a differentiable path, with score function*

and is not necessary to set up the model.

$g'\dot{\ell}_\theta + \tilde{h}_0 + \sum_{k=1}^K \tilde{h}_k$, where $\dot{\ell}_\theta \coloneqq \nabla_\gamma \log p_\theta$, $\tilde{h}_0(W) \coloneqq h_0(\tilde{X})$ and $\tilde{h}_k(W) \coloneqq h_k(e_k' A(\alpha,\sigma)V_\theta)$. $\dot{\ell}_\theta$ has the form $\dot{\ell}_\theta = (\dot{\ell}_{\theta,\alpha}', \dot{\ell}_{\theta,b}', \dot{\ell}_{\theta,\sigma}')'$, with

$$\begin{aligned}\dot{\ell}_{\theta,\alpha,l}(W) \coloneqq& \sum_{k=1}^K \sum_{j=1,j\neq k}^K \zeta_{l,k,j}^\alpha(\alpha,\sigma)\phi_k(e_k'A(\alpha,\sigma)V_\theta)e_j'A(\alpha,\sigma)V_\theta \\ &+ \sum_{k=1}^K \zeta_{l,k,k}^\alpha(\alpha,\sigma)[\phi_k(e_k'A(\alpha,\sigma)V_\theta)e_k'A(\alpha,\sigma)V_\theta + 1] \\ \dot{\ell}_{\theta,\sigma,l}(W) \coloneqq& \sum_{k=1}^K \sum_{j=1,j\neq k}^K \zeta_{l,k,j}^\sigma(\alpha,\sigma)\phi_k(e_k'A(\alpha,\sigma)V_\theta)e_j'A(\alpha,\sigma)V_\theta \\ &+ \sum_{k=1}^K \zeta_{l,k,k}^\sigma(\alpha,\sigma)[\phi_k(e_k'A(\alpha,\sigma)V_\theta)e_k'A(\alpha,\sigma)V_\theta + 1],\end{aligned}$$

and

$$\dot{\ell}_{\theta,b}(W)' \coloneqq -\sum_{k=1}^K \phi_k\left(e_k'A(\alpha,\sigma)V_\theta\right) e_k'A(\alpha,\sigma)[X' \otimes I_K].$$

Proof. Let $g = (a, \varrho, s) \in \mathbb{R}^{L_\alpha} \times \mathbb{R}^{L_b} \times \mathbb{R}^{L_\sigma}$. The log density of W under θ_t is then

$$\begin{aligned}\ell_{\theta_t}(W) &= \log p_{\theta_t}(W) \\ &= \log \eta_0(\tilde{X}) + \log(1 + th_0(\tilde{X})) + \log|\det(A(\alpha + ta, \sigma + ts))| \\ &\quad + \sum_{k=1}^K \log \eta_k\left(e_k'A(\alpha + ta, \sigma + ts)(Y - BX - t\operatorname{vec}^{-1}(\varrho)X)\right) \\ &\quad + \sum_{k=1}^K \log\left(1 + th_k\left(e_k'A(\alpha + ta, \sigma + ts)(Y - BX - t\operatorname{vec}^{-1}(\varrho)X)\right)\right),\end{aligned}$$

By Lemma S10, $t \mapsto \sqrt{p_{\theta_t}}$ is continuously differentiable (pointwise) in a neighbourhood $\mathcal{V}$ of 0. Moreover, if we define $q_t(W) \coloneqq \frac{\partial \log p_{\theta_x}(W)}{\partial x}\big|_{x=t}$ and $Q_t \coloneqq P_{\theta_t} q_t(W)^2$, Q_t is finite and continuous in a neighbourhood of 0 by the uniformly integrability of $\{q_t(W)^2 : t \in \mathcal{V}\}$ along with the pointwise continuity of $t \mapsto q_t(W)$, both of which follow from Lemma S10.

Hence, by Lemma 1.8 in van der Vaart (2002), $t \mapsto P_{\theta_t}$ is a differentiable path with score function given by the derivative of $\ell_{\theta_t}(W)$ at $t=0$, which is:

$$\begin{aligned}
&\sum_{k=1}^{K}\phi_k\left(e_k'A(\alpha,\sigma)V_\theta\right)e_k'\sum_{l=1}^{L_\alpha}a_l D_{\alpha,l}(\alpha,\sigma)V_\theta+\sum_{l=1}^{L_\alpha}a_l\operatorname{tr}(A(\alpha,\sigma)^{-1}D_{\alpha,l}(\alpha,\sigma))\\
&+\sum_{k=1}^{K}\phi_k\left(e_k'A(\alpha,\sigma)V_\theta\right)e_k'\sum_{l=1}^{L_\sigma}s_l D_{\sigma,l}(\alpha,\sigma)V_\theta+\sum_{l=1}^{L_\sigma}s_l\operatorname{tr}(A(\alpha,\sigma)^{-1}D_{\sigma,l}(\alpha,\sigma))\\
&-\sum_{k=1}^{K}\phi_k\left(e_k'A(\alpha,\sigma)V_\theta\right)e_k'A(\alpha,\sigma)[X'\otimes I_K]\varrho+h_0(\tilde{X})+\sum_{k=1}^{K}h_k\left(e_k'A(\alpha,\sigma)V_\theta\right),
\end{aligned}\tag{S5}$$

with $D_{x,l}(\alpha,\sigma)=\nabla_{x_l}A(\alpha,\sigma)$ for any $x\in\{\alpha,\sigma\}$ and any l in $\{1,\ldots,L_\alpha\}$ or $\{1,\ldots,L_\sigma\}$ as appropriate. We can re-write the two expressions involving the trace as follows: for any $x\in\{\alpha,\sigma\}$ and appropriate index l we have

$$\begin{aligned}
&\sum_{k=1}^{K}\phi_k(e_k'A(\alpha,\sigma)V_\theta)e_k'D_{x,l}(\alpha,\sigma)V_\theta+\operatorname{tr}(A(\alpha,\sigma)^{-1}D_{x,l}(\alpha,\sigma))\\
&\quad=\sum_{k=1}^{K}\phi_k(e_k'A(\alpha,\sigma)V_\theta)e_k'D_{x,l}(\alpha,\sigma)A(\alpha,\sigma)^{-1}\epsilon+\operatorname{tr}(D_{x,l}(\alpha,\sigma)A(\alpha,\sigma)^{-1})\\
&\quad=\sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K}\zeta_{l,k,j}^{x}(\alpha,\sigma)\phi_k(e_k'A(\alpha,\sigma)V_\theta)e_j'A(\alpha,\sigma)V_\theta\\
&\qquad+\sum_{k=1}^{K}\zeta_{l,k,k}^{x}(\alpha,\sigma)[\phi_k(e_k'A(\alpha,\sigma)V_\theta)e_k'A(\alpha,\sigma)V_\theta+1],
\end{aligned}$$

for $\zeta_{l,k,j}^{x}(\alpha,\sigma):=e_k'D_{x,l}(\alpha,\sigma)A(\alpha,\sigma)^{-1}e_j$. We may therefore write the derivative (S5) as $a'\dot{\ell}_{\theta,\alpha}+\varrho'\dot{\ell}_{\theta,b}+s'\dot{\ell}_{\theta,\sigma}+\dot{\ell}_{\theta,\eta,h}$ where

$$\dot{\ell}_{\theta,\eta,h}(W):=h_0(\tilde{X})+\sum_{k=1}^{K}h_k\left(e_k'A(\alpha,\sigma)V_\theta\right)=\tilde{h}_0(W)+\sum_{k=1}^{K}\tilde{h}_k(W).\tag{S6}$$

An elementary calculation reveals that $g'\dot{\ell}_\theta=a'\dot{\ell}_{\theta,\alpha}+\varrho'\dot{\ell}_{\theta,b}+s'\dot{\ell}_{\theta,\sigma}$.

□

As shown in Lemma S1, the score functions corresponding to η are $\dot{\ell}_{\theta,\eta,h}$ as defined in (S6), for h ranging over H. These are collected in the set $\mathcal{T}$, as defined in equation (6).

The next Lemma establishes a uniform local asymptotic normality result for (a localised version of) our model. For this we need to specify the notion of convergence on $\mathcal{V}:=\mathbb{R}^L\times H$.

We equip the product space $\mathcal{V}$ with the norm[S2]

$$\|(g,h)\| \coloneqq \sqrt{\|g\|^2 + \|\tilde{h}_0\|^2_{L_2(P_\theta)} + \sum_{k=1}^{K} \|\tilde{h}_k\|_{L_2(P_\theta)}\}^2} \ .$$

Lemma S2. *Suppose that Assumptions 1 and 2 hold and that (α, β) is an interior point of $\mathcal{A} \times \mathcal{B}$. For $(g,h) \in \mathcal{V}$ let*

$$\theta_n(g,h) \coloneqq \theta + n^{-1/2}(g, \eta_0 h_0, \ldots, \eta_K h_K).$$

For any convergent sequence $(g_n, h_n) \to (g,h)$ (all in $\mathcal{V}$), define R_n as

$$R_n \coloneqq \log \prod_{i=1}^{n} \frac{p_{\theta_n(g_n,h_n)}(W_i)}{p_\theta(W_i)} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g'\dot{\ell}_\theta(W_i) + \sum_{k=0}^{K} \tilde{h}_k(W_i) \right] + \frac{1}{2}\mathbb{E}\left[g'\dot{\ell}_\theta(W_i) + \sum_{k=0}^{K} \tilde{h}_k(W_i) \right]^2 .$$

Then,

1. *$R_n \xrightarrow{P_\theta} 0$,*

2. *Under P_θ,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g'\dot{\ell}_\theta(W_i) + \sum_{k=0}^{K} \tilde{h}_k(W_i) \right] \rightsquigarrow \mathcal{N}\left(0, \mathbb{E}\left[g'\dot{\ell}_\theta(W_i) + \sum_{k=0}^{K} \tilde{h}_k(W_i) \right]^2 \right),$$

3. *The (product) measures $P^n_{\theta_n}$ and P^n_θ are mutually contiguous.*

Proof. Part 2 follows from Lemma S1 in combination with the Lindenberg-Lévy central limit theorem and Lemma 1.7 of van der Vaart (2002). For Part 1, we first note that in the special case where $(g_n, h_n) = (g, h)$ for all $n \in \mathbb{N}$, $R_n \xrightarrow{P_\theta} 0$ follows by combining Lemma S1 with Lemma 1.9 in van der Vaart (2002). For the general case, note that by Lemma S11 (i) the functions $(g,h) \mapsto \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g'\dot{\ell}_\theta + \sum_{k=0}^{K} \tilde{h}_k \right]$ (i.e. indexed by n) are equicontinuous on compacts in $L_2(P_\theta)$ and (ii) the functions $(g,h) \mapsto P^n_{\theta_n(g,h)}$ (i.e. indexed by n) are equicontinuous on compacts in the total variation metric. By (i), the i.i.d. assumption and

[S2] Each $\tilde{h}_k$ is as defined in the statement of Lemma S1.

Lemma 1.7 in van der Vaart (2002)

$$
\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}\left[(g-g_n)'\dot{\ell}_\theta(W_i) + \sum_{k=0}^{K}\left(\tilde{h}_k(W_i) - \tilde{h}_{n,k}(W_i)\right)\right]^2 \\
&\quad = \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[(g-g_n)'\dot{\ell}_\theta(W_i) + \sum_{k=0}^{K}\left(\tilde{h}_k(W_i) - \tilde{h}_{n,k}(W_i)\right)\right]\right]^2 \qquad \text{(S7)} \\
&\quad = 0.
\end{aligned}
$$

By (ii) one has $\lim_{n\to\infty} d_{TV}(P^n_{\theta_n(g_n,h_n)}, P^n_{\theta_n(g,h)}) = 0$ where d_{TV} indicates the total variation metric. This implies (cf. Theorem 80.13 in Strasser (1985))

$$
\log \prod_{i=1}^{n} \frac{p_{\theta_n(g_n,h_n)}(W_i)}{p_\theta(W_i)} - \log \prod_{i=1}^{n} \frac{p_{\theta_n(g,h)}(W_i)}{p_\theta(W_i)} = o_{P_\theta^n}(1).
$$

Combine the preceding two displays with the previously demonstrated result for the special case where $(g_n, h_n) = (g, h)$ for all $n \in \mathbb{N}$ to conclude. Part 3 then follows by combining Parts 1 and 2 with Example 6.5 in van der Vaart (1998). □

S1.2 Orthogonality and the effective score

We now derive the effective score for α, i.e. $\tilde{\kappa}_\theta$. By definition, this is the orthogonal projection of the score function for the parameter of interest, i.e. $\dot{\ell}_{\theta,\alpha}$, on the orthocomplement (in $L_2(P_\theta)$) of the space spanned by the score functions for all nuisance parameters, i.e. $\dot{\ell}_{\theta,\sigma}$, $\dot{\ell}_{\theta,b}$ and $\dot{\ell}_{\theta,\eta,h}$.[S3] That is, collecting the scores for the nuisance parameters as

$$
\mathcal{S} := \mathrm{Span}(\dot{\ell}_{\theta,b}) + \mathrm{Span}(\dot{\ell}_{\theta,\sigma}) + \mathcal{T} \subset L_2(P_\theta),
$$

where $\mathcal{T}$ is defined in (6) and collects the scores corresponding to η, one has

$$
\tilde{\kappa}_{\theta,l} := \Pi\left(\dot{\ell}_{\theta,\alpha,l}\middle|\mathcal{S}^\perp\right),
$$

for each $l = 1, \ldots, L_\alpha$.

It is convenient to calculate this projection in two steps (see Bickel et al., 1998, p. 74). Firstly we calculate the effective score for the Euclidean parameters γ, i.e. the orthogonal projection of $(\dot{\ell}'_{\theta,\alpha}, \dot{\ell}'_{\theta,\sigma}, \dot{\ell}'_{\theta,b})'$ onto the orthocomplement of the space spanned by the score functions for the infinite dimensional parameter η, i.e. $\mathcal{T}^\perp$. We denote this by

[S3] The terminology "effective score" is taken from Choi, Hall and Schick (1996); much of the semiparametric literature calls this object the "efficient score" (e.g. Bickel et al., 1998; van der Vaart, 1998).

$\tilde{\ell}_\theta \coloneqq (\tilde{\ell}'_{\theta,\alpha}, \tilde{\ell}'_{\theta,\sigma}, \tilde{\ell}'_{\theta,b})' = (\tilde{\ell}'_{\theta,\alpha}, \tilde{\ell}'_{\theta,\beta})'$, i.e. for any $x \in \{\alpha, \sigma, b\}$ and l in $\{1, \ldots, L_x\}$

$$\tilde{\ell}_{\theta,x,l} = \Pi\left(\dot{\ell}_{\theta,x,l} \middle| \mathcal{T}^\perp\right). \tag{S8}$$

For the second step, we may partition

$$\tilde{\ell}_\theta = \left(\tilde{\ell}'_{\theta,\alpha}, \tilde{\ell}'_{\theta,\beta}\right)' \quad \text{and} \quad \tilde{I}_\theta = \begin{bmatrix} \tilde{I}_{\theta,\alpha\alpha} & \tilde{I}_{\theta,\alpha\beta} \\ \tilde{I}_{\theta,\beta\alpha} & \tilde{I}_{\theta,\beta\beta} \end{bmatrix}, \tag{S9}$$

with $\tilde{I}_\theta \coloneqq P_\theta[\tilde{\ell}_\theta \tilde{\ell}'_\theta]$. If $\tilde{I}_{\theta,\beta\beta}$ is nonsingular,[S4] we can (orthogonally) project once more to obtain the effective score function for α:[S5]

$$\tilde{\kappa}_\theta = \tilde{\ell}_{\theta,\alpha} - \tilde{I}_{\theta,\alpha\beta}\tilde{I}^{-1}_{\theta,\beta\beta}\tilde{\ell}_{\theta,\beta}, \tag{S10}$$

which has corresponding effective information matrix

$$\tilde{\mathcal{I}}_\theta \coloneqq \tilde{I}_{\theta,\alpha\alpha} - \tilde{I}_{\theta,\alpha\beta}\tilde{I}^{-1}_{\theta,\beta\beta}\tilde{I}_{\theta,\beta\alpha}. \tag{S11}$$

Lemma S3. *Suppose Assumptions 1 and 2 hold. Then the components of $\tilde{\ell}_\theta$ are as follows. For $x = \alpha$ or $x = \sigma$,*

$$\begin{aligned}\tilde{\ell}_{\theta,x,l}(W) = &\sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K} \zeta^x_{l,k,j}(\alpha,\sigma)\phi_k(e'_k A(\alpha,\sigma)V_\theta)e'_j A(\alpha,\sigma)V_\theta \\ &+ \sum_{k=1}^{K} \zeta^x_{l,k,k}(\alpha,\sigma)(\tau_{k,1}e'_k A(\alpha,\sigma)V_\theta + \tau_{k,2}\kappa(e'_k A(\alpha,\sigma)V_\theta)),\end{aligned}$$

[S4] If $\tilde{I}_{\theta,\beta\beta}$ is singular, we may drop components from $\tilde{\ell}_{\theta,\beta}$ until the remaining components form a linearly independent collection which span the same subspace of $L_2(P_\theta)$ as $\tilde{\ell}_{\theta,\beta}$. The corresponding variance matrix of this smaller vector will be non-singular and $\tilde{\ell}_{\theta,\beta}$ can be replaced throughout by this smaller vector.

[S5] For any $l = 1, \ldots, L_\alpha$, one has that

$$\kappa_{\theta,l} = \tilde{\ell}_{\theta,\alpha,l} - e'_l\tilde{I}_{\theta,\alpha\beta}\tilde{I}^{-1}_{\theta,\beta\beta}\tilde{\ell}_{\theta,\beta} = \Pi\left(\tilde{\ell}_{\theta,\alpha,l} \middle| \left[\text{Span}\left(\tilde{\ell}_{\theta,\beta}\right)\right]^\perp\right).$$

with l in $\{1,\ldots,L_\alpha\}$ or $\{1,\ldots,L_\sigma\}$ (respectively); for $x=b$,

$$\begin{aligned}\tilde{\ell}_{\theta,b}(W) = &-\sum_{k=1}^{K}\phi_k(e_k'A(\alpha,\sigma)V_\theta)e_k'A(\alpha,\sigma)\left([X'\otimes I_K]-\mathbb{E}[(X'\otimes I_K)]\right)\\ &+\sum_{k=1}^{K}e_k'A(\alpha,\sigma)\mathbb{E}[(X'\otimes I_K)](\varsigma_{k,1}e_k'A(\alpha,\sigma)V_\theta+\varsigma_{k,2}\kappa(e_k'A(\alpha,\sigma)V_\theta));\end{aligned}$$

where the expectations are taken under P_θ and

$$\tau_k := M_k^{-1}\begin{pmatrix}0\\-2\end{pmatrix},\ \varsigma_k := M_k^{-1}\begin{pmatrix}1\\0\end{pmatrix},\ \textit{for } M_k := \begin{pmatrix}1 & \mathbb{E}[\epsilon_k^3]\\ \mathbb{E}[\epsilon_k^3] & \mathbb{E}[\epsilon_k^4]-1\end{pmatrix}.$$

Proof. For each $h_k\in H_k$, define the corresponding $\tilde{h}_k$ as in the statement of Lemma S1 and let $\tilde{H}_k$ collect all such $\tilde{h}_k$ formed with h_k ranging over H_k.[S6] By the definition of $\tilde{\ell}_\theta$ in equation (S8) and Theorem 4.11 in Rudin (1987) it suffices to show that each such component is (a) in $(\tilde{H}_0+\cdots+\tilde{H}_K)^\perp$ and (b) $\dot{\ell}_{\theta,x}-\tilde{\ell}_{\theta,x}\in \mathrm{cl}(\tilde{H}_0+\cdots+\tilde{H}_K)$, the form of which is given in Lemma S12.

Case 1: $x=\alpha,\sigma$. For (a) note that if $j\neq k$, then

$$\begin{aligned}\mathbb{E}\left[\zeta^x_{l,k,j}(\alpha,\sigma)\phi_k(\epsilon_k)\epsilon_j h_0(\tilde{X})\right] &= \mathbb{E}\left[\zeta^x_{l,k,j}(\alpha,\sigma)\phi_k(\epsilon_k)h_0(\tilde{X})\right]\mathbb{E}[\epsilon_j]=0\\ \mathbb{E}\left[\zeta^x_{l,k,j}(\alpha,\sigma)\phi_k(\epsilon_k)\epsilon_j h_m(\epsilon_m)\right] &= \mathbb{E}\left[\zeta^x_{l,k,j}(\alpha,\sigma)\right]\mathbb{E}\left[\phi_k(\epsilon_k)\epsilon_j h_m(\epsilon_m)\right]=0\end{aligned}$$

where the last equality follows from independence and the fact that m must differ from one of k,j. Additionally, by independence and our moment assumptions (i.e. Assumption 2)

$$\mathbb{E}\left[\left(\zeta^x_{l,k,j}(\alpha,\sigma)[\tau_{k,1}\epsilon_k+\tau_{k,2}\kappa(\epsilon_k)]\right)h_0(\tilde{X})\right]=\zeta^x_{l,k,j}(\alpha,\sigma)\mathbb{E}\left[\tau_{k,1}\epsilon_k+\tau_{k,2}\kappa(\epsilon_k)\right]\mathbb{E}[h_0(\tilde{X})]=0,$$

and again using independence and the definition of H_k,

$$\mathbb{E}\left[\zeta^x_{l,k,j}(\alpha,\sigma)[\tau_{k,1}\epsilon_k+\tau_{k,2}\kappa(\epsilon_k)]h_j(\epsilon_j)\right]=\zeta^x_{l,k,j}(\alpha,\sigma)\mathbb{E}\left[\left(\tau_{k,1}\epsilon_k+\tau_{k,2}\kappa(\epsilon_k)\right)h_j(\epsilon_j)\right]=0.$$

Since $\epsilon_k=e_k'A(\alpha,\sigma)V_\theta$, these observations and the form of $\tilde{\ell}_{\theta,x}$ establish (a). For (b), it suffices to show that

$$f_k(\epsilon_k):=\phi_k(\epsilon_k)\epsilon_k+1-\tau_{k,1}\epsilon_k-\tau_{k,2}\kappa(\epsilon_k)\in H_k.$$

[S6] That is, for each $h_0\in H_0$ define $\tilde{h}_0:\mathcal{W}\to\mathbb{R}$ acccording to $\tilde{h}_0(W):=h_0(\tilde{X})$ and let $\tilde{H}_0$ collect the $\tilde{h}_0$ functions so formed. Similarly, for each $h_k\in H_k$ ($k=1,\ldots,K$), define $\tilde{h}_k:\mathcal{W}\to\mathbb{R}$ according to $\tilde{h}_k(W):=h_k(e_k'A(\alpha,\sigma)V_\theta)$ and let let $\tilde{H}_k$ collect the $\tilde{h}_k$ functions so formed.

That $\mathbb{E}[f_k(\epsilon_k)] = 0$ and $\mathbb{E}[f_k(\epsilon_k)^2] < \infty$ follows immediately from Assumption 2. That additionally $\mathbb{E}[f_k(\epsilon_k)\epsilon_k] = \mathbb{E}[f_k(\epsilon_k)\kappa(\epsilon_k)] = 0$ is ensured by the choice of τ_k.

Case 2: $x = b$. For (a) let $m(X) \coloneqq A(\alpha, \sigma)(X' \otimes I_K)$ and $\mu = \mathbb{E}[m(X)]$. Then,

$$\begin{aligned}
\mathbb{E}[\phi_k(\epsilon_k)e_k'(m(X) - \mu)h_0(\tilde{X})] &= \mathbb{E}[\phi_k(\epsilon_k)]\mathbb{E}[e_k'(m(X) - \mu)h_0(\tilde{X})] = 0 \\
\mathbb{E}[\phi_k(\epsilon_k)e_k'(m(X) - \mu)h_j(\epsilon_j)] &= \mathbb{E}[\phi_k(\epsilon_k)h_j(\epsilon_j)]\mathbb{E}[e_k'(m(X) - \mu)] = 0 \\
\mathbb{E}[e_k'\mu\left(\varsigma_{k,1}\epsilon_k + \varsigma_{k,2}\kappa(\epsilon_k)\right)h_0(\tilde{X})] &= e_k'\mu\mathbb{E}[\varsigma_{k,1}\epsilon_k + \varsigma_{k,2}\kappa(\epsilon_k)]\mathbb{E}[h_0(\tilde{X})] = 0;
\end{aligned}$$

for $k \neq j$ by independence

$$\mathbb{E}[e_k'\mu\left(\varsigma_{k,1}\epsilon_k + \varsigma_{k,2}\kappa(\epsilon_k)\right)h_j(\epsilon_j)] = e_k'\mu\mathbb{E}[\varsigma_{k,1}\epsilon_k + \varsigma_{k,2}\kappa(\epsilon_k)]\mathbb{E}[h_j(\epsilon_j)] = 0$$

whilst for $k = j$, the definition of H_k ensures that

$$\mathbb{E}[e_k'\mu\left(\varsigma_{k,1}\epsilon_k + \varsigma_{k,2}\kappa(\epsilon_k)\right)h_k(\epsilon_k)] = e_k'\mu\mathbb{E}[\varsigma_{k,1}\epsilon_k h_k(\epsilon_k) + \varsigma_{k,2}\kappa(\epsilon_k)h_k(\epsilon_k)] = 0.$$

Since $\epsilon_k = e_k'A(\alpha, \sigma)V_\theta$, these observations and the form of $\tilde{\ell}_{\theta,b}$ establish (a). For (b) it suffices to show that

$$q_k(\epsilon_k) \coloneqq \left(\phi_k(\epsilon_k) + \varsigma_{k,1}\epsilon_k + \varsigma_{k,2}\kappa(\epsilon_k)\right)(-e_k'\mu) \in H_k.$$

That $\mathbb{E}[q_k(\epsilon_k)] = 0$ and $\mathbb{E}[q_k(\epsilon_k)^2] < \infty$ follows immediately from Assumption 2. That additionally $\mathbb{E}[q_k(\epsilon_k)\epsilon_k] = \mathbb{E}[q_k(\epsilon_k)\kappa(\epsilon_k)] = 0$ is ensured by the choice of ς_k. □

S1.3 Proof of Theorem 1

S1.3.1 Log density score estimation

As discussed just prior to Assumption 3, the log density score estimator in (11) may be replaced by an alternative estimator, provided it satisfies some high level conditions. These are given in the following assumption.

Assumption S1. *Let ν_n be as in Assumption 3. We have estimators $\hat{\phi}_{k,n,\gamma}$ such that for (a) any sequence with elements $\theta_n = (\alpha_0, \beta_n, \eta) \in \Theta$ where $(\beta_n)_{n\in\mathbb{N}}$ is a deterministic sequence with $\sqrt{n}\|\beta_n - \beta\| = O(1)$ and (b) any array $(Z_{n,i})_{n\in\mathbb{N},i\leq n}$ with i.i.d. rows and such that $\mathbb{E}Z_{n,i} = 0$, $\sup_{n\in\mathbb{N}}\mathbb{E}Z_{n,i}^2 < \infty$ and $Z_{n,i} \perp\!\!\!\perp \epsilon_{i,k}$ for each n, i, and k,*

$$\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n}V_{\theta_n,i}) - \phi_k(A_{k,\gamma_n}V_{\theta_n,i})\right]Z_{n,i} = o_{P_{\theta_n}^n}(n^{-1/2}), \tag{S12}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\left[\hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n}V_{\theta_n,i}) - \phi_k(A_{k,\gamma_n}V_{\theta_n,i})\right] Z_{n,i}\right)^2 = o_{P^n_{\theta_n}}(\nu_n). \tag{S13}$$

where $A_{k,\gamma_n} := e_k' A(\alpha_0, \sigma_n)$, $V_{\theta_n,i} := Y_i - \mathrm{vec}^{-1}(b_n)X_i$.

The following Lemma verifies that, under Assumptions 2 and 3, the log density score estimator in (11) satisfies Assumption S1. Its proof is given in Section S6.

Lemma S4. *Suppose Assumptions 2 and 3 hold. Then,* $\hat{\phi}_{k,n,\gamma} := \hat{\phi}_{k,n}$ *as defined in* (11) *satisfies Assumption S1.*

S1.3.2 Proof of Theorem 1

In order to prove Theorem 1, we first establish two results which give high level conditions under which Theorem 1 holds. The proof of Theorem 1 then consists of verifying the required high level conditions under our primitive assumptions. Let us first recall the definitions of various objects which were introduced in Section 3.

We have that $\tilde{\ell}_\theta$ denotes the *effective score* for Euclidean parameter vector $\gamma = (\alpha, \beta)$, evaluated at θ (as defined in (S8) and derived in Lemma S3). The *effective information* for γ is denoted $\tilde{I}_\theta := P_\theta[\tilde{\ell}_\theta \tilde{\ell}_\theta']$. Given a $\gamma = (\alpha, \beta)$, these objects are estimated by $\hat{\ell}_{n,\gamma} = \hat{\ell}_{n,\gamma}(W_1, \ldots, W_n)$ and $\hat{I}_{n,\gamma} = \hat{I}_{n,\gamma}(W_1, \ldots, W_n)$, respectively. Each of these objects can be partitioned conformally with (α, β):

$$\tilde{\ell}_\theta = \begin{pmatrix} \tilde{\ell}_{\theta,\alpha} \\ \tilde{\ell}_{\theta,\beta} \end{pmatrix}, \ \hat{\ell}_{n,\gamma} = \begin{pmatrix} \hat{\ell}_{n,\gamma,\alpha} \\ \hat{\ell}_{n,\gamma,\beta} \end{pmatrix}, \ \tilde{I}_\theta = \begin{pmatrix} \tilde{I}_{\theta,\alpha\alpha} & \tilde{I}_{\theta,\alpha\beta} \\ \tilde{I}_{\theta,\beta\alpha} & \tilde{I}_{\theta,\beta\beta} \end{pmatrix}, \ \text{ and } \hat{I}_{n,\gamma} = \begin{pmatrix} \hat{I}_{n,\gamma,\alpha\alpha} & \hat{I}_{n,\gamma,\alpha\beta} \\ \hat{I}_{n,\gamma,\beta\alpha} & \hat{I}_{n,\gamma,\beta\beta} \end{pmatrix}.$$

The effective score for α is $\tilde{\kappa}_\theta := \tilde{\ell}_{\theta,\alpha} - \tilde{I}_{\theta,\alpha\beta}\tilde{I}_{\theta,\beta\beta}^{-1}\tilde{\ell}_{\theta,\beta}$, with corresponding effective information $\tilde{\mathcal{I}}_\theta := \tilde{I}_{\theta,\alpha\alpha} - \tilde{I}_{\theta,\alpha\beta}\tilde{I}_{\theta,\beta\beta}^{-1}\tilde{I}_{\theta,\beta\alpha}$.[S7] For a given γ, the estimator of $\tilde{\kappa}_\theta$ is

$$\hat{\kappa}_{n,\gamma} := \hat{\ell}_{n,\gamma,\alpha} - \hat{I}_{n,\gamma,\alpha\beta}\hat{I}_{n,\gamma,\beta\beta}^{-1}\hat{\ell}_{n,\gamma,\beta}.$$

The estimator of the effective information for α, $\tilde{\mathcal{I}}_\theta$, is formed in two steps. Firstly, the preliminary estimate $\check{\mathcal{I}}_{n,\gamma} := \hat{I}_{n,\gamma,\alpha\alpha} - \hat{I}_{n,\gamma,\alpha\beta}\hat{I}_{n,\gamma,\beta\beta}^{-1}\hat{I}_{n,\gamma,\beta\beta}$ is formed by replacing population quantities by their sample equivalents. Secondly, the regularized estimator $\hat{\mathcal{I}}_{n,\gamma}$ is formed as in (15): let $\check{U}_{n,\gamma}\check{\Lambda}_{n,\gamma}\check{U}_{n,\gamma}'$ be the eigendecomposition of the initial estimator $\check{\mathcal{I}}_{n,\gamma}$. $\check{\Lambda}_{n,\gamma}$ is a diagonal matrix with (i,i)th element $\check{\lambda}_{n,\gamma,i}$. Then the estimator is:

$$\hat{\mathcal{I}}_{n,\gamma}^t = \check{U}_{n,\gamma}\hat{\Lambda}_{n,\gamma}(\nu_n^{1/2})\check{U}_{n,\gamma}' \ ,$$

[S7] Here it is assumed that $\tilde{I}_{\theta,\beta\beta}$ is non-singular; cf. footnote S4.

where $\hat{\Lambda}_{n,\gamma}(\nu_n^{1/2})$ is a diagonal matrix with the $\nu_n^{1/2}$-truncated eigenvalues of $\hat{\mathcal{I}}_{n,\gamma}$ on the main diagonal, i.e. the (i,i)–th element of $\hat{\Lambda}_n(\nu_n^{1/2})$ is $\mathbf{1}(\check{\lambda}_{n,\gamma,i} \geq \nu_n^{1/2})$. The rank estimator used is $\hat{r}_{n,\gamma} = \operatorname{rank}(\hat{\mathcal{I}}^t_{n,\gamma})$. Finally, the effective score statistic (for a given γ) is given by

$$\hat{S}_{n,\gamma} \coloneqq n\left(\mathbb{P}_n\hat{\kappa}_{n,\gamma}\right)' \hat{\mathcal{I}}^{t,\dagger}_{n,\gamma}\left(\mathbb{P}_n\hat{\kappa}_{n,\gamma}\right),$$

where $\hat{\mathcal{I}}^{t,\dagger}_{n,\gamma}$ is the Moore – Penrose psuedoinverse of $\hat{\mathcal{I}}^t_{n,\gamma}$.

Theorem S1. *Suppose that for any deterministic sequence $(\tilde{\theta}_n)_{n\in\mathbb{N}}$ in Θ with elements $\tilde{\theta}_n = (\alpha, \beta_n, \eta)$ such that $\sqrt{n}\|\beta_n - \beta\| = O(1)$ the following conditions hold:*

1. *The functions $\tilde{\ell}_{\theta_n}$ satisfy*

$$\sqrt{n}\mathbb{P}_n\left[\tilde{\ell}_{\tilde{\theta}_n} - \tilde{\ell}_\theta\right] + \sqrt{n}\tilde{I}_\theta\begin{pmatrix}0\\ \beta_n - \beta\end{pmatrix} = o_{P^n_\theta}(1); \tag{S14}$$

2. *The estimators $\hat{\ell}_{n,\gamma_n}$ satisfy $\sqrt{n}\mathbb{P}_n\left[\hat{\ell}_{n,\gamma_n} - \tilde{\ell}_{\tilde{\theta}_n}\right] = o_{P^n_{\tilde{\theta}_n}}(1)$;*
3. *The estimators $\hat{I}_{n,\gamma_n}$ satisfy $\|\hat{I}_{n,\gamma_n} - \tilde{I}_\theta\|_2 = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2})$ for a non-negative sequence $(\nu_n)_{n\in\mathbb{N}}$ with $\nu_n \to 0$;*

where $\gamma_n \coloneqq (\alpha, \beta_n)$, $\theta \coloneqq (\alpha, \beta, \eta)$ and $\tilde{I}_\theta \coloneqq P_\theta[\tilde{\ell}_\theta\tilde{\ell}_\theta']$. Moreover, suppose that for $(g_n, h_n) \to (g, h)$ (all in $\mathcal{V}$) and some $\sigma(g,h) \in (0,\infty)$, under P^n_θ

$$\left(\sqrt{n}\mathbb{P}_n\tilde{\ell}_\theta,\ \log\prod_{i=1}^n \frac{p_{\theta_n(g_n,h_n)}}{p_\theta}\right) \rightsquigarrow \mathcal{N}\left(\begin{pmatrix}0\\ -\frac{1}{2}\sigma(g,h)\end{pmatrix}, \begin{pmatrix}\tilde{I}_\theta & \tilde{I}_\theta g\\ g'\tilde{I}_\theta & \sigma(g,h)\end{pmatrix}\right), \tag{S15}$$

where $\theta_n(g,h)$ is as in Lemma S2. Suppose that initial estimators $\hat{\beta}_n$ are available with $\sqrt{n}\|\hat{\beta}_n - \beta\| = O_{P^n_\theta}(1)$ and let $\bar{\beta}_n$ be a discretised version of this which takes values in $\mathsf{G}_n \coloneqq n^{-1/2}C\mathbb{Z}^{L_\beta}$ for some $C \in (0,\infty)$.[S8] *Then, if $\bar{\gamma}_n \coloneqq (\alpha, \bar{\beta}_n)$ and $r \coloneqq \operatorname{rank}\tilde{\mathcal{I}}_\theta$,*

$$\sqrt{n}\mathbb{P}_n\hat{\kappa}_{n,\bar{\gamma}_n} = \sqrt{n}\mathbb{P}_n\tilde{\kappa}_\theta + o_{P^n_{\theta_n(g_n,h_n)}}(1) \rightsquigarrow \mathcal{N}(0,\tilde{\mathcal{I}}_\theta), \quad \text{and} \quad \hat{S}_{n,\bar{\gamma}_n} \rightsquigarrow \chi^2_r, \tag{S16}$$

under any $P^n_{\theta_n(g_n,h_n)}$ such that $(g_n,h_n) \to (g,h)$ (all in $\mathcal{V}$) with $g = (0,(b,s)')' \in \mathbb{R}^{L_\alpha} \times \mathbb{R}^{L_\beta}$. Additionally, under any $P^n_{\theta_n(g_n,h_n)}$ such that $(g_n,h_n) \to (g,h)$ (all in $\mathcal{V}$),

$$\hat{r}_{n,\bar{\gamma}_n} \xrightarrow{P^n_{\theta_n(g_n,h_n)}} r. \tag{S17}$$

[S8] That is, $\bar{\beta}_n$ is the nearest element in G_n to $\hat{\beta}_n$.

Proof. Step 1: Let $d_n \coloneqq \sqrt{n}(\beta_n - \beta)$. By arguing along subsequences if necessary we may assume without loss of generality that $d_n \to d$. Hence for $g_n^\diamond \coloneqq (0, d_n')' \to (0, d')' =: g^\diamond$, $\tilde{\theta}_n = \theta_n(g_n^\diamond, 0)$. By condition (S15) and Example 6.5 in van der Vaart (1998), $P_{\tilde{\theta}_n}^n \lhd\rhd P_\theta^n$ and so, given the assumed convergences in conditions 2 and 3, we have

$$\sqrt{n}\mathbb{P}_n\left[\hat{\ell}_{n,\gamma_n} - \tilde{\ell}_{\tilde{\theta}_n}\right] = o_{P_\theta^n}(1) \quad \text{and} \quad \|\hat{I}_{n,\gamma_n} - \tilde{I}_\theta\|_2 = o_{P_\theta^n}(\nu_n^{1/2}).$$

Step 2: We show that the convergences in the preceding display and equation (S14) continue to hold if γ_n (and $\theta_n = (\gamma_n, \eta)$) is replaced by $\bar{\gamma}_n$ (and $\bar{\theta}_n = (\bar{\gamma}_n, \eta)$) as in the statement of the theorem. Let $\gamma^\star = (\alpha, \beta^\star)$ and $\theta^\star = (\gamma^\star, \eta)$ and define

$$\begin{aligned}
R_{n,1}(\beta^\star) &\coloneqq \sqrt{n}\mathbb{P}_n\left[\tilde{\ell}_{\theta^\star} - \tilde{\ell}_\theta\right] + \sqrt{n}\tilde{I}_\theta \begin{pmatrix} 0 \\ \beta^\star - \beta \end{pmatrix} \\
R_{n,2}(\beta^\star) &\coloneqq \sqrt{n}\mathbb{P}_n\left[\hat{\ell}_{n,\gamma^\star} - \tilde{\ell}_{\theta^\star}\right] \\
R_{n,3}(\beta^\star) &\coloneqq \nu_n^{-1/2}\left[\hat{I}_{n,\gamma^\star} - \tilde{I}_\theta\right].
\end{aligned}$$

For any $\varepsilon > 0$ there is an M such that $P_\theta^n(\sqrt{n}\|\hat{\beta}_n - \beta\| > M) < \epsilon$. Moreover, whenever $\sqrt{n}\|\hat{\beta}_n - \beta\| \le M$ then $\bar{\beta}_n \in \mathsf{G}_n^M \coloneqq \{\upbeta \in \mathsf{G}_n : \|\upbeta - \beta\| \le n^{-1/2}M\}$. For fixed M, the cardinality $|\mathsf{G}_n^M| < \infty$ of this set is bounded independently of n, say by G^M. For any $\upsilon > 0$,

$$\begin{aligned}
P_\theta^n\left(\|R_{n,i}(\bar{\beta}_n)\| > \upsilon\right) &\le \varepsilon + \sum_{\beta_n^\star \in \mathsf{G}_n^M} \left(\{\|R_{n,i}(\beta_n^\star)\| > \upsilon\} \cap \bar{\beta}_n = \beta_n^\star\right) \\
&\le \varepsilon + \sum_{\beta_n^\star \in \mathsf{G}_n^M} \left(\|R_{n,i}(\beta_n^\star)\| > \upsilon\right) \\
&\le \varepsilon + \mathsf{G}^M P_\theta^n\left(\|R_{n,i}(\beta_n^\diamond)\| > \upsilon\right),
\end{aligned}$$

where $\breve{\beta}_n \in \mathsf{G}_n^M$ is the maximiser of $\beta^\star \mapsto P_\theta^n(\|R_{n,i}(\beta^\star)\| > \upsilon)$. As $\breve{\beta}_n \in \mathsf{G}_n^M$, $\breve{\theta}_n \coloneqq (\alpha, \breve{\beta}_n, \eta)$ is a deterministic sequence with $\sqrt{n}\|\breve{\beta}_n - \beta\| = O_{P_\theta^n}(1)$. Thus, by equation (S14) and Step 1,

$$\begin{aligned}
\sqrt{n}\mathbb{P}_n\left[\tilde{\ell}_{\bar{\theta}_n} - \tilde{\ell}_\theta\right] + \sqrt{n}\tilde{I}_\theta \begin{pmatrix} 0 \\ \bar{\beta}_n - \beta \end{pmatrix} &= o_{P_\theta^n}(1); \\
\sqrt{n}\mathbb{P}_n\left[\hat{\ell}_{n,\bar{\gamma}_n} - \tilde{\ell}_{\bar{\theta}_n}\right] &= o_{P_\theta^n}(1); \\
\|\hat{I}_{n,\bar{\gamma}_n} - \tilde{I}_\theta\|_2 &= o_{P_\theta^n}(\nu_n^{1/2}).
\end{aligned} \tag{S18}$$

Step 3: Combine the first two lines of (S18) to obtain

$$\sqrt{n}\mathbb{P}_n\left[\hat{\ell}_{n,\bar{\gamma}_n} - \tilde{\ell}_\theta\right] = -\sqrt{n}\tilde{I}_\theta \begin{pmatrix} 0 \\ \bar{\beta}_n - \beta \end{pmatrix} + o_{P_\theta^n}(1).$$

By the third line of (S18),

$$\hat{\mathcal{K}}_{n,\bar{\gamma}_n} := \begin{bmatrix} I & -\hat{I}_{n,\bar{\gamma}_n,\alpha\beta}\hat{I}^{-1}_{n,\bar{\gamma}_n,\beta\beta} \end{bmatrix} \xrightarrow{P_\theta^n} \tilde{\mathcal{K}}_\theta := \begin{bmatrix} I & -\tilde{I}_{\theta,\alpha\beta}\tilde{I}^{-1}_{\theta,\beta\beta} \end{bmatrix}.$$

By (S15) and Example 6.5 in van der Vaart (1998), $P^n_{\theta_n(g_n,h_n)} \triangleleft\triangleright P^n_\theta$. In combination with the preceding two displays this gives

$$\begin{aligned}
&\sqrt{n}\mathbb{P}_n\left[\hat{\kappa}_{n,\bar{\gamma}_n} - \tilde{\kappa}_\theta\right] \\
&\quad = \left[\hat{\mathcal{K}}_{n,\bar{\gamma}_n} - \tilde{\mathcal{K}}_\theta\right]\sqrt{n}\mathbb{P}_n\left[\hat{\ell}_{n,\bar{\gamma}_n} - \tilde{\ell}_\theta\right] + \tilde{\mathcal{K}}_\theta\sqrt{n}\mathbb{P}_n\left[\hat{\ell}_{n,\bar{\gamma}_n} - \tilde{\ell}_\theta\right] + \left[\hat{\mathcal{K}}_{n,\bar{\gamma}_n} - \tilde{\mathcal{K}}_\theta\right]\sqrt{n}\mathbb{P}_n\tilde{\ell}_\theta \\
&\quad = -\tilde{\mathcal{K}}_\theta\tilde{I}_\theta\begin{pmatrix} 0 \\ \sqrt{n}(\bar{\beta}_n - \beta) \end{pmatrix} + o_{P^n_{\theta_n(g_n,h_n)}}(1) \\
&\quad = -\begin{pmatrix} \tilde{\mathcal{I}}_\theta & 0 \end{pmatrix}\begin{pmatrix} 0 \\ \sqrt{n}(\bar{\beta}_n - \beta) \end{pmatrix} + o_{P^n_{\theta_n(g_n,h_n)}}(1) \\
&\quad = o_{P^n_{\theta_n(g_n,h_n)}}(1).
\end{aligned}$$

By (S15) and Le Cam's third Lemma (e.g. van der Vaart, 1998, Example 6.7),

$$\sqrt{n}\mathbb{P}_n\tilde{\kappa}_\theta = \mathcal{K}_\theta\sqrt{n}\mathbb{P}_n\tilde{\ell}_\theta \rightsquigarrow \mathcal{K}_\theta\mathcal{Z}, \qquad \text{where } \mathcal{Z} \sim \mathcal{N}(\tilde{I}_\theta g, \tilde{I}_\theta)$$

under any $P^n_{\theta_n(g_n,h_n)}$ with $(g_n, h_n) \to (g, h)$ (all in $\mathcal{V}$). $\mathcal{K}_\theta\tilde{I}_\theta\mathcal{K}'_\theta = \tilde{\mathcal{I}}$ and with $g = (0, (b, s)')'$,

$$\mathcal{K}_\theta\tilde{I}_\theta g = \begin{pmatrix} \tilde{\mathcal{I}}_\theta & 0 \end{pmatrix}\begin{pmatrix} 0 \\ (b, s)' \end{pmatrix} = 0.$$

We conclude that

$$\hat{\kappa}_{n,\bar{\gamma}_n} \rightsquigarrow \mathcal{N}(0, \tilde{\mathcal{I}}_\theta) \quad \text{under } P^n_{\theta_n(g_n,h_n)}. \tag{S19}$$

For the final part of the proof, note that since any submatrix has a smaller operator norm than the original matrix and the matrix inverse is Lipschitz continuous at a non-singular matrix, the third line of (S18) implies that

$$\|\hat{\mathcal{I}}_{n,\bar{\gamma}_n} - \tilde{\mathcal{I}}_\theta\|_2 = o_{P^n_\theta}(\nu_n^{1/2}).$$

Therefore, by Proposition S1 and $P^n_{\theta_n(g_n,h_n)} \lhd\rhd P^n_\theta$,

$$\hat{\mathcal{I}}^{t,\dagger}_{n,\bar{\gamma}_n} \xrightarrow{P^n_{\theta_n(g_n,h_n)}} \tilde{\mathcal{I}}^\dagger_\theta \quad \text{and} \quad \hat{r}_{n,\bar{\gamma}_n} \xrightarrow{P^n_{\theta_n(g_n,h_n)}} r,$$

which gives (S17). For the final part of (S16), combine the preceding display with the weak convergence result in equation (S19) and Theorem 9.2.2 in Rao and Mitra (1971). □

Corollary 1. *In the setting of Theorem S1, let c_n be the $1-a$ quantile of the $\chi^2_{r_n}$ distribution for any $a \in (0,1)$ and*

$$\Theta_{0,n} = \left\{(\alpha_0, \beta + d/\sqrt{n}, \eta(1+h/\sqrt{n}) : d \in D^\star, h \in H^\star\right\},$$

where $D^\star$ is a bounded subset of $\mathbb{R}^{L_\beta}$ and $H^\star$ is a compact subset of H.[S9] *Then,*

$$\lim_{n\to\infty} \sup_{\vartheta \in \Theta_{0,n}} P^n_\vartheta\left(\hat{S}_{n,\bar{\gamma}_n} > c_n\right) \leq a,$$

with inequality only if $r = 0$.

Proof. Set $\hat{S}_n \coloneqq \hat{S}_{n,\bar{\gamma}_n}$, $\hat{r}_n \coloneqq \hat{r}_{n,\bar{\gamma}_n}$ and $\varphi_n \coloneqq \mathbf{1}\{\hat{S}_n > c_n\}$. Let g, h be such that $g = (0, d)$, $d \in D^\star$ and $h \in H^\star$. Since $\hat{r}_n \xrightarrow{P^n_\theta} r$ (by Theorem S1), the events $E_n \coloneqq \{\hat{r}_n = r\}$ satisfy $P^n_\theta E_n \to 1$. Thus $c_n \xrightarrow{P^n_\theta} c$, the $1-a$ quantile of a χ^2_r random variable. We now split into cases.

Case 1: $r > 0$. By Theorem S1

$$\hat{S}_n - c_n \rightsquigarrow \mathcal{Z} - c \text{ under } P^n_{\theta_n(g,h)} \text{ as } n \to \infty,$$

with $\mathcal{Z} \sim \chi^2_r$. Since this is a continuous distribution

$$\lim_{n\to\infty} P^n_{\theta_n(g,h)}\varphi_n = a.$$

Case 2: $r = 0$. On E_n, $\hat{r}_n = 0 \implies \hat{\mathcal{I}}_{n,\bar{\gamma}_n} = 0 \implies \hat{S}_n = 0 \implies \varphi_n = 0$, whilst $P^n_{\theta_n(g,h)} E_n \to 1$ by the contiguity which follows from (S15) and Example 6.5 in van der Vaart (1998). Thus

$$\lim_{n\to\infty} P^n_{\theta_n(g,h)}\varphi_n = 0.$$

These two limiting statements continue to hold under any convergent sequence $(g_n, h_n) \to (g, h)$, with each $g_n = (0, d_n)$ for $d_n \in D^\star$ and $h_n \in H^\star$ and $(g, h) \in \operatorname{cl} D^\star \times H^\star$, as

[S9] See the discussion immediately preceding Lemma S2 for the norm used on H.

follows directly from $d_{TV}(P^n_{\theta_n(g_n,h_n)}, P^n_{\theta_n(g,h)}) \to 0$ as shown in Lemma S11. Considering such convergent sequences is sufficient since each $(g_n, h_n) \in \{0\} \times \operatorname{cl} D^\star \times H^\star$, which is compact. $\square$

We next prove our main Theorem by verifying the conditions of Corollary 1.

Proof of Theorem 1. It suffices to show the conditions of Corollary 1 hold. There are 4 conditions which we verify in order: items 1, 2, 3 & equation (S15) of the statement of Theorem S1.

Condition 1: Let $d_n := \sqrt{n}(\beta_n - \beta)$ and $g_n = (0, d_n)$. Then $\tilde{\theta}_n = \theta_n(g_n, 0)$. By arguing along subsequences if necessary we may assume without loss of generality that $d_n \to d$. By Theorem 12.14 in Rudin (1991),

$$P_\theta\left[\tilde{\ell}_\theta \dot{\ell}_\theta'\right] g = \tilde{I}_\theta \begin{pmatrix} 0 \\ d \end{pmatrix} = \tilde{I}_\theta \begin{pmatrix} 0 \\ \sqrt{n}(\beta_n - b) \end{pmatrix} + o(1). \tag{S20}$$

Given this, condition 1 follows by Proposition A.10 in van der Vaart (1988), the hypotheses of which are verified by Lemmas S1, S13 and S14.

Condition 2: This follows by repeated addition and subtraction along with the convergence in probability and stochastic boundedness results of Lemma S15, Lemma S4, the moment conditions in Assumption 2 and the boundedness of $A(\alpha, \sigma_n)$, $A(\alpha, \sigma_n)^{-1}$ and $D_{x,l}(\alpha, \sigma_n)$ (for $x \in \{\alpha, \sigma\}$), which follows as each of these functions is continuous by Assumption 1 and $(\sigma_n)_{n\in\mathbb{N}}$ is a convergent sequence.

Condition 3: Let $\breve{I}_{n,\theta_n} := \frac{1}{n}\sum_{i=1}^n \tilde{\ell}_{\tilde{\theta}_n} \tilde{\ell}_{\tilde{\theta}_n}'$. By repeated addition and subtraction along with the results of Lemmas S4, S17 and S18,

$$\frac{1}{n}\sum_{i=1}^n \|\tilde{\ell}_{\tilde{\theta}_n} - \hat{\ell}_{n,\gamma_n}\|^2 = o_{P^n_{\tilde{\theta}_n}}(\nu_n).$$

This and Lemma S13 imply that

$$\begin{aligned}\left\|\hat{I}_{n,\gamma_n}-\breve{I}_{n,\tilde{\theta}_n}\right\|_2 &= \left\|\frac{1}{n}\sum_{i=1}^n \hat{\ell}_{n,\gamma_n}\left(\hat{\ell}_{n,\gamma_n}-\tilde{\ell}_{\tilde{\theta}_n}\right)'+\left(\hat{\ell}_{n,\gamma_n}-\tilde{\ell}_{\tilde{\theta}_n}\right)\tilde{\ell}'_{\tilde{\theta}_n}\right\|_2\\
&\leq \frac{1}{n}\sum_{i=1}^n\left\|\hat{\ell}_{n,\gamma_n}\left(\hat{\ell}_{n,\gamma_n}-\tilde{\ell}_{\tilde{\theta}_n}\right)'\right\|_2+\frac{1}{n}\sum_{i=1}^n\left\|\left(\hat{\ell}_{n,\gamma_n}-\tilde{\ell}_{\tilde{\theta}_n}\right)\tilde{\ell}'_{\tilde{\theta}_n}\right\|_2\\
&\leq \left(\frac{1}{n}\sum_{i=1}^n\left\|\hat{\ell}_{n,\gamma_n}\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n\left\|\hat{\ell}_{n,\gamma_n}-\tilde{\ell}_{\tilde{\theta}_n}\right\|^2\right)^{1/2}\\
&\quad+\left(\frac{1}{n}\sum_{i=1}^n\left\|\hat{\ell}_{n,\gamma_n}-\tilde{\ell}_{\tilde{\theta}_n}\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n\left\|\tilde{\ell}_{\tilde{\theta}_n}\right\|^2\right)^{1/2}\\
&= o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2}).\end{aligned}$$

To complete the demonstration of Condition 3, we show that the right hand side terms in

$$\|\breve{I}_{n,\tilde{\theta}_n}-\tilde{I}_\theta\|\leq\left\|\mathbb{P}_n\left[\tilde{\ell}_{\tilde{\theta}_n}\tilde{\ell}'_{\tilde{\theta}_n}-P_{\tilde{\theta}_n}\left[\tilde{\ell}_{\tilde{\theta}_n}\tilde{\ell}'_{\tilde{\theta}_n}\right]\right]\right\|+\left\|P_{\tilde{\theta}_n}\left[\tilde{\ell}_{\tilde{\theta}_n}\tilde{\ell}'_{\tilde{\theta}_n}\right]-P_\theta\left[\tilde{\ell}_\theta\tilde{\ell}'_\theta\right]\right\|$$

are respectively $o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2})$ and $o(\nu_n^{1/2})$. Under $P^n_{\tilde{\theta}_n}$, each $e_k'A(\alpha,\sigma_n)V_{\tilde{\theta}_n,i}$ has the same law as $\epsilon_{k,i}$ $(k=1,\ldots,K)$, whilst the same is true for $A(\alpha,\sigma)V_{\theta,i}$ under P^n_θ. This, $\sqrt{n}\|\beta_n-\beta\|=O(1)$ and the local Lipschitz continuity of each $\beta\mapsto\zeta^x_{l,j,k}(\alpha,\sigma)$ and $\beta\mapsto A(\alpha,\sigma)$ yield that the rightmost term is $O(n^{-1/2})=o(\nu_n^{1/2})$. For the first term on the right hand side we note that $\sup_{n\in\mathbb{N}}P_{\tilde{\theta}_n}\|\tilde{\ell}_{\tilde{\theta}_n}\tilde{\ell}'_{\tilde{\theta}_n}\|^{2+\delta/2}<\infty$ by Lemma S13. This is sufficient as either $1+\delta/4>p=2$, in which case $\mathbb{P}_n\left[\tilde{\ell}_{\tilde{\theta}_n}\tilde{\ell}'_{\tilde{\theta}_n}-P_{\tilde{\theta}_n}\left[\tilde{\ell}_{\tilde{\theta}_n}\tilde{\ell}'_{\tilde{\theta}_n}\right]\right]=O_{P^n_{\tilde{\theta}_n}}(n^{-1/2})=o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2})$ by Lindeberg's CLT or $p=1+\delta/4\in(1,2)$ whence $\mathbb{P}_n\left[\tilde{\ell}_{\tilde{\theta}_n}\tilde{\ell}'_{\tilde{\theta}_n}-P_{\tilde{\theta}_n}\left[\tilde{\ell}_{\tilde{\theta}_n}\tilde{\ell}'_{\tilde{\theta}_n}\right]\right]=O_{P^n_{\tilde{\theta}_n}}(n^{(1-p)/p})=o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2})$ by a Marcinkiewicz – Zygmund style weak law of large numbers for triangular arrays.[S10]

Condition 4: By Lemma S1, Lemma 1.7 of van der Vaart (2002) and Theorem I.2.7 of Conway (1985), the random vector

$$\left(\tilde{\ell}_\theta(W_i),\ g'\dot{\ell}_\theta(W_i)+\sum_{k=0}^K\tilde{h}_k(W_i)\right)$$

is zero mean and has a finite variance matrix under P_θ. By the definition of $\tilde{\ell}_\theta$ as an

[S10] A formal statemement is as follows: Let $(X_{n,i})_{n\in\mathbb{N},1\leq i\leq n}$ be a triangular array of zero-mean random variables, i.i.d. along rows. Let $S_n\coloneqq\sum_{i=1}^n X_{n,i}$. If $\sup_{n\in\mathbb{N}}\mathbb{E}|X_{n,1}|^p<\infty$ for $p\in(1,2)$, then $S_n/n^{1/p}$ converges to zero in probability as $n\to\infty$. For the case of an i.i.d. sequence (in place of a triangular array) this result is recorded as, for example, Theorem 6.3.2 of Gut (2005); the proof given there extends essentially verbatim to the case considered here.

orthogonal projection and Theorem 12.14 in Rudin (1991), one has

$$P_\theta\left[\tilde{\ell}_\theta\left(g'\dot{\ell}_\theta + \sum_{k=0}^{K}\tilde{h}_k\right)\right] = P_\theta\left[\tilde{\ell}_\theta\dot{\ell}_\theta'\right]g = \tilde{I}_\theta g.$$

Therefore, by the central limit theorem, under P_θ^n

$$\sqrt{n}\mathbb{P}_n\left(\tilde{\ell}_\theta,\ g'\dot{\ell}_\theta + \sum_{k=0}^{K}\tilde{h}_k\right) \rightsquigarrow \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\tilde{I}_\theta & \tilde{I}_\theta g\\ g'\tilde{I}_\theta & \sigma(g,h)\end{pmatrix}\right), \tag{S21}$$

where

$$\sigma(g,h) \coloneqq P_\theta\left[g'\dot{\ell}_\theta + \sum_{k=0}^{K}\tilde{h}_k\right]^2.$$

Combination of this with Lemma S2 and equation (S7) verifies (S15). □

S2 A more general model

S2.1 Model setup, ULAN and the effective score

In this section we extend the approach in the main paper to the more general model:

$$Y_i = B(b, X_i) + A(\alpha, \sigma, X_i)^{-1}\epsilon_i\ , \qquad i = 1, \dots, n\ , \tag{S22}$$

under Assumptions S2 and S3 below, which are weakened versions of Assumptions 1 and 2 respectively. This version of the model allows (a) (parametric) conditional heteroskedasticity in the reduced form error $A(\alpha, \sigma, X_i)^{-1}\epsilon_i$ and (b) the conditional mean $\mathbb{E}[Y_i|X_i] = B(b, X_i)$ to be a non-linear function of X_i, known up to a finite dimensional parameter b.

Assumption S2. *Suppose that for all $(\alpha, \beta) \in \mathcal{A} \times \mathcal{B}$,*

1. *$A(\alpha, \sigma, X)$ is non-singular for all X;*
2. *$(\alpha, \sigma) \mapsto A(\alpha, \sigma, X)$ and $b \mapsto B(b, X)$ are continuously differentiable for all X.*

Define the partial derivative matrices $D_{\alpha,l}(\alpha, \sigma, X) = \partial A(\alpha, \sigma, X)/\partial\alpha_l$, for $l = 1, \dots, L_\alpha$ $D_{\sigma,l}(\alpha, \sigma, X) = \partial A(\alpha, \sigma, X)/\partial\sigma_l$, for $l = 1, \dots, L_\sigma$ and $D_{b,l} \coloneqq \partial B(b, X)/\partial b_l$ for $l = 1, \dots, L_b$. Further, for each $k, j \in \{1, \dots, K\}$, $l \in \{1, \dots, L_\alpha\}$ and $m \in \{1, \dots, L_\sigma\}$ define $\zeta^\alpha_{l,k,j}(\alpha, \sigma, X) \coloneqq e_k' D_{\alpha,l}(\alpha, \sigma, X)A(\alpha, \sigma, X)^{-1}e_j$ and $\zeta^\sigma_{m,k,j} \coloneqq e_k' D_{\sigma,m}(\alpha, \sigma, X)A(\alpha, \sigma, X)^{-1}e_j$. With this notation, for all $(\alpha, \beta) \in \mathcal{A} \times \mathcal{B}$

3. $(\alpha, \sigma) \to \zeta^{\alpha}_{l,k,j}(\alpha, \sigma, X)$ *and* $(\alpha, \sigma) \to \zeta^{\sigma}_{m,k,j}(\alpha, \sigma, X)$ *are locally Lipschitz continuous for all* j, k, l, m *considered and all* X.

4. $\|A(\alpha, \sigma, X)\|$, $\|A(\alpha, \sigma, X)^{-1}\|$, $\|D_{\alpha,l}(\alpha, \sigma, X)\|$ *and* $\|D_{\sigma,l}(\alpha, \sigma, X)\|$ *are locally (in* (α, σ)*) bounded.*

Assumption S3. *For* $\epsilon_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,K})'$ *in model (S22), each component* $\epsilon_{i,k}$ *has a continuously differentiable root density (with respect to Lebesgue measure on* $\mathbb{R}$*). We write the density as* η_k *with log density score* $\phi_k(x) = \partial \log \eta_k(x)/\partial x$*. We assume that for all* $k = 1, \ldots, K$ *and some* $\delta > 0$

1. $\mathbb{E}\epsilon_{i,k} = 0$, $\mathbb{E}\epsilon^2_{i,k} = 1$, $\mathbb{E}\epsilon^{4+\delta}_{i,k} < \infty$, $\mathbb{E}(\epsilon^4_{i,k}) - 1 > \mathbb{E}(\epsilon^3_{i,k})^2$, *and* $\mathbb{E}\phi^{4+\delta}_k(\epsilon_{i,k}) < \infty$;

2. $\mathbb{E}\phi_k(\epsilon_{i,k}) = 0$, $\mathbb{E}\phi_k(\epsilon_{i,k})\epsilon_{i,k} = -1$, $\mathbb{E}\phi_k(\epsilon_{i,k})\epsilon^2_{i,k} = 0$ *and* $\mathbb{E}\phi_k(\epsilon_{i,k})\epsilon^3_{i,k} = -3$;

3. $\epsilon_{i,k}$ *is independent of* $\epsilon_{i,l}$ *for all* $k \neq l$;

4. $\eta_0 \in \mathscr{Z}$ *is a density function (with respect to Lebesgue measure on* $\mathbb{R}^{d-1}$*) such that if* $\tilde{X}_i \sim \eta_0$, $\mathbb{E}[\|D_{b,l}(b + \varrho, X_i)\|^{4+\delta}] \leq \overline{D}_{b,l}(b) < \infty$ *for all* $b \in \mathcal{B}$*, all* ϱ *in a neighbourhood of zero and all* $l = 1, \ldots, L_b$;

5. ϵ_i *and* $\tilde{X}_i$ *are independent.*

Remark 1. *If* $A(\alpha, \sigma, X) = A(\alpha, \sigma)$ *and* $B(b, X) = \mathrm{vec}^{-1}(b)X$ *then Assumptions S2 and S3 are implied by Assumptions 1 and 2 respectively.*

Formally, the considered model is the collection

$$\mathcal{P}_\Theta = \{P_\theta : \theta \in \Theta\} \,, \tag{S23}$$

where each P_θ is the law of the data $W_i = (Y_i, \tilde{X}_i)$ which lies in $\mathcal{W} \subset \mathbb{R}^{K+d-1}$. The parameter space Θ has the form $\Theta = \mathcal{A} \times \mathcal{B} \times \mathcal{H}$, where $\mathcal{A} \subset \mathbb{R}^{L_\alpha}$, $\mathcal{B} \subset \mathbb{R}^{L_\beta}$. $\mathcal{H}$ has the form $\mathscr{Z} \times \prod_{k=1}^K \mathscr{H}$, where $\mathscr{Z}$ is the space of density functions η_0 and $\mathscr{H}$ is the space of density functions η_k such that if $\tilde{X} \sim \eta_0$ and $\epsilon_k \sim \eta_k$ then Assumption S3 parts 1, 3, 4 and 5 hold.[S11].

We write a typical element of Θ as $\theta = (\alpha, \beta, \eta)$, where $\beta = (b', \sigma')'$ and it is understood that $\alpha \in \mathcal{A}$, $\beta \in \mathcal{B}$ and $\eta \in \mathcal{H}$. In what follows we will let $V_{\theta,i} \coloneqq Y_i - B(b, X_i)$ be

[S11] Part 2 of Assumption S3 serves to simplify the form of the effective score function derived in Lemma S7 and is not necessary to set up the model.

the reduced form error so that $A(\alpha, \sigma, X_i)V_{\theta,i} = \epsilon_i$. Each P_θ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^{K+d-1}$, with (Lebesgue) density given by

$$p_\theta(W_i) = |\det A(\alpha, \sigma, X_i)| \prod_{k=1}^{K} \eta_k(e_k' A(\alpha, \sigma, X_i)V_{\theta,i}) \times \eta_0(\tilde{X}_i) \ , \tag{S24}$$

and hence log-density

$$\ell_\theta(W_i) = \log |\det A(\alpha, \sigma, X_i)| + \sum_{k=1}^{K} \log \eta_k(e_k' A(\alpha, \sigma, X_i)V_{\theta,i}) + \log \eta_0(\tilde{X}_i) \ . \tag{S25}$$

The differentiable paths we consider have the following form.

Let $H = H_0 \times \prod_{k=1}^{K} H_k$, where each H_k is as defined following (6). Given a direction $(g, h) \in \mathbb{R}^L \times H$, the measures P_t are those corresponding to the density with form as in (S24) evaluated at $\theta_t := (\gamma + tg, \eta_t)$ where the k-th coordinate of η_t is $\eta_{k,t}^{h_k} := \eta_k(1 + th_k)$ $(k = 0, \ldots, K)$.

We have the following analogues of Lemmas S1, S2 and S3.

Lemma S5. *Suppose Assumptions S2 and S3 hold and that (α, β) is an interior point of $\mathcal{A} \times \mathcal{B}$. For each $(g, h) \in \mathbb{R}^L \times H := \mathcal{V}$, the map $t \mapsto P_{\theta_t}$ is a differentiable path, with score function $g'\dot{\ell}_\theta + \tilde{h}_0 + \sum_{k=1}^{K} \tilde{h}_k$, where $\dot{\ell}_\theta := \nabla_\gamma \log p_\theta$, $\tilde{h}_0(W) := h_0(\tilde{X})$ and $\tilde{h}_k(W) := h_k(e_k' A(\alpha, \sigma, X)V_\theta)$. $\dot{\ell}_\theta$ has the form $\dot{\ell}_\theta = (\dot{\ell}_{\theta,\alpha}', \dot{\ell}_{\theta,b}', \dot{\ell}_{\theta,\sigma}')'$, with*

$$\begin{aligned}
\dot{\ell}_{\theta,\alpha,l}(W) &:= \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta_{l,k,j}^{\alpha}(\alpha, \sigma, X)\phi_k(e_k' A(\alpha, \sigma, X)V_\theta)e_j' A(\alpha, \sigma, X)V_\theta \\
&\qquad + \sum_{k=1}^{K} \zeta_{l,k,k}^{\alpha}(\alpha, \sigma, X)[\phi_k(e_k' A(\alpha, \sigma, X)V_\theta)e_k' A(\alpha, \sigma, X)V_\theta + 1]; \\
\dot{\ell}_{\theta,b,l}(W) &:= -\sum_{k=1}^{K} \phi_k\left(e_k' A(\alpha, \sigma, X)V_\theta\right) e_k' A(\alpha, \sigma, X)D_{b,l}(b, X); \\
\dot{\ell}_{\theta,\sigma,l}(W) &:= \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta_{l,k,j}^{\sigma}(\alpha, \sigma, X)\phi_k(e_k' A(\alpha, \sigma, X)V_\theta)e_j' A(\alpha, \sigma, X)V_\theta \\
&\qquad + \sum_{k=1}^{K} \zeta_{l,k,k}^{\sigma}(\alpha, \sigma, X)[\phi_k(e_k' A(\alpha, \sigma, X)V_\theta)e_k' A(\alpha, \sigma, X)V_\theta + 1].
\end{aligned}$$

Proof. Let $g = (a, \varrho, s) \in \mathbb{R}^{L_\alpha} \times \mathbb{R}^{L_b} \times \mathbb{R}^{L_\sigma}$. The log density of W under θ_t is then

$$\begin{aligned}\ell_{\theta_t}(W) &= \log p_{\theta_t}(W)\\ &= \log \eta_0(\tilde{X}) + \log(1 + th_0(\tilde{X})) + \log|\det(A(\alpha + ta, \sigma + ts, X))|\\ &\quad + \sum_{k=1}^{K} \log \eta_k\left(e_k' A(\alpha + ta, \sigma + ts, X)(Y - B(b + t\varrho, X))\right)\\ &\quad + \sum_{k=1}^{K} \log\left(1 + th_k\left(e_k' A(\alpha + ta, \sigma + ts, X)(Y - B(b + t\varrho, X))\right)\right).\end{aligned}$$

By Lemma S10, $t \mapsto \sqrt{p_{\theta_t}}$ is continuously differentiable (pointwise) in a neighbourhood $\mathcal{V}$ of 0. Moreover, if we define $q_t(W) \coloneqq \frac{\partial \log p_{\theta_x}(W)}{\partial x}\big|_{x=t}$ and $Q_t \coloneqq P_{\theta_t} q_t(W)^2$, Q_t is finite and continuous in a neighbourhood of 0 by the uniformly integrability of $\{q_t(W)^2 : t \in \mathcal{V}\}$ along with the pointwise continuity of $t \mapsto q_t(W)$, both of which follow from Lemma S10.

$$\begin{aligned}&\sum_{k=1}^{K} \phi_k\left(e_k' A(\alpha, \sigma, X) V_\theta\right) e_k' \sum_{l=1}^{L_\alpha} a_l D_{\alpha,l}(\alpha, \sigma, X) V_\theta + \sum_{l=1}^{L_\alpha} a_l \operatorname{tr}(A(\alpha, \sigma, X)^{-1} D_{\alpha,l}(\alpha, \sigma, X))\\ &+ \sum_{k=1}^{K} \phi_k\left(e_k' A(\alpha, \sigma, X) V_\theta\right) e_k' \sum_{l=1}^{L_\sigma} s_l D_{\sigma,l}(\alpha, \sigma, X) V_\theta + \sum_{l=1}^{L_\sigma} s_l \operatorname{tr}(A(\alpha, \sigma, X)^{-1} D_{\sigma,l}(\alpha, \sigma, X))\\ &- \sum_{k=1}^{K} \phi_k\left(e_k' A(\alpha, \sigma, X) V_\theta\right) e_k' A(\alpha, \sigma, X) \sum_{l=1}^{L_b} \varrho_l D_{b,l}(b, X)\\ &+ h_0(\tilde{X}) + \sum_{k=1}^{K} h_k\left(e_k' A(\alpha, \sigma, X) V_\theta\right).\end{aligned} \tag{S26}$$

We can re-write the two expressions involving the trace as follows: for any $x \in \{\alpha, \sigma\}$ and appropriate index l we have

$$\begin{aligned}&\sum_{k=1}^{K} \phi_k(e_k' A(\alpha, \sigma, X) V_\theta) e_k' D_{x,l}(\alpha, \sigma, X) V_\theta + \operatorname{tr}(A(\alpha, \sigma, X)^{-1} D_{x,l}(\alpha, \sigma, X))\\ &\quad = \sum_{k=1}^{K} \phi_k(e_k' A(\alpha, \sigma, X) V_\theta) e_k' D_{x,l}(\alpha, \sigma, X) A(\alpha, \sigma, X)^{-1} \epsilon + \operatorname{tr}(D_{x,l}(\alpha, \sigma, X) A(\alpha, \sigma, X)^{-1})\\ &\quad = \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta_{l,k,j}^{x}(\alpha, \sigma, X) \phi_k(e_k' A(\alpha, \sigma, X) V_\theta) e_j' A(\alpha, \sigma, X) V_\theta\\ &\qquad + \sum_{k=1}^{K} \zeta_{l,k,k}^{x}(\alpha, \sigma, X)[\phi_k(e_k' A(\alpha, \sigma, X) V_\theta) e_k' A(\alpha, \sigma, X) V_\theta + 1],\end{aligned}$$

for $\zeta^x_{l,k,j}(\alpha, \sigma, X) := e_k' D_{x,l}(\alpha, \sigma, X) A(\alpha, \sigma, X)^{-1} e_j$. We may therefore write the derivative (S26) as $a'\dot{\ell}_{\theta,\alpha} + \varrho'\dot{\ell}_{\theta,b} + s'\dot{\ell}_{\theta,\sigma} + \dot{\ell}_{\theta,\eta,h}$ where

$$\dot{\ell}_{\theta,\eta,h} := h_0(\tilde{X}) + \sum_{k=1}^{K} h_k\left(e_k' A(\alpha, \sigma, X) V_\theta\right) = \tilde{h}_0(W) + \sum_{k=1}^{K} \tilde{h}_k(W).$$

An elementary calculation reveals that $g'\dot{\ell}_\theta = a'\dot{\ell}_{\theta,\alpha} + \varrho'\dot{\ell}_{\theta,b} + s'\dot{\ell}_{\theta,\sigma}$. □

Lemma S6. *Suppose that Assumptions S2 and S3 hold and that (α, β) is an interior point of $\mathcal{A} \times \mathcal{B}$. For $(g, h) \in \mathcal{V}$ let*

$$\theta_n(g, h) := \theta + n^{-1/2}(g, \eta_0 h_0, \ldots, \eta_K h_K).$$

For any convergent sequence $(g_n, h_n) \to (g, h)$ (all in $\mathcal{V}$), define R_n as

$$R_n := \log \prod_{i=1}^{n} \frac{p_{\theta_n(g_n,h_n)}(W_i)}{p_\theta(W_i)} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g'\dot{\ell}_\theta(W_i) + \sum_{k=0}^{K} \tilde{h}_k(W_i) \right] + \frac{1}{2}\mathbb{E}\left[g'\dot{\ell}_\theta(W_i) + \sum_{k=0}^{K} \tilde{h}_k(W_i) \right]^2.$$

Then,

1. *$R_n \xrightarrow{P_\theta} 0$,*
2. *Under P_θ,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g'\dot{\ell}_\theta(W_i) + \sum_{k=0}^{K} \tilde{h}_k(W_i) \right] \rightsquigarrow \mathcal{N}\left(0, \mathbb{E}\left[g'\dot{\ell}_\theta(W_i) + \sum_{k=0}^{K} \tilde{h}_k(W_i) \right]^2\right),$$

3. *The (product) measures $P^n_{\theta_n}$ and P^n_θ are mutually contiguous.*

Proof. The proof proceeds verbatim as that of Lemma S2 on replacing Lemma S1 with Lemma S5. □

Lemma S7. *Suppose Assumptions S2 and S3 hold. Then the components of $\tilde{\ell}_\theta$ are as follows.*

For $x = \alpha$ or $x = \sigma$,

$$\begin{aligned}
\tilde{\ell}_{\theta,x,l}(W) = &\sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta^x_{l,k,j}(\alpha,\sigma,X)\phi_k(e_k' A(\alpha,\sigma,X)V_\theta)e_j' A(\alpha,\sigma,X)V_\theta \\
&+ \sum_{k=1}^{K} \left(\zeta^x_{l,k,k}(\alpha,\sigma,X) - \mathbb{E}\left[\zeta^x_{l,k,k}(\alpha,\sigma,X)\right]\right)\left[\phi_k(e_k' A(\alpha,\sigma,X)V_\theta)e_k' A(\alpha,\sigma,X)V_\theta + 1\right] \\
&+ \sum_{k=1}^{K} \mathbb{E}\left[\zeta^x_{l,k,k}(\alpha,\sigma,X)\right](\tau_{k,1}e_k' A(\alpha,\sigma,X)V_\theta + \tau_{k,2}\kappa(e_k' A(\alpha,\sigma,X)V_\theta)),
\end{aligned}$$

with l in $\{1,\ldots,L_\alpha\}$ or $\{1,\ldots,L_\sigma\}$ (respectively); for $l = 1,\ldots,L_b$,

$$\begin{aligned}
\tilde{\ell}_{\theta,b,l}(W) = &-\sum_{k=1}^{K} \phi_k(e_k' A(\alpha,\sigma,X)V_\theta)e_k' \left(A(\alpha,\sigma,X)D_{b,l}(b,X) - \mathbb{E}[A(\alpha,\sigma,X)D_{b,l}(b,X)]\right) \\
&+ \sum_{k=1}^{K} e_k' \mathbb{E}[A(\alpha,\sigma,X)D_{b,l}(b,X)](\varsigma_{k,1}e_k' A(\alpha,\sigma,X)V_\theta + \varsigma_{k,2}\kappa(e_k' A(\alpha,\sigma,X)V_\theta));
\end{aligned}$$

where the expectations are taken under P_θ and

$$\tau_k := M_k^{-1}\begin{pmatrix} 0 \\ -2 \end{pmatrix}, \ \varsigma_k := M_k^{-1}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ \text{for } M_k := \begin{pmatrix} 1 & \mathbb{E}[\epsilon_k^3] \\ \mathbb{E}[\epsilon_k^3] & \mathbb{E}[\epsilon_k^4] - 1 \end{pmatrix}.$$

Proof. For each $h_k \in H_k$, define the corresponding $\tilde{h}_k$ as in the statement of Lemma S5 and let $\tilde{H}_k$ collect all such $\tilde{h}_k$ formed with h_k ranging over H_k.[S12] By the definition of $\tilde{\ell}_\theta$ in equation (S8) and Theorem 4.11 in Rudin (1987) it suffices to show that each such component is (a) in $(\tilde{H}_0 + \cdots + \tilde{H}_K)^\perp$ and (b) $\dot{\ell}_{\theta,x} - \tilde{\ell}_{\theta,x} \in \mathrm{cl}(\tilde{H}_0 + \cdots + \tilde{H}_K)$, the form of which is given in Lemma S12.

Case 1: $x = \alpha$ or $x = \sigma$. For (a) note that if $j \neq k$, then

$$\begin{aligned}
\mathbb{E}\left[\zeta^x_{l,k,j}(\alpha,\sigma,X)\phi_k(\epsilon_k)\epsilon_j h_0(\tilde{X})\right] &= \mathbb{E}\left[\zeta^x_{l,k,j}(\alpha,\sigma,X)\phi_k(\epsilon_k)h_0(\tilde{X})\right]\mathbb{E}[\epsilon_j] = 0 \\
\mathbb{E}\left[\zeta^x_{l,k,j}(\alpha,\sigma,X)\phi_k(\epsilon_k)\epsilon_j h_m(\epsilon_m)\right] &= \mathbb{E}\left[\zeta^x_{l,k,j}(\alpha,\sigma,X)\right]\mathbb{E}\left[\phi_k(\epsilon_k)\epsilon_j h_m(\epsilon_m)\right] = 0
\end{aligned}$$

where the last equality follows from independence and the fact that m must differ from one of k, j. Additionally, writing $\tilde{\zeta}^x_{l,k,j}(X) := \zeta^x_{l,k,j}(\alpha,\sigma,X) - \mathbb{E}[\zeta^x_{l,k,j}(\alpha,\sigma,X)]$ and $\bar{\zeta}^x_{l,k,j} :=$

[S12] That is, for each $h_0 \in H_0$ define $\tilde{h}_0 : \mathcal{W} \to \mathbb{R}$ acccording to $\tilde{h}_0(W) := h_0(\tilde{X})$ and let $\tilde{H}_0$ collect the $\tilde{h}_0$ functions so formed. Similarly, for each $h_k \in H_k$ ($k = 1,\ldots,K$), define $\tilde{h}_k : \mathcal{W} \to \mathbb{R}$ according to $\tilde{h}_k(W) := h_k(e_k' A(\alpha,\sigma,X)V_\theta)$ and let let $\tilde{H}_k$ collect the $\tilde{h}_k$ functions so formed.

$\mathbb{E}[\zeta^x_{l,k,j}(\alpha,\sigma,X)]$, by independence and our moment assumptions (i.e. Assumption S3)

$$\begin{aligned}
&\mathbb{E}\left[\left(\tilde{\zeta}^x_{l,k,j}(X)[\phi_k(\epsilon_k)\epsilon_k+1]+\bar{\zeta}^x_{l,k,j}[\tau_{k,1}\epsilon_k+\tau_{k,2}\kappa(\epsilon_k)]\right)h_0(\tilde{X})\right]\\
&=\mathbb{E}\left[\tilde{\zeta}^x_{l,k,j}(X)h_0(\tilde{X})\right]\mathbb{E}\left[\phi_k(\epsilon_k)\epsilon_k+1\right]+\bar{\zeta}^x_{l,k,j}\mathbb{E}\left[\tau_{k,1}\epsilon_k+\tau_{k,2}\kappa(\epsilon_k)\right]\mathbb{E}[h_0(\tilde{X})]\\
&=0,
\end{aligned}$$

and again using independence and the definition of H_k,

$$\begin{aligned}
&\mathbb{E}\left[\left(\tilde{\zeta}^x_{l,k,j}(X)[\phi_k(\epsilon_k)\epsilon_k+1]+\bar{\zeta}^x_{l,k,j}[\tau_{k,1}\epsilon_k+\tau_{k,2}\kappa(\epsilon_k)]\right)h_j(\epsilon_j)\right]\\
&=\mathbb{E}\left[\tilde{\zeta}^x_{l,k,j}(X)\right]\mathbb{E}\left[(\phi_k(\epsilon_k)\epsilon_k+1)h_j(\epsilon_j)\right]+\bar{\zeta}^x_{l,k,j}\mathbb{E}\left[\left(\tau_{k,1}\epsilon_k+\tau_{k,2}\kappa(\epsilon_k)\right)h_j(\epsilon_j)\right]\\
&=0.
\end{aligned}$$

Since $\epsilon_k=e_k'A(\alpha,\sigma,X)V_\theta$, these observations and the form of $\tilde{\ell}_{\theta,x}$ establish (a). For (b), it suffices to show that

$$f_k(\epsilon_k)\coloneqq\phi_k(\epsilon_k)\epsilon_k+1-\tau_{k,1}\epsilon_k-\tau_{k,2}\kappa(\epsilon_k)\in H_k.$$

That $\mathbb{E}[f_k(\epsilon_k)]=0$ and $\mathbb{E}[f_k(\epsilon_k)^2]<\infty$ follows immediately from Assumption S3. That additionally $\mathbb{E}[f_k(\epsilon_k)\epsilon_k]=\mathbb{E}[f_k(\epsilon_k)\kappa(\epsilon_k)]=0$ is ensured by the choice of τ_k.

Case 2: $x=b$. For (a) let $m(X)\coloneqq A(\alpha,\sigma,X)D_{b,l}(b,X)$ and $\mu=\mathbb{E}[m(X)]$. Then,

$$\begin{aligned}
\mathbb{E}[\phi_k(\epsilon_k)e_k'(m(X)-\mu)h_0(\tilde{X})]&=\mathbb{E}[\phi_k(\epsilon_k)]\mathbb{E}[e_k'(m(X)-\mu)h_0(\tilde{X})]=0\\
\mathbb{E}[\phi_k(\epsilon_k)e_k'(m(X)-\mu)h_j(\epsilon_j)]&=\mathbb{E}[\phi_k(\epsilon_k)h_j(\epsilon_j)]\mathbb{E}[e_k'(m(X)-\mu)]=0\\
\mathbb{E}[e_k'\mu\left(\varsigma_{k,1}\epsilon_k+\varsigma_{k,2}\kappa(\epsilon_k)\right)h_0(\tilde{X})]&=e_k'\mu\mathbb{E}[\varsigma_{k,1}\epsilon_k+\varsigma_{k,2}\kappa(\epsilon_k)]\mathbb{E}[h_0(\tilde{X})]=0;
\end{aligned}$$

for $k\neq j$ by independence

$$\mathbb{E}[e_k'\mu\left(\varsigma_{k,1}\epsilon_k+\varsigma_{k,2}\kappa(\epsilon_k)\right)h_j(\epsilon_j)]=e_k'\mu\mathbb{E}[\varsigma_{k,1}\epsilon_k+\varsigma_{k,2}\kappa(\epsilon_k)]\mathbb{E}[h_j(\epsilon_j)]=0$$

whilst for $k=j$, the definition of H_k ensures that

$$\mathbb{E}[e_k'\mu\left(\varsigma_{k,1}\epsilon_k+\varsigma_{k,2}\kappa(\epsilon_k)\right)h_k(\epsilon_k)]=e_k'\mu\mathbb{E}[\varsigma_{k,1}\epsilon_kh_k(\epsilon_k)+\varsigma_{k,2}\kappa(\epsilon_k)h_k(\epsilon_k)]=0.$$

Since $\epsilon_k=e_k'A(\alpha,\sigma,X)V_\theta$, these observations and the form of $\tilde{\ell}_{\theta,b}$ establish (a). For (b) it suffices to show that

$$q_k(\epsilon_k) \coloneqq (\phi_k(\epsilon_k) + \varsigma_{k,1}\epsilon_k + \varsigma_{k,2}\kappa(\epsilon_k))\,(-e_k'\mu) \in H_k.$$

That $\mathbb{E}[q_k(\epsilon_k)] = 0$ and $\mathbb{E}[q_k(\epsilon_k)^2] < \infty$ follows immediately from Assumption S3. That additionally $\mathbb{E}[q_k(\epsilon_k)\epsilon_k] = \mathbb{E}[q_k(\epsilon_k)\kappa(\epsilon_k)] = 0$ is ensured by the choice of ς_k. □

S2.2 Log density score estimation

We work with a high level condition analagous to Assumption S1, adapted to the more general setting of equation (S22).

Assumption S4. *Let ν_n be as in Assumption 3. We have estimators $\hat{\phi}_{k,n,\gamma}$ such that for (a) any sequence with elements $\theta_n = (\alpha_0, \beta_n, \eta) \in \Theta$ where $(\beta_n)_{n\in\mathbb{N}}$ is a deterministic sequence with $\sqrt{n}\|\beta_n - \beta\| = O(1)$ and (b) any array $(Z_{n,i})_{n\in\mathbb{N}, i\leq n}$ with i.i.d. rows and such that $\mathbb{E}Z_{n,i} = 0$, $\sup_{n\in\mathbb{N}} \mathbb{E}Z_{n,i}^2 < \infty$ and $Z_{n,i} \perp\!\!\!\perp \epsilon_{i,k}$ for each n, i, and k,*

$$\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n,i}V_{\theta_n,i}) - \phi_k(A_{k,\gamma_n,i}V_{\theta_n,i})\right] Z_{n,i} = o_{P^n_{\theta_n}}(n^{-1/2}), \tag{S27}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\left[\hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n,i}V_{\theta_n,i}) - \phi_k(A_{k,\gamma_n,i}V_{\theta_n,i})\right] Z_{n,i}\right)^2 = o_{P^n_{\theta_n}}(\nu_n). \tag{S28}$$

where $A_{k,\gamma_n,i} \coloneqq e_k' A(\alpha_0, \sigma_n, X_i)$, $V_{\theta_n,i} \coloneqq Y_i - B(b_n, X_i)$.

We additionally impose the following condition, which is necessary in this more general setup, due to the term

$$\sum_{k=1}^{K}\left(\zeta^x_{l,k,k}(\alpha,\sigma,X) - \mathbb{E}\left[\zeta^x_{l,k,k}(\alpha,\sigma,X)\right]\right)\left[\phi_k(e_k'A(\alpha,\sigma,X)V_\theta)e_k'A(\alpha,\sigma,X)V_\theta + 1\right],$$

which appears in $\tilde{\ell}_{\theta,x}$ for $x \in \{\alpha, \sigma\}$ when $A(\alpha,\sigma,X)$ depends on X.[S13]

Assumption S5. *In the context of Assumption S4, additionally*

$$\frac{1}{n}\sum_{i=1}^{n}\left(\left[\hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n,i}V_{\theta_n,i}) - \phi_k(A_{k,\gamma_n,i}V_{\theta_n,i})\right] A_{k,\gamma_n,i}V_{\theta_n,i}\right)^2 = o_{P^n_{\theta_n}}(\nu_n). \tag{S29}$$

Lemmas S8 and S9 below demonstrate that the estimator defined in (11) satisfies the high-level conditions in Assumptions S4 and S5 provided Assumption S3 holds along with

[S13] Compare the forms of the effective scores given in Lemmas S3 and S7.

Assumption 3 and some additional conditions given in the statement of Lemma S9. The proofs of these Lemmas are given in Section S6 below.

Lemma S8. *Suppose Assumptions S3 and 3 hold. Then, $\hat{\phi}_{k,n,\gamma}$ as defined in (11) satisfies Assumption S4.*

Lemma S9. *Suppose Assumptions S3 and 3 hold. Additionally suppose that for some $\mathsf{M}_{k,n} \geq \max\{|\Xi^L_{k,n}|, |\Xi^U_{k,n}|\}$,*

1. $\delta_{k,n}^{-3}\Delta_{k,n}\mathbb{E}\left[\epsilon_{i,k}^2 \mathbf{1}\{|\epsilon_{i,k}| > \mathsf{M}_{k,n}\}\right] = o(\nu_n)$*;*
2. $\mathbb{E}\left[\epsilon_{i,k}^4 \mathbf{1}\{|\epsilon_{i,k}| > \mathsf{M}_{k,n}\}\right] = o(\nu_n^2)$*;*
3. $\mathsf{M}_{k,n}^2 \|\phi_{k,n}^{(3)}\|_\infty^2 \delta_{k,n}^6 = o(\nu_n)$*.*

Then, $\hat{\phi}_{k,n,\gamma}$ as defined in (11) satisfies Assumption S5.

Remark 2. *For $\varrho < \rho$ where $\mathbb{E}|\epsilon_k|^\rho < \infty$, one has*

$$\mathbb{E}[|\epsilon_k|^\varrho \mathbf{1}\{|\epsilon_k| > \mathsf{M}_{k,n}\}] = \mathbb{E}\left[|\epsilon_k|^\rho |\epsilon_k|^{\varrho-\rho}\mathbf{1}\{|\epsilon_k| > \mathsf{M}_{k,n}\}\right] \leq \mathbb{E}|\epsilon_k|^\rho \mathsf{M}_{k,n}^{\varrho-\rho},$$

and thus the speed at which $\mathsf{M}_{k,n}$ is required to increase to satisfy conditions 1, 2 in Lemma S9 decreases with the number of finite moments of ϵ_k.

S2.3 The test and its asymptotic properties

Since $\tilde{\ell}_\theta$ has a slightly different form in the setting considered in this section (compared to that considered in the main text; compare Lemmas S3 and S7), we amend our estimator $\hat{\ell}_{n,\gamma}$ accordingly. First let $\hat{\tau}_{k,n,\gamma}$ and $\hat{\varsigma}_{k,n,\gamma}$ be given by

$$\hat{\tau}_{k,n,\gamma} = \hat{M}_{k,n,\gamma}^{-1}\begin{pmatrix} 0 \\ -2 \end{pmatrix}, \quad \hat{\varsigma}_{k,n,\gamma} = \hat{M}_{k,n,\gamma}^{-1}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \hat{M}_{k,n,\gamma} = \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} 1 & (A_{k,\gamma,i}V_{\gamma,i})^3 \\ (A_{k,\gamma,i}V_{\gamma,i})^3 & (A_{k,\gamma,i}V_{\gamma,i})^4 - 1 \end{pmatrix}.$$

The estimators for the components corresponding to α and σ are:

$$\begin{aligned}\hat{\ell}_{n,\gamma,\alpha,l}(W_i) &:= \sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K} \zeta^{\alpha}_{l,k,j,\gamma,i}\hat{\phi}_{k,n,\gamma}(A_{k,\gamma,i}V_{\gamma,i})A_{j,\gamma,i}V_{\gamma,i} \\ &\quad + \sum_{k=1}^{K}\left(\zeta^{\alpha}_{l,k,k,\gamma,i} - \bar{\zeta}^{\alpha}_{l,k,k,n,\gamma}\right)\left(\hat{\phi}_{k,n,\gamma}(A_{k,\gamma,i}V_{\gamma,i})A_{k,\gamma,i}V_{\gamma,i} + 1\right) \\ &\quad + \sum_{k=1}^{K}\bar{\zeta}^{\alpha}_{l,k,k,n,\gamma}\left(\hat{\tau}_{k,n,\gamma,1}A_{k,\gamma,i}V_{\gamma,i} + \hat{\tau}_{k,n,\gamma,2}\kappa(A_{k,\gamma,i}V_{\gamma,i})\right); \\ \hat{\ell}_{n,\gamma,\sigma,l}(W_i) &:= \sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K} \zeta^{\sigma}_{l,k,j,\gamma,i}\hat{\phi}_{k,n,\gamma}(A_{k,\gamma,i}V_{\gamma,i})A_{j,\gamma,i}V_{\gamma,i} \\ &\quad + \sum_{k=1}^{K}\left(\zeta^{\sigma}_{l,k,k,\gamma,i} - \bar{\zeta}^{\sigma}_{l,k,k,n,\gamma}\right)\left(\hat{\phi}_{k,n,\gamma}(A_{k,\gamma,i}V_{\gamma,i})A_{k,\gamma,i}V_{\gamma,i} + 1\right) \\ &\quad + \sum_{k=1}^{K}\bar{\zeta}^{\sigma}_{l,k,k,n,\gamma}\left(\hat{\tau}_{k,n,\gamma,1}A_{k,\gamma,i}V_{\gamma,i} + \hat{\tau}_{k,n,\gamma,2}\kappa(A_{k,\gamma,i}V_{\gamma,i})\right);\end{aligned}, \tag{S30}$$

with $\zeta^{\alpha}_{l,k,j,\gamma,i} := \zeta^{\alpha}_{l,k,j}(\alpha,\sigma,X_i)$, $\bar{\zeta}^{\alpha}_{l,k,j,n,\gamma} := \frac{1}{n}\sum_{i=1}^{n}\zeta^{\alpha}_{l,k,j,\gamma,i}$, $A_{k,\gamma,i} := e_k' A(\alpha,\sigma,X_i)$, $V_{\gamma,i} := V_{\theta,i} := Y_i - BX_i$, $\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$. For the components corresponding to b,

$$\begin{aligned}\hat{\ell}_{n,\gamma,b}(W_i) &:= -\sum_{k=1}^{K}\hat{\phi}_{k,n,\gamma}(A_{k,\gamma,i}V_{\gamma,i})\left(A_{k,\gamma,i}(X_i'\otimes I_K) - \frac{1}{n}\sum_{i=1}^{n}[A_{k,\gamma,i}(X_i'\otimes I_K)]\right) \\ &\quad + \sum_{k=1}^{K}\left(\frac{1}{n}\sum_{i=1}^{n}[A_{k,\gamma,i}(X_i'\otimes I_K)]\right)\left(\hat{\varsigma}_{k,n,\gamma,1}A_{k,\gamma,i}V_{\gamma,i} + \hat{\varsigma}_{k,n,\gamma,2}\kappa\left(A_{k,\gamma,i}V_{\gamma,i}\right)\right).\end{aligned} \tag{S31}$$

The estimator $\hat{I}_{n,\gamma}$ is given by

$$\hat{I}_{n,\gamma} := \frac{1}{n}\sum_{i=1}^{n}\hat{\ell}_{n,\gamma}(W_i)\hat{\ell}_{n,\gamma}(W_i)'.$$

Remark 3. *If $A(\alpha,\sigma,X) = A(\alpha,\sigma)$ and $B(b,X) = \mathrm{vec}^{-1}(b)X$ (as considered in the main text), the estimators given in* (S30) *and* (S31) *are numerically identical to those in* (9).

$\hat{S}_\gamma$ is then defined as in (14) and we have the following Theorem (cf. Theorem 1), the proof of which is analogous to that of Theorem 1.

Theorem S2. *Suppose that Assumptions S2, S3, S4 and S5 hold and suppose that β is an interior point of $\mathcal{B}$. Let $r_n = \mathrm{rank}(\hat{\mathcal{I}}^t_{\bar{\gamma}})$ and denote by c_n the $1-a$ quantile of the $\chi^2_{r_n}$*

distribution, for any $a \in (0,1)$. Then

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta_{0,n}} P_\theta(\hat{S}_{\bar{\gamma}} > c_n) \leq a,$$

with inequality only if $\operatorname{rank}(\tilde{\mathcal{I}}_{\theta_0}) = 0$ *where* $\theta_0 = (\alpha_0, \beta, \eta)$.

Proof. It suffices to show the conditions of Corollary 1 hold. There are 5 conditions which we verify in order: items 1, 2, 3 & equation (S15) of the statement of Theorem S1.

Condition 1: This follows verbatim as the demonstration of Condition 1 in the proof of Theorem 1 on replacing Lemma S1 with Lemma S5.

Condition 2: This follows by repeated addition and subtraction along with the convergence in probability and stochastic boundedness results of Lemma S15, the moment conditions in Assumption S3 and the local boundedness given by Assumption S2 Part 4.

Condition 3: This follows verbatim as the demonstration of Condition 3 in the proof of Theorem 1 on replacing "the local Lipschitz continuity of each $\beta \mapsto \zeta^x_{l,j,k}(\alpha,\sigma)$ and $\beta \mapsto A(\alpha,\sigma)$" with "the local Lipschitz continuity of each $\beta \mapsto \zeta^x_{l,j,k}(\alpha,\sigma,X)$ and $\beta \mapsto A(\alpha,\sigma,X)$" and removing the reference to Lemma S4.[S14]

Condition 4: This follows verbatim as the demonstration of Condition 4 in the proof of Theorem 1 on replacing Lemmas S1 and S2 with Lemmas S5 and S6.

□

S3 Supporting results for the main Theorems

The following supporting results apply to the model introduced in Section S2. The model considered in the main text is a special case of this model with $A(\alpha,\sigma,X) = A(\alpha,\sigma)$ and $B(b,X) = \operatorname{vec}^{-1}(b)X$, for which Assumptions 1, 2 and S1 imply S2, S3 and S4 respectively. In consequence the results in this section apply a fortiori to the case considered in the main text.

Lemma S10. *Suppose that Assumptions S2 and S3 hold and that (α,β) is an interior point of $\mathcal{A}\times\mathcal{B}$. Let $\varphi(g,h) = (g, \eta_0 h_0, \ldots, \eta_K h_K)$. Then*

1. *$t \mapsto \sqrt{p_{\theta+t\varphi(g,h)}(w)}$ is (pointwise) continuously differentiable in a neighbourhood $\mathcal{U} \subset [0,\infty)$ of zero.*[S15]

Moreover, if we define $q_{\theta,(g,h),u}(w) := \frac{\partial \log p_{\theta+t\varphi(g,h)}(w)}{\partial t}|_{t=u}$, then

[S14]Lemmas S8 and S9 are not necessary here since the high level Assumptions S4 and S5 are directly assumed.

[S15]If $\theta + t\varphi(g,h) \in \Theta$ for all $t \in [0,1]$, $\mathcal{U}$ may be taken to include $[0,1]$.

2. $\{q_{\theta,(g,h),u}(W)^2 : u \in \mathcal{V}\}$ *is uniformly* $P_{\theta+u\varphi(g,h)}$ *– integrable for some neighbourhood of zero* $\mathcal{V} \subset \mathcal{U}$.

Proof. For all sufficiently small t, $\theta + t\varphi(g,h) \in \Theta$; in such an interval, the continuous differentiability follows directly from Assumptions S2 and S3 along with the definition of H.

Under $P_{\theta+u\varphi(g,h)}$, $q_{\theta,(g,h),u}(W)$ has the same law as

$$
\begin{aligned}
Z_u &:= \frac{h_0(\tilde{X})}{1 + uh_0(\tilde{X})} + \sum_{k=1}^{K} \frac{h_k(\epsilon_k) + uh_k'(\epsilon_k)e_k'[\mathsf{D}_{1,u}V_{\theta+u\varphi(g,h)} + \mathsf{D}_{2,u}]}{1 + uh_k(\epsilon_k)} \\
&\quad + \operatorname{tr}(A(\alpha + ua, \sigma + us, X)^{-1}\mathsf{D}_{1,\mathsf{u}}) + \sum_{k=1}^{K} \phi_k(\epsilon_k)e_k'[\mathsf{D}_{1,u}V_{\theta+u\varphi(g,h)} + \mathsf{D}_{2,u}].
\end{aligned} \tag{S32}
$$

where

$$
\mathsf{D}_{1,u} := \sum_{l=1}^{L_\alpha} a_l D_{\alpha,l}(\alpha + ua, \sigma + us, X) + \sum_{l=1}^{L_\sigma} s_l D_{\sigma,l}(\alpha + ua, \sigma + us, X)
$$

and

$$
\mathsf{D}_{2,u} := A(\alpha + ua, \sigma + us, X) \sum_{l=1}^{L_b} \varrho_l D_{b,l}(b + u\varrho, X).
$$

The definition of H ensures that for all sufficiently small u (i.e. $u \in \mathcal{V}$), the denominators $1 + uh_0(\tilde{X})$ and $1 + uh_k(\epsilon_k)$ are bounded, as are $h_0(\tilde{X})$, $h_k(\epsilon_k)$ and $uh_k'(\epsilon_k)$. Assumption S2 ensures the same is true of $\mathsf{D}_{1,u}$, the trace term, $A(\alpha + ta, \sigma + ts, X)$ and its inverse. These bounds, along with the finite moments given by Assumption S3 allow the application of Jensen's and Hölder's inequalties to obtain that $\sup_{u\in\mathcal{V}} \mathbb{E}|Z_u|^{2+\delta/2} < \infty$, implying the claimed uniform integrability. $\square$

Lemma S11. *Suppose that Assumptions S2 and S3 hold and let* $\mathcal{V} = \mathbb{R}^L \times H$ *be equipped with the norm*[S16]

$$
\|(g,h)\| := \sqrt{\|g\|^2 + \sum_{k=0}^{K} \|\tilde{h}_k\|^2_{L_2(P_\theta)}}\ .
$$

Then, the functions $(g,h) \mapsto \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g'\dot{\ell}_\theta + \sum_{k=0}^{K} \tilde{h}_k\right]$ *(i.e. indexed by* n*) are equicontinuous on compacts in* $L_2(P_\theta)$ *and the functions* $(g,h) \mapsto P^n_{\theta_n(g,h)}$ *(i.e. indexed by* n*) are equicontinuous on compacts in the total variation metric.*

Proof. For any (g,h), $(g^\star,h^\star) \in \mathcal{V}$, by the fact the observations are i.i.d. and any $h \in H$ is

[S16] Each $\tilde{h}_k$ is as defined in the statement of Lemma S5.

mean zero, as is $\dot{\ell}_\theta$,

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[(g^\star - g)'\dot{\ell}_\theta + \sum_{k=0}^{K}(\tilde{h}_k^\star - \tilde{h}_k)\right]\right\|_{L_2(P_\theta^n)}^2 = \left\|(g^\star - g)'\dot{\ell}_\theta + \sum_{k=0}^{K}(\tilde{h}_k^\star - \tilde{h}_k)\right\|_{L_2(P_\theta)}^2 .$$

Therefore, left hand side in the display above can be made arbitrarily small, uniformly in n, by taking $\|(g^\star, h^\star) - (g, h)\|$ sufficiently small and hence the first claim holds. For the second claim we note that each $(g, h) \mapsto P^n_{\theta_n(g,h)}$ is continuous by the pointwise continuity of the densities and Scheffé's Lemma. Then, let $K \subset \mathcal{V} = \mathbb{R}^L \times H$ be compact. We will now show that for any convergent sequence $(g_n, h_n) \to (g, h)$ in K, $d_{TV}(P^n_{\theta_n(g_n,h_n)}, P^n_{\theta_n(g,h)}) \to 0$ as $n \to \infty$.[S17] For this, by Lemma S21 and the triangle inequality, it is sufficient to show that

$$\log\frac{p^n_{\theta_n(g_n,h_n)}}{p_{\theta_n(g_n,h)}} = o_{P^n_{\theta_n(g_n,h)}}(1), \qquad \log\frac{p^n_{\theta_n(g_n,h)}}{p_{\theta_n(g,h)}} = o_{P^n_{\theta_n(g,h)}}(1). \tag{S33}$$

For these we first note that since h_k is bounded,

$$\begin{aligned}\left\|\tilde{h}_{k,n} - \tilde{h}_k\right\|_{L_2(P^n_{\theta_n(g_n,h)})}^2 &= \int \left[h_{n,k}(x) - h_k(x)\right]^2 \eta_k(x)(1 + h_k(x)/\sqrt{n})\,\mathrm{d}x \\ &\le \|h_{n,k} - h_k\|_{L_2(P^n_\theta)} + \|h_{n,k} - h_k\|_{L_2(P^n_\theta)}\|h_k\|_{L_\infty(P^n_\theta)}/\sqrt{n}.\end{aligned} \tag{S34}$$

Next introduce the notation:[S18]

$$u_{k,n,i} \coloneqq \begin{cases} e_k' A(\theta_n(g_n,h),X)V_{\theta_n(g_n,h),i} = e_k' A(\theta_n(g_n,h_n),X)V_{\theta_n(g_n,h_n),i} & \text{if } k = 1,\dots,K \\ \tilde{X}_i & \text{if } k = 0\end{cases}.$$

Equation (S34) implies that $(\tilde{h}_{k,n})_{n\in\mathbb{N}}$ is uniformly square $P^n_{\theta_n(g_n,h)}$ integrable, and hence the Lindeberg condition holds for $h_{k,n}(u_{k,n,i})/\sqrt{n}$. In particular, under $P^n_{\theta_n(g_n,h)}$,

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

[S17]That this convergence holds for any convergent sequence in a compact subset K is equivalent to equicontinuity on K, given the continuity of $(g, h) \mapsto P^n_{\theta_n(g,h)}$ already noted.

[S18]$A(\theta, X) \coloneqq A(\alpha, \sigma, X)$.

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

$$\max_{1\le i\le n} \frac{|h_{k,n}(u_{k,n,i})|}{\sqrt{n}} \xrightarrow{P^n_{\theta_n(g_n,h)}} 0. \tag{S35}$$

Then, to prove the first claim in (S33) observe

$$\log \frac{p^n_{\theta_n(g_n,h_n)}}{p^n_{\theta_n(g_n,h)}} = \sum_{k=0}^{K}\sum_{t=1}^{n} \log(1 + h_{k,n}(u_{k,n,i})/\sqrt{n}) - \log(1 + h_k(u_{k,n,i})/\sqrt{n}),$$

hence it suffices to show that each

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

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

$$\begin{aligned} E_n &\coloneqq \left\{ \max_{1\le i\le n} |h_{k,n}(u_{k,n,i})|/\sqrt{n} \le \varepsilon \right\}; \\ F_n &\coloneqq \left\{ \max_{1\le i\le n} |h_k(u_{k,n,i})|/\sqrt{n} \le \varepsilon \right\}. \end{aligned}$$

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 equation S35. 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

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

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

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

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

$$\sum_{i=1}^{n}\left|R\left(\frac{h_{k,n}(u_{k,n,i})}{\sqrt{n}}\right)\right| \leq \sum_{i=1}^{n}\left|\frac{h_{k,n}(u_{k,n,i})}{\sqrt{n}}\right|\left|\frac{h_{k,n}(u_{k,n,i})^2}{n}\right| \leq \max_{1\leq i\leq n}\frac{|h_{k,n}(u_{k,n,i})|}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n}h_{k,n}(u_{k,n,i})^2.$$

By Markov's inequality and equations (S34), (S35), 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,

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

and it remains to show that $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}h_{k,n}(u_{k,n,i})-h_k(u_{k,n,i})$ and $\frac{1}{n}\sum_{i=1}^{n}[h_{k,n}(u_{k,n,i})^2-h_k(u_{k,n,i})^2]$ also converge to zero in probability under $P^n_{\theta_n(g_n,h)}$. The second of these follows directly from (S34), Markov's inequality and the reverse triangle inequality since

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

For the remaining term, we start by noting that

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

so

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

Thus it suffices to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}h_{k,n}(u_{k,n,i})-h_k(u_{k,n,i})\xrightarrow{P^n_{\theta_n(g_n,h)}}0,$$

for $h_{k,n}(u_{k,n,i}):=h_{k,n}(u_{k,n,i})-\mathbb{E}\left[h_{k,n}(u_{k,n,i})\right]$ and $h_k(u_{k,n,i}):=h_{k,n}(u_{k,n,i})-\mathbb{E}\left[h_k(u_{k,n,i})\right]$. By the reverse triangle inequality and (S34),

$$\mathbb{E}\left[(h_{k,n}(u_{k,n,i})-h_k(u_{k,n,i}))^2\right]\to 0,\quad\text{uniformly in } i.$$

Using this, the independence of the W_i and Markov's inequality:

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

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

For the second condition in (S33), by Lemma S6 part 3 $P^n_{\theta_n(g,h)} \triangleleft\triangleright P^n_\theta$.[S19] Hence it suffices to show that $\log \frac{p^n_{\theta_n(g_n,h)}}{p_{\theta_n(g,h)}} = o_{P^n_\theta}(1)$. We first show that,

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

where the expectations are taken under P^n_θ. Here we may proceed analogously to Lemma S5. In particular, by an argument analogous to that showing condition 1 in Lemma S10, $g \mapsto \sqrt{p_{\theta_n(g,0)}}$ is continuously differentiable, whilst an argument analogous to that showing condition 2 in Lemma S10 yields that $\{q_{\theta,(g,0)}(W)^2 : g \in \mathcal{U}\}$ is uniformly $P_{\theta+\varphi(g,0)}$ – integrable for some neighbourhood $\mathcal{U} \subset \mathbb{R}^L$ of 0. Application of Lemma 7.6 and Theorem 7.2 in van der Vaart (1998) then yields the two likelihood expansions in the display above. To complete the proof set

$$\tilde{u}_{k,n,i} := e_k' A(\theta_n(g_n,h),X)V_{\theta_n(g_n,h),i}, \qquad u_{k,n,i} := e_k' A(\theta_n(g,h),X)V_{\theta_n(g,h),i},$$

and observe that

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

where the bracketed term is $o_{P^n_\theta}(1)$ by the preceding argument. Hence it suffices to show that an arbitrary k-th element of the outer sum on the right hand side is also $o_{P^n_\theta}(1)$. Let

[S19] The present Lemma is used in the proof of Lemma S6, but is used only to handle the case where (g_n, h_n) are not constant in n, which is the relevant case here.

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

$$E_n := \left\{ \max_{1\le i\le n} |h_k(\tilde{u}_{k,n,i})|/\sqrt{n} \le \varepsilon \right\}, \quad F_n := \left\{ \max_{1\le i\le n} |h_k(u_{k,n,i})|/\sqrt{n} \le \varepsilon \right\}.$$

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

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

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 $A(\theta, \sigma, X)$ is bounded on a neighbourhood of (θ, σ) (by Assumption S2). Combine this with the boundedness of h_k' and the mean value theorem to conclude that

$$|h_k(\tilde{u}_{k,n,i}) - h_k(u_{k,n,i})| \lesssim n^{-1/2}\|g_n - g\| \left[\|\epsilon_i\| + \sqrt{\sum_{l=1}^{L_b} D_{b,l}(b + \varrho_{l,n}, X_i)^2}\right],$$

for some $\varrho_{l,n}$ with $\|\varrho_{l,n}\| \le \|g_n - g\|$. Since h_k is bounded,

$$|h_k(\tilde{u}_{k,n,i})^2 - h_k(u_{k,n,i})^2| \lesssim n^{-1/2}\|g_n - g\| \left[\|\epsilon_i\| + \sqrt{\sum_{l=1}^{L_b} D_{b,l}(b + \varrho_{l,n}, X_i)^2}\right].$$

Therefore, using the moment bounds in Assumption S3 parts 1 and 4

$$\begin{aligned}
&\sum_{i=1}^n \left| \frac{h_k(\tilde{u}_{k,n,i}) - h_k(u_{k,n,i})}{\sqrt{n}} - \frac{1}{2}\frac{h_k(\tilde{u}_{k,n,i})^2 - h_k(u_{k,n,i})^2}{n}\right| \\
&\qquad \lesssim \|g_n - g\| \left(1 + \frac{1}{\sqrt{n}}\right) \frac{1}{n}\sum_{i=1}^n \left[\|\epsilon_i\| + \sqrt{\sum_{l=1}^{L_b} D_{b,l}(b + \varrho_{l,n}, X_i)^2}\right] = o_{P_\theta^n}(1).
\end{aligned}$$

This completes the demonstration of (S33) and hence the proof. ◻

Lemma S12. *Suppose that Assumptions S2 and S3 hold. Then,*

1. $\operatorname{cl} H_0$ *is the space of functions* $h_0 : \mathbb{R}^{d-1} \to \mathbb{R}$ *such that* $\mathbb{E}h_0(\tilde{X}_i)^2 < \infty$, $\mathbb{E}h_0(\tilde{X}) = 0$*;*

2. *For* $k = 1, \ldots, K$, $\operatorname{cl} H_k$ *is the space of functions* $h_k : \mathbb{R} \to \mathbb{R}$ *such that* $\mathbb{E}h_k(\epsilon_k)^2 < \infty$,

$$\mathbb{E}[h_k(\epsilon_k)] = \mathbb{E}[\epsilon_k h_k(\epsilon_k)] = \mathbb{E}[\kappa(\epsilon_k) h_k(\epsilon_{i,k})] = 0.$$

Additionally, define $H_0^\star$ *as the space of functions* $\tilde{h}_0(W) := h_0(\tilde{X})$ *for* $h_0 \in \operatorname{cl} \tilde{H}_0$ *and* $H_k^\star$ *as the space of functions* $\tilde{h}_k(W) := h_k(e_k' A(\alpha, \sigma, X) V_\theta)$ *for* $h_k \in \operatorname{cl} \tilde{H}_k$ $(k = 1, \ldots, K)$*. Then*

$$H^\star := H_0^\star + \cdots + H_K^\star \subset L_2(P_\theta) \quad \text{and} \quad H^\star = \operatorname{cl}(\tilde{H}_0 + \cdots + \tilde{H}_K).$$

Proof. For 1 & 2 let H_k^* denote the set of functions described in the statement (for $k = 0, \ldots, K$). Clearly any convergent sequence in this space has a limit also in this space and hence H_k^* is closed. For any $h_k \in \tilde{H}_k^*$ there is a sequence $(h_{k,n})_{n\in\mathbb{N}}$ such that each $h_{k,n} \in H_k$ and $h_{k,n} \to h_k$ in squared mean (e.g. Newey, 1991, Lemma C.7) and hence $\operatorname{cl} H_k = H_k^*$.[S20]

For the second part, the first claim follows since $e_k' A(\alpha, \sigma, X) V_\theta$ has the same law as ϵ_k under P_θ and hence each $P_\theta[\tilde{h}_k(W)^2] < \infty$. For the second claim, as $\tilde{X}, \epsilon_1, \ldots, \epsilon_K$ are independent, $\tilde{H}_0^\star, \ldots, \tilde{H}_K^\star$ are pairwise orthogonal. As the (finite) sum of closed pairwise orthogonal subspaces is closed (e.g. Conway, 1985, p. 39) we have that $\operatorname{cl}(\tilde{H}_0 + \ldots + \tilde{H}_K) \subset H^\star$. For the reverse inclusion let $\tilde{h} = \sum_{k=0}^K \tilde{h}_k \in H^\star$. By the definition of $H^\star$ there are $\tilde{h}_{0,n}(W) := h_{0,n}(\tilde{X})$ such that $\tilde{h}_{0,n} \in \tilde{H}_0$ and $P_\theta \left[\tilde{h}_{0,n}(W) - \tilde{h}_0(W)\right]^2 \to 0$ and $\tilde{h}_{k,n}(W) := h_{k,n}(e_k' A(\alpha, \sigma, X) V_\theta)$ such that $\tilde{h}_{k,n} \in H_k$ and $P_\theta \left[\tilde{h}_{0,k}(W) - \tilde{h}_k(W)\right]^2 \to 0$. Hence $\tilde{h}_n := \sum_{k=0}^K \tilde{h}_{k,n} \in \tilde{H}_0 + \ldots + \tilde{H}_K$ and converges to $\tilde{h}$, implying that $\tilde{h} \in \operatorname{cl}(\tilde{H}_0 + \cdots + \tilde{H}_K)$. □

Lemma S13. *Suppose that Assumptions S2 and S3 hold. Then* $\sup_{n\in\mathbb{N}} P_{\tilde{\theta}_n} \|\tilde{\ell}_{\tilde{\theta}_n}\|^{2+\delta/2} < \infty$ *and hence* $(\|\tilde{\ell}_{\tilde{\theta}_n}\|^2)_{n\in\mathbb{N}}$ *is uniformly* $P_{\tilde{\theta}_n}$*–integrable.*

Proof. As each component of $\tilde{\ell}_{\tilde{\theta}_n}$ lies in $L_2(P_{\tilde{\theta}_n})$ by its definition as an orthogonal projection, it suffices to show that $\limsup_{n\in\mathbb{N}} P_{\tilde{\theta}_n} \left[\|\tilde{\ell}_{\tilde{\theta}_n}\|^{2+\delta/2}\right] < \infty$. Let $d_n := (b_n, s_n) := \sqrt{n}(\beta_n - \beta)$, with $b_n \in \mathbb{R}^{L_b}$ and $s_n \in \mathbb{R}^{L_\sigma}$, so that $\tilde{\theta}_n = \theta_n(g_n, 0)$ with $g_n = (0, b_n, s_n)$. Then, under $P_{\tilde{\theta}_n}$, $e_k' A(\alpha, \sigma + s_n/\sqrt{n}, X) V_{\tilde{\theta}_n}$ has the same law as ϵ_k. This, along with the observations that $\mathbb{E}[|\phi_k(\epsilon_k)|^{4+\delta}] < \infty$, $\mathbb{E}|\epsilon_k|^{4+\delta}$ (both for $k = 1, \ldots, K$), $\mathbb{E}[\|D_{b,l}(b, X)\|^{4+\delta}] < \infty$ and the local boundedness conditions in Assumption S2 part 4 allow the application of Jensen's and Hölder's inequalities to conclude that $\limsup_{n\in\mathbb{N}} P_{\tilde{\theta}_n} \left[\|\tilde{\ell}_{\tilde{\theta}_n}\|^{2+\delta/2}\right] < \infty$ as desired. □

[S20] The required non-singularity condition for $q(\epsilon_k) = (1, \epsilon_k, \kappa(\epsilon_k))'$ is satisfied under the condition $\mathbb{E}(\epsilon_k^4) - 1 > \mathbb{E}(\epsilon_k^3)^2$ imposed in Assumption S3.

Lemma S14. *Suppose that Assumptions S2 and S3 hold. Then,*

$$\lim_{n\to\infty}\int\left\|\tilde{\ell}_{\tilde{\theta}_n}\sqrt{p_{\tilde{\theta}_n}}-\tilde{\ell}_{\theta}\sqrt{p_{\theta}}\right\|^2 \mathrm{d}\lambda = 0.$$

Proof. Re-write the integral as

$$\int\left\|\tilde{\ell}_{\tilde{\theta}_n}\sqrt{p_{\tilde{\theta}_n}}-\tilde{\ell}_{\theta}\sqrt{p_{\theta}}\right\|^2 \mathrm{d}\lambda = \sum_{l=1}^{L}\int\left[\tilde{\ell}_{\tilde{\theta}_n,l}\sqrt{p_{\tilde{\theta}_n}}-\tilde{\ell}_{\theta,l}\sqrt{p_{\theta}}\right]^2 \mathrm{d}\lambda. \tag{S36}$$

It is evidently sufficient to show that each of the integrals in the sum on the rhs converges to zero. For this note that inspection of the forms of $\tilde{\ell}_\theta$ and p_θ reveals that $\tilde{\ell}_{\tilde{\theta}_n} \to \tilde{\ell}_\theta$ and $p_{\tilde{\theta}_n} \to p_\theta$ pointwise. Hence each $\tilde{\ell}_{\tilde{\theta}_n,l}\sqrt{p_{\tilde{\theta}_n}} \to \tilde{\ell}_{\theta,l}\sqrt{p_\theta}$ pointwise and, by Scheffé's Lemma, $P_{\tilde{\theta}_n} \xrightarrow{TV} P_\theta$. Combine this observation with Lemma S13 and Corollary 2.9 in Feinberg, Kasyanov and Zgurovsky (2016) to obtain $\lim_{n\to\infty}\int|\tilde{\ell}_{\tilde{\theta}_n,l}\sqrt{p_{\tilde{\theta}_n}}|^2\,\mathrm{d}\lambda = \int|\tilde{\ell}_{\theta,l}\sqrt{p_\theta}|^2 < \infty$. Apply Proposition 2.29 in van der Vaart (1998) to conclude. □

Lemma S15. *Suppose that Assumptions S2, S3 and S4 hold. Then, for each (k,j) with $k \neq j$, each l, each $x \in \{\alpha,\sigma\}$ and each $\varrho \in \{\tau,\varsigma\}$, the following terms are $o_{P^n_{\tilde{\theta}_n}}(1)$:*

1. $\bar{\zeta}^x_{l,k,k,n,\gamma_n} - P_{\tilde{\theta}_n}\left[\zeta^x_{l,k,k,\gamma_n,i}\right]$;
2. $\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\phi_k(A_{k,\gamma_n i}V_{\gamma_n,i}) - \hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n i}V_{\gamma_n,i})\right)\zeta^x_{l,k,j,i}A_{j,\gamma_n,i}V_{\gamma_n,i}$;
3. $\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\phi_k(A_{k,\gamma_n i}V_{\gamma_n,i}) - \hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n i}V_{\gamma_n,i})\right)A_{k,\gamma_n,i}V_{\gamma_n,i}\left(\zeta^x_{l,k,j,\gamma_n,i} - \bar{\zeta}^x_{l,k,j,n,\gamma_n}\right)$;
4. $\frac{1}{\sqrt{n}}\sum_{i=1}^n([\hat{\varrho}_{k,n,\gamma_n,1} - \varrho_{k,1}]A_{k,\gamma_n,i}V_{\gamma_n,i} + [\hat{\varrho}_{k,n,\gamma_n,2} - \varrho_{k,2}]\kappa(A_{k,\gamma_n,i}V_{\gamma_n,i}))$;
5. $\frac{1}{n}\sum_{i=1}^n[A_{k,\gamma_n,i}D_{b,l}(b_n,X_i)] - P_{\tilde{\theta}_n}[A_{k,\gamma_n,i}D_{b,l}(b_n,X_i)]$;
6. $\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\phi_k(A_{k,\gamma_n i}V_{\gamma_n,i}) - \hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n i}V_{\gamma_n,i})\right)\left([A_{k,\gamma_n,i}D_{b,l}(b_n,X_i)] - \frac{1}{n}\sum_{i=1}^n[A_{k,\gamma_n,i}D_{b,l}(b_n,X_i)]\right)$;

and the following terms are $O_{P^n_{\tilde{\theta}_n}}(1)$:

7. $\frac{1}{\sqrt{n}}\sum_{i=1}^n(\phi_k(A_{k,\gamma_n i}V_{\gamma_n,i})A_{k,\gamma_n i}V_{\gamma_n,i} + 1)$;
8. $\frac{1}{\sqrt{n}}\sum_{i=1}^n\varrho_{k,1}A_{k,\gamma_n,i}V_{\gamma_n,i} + \varrho_{k,2}\kappa(A_{k,\gamma_n,i}V_{\gamma_n,i})$;
9. $\frac{1}{\sqrt{n}}\sum_{i=1}^n\phi_k\left(A_{k,\gamma_n i}V_{\gamma_n,i}\right)$.

Proof. Under $P_{\tilde{\theta}_n}$, $\tilde{X}$ is distributed according to the density η_0 whilst $A_{k,\gamma_n,i}V_{\gamma_n,i}$ has the same law as ϵ_k. We will use these facts without explicit reference in the rest of the proof.

1. The triangular array $(\zeta^x_{l,k,k,\gamma_n,i})_{n\in\mathbb{N},i=1,\dots,n}$ has i.i.d. rows and the variance of $\zeta^x_{l,k,k,\gamma_n,i}$ is bounded above uniformly in n by Assumption S2. The claim then follows from a WLLN for triangular arrays (e.g. Durrett, 2019, Theorem 2.2.6).

2. Let $Z_{n,i} := \zeta^x_{l,k,j,i} A_{j,\gamma_n,i} V_{\gamma_n,i}$. The triangular array $(Z_{n,i})_{n\in\mathbb{N},i=1,\dots,n}$ has i.i.d. rows, $Z_{n,i} \perp\!\!\!\perp \epsilon_{i,k}$, $Z_{n,i}$ is mean zero and the variance of $Z_{n,i}$ is bounded above uniformly in n by Assumptions S2 and S3. The claim then follows by Assumption S4.

3. By Cauchy – Schwarz one has

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\phi_k(A_{k,\gamma_n i}V_{\gamma_n,i}) - \hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n i}V_{\gamma_n,i})\right) A_{k,\gamma_n,i}V_{\gamma_n,i}\left(\zeta^x_{l,k,j,\gamma_n,i} - \bar{\zeta}^x_{l,k,j,n,\gamma_n}\right) \\
&\quad \le \left[\frac{1}{n}\sum_{i=1}^{n}\left(\phi_k(A_{k,\gamma_n i}V_{\gamma_n,i}) - \hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n i}V_{\gamma_n,i})\right)^2\left(\zeta^x_{l,k,j,\gamma_n,i} - \bar{\zeta}^x_{l,k,j,n,\gamma_n}\right)^2\right]^{1/2} \\
&\qquad \times \left[\frac{1}{n}\sum_{i=1}^{n}(A_{k,\gamma_n,i}V_{\gamma_n,i})^2\right]^{1/2}.
\end{aligned}
$$

Take $Z_{n,i} := \zeta^x_{l,k,j,\gamma_n,i} - \bar{\zeta}^x_{l,k,j,n,\gamma_n}$. The triangular array $(Z_{n,i})_{n\in\mathbb{N},i=1,\dots,n}$ has i.i.d. rows, $Z_{n,i} \perp\!\!\!\perp \epsilon_{i,k}$, $Z_{n,i}$ is mean zero and the variance of $Z_{n,i}$ is bounded above uniformly in n by Assumption S2. Therefore, the first factor on the right hand side is $o_{P^n_{\tilde{\theta}_n}}(1)$ by Assumption S4. The second right hand side factor is $O_{P^n_{\tilde{\theta}_n}}(1)$ by Assumption S3.

4. $\varrho_{k,n,\gamma_n} \xrightarrow{P^n_{\tilde{\theta}_n}} \varrho_k$ by Lemma S16. Assumption S3 and the central limit theorem imply that $\frac{1}{\sqrt{n}}\sum_{i=1}^n A_{k,\gamma_n,i}V_{\gamma_n,i}$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^n \kappa(A_{k,\gamma_n,i}V_{\gamma_n,i})$ are $O_{P_{\tilde{\theta}_n}}(1)$.

5. Let $U_{n,i} := \mathrm{vec}(A_{k,\gamma_n,i}D_{b,l}(b_n,X))$. Then for each component $U_{n,i,l}$, $(U_{n,i,l})_{n\in\mathbb{N},i=1,\dots,n}$ is a triangular array with i.i.d. rows and the variance of $U_{n,i,l}$ is bounded above uniformly in n by Assumptions S2 and S3. The claim then follows from a WLLN for triangular arrays (e.g. Durrett, 2019, Theorem 2.2.6).

6. Put $Z_{n,i} := [A_{k,\gamma_n,i}D_{b,l}(b_n,X)] - \frac{1}{n}\sum_{i=1}^n [A_{k,\gamma_n,i}D_{b,l}(b_n,X)]$. Then, the triangular array $(Z_{n,i})_{n\in\mathbb{N},i=1,\dots,n}$ has i.i.d. rows, $Z_{n,i} \perp\!\!\!\perp \epsilon_{i,k}$, $Z_{n,i}$ is mean zero and the variance of $Z_{n,i}$ is bounded above uniformly in n by Assumptions S2 and S3. The claim follows by Assumption S4.

Each of the remaining items follow from the central limit theorem given Assumption S3. $\square$

Lemma S16. *If Assumption S3 holds,* $\|\varrho_{k,n,\gamma_n} - \varrho_k\| = o_{P^n_{\tilde{\theta}_n}}(\nu_{n,p})$ *for* $\varrho \in \{\tau, \varsigma\}$.[S21]

Proof. Under $P_{\tilde{\theta}_n}$, $\hat{M}_{k,n,\gamma_n}$ has the same law as $M_{k,n} := \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} 1 & \epsilon_{i,k}^3 \\ \epsilon_{i,k}^3 & \epsilon_{i,k}^4 - 1 \end{pmatrix}$. Therefore, it suffices to show that $\|M_{k,n}^{-1}w - M_k^{-1}w\| = o_{P^n_{\tilde{\theta}_n}}(\nu_{n,p})$ for any fixed $w \in \mathbb{R}^2$. Since the map $M \mapsto M^{-1}$ is Lipschitz continuous at a positive definite matrix,

$$\|M_{k,n}^{-1}w - M_k^{-1}w\|_2 \leq \|w\|\|M_{k,n}^{-1} - M_k^{-1}\|_2 \lesssim \|M_{k,n} - M_k\|_2,$$

and thus it suffices to show that $\|M_{k,n} - M_k\|_2 = o_{P^n_{\tilde{\theta}_n}}(\nu_{n,p})$. If $\upsilon := \delta/4 \geq 1$, we have that by Theorem 2.5.11 in Durrett (2019)

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n [\epsilon_{i,k}^3 - \mathbb{E}(\epsilon_{i,k}^3)] &= o_{P^n_{\tilde{\theta}_n}}\left(n^{-1/2}\log(n)^{1/2+\rho}\right) \\
\frac{1}{n}\sum_{i=1}^n [\epsilon_{i,k}^4 - \mathbb{E}(\epsilon_{i,k}^4)] &= o_{P^n_{\tilde{\theta}_n}}\left(n^{-1/2}\log(n)^{1/2+\rho}\right)
\end{aligned}$$

for any $\rho > 0$, which implies that

$$\|M_{k,n} - M_k\|_2 \leq \|M_{k,n} - M_k\|_F = o_{P^n_{\tilde{\theta}_n}}\left(n^{-1/2}\log(n)^{1/2+\rho}\right).$$

If $0 < \upsilon < 1$, by Theorems 2.5.11 & 2.5.12 in Durrett (2019), for any $\rho > 0$,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n [(\epsilon_{i,k})^3 - \mathbb{E}(\epsilon_{i,k})^3] &= \begin{cases} o_{P^n_{\tilde{\theta}_n}}\left(n^{-1/2}\log(n)^{1/2+\rho}\right) & \text{if } \upsilon \in [1/2, 1) \\ o_{P^n_{\tilde{\theta}_n}}\left(n^{\frac{1-p}{p}}\right) & \text{if } \upsilon \in (0, 1/2) \end{cases}, \\
\frac{1}{n}\sum_{i=1}^n [(\epsilon_{i,k})^4 - \mathbb{E}(\epsilon_{i,k})^4] &= o_{P^n_{\tilde{\theta}_n}}\left(n^{\frac{1-p}{p}}\right).
\end{aligned}$$

which together imply that

$$\|M_{k,n} - M_k\|_2 \leq \|M_{k,n} - M_k\|_F = o_{P^n_{\tilde{\theta}_n}}\left(n^{\frac{1-p}{p}}\right).$$

□

Lemma S17. *Suppose that Assumptions S2, S3 and S4 hold. Then, for each* (k, j) *with* $k \neq j$*, each* l*, each* $x \in \{\alpha, \sigma\}$ *and each* $\varrho \in \{\tau, \varsigma\}$*, the following terms are* $o_{P^n_{\tilde{\theta}_n}}(\nu_n)$*:*

1. $\frac{1}{n}\sum_{i=1}^n \left(\phi_k(A_{k,\gamma_n,i}V_{\gamma_n,i}) - \hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n,i}V_{\gamma_n,i})\right)^2 \left(A_{j,\gamma_n,i}V_{\gamma_n,i}\zeta^x_{l,k,j,\gamma_n,i}\right)^2$;

[S21] $\nu_{n,p}$ is as defined in Assumption 3: $p := \min\{1 + \delta/4, 2\}$ and $\nu_{n,p} := \begin{cases} n^{(1-p)/p} & \text{for } p \in (1, 2) \\ n^{-1/2}\log(n)^{1/2+\rho} & \text{for } p = 2 \end{cases}$, for some $\rho > 0$.

2. $\left(P_{\tilde{\theta}_n}[\zeta^x_{l,k,k,\gamma_n,i}] - \bar{\zeta}^x_{l,k,k,n,\gamma_n}\right)^2$;

3. $\frac{1}{n}\sum_{i=1}^n \left(\phi_k(A_{k,\gamma_n,i}V_{\gamma_n,i})A_{k,\gamma_n,i}V_{\gamma_n,i} + 1\right)^2 \left(P_{\tilde{\theta}_n}[\zeta^x_{l,k,k,\gamma_n,i}] - \bar{\zeta}^x_{l,k,k,n,\gamma_n}\right)^2$;

4. $\frac{1}{n}\sum_{i=1}^n \left(\varrho_{k,1}A_{k,\gamma_n,i}V_{\gamma_n,i} + \varrho_{k,2}\kappa(A_{k,\gamma_n,i}V_{\gamma_n,i})\right)^2 \left(P_{\tilde{\theta}_n}[\zeta^x_{l,k,k,\gamma_n,i}] - \bar{\zeta}^x_{l,k,k,n,\gamma_n}\right)^2$;

5. $\frac{1}{n}\sum_{i=1}^n \left(\bar{\zeta}^x_{l,k,k,n,\gamma_n}\left([\hat{\varrho}_{k,n,\gamma_n,1} - \varrho_{k,1}]A_{k,\gamma_n,i}V_{\gamma_n,i} + [\hat{\varrho}_{k,n,\gamma_n,2} - \varrho_{k,2}]\kappa(A_{k,\gamma_n,i}V_{\gamma_n,i})\right)\right)^2$;

6. $\left(P_{\tilde{\theta}_n}\left[A_{k,\gamma_n,i}D_{b,l}(b_n, X_i)\right] - [\mathsf{ADbX}]_n\right)^2$;

7. $\frac{1}{n}\sum_{i=1}^n \left(\varrho_{k,1}A_{k,\gamma_n,i}V_{\gamma_n,i} + \varrho_{k,2}\kappa(A_{k,\gamma_n,i}V_{\gamma_n,i})\right)^2 \left(P_{\tilde{\theta}_n}\left[A_{k,\gamma_n,i}D_{b,l}(b_n, X_i)\right] - [\mathsf{ADbX}]_n\right)^2$;

8. $\frac{1}{n}\sum_{i=1}^n \left([\mathsf{ADbX}]_n\left([\hat{\varrho}_{k,n,\gamma_n,1} - \varrho_{k,1}]A_{k,\gamma_n,i}V_{\gamma_n,i} + [\hat{\varrho}_{k,n,\gamma_n,2} - \varrho_{k,2}]\kappa(A_{k,\gamma_n,i}V_{\gamma_n,i})\right)\right)^2$

9. $\frac{1}{n}\sum_{i=1}^n \left(\phi_k(A_{k,\gamma_n,i}V_{\gamma_n,i}) - \hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n,i}V_{\gamma_n,i})\right)^2 \left(A_{k,\gamma_n,i}D_{b,l}(b_n, X_i) - [\mathsf{ADbX}]_n\right)^2$;

10. $\frac{1}{n}\sum_{i=1}^n \left(\phi_k(A_{k,\gamma_n,i}V_{\gamma_n,i})\right)^2 \left(P_{\tilde{\theta}_n}\left[A_{k,\gamma_n,i}D_{b,l}(b_n, X_i)\right] - [\mathsf{ADbX}]_n\right)^2$,

where $[\mathsf{ADbX}]_n \coloneqq \frac{1}{n}\sum_{i=1}^n A_{k,\gamma_n,i}D_{b,l}(b_n, X_i)$.

Proof. Under $P_{\tilde{\theta}_n}$, $\tilde{X}$ is distributed according to the density η_0 whilst $A_{k,\gamma_n,i}V_{\gamma_n,i}$ has the same law as ϵ_k. We will use these facts without explicit reference in the rest of the proof.

1. Let $Z_{n,i} \coloneqq A_{j,\gamma_n,i}V_{\gamma_n,i}\zeta^x_{l,k,j,\gamma_n,i}$. This is independent of $\epsilon_{i,k}$, is mean-zero and has variance bounded above uniformly in n by Assumptions S2 and S3. The claim then follows by Assumption S4.

2. Let $Z_{n,i} \coloneqq \left(\zeta^x_{l,k,k,\gamma_n,i} - P_{\tilde{\theta}_n}[\zeta^x_{l,k,k,\gamma_n,i}]\right)$ and note that $\sup_{n\in\mathbb{N}} \mathbb{E}Z_{n,i}^{2+\varepsilon} < \infty$ for a $\varepsilon > 0$ (by Assumption S2). By the Lindeberg CLT one then has that $\sum_{i=1}^n Z_{n,i} = O_{P^n_{\tilde{\theta}_n}}(\sqrt{n})$ and hence $\left(P_{\tilde{\theta}_n}[\zeta^x_{l,k,k,\gamma_n,i}] - \bar{\zeta}^x_{l,k,k,n,\gamma_n}\right)^2 = o_{P^n_{\tilde{\theta}_n}}(\nu_n)$.

3. By Assumption S3, $\frac{1}{n}\sum_{i=1}^n \left(\phi_k(A_{k,\gamma_n,i}V_{\gamma_n,i})A_{k,\gamma_n,i}V_{\gamma_n,i} + 1\right)^2 = O_{P_{\tilde{\theta}_n}}(1)$. Use 2.

4. By Assumption S3, $\frac{1}{n}\sum_{i=1}^n \left(\varrho_{k,1}A_{k,\gamma_n,i}V_{\gamma_n,i} + \varrho_{k,2}\kappa(A_{k,\gamma_n,i}V_{\gamma_n,i})\right)^2 = O_{P_{\tilde{\theta}_n}}(1)$. Use 2.

5. By Assumption S2, $\bar{\zeta}^x_{l,k,k,n,\gamma_n}$ is bounded uniformly for all sufficiently large n. By Assumption S3, $\frac{1}{n}\sum_{i=1}^n (A_{k,\gamma_n,i}V_{\gamma_n,i})^2$ and $\frac{1}{n}\sum_{i=1}^n \kappa(A_{k,\gamma_n,i}V_{\gamma_n,i})^2$ are $O_{P^n_{\tilde{\theta}_n}}(1)$. Combine with Lemma S16.

6. Let $Z_{n,i} \coloneqq \left(A_{k,\gamma_n,i}D_{b,l}(b_n, X_i) - P_{\tilde{\theta}_n}\left[A_{k,\gamma_n,i}D_{b,l}(b_n, X_i)\right]\right)$ and note that $\sup_{n\in\mathbb{N}} \mathbb{E}Z_{n,i}^{2+\varepsilon} < \infty$ for a $\varepsilon > 0$ (by Assumptions S2 and S3). By the Lindeberg CLT one then has that $\sum_{i=1}^n Z_{n,i} = O_{P^n_{\tilde{\theta}_n}}(\sqrt{n})$ and hence $\left(P_{\tilde{\theta}_n}\left[A_{k,\gamma_n,i}D_{b_n,l}(b_n, X_i)\right] - [\mathsf{ADbX}]_n\right)^2 = o_{P^n_{\tilde{\theta}_n}}(\nu_n)$.

7. By Assumption S3, $\frac{1}{n}\sum_{i=1}^n \left(\varrho_{k,1}A_{k,\gamma_n,i}V_{\gamma_n,i} + \varrho_{k,2}\kappa(A_{k,\gamma_n,i}V_{\gamma_n,i})\right)^2 = O_{P_{\tilde{\theta}_n}}(1)$. Use 6.

8. Take $[\mathsf{ADbX}_n]$ out of the summation. By Assumption S3 and 6. this is $O_{P^n_{\tilde{\theta}_n}}(1)$. By Assumption S3, $\frac{1}{n}\sum_{i=1}^n (A_{k,\gamma_n,i}V_{\gamma_n,i})^2$ and $\frac{1}{n}\sum_{i=1}^n \kappa(A_{k,\gamma_n,i}V_{\gamma_n,i})^2$ are $O_{P^n_{\tilde{\theta}_n}}(1)$. Combine with Lemma S16.

9. For $Z_{n,i} := A_{k,\gamma_n,i}D_{b,l}(b_n, X_i) - [\mathsf{ADbX}]_n$, $Z_{n,i}$ is independent of $\epsilon_{i,k}$, mean-zero and has variance bounded uniformly in n by Assumptions S2 and S3. The claim follows from Assumption S4.

10. $\frac{1}{n}\sum_{i=1}^n \left(\phi_k(A_{k,\gamma_n,i}V_{\gamma_n,i})\right)^2 = O_{P^n_{\tilde{\theta}_n}}(1)$ by Assumption S3. Use 6.

□

Lemma S18. *Suppose that Assumptions S2 and S5 hold. Then, for each k, each l, each $x \in \{\alpha, \sigma\}$,*

$$\frac{1}{n}\sum_{i=1}^n \left(\phi_k(A_{k,\gamma_n,i}V_{\gamma_n,i}) - \hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n,i}V_{\gamma_n,i})\right)^2 \left(A_{k,\gamma_n,i}V_{\gamma_n,i}\left[\zeta^x_{l,k,k,\gamma_n,i} - \bar{\zeta}^x_{l,k,k,n,\gamma_n}\right]\right)^2 = o_{P^n_{\tilde{\theta}_n}}(\nu_n).$$

Proof. By Assumption S2, $\left[\zeta^x_{l,k,k,\gamma_n,i} - \bar{\zeta}^x_{l,k,k,n,\gamma_n}\right]^2$ is uniformly bounded for all large enough n. Hence it suffices that by Assumption S5,

$$\frac{1}{n}\sum_{i=1}^n \left(\phi_k(A_{k,\gamma_n,i}V_{\gamma_n,i}) - \hat{\phi}_{k,n,\gamma_n}(A_{k,\gamma_n,i}V_{\gamma_n,i})\right)^2 \left(A_{k,\gamma_n,i}V_{\gamma_n,i}\right)^2 = o_{P^n_{\tilde{\theta}_n}}(\nu_n).$$ □

S4 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 $\tilde{I}_{\theta,\alpha\alpha}$ becomes singular, which implies the singularity of $\tilde{\mathcal{I}}_\theta$ if $\tilde{I}_{\theta,\beta\beta}$ is non-singular (cf. Propositions 8.2.4 and 8.2.8 in Bernstein (2009)). Second, we provide an explicit example of a density which satisfies the first part of the Assumption 2 but not the second. Third we prove that if Assumption 2 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 S19. *Consider the LSEM model (3) and suppose that Assumptions 1 and 2 hold. Define the random vector Q in $\mathbb{R}^{K^2}$ as*

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

where the j-th element of Q_k for $j \in [K]$ is given by

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

Next define the matrix $\zeta \in \mathbb{R}^{K^2 \times L_\alpha}$ according to

$$\zeta = (\operatorname{vec}([D_{\alpha,1}(\alpha,\sigma)A(\alpha,\sigma)^{-1}]'), \ldots, \operatorname{vec}([D_{\alpha,L_\alpha}(\alpha,\sigma)A(\alpha,\sigma)^{-1}]')).$$

Then where $\tilde{\ell}_\theta$ is the effective score function as defined in lemma S3, the law of $\tilde{\ell}_{\theta,1}$ under P_θ is equal to that of $\zeta' Q$. Moreover,

(i) $\mathbb{E}QQ'$ 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 $[\mathbb{E}\phi_k^2(\epsilon_k)][\mathbb{E}\phi_j^2(\epsilon_j)] \neq 1$.

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

(iii) If $\operatorname{rank}(\zeta) < L_\alpha$ then $\tilde{I}_{\theta,\alpha\alpha}$ is singular.

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

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

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

Proof. For (i), let j,k,m,i all be in $[K]$. We will consider the entries of the matrix $\mathbb{E}QQ'$, which are of the form $\mathbb{E}[Q_{k,j}Q_{m,i}]$. In particular, the s,t-th element of the matrix is given by the form $\mathbb{E}[Q_{k,j}Q_{m,i}]$ where $(k-1)K + j = s$ and $(m-1)K + i = t$. If $k = j = m = i$ we have $s = t$ and $\mathbb{E}[Q_{k,j}Q_{m,i}] = \mathbb{E}[\tau_{k,1}\epsilon_k + \tau_{k,2}\kappa(\epsilon_k)]^2$. The other diagonal entries occur when $k = m \neq j = i$, and have the form $\mathbb{E}[Q_{k,j}Q_{m,i}] = \mathbb{E}[\phi_k^2(\epsilon_k)]$. 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 $\mathbb{E}[Q_{k,j}Q_{m,i}] = \mathbb{E}[\phi_k(\epsilon_k)\epsilon_k]\mathbb{E}[\phi_k(\epsilon_m)\epsilon_m] = 1$ by assumption 2.

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 $\mathbb{E}\phi_k^2(\epsilon_k)$ 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 $\mathbb{E}\phi_k^2(\epsilon_k)$ 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 $\mathbb{E}\phi_j^2(\epsilon_j)$ 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 $[\mathbb{E}\phi_k^2(\epsilon_k)][\mathbb{E}\phi_j^2(\epsilon_j)] \neq 1$.

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

For the remaining parts note first that

$$\tilde{I}_{\theta,\alpha\alpha} = \mathbb{E}\tilde{\ell}_{\theta,1}\tilde{\ell}_{\theta,1}' = \zeta'\left[\mathbb{E}QQ'\right]\zeta,$$

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

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

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

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

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

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

which is clearly singular with rank 3. We have

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

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 $\mathbb{E}\phi_l^2(\epsilon_l) = 1$. $E_\theta QQ'$ is singular by (i) and hence by (iv) $\tilde{I}_{\theta,\alpha\alpha}$ is singular. $\square$

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

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 $r = 0$.

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 1 of assumption 2 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. □

Lemma S21. *Suppose that* P_n *and* Q_n *are probability measures (with each pair* (P_n, Q_n) *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.*[S22] *If*

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

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

Proof. By the continuous mapping theorem

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

[S22] l_n may be defined arbitrarily when $p_n = 0$.

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 by Slutsky's Theorem

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

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. $\square$

S5 A consistent estimator of the Moore – Penrose psuedoinverse

As is well known, the Moore – Penrose psuedoinverse of a matrix is not a continuous function on the space of positive semi-definite matrices (see e.g. Ben-Israel and Greville, 2003, Section 6.6). In consequence, if one has a consistent estimator $\check{M}_n$ of some matrix M, it need not follow that $\check{M}^\dagger$ is consistent for $M^\dagger$. A necessary and sufficient condition for this convergence in probability to occur is that $\operatorname{rank}(\check{M}_n) = \operatorname{rank}(M)$ with probability approaching one as $n \to \infty$ (Andrews, 1987, Theorem 2).

Here we provide a simple construction, based on the knowledge of the speed of convergence of $\check{M}_n$ to M, which results in an estimator $\hat{M}_n$ which is consistent for M and satisfies $\operatorname{rank} \hat{M}_n = \operatorname{rank} M$ with probability approaching one as $n \to \infty$ and, in consequence, $\hat{M}_n^\dagger$ is consistent for $M^\dagger$.

The construction proposed here is very similar to a special case of that considered by Dufour and Valéry (2016). We provide a direct proof for this construction rather than relying on Proposition 9.1 in Dufour and Valéry (2016) as the latter would require the introduction of an additional rate term (b_n in their notation) which satisfies a given condition (their Assumption 2.2). For our purposes we need only a single rate term (essentially the equivalent of c_n in their notation) and thus there are fewer conditions to verify.

In particular, suppose that the sequence of (random) positive semi-definite (symmetric) matrices $(\check{M}_n)_{n\in\mathbb{N}}$ (of fixed dimension $L \times L$) satisfy

$$P_n\left(\|\check{M}_n - M_n\|_2 < \nu_n\right) \to 1, \tag{S37}$$

for a sequence $(P_n)_{n\in\mathbb{N}}$ of probability measures, a known non-negative sequence $\nu_n \to 0$ and a sequence of deterministic matrices $M_n \to M$ with $\operatorname{rank}(M_n) = \operatorname{rank}(M)$ for all sufficiently large n.[S23] Let $\check{M}_n = \check{U}_n \check{\Lambda}_n \check{U}_n'$ be the corresponding eigendecompositions and define

$$\hat{M}_n := \check{U}_n \Lambda_n(\nu_n) \check{U}_n' \ , \tag{S38}$$

where $\Lambda_n(\nu_n)$ is a diagonal matrix with the ν_n-truncated eigenvalues of $\check{M}_n$ on the main diagonal and $\check{U}_n$ is the matrix of corresponding orthonormal eigenvectors. That is, if $(\check{\lambda}_{n,i})_{i=1}^L$ denote the non-increasing eigenvalues of $\check{M}_n$, then the (i,i)-th element of $\Lambda_n(\nu_n)$ is $\check{\lambda}_{n,i}\mathbf{1}(\check{\lambda}_{n,i} \geq \nu_n)$.

Proposition S1. *If* (S37) *holds,* $M_n \to M$ *and for all* n *greater than some* $N \in \mathbb{N}$ $\operatorname{rank}(M_n) = \operatorname{rank}(M)$, *then* $\hat{M}_n \xrightarrow{P_n} M$ *and*

$$P_n\left(\operatorname{rank}(\hat{M}_n) = \operatorname{rank}(M)\right) \to 1,$$

where $\hat{M}_n$ *is defined as in* (S38). *In consequence,*

$$\hat{M}_n^\dagger \xrightarrow{P_n} M^\dagger.$$

Proof. Throughout let $\hat{r}_n := \operatorname{rank}(\hat{M}_n)$, $r := \operatorname{rank}(M)$, $R_n := \{\hat{r}_n = r\}$ and $\lambda_l, \lambda_{n,l}, \check{\lambda}_{n,l}$ and $\hat{\lambda}_{n,l}$ respectively the l-th largest eigenvalue of M, M_n, $\check{M}_n$ and $\hat{M}_n$.

Start with the case $r = 0$. By Weyl's perturbation theorem (e.g. Bhatia, 1997, Corollary III.2.6) and the fact that $M_n = 0$ for all n larger than some $N \in \mathbb{N}$,

$$P_n(R_n) = P_n\left(\max_{l=1,\dots,L} |\check{\lambda}_{n,l}| < \nu_n\right) \geq P_n(\|\check{M}_n - M_n\|_2 < \nu_n) \to 1.$$

On the sets R_n we have that $\hat{M}_n = 0 = M$ and so $\hat{M}_n \xrightarrow{P_n} M$ as $P(R_n) \to 1$.

Now suppose that $r > 0$. let $\underline{\nu} := \lambda_r/2 > 0$ and note that (S37) implies that $\|\check{M}_n - M_n\|_2 = o_{P_n}(1)$ and so, by Weyl's perturbation theorem, $\max_{l=1,\dots,L} |\check{\lambda}_{n,l} - \lambda_{n,l}| \leq \|\check{M}_n - M_n\|_2 = o_{P_n}(1)$. Hence, defining $E_n := \{\check{\lambda}_{n,r} \geq \nu_n\}$, for n large enough such that $\nu_n < \underline{\nu}$

[S23] (S37) is implied by $\|\check{M}_n - M_n\| = o_{P_n}(\nu_n)$ for any matrix norm. Moreover, the existence of such a sequence $(\nu_n)_{n\in\mathbb{N}}$ is guaranteed if $\|\check{M}_n - M_n\|_2 \to 0$ in P_n-probability, however its explicit knowledge is necessary to perform the subsequent construction. In most cases $M_n = M$ for all $n \in \mathbb{N}$.

and $\|M_n - M\|_2 < \underline{\nu}/2$ we have

$$P_n(E_n) = P_n\left(\check{\lambda}_{n,r} \geq \nu_n\right) \geq P_n\left(\check{\lambda}_{n,r} \geq \underline{\nu}\right) \geq P_n\left(|\check{\lambda}_{n,r} - \lambda_{n,r}| < \underline{\nu}/2\right) \to 1.$$

If $r = L$ we have that $R_n \supset E_n$ and therefore $P_n(R_n) \to 1$. Additionally, if $\check{\lambda}_{n,L} \geq \nu_n$ then $\hat{\lambda}_{n,l} = \check{\lambda}_{n,l}$ for each $l = 1, \ldots, L$ and hence $\hat{M}_n = \check{M}_n$, implying $\|\hat{M}_n - M\|_2 \leq \|\check{M}_n - M_n\|_2 + \|M_n - M\|_2 = o_{P_n}(1)$.

Now suppose instead that $r < L$ and define $F_n \coloneqq \{\check{\lambda}_{n,r+1} < \nu_n\}$. It follows by Weyl's perturbation theorem and the fact that $\lambda_{n,l} = 0$ for $l > r$ and $n \geq N$ that as $n \to \infty$

$$P_n(F_n) = P_n(\check{\lambda}_{n,r+1} < \nu_n) \geq P_n(\|\check{M}_n - M_n\|_2 < \nu_n) \to 1.$$

Since $R_n \supset E_n \cap F_n$, this implies that $P_n(R_n) \to 1$ as $n \to \infty$. Additionally, if $\check{\lambda}_{n,r} \geq \nu_n$, $\check{\lambda}_{n,r+1} < \nu_n$ and $\|\check{M}_n - M\|_2 \leq \upsilon$, we have that $\hat{\lambda}_{n,k} = \check{\lambda}_{n,k}$ for $k \leq r$ and $\hat{\lambda}_{n,l} = 0 = \lambda_l$ for $l > r$ and so

$$\|\Lambda_n(\nu_n) - \Lambda\|_2 = \max_{l=1,\ldots,r} |\hat{\lambda}_{n,l} - \lambda_l| = \max_{l=1,\ldots,r} |\check{\lambda}_{n,l} - \lambda_l| \leq \|\check{\Lambda}_n - \Lambda\|_2 \leq \|\check{M}_n - M\|_2 \leq \upsilon,$$

and hence $\{\|\check{M}_n - M\|_2 \leq \upsilon\} \cap E_n \cap F_n \subset \{\|\Lambda_n(\nu_n) - \Lambda\|_2 \leq \upsilon\}$, from which it follows that $\Lambda_n(\nu_n) \xrightarrow{P_n} \Lambda$ as $\|\check{M}_n - M\|_2 \leq \|\check{M}_n - M_n\|_2 + \|M_n - M\|_2 \xrightarrow{P_n} 0$. Suppose that $(\lambda_1, \ldots, \lambda_r)$ consists of s distinct eigenvalues with values $\lambda^1 > \lambda^2 > \cdots > \lambda^s$ and multiplicities $\mathfrak{m}_1, \ldots, \mathfrak{m}_s$ (each at least one).[S24] $\lambda^{s+1} = 0$ is an eigenvalue with multiplicity $\mathfrak{m}_{s+1} = L - r$. Let l_i^k for $k = 1, \ldots, s+1$ and $i = 1, \ldots, \mathfrak{m}_k$ denote the column indices of the eigenvectors in U corresponding to each λ^k. For each λ^k, the total eigenprojection is $\Pi_k \coloneqq \sum_{i=1}^{\mathfrak{m}_k} u_{l_i^k} u'_{l_i^k}$.[S25] Total eigenprojections are continuous.[S26] Therefore, if we construct $\Pi_{n,k}$ in in an analogous fashion to Π_k but replace columns of U with columns of $\check{U}_n$, we have $\Pi_{n,k} \xrightarrow{P_n} \Pi_k$ for each $k = 1, \ldots, s+1$ since $\check{M}_n \xrightarrow{P_n} M$. Spectrally decompose M as $M = \sum_{k=1}^{s} \lambda^k \Pi_k$, where the sum runs to s rather than $s+1$ since $\lambda^{s+1} = 0$. Then,

$$\hat{M}_n = \sum_{k=1}^{s+1} \sum_{i=1}^{\mathfrak{m}_k} \hat{\lambda}_{n,l_i^k} u_{n,l_i^k} u'_{n,l_i^k} = \sum_{k=1}^{s+1} \sum_{i=1}^{\mathfrak{m}_k} (\hat{\lambda}_{n,l_i^k} - \lambda^k) u_{n,l_i^k} u'_{n,l_i^k} + \sum_{k=1}^{s} \lambda^k \Pi_{n,k},$$

[S24] The superscripts on the λs are indices, not exponents.
[S25] See e.g Chapter 8.8 of Magnus and Neudecker (2019).
[S26] E.g. Theorem 8.7 of Magnus and Neudecker (2019).

whence

$$\|\hat{M}_n - M\|_2 \leq \sum_{k=1}^{s+1}\sum_{i=1}^{\mathfrak{m}_k} |\hat{\lambda}_{n,l_i^k} - \lambda^k| \|u_{n,l_i^k}u'_{n,l_i^k}\|_2 + \sum_{k=1}^{s} |\lambda^k| \|\Pi_{n,k} - \Pi_k\|_2 \xrightarrow{P_n} 0,$$

by $\hat{\Pi}_{n,k} \xrightarrow{P_n} \Pi_k$, $\hat{\Lambda}_n(\boldsymbol{\nu}_n) \xrightarrow{P_n} \Lambda$ and since we have $\|u_{n,l_i^k}u'_{n,l_i^k}\|_2 = 1$ for any i, k, n. Combine this with $P_n(R_n) \to 1$ and Lemma 1 in Andrews (1987) to conclude. $\square$

S6 Log density score estimation

In this section we discuss the details for the estimation of the log density scores ϕ_k. We first provide a detailed description of the construction of the estimator (11). Secondly we provide a proofs of Lemma S4, i.e. we show that this estimate satisfies Assumption S1. Thirdly we provide proofs of Lemmas S8 and S9. The analysis here (in addition to the proposed estimator) is based on Chen and Bickel (2006) and Jin (1992), with small tweaks to fit the setup of the present paper.

S6.1 B-spline based log density score estimation

For $\xi_1 < \cdots < \xi_N$ a knot sequence, the first order B-splines are defined according to $b_i^{(1)}(x) := \mathbf{1}_{[\xi_i, \xi_{i+1})}(x)$. Subsequent order B-splines can be computed according to the recurrence relation

$$b_i^{(l)}(x) = \frac{x - \xi_i}{\xi_{i+l-1} - \xi_i} b_i^{(l-1)}(x) + \frac{\xi_{i+l} - x}{\xi_{i+l} - \xi_{i+1}} b_{i+1}^{(l-1)}(x), \tag{S39}$$

for $l > 1$ and $i = 1, \ldots, N - l$. A l-th order B-spline is $l - 2$ times differentiable in x with first derivative

$$c_i^{(l)}(x) = \frac{l - 1}{\xi_{i+l-1} - \xi_i} b_i^{(l-1)}(x) - \frac{l - 1}{\xi_{i+l} - \xi_{i+1}} b_{i+1}^{(l-1)}(x). \tag{S40}$$

See de Boor (2001) for more details on B-splines.

Let $b_{k,n} = (b_{k,n,1}, \ldots, b_{k,n,\mathsf{B}_{k,n}})'$ be a collection of $\mathsf{B}_{k,n}$ cubic (i.e. 4-th order) B-splines and let $c_{k,n} = (c_{k,n,1}, \ldots, c_{k,n,\mathsf{B}_{k,n}})'$ be their derivatives: $c_{k,n,i}(x) := \frac{\mathrm{d}b_{k,n,i}(x)}{\mathrm{d}x}$ for each $i \in \{1, \ldots, \mathsf{B}_{k,n}\}$. The knots of the splines, $\xi_{k,n} = (\xi_{k,n,i})_{i=1}^{K_{k,n}}$ are equally spaced in $[\Xi_{k,n}^L, \Xi_{k,n}^U]$ with $\delta_{k,n} := \xi_{k,n,i+1} - \xi_{k,n,i} > 0$.[S27] For each (k, n) pair the relationships between the number of knots ($K_{k,n}$), the number of spline functions ($\mathsf{B}_{k,n}$) and $\delta_{k,n}$ are given by $\mathsf{B}_{k,n} = K_{k,n} - 4$ and $K_{k,n} = 1 + (\Xi_{k,n}^U - \Xi_{k,n}^L)/\delta_{k,n}$.[S28]

[S27]For each $k = 1, \ldots, K$ the sequences $(\Xi_{k,n}^L)_{n\in\mathbb{N}}$, $(\Xi_{k,n}^U)_{n\in\mathbb{N}}$, $(\mathsf{B}_{k,n})_{n\in\mathbb{N}}$ and $(\delta_{k,n})_{n\in\mathbb{N}}$ are deterministic.

[S28]Implicitly we choose $K_{k,n}$ and the endpoints and $\delta_{k,n}$ adjusts such that these formulae hold; this way

Since the B-splines vanish at infinity for any $n \in \mathbb{N}$, integration by parts gives that

$$\begin{aligned} &\int (\phi_k(z) - \psi_{k,n}' b_{k,n}(z))^2 \eta_k(z) \, \mathrm{d}z \\ &\quad = \int \phi_k(z)^2 \eta_k(z) \, \mathrm{d}z + \int (\psi_{k,n}' b_{k,n})^2 \eta_k(z) \, \mathrm{d}z + 2 \int \psi_{k,n}' c_{k,n}(z) \eta_k(z) \, \mathrm{d}z \\ &\quad = \mathbb{E}\phi_k(\epsilon_k)^2 + \psi_{k,n}' \mathbb{E}[b_{k,n}(\epsilon_k) b_{k,n}(\epsilon_k)'] \psi_{k,n} + 2\psi_{k,n}' \mathbb{E} c_{k,n}(\epsilon_k), \end{aligned} \tag{S41}$$

where we integrate over the support of $\phi_{k,n}$ (which is also the support of $b_{k,n}$ and $c_{k,n}$). This mean-squared error is minimsed by:[S29]

$$\psi_{k,n} \coloneqq -\mathbb{E}[b_{k,n}(\epsilon_k) b_{k,n}(\epsilon_k)']^{-1} \mathbb{E}[c_{k,n}(\epsilon_k)]. \tag{S42}$$

Replace the population expectations with sample counterparts to define the estimate of $\psi_{k,n}$

$$\hat{\psi}_{k,n,\gamma} \coloneqq - \left[\frac{1}{n} \sum_{i=1}^{n} b_{k,n}(A_{n,k,i} V_{\theta_n,i}) b_{k,n}(A_{n,k,i} V_{\theta_n,i})' \right]^{-1} \frac{1}{n} \sum_{i=1}^{n} c_{k,n}(A_{n,k,i} V_{\theta_n,i}), \tag{S43}$$

where $A_{n,k,i}$ and $V_{\theta_n,i}$ are defined as in Assumption S1. The estimate for ϕ_k is

$$\hat{\phi}_{k,n,\gamma}(z) \coloneqq \hat{\psi}_{k,n,\gamma}' b_{k,n}(z) \ . \tag{S44}$$

We note that computing (S44) effective only requires computing the B-spline regression coefficients $\hat{\psi}_{k,n,\gamma}$ in (S43). To implement the score test we need to estimate K density scores, hence the computational cost is quite modest.

S6.2 Proof of Lemmas S4, S8 & S9

Proof of Lemma S4. Under P_{θ_n}, $A_{k,\gamma_n,i} V_{\theta_n,i} \simeq \epsilon_k \sim \eta_k$. We start by showing that $\hat{\phi}_{k,n} \coloneqq \hat{\phi}_{k,n,\gamma_n}$ (where $\gamma_n = (\alpha_0, \beta_n)$) satisfies equation (S12). We have

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^{n} \hat{\phi}_{k,n}(\epsilon_{i,k}) Z_{n,i} - \phi_k(\epsilon_{i,k}) Z_{n,i} \right| \leq & \left| \frac{1}{n} \sum_{i=1}^{n} \left[\hat{\phi}_{k,n}(\epsilon_{i,k}) - \tilde{\phi}_{k,n}(\epsilon_{i,k}) \right] Z_{n,i} \right| \\ & + \left| \frac{1}{n} \sum_{i=1}^{n} \left[\tilde{\phi}_{k,n}(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k}) \right] Z_{n,i} \right| \\ & + \left| \frac{1}{n} \sum_{i=1}^{n} \left[\phi_{k,n}(\epsilon_{i,k}) - \phi_k(\epsilon_{i,k}) \right] Z_{n,i} \right|, \end{aligned} \tag{S45}$$

we do not need to adjust anything to ensure these are integers.

[S29] This differs from the expression in Chen and Bickel (2006) by a factor of -1 as they estimate $-\phi_k$.

where $\phi_{k,n} := \phi_k \mathbf{1}_{[\Xi^L_{k,n}, \Xi^U_{k,n}]}$ as in Assumption 3, $\tilde{\phi}_{k,n}(z) := \psi'_{k,n} b_{k,n}(z)$ and $\hat{\phi}_{k,n}(z) := \hat{\psi}'_{k,n,\gamma_n} b_{k,n}(z)$. To establish (S12) it suffices to show that each of these three terms on the right hand side are $o_{\mathrm{P}}(n^{-1/2})$.[S30]

For the last term in (S45), by assumption $\mathbb{E}\mathbf{1}\{\epsilon_{i,k} \notin [\Xi^L_{k,n}, \Xi^U_{k,n}]\} \downarrow 0$ and hence by independence, Cauchy-Schwarz and $\sup_{n \in \mathbb{N}} \mathbb{E} Z^2_{n,i} < \infty$,

$$\begin{aligned} \mathbb{E}\left([\phi_{k,n}(\epsilon_{i,k}) - \phi_k(\epsilon_{i,k})]^2 Z^2_{n,i}\right) &= \mathbb{E}\left[\phi_k(\epsilon_{i,k})^2 \mathbf{1}\{\epsilon_{i,k} \notin [\Xi^L_{k,n}, \Xi^U_{k,n}]\}\right] \mathbb{E} Z^2_{n,i} \\ &\leq \left[\mathbb{E}\phi_k(\epsilon_{i,k})^4\right]^{1/2} \left[\mathbb{E}\mathbf{1}\{\epsilon_{i,k} \notin [\Xi^L_{k,n}, \Xi^U_{k,n}]\}\right]^{1/2} \mathbb{E} Z^2_{n,i} \\ &\to 0. \end{aligned} \tag{S46}$$

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

$$\mathrm{P}\left(\left|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\phi_{k,n}(\epsilon_{i,k}) - \phi_k(\epsilon_{i,k})] Z_{n,i}\right| > \upsilon\right) \leq \frac{n\mathbb{E}\left([\phi_{k,n}(\epsilon_{i,k}) - \phi_k(\epsilon_{i,k})]^2 Z^2_{n,i}\right)}{n\upsilon} \to 0.$$

For the second term, we note that by our hypotheses and lemma S22 we have

$$\begin{aligned} \mathbb{E}\left([\tilde{\phi}_{k,n}(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})]^2 Z^2_{n,i}\right) &= \mathbb{E}\left([\tilde{\phi}_{k,n}(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})]^2\right) \mathbb{E} Z^2_{n,i} \\ &\leq C^2 \delta^6_{k,n} \|\phi^{(3)}_{k,n}\|^2_\infty \mathbb{E} Z^2_{n,i} \to 0 \end{aligned}, \tag{S47}$$

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

$$\mathrm{P}\left(\left|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\tilde{\phi}_{k,n}(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})] Z_{n,i}\right| > \upsilon\right) \leq \frac{n\mathbb{E}\left([\tilde{\phi}_k(\epsilon_{i,k}) - \phi_{k,n}(\epsilon_{i,k})]^2 Z^2_{n,i}\right)}{n\upsilon} \to 0.$$

For the first term, by Cauchy-Schwarz

$$\left|\frac{1}{n} \sum_{i=1}^{n} \left[\hat{\phi}_{k,n}(\epsilon_{i,k}) - \tilde{\phi}_{k,n}(\epsilon_{i,k})\right] Z_{n,i}\right| \leq \|\hat{\psi}_{k,n} - \psi_{k,n}\|_2 \left\|\frac{1}{n} \sum_{i=1}^{n} b_{k,n}(\epsilon_{i,k}) Z_{n,i}\right\|_2 = o_{\mathrm{P}}(n^{-1/2}),$$

by lemmas S23 and S24.

[S30] Here we implicitly assume (without loss of generality) that all the ϵ_i and $Z_{n,i}$ are defined on a common probability space $(\Omega, \mathcal{F}, \mathrm{P})$.

Next, we show that $\hat{\phi}_{k,n}$ satisfies equation (S13). We have:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(\left[\hat{\phi}_{k,n}(\epsilon_{i,k})-\phi_k(\epsilon_{i,k})\right]Z_{n,i}\right)^2 \leq \frac{4}{n}\sum_{i=1}^{n}&\left[\hat{\phi}_{k,n}(\epsilon_{i,k})-\tilde{\phi}_{k,n}(\epsilon_{i,k})\right]^2 Z_{n,i}^2\\ &+\frac{4}{n}\sum_{i=1}^{n}\left[\tilde{\phi}_{k,n}(\epsilon_{i,k})-\phi_{k,n}(\epsilon_{i,k})\right]^2 Z_{n,i}^2\\ &+\frac{4}{n}\sum_{i=1}^{n}\left[\phi_{k,n}(\epsilon_{i,k})-\phi_k(\epsilon_{i,k})\right]^2 Z_{n,i}^2.\end{aligned}\tag{S48}$$

We will show that (1/4 of) each of the right hand side terms is $o_{\mathrm{P}}(\nu_n)$ under our assumptions, which is sufficient for equation (S13). For the last term, for any $\upsilon > 0$, by Markov's inequality, independence and Cauchy-Schwarz we have

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left[\phi_{k,n}(\epsilon_{i,k})-\phi_k(\epsilon_{i,k})\right]^2 Z_{n,i}^2\right|>\upsilon\nu_n\right)\lesssim\frac{\left[\mathbb{E}\mathbf{1}\{\epsilon_{i,k}\notin[\Xi_{k,n}^L,\Xi_{k,n}^U]\}\right]^{1/2}\mathbb{E}Z_{n,i}^2}{\upsilon\nu_n}=o(1).$$

For the second term, for any $\upsilon > 0$, by Markov's inequality, independence and lemma S22:

$$\begin{aligned}\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left[\tilde{\phi}_{k,n}(\epsilon_{i,k})-\phi_{k,n}(\epsilon_{i,k})\right]^2 Z_{n,i}^2\right|>\upsilon\nu_n\right)&\leq\frac{\mathbb{E}\left(\left[\tilde{\phi}_{k,n}(\epsilon_{i,k})-\phi_{k,n}(\epsilon_{i,k})\right]^2\right)\mathbb{E}Z_{n,i}^2}{\upsilon\nu_n}\\ &\leq\frac{C\delta_{k,n}^6\|\phi_{k,n}^{(3)}\|_\infty^2\mathbb{E}Z_{n,i}^2}{\upsilon\nu_n}\\ &=o(1).\end{aligned}$$

Finally, for the first term in the decomposition, by lemma S24 and Assumption 3-part (ii):

$$\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\phi}_{k,n}(\epsilon_{i,k})-\tilde{\phi}_{k,n}(\epsilon_{i,k})\right]^2 Z_{n,i}^2\leq\|\hat{\psi}_{k,n}-\psi_{k,n}\|_2^2\left[\frac{1}{n}\sum_{i=1}^{n}\|b_{k,n}(\epsilon_{i,k})\|_2^2 Z_{n,i}^2\right]=o_{\mathrm{P}}(\nu_n).\quad\square$$

Proof of Lemma S8. The proof proceeds verbatim as that of Lemma S4 once references to equations (S12), (S13) are replaced by equations (S27), (S28) since under the conditions of the present Lemma, one still has $A_{n,\gamma_n,i}V_{\theta_n,i}\simeq\epsilon_k\sim\eta_k$ under P_{θ_n}. $\square$

Proof of Lemma S9. We use a similar decomposition to as in the Proof of Lemma S4:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\left[\hat{\phi}_{k,n}(\epsilon_{i,k})-\phi_k(\epsilon_{i,k})\right]\epsilon_{k,i}\right)^2 \le \frac{4}{n}\sum_{i=1}^{n}\left[\hat{\phi}_{k,n}(\epsilon_{i,k})-\tilde{\phi}_{k,n}(\epsilon_{i,k})\right]^2\epsilon_{k,i}^2 + \frac{4}{n}\sum_{i=1}^{n}\left[\tilde{\phi}_{k,n}(\epsilon_{i,k})-\phi_{k,n}(\epsilon_{i,k})\right]^2\epsilon_{k,i}^2 + \frac{4}{n}\sum_{i=1}^{n}\left[\phi_{k,n}(\epsilon_{i,k})-\phi_k(\epsilon_{i,k})\right]^2\epsilon_{k,i}^2. \tag{S49}$$

We will show that (1/4 of) each of the right hand side terms is $o_{\mathrm{P}}(\nu_n)$ under our assumptions, which is sufficient for equation (S29), since under P_{θ_n}, $A_{k,\gamma_n,i}V_{\theta_n,i} \simeq \epsilon_k \sim \eta_k$. For the last term, for any $\upsilon > 0$, by Markov's inequality, Cauchy – Schwarz and the first additional condition in Lemma S9 we have

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left[\phi_{k,n}(\epsilon_{i,k})-\phi_k(\epsilon_{i,k})\right]^2\epsilon_{k,i}^2\right| > \upsilon\nu_n\right) \lesssim \frac{\left(\mathbb{E}\left[\epsilon_{k,i}^4\mathbf{1}\{\epsilon_{i,k}\notin[\Xi_{k,n}^L,\Xi_{k,n}^U]\}\right]\right)^{1/2}}{\upsilon\nu_n} = o(1).$$

For the second term, first note that by Lemma S24

$$\tilde{\phi}_{k,n}(\epsilon_{i,k})^2 \le \|\psi_{k,n}\|_2^2\|b_{k,n}(\epsilon_{i,k})\|_2^2 \le \|\psi_{k,n}\|_2^2 \le \|\Gamma_{k,n}^{-1}\|_2^2\|C_{k,n}\|_2^2 = O(\delta_{k,n}^{-3}\Delta_{k,n}).$$

Thus, for any $\upsilon > 0$, by Markov's inequality, Cauchy – Schwarz, the additional conditions in Lemma S9 and Lemma S22:

$$\begin{aligned}
&\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left[\tilde{\phi}_{k,n}(\epsilon_{i,k})-\phi_{k,n}(\epsilon_{i,k})\right]^2\epsilon_{k,i}^2\right| > \upsilon\nu_n\right)\\
&\le \frac{\mathbb{E}\left([\tilde{\phi}_{k,n}(\epsilon_{i,k})-\phi_{k,n}(\epsilon_{i,k})]^2\epsilon_{k,i}^2\right)}{\upsilon\nu_n}\\
&\le \frac{\mathsf{M}_{k,n}^2\mathbb{E}\left([\tilde{\phi}_{k,n}(\epsilon_{i,k})-\phi_{k,n}(\epsilon_{i,k})]^2\right)}{\upsilon\nu_n} + \frac{\mathbb{E}\left[(\tilde{\phi}_{k,n}(\epsilon_{i,k})^2+\phi_k(\epsilon_{i,k})^2)\epsilon_{i,k}^2\mathbf{1}\{|\epsilon_{i,k}|>\mathsf{M}_{k,n}\}\right]}{\upsilon\nu_n}\\
&\lesssim \frac{\mathsf{M}_{k,n}^2C\delta_{k,n}^6\|\phi_{k,n}^{(3)}\|_\infty^2}{\upsilon\nu_n} + \frac{\delta_{k,n}^{-3}\Delta_{k,n}\mathbb{E}\left[\epsilon_{i,k}^2\mathbf{1}\{|\epsilon_{i,k}|>\mathsf{M}_{k,n}\}\right]}{\upsilon\nu_n} + \frac{\left[\mathbb{E}\left(\epsilon_{i,k}^4\mathbf{1}\{|\epsilon_{i,k}|>\mathsf{M}_{k,n}\}\right)\right]^{1/2}}{\upsilon\nu_n}\\
&= o(1).
\end{aligned}$$

Finally, for the first term in the decomposition, by lemma S24, $\|b_{k,n}(\epsilon_{i,k})\|_2^2 \le 1$ (e.g. de Boor, 2001, equation (36), p. 96), Assumption S3, the WLLN and Assumption 3-part (ii)

$$\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\phi}_{k,n}(\epsilon_{i,k})-\tilde{\phi}_{k,n}(\epsilon_{i,k})\right]^2\epsilon_{k,i}^2 \leq \|\hat{\psi}_{k,n,\gamma_n}-\psi_{k,n}\|_2^2\left[\frac{1}{n}\sum_{i=1}^{n}\epsilon_{k,i}^2\right] = o_{\mathrm{P}}(\nu_n). \qquad \square$$

S6.3 Technical lemmas

Lemma S22 (Cf. Lemma A.5, Chen and Bickel, 2006)**.** *Let $\phi_{k,n}$ be defined as in Assumption 3 and $\tilde{\phi}_{k,n} := \psi'_{k,n} b_{k,n}$. If part (iv) of Assumption 3 holds,*

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

Proof. By the definition of $\tilde{\phi}_{k,n}$ and lemma S26 we have

$$\mathbb{E}\left(\tilde{\phi}_{k,n}(\epsilon_{i,k})-\phi_{k,n}(\epsilon_{i,k})\right)^2 = \inf_{g\in\mathscr{G}_4(\xi_{k,n})}\mathbb{E}\left(g(\epsilon_{i,k})-\phi_{k,n}(\epsilon_{i,k})\right)^2 \leq C^2\delta_{k,n}^6\|\phi_{k,n}^{(3)}\|_\infty^2,$$

where the equality follows since $\psi_{k,n}$ is the minimiser of (S41) where we integrate over the support of $\phi_{k,n}$ (which is also the support of $b_{k,n}$ and $c_{k,n}$). $\square$

Lemma S23 (Cf. Lemma A.3, Chen and Bickel, 2006)**.** *Suppose assumptions 2 (or S3) and 3 hold. If $Z_{n,i}$ is independent of $\epsilon_{i,k}$ and $\sup_{n\in\mathbb{N},i\leq 1,\ldots,n}\mathbb{E}Z_{n,i}^2<\infty$, then*

$$\left\|\frac{1}{n}\sum_{i=1}^{n}b_{k,n}(\epsilon_{i,k})Z_{n,i}\right\|_2 = O_{\mathrm{P}}(n^{-1/2}).$$

Proof. By $\sum_{m=1}^{\mathsf{B}_{k,n}} b_{k,n,m}(x)^2 \leq 1$ (e.g. de Boor, 2001, equation (36), p. 96) and our hypotheses

$$\mathbb{E}\left(\left\|\frac{1}{n}\sum_{i=1}^{n}b_{k,n}(\epsilon_{i,k})Z_{n,i}\right\|_2^2\right) = \frac{1}{n}\mathbb{E}\left(\sum_{m=1}^{\mathsf{B}_{k,n}}b_{n,k,m}(\epsilon_{i,k})^2\right)\mathbb{E}Z_{n,i}^2 \leq \frac{\mathbb{E}Z_{n,i}^2}{n}.$$

Fix $\epsilon>0$ and take $M>0$ large enough such that $\sup_{n\in\mathbb{N},i\leq 1,\ldots,n}\mathbb{E}Z_{n,i}^2/M^2<\epsilon$. Markov's inequality yields

$$\mathrm{P}\left(\sqrt{n}\left\|\frac{1}{n}\sum_{i=1}^{n}b_{k,n}(\epsilon_{i,k})Z_{n,i}\right\|_2 > M\right) \leq \frac{\mathbb{E}\left(n\left\|\frac{1}{n}\sum_{i=1}^{n}b_{k,n}(\epsilon_{i,k})Z_{n,i}\right\|_2^2\right)}{M^2} \leq \frac{\mathbb{E}Z_{n,i}^2}{M^2} < \epsilon.$$

$\square$

Lemma S24 (Cf. Lemma A.2, Chen and Bickel, 2006)**.** *Suppose that Assumptions 2 (or S3) and 3 hold. Then, for*

$$\hat{\Gamma}_{k,n} := \frac{1}{n}\sum_{i=1}^{n} b_{k,n}(\epsilon_{i,k}) b_{k,n}(\epsilon_{i,k})', \quad \Gamma_{k,n} := \mathbb{E}[b_{k,n}(\epsilon_k) b_{k,n}(\epsilon_k)'],$$

and

$$\hat{C}_{k,n} := \frac{1}{n}\sum_{i=1}^{n} c_{k,n}(\epsilon_{i,k}), \quad C_{k,n} := \mathbb{E}[c_{k,n}(\epsilon_k)],$$

we have that

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

(ii) $\|\hat{C}_{k,n} - C_{k,n}\|_2 = O_{\mathrm{P}}\left(\sqrt{\frac{\mathsf{B}_{k,n}\log \mathsf{B}_{k,n}}{n\delta_{k,n}^2}}\right)$,

(iii) $\|\hat{\Gamma}_{k,n} - \Gamma_{k,n}\|_2 = O_{\mathrm{P}}\left(\sqrt{\frac{\mathsf{B}_{k,n}\log \mathsf{B}_{k,n}}{n}}\right)$,

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

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

In particular, $\|\hat{\Gamma}_{k,n}^{-1}\hat{C}_{k,n} - \psi_{k,n}\|_2 = O_{\mathrm{P}}(n^{-1/2}\Delta_{k,n}\delta_{k,n}^{-4}(\Delta_{k,n}\delta_{k,n}^{-1})^{\iota}) = o_{\mathrm{P}}(1)$ *and* $\|\hat{\Gamma}_{k,n}\|_2 = o_{\mathrm{P}}(1)$.

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) $c_{k,n,i} = \left(b_{k,n,i}^{(3)} - b_{k,n,i+1}^{(3)}\right)/\delta_{k,n}$. We have, for large enough $n \in \mathbb{N}$,

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

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 $[0,1]$.[S31] It follows immediately that for large enough $n \in \mathbb{N}$,

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

from which (i) follows.

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

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

Therefore,

$$\begin{aligned}\mathrm{P}\left(\|\hat{C}_{k,n} - C_{k,n}\|_2 \geq t\right) &\leq \sum_{m=1}^{\mathsf{B}_{k,n}} \mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} c_{k,n,m}(\epsilon_{i,k}) - \mathbb{E}c_{n,k,m}(\epsilon_{i,k})\right| \geq \frac{t}{\sqrt{\mathsf{B}_{k,n}}}\right) \\ &\leq 2\mathsf{B}_{k,n}\exp(-nt^2\mathsf{B}_{k,n}^{-1}\delta_{k,n}^2/2),\end{aligned}$$

and so for any fixed $\epsilon > 0$ we can take $t = \sqrt{\frac{4\mathsf{B}_{k,n}\log \mathsf{B}_{k,n}}{n\delta_{k,n}^2}}$ to obtain (ii) as then

$$\mathrm{P}\left(\|\hat{C}_{k,n} - C_{k,n}\|_2 \geq t\right) \leq 2\mathsf{B}_{k,n}^{-1} \to 0.$$

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

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} b_{k,n,m}(\epsilon_{i,k})b_{k,n,s}(\epsilon_{i,k}) - \mathbb{E}[b_{k,n,m}(\epsilon_{i,k})b_{k,n,s}(\epsilon_{i,k})]\right| \geq t\right) \leq 2\exp(-2nt^2).$$

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)

[S31] This is evident from their definition. See also property (36) (p. 96) of de Boor (2001).

entries where $|m-s|>3$ (de Boor, 2001, (20), p. 91) we have that

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

Putting $t=\sqrt{\frac{7\mathsf{B}_{k,n}\log \mathsf{B}_{k,n}}{n}}$ we obtain (iii) as

$$\mathrm{P}\left(\|\hat{\Gamma}_{k,n}-\Gamma_{k,n}\|_2 \geq t\right) \leq 14\mathsf{B}_{k,n}^{-1} \to 0.$$

Since $\Gamma_{k,n}$ is symmetric and positive (semi-)definite we have that:[S32]

$$\|\Gamma_{k,n}\|_2 \leq \|\Gamma_{k,n}\|_\infty = \max_{m=1,\ldots,\mathsf{B}_{k,n}} \sum_{s=1}^{\mathsf{B}_{k,n}} \mathbb{E} b_{n,k,m}(\epsilon_k) b_{k,n,s}(\epsilon_k).$$

Then, since for any $z\in\mathbb{R}$, each row of $b_{k,n}(z)b_{k,n}(z)'$ has at most 7 non-zero entries,[S33] all of which are bounded above by 1 we have

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

which yields (iv) in conjunction with requirement (iii) of Assumption 3.

By Assumption 3 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 $\int b_{k,n}b_{k,n}'(\eta-c\delta_{k,n})\lambda = \int (b_{k,n}\sqrt{\eta-c\delta_{k,n}})(b_{k,n}\sqrt{\eta-c\delta_{k,n}})'\lambda$. 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

[S32] See e.g. Theorem 5.6.9 in Horn and Johnson (2013).
[S33] $b_{k,n,m}(z)=0$ outside $[\xi_{k,n,m},\xi_{k,n,m+4})$. See (20) on p. 91 in de Boor (2001).

the matrix $\int b_{k,n}b'_{k,n}(\eta - c\delta_k)\,\mathrm{d}\lambda$ 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 S25. Hence, by the Rayleigh quotient theorem (see e.g. Theorem 4.2.2 in Horn and Johnson, 2013) and lemma S25

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

for a $v > 0$, which may be used to conclude that (v) holds via

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

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

$$\|\hat{C}_{k,n}\|_2 \leq \|\hat{C}_{k,n} - C_{k,n}\|_2 + \|C_{k,n}\|_2 = O_{\mathrm{P}}\left(\sqrt{\frac{\mathsf{B}_{k,n}\log \mathsf{B}_{k,n}}{n\delta_{k,n}^2}}\right) + O\left(\delta_{k,n}\sqrt{\mathsf{B}_{k,n}}\right) = O_{\mathrm{P}}\left(\delta_{k,n}\sqrt{\mathsf{B}_{k,n}}\right),$$

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

$$\begin{aligned}
\|\hat{\Gamma}_{k,n}^{-1}\|_2 &\leq \|\Gamma_{k,n}^{-1}(I + [\hat{\Gamma}_{k,n} - \Gamma_{k,n}]\Gamma_{k,n}^{-1})^{-1}\|_2 \\
&\leq \|\Gamma_{k,n}^{-1}\|_2\|(I + [\hat{\Gamma}_{k,n} - \Gamma_{k,n}]\Gamma_{k,n}^{-1})^{-1}\|_2 \\
&\leq \|\Gamma_{k,n}^{-1}\|_2\left(1 - \|[\hat{\Gamma}_{k,n} - \Gamma_{k,n}]\Gamma_{k,n}^{-1}\|_2\right)^{-1} \\
&\leq \|\Gamma_{k,n}^{-1}\|_2\left(1 - \|\hat{\Gamma}_{k,n} - \Gamma_{k,n}\|_2\|\Gamma_{k,n}^{-1}\|_2\right)^{-1} \\
&= O_{\mathrm{P}}(\delta_{k,n}^{-2}).
\end{aligned} \tag{S50}$$

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

$$\begin{aligned}
\|\hat{\psi}_{k,n} - \psi_{k,n}\|_2 &= \|\hat{\Gamma}_{k,n}^{-1}\hat{C}_{k,n} - \Gamma_{k,n}^{-1}C_{k,n}\|_2 \\
&\leq \|(\hat{\Gamma}_{k,n}^{-1} - \Gamma_{k,n}^{-1})\hat{C}_{k,n}\|_2 + \|\Gamma_{k,n}^{-1}(\hat{C}_{k,n} - C_{k,n})\|_2 \\
&\leq \|\Gamma_{k,n}^{-1}\|_2\|\Gamma_{k,n} - \hat{\Gamma}_{k,n}\|_2\|\hat{\Gamma}_{k,n}^{-1}\|_2\|\hat{C}_{k,n}\|_2 + \|\Gamma_{k,n}^{-1}\|_2\|\hat{C}_{k,n} - C_{k,n}\|_2 \\
&= O_{\mathrm{P}}\left(\sqrt{\frac{\mathsf{B}_{k,n}^2\log \mathsf{B}_{k,n}}{\delta_{k,n}^6 n}}\right) + O_{\mathrm{P}}\left(\sqrt{\frac{\mathsf{B}_{k,n}\log \mathsf{B}_{k,n}}{\delta_{k,n}^6 n}}\right) \\
&= o_{\mathrm{P}}(1),
\end{aligned}$$

by Assumption 3 part (ii), since we have $\mathsf{B}_{k,n} \leq \Delta_{k,n}\delta_{k,n}^{-1}$ and hence the dominant term above vanishes since for all large enough n,

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

Finally, by (iii) and (iv) and Assumption 3 part (ii) we have

$$\|\hat{\Gamma}_{k,n}\|_2 \leq \|\hat{\Gamma}_{k,n} - \Gamma_{k,n}\|_2 + \|\Gamma_{k,n}\|_2 = O_{\mathrm{P}}\left(\sqrt{\frac{\mathsf{B}_{k,n}\log \mathsf{B}_{k,n}}{n}}\right) + O(\delta_{k,n}) = o_{\mathrm{P}}(1),$$

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

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

Lemma S25. *The smallest eigenvalue of the $\mathsf{B}_{k,n} \times \mathsf{B}_{k,n}$ Gram matrix $\tilde{\Gamma}_{k,n} \coloneqq \int b_{k,n}b_{k,n}' \,\mathrm{d}\lambda$ satisfies*

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

for a $\upsilon > 0$.

Proof. Since $b_{k,n,m}(x)b_{k,n,s}(x)$ is non-zero only for $|m-s| \leq 3$ and each $b_{k,n,m}$ is non-zero only on $[\xi_{m,k,n}, \xi_{m+4,k,n})]$ (e.g. (20) p. 91 of de Boor, 2001), $\tilde{\Gamma}_{k,n}$ is a symmetric banded Toeplitz matrix.[S34] Its entries can be computed by direct integration:

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

Let $f_0 \coloneqq \frac{151}{315}$, $f_1 \coloneqq f_{-1} \coloneqq \frac{397}{1680}$, $f_2 \coloneqq f_{-2} \coloneqq \frac{1}{42}$ and $f_3 \coloneqq f_{-3} \coloneqq \frac{1}{5040}$ and let $f_s \coloneqq 0$ for $|s| > 3$. Now, let $f(\theta) \coloneqq \sum_{s=-3}^{3} f_s e^{i(s\theta)}$. Then, $\tilde{\Gamma}_{k,n}/\delta_{k,n}$ is then the matrix generated by f in the sense that $\tilde{\Gamma}_{k,n}/\delta_{k,n} = \mathscr{T}_n(f) \coloneqq \sum_{s=-\min(\mathsf{B}_{k,n}-1,3)}^{\min(\mathsf{B}_{k,n}-1,3)} f_k J_n^s$ where each J_n^s is the $\mathsf{B}_{k,n} \times \mathsf{B}_{k,n}$ matrix which is zero everywhere except for the (i,j)-th entries where $i - j = s$, where it has a value of 1.[S35] Since $f \in L_1([-\pi,\pi])$ and is real on $[-\pi,\pi]$ by Theorem 6.1 in Garoni and Serra-

[S34] 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.

[S35] See section 6.1 in Garoni and Serra-Capizzano (2017), noting that it is clear that $f \in L_1([-\pi,\pi])$.

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 $\upsilon := \inf_{\theta\in[-\pi,\pi]} f(\theta) > 0$. $\square$

Lemma S26. *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 $\mathscr{G}_l(\xi)$ be the linear space formed by degree l splines with knots ξ. Then, if $f \in C^{l-1}[a,b]$ we have that*

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

where c_l depends only on l.

Proof. This is a special case of Theorem 20.3 in Powell (1981). $\square$

S7 Power optimality under strong identification

In either the setting considered in the main text or that introduced in Section S2, consider local alternatives of the type given in (17). We now prove the limiting power statements claimed in equations (18), (19) and (20).

Proposition S2. *Suppose that Assumptions 1, 2 and 3 (or S2, S3, S4 and S5) hold, $\alpha \in \mathbb{R}$ and $\tilde{\mathcal{I}}_\theta > 0$. Then, (18) holds.*

Proof. Apply Proposition S3 in the case where $L_\alpha = 1$ to obtain

$$\lim_{n\to\infty} P^n_{\theta_n(q,d,h)}\varphi_n = 1 - \mathrm{P}\left(\chi^2_1(\tilde{\mathcal{I}}_\theta q^2) \leq c_a\right).$$

The right hand side is the power function of the test $\psi(Z) := \mathbf{1}\{Z^2 > c_a\}$ for $Z \sim \mathcal{N}(\tilde{\mathcal{I}}_\theta^{1/2}q, 1)$. If $X = Z - \tilde{\mathcal{I}}_\theta^{1/2}q$, then

$$\psi(Z) = \mathbf{1}\{(X - \tilde{\mathcal{I}}_\theta^{1/2}q)^2 > c_a\} = \mathbf{1}\{|X - \tilde{\mathcal{I}}_\theta^{1/2}q| > z_{a/2}\}, \qquad X \sim \mathcal{N}(0,1),$$

hence $\mathbb{E}\psi(Z)$ is (18). $\square$

Proposition S3. *Suppose that Assumptions 1, 2 and 3 (or S2, S3, S4 and S5) hold and $\tilde{\mathcal{I}}_\theta$ is positive definite. Then, (19) holds.*

Proof. The proof of Theorem 1 (or Theorem S2) showed that the conditions of Theorem S1 hold. Therefore, by (S17), c_n is equal to the $1-a$ quantile of a $\chi^2_{L_\alpha}$ distribution with probability 1 for all large enough n. By (S15), (S16), Le Cam's third lemma (e.g. Example 6.7 in van der Vaart (1998)) and Theorem 12.14 in Rudin (1991),

$$\sqrt{n}\mathbb{P}_n\hat{\kappa}_{n,\bar{\gamma}_n} \rightsquigarrow \mathcal{N}(\tilde{\mathcal{I}}_\theta q, \tilde{\mathcal{I}}_\theta) \qquad \text{under } P^n_{\theta_n(q,d,h)}.$$

By condition 3, the mutual contiguity which follows from (S15) and Example 6.5 in van der Vaart (1998), Proposition S1 and Theorem 9.2.3 in Rao and Mitra (1971)

$$\hat{S}_{n,\bar{\gamma}_n} \rightsquigarrow \chi^2_{L_\alpha}(q'\tilde{\mathcal{I}}_\theta q) \qquad \text{under } P^n_{\theta_n(q,d,h)},$$

from which the result follows. □

Proposition S4. *Suppose that Assumptions 1, 2 and 3 (or S2, S3, S4 and S5) hold and $\tilde{\mathcal{I}}_\theta$ is positive definite. Then, (20) holds.*

Proof. By arguing exactly as in Proposition S3 with convergent sequences $(q_n, g_n, h_n) \to (q, d, h)$ replacing the fixed (q, d, h) in that Proposition one obtains that

$$\hat{S}_{n,\bar{\gamma}_n} \rightsquigarrow \chi^2_{L_\alpha}(q'\tilde{\mathcal{I}}_\theta q) \qquad \text{under } P^n_{\theta_n(q_n,d_n,h_n)},$$

and hence

$$\lim_{n\to\infty} P^n_{\theta_n(q_n,g_n,h_n)}\varphi_n = 1 - \mathrm{P}\left(\chi^2_{L_\alpha}(q'\tilde{\mathcal{I}}_\theta q) \le c_a\right), \tag{S51}$$

with c_a the $1 - a$ quantile of a $\chi^2_{L_\alpha}$ distribution. The proof is completed by a standard subsequence argument. Note first that the map $(q, d, h) \mapsto q'\tilde{\mathcal{I}}_\theta q$ from $\mathcal{V} \to \mathbb{R}$ is continuous. As $K^\star_u$ is compact this function attains its infimum, hence

$$u = \inf\{q'\tilde{\mathcal{I}}_\theta q : (q, d, h) \in K^\star_u\} = \min\{q'\tilde{\mathcal{I}}_\theta q : (q, d, h) \in K^\star_u\}.$$

Taking $(q_\star, d_\star, h_\star) \in K^\star_u$ such that $q_\star\tilde{\mathcal{I}}_\theta q_\star = u$, we have by (S51)

$$\limsup_{n\to\infty} \inf_{(q,d,h)\in K^\star_u} P^n_{\theta_n(q,d,h)}\varphi_n \le \lim_{n\to\infty} P^n_{\theta_n(q_\star,d_\star,h_\star)}\varphi_n = 1 - \mathrm{P}\left(\chi^2_{L_\alpha}(u) \le c_a\right) =: \mathcal{R}. \tag{S52}$$

There is a sequence $(v_n)_{n\in\mathbb{N}} \subset K^\star_u$ and a subsequence $(n_j)_{j\in\mathbb{N}}$ such that

$$\lim_{j\to\infty} v_{n_j} = v_\star = (q_\star, d_\star, h_\star) \in K^\star_u$$

and

$$\mathcal{S} := \liminf_{n\to\infty} \inf_{(q,d,h)\in K^\star_u} P^n_{\theta_n(q,d,h)}\varphi_n = \lim_{j\to\infty} P^{n_j}_{\theta_{n_j}(q_{n_j},d_{n_j},h_{n_j})}\varphi_{n_j}. \tag{S53}$$

Construct a new sequence $(v_m^*)_{m\in\mathbb{N}}$ as follows. For all $m \in [n_j, n_{j+1}) \cap \mathbb{N}$ for some $j \in \mathbb{N}$ put $v_m^* = v_{n_j}$ and for $m = 1, \ldots, n_1$ put $v_m^* = v_{n_1}$. By construction $\lim_{m\to\infty} v_m^* = v_\star$. By (S51)

$$\lim_{m\to\infty} P^m_{\theta_m(v_m^*)}\varphi_m = 1 - \mathrm{P}\left(\chi^2_r\left(u^\star\right) \leq c_a\right) \geq \mathcal{R}, \quad \text{with} \quad u^\star = (q_\star)'\tilde{\mathcal{I}}_\theta q_\star \geq u.$$

For any $\varepsilon > 0$, there is a $M \in \mathbb{N}$ such that if $m \geq M$, $P^m_{\theta_m(v_m^*)}\varphi_m \geq \mathcal{R} - \varepsilon$ by the preceding display. Taking a subsequence n_{j_k} such that for all $k \in \mathbb{N}$ we have $m_k = n_{j_k} \geq M$ gives

$$\mathcal{S} = \mathcal{S} - P^{n_{j_k}}_{\theta_{n_{j_k}}(v^*_{n_{j_k}})}\varphi_{n_{j_k}} + P^{m_k}_{\theta_{m_k}(v^*_{m_k})}\varphi_{m_k} \geq \mathcal{S} - P^{n_{j_k}}_{\theta_{n_{j_k}}(v^*_{n_{j_k}})}\varphi_{n_{j_k}} + \mathcal{R} - \varepsilon.$$

Take $k \to \infty$ to conclude (via (S53)) that $\mathcal{S} \geq \mathcal{R} - \varepsilon$. Since $\varepsilon > 0$ was arbitrary, it follows that $\mathcal{S} \geq \mathcal{R}$. Combine with equations (S52) and (S53) to obtain (20). □

S8 Additional simulation results

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

S8.1 Truncation in the baseline model

In our main simulations we truncated the effective information matrix estimate at machine precision, i.e. $\nu_n^{1/2} = 10^{-308}$. Here we investigate the sensitivity of the rejection frequencies to this choice. Specifically, we replicate Table 2 from the main text, fixing $B = 6$, but allowing for different truncation rates $\nu_n^{1/2} = 10^{-308}, 10^{-5}, 10^{-1}$.[S36] The value 10^{-1} is a high truncation value which implies that we end up truncating often when all densities are Gaussian. The results are shown in Table S1.

We find that the results are not sensitive to the truncation parameter choice. Comparing machine precision to $\nu_n^{1/2} = 10^{-5}$ yields no differences at all, whereas $\nu_n^{1/2} = 10^{-1}$ makes the test slightly conservative. Closer inspection reveals that the under rejection is due to cases where all eigenvalues are truncated and hence $\mathrm{rank}(\hat{\mathcal{I}}^t_{\hat{\gamma}}) = 0$. In Theorem 1 this corresponds to the conservative case.

S8.2 Additional power results for the baseline model

Figure 4 in the main text compared the power of different tests for the baseline model $Y_i = A^{-1}\epsilon_i$ for the case where $n = 1000$. Here we show the results for $n = 200$ and $n = 500$.

[S36] Recall that the specification corresponds to the baseline model $Y_i = A^{-1}\epsilon_i$, with A a rotation matrix parametrized by the Cayley transform. The first shock is always drawn from a Gaussian distribution whereas the remaining $k = 2, \ldots, K$ are from different distributions whose densities are shown in Figure 3.

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 S^{gmm} test over-rejects 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.

S8.3 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 null rejection frequency of the test remains well controlled. These findings persist when we increase to $n = 500$, though the power does improve as one would expect.

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

S8.4 Heteroskedastic LSEM model

In this section we study the empirical rejection frequency (under the null) of the semi-parametric score test for the heteroskedastic baseline model. Specifically we consider

$$Y_i = A(\alpha, \sigma, X_i)^{-1}\epsilon_i \qquad A(\alpha, \sigma, X_i)^{-1} = L(\sigma)D(\sigma, \tilde{X}_i)^{1/2}R(\alpha)' \,, \tag{S54}$$

where $R(\alpha)$ is a rotation matrix parametrized by the Cayley transformation of a skew-symmetric matrix (e.g. Gouriéroux, Monfort and Renne, 2017), $L(\sigma)$ is lower triangular with positive diagonal elements and $D(\sigma, \tilde{X}_i)$ is a diagonal matrix with diagonal elements given by

$$[D(\sigma, \tilde{X}_i)]_{jj} = \exp\left(\sigma'_{j1}\tilde{X}_i\right) \ , \qquad j = 1, \ldots, K \ ,$$

where σ_{j1} is a $(d-1) \times 1$ parameter vector. Note that the average scaling of the errors is captured by $L(\sigma)$ and $D(\sigma, \tilde{X}_i)$ is the only heteroskedastic part. More elaborate specifications that allow off-diagonal elements of L to depend on X_i are also possible.

The results for different sample sizes, dimensions K and number of explanatory variables are shown in Table S2. Overall, we find a similar pattern as for the LSEM model from the main text (cf Table 4). When $K = 5$ and the sample size is small, i.e. $n = 200$, the test tends to over-reject. The over-rejection vanishes for larger sample sizes. A slight difference is observed for heavy tailed densities (e.g. $t(5)$) where even with $n = 1000$ there is still some over-rejection.

S9 Additional empirical results

In this section we present some additional results for the returns to schooling application of section 6. Specifically, we consider the more flexible model from Section S2 which allows for conditional heteroskedasticity.

Starting from the baseline linear IV model with a possibly scalar endogenous instrument:

$$\begin{array}{rcl} y_i & = & \alpha_1 w_i + b'_y X_i + u_i \\ w_i & = & \pi z_i + b'_w X_i + v_i \\ z_i & = & B_z X_i + (\alpha_2/\sigma_u) u_i + e_i \end{array} \quad , \tag{S55}$$

We now allow the scaling of the errors σ_u, σ_v and σ_e to be a flexible functions of X_i. Specifically, we follow Wooldridge (2012, Chapter 8) and model the scales using flexible functions, i.e.

$$\sigma_j(X_i) = \sigma_{j,0} \exp\left(\sigma_{j1}\tilde{X}_{i,1} + \ldots + \sigma_{jd}\tilde{X}_{i,d-1}\right) \ , \qquad j = u, v, e \ ,$$

see also Romano and Wolf (2017) for more elaborate specifications. The coefficients σ_{ik} are estimated along with the other well identified parameters. Following (23) we write the model in our general form

$$Y_i = BX_i + A^{-1}(\alpha, \sigma, X_i)\epsilon_i \ , \tag{S56}$$

$$A^{-1}(\alpha, \sigma, X_i) = \begin{bmatrix} \sigma_u(X_i) + \alpha_1\sigma_v(X_i)\rho + \alpha_1\pi\alpha_2 & \alpha_1\sqrt{1-\rho^2}\sigma_v(X_i) & \alpha_1\pi\sigma_e(X_i) \\ \rho\sigma_v(X_i) + \pi'\alpha_2 & \sqrt{1-\rho^2}\sigma_v(X_i) & \pi'\sigma_e(X_i) \\ \alpha_2 & 0 & \sigma_e(X_i) \end{bmatrix} ,$$

which shows that the model is a special case of (S22). For this specification we reconstruct the confidence set for $\alpha = (\alpha_1, \alpha_2)$. The result is shown in Figure S5.

We find that the confidence region is quite similar when compared to the homoskedastic one. The volume is slightly smaller and there is more mass on the probability that α_2 is positive. Importantly however, the main conclusion remains the same. Even when relaxing the instrument validity assumption the effect of education is positive and quite precisely identified.

An obvious caveat is that this result is obtained under the additional assumption that the model for heteroskedasticity is correctly specified. An open question is how to handle model mis-specification in the class semi-parametric LSEM models. We leave this for future research.

References

Andrews, Donald W. K. 1987. "Asymptotic Results for Generalized Wald Tests." *Econometric Theory*, 3(3): 348–358.

Ben-Israel, A., and T. N. E. Greville. 2003. *Generalized Inverses: Theory and Applications.* New York, NY, USA:Springer.

Bernstein, D. S. 2009. *Matrix Mathmatics: Theory, Facts, and Formulas.* Princeton, NJ, USA:Princeton University Press.

Bhatia, R. 1997. *Matrix Analysis.* New York, NY, USA:Springer.

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.

Chen, A., and P. J. Bickel. 2006. "Efficient Independent Component Analysis." *Annals of Statistics*, 34(6): 2825–2855.

Choi, Sungsub, W. J. Hall, and Anton Schick. 1996. "Asymptotically uniformly most powerful tests in parametric and semiparametric models." *Annals of Statistics*, 24(2): 841–861.

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.

Dufour, Jean-Marie, and Pascale Valéry. 2016. "Rank-robust Regularized Wald-type tests." Working paper.

Durrett, Rick. 2019. *Probability Theory and Examples.* . 5th ed., Cambridge, UK:Cambridge University Press.

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.

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

Gouriéroux, C., A. Monfort, and J-P. Renne. 2017. “Statistical inference for independent component analysis: Application to structural VAR models.” *Journal of Econometrics*, 196: 111–126.

Gut, Allan. 2005. *Probability: A Graduate Course. Springer Texts in Statistics*, Springer.

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

Jin, K. 1992. “Empirical Smoothing Parameter Selection In Adaptive Estimation.” *Annals of Statistics*, 20(4): 1844–1874.

Magnus, J. R., and H. Neudecker. 2019. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley & Sons.

Newey, Whitney K. 1991. “Estimation of Tobit models under conditional symmetry.” In *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics. International Symposia in Economic Theory and Econometrics*, , ed. William A. Barnett, James Powell and George E. Tauchen. Cambridge University Press.

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

Rao, C. R., and S. K. Mitra. 1971. *Generalized Inverse of Matrices and its Applications.* New York, NY, USA:John Wiley & Sons, Inc.

Romano, Joseph P., and Michael Wolf. 2017. “Resurrecting weighted least squares.” *Journal of Econometrics*, 197(1): 1–19.

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

Rudin, W. 1991. *Functional analysis.* . 2 ed., McGraw Hill, Inc.

Strasser, Helmut. 1985. *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory. De Gruyter studies in mathematics*, W. de Gruyter.

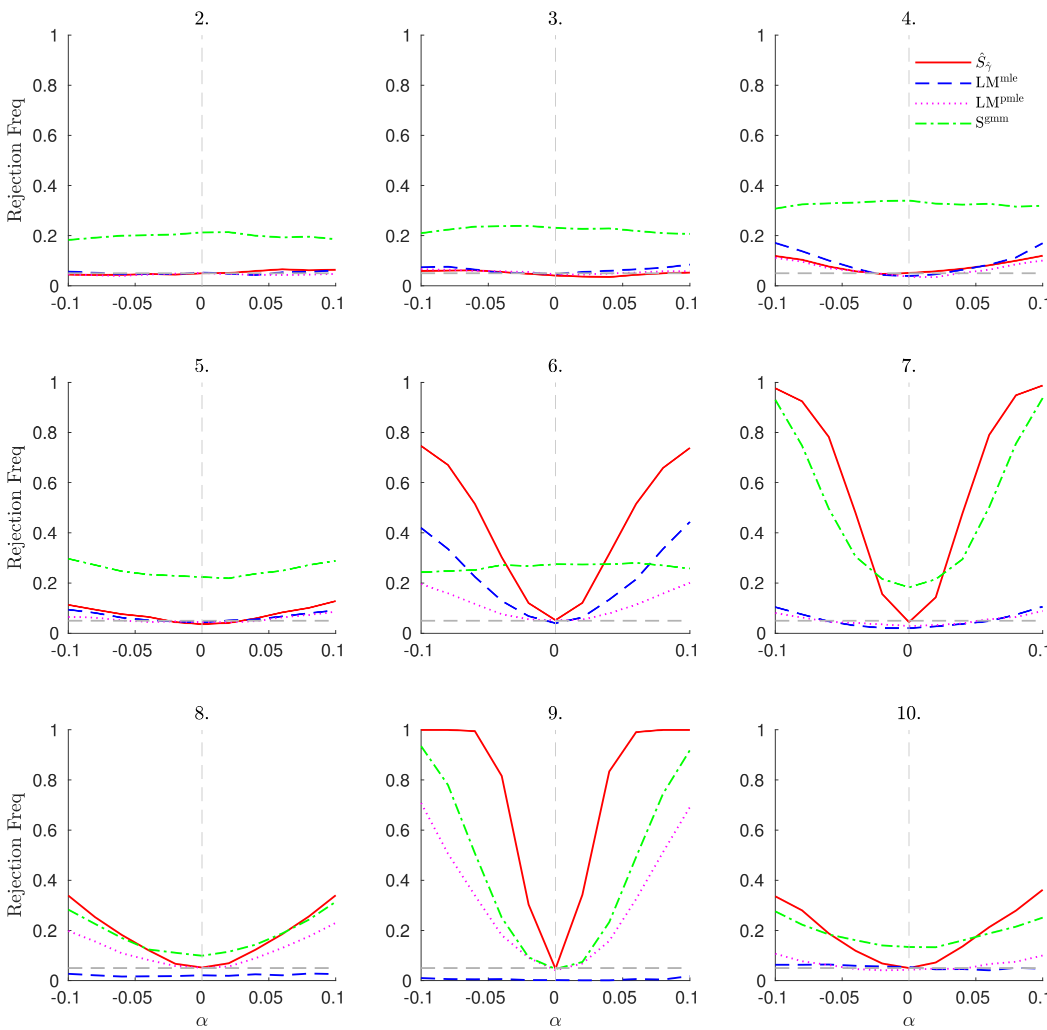
van der Vaart, A. W. 1988. *Statistical Estimation in Large Parameter Spaces. CWI Tracts*, Amsterdam:Centrum voor Wiskunde en Informatica.

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.

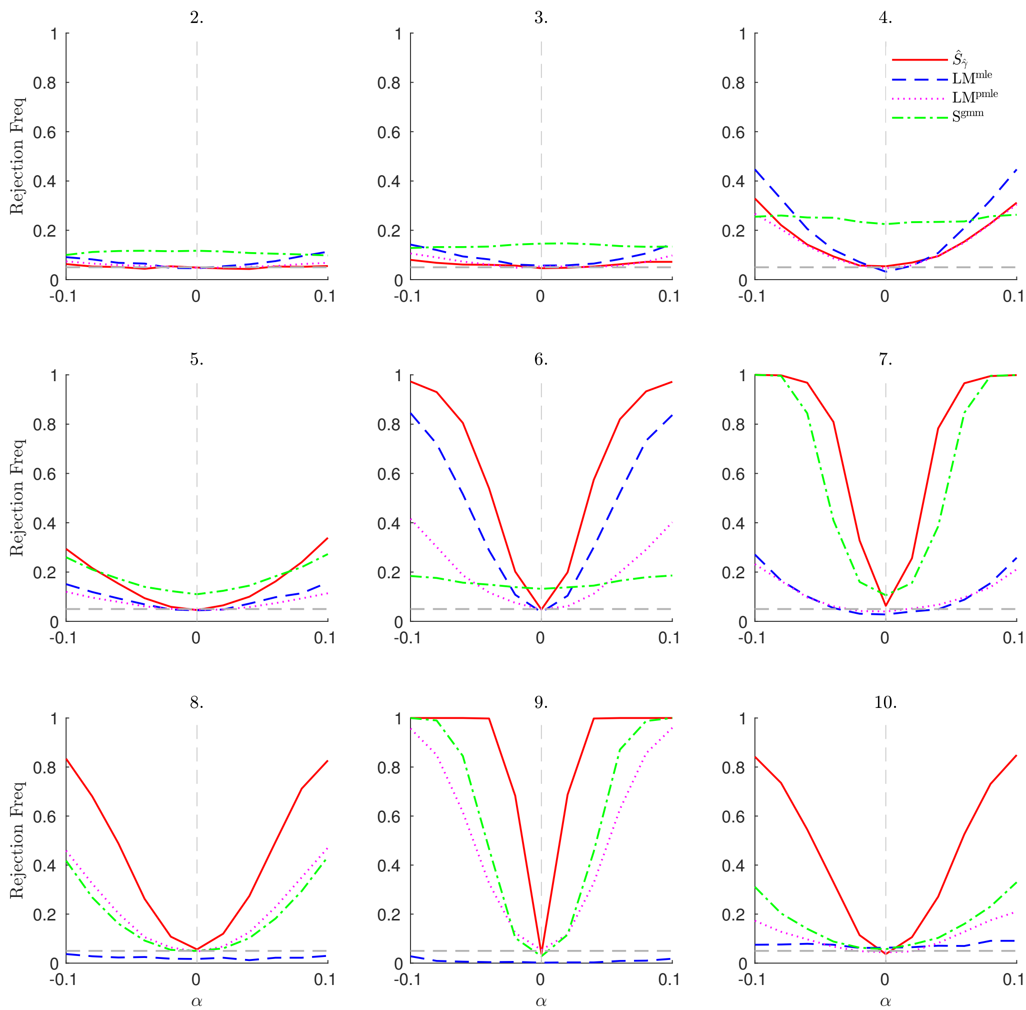
Wooldridge, Jeffrey Marc. 2012. *Introductory Econometrics: A Modern Approach (5th edition).* South-Western.

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



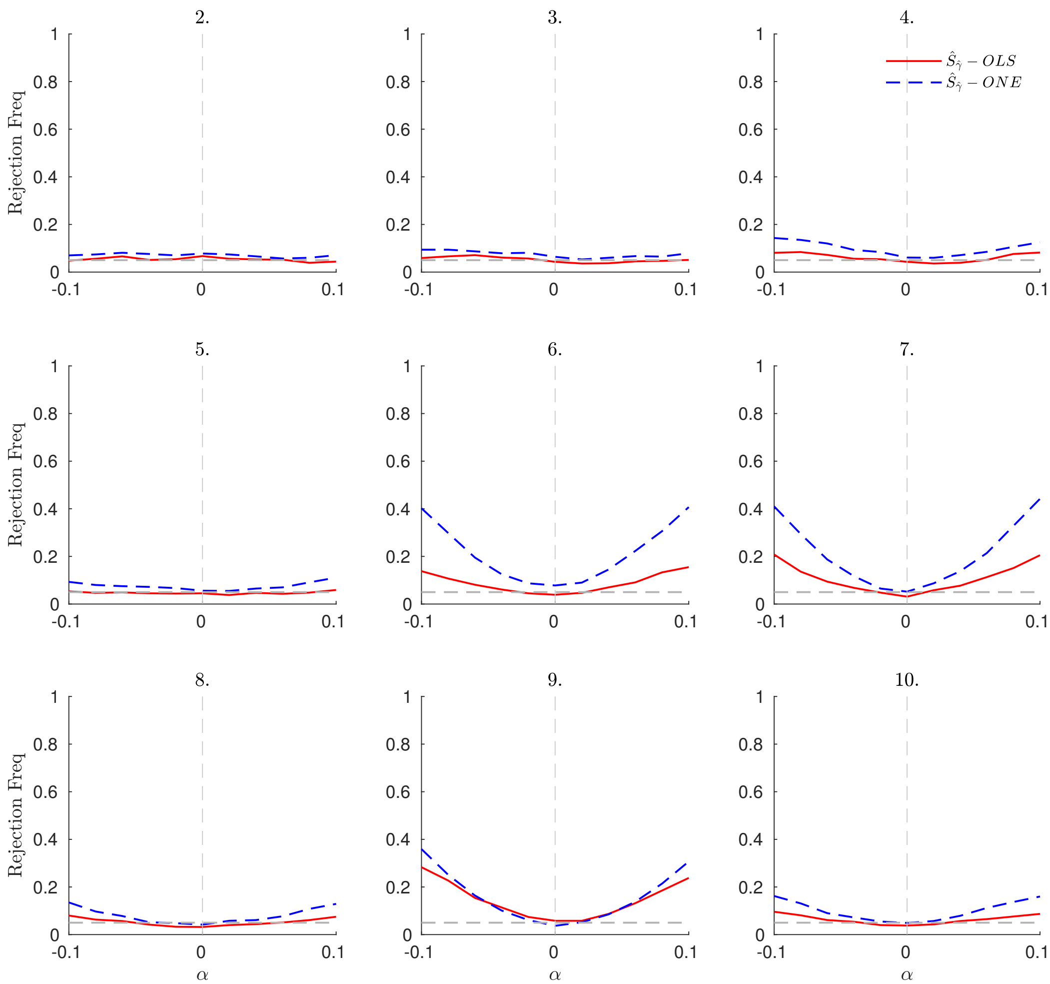
Notes: Empirical power curves for the baseline model with $k = 2$ and $n = 200$. Each plot corresponds to the choice for densities ϵ_k, for $k \geq 2$, where the numbers correspond to the different densities listed in Figure 3. The solid red line corresponds to $S_{\hat{\gamma}}$, the dashed blue line 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 $n = 500$



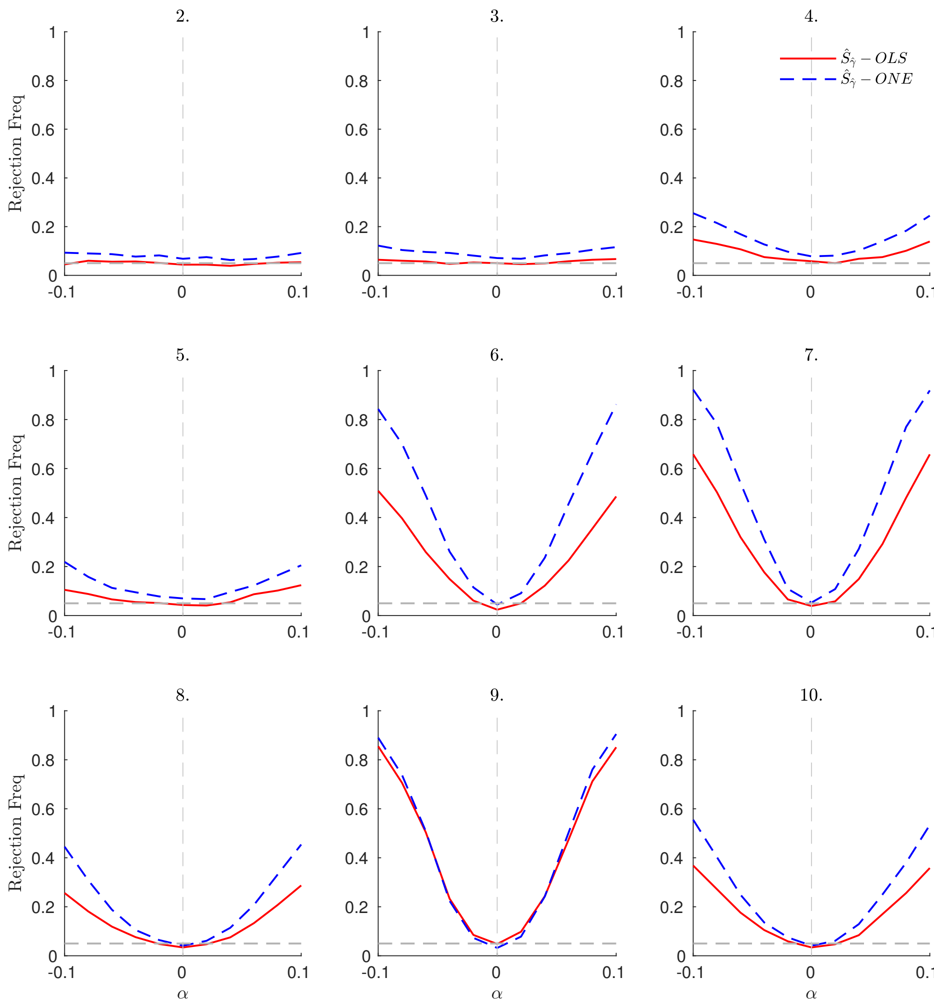
Notes: Empirical power curves for the baseline model with $k = 2$ and $n = 500$. Each plot corresponds to the choice for densities ϵ_k, for $k \geq 2$, where the numbers correspond to the different densities listed in Figure 3. The solid red line corresponds to $S_{\hat{\gamma}}$, the dashed blue line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

Figure S3: Power LSEM $n = 200$



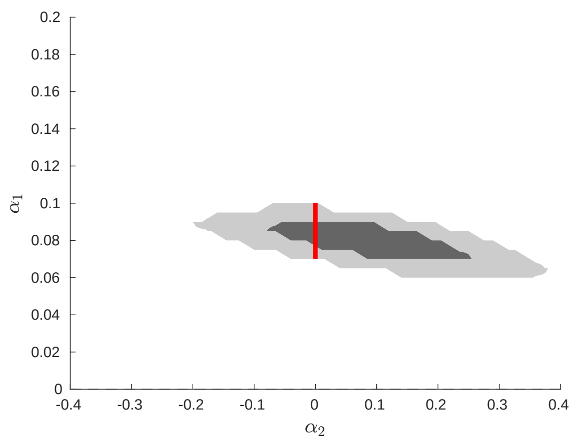
Notes: Empirical power curves for the LSEM model with $k = 2$, $d = 2$ and $n = 200$. Each plot corresponds to the choice for densities $\epsilon_{i,k}$, for $k \geq 2$, where the numbers correspond to the different densities shown in Figure 3. The solid red line corresponds to the empirical rejection 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 $n = 500$

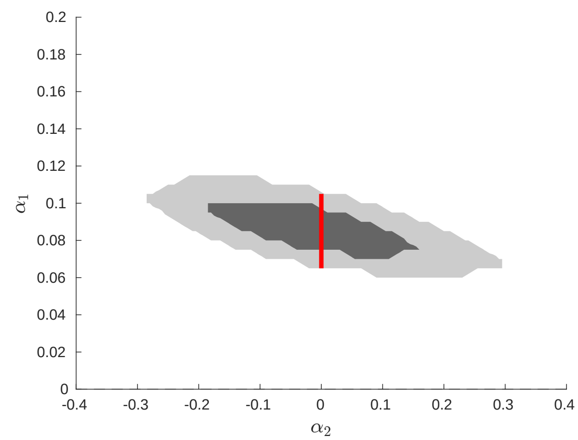


Notes: Empirical power curves for the LSEM model with $k = 2$, $d = 2$ and $n = 500$. Each plot corresponds to the choice for densities $\epsilon_{i,k}$, for $k \geq 2$, where the numbers correspond to the different densities shown in Figure 3. The solid red line corresponds to the empirical rejection 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.

Figure S5: Confidence sets: returns to schooling



(a) Heteroskedastic



(b) Homoskedastic

Notes: We show 95% (light gray) and 67% (dark gray) confidence sets for $\alpha = (\alpha_1, \alpha_2)$, where α_1 captures the effect of education on log wages and α_2 capture the correlation between the instrument (proximity to schooling interacted with parental education) and the error of the log wage equation. The red line indicates the confidence interval under the restriction of instrument exogeneity, i.e. $\alpha_2 = 0$. Figure (a) shows the result after inverting the $\hat{S}_{\hat{\gamma}}$ test statistic with heteroskedastic errors. Figure (b) shows the result after inverting the same test statistic but with homoskedastic errors.

Table S1: Rejection Frequencies $\hat{S}_{\hat{\gamma}}$ test for Baseline model: truncation

n	K	$\nu_n^{1/2}$	1	2	3	4	5	6	7	8	9	10
200	2	10^{-308}	0.051	0.047	0.048	0.041	0.050	0.049	0.047	0.049	0.050	0.044
200	2	10^{-5}	0.051	0.047	0.048	0.041	0.050	0.049	0.047	0.049	0.050	0.044
200	2	10^{-1}	0.051	0.047	0.048	0.041	0.050	0.049	0.047	0.049	0.050	0.044
200	3	10^{-308}	0.046	0.041	0.049	0.036	0.045	0.052	0.046	0.048	0.049	0.047
200	3	10^{-5}	0.046	0.041	0.049	0.036	0.045	0.052	0.046	0.048	0.049	0.047
200	3	10^{-1}	0.046	0.043	0.049	0.036	0.044	0.052	0.046	0.049	0.049	0.045
200	5	10^{-308}	0.034	0.040	0.037	0.037	0.034	0.044	0.041	0.048	0.044	0.042
200	5	10^{-5}	0.034	0.040	0.037	0.037	0.034	0.044	0.041	0.048	0.044	0.042
200	5	10^{-1}	0.041	0.039	0.040	0.036	0.037	0.047	0.042	0.050	0.044	0.040
500	2	10^{-308}	0.050	0.044	0.052	0.045	0.051	0.052	0.052	0.043	0.049	0.049
500	2	10^{-5}	0.050	0.044	0.052	0.045	0.051	0.052	0.052	0.043	0.049	0.049
500	2	10^{-1}	0.050	0.044	0.052	0.045	0.031	0.052	0.052	0.043	0.049	0.049
500	3	10^{-308}	0.048	0.046	0.040	0.047	0.050	0.055	0.054	0.047	0.051	0.048
500	3	10^{-5}	0.048	0.046	0.040	0.047	0.050	0.055	0.054	0.047	0.051	0.048
500	3	10^{-1}	0.038	0.048	0.042	0.045	0.050	0.055	0.054	0.047	0.051	0.051
500	5	10^{-308}	0.042	0.038	0.041	0.039	0.045	0.050	0.040	0.050	0.052	0.043
500	5	10^{-5}	0.042	0.038	0.041	0.039	0.045	0.050	0.040	0.050	0.052	0.043
500	5	10^{-1}	0.043	0.034	0.050	0.040	0.047	0.051	0.041	0.050	0.052	0.042
1000	2	10^{-308}	0.056	0.048	0.045	0.047	0.050	0.053	0.049	0.049	0.045	0.050
1000	2	10^{-5}	0.056	0.048	0.045	0.047	0.050	0.053	0.049	0.049	0.045	0.050
1000	2	10^{-1}	0.010	0.048	0.041	0.047	0.050	0.053	0.049	0.049	0.045	0.050
1000	3	10^{-308}	0.046	0.044	0.046	0.042	0.049	0.050	0.046	0.051	0.049	0.047
1000	3	10^{-5}	0.046	0.040	0.046	0.042	0.049	0.050	0.046	0.051	0.049	0.047
1000	3	10^{-1}	0.039	0.044	0.035	0.043	0.049	0.050	0.046	0.050	0.049	0.047
1000	5	10^{-308}	0.044	0.042	0.043	0.038	0.045	0.050	0.043	0.050	0.049	0.046
1000	5	10^{-5}	0.044	0.042	0.043	0.038	0.045	0.050	0.043	0.050	0.049	0.046
1000	5	10^{-1}	0.043	0.050	0.044	0.036	0.050	0.053	0.042	0.053	0.049	0.047

Notes: The table shows the empirical rejection frequencies for the $S_{\hat{\gamma}}$ test based on $S = 5,000$ Monte Carlo replications for the baseline model $Y_i = A^{-1}\epsilon_i$. The test has nominal level $a = 0.05$. The columns denote the sample size n, the dimension of the model K, the truncation rate $\nu_n^{1/2}$ and the choice for densities ϵ_{ik}, for $k \geq 2$, where the numbers correspond to the different densities shown in Figure 3.

Table S2: REJECTION FREQUENCIES $\hat{S}_{\hat{\gamma}}$ TEST FOR HETEROSKEDASTIC MODEL

n	K	d	1	2	3	4	5	6	7	8	9	10
200	2	2	0.061	0.061	0.065	0.072	0.054	0.053	0.054	0.040	0.056	0.045
200	2	3	0.063	0.069	0.070	0.085	0.067	0.061	0.058	0.047	0.062	0.051
200	3	2	0.074	0.088	0.092	0.127	0.076	0.071	0.081	0.047	0.081	0.056
200	3	3	0.079	0.093	0.103	0.145	0.080	0.078	0.082	0.044	0.081	0.065
200	5	2	0.126	0.167	0.197	0.279	0.132	0.097	0.068	0.056	0.057	0.080
200	5	3	0.151	0.180	0.209	0.307	0.151	0.107	0.065	0.062	0.059	0.080
500	2	2	0.050	0.060	0.057	0.075	0.058	0.054	0.035	0.045	0.061	0.051
500	2	3	0.054	0.060	0.062	0.079	0.063	0.055	0.040	0.048	0.052	0.050
500	3	2	0.061	0.074	0.079	0.110	0.060	0.063	0.044	0.046	0.078	0.051
500	3	3	0.070	0.079	0.084	0.115	0.064	0.058	0.052	0.048	0.074	0.050
500	5	2	0.084	0.113	0.139	0.201	0.091	0.075	0.050	0.060	0.097	0.069
500	5	3	0.094	0.132	0.158	0.229	0.095	0.090	0.047	0.053	0.091	0.061
1000	2	2	0.059	0.060	0.057	0.066	0.053	0.050	0.026	0.040	0.057	0.045
1000	2	3	0.055	0.055	0.062	0.072	0.049	0.053	0.027	0.046	0.054	0.053
1000	3	2	0.056	0.062	0.069	0.087	0.056	0.056	0.030	0.047	0.072	0.050
1000	3	3	0.053	0.067	0.076	0.102	0.054	0.055	0.035	0.045	0.065	0.057
1000	5	2	0.071	0.092	0.101	0.150	0.074	0.051	0.048	0.042	0.051	0.051
1000	5	3	0.072	0.092	0.100	0.145	0.071	0.052	0.049	0.046	0.052	0.050

Notes: The table shows the empirical rejection frequencies for the $S_{\hat{\gamma}}$ test based on $S = 5,000$ Monte Carlo replications for the heteroskedastic model $Y_i = A(\alpha, \sigma, X_i)^{-1}\epsilon_i$. The test has nominal level $a = 0.05$. The columns denote the sample size n, the dimension of the model K, the number explanatory variables d and the choice for densities ϵ_{ik}, for $k \geq 2$, where the numbers correspond to the different densities shown in Figure 3.