Supplementary material for "Locally regular and efficient tests in potentially non-regular semiparametric models"

Adam Lee*

February 7, 2024

Abstract

This Supplementary Material contains the following sections:

S1: Details on notation
S2: Additional results
S3: Additional details and proofs for the examples
S4: Additional simulation details & results
S5: Tables and Figures

*BI Norwegian Business School, adam.lee@bi.no.

S1 Notation

$x \coloneqq y$ means that x is defined to be y. The Lebesgue measure on $\mathbb{R}^K$ is denoted by λ_K or λ if the dimension is clear from context. The standard basis vectors in $\mathbb{R}^K$ are $e_1, \ldots, e_K$. For any matrix M, $M^\dagger$ is its Moore – Penrose pseudoinverse. We 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. If S is a subset of a vector space, $\operatorname{lin} S$ or $\operatorname{Span} S$ means the linear span of S. If S is a subset of a topological vector space, $\overline{\operatorname{lin}}\ S$ or $\operatorname{cl}\operatorname{Span} S$ means the closure of the linear span of S. 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 $\Pi(x|S)$. The total variation distance between measures P and Q defined on the measurable space $(\Omega, \mathcal{F})$ is $d_{TV}(P, Q) = \sup_{A\in\mathcal{F}} |P(A) - Q(A)|$. $\overset{P_n}{\rightsquigarrow}$ denotes weak convergence under the sequence of measures P_n. If the sequence of measures is clear from context, we write just $\rightsquigarrow$.

S2 Additional results

S2.1 A consistent estimator of the Moore – Penrose pseudoinverse

As is well known, the Moore – Penrose pseudoinverse 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 record a construction given in the supplementary appendix to Lee and Mesters (2024a) 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 $M^\dagger$. This construction requires an initial estimator with a known rate of convergence and is based on a spectral cut-off regularisation scheme. It is very similar to that considered in Lütkepohl and Burda (1997) and is a special case of the larger class of regularisation schemes considered by Dufour and Valéry (2016). That this results in an estimator with the claimed properties is recorded in Proposition S2.1 below, which is proven in Lee and Mesters (2024b).[1]

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

$$P_n\left(\|\check{M}_n - M_n\|_2 < \nu_n\right) \to 1, \tag{S1}$$

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.[2] Let $\check{M}_n = \check{U}_n \check{\Lambda}_n \check{U}_n'$ be the corresponding eigendecompositions and define

$$\hat{M}_n := \check{U}_n \Lambda_n(\nu_n) \check{U}_n' , \tag{S2}$$

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 S2.1 (Proposition S1 in Lee and Mesters (2024a)): *If* (S1) *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*

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

[1] Dufour and Valéry (2016) prove an analogous result (their Proposition 9.1) for a broader class of regularisation schemes. However the statement of their result involves an additional rate term (which satisfies their Assumption 2) as compared to the result stated in Proposition S2.1.

[2] (S1) 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 $n \in \mathbb{N}$.

where $\hat{M}_n$ is defined as in (S2). In consequence,

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

S2.2 The quotient space $\mathbb{H}_\gamma$

I first briefly recall some preliminaries regarding quotient spaces, for the convenience of the reader. Following this a lemma used during the development of the power bounds is established.

S2.2.1 Preliminaries on quotient spaces

Let X be a linear space and V a subspace of X. The quotient of X by V, $X \,/\, V$ is a linear space whose elements are the cosets $[x] \coloneqq x + V \coloneqq \{x + n : n \in V\}$ (for $x \in X$). x (or any other member of $[x]$) is a coset representative. The zero vector in $X \,/\, V$ is $[0] = 0 + V = V$.

Vector addition and scalar multiplication are defined according to:

$$[x+y] \coloneqq [x] + [y], \quad [ax] \coloneqq a[x] \quad \text{for all } [x],[y] \in X \,/\, V \text{ and all } a \in \mathbb{R}.$$

The map $\pi_V : X \to X \,/\, V$ defined by $\pi_V(x) \coloneqq [x]$ is the *natural projection* or *quotient map*. π_V is a surjective linear transformation with $\ker \pi_N = V$ (Roman, 2005, Theorem 3.2).

There are two main cases of interest in the present paper. In the first, X is a linear space equipped with a positive semi-definite symmetric bilinear form, $\langle \cdot \,,\, \cdot \rangle$. Let $\|\cdot\|$ be the corresponding semi-norm formed in the usual way: $\|x\| \coloneqq \sqrt{\langle x \,,\, x \rangle}$ and let $V \coloneqq \{x \in X : \|x\| = 0\}$, which is evidently a subspace of X. We define an inner product on $X \,/\, V$ as follows:

$$\langle [x] \,,\, [y] \rangle_V \coloneqq \langle x \,,\, y \rangle \,.$$

Symmetry and linearity in the first argument of $\langle \cdot \,,\, \cdot \rangle_V$ follow from the corresponding properties of $\langle \cdot \,,\, \cdot \rangle$. For positive definiteness, suppose that $[x] \neq [0]$. Then, $\langle x \,,\, x \rangle_V = \|x\|^2 > 0$ since $x \notin \ker \pi_V = V$. In consequence $(X \,/\, V, \langle \cdot \,,\, \cdot \rangle_V)$ is an inner product (pre-Hilbert) space. The induced norm on $X \,/\, V$ evidently satisfies $\|[x]\|_V = \|\pi_V(x)\|_V = \|x\|$.

In the second case of interest, X is a linear space equipped with a semi-norm $\|\cdot\|$. Let $V \coloneqq \{x \in X : \|x\| = 0\}$, which is evidently a subspace of X. We define a norm on $X\,/\,V$ as follows:

$$\|[x]\|_V \coloneqq \|x\|.$$

This definition ensures that $(X\,/\,V, \|\cdot\|_V)$ is a normed space (Rudin, 1991, 1.43).

S2.2.2 A lemma on the kernel of π_1

Lemma S2.1: *Suppose Assumption 3.1 holds and B_η is a linear space. Let π_1' denote the restriction of π_1 to $\mathbb{H}_\gamma$. Then, the closure of $\ker\pi_1'$ in $\overline{\mathbb{H}_\gamma}$ is $\ker\pi_1$.*

Proof. Since π_1 is continuous, $\ker\pi_1 = \pi_1^{-1}(\{0\})$ is closed. Hence it suffices to show that

$$\ker\pi_1 = \{[h] \in \overline{\mathbb{H}_\gamma} : [h] = [0, b]\} \subset \operatorname{cl}\ker\pi_1' = \operatorname{cl}\{[h] \in \mathbb{H}_\gamma : [h] = [0, b]\}.$$

Let $[h] = [0, b] \in \ker\pi_1$. There is a sequence $\mathbb{H}_\gamma \ni [h_n] = [t_n, b_n] \to [h]$. Decomposing the norm, we have that

$$\begin{aligned}
\|[h_n] - [h]\|_\gamma &= \|\Pi^\perp[h_n] - \Pi^\perp[h]_\gamma\|_\gamma + \Pi[h_n] - \Pi[h]\|_\gamma \\
&= t_n'\tilde{\mathcal{I}}_\gamma t_n + \|\Pi[t_n, 0] + [0, b_n] - [0, b]\|_\gamma \\
&= t_n'\tilde{\mathcal{I}}_\gamma t_n + \left\| \sum_{j=1}^{d_\theta} t_{n,j}\Pi[e_j, 0] + [0, b_n] - [0, b] \right\|_\gamma \\
&= t_n'\tilde{\mathcal{I}}_\gamma t_n + \|t_n'\boldsymbol{e} + [0, b_n] - [0, b]\|_\gamma,
\end{aligned}$$

with $\boldsymbol{e} = (\Pi[e_1, 0], \ldots, \Pi[e_{d_\theta}, 0])'$, where e_j is the j-th canonical basis vector in $\mathbb{R}^{d_\theta}$. For each $n \in \mathbb{N}$, there are $\check{\boldsymbol{e}}_n = ([0, \check{b}_{1,n}], \ldots, [0, \check{b}_{d_\theta,n}])'$ with each $[0, \check{b}_{j,n}] \in \mathbb{H}_\gamma$ such that

$$\left\|[0, \check{b}_{j,n}] - \Pi[e_j, 0]\right\| \le \frac{1}{n|t_{n,j}|}.$$

Hence, putting $[0, \tilde{b}_n] \coloneqq t_n'\check{\boldsymbol{e}}_n + [0, b_n]$, we have

$$\|[0, \tilde{b}_n] - [0, b]\|_\gamma \le \|t_n'\check{\boldsymbol{e}}_n - t_n'\boldsymbol{e}\|_\gamma + \|t_n'\boldsymbol{e} + [0, b_n] - [0, b]\|_\gamma \le \frac{d_\theta}{n} + o(1) = o(1).$$

Since each $[0, \tilde{b}_n] \in \ker\pi_1'$, the limit $[0, b] \in \operatorname{cl}\ker\pi_1'$. □

S2.3 Uniform Local Asymptotic Normality

The equivalence discussed in Remark 3.2 is proved in Proposition S2.2 below, which is an adaptation of Theorem 80.13 in Strasser (1985). H_γ is assumed to be a subset (containing 0) of a linear space equipped with some pseudometric.

Assumption S2.1 (Uniform local asymptotic normality): *$L_{n,\gamma}(h)$ satisfies*

$$L_{n,\gamma}(h) = \Delta_{n,\gamma} h - \frac{1}{2}\|\Delta_{n,\gamma} h\|^2 + R_{n,\gamma}(h),$$

where $h = (\tau, b)$, $\Delta_{n,\gamma} : \overline{\text{lin}}\ H_\gamma \to L_2^0(P_{n,\gamma})$ are bounded linear maps and for any $h_n \to h$ in H_γ, $R_n(\gamma, h_n) \xrightarrow{P_{n,\gamma}} 0$. Additionally, suppose that for each $h_n \to h$ in H_γ, $(\Delta_{n,\gamma} h_n)_{n\in\mathbb{N}}$ is uniformly square $P_{n,\gamma}$-integrable and

$$(\Delta_{n,\gamma} h_n, \Delta_{n,\gamma} h)' \overset{P_{n,\gamma}}{\rightsquigarrow} \mathcal{N}\left(0, \sigma_\gamma(h) \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]\right), \qquad \sigma_\gamma(h) := \lim_{n\to\infty} \|\Delta_{n,\gamma} h\|^2.$$

Remark S2.1: *Assumption S2.1 ensures that the pairs of sequences $(P_{n,\gamma})_{n\in\mathbb{N}}$ and $(P_{n,\gamma,h_n})_{n\in\mathbb{N}}$ are mutually contiguous for any $h_n \to h \in H_\gamma$ (see e.g. van der Vaart, 1998, Example 6.5).*

Remark S2.2: *In Assumption S2.1, the assumption of joint convergence of $(\Delta_n h_n, \Delta_n h)'$ is nedeed only because H_γ is not required to be linear. If H_γ is a linear space this follows from the Cramér – Wold Theorem given the definition of $\sigma_\gamma(h)$.*

Remark S2.3: *If $(\Delta_{n,\gamma})_{n\in\mathbb{N}}$ is asymptotically equicontinuous on compact subsets $K \subset H_\gamma$, then $h_n \to h$ in H_γ implies $\|\Delta_{n,\gamma}(h_n - h)\| \to 0$. In consequence $(\Delta_{n,\gamma} h)_{n\in\mathbb{N}}$ being uniformly square $P_{n,\gamma}$-integrable and $\Delta_{n,\gamma} h \overset{P_{n,\gamma}}{\rightsquigarrow} \mathcal{N}(0, \sigma_\gamma(h))$ for each $h \in H_\gamma$, suffices for $(\Delta_{n,\gamma} h_n)_{n\in\mathbb{N}}$ being uniformly square $P_{n,\gamma}$-integrable and*

$$\begin{aligned}(\Delta_{n,\gamma} h_n,\ \Delta_{n,\gamma} h)' &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} (\Delta_{n,\gamma} h_n - \Delta_{n,\gamma} h,\ \Delta_{n,\gamma} h)' \\ &\overset{P_{n,\gamma}}{\rightsquigarrow} \mathcal{N}\left(0, \left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 0 & 0 \\ 0 & \sigma_\gamma(h) \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]'\right) \\ &= \mathcal{N}\left(0, \sigma_\gamma(h) \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]\right),\end{aligned}$$

for any $h_n \to h \in H_\gamma$.[3]

If H_γ is a Banach space metrised by its norm, the equicontinuity of $(\Delta_{n,\gamma})_{n\in\mathbb{N}}$ is guaranteed as uniform boundedness of $(\Delta_{n,\gamma})_{n\in\mathbb{N}}$ (hence equicontinuity on H_γ) is implied by uniform square $P_{n,\gamma}$-integrability of $(\Delta_{n,\gamma} h)_{n\in\mathbb{N}}$ for $h \in H_\gamma$.

Proposition S2.2: *Assumption S2.1 is equivalent to Assumption 3.1 plus asymptotic equicontinuity on compact subsets $K \subset H_\gamma$ of $(\Delta_{n,\gamma})_{n\in\mathbb{N}}$ and $(h \mapsto P_{n,\gamma,h})_{n\in\mathbb{N}}$, where the metric on the relevant space of probability measures is d_{TV}.*

[3] Each $\Delta_{n,\gamma} h_n \in L_2(P_{n,\gamma})$ by definition.

Proof. Suppose first that Assumption 3.1 and the asymptotic equicontinuity conditions hold. Let $h_n \to h$ in H_γ. By asymptotic equicontinuity of $(h \mapsto P_{n,\gamma,h})_{n\in\mathbb{N}}$,

$$\lim_{n\to\infty} d_{TV}(P_{n,\gamma,h_n}, P_{n,\gamma,h}) = 0 \quad \implies \quad \lim_{n\to\infty} \int \left| \frac{p_{n,\gamma,h_n}}{p_{n,\gamma,0}} - \frac{p_{n,\gamma,h}}{p_{n,\gamma,0}} \right| \mathrm{d}P_{n,\gamma,0} = 0.$$

In combination with (compact) asymptotic equicontinuity of $(\Delta_{n,\gamma})_{n\in\mathbb{N}}$, this yields

$$R_{n,\gamma}(h_n) - R_{n,\gamma}(h) = L_{n,\gamma}(h_n) - L_{n,\gamma}(h) + o_{P_{n,\gamma}}(1) = o_{P_{n,\gamma}}(1).$$

That $(\Delta_{n,\gamma}h_n)_{n\in\mathbb{N}}$ is uniformly square $P_{n,\gamma}$-integrable and the required joint weak convergence under $P_{n,\gamma}$ follows from the asymptotic equicontinuity of $(\Delta_{n,\gamma})_{n\in\mathbb{N}}$ on compacts as discussed in Remark S2.3.

For the converse suppose that Assumption S2.1 holds. We need to prove only the asymptotic equicontinuity conditions. It suffices to show that (i) $\|\Delta_{n,\gamma}(h_n - h)\| \to 0$ and (ii) $d_{TV}(P_{n,\gamma,h_n}, P_{n,\gamma,h})$ for any $h_n \to h$ with $h_n, h \in K \subset H_\gamma$. (i) holds since for any convergent $h_n \to h \in H_\gamma$ we have $\Delta_{n,\gamma}(h_n - h) = (1,-1)(\Delta_{n,\gamma}h_n, \Delta_{n,\gamma}h)' \xrightarrow{P_{n,\gamma}} 0$ and so by the square uniform integrability and e.g. Theorem 2.7 in Serfozo (1982), $\|\Delta_{n,\gamma}(h_n - h)\|^2 \to 0$. That (ii) holds follows from Lemma S2.4 and

$$\begin{aligned} L_{n,\gamma}(h_n) - L_{n,\gamma}(h) &= \Delta_{n,\gamma}h_n - \frac{1}{2}\|\Delta_{n,\gamma}h_n\|^2 + R_{n,\gamma}(h_n) - \left[\Delta_{n,\gamma}h - \frac{1}{2}\|\Delta_{n,\gamma}h\|^2 + R_{n,\gamma}(h)\right] \\ &= o_{P_{n,\gamma}}(1), \end{aligned}$$

since $R_{n,\gamma}(h_n) = o_{P_{n,\gamma}}(1)$, $R_{n,\gamma}(h) = o_{P_{n,\gamma}}(1)$ and $\|\Delta_{n,\gamma}(h_n - h)\|^2 \to 0$. □

The following Lemmas provide conditions which can be useful for demonstrating (compact) equicontinuity in the i.i.d. case.

Lemma S2.2: *If H_γ is a linear space and $\Delta_{n,\gamma}h = \frac{1}{\sqrt{n}}\sum_{i=1}^n A_\gamma h$ for some bounded linear map $A_\gamma : H_\gamma \to L_2(P_\theta)$, then $(h \mapsto \Delta_{n,\gamma}h)_{n\in\mathbb{N}}$ is equicontinuous on compact subsets of H_γ in $L_2(P_\theta)$.*

Proof. Let $K \subset H_\gamma$ be compact and note that for any $h_n \to h$ (all in K),

$$\|\Delta_{n,\gamma}(h_n - h)\| = \|\mathbb{G}_n A_\gamma(h_n - h)\| = \|A_\gamma(h_n - h)\| \leq \|A_\gamma\|\|h_n - h\| \to 0. \quad \square$$

Lemma S2.3: *Let Γ, Γ' be subsets of linear spaces, H a subset of a seminormed linear space H' and $\mathcal{W} \subset \mathbb{R}^K$. Suppose $\mathcal{V}$ is the intersection of a neighbourhood of 0 in H' with H and $\varphi : H \mapsto \Gamma'$ is a linear map such that $\gamma + s\varphi(h) \in \Gamma$ for* *each $s \in [0,1]$, $h \in \mathcal{V}$ and $\gamma \in \Gamma$. Let $\mathcal{U} := \{\gamma + \varphi(h) : h \in \mathcal{V}\}$. If*

(i) For each $\gamma \in \Gamma$, P_γ is a probability measure on $(\mathcal{W}, \mathcal{B}(\mathcal{W}))$, dominated by a σ – finite measure ν, with corresponding density p_γ;

(ii) $t \mapsto \sqrt{p_{\gamma_\star + t\varphi(h)}(w)}$ is absolutely continuous on $[0,1]$ for all $\gamma_\star \in \mathcal{U}$, $h \in \mathcal{V}$ and $w \in \mathcal{W}$;

(iii) For all $\gamma_\star \in \mathcal{U}$, $h \in \mathcal{V}$, $q_{\gamma_\star,h,s}(w) = \frac{\partial \log p_{\gamma_\star + t\varphi(h)}(w)}{\partial t}|_{t=s}$ satisfies

$$P_{\gamma_\star + s\varphi(h)}[q_{\gamma_\star,h,s}]^2 \leq m(s)\|h\|^2,$$

for $m : [0,1] \to [0,\infty)$ with $\int_0^1 m(s)\,\mathrm{d}s < \infty$.[4]

Then there is an N such that for all $n \geq N$, the functions $h \mapsto P_{n,\gamma,h} := P^n_{\gamma+\varphi(h/\sqrt{n})}$ are d_{TV}–Lipschitz with a common Lipschitz constant. Consequently, $\{h \mapsto P_{n,\gamma,h} : n \geq N\}$ is uniformly equicontinuous in d_{TV}.

Proof. Let $h^\star \in H$ and let $\tilde{h} := (h^\star - h)$. For all large enough n, $h/\sqrt{n}$ and $\tilde{h}/\sqrt{n}$ belong to $\mathcal{V}$ and hence also $\gamma_{0,n} := \gamma + \varphi(h/\sqrt{n}) \in \mathcal{U}$. Therefore, by (ii) $q_{\gamma_{0,n},\tilde{h}/\sqrt{n},s}(w)$ exists for almost every $s \in (0,1)$ and

$$\begin{aligned}\sqrt{p_{\gamma+\varphi(h^\star/\sqrt{n})}(w)} - \sqrt{p_{\gamma+\varphi(h/\sqrt{n})}(w)} &= \sqrt{p_{\gamma_{0,n}+\varphi(\tilde{h}/\sqrt{n})}(w)} - \sqrt{p_{\gamma_{0,n}}(w)} \\ &= \frac{1}{2}\int_0^1 q_{\gamma_{0,n},\tilde{h}/\sqrt{n},s}(w)\sqrt{p_{\gamma_{0,n}+s\varphi(\tilde{h}/\sqrt{n})}(w)}\,\mathrm{d}s.\end{aligned}$$

By Fubini's Theorem, Jensen's inequality and (iii),

$$\begin{aligned}\int\left(\sqrt{p_{\gamma+\varphi(h^\star/\sqrt{n})}(w)} - \sqrt{p_{\gamma+\varphi(h/\sqrt{n})}(w)}\right)^2 \mathrm{d}\nu &\leq \frac{1}{4}\int\int_0^1 q_{\gamma_{0,n},\tilde{h}/\sqrt{n},s}(w)^2 p_{\gamma_{0,n}+s\varphi(\tilde{h}/\sqrt{n})}(w)\,\mathrm{d}s\,\mathrm{d}\nu \\ &= \frac{1}{4}\int_0^1\int q_{\gamma_{0,n},\tilde{h}/\sqrt{n},s}(w)^2 p_{\gamma_{0,n}+s\varphi(\tilde{h}/\sqrt{n})}(w)\,\mathrm{d}\nu\,\mathrm{d}s \\ &\leq \frac{1}{4}\int_0^1 m(s)\frac{\|\tilde{h}\|^2}{n}\,\mathrm{d}s \\ &= \frac{M\|h^\star - h\|^2}{4n},\end{aligned}$$

where $M := \int_0^1 m(s)\,\mathrm{d}s$. Therefore, by Lemmas 2.15 and 2.17 of Strasser (1985),

$$d_{TV}(P_{n,\gamma,h^\star}, P_{n,\gamma,h}) \leq \frac{\sqrt{2M}}{2}\|h^\star - h\|. \qquad \square$$

[4] $q_{\gamma_\star,h,s}(w)$ exists almost everywhere on $(0,1)$ under (ii).

S2.4 Convergence of log-likelihood ratios and convergence in total variation

LEMMA S2.4: *Suppose that for* $h_n, g \in H_\gamma$, $P_{n,\gamma,g} \lhd P_{n,\gamma}$ *and*

$$L_{n,\gamma}(h_n) - L_{n,\gamma}(g) = o_{P_{n,\gamma}}(1).$$

Then $d_{TV}(P_{n,\gamma,h_n}, P_{n,\gamma,g}) \to 0$.

Proof. By the continuous mapping theorem and Le Cam's first lemma (e.g. van der Vaart, 1998, Lemma 6.4),

$$\frac{p_{n,\gamma,h_n}}{p_{n,\gamma,g}} = \exp\left(L_{n,\gamma}(h_n) - L_{n,\gamma}(g)\right) \xrightarrow{P_{n,\gamma,g}} 1.$$

By Le Cam's first lemma again, $P_{n,\gamma,h_n} \lhd P_{n,\gamma,g}$. 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 $P_{n_m,\gamma,g}$. Therefore by Slutsky's Theorem

$$\left(\phi_{n_m}, \frac{p_{n_m,\gamma,h_{n_m}}}{p_{n_m,\gamma,g}}\right) \rightsquigarrow (\phi, 1) \quad \text{under } P_{n_m,\gamma,g}.$$

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

$$\lim_{m\to\infty}\left[P_{n_m,\gamma,h_{n_m}}\phi_{n_m} - P_{n_m,\gamma,g}\phi_{n_m}\right] = 0.$$

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

COROLLARY S2.1: *Suppose that Assumption 3.1 holds and* H_η *is a linear space equipped with the semi-norm* $\|\cdot\|_\gamma$. *Then, if* $h, g \in H_\gamma$ *are such that* $\|h-g\|_\gamma = 0$, $d_{TV}(P_{n,\gamma,h}, P_{n,\gamma,g}) \to 0$.

Proof. By Assumption 3.1, the reverse triangle inequality and $\sigma_\gamma(h-g) = \|h-g\|_\gamma$ we have that $L_{n,\gamma}(h) - L_{n,\gamma}(g) = o_{P_{n,\gamma}}(1)$. Noting Remark 3.1, apply Lemma S2.4 with each $h_n = h$. □

S2.5 Orthogonal projections

Proposition S2.3: *Let H_1 and H_2 be subspaces of a Hilbert space H. For any $h \in H$, let $\breve{h} := \Pi[h|H_1^{\perp}]$. Then,*

$$\Pi\left[h|(H_1+H_2)^{\perp}\right] = \breve{h} - \Pi\left[h\middle|\mathrm{cl}(H_1+H_2)\cap H_1^{\perp}\right] = \breve{h} - \Pi\left[\breve{h}\middle|\mathrm{cl}(H_1+H_2)\cap H_1^{\perp}\right].$$

Proof. That the last equality in the display holds is an immediate consequence of the fact that $\breve{h} = h - \Pi[h|H_1]$ and $H_1 \perp H_1^{\perp}$ by definition. For the first equality, by (A.2.11) of Proposition A.2.4 in Bickel, Klaassen, Ritov, and Wellner (1998),

$$\begin{aligned}\Pi\left[h|\,\mathrm{cl}(H_1+H_2)\right] &= \Pi\left[h|\,\mathrm{cl}(H_1+H_2)\cap[\mathrm{cl}\,H_1]^{\perp}\right] + \Pi\left[h|\,\mathrm{cl}\,H_1\right] \\ &= \Pi\left[h|\,\mathrm{cl}(H_1+H_2)\cap H_1^{\perp}\right] + \Pi\left[h|\,\mathrm{cl}\,H_1\right].\end{aligned}$$

Hence

$$\begin{aligned}\Pi\left[h|(H_1+H_2)^{\perp}\right] &= h - \Pi\left[h|\,\mathrm{cl}(H_1+H_2)\right] \\ &= h - \Pi\left[h|\,\mathrm{cl}\,H_1\right] - \Pi\left[h|\,\mathrm{cl}(H_1+H_2)\cap H_1^{\perp}\right] \\ &= \breve{h} - \Pi\left[h|\,\mathrm{cl}(H_1+H_2)\cap H_1^{\perp}\right].\end{aligned}$$

□

Proposition S2.4: *Let H_1 and H_2 be subspaces of a Hilbert space H. Define $\breve{h} := \Pi[h|H_1^{\perp}]$ for $h \in H$. Then*

$$\mathrm{cl}(H_1+H_2)\cap H_1^{\perp} = \mathrm{cl}\{\breve{h} : h \in H_2\}.$$

Proof. Firstly let $f \in \mathrm{cl}(H_1+H_2)\cap H_1^{\perp}$. So there are scalars a_i, such that $f = \sum_{i=1}^{\infty} a_i(h_{1,i}+h_{2,i})$ with each $h_{1,i} \in H_1$, $h_{2,i} \in H_2$ and $f \in H_1^{\perp}$. Let $g := -\sum_{i=1}^{\infty} a_i \Pi[h_{2,i}|\,\mathrm{cl}\,H_1]$. Suppose that $f_1 := \sum_{i=1}^{\infty} a_i h_{1,i} \neq g$. Then if $e := f_1 - g$ one has $e \in \mathrm{cl}\,H_1$ and $\|e\| > 0$. Therefore,

$$\begin{aligned}\langle f\,,\,e\rangle &= \sum_{i=1}^{\infty} a_i\,\langle h_{2,i}\,,\,e\rangle - \sum_{i=1}^{\infty} a_i\,\langle \Pi[h_{2,i}|\,\mathrm{cl}\,H_1]\,,\,e\rangle + \langle e\,,\,e\rangle \\ &= \sum_{i=1}^{\infty} a_i\left\langle \breve{h}_{2,i}\,,\,e\right\rangle + \langle e\,,\,e\rangle \\ &= \langle e\,,\,e\rangle \\ &> 0.\end{aligned}$$

But this contradicts the fact that that $f \in H_1^\perp$ as $e \in \operatorname{cl} H_1$. Hence

$$f = \sum_{i=1}^{\infty} a_i \left[h_{2,i} - \Pi[h_{2,i}|\operatorname{cl} H_1]\right] = \sum_{i=1}^{\infty} a_i \breve{h}_{2,i},$$

thus f is in $\operatorname{cl}\{\breve{h} : h \in H_2\}$.[5]

Conversely, let $f \in \operatorname{cl}\{\breve{h} : h \in H_2\}$. Then

$$f = \sum_{i=1}^{\infty} a_i \breve{h}_{2,i} = \sum_{i=1}^{\infty} a_i \left[h_{2,i} - \Pi[h_{2,i}|\operatorname{cl} H_1]\right]$$

Since each $\breve{h}_{2,i} = h_{2,i} - \Pi[h_{2,i}|\operatorname{cl} H_1] \in \operatorname{cl} H_1 + \operatorname{cl} H_2 \subset \operatorname{cl}(H_1 + H_2)$, $f \in \operatorname{cl}(H_1 + H_2)$. Moreover, by definition, $\breve{h}_{2,i} \in H_1^\perp$. Therefore for any $h_1 \in H_1$,

$$\langle f \, , \, h_1 \rangle = \sum_{i=1}^{\infty} a_i \left\langle \breve{h}_{2,i} \, , \, h_1 \right\rangle = 0,$$

hence $f \in H_1^\perp$. That is, $f \in \operatorname{cl}(H_1 + H_2) \cap H_1^\perp$. □

Lemma S2.5: *Let X be an integrable random vector in $\mathbb{R}^K$ defined on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and let $\mathcal{F}_0$ be a sub σ – field of $\mathcal{F}$. Then, $\mathbb{E}[X|\mathcal{F}_0] = 0$ (P-almost surely) if and only if $\mathbb{E}[XZ] = 0$ for all bounded $\mathcal{F}_0$ – measurable random variables Z.*

Proof. Suppose that $\mathbb{E}[X|\mathcal{F}_0] = 0$. We have

$$\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[XZ|\mathcal{F}_0]] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_0]Z] = 0.$$

Conversely suppose that $\mathbb{E}[XZ] = 0$ for all bounded $\mathcal{F}_0$ – measurable random variables Z. Let $A \in \mathcal{F}_0$ and set $Z \coloneqq \mathbf{1}_A$. Clearly $\mathbb{E}Z^2 \leq 1$. Let Y be any of the conditional expectations $\mathbb{E}[X|\mathcal{F}_0]$. Then, by definition,

$$\int_A Y \,\mathrm{dP} = \int_A X \,\mathrm{dP} = \int XZ \,\mathrm{dP} = \mathbb{E}[XZ] = 0.$$

Now, suppose $\{Y \neq 0\}$ has positive measure. Then one of $\{Y > 0\}$ or $\{Y < 0\}$ must. Say the first, the argument for the latter is analogous. This is $\{Y > 0\} = E = \cup_{n \geq 1} E_n$ for $E_n \coloneqq \{Y > 1/n\}$. So one E_k at least has positive measure. So

[5] That $\{\breve{h} : h \in H_2\}$ is a linear subspace (and hence equal to its linear span) is clear as it is the image of the linear subspace H_2 under the linear operator $\Pi_1 \coloneqq \Pi[\cdot|H_1^\perp]$. Hence $\operatorname{cl}\{\breve{h} : h \in H_2\}$ is the closed linear span of $\{\breve{h} : h \in H_2\}$.

$\int_E Y \, \mathrm{dP} \geq \int_{E_k} Y \, \mathrm{dP} \geq \int_{E_k} 1/k \, \mathrm{dP} = \mathrm{P}(E_k)/k > 0$. But this is a contradiction since $E \in \mathcal{F}_0$. □

COROLLARY S2.2: *Let (U, X) be a random vector on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ with $U \in L_2(\mathrm{P})$ and $\mathbb{E}[UU'|X]$ non – singular almost surely. Let $B \subset L_2(\Omega, \sigma(U, X), \mathrm{P})$ be the set of bounded functions b of (u, x) such that $\mathbb{E}[b(U, X)U|X] = 0$. Then*

$$\mathrm{cl}\, B = \{UZ : Z \textit{ is a bounded, } \sigma(X)\textit{–measurable random variable}\}^{\perp}.$$

Proof. Suppose that $b \in B$. Then $\mathbb{E}[b(U, X)UZ] = \mathbb{E}[\mathbb{E}[b(U, X)U|X]Z] = 0$. Conversely suppose that $b \in L_2(\Omega, \sigma(U, X), \mathrm{P})$ is such that $\mathbb{E}[b(U, X)UZ] = 0$ for Z any bounded $\sigma(X)$ – measurable random variable. By Lemma S2.5 this implies that $\mathbb{E}[b(U, X)U|X] = 0$, whence by Lemma C.7 in Newey (1991) $b \in \mathrm{cl}\, B$. □

THEOREM S2.1: *Let H be a Hilbert space. Let $(h_n)_{n\in\mathbb{N}}$ be a sequence in H, $h \in H$, $(L_n)_{n\in\mathbb{N}}$ be a sequence of closed (proper) linear subspaces of H and L a closed (proper) linear subspace of H. Set $g_n := \Pi(h_n|L_n)$ and $g := \Pi(h|L)$. If*

(i) $h_n \to h$;

(ii) for each $f \in L$, there is a sequence $(f_n)_{n\in\mathbb{N}}$ and a $N \in \mathbb{N}$ such that $f_n \to f$ and $f_n \in L_n$ for $n \geq N$,

then $g_n \to g$.

Proof. Let Π_n denote the projection onto L_n and Π that onto L. We consider first the case where $h_n = h$ for each $n \in \mathbb{N}$. Then $g_n = \Pi_n h$. We will show that any subsequence of $(g_n)_{n\in\mathbb{N}}$ has a further subsequence which converges to g. In particular, any subsequence of $(g_n)_{n\in\mathbb{N}}$ is bounded since $\|\Pi_n\| = 1$. Hence it has a weakly convergent subsequence, say $(g_{n_k})_{k\in\mathbb{N}}$ (Royden and Fitzpatrick, 2010, Theorem 16.6). Let $g^\star$ be this weak limit. By self-adjointness and idempotency of orthogonal projections

$$\langle g_{n_k} \, , \, g_{n_k} \rangle = \langle \Pi_{n_k} h \, , \, \Pi_{n_k} h \rangle = \langle h \, , \, \Pi_{n_k} h \rangle \to \langle h \, , \, g^\star \rangle \, . \tag{S3}$$

Let $f \in L$. By hypothesis there is a sequence $(f_n)_{n\in\mathbb{N}}$ with $f_n \to f$ and $f_n \in L_n$ for all sufficiently large n. Hence $f_{n_k} \to f$ and $f_{n_k} \in L_{n_k}$ for all sufficiently large k. Therefore, since $h - \Pi_{n_k} h \rightharpoonup h - g^\star$,

$$\langle h - g^\star \, , \, f \rangle = \lim_{k\to\infty} \langle h - g_{n_k} \, , \, f_{n_k} \rangle = 0$$

by Proposition 16.7 in Royden and Fitzpatrick (2010) and the fact that $h - g_{n_k} \in$

$L_{n_k}^\perp$ for each k. In consequence $g^\star = \Pi h = g$. Therefore, by self-adjointness and idempotency of Π and (S3)

$$\lim_{k\to\infty} \langle g_{n_k} \,,\, g_{n_k}\rangle = \langle h \,,\, \Pi h\rangle = \langle \Pi h \,,\, \Pi h\rangle = \langle g \,,\, g\rangle \,,$$

and hence by the Radon – Riesz Theorem (e.g. Royden and Fitzpatrick, 2010, p. 315) $g_{n_k} \to g$. As the initial subsequence was arbitrary it follows that $g_n \to g$.

To complete the proof let $h_n \to h$ be an arbitrary convergent sequence. Then

$$\|g_n - g\| = \|\Pi_n h_n - \Pi h\| \le \|\Pi_n h_n - \Pi_n h\| + \|\Pi_n h - \Pi h\| \le \|h_n - h\| + \|\Pi_n h - \Pi h\|.$$

The first term on the right hand side converges to zero by assumption; the second by the case with $h_n = h$ proven above. □

S2.6 Uniform results under a measure structure

Assume that $(H_\gamma, \mathcal{S}, Q)$ is a finite measure space, where both the σ-algebra $\mathcal{S}$ on H_γ and the finite measure Q are arbitrary. With such a structure uniformity over "large" subsets holds under measurability assumptions as a consequence of the pointwise result recorded in Remark 3.4 and and Egorov's Theorem (see e.g. Dudley, 2002, Theorem 7.5.1).

Corollary S2.3: *Suppose the conditions of Theorem 3.1 hold, $(H_\gamma, \mathcal{S}, Q)$ is a finite measure space and the functions $h = (\tau, b) \mapsto \pi_n(\tau, b)$ are measurable. Then, for any $\varepsilon > 0$ there is a $K \in \mathcal{S}$ such that $Q(H_\gamma \setminus K) < \varepsilon$ and*

$$\lim_{n\to\infty} \sup_{(\tau,b)\in K} |\pi_n(\tau, b) - \pi(\tau)| = 0,$$

for π_n and π as defined in Remark 3.4.

Proof. Let $\pi_n(h) \coloneqq P_{n,\gamma,h}\psi_{n,\theta}$ and $\pi(h) \coloneqq 1 - \mathrm{P}(\chi_r^2(a) \le c_r)$ if $r \ge 1$ or $\pi(h) \coloneqq 0$ if $r = 0$. By Remark 3.4, $\pi_n(h) \to \pi(h)$ pointwise in $h \in H_\gamma$. π is measurable as the pointwise limit of measurable functions (e.g. Dudley, 2002, Theorem 4.2.2). By Egorov's theorem (e.g. Dudley, 2002, Theorem 7.5.1), $\pi_n(h) \to \pi(h)$ uniformly on a K satisfying the given requirements. □

One set of sufficient conditions for the measurability requirement in the statement of Corollary S2.3 is: H_γ is a topological space, $\mathcal{S}$ its Borel σ-algebra and each $h \mapsto P_{n,\gamma,h}\psi_{n,\theta} = \pi_n(h)$ is continuous. The last requirement holds a fortiori if each $h \mapsto P_{n,\gamma,h}$ is continuous in total variation.

If one requires only a uniform version of Corollary 3.1, one may place the measure structure only on $H_{\gamma,0}$.

Corollary S2.4: *Suppose that the conditions of Corollary 3.1 hold, that $(H_{\gamma,0}, \mathcal{S}, Q)$ is a finite measure space and that the functions $h = (\tau, b) \mapsto \pi_n(\tau, b)$ are measurable. Then, for any $\varepsilon > 0$ there is a $K \in \mathcal{S}$ such that $Q(H_{\gamma,0} \setminus K) < \varepsilon$ and*

$$\lim_{n\to\infty} \sup_{h\in K} P_{n,\gamma,h}\psi_{n,\theta} = \begin{cases} \alpha & \text{if } r \geq 1 \\ 0 & \text{if } r = 0 \end{cases}.$$

Proof. Let $\pi_n(h) := P_{n,\gamma,h}\psi_{n,\theta}$ and $\pi(h) := \alpha$ if $r \geq 1$ or $\pi(h) := 0$ if $r = 0$. By Corollary 3.1, $\pi_n(h) \to \pi(h)$ pointwise in $h \in H_{\gamma,0}$. Since π is a constant function it is measurable. By Egorov's theorem (e.g. Dudley, 2002, Theorem 7.5.1), $\pi_n(h) \to \pi(h) = c$ uniformly on a K satisfying the given requirements. Hence,

$$\lim_{n\to\infty} \sup_{h\in K} \pi_n(h) \leq \lim_{n\to\infty} \sup_{h\in K} |\pi_n(h) - c| + c = c = \pi(h). \qquad \square$$

S2.7 Attaining the power bounds

The proof of Theorem 3.5 relies on the following result.

Theorem S2.2: *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, H a linear space and $B \subset H$ a linear subspace of H. Suppose that G_n is a Gaussian process on $(\Omega, \mathcal{F}, \mathbb{P})$ with index set H and covariance kernel K_n for each $n \in \mathbb{N}$ and that G is a Gaussian process on $(\Omega, \mathcal{F}, \mathbb{P})$ with index set H and covariance kernel K. Suppose that $K_n(h, g) \to K(h, g)$, $h, g \in H$. Let H be equipped with the positive semi - definite, symmetric bilinear form defined as $\langle h\,,\, g\rangle := K(h, g)$ and suppose that H is separable under the induced pseudometric.*

Fix $h, g \in H$ and define

$$X_n := (\mathsf{G}_n h, \mathbb{E}[\mathsf{G}_n g | \mathscr{G}_n]), \qquad X := (\mathsf{G}h, \mathbb{E}[\mathsf{G}g | \mathscr{G}]),$$

where $\mathscr{G}_n := \sigma(\{\mathsf{G}_n f : f \in B\})$ and $\mathscr{G}_n := \sigma(\{\mathsf{G} f : f \in B\})$. Then $X_n \rightsquigarrow X$.

Proof. We first note that $\mathrm{cl}\{\mathsf{G}b : b \in B\}$ is a Hilbert space when viewed as a subspace of L_2, i.e. once functions a.e. equal have been identified. Hence an orthonormal basis $(b_j)_{j\in\mathbb{N}} \subset \mathrm{cl}\, B$ of $\mathrm{cl}\{\mathsf{G}b : b \in B\} \subset L_2$ exists. Let $\mathcal{G}_n := \sigma(\{\mathsf{G}_n b_i : i \in \mathbb{N}\})$ and $\mathcal{G} := \sigma(\{\mathsf{G}b_i : i \in \mathbb{N}\})$. Let $B_m := (b_1, \ldots, b_m)$, $\mathcal{G}_n^m :=$

$\sigma(\{\mathsf{G}_n b : b \in B_m\})$ and $\mathcal{G}^m := \sigma(\{\mathsf{G}b : b \in B_m\})$. Define

$$X_n^m := (\mathsf{G}_n h, \mathbb{E}[\mathsf{G}_n g|\mathcal{G}_n^m])\,, \qquad X^m := (\mathsf{G}h, \mathbb{E}[\mathsf{G}g|\mathcal{G}^m])\,.$$

Since G_n, G are Gaussian processes, the conditional expectations in the preceding display can be written in closed form. Specifically let $Z_n^m := (\mathsf{G}_n h, \mathsf{G}_n g, \mathsf{G}_n b_1, \ldots, \mathsf{G}_n b_m)'$ and $Z^m := (\mathsf{G}_n h, \mathsf{G}g, \mathsf{G}_n b, \ldots, \mathsf{G}_n b)'$. Then

$$Z_n^m \sim \mathcal{N}(0, \Sigma_n^m), \quad Z^m \sim \mathcal{N}(0, \Sigma^m).$$

Partition Σ^m so that it is conformal with $Z_1^m = \mathsf{G}h$, $Z_2^m = \mathsf{G}g$ and $Z_3^m = (\mathsf{G}_n b, \ldots, \mathsf{G}_n b)'$, i.e.

$$\Sigma^m = \begin{bmatrix} [\Sigma^m]_{1,1} & [\Sigma^m]_{1,2} & [\Sigma^m]_{1,3} \\ [\Sigma^m]_{2,1} & [\Sigma^m]_{2,2} & [\Sigma^m]_{2,3} \\ [\Sigma^m]_{3,1} & [\Sigma^m]_{3,2} & [\Sigma^m]_{3,3} \end{bmatrix},$$

and similarly for Σ_n^m and Z_n^m. Then we have

$$X_n^m = (\mathsf{G}_n h,\, \mathbb{E}[\mathsf{G}_n g|\mathcal{G}_n^m]) = \left(Z_{n,1}^m,\, Z_{n,2}^m - [\Sigma_n^m]_{2,3}[\Sigma_n^m]_{3,3}^{-1} Z_{n,3}^m\right),$$

and similarly

$$X^m = (\mathsf{G}h,\, \mathbb{E}[\mathsf{G}g|\mathcal{G}^m]) = \left(Z_1^m,\, Z_2^m - [\Sigma^m]_{2,3}[\Sigma^m]_{3,3}^{-1} Z_3^m\right).$$

Since $K_n(h_1, h_2) \to K(h_1, h_2)$ for all $h_1, h_2 \in H$, $\Sigma_n^m \to \Sigma^m$ as $n \to \infty$ and therefore the inverses in the preceding displays exist for all sufficiently large n since B_m is orthonormal. By $\Sigma_n^m \to \Sigma^n$, Levy's continuity Theorem and the Cramér – Wold Theorem, $Z_n^m \rightsquigarrow Z^m$. Hence,

$$X_n^m \rightsquigarrow X^m. \tag{S4}$$

Let Π^m be the orthogonal projection onto $S_m := \text{Span}\{\mathsf{G}b : b \in B_m\}$. Then,

$$X^m = \mathbb{E}[\mathsf{G}g|\mathcal{G}^m] = \mathbb{E}[\mathsf{G}g|\mathscr{G}^m] = \Pi^m \mathsf{G}g,$$

by Theorem 9.1 in Janson (1997). The S_m are such that $S_m \subset S_{m+1}$ and $S := \text{cl}\{\mathsf{G}b : b \in B\} = \text{cl}\cup_{m\in\mathbb{N}} S_m$. By Theorem S2.1 and Theorem 9.1 in Janson (1997),

$\|\mathbb{E}[\mathsf{G}g|\mathscr{G}^m] - \mathbb{E}[\mathsf{G}g|\mathscr{G}]\|_{L_2} = \|\Pi^m \mathsf{G}g - \Pi \mathsf{G}g\|_{L_2} \to 0$ and so

$$X^m \rightsquigarrow X. \tag{S5}$$

Define $Y_n := \mathbb{E}[\mathsf{G}_n h|\mathcal{G}_n]$, $Y_n^m := \mathbb{E}[\mathsf{G}_n h|\mathcal{G}_n^m]$, $Y^m := \mathbb{E}[\mathsf{G}h|\mathcal{G}^m]$ and $Y := \mathbb{E}[\mathsf{G}h|\mathcal{G}]$. By Theorem 9.1 in Janson (1997), $Y_n \in \mathrm{cl}\{\mathsf{G}_n b : b \in B\}$ and $Y_n^m \in \{\mathsf{G}_n b : b \in B_m\}$, hence

$$Y_n - Y_n^m \sim \mathcal{N}(0, \sigma_{n,m}^2), \qquad \sigma_{n,m}^2 := \mathrm{Var}(Y_n - Y_n^m).$$

By Theorem 9.1 in Janson (1997), $Y_n = \mathbb{E}[\mathsf{G}_n h|\mathscr{G}_n]$ and $Y_n^m = \mathbb{E}[\mathsf{G}_n h|\mathscr{G}_n^m]$. Therefore,

$$\mathbb{P}\left(\|X_n - X_n^m\| > \varepsilon\right) = \mathbb{P}\left(|Y_n - Y_n^m| > \varepsilon\right) \leq C \exp\left(-\frac{\varepsilon^2}{\sigma_{n,m}^2}\right). \tag{S6}$$

We show next that $\sigma_{n,m}^2 \to \sigma_m^2 := \mathrm{Var}(Y - Y^m)$. For this let $f_0 := h$, $f_i := b_i$, $i \in \mathbb{N}$. Consider the restricted processes $F_n := (F_{n,i})_{i\in\mathbb{N}}$ and $F := (F_i)_{i\in\mathbb{N}}$ where $F_{n,i} := \mathsf{G}_n f_{i+1}$ and $F_i := \mathsf{G} f_{i+1}$. F_n and F are random elements in $(\mathbb{R}^\infty, d)$ where d is the metric given in Example 1.2 of Billingsley (1999). Hence $F_n \rightsquigarrow F$ in $(\mathbb{R}^\infty, d)$ by Example 2.4 of Billingsley (1999). By this and the fact that $(\mathbb{R}^\infty, d)$ is separable (e.g. Billingsley, 1999, Example 1.2), the Skorohod representation Theorem (e.g. Billingsley, 1999, Theorem 6.7) yields random elements $\tilde{F}_n$ and $\tilde{F}$ defined on a common probability space such that $\tilde{F}_n \to \tilde{F}$ surely and with $\mathcal{L}(\tilde{F}) = \mathcal{L}(F)$ and $\mathcal{L}(\tilde{F}_n) = \mathcal{L}(F_n)$. Thus $\tilde{F}_n$ and $\tilde{F}$ are Gaussian processes. In particular, $\mathrm{Cov}(\tilde{F}_{n,i}, \tilde{F}_{n,j}) = K_n(f_i, f_j) \to K(f_i, f_j) = \mathrm{Cov}(\tilde{F}_i, \tilde{F}_j)$ which implies each $(\tilde{F}_{n,i})_{n\in\mathbb{N}}$ is uniformly square integrable. As $(\mathbb{R}^\infty, d)$ has the topology of pointwise convergence (Billingsley, 1999, Example 1.2), each $\tilde{F}_{n,i} \to \tilde{F}_i$ surely. Hence $\tilde{F}_{n,i} \xrightarrow{L_2} \tilde{F}_i$. By the equality in law one has that

$$\tilde{Y}_n^m := \mathbb{E}[\tilde{F}_{n,1}|\sigma(\{\tilde{F}_{n,i} : 2 \leq i \leq m\})] \sim Y_n^m, \qquad \tilde{Y}_n := \mathbb{E}[\tilde{F}_{n,1}|\sigma(\{\tilde{F}_{n,i} : i \in \mathbb{N}, i \neq 1\})] \sim Y_n$$

and

$$\tilde{Y}^m := \mathbb{E}[\tilde{F}_1|\sigma(\{\tilde{F}_i : 2 \leq i \leq m\})] \sim Y^m, \qquad \tilde{Y} := \mathbb{E}[\tilde{F}_1|\sigma(\{\tilde{F}_i : i \in \mathbb{N}, i \neq 1\})] \sim Y.$$

Let $\tilde{S}_n^m := \mathrm{Span}\{\tilde{F}_{n,i} : 2 \leq i \leq m\}$, $\tilde{S}_n := \mathrm{cl}\,\mathrm{Span}\{\tilde{F}_{n,i} : i \in \mathbb{N}, i \neq 1\}$, $\tilde{S}^m := \mathrm{Span}\{\tilde{F}_i : 2 \leq i \leq m\}$ and $\tilde{S} := \mathrm{cl}\,\mathrm{Span}\{\tilde{F}_i : i \in \mathbb{N}, i \neq 1\}$ all considered as subsets of L_2. Then

$$\tilde{Y}_n^m = \Pi[\tilde{F}_{n,1}|S_n^m], \qquad \tilde{Y}_n = \Pi[\tilde{F}_{n,1}|S_n], \qquad \tilde{Y}^m = \Pi[\tilde{F}_1|S^m], \qquad \tilde{Y} = \Pi[\tilde{F}_1|S],$$

By Theorem 9.1 in Janson (1997). We will apply Theorem S2.1 twice (in L_2). It is straightforward to check the hypotheses are satisfied with (i) $L_n := \tilde{S}_n^m$, $L := \tilde{S}^m$; (ii) $L_n := \tilde{S}_n$, $L := \tilde{S}$ and $h_n := \tilde{F}_{n,1}$, $h := \tilde{F}_1$ in both cases. Then by Theorem S2.1,

$$\|\tilde{Y}_n - \tilde{Y}_n^m - (\tilde{Y} - \tilde{Y}^m)\|_{L_2} \le \|\tilde{Y}_n - \tilde{Y}\|_{L_2} + \|\tilde{Y}_n^m - \tilde{Y}^m\|_{L_2} \to 0,$$

hence $\sigma_{n,m}^2 = \operatorname{Var}(Y_n - Y_n^m) = \operatorname{Var}(\tilde{Y}_n - \tilde{Y}_n^m) \to \operatorname{Var}(\tilde{Y} - \tilde{Y}^m) = \operatorname{Var}(Y - Y^m) = \sigma_m^2$.

The penultimate step is to show that $\sigma_m^2 \to 0$. For this, we note that $Y = \Pi[\mathsf{G}h | \operatorname{cl}\{\mathsf{G}b : b \in B\}]$ and $Y^m = \Pi[\mathsf{G}h | \operatorname{Span}\{\mathsf{G}b : b \in B_m\}]$. Set $L_m := \operatorname{Span}\{\mathsf{G}b : b \in B_m\}$ and $L := \operatorname{cl}\{\mathsf{G}b : b \in B\}$. It is easy to check the hypotheses of Theorem S2.1 (with m in place of n) hold, with the second following from the choice of B_m. Hence $Y^m \xrightarrow{L_2} Y$ and so $\sigma_m^2 = \operatorname{Var}(Y - Y^m) \to 0$. In conjunction with (S6) we have

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\|X_n - X_n^m\| > \varepsilon\right) \le \lim_{m\to\infty} \limsup_{n\to\infty} C \exp\left(-\frac{\varepsilon^2}{\sigma_{n,m}^2}\right) = 0. \qquad \text{(S7)}$$

The result now follows by applying Theorem 3.2 in Billingsley (1999), noting that equations (S4), (S5) and (S7) verify the required hypotheses. □

S3 Additional details and proofs for the examples

S3.1 Single index model

S3.1.1 Proofs of results in the main text

Proof of Proposition 4.1. We verify the conditions of Lemma 3.8. That (each component of) $g_{n,\gamma} \in L_2^0(P_\gamma)$ follows from the facts that under Assumption 4.1, for $W \sim P_\gamma$, $\mathbb{E}\left[g_\gamma(W)\right] = \mathbb{E}\left[\mathbb{E}\left[g_\gamma(W)|X\right]\right] = 0$ and

$$\mathbb{E}\left[g_{\gamma,k}(W)^2\right] \lesssim \mathbb{E}\left[\epsilon^2 \left(X_{2,k} - \frac{\mathbb{E}[\omega(X)X_{2,k}|V_\theta]}{\mathbb{E}[\omega(X)|V_\theta]}\right)^2\right] \lesssim \mathbb{E}X_{2,k}^2 < \infty,$$

by $\omega : \mathbb{R}^K \to [\underline{\omega}, \overline{\omega}]$ and the first part of (29). For any $b \in B_\gamma$, if $W \sim P_\gamma$, $\mathbb{E}[\epsilon b_2(\epsilon, X)|X] = 0$ by (26) and hence

$$\begin{aligned}
\mathbb{E}\left[g_\gamma(W)[D_\gamma b](W)\right] &= \mathbb{E}\left[\omega(X)\left(\epsilon f'(V_\theta)\left(X_2 - \frac{\mathbb{E}[\omega(X)X_2|V_\theta]}{\mathbb{E}[\omega(X)|V_\theta]}\right)\right)(-\phi(\epsilon, X)b_1(V_\theta) + b_2(\epsilon, X))\right] \\
&= \mathbb{E}\left[-\mathbb{E}\left[\epsilon\phi(\epsilon, X)|X\right] f'(V_\theta)\omega(X)\left(X_2 - \frac{\mathbb{E}[\omega(X)X_2|V_\theta]}{\mathbb{E}[\omega(X)|V_\theta]}\right) b_1(V_\theta)\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}\left[\epsilon b_2(\epsilon, X)|X\right] f'(V_\theta)\omega(X)\left(X_2 - \frac{\mathbb{E}[\omega(X)X_2|V_\theta]}{\mathbb{E}[\omega(X)|V_\theta]}\right)\right] \\
&= \mathbb{E}\left[f'(V_\theta)\left(\mathbb{E}[\omega(X)X_2|V_\theta] - \frac{\mathbb{E}[\omega(X)|V_\theta]\mathbb{E}[\omega(X)X_2|V_\theta]}{\mathbb{E}[\omega(X)|V_\theta]}\right) b_1(V_\theta)\right] \\
&= 0. \qquad \square
\end{aligned}$$

Proof of Proposition 4.2. For condition (i) of Asssumption 3.3, we start by observing that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{n,\theta,i} - g_\gamma(W_i) = \sum_{l=1}^5 a_l \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{m_n} R_{l,n,i} + \sum_{i=m_n+1}^{n} R_{l,n,i}\right],$$

for some $a_j \in \{-1, 1\}$ and

$$\begin{aligned}
R_{1,n,i} &:= \omega(X_i)(\hat{f}_{n,i}(V_{\theta,i}) - f(V_{\theta,i}))f'(V_{\theta,i})(X_{2,i} - Z_0(V_{\theta,i})) \\
R_{2,n,i} &:= \omega(X_i)(Y_i - f(V_{\theta,i}))\left(f'(V_{\theta,i}) - \widehat{f'}_{n,i}(V_{\theta,i})\right)(X_{2,i} - Z_0(V_{\theta,i})) \\
R_{3,n,i} &:= \omega(X_i)(Y_i - f(V_{\theta,i}))\widehat{f'}_{n,i}(V_{\theta,i})\left(\hat{Z}_{0,n,i}(V_{\theta,i}) - Z_0(V_{\theta,i})\right) \\
R_{4,n,i} &:= \omega(X_i)(\hat{f}_{n,i}(V_{\theta,i}) - f(V_{\theta,i}))\left(f'(V_{\theta,i}) - \widehat{f'}_{n,i}(V_{\theta,i})\right)(X_{2,i} - Z_0(V_{\theta,i})) \\
R_{5,n,i} &:= \omega(X_i)(\hat{f}_{n,i}(V_{\theta,i}) - f(V_{\theta,i}))\widehat{f'}_{n,i}(V_{\theta,i})\left(\hat{Z}_{0,n,i}(V_{\theta,i}) - Z_0(V_{\theta,i})\right).
\end{aligned}$$

It suffices to verify that one of conditions (i) or (ii) of Lemma S3.3 is satisfied with $Y_{n,i} = R_{l,n,i}$ with i ranging over either $1, \ldots, m_n$ or $m_n + 1, \ldots, n$. For $l = 4, 5$ I show that condition (i) holds and for $l = 1, 2, 3$, I show that condition (ii) is satisfied.

First suppose that $1 \leq i \leq m_n$. Let $\mathcal{C}_n = (W_{m_n+1}, \ldots, W_n)$ and note that each $\hat{Z}_{k,n,i}(V_{\theta,i})$ is $\sigma(V_{\theta,i}, \mathcal{C}_n)$ – measurable for $k = 0, 1, 2, 3, 4$. Additionally $f(V_{\theta,i})$, $f'(V_{\theta,i})$, $\omega(X_i)$ and $X_{2,i} - Z_0(V_{\theta,i})$ are bounded uniformly in i under our hypotheses and there is a sequence of events E_n with probability approaching one on which $\mathsf{R}_{l,n,i} \leq r_n$ and for all large enough $n \in \mathbb{N}$, $\hat{f}_{n,i}(V_{\theta,i})$, $\widehat{f'}_{n,i}(V_{\theta,i})$, $\hat{Z}_{1,n,i}(V_{\theta,i})$ are bounded above uniformly in i and $\hat{Z}_{2,n,i}(V_{\theta,i})$ is bounded above and below uniformly in i. Since

$$\hat{Z}_{0,n,i}(V_{\theta,i})-Z_0(V_{\theta,i}) = \frac{(\hat{Z}_{1,n,i}(V_{\theta,i}) - Z_1(V_{\theta,i}))Z_2(V_{\theta,i}) + (Z_2(V_{\theta,i}) - \hat{Z}_{2,n,i}(V_{\theta,i}))Z_1(V_{\theta,i})}{Z_2(V_{\theta,i})\hat{Z}_{2,n,i}(V_{\theta,i})}$$

on these sets we also have

$$\mathbb{E}\left[\left\|\hat{Z}_{0,n,i}(V_{\theta,i}) - Z_0(V_{\theta,i})\right\|^2 \middle| \mathcal{C}_n\right] \lesssim r_n^2. \tag{S8}$$

$l = 1$: the first part of condition (ii) follows by the law of iterated expectations and independence since $\mathbb{E}[\omega(X_i)(X_{2,i}-Z_0(V_{\theta,i}))|V_{\theta,i}] = 0$. The second part follows with $\delta_n \lesssim r_n^2$ due to the uniform boundedness noted above and $\mathsf{R}_{3,n,i} \leq r_n$ on E_n.

$l = 2$: the first part of condition (ii) follows by the law of iterated expectations and independence since $\mathbb{E}[\epsilon_i|X_i] = 0$. The second part follows with $\delta_n \lesssim r_n^2$ due to the uniform boundedness noted above, $\mathbb{E}[\epsilon^2|X] \leq C$ from equation (29) and $\mathsf{R}_{4,n,i} \leq r_n$ on E_n.

$l = 3$: the first part of condition (ii) follows by the law of iterated expectations and independence since $\mathbb{E}[\epsilon_i|X_i] = 0$. The second part follows with $\delta_n \lesssim r_n^2$ due to the uniform boundedness noted above, $\mathbb{E}[\epsilon^2|X] \leq C$ and equation (S8) which holds on E_n.

$l = 4$: By the uniform boundedness noted above and the (conditional) Cauchy – Schwarz inequality,

$$\mathbb{E}[\|R_{4,n,i}\||\mathcal{C}_n] \lesssim \mathbb{E}\left[\left|\hat{f}_{n,i}(V_{\theta,i}) - f(V_{\theta,i})\right| \left|f'(V_{\theta,i}) - \widehat{f'}_{n,i}(V_{\theta,i})\right| \middle| \mathcal{C}_n\right],$$

and the right hand side is upper bounded by $\mathsf{R}_{3,n,i}\mathsf{R}_{4,n,i} = o(n^{-1/2})$ on E_n.

$l = 5$: By the uniform boundedness noted above and the (conditional) Cauchy – Schwarz inequality,

$$\mathbb{E}[\|R_{4,n,i}\||\mathcal{C}_n] \lesssim \mathbb{E}\left[\left|\hat{f}_{n,i}(V_{\theta,i}) - f(V_{\theta,i})\right| \left\|\hat{Z}_{0,n,i}(V_{\theta,i}) - Z_0(V_{\theta,i})\right\| \middle| \mathcal{C}_n\right],$$

and, given equation (S8), the right hand side is upper bounded by some constant multiple of $r_n\mathsf{R}_{3,n,i} = o(n^{-1/2})$ on E_n.

The case where $m_n + 1 \leq i \leq n$ is analogous with $\mathcal{C}_n = (W_1, \ldots, W_{m_n})$.

For part (ii) of Asssumption 3.3, we show that $\|\check{V}_{n,\theta} - V_\gamma\| = o_{P_\gamma^n}(\nu_n)$. This suffices since in the case where $\hat{V}_{n,\theta}$ is constructed as in subsection S2.1, it implies that equation (S2) holds and hence the desired result follows by by Proposition S2.1. In the case where V_γ is full rank and $\check{V}_{n,\theta} = \hat{V}_{n,\theta}$ this directly gives consistency of $\hat{V}_{n,\theta}$ for V_γ and the result follows by the continuous mapping theorem.

For $\breve{V}_\gamma := \mathbb{P}_n g_\gamma g_\gamma'$,

$$\breve{V}_{n,\theta} - V_\gamma = \breve{V}_{n,\theta} - \breve{V}_\gamma + \breve{V}_\gamma - V_\gamma = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{g}_{n,\theta,i}\hat{g}_{n,\theta,i}' - g_\gamma(W_i)g_\gamma(W_i)'\right] + \frac{1}{\sqrt{n}}\mathbb{G}_n[g_\gamma g_\gamma'].$$

For the second term on the right hand side, $P_\gamma[(g_{\gamma,l}g_{\gamma,k})^2] < \infty$ by $\mathbb{E}[\epsilon^4] < \infty$ and the boundedness of all the other terms in (28) under Assumption 4.3. Hence by the central limit theorem, $\frac{1}{\sqrt{n}}\mathbb{G}_n[g_\gamma g_\gamma'] = O_{P_\gamma^n}(n^{-1/2})$. For the remaining term,

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{g}_{n,\theta,i,k} - g_{\gamma,k}(W_i)\right)^2 \lesssim \sum_{l=1}^{5}\frac{1}{n}\left[\sum_{i=1}^{m_n} R_{l,n,i,k}^2 + \sum_{i=m_n+1}^{n} R_{l,n,i,k}^2\right].$$

For $l = 1, 2, 3$ we showed above that if $1 \leq i \leq m_n$ and $\mathcal{C}_n = (W_{m_n+1}, \ldots, W_n)$ then $\mathbb{E}[R_{l,n,i,k}^2|\mathcal{C}_n] \lesssim r_n^2$ on E_n. We will show this also holds for $l = 4, 5$ with $1 \leq i \leq m_n$ and $\mathcal{C}_n = (W_{m_n+1}, \ldots, W_n)$ (the case with $m_n + 1 \leq i \leq n$ with $\mathcal{C}_n = (W_1, \ldots, W_{m_n})$ is once again analogous). For $l = 4$ or $l = 5$, by the uniform boundedness (for all large enough n) we have

$$\mathbb{E}\left[R_{l,n,i,k}^2|\mathcal{C}_n\right] \lesssim \mathbb{E}\left[\left(\hat{f}_{n,i}(V_{\theta,i}) - f(V_{\theta,i})\right)^2\middle|\mathcal{C}_n\right],$$

and the right hand side term is bounded above by r_n^2 on E_n. Hence, by Markov's inequality $\frac{1}{n}\left[\sum_{i=1}^{m_n} R_{l,n,i,k}^2 + \sum_{i=m_n+1}^{n} R_{l,n,i,k}^2\right] = O_{P_\gamma^n}(r_n^2)$ for $l = 1, \ldots, 5$, which implies that the same is true of $\frac{1}{n}\sum_{i=1}^{n}\|\hat{g}_{n,\theta,i,} - g_\gamma(W_i)\|^2$. Therefore, by Cauchy – Schwarz

$$\begin{aligned}
\left\|\breve{V}_{n,\theta} - \breve{V}_\gamma\right\|_2 &= \left\|\frac{1}{n}\sum_{i=1}^{n}\hat{g}_{n,\theta,i}\left(\hat{g}_{n,\theta,i} - g_\gamma(W_i)\right)' + \left(\hat{g}_{n,\theta,i} - g_\gamma(W_i)\right)g_\gamma(W_i)'\right\|_2 \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\left\|\hat{g}_{n,\theta,i}\left(\hat{g}_{n,\theta,i} - g_\gamma(W_i)\right)'\right\|_2 + \frac{1}{n}\sum_{i=1}^{n}\left\|\left(\hat{g}_{n,\theta,i} - g_\gamma(W_i)\right)g_\gamma(W_i)'\right\|_2 \\
&\leq \left(\frac{1}{n}\sum_{i=1}^{n}\|\hat{g}_{n,\theta,i}\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\|\hat{g}_{n,\theta,i} - g_\gamma(W_i)\|^2\right)^{1/2} \\
&\quad + \left(\frac{1}{n}\sum_{i=1}^{n}\|\hat{g}_{n,\theta,i} - g_\gamma(W_i)\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\|g_\gamma(W_i)\|^2\right)^{1/2} \\
&= O_{P_\gamma^n}(r_n).
\end{aligned}$$

□

S3.1.2 The LAN condition

Here I provide examples of local perturbations $P_{n,\gamma,h}$ and lower level conditions under which the LAN condition in Assumption 4.2 holds. I will consider two models, (A) and (B). (B) is restricted such that f is an increasing function. In both cases the function $\varphi_{n,2}(b_1, b_2)$ has the form

$$\varphi_{n,2}(b_1, b_2) = (b_1,\, b_2\zeta)/\sqrt{n} \tag{S9}$$

and $B_{\gamma,2}$ is taken to be the set of functions $b_2 : \mathbb{R}^{1+K} \to \mathbb{R}$ such that b_2 is bounded, $e \mapsto b_2(e, x)$ is continuously differentiable with bounded derivative and equation (26) holds.[6]

In case (A), $B_{\gamma,1} = B^A_{\gamma,1} := C^1_b(\mathscr{D})$, the class of functions which are bounded and continuously differentiable with bounded derivative on $\mathscr{D}$. In case (B) $B_{\gamma,1} = B^B_{\gamma,2} := C^1_b(\mathscr{D}) \cap \mathscr{I}(\mathscr{D})$, for $\mathscr{I}(\mathscr{D})$ the set of functions from $\mathscr{D} \to \mathbb{R}$ which are monotone increasing, ensuring that $f + b_1/\sqrt{n}$ is always a monotone increasing function.

Proposition S3.1: *Let $H_\gamma = \mathbb{R}^{d_\theta} \times B^A_{\gamma,1} \times B_{\gamma,2}$ and $H_\gamma = \mathbb{R}^{d_\theta} = B^B_{\gamma,1} \times B_{\gamma,2}$ in Model A and B respectively. Then, if Assumption 4.1 holds, $e \mapsto \sqrt{\zeta(e,x)}$ is continuously differentiable, $\mathcal{W}_n = \prod_{i=1}^n \mathbb{R}^{1+K}$, ν is invariant under the function $F_\gamma(y, x) := (y - f(x_1 + x_2'\theta), x)$ for any $\gamma \in \Gamma$ and $p_{n,\gamma,h} = p^n_{\gamma+\varphi_n(h)}$ with p_γ as in (23) and φ_n as given by equations (25) & (S9), Assumption 4.2 holds for both Model A and Model B.*

Proof. The product space and product measure parts of Assumption 3.5 holds by the corresponding Assumptions in the Proposition. That each $P_\gamma \ll \nu$ follows from the definition of p_γ in (23).

Define $\gamma_t(h) := \gamma + t(\tau, b_1, b_2\zeta)$ for $h = (\tau, b_1, b_2)$ and $t \in [0, \infty)$. We first note that – as is easy to check – the corresponding measures $P_{\gamma_t(h)} \in \{P_\gamma : \gamma \in \Gamma\}$ for all small enough t for both Model A and Model B. We now verify the conditions of Lemma 1.8 in van der Vaart (2002). Firstly, $t \mapsto \sqrt{p_{\gamma_t(h)}}$ is continuously differentiable everywhere since

$$\sqrt{p_{\gamma_t(h)}(W)} = \sqrt{\zeta(Y - f(V_{\theta+t\tau}) - tb_1(V_{\theta+t\tau}), X)}\sqrt{(1 + tb_2(Y - f(V_{\theta+t\tau}) - tb_1(V_{\theta+t\tau}), X))},$$

which is a composition of continuously differentiable functions for t small enough that $(1+tb_2)$ is bounded away from zero. This ensures that $q_t(W) := \frac{\mathrm{d}\log p_{\gamma_s(h)(W)}}{\mathrm{d}s}|_{s=t}$

[6] Motivation for the conditions in (26) is given following Proposition S3.1.

is defined for small enough t. Writing $v_t \coloneqq V_{\theta+t\tau}$ and $e_t \coloneqq Y - f(v_t) - tb_1(v_t)$ this has the form

$$\begin{aligned} q_t(W) \coloneqq &- \phi(e_t, X)[f'(v_t)X_2'\tau + tb_1'(v_t)X_2'\tau + b_1(v_t)] \\ &+ \frac{b_2(e_t, X) - tb_2'(e_t, X)[f'(v_t)X_2'\tau + tb_1'(v_t)X_2'\tau + b_1(v_t)]}{1 + tb_2(e_t, X)}. \end{aligned} \tag{S10}$$

By inspection, this is continuous everywhere as the composition of continuous functions. For some $\rho > 0$, note that by the boundedness of f', b_1, b_1', b_2, b_2' and $1/(1 + tb_2)$ and equation (24) for some positive constant $C < \infty$,

$$\int |q_t(W)|^{2+\rho} \, \mathrm{d}P_{\gamma_t(h)} \leq C\mathbb{E}\left[\left(|\phi(\epsilon, X)|^{2+\rho} + 1\right) \|X\|^{2+\rho}\right] < \infty.$$

This implies that for any $t_n \to t$, $(q_{t_n}(W)^2)_{n\in\mathbb{N}}$ is uniformly $P_{\gamma_{t_n}(h)}$ – integrable. Combination with $q_{t_n}(W)^2 \to q_t(W)^2$ (everywhere) yields

$$\int q_{t_n}(W)^2 p_{\gamma_{t_n}(h)}(W)\lambda \to \int q_t(W)^2 p_{\gamma_t(h)}(W)\lambda.$$

Applying Lemma 1.8 in van der Vaart (2002) demonstrates that equation (20) holds, with $A_\gamma h$ as in (21). Lemma 1.7 of van der Vaart (2002) ensures that $A_\gamma h \in L_2^0(P_\gamma)$. The form of $A_\gamma h$ reveals that it is a linear map on H_η. That it is bounded follows from

$$\|A_\gamma h\|^2 \leq C_1\mathbb{E}\left[\phi(\epsilon, X)^2\|X\|^2\right] \|\tau\|^2 + \mathbb{E}\left[\phi(\epsilon, X)^2\right] \|b_1\|^2 + \|b_2\|^2 \leq C_2\|h\|^2,$$

where $C_1, C_2 \in (0, \infty)$ are positive constants. Apply Lemma 3.7 to complete the proof. □

Motivation for the conditions on b_2 Equation (26) imposes 3 conditions on the functions b_2, i.e. the score functions corresponding to the perturbation of the density function ζ. As the first and last such conditions, i.e. that b_2 is mean zero with finite second moment is a requirement of *any* score function (cf. Assumption 3.1 or Lemma 1.7 in van der Vaart, 2002) here I discuss the motivation for condition that $\mathbb{E}[\epsilon b_2(\epsilon, X)|X] = 0$. In particular, I will heuristically argue that "well-behaved" parametric submodels lead to scores with this property. Let $\zeta_\beta(e, x)$ denote a parametric family of density functions of (ϵ, X) with respect to $\nu_1 \otimes \nu_2$ such that the marginal density of X, $\iota(x) = \int \zeta_\beta(e, x) \, \mathrm{d}\nu_1(e)$, does not depend on β. Provided the parametric family is sufficiently well – behaved, scores for β in the model $\{\zeta_\beta : \beta \in \mathcal{B}\}$ for $\mathcal{B}$ some open set, have the form $\nabla_\beta \log \zeta_\beta = \varphi_\beta$. The conditional expectation of this score can be written (X-a.s.)

$$\int e\varphi_\beta(e,X)\frac{\zeta_\beta(e,X)}{\int \zeta_\beta(e,X)\,\mathrm{d}\nu_1(e)}\,\mathrm{d}\nu_1(e) = \int e\frac{\nabla_\beta\zeta_\beta(e,X)}{\zeta_\beta(e,X)}\frac{\zeta_\beta(e,X)}{\int \zeta_\beta(e,X)\,\mathrm{d}\nu_1(e)}\,\mathrm{d}\nu_1(e)$$
$$= \int e\frac{\nabla_\beta\zeta_\beta(e,X)}{\int \zeta_\beta(e,X)\,\mathrm{d}\nu_1(e)}\,\mathrm{d}\nu_1(e).$$

Provided the derivatives exist, since $\nabla_\beta e = 0$ and $\nabla_\beta \iota(x) = 0$,

$$\nabla_\beta e\frac{\zeta_\beta(e,x)}{\int \zeta_\beta(e,x)\,\mathrm{d}\nu_1(e)} = e\frac{\nabla_\beta\zeta_\beta(e,x)}{\int \zeta_\beta(e,x)\,\mathrm{d}\nu_1(e)}.$$

Additionally (X-a.s.)

$$\mathbb{E}[\epsilon|X] = \int e\frac{\zeta_\beta(e,X)}{\int \zeta_\beta(e,X)\,\mathrm{d}\nu_1(e)}\,\mathrm{d}\nu_1(e) = 0 \quad \Longrightarrow \quad \nabla_\beta \int e\frac{\zeta_\beta(e,X)}{\int \zeta_\beta(e,X)\,\mathrm{d}\nu_1(e)}\,\mathrm{d}\nu_1(e) = 0.$$

If the last derivative can be taken inside the integral, combination of these displays yields

$$\int e\varphi_\beta(e,X)\frac{\zeta_\beta(e,X)}{\int \zeta_\beta(e,X)\,\mathrm{d}\nu_1(e)}\,\mathrm{d}\nu_1(e) = \int \nabla_\beta e\frac{\zeta_\beta(e,X)}{\int \zeta_\beta(e,X)\,\mathrm{d}\nu_1(e)}\,\mathrm{d}\nu_1(e) = 0.$$

Thus any score φ_β in such a well-behaved parametric submodel must satisfy the property imposed on b_2.

Ensuring p_γ is a probability density That p_γ is a valid probability density with $\tilde{\nu} = \nu$ follows immediately from (23) if ν is invariant with respect to

$$F_\gamma(y,x) = (y - f(x_1 + x_2'\theta), x) \tag{S11}$$

for each $\gamma \in \Gamma$, i.e. $[\nu \circ F_\gamma^{-1}] = \nu$. In such a case, clearly $p_\gamma \geq 0$ by (23) and by the invariance

$$\int p_\gamma\,\mathrm{d}\nu = \int \zeta \circ F_\gamma\,\mathrm{d}\nu = \int \zeta\,\mathrm{d}[\nu \circ F_\gamma^{-1}] = \int \zeta\,\mathrm{d}\nu = 1.$$

Such invariance of ν holds in important special cases. Specifically, suppose that $\epsilon|X$ has conditional density $\zeta_{1,x}$ with respect to λ and ζ_2 is the marginal density of X with respect to ν_2. Then $\zeta(e,x) = \zeta_{1,x}(e)\zeta_2(x)$ is a density with respect to $\nu := \lambda \otimes \nu_2$ To see that ν is invariant under F_γ, let $A = A_1 \times A_2 \subset \mathbb{R} \times \mathbb{R}^K$ be a measurable rectangle and define $G_{\gamma,x}(z) := z - f(x_1 + x_2'\theta)$. Then $(e,x) \in F_\gamma(A)$

if and only if $x \in A_2$ and $e \in G_{\gamma,x}(A_1)$. Hence, by Tonelli's Theorem

$$\nu(F_\gamma(A)) = \int \mathbf{1}_{A_2}(x) \left[\int \mathbf{1}_{G_{\gamma,x}(A_1)}(e) \, \mathrm{d}\lambda(e) \right] \mathrm{d}\nu_2(x) = \int_{A_2} \lambda(G_{\gamma,x}(A_1)) \, \mathrm{d}\nu_2(x).$$

Since $J_{G_{\gamma,x}}(z) = 1$, by change of variables

$$\lambda(G_{\gamma,x}(A_1)) = \int_{G_{\gamma,x}(A_1)} \mathrm{d}\lambda = \int_{A_1} |\det J_{G_{\gamma,x}}| \, \mathrm{d}\lambda = \int_{A_1} \mathrm{d}\lambda = \lambda(A_1),$$

for each x. Hence,

$$[\nu \circ F_\gamma^{-1}](A) = \nu(F_\gamma(A)) = \int_{A_2} \lambda(G_{\gamma,x}(A_1)) \, \mathrm{d}\nu_2 = \int_{A_2} \lambda(A_1) \, \mathrm{d}\nu_2 = \lambda(A_1) \times \nu_2(A_2) = \nu(A).$$

Since the measurable rectangles form a separating class, it follows that $\nu \circ F_\gamma^{-1} = \nu$.

S3.1.3 Uniform local regularity

As noted in section 4.1, in the single index model the proposed $\mathrm{C}(\alpha)$ test is locally regular as defined in Definition 2. Here I discuss strengthening this to the uniform local regularity required by Definition 2.2. I will place a psuedometric structure on H_γ as defined in Proposition S3.1 for Model A.[7]

There are many possible options. I outline two possibilities to establish asymptotic equicontinuity in total variation of $h \mapsto P_{n,\gamma,h}$. The first establishes this on $H_{\gamma,0}$, and thus may be used to establish that the proposed test controls the null rejection probability in a locally uniform manner (Corollary 3.4). The second requires a stronger norm on $B_{\gamma,1}$, but establishes the asymptotic equicontinuity on all subsets of the form $K_\gamma^B := \{(\tau, b_1, b_2) \in H_\gamma : \sup_{v \in \mathscr{D}} |b_1'(v)| \leq B\} \subset H_\gamma$.

Asymptotic equicontinuity on $H_{\gamma,0}$ One seemingly natural approach is as follows. Under Assumption 4.1, $\int \phi(e,x)^2 \zeta(e,x) \, \mathrm{d}\nu < \infty$ and hence $\int b_1(x_1 + x_2'\theta)^2 \phi(e,x)^2 \zeta(e,x) \, \mathrm{d}\nu < \infty$ since b_1 is bounded on $\mathscr{D}$. In consequence $H_{\gamma,0}$ is a subspace of the linear space

$$\mathcal{S}_\gamma := \{0\} \times \left\{ b_1 \in C_b^1(\mathscr{D}) : b_1(x_1 + x_2'\theta) =: \breve{b}_1(e,x) \in L_2(\mathbb{R}^{1+K}, \phi^2 \zeta) \right\} \times L_2(\mathbb{R}^{1+K}, \zeta).$$

[7] It suffices to consider Model A since the corresponding $H_\gamma = H_\gamma^A = \mathbb{R}^{d_\theta} \times B_{\gamma,1}^A \times B_{\gamma,2}$ is a superset of $H_\gamma^B := \mathbb{R}^{d_\theta} \times B_{\gamma,1}^B \times B_{\gamma,2}$, the space considered in Model B. Hence if the collection formed by the functions $h \mapsto P_{n,\gamma,h}$ with domain H_γ are shown to be asymptotically equicontinuous on a subset $K \subset H_\gamma$, they are also a fortiori asymptotically equicontinuous on $K \cap H_\gamma^B$ when their domain is restricted to H_γ^B.

$\mathcal{S}_\gamma$ can be equipped with the seminorm:[8]

$$\|h\| := \|(0, b_1, b_2)\| := \sqrt{\int b_1(x_1 + x_2'\theta)^2 \phi(e, x)^2 \zeta(e, x) \, \mathrm{d}\nu + \|b_2\|^2_{L_2(\mathbb{R}^{1+K}, \zeta)}}. \quad \text{(S12)}$$

I will additionally assume that

$$\mathbb{E}\left[\phi(\epsilon, X)^2 | X\right] \geq c > 0, \qquad (\epsilon, X) \sim \zeta. \quad \text{(S13)}$$

This ensures that the $L_2(\mathbb{R}^{1+K}, \phi^2\zeta)$ norm is stronger than the $L_2(\mathbb{R}^{1+K}, \zeta)$ norm on the subspace we consider.[9] The condition in (S13) is mild: if $\epsilon | X$ has a Lebesgue density $\zeta_{\epsilon|X}$, then $\mathbb{E}\left[\phi(\epsilon, X)^2 | X = x\right]$ is the Fisher information for the location family with densities $\zeta_{\epsilon|X}(e - \mu, x)$ (for $\mu \in \mathbb{R}$). Under very weak regularity conditions on $e \mapsto \zeta_{\epsilon|X}(e, x)$, this is bounded below by $1/\mathbb{E}[\epsilon^2 | X = x]$ (e.g. Kagan, Linnik, and Rao, 1973, Theorem 13.1.1), which is itself lower bounded under the condition imposed in equation (29).

$H_{\gamma,0}$ can then be considered as a pseudometric space with the induced pseudometric $d_1(h_1, h_2) = \|h_1 - h_2\|$. With this structure I obtain the asymptotic equicontinuity in total variation required by Lemma 3.1 under Assumptions 4.1, 4.2, (S13) and that $\sqrt{\zeta(e, x)}$ is continuously differentiable in its first argument as in Proposition S3.1.

Proposition S3.2: *Let $H_{\gamma,0} = \{0\} \times B^A_{\gamma,1} \times B_{\gamma,2}$ and equip it with the pseudometric d_1 induced by the seminorm (S12). Suppose that Assumptions 4.1, 4.2 & equation (S13) hold and that $e \mapsto \sqrt{\zeta(e, x)}$ is continuously differentiable. Then $h \mapsto P_{n,\gamma,h}$ is asymptotically equicontinuous in total variation on $H_{\gamma,0}$.*

Proof. It suffices to show $\lim_{n\to\infty} d_{TV}(P_{n,\gamma,h_n}, P_{n,\gamma,h}) = 0$ for $h_n \to h$, $h_n, h \in H_{\gamma,0}$. Let $q_{n,\gamma,b_1} := p_{(\theta, f + b_1/\sqrt{n}, \zeta)}$ such that

$$\frac{p_{n,\gamma,h_n}(W)}{p_{n,\gamma,h}(W)} = \frac{q_{n,\gamma,b_{1,n}}(W)}{q_{n,\gamma,b_1}(W)} \times \frac{1 + b_{2,n}(Y - f(V_\theta) - b_{1,n}(V_\theta)/\sqrt{n}, X)}{1 + b_2(Y - f(V_\theta) - b_1(V_\theta)/\sqrt{n}, X)}.$$

[8] Here $L_2(A, w) := L_2(A, \mathcal{A}, \mu, w)$ for a measure space $(A, \mathcal{A}, \mu)$ and a weight function $w : A \to [0, \infty)$ is the weighted L_2 space consisting of all functions f such that $\int f^2 w \, \mathrm{d}\mu < \infty$. This is just a L_2 space: one has that $L_2(A, \mathcal{A}, \mu, w) = L_2(A, \mathcal{A}, \nu)$ where $\nu(S) := \int_S w \, \mathrm{d}\mu$. In the main text, I omit the σ-field $\mathcal{A}$ and measure μ from the notation: the σ-field is always the Borel σ-field and the measure should be clear from context.

[9] If there exists a $C > 0$ such that $\mathbb{E}\left[\phi(\epsilon, X)^2 | X\right] \leq C < \infty$, then the converse also holds and the two norms are equivalent on this subspace.

By Assumption 4.2, Remark 3.1 and Lemma S2.4 it suffices to show that each of

$$\sum_{i=1}^{n} \log \frac{q_{n,\gamma,b_{1,n}}(W_i)}{q_{n,\gamma,b_1}(W_i)}, \quad \sum_{i=1}^{n} \log \frac{1 + b_{2,n}(Y_i - f(V_{\theta,i}) - b_{1,n}(V_{\theta,i})/\sqrt{n}, X_i)/\sqrt{n}}{1 + b_2(Y_i - f(V_{\theta,i}) - b_1(V_{\theta,i})/\sqrt{n}, X_i)/\sqrt{n}}, \tag{S14}$$

are $o_{P_{n,\gamma}}(1)$. For the former we use Lemma S2.3 applied to the measures Q_{n,γ,b_1} on $\mathcal{W}^n$ which correspond to q^n_{n,γ,b_1}; note that $Q_{n,\gamma,b_1} = P_{n,\gamma,(0,b_1,0)}$. The set Γ is a subset of a linear space by Asssumption 4.1. Equip the linear space $C^1_b(\mathscr{D})$ with the seminorm

$$\|b_1\| := \sqrt{\int b_1(x_1 + x_2'\theta)^2 \phi(e,x)^2 \zeta(e,x)\,\mathrm{d}\nu},$$

and let φ be the linear map $\varphi(b_1) := (0, b_1, 0)$. Similarly to as noted in the proof of Proposition S3.1, each $\gamma + s\varphi(b_1) \in \Gamma$ if $\mathcal{V}$ is taken to be a small enough neighbourhood of 0 in $C^1_b(\mathscr{D})$; this also ensures that condition (i) of Lemma S2.3 holds given Assumption 4.1. Condition (ii) holds by the Assumption that $e \mapsto \sqrt{\zeta(e,x)}$ is continuously differentiable and the chain rule. For condition (iii) with $u_{\gamma,b_1,s}(w) := \frac{\partial \log p_{\gamma + t(0,b_1,0)}(w)}{\partial t}_{|t=s}$ we have (cf. equation (S10))

$$u_{\gamma_\star,b_1,s}(W) = -\phi(Y - f(V_\theta) - \tilde{b}_1(V_\theta) - sb_1(V_\theta), X) b_1(V_\theta),$$

for $\gamma_\star = \gamma + (0, \tilde{b}_1, 0)$, $\tilde{b}_1 \in \mathcal{V}$. Under $P_{\gamma_\star + s\varphi(b_1)}$,

$$\left(Y - f(V_\theta) - \tilde{b}_1(V_\theta) - sb_1(V_\theta),\, X\right) \sim \zeta.$$

Hence, with $(\epsilon, X) \sim \zeta$,

$$P_{\gamma_\star + s\varphi(b_1)}\left[u^2_{\gamma_\star,b_1,s}\right] \leq \mathbb{E}\left[\phi(\epsilon, X)^2 b_1(V_\theta)^2\right] = \|b_1\|^2.$$

By Lemma S2.3 one then has that $\{b_1 \mapsto Q_{n,\gamma,b_1} : n \geq N\}$ is equicontinuous in total variation for some $N \in \mathbb{N}$. By the definition of the seminorm in (S12), if $h_n \to h$, then $b_{1,n} \to b_1$ and so $d_{TV}(Q_{n,\gamma,b_1}, Q_{n,\gamma,b_{1,n}}) \to 0$. By Lemma 2.4 in Strasser (1985)

$$\lim_{n\to\infty} \int \left| \frac{q^n_{n,\gamma,b_{1,n}}}{p^n_{n,\gamma}} - \frac{q^n_{n,\gamma,b_1}}{p^n_{n,\gamma}} \right| \mathrm{d}P_{n,\gamma} = 0 \implies \prod_{i=1}^{n} \frac{q_{n,\gamma,b_{1,n}}(W_i)}{p_{n,\gamma}(W_i)} - \prod_{i=1}^{n} \frac{q_{n,\gamma,b_1}(W_i)}{p_{n,\gamma}(W_i)} = o_{P_{n,\gamma}}(1),$$

and hence

$$\sum_{i=1}^{n}\left[\log q_{n,\gamma,b_{1,n}}(W_i) - \log q_{n,\gamma,b_1}(W_i)\right] = \sum_{i=1}^{n}\log\frac{q_{n,\gamma,b_{1,n}}(W_i)}{q_{n,\gamma,b_1}(W_i)} = o_{P_{n,\gamma}}(1)$$

by the continuous mapping theorem, which establishes the required condition for the first term in (S14). For the second term, let $\tilde{h}_n := (0, b_{1,n}, 0)$. Since $d_{TV}(P_{n,\gamma,(0,b_1,0)}, P_{n,\gamma,\tilde{h}_n}) \to 0$, $P_{n,\gamma,(0,b_1,0)} \triangleleft \triangleright P_{n,\gamma,\tilde{h}_n}$. Combined with Remark 3.1 and Assumption 4.2 this reveals that it suffices to show that that the sum on the right hand side of (S14) converges to zero in $P_{n,\gamma,\tilde{h}_n}$ – probability. For this, we verify the conditions of Lemma S3.4, which we will apply with the arrays formed by $(e_{n,b_{1,n},i}, X_i)$ and $(e_{n,b_1,i}, X_i)$, where $e_{n,b_1,i} := Y_i - f(V_{\theta,i}) - b_1(V_{\theta,i})/\sqrt{n}$. Under $P_{n,\gamma,\tilde{h}_n}$, $(e_{n,b_{1,n},i}, X_i) \sim \zeta$ and so

$$\mathbb{E}\left[b_{2,n}(e_{n,b_{1,n},i}, X_i) - b_2(e_{n,b_{1,n},i}, X_i)\right]^2 \leq \|b_{2,n} - b_2\|^2_{L_2(\zeta)} \to 0 \qquad \text{(S15)}$$

Note that

$$e_{n,b_{1,n},i} - e_{n,b_1,i} = \frac{b_1(V_{\theta,i}) - b_{1,n}(V_{\theta,i})}{\sqrt{n}},$$

and since b_2 is continuously differentiable in the first argument, with bounded derivative,

$$|b_2(e_{n,b_{1,n},i}, X_i) - b_2(e_{n,b_1,i}, X_i)| \lesssim \frac{|b_1(V_{\theta,i}) - b_{1,n}(V_{\theta,i})|}{\sqrt{n}}. \qquad \text{(S16)}$$

By equation (S13), $\mathbb{E}[b_{1,n}(V_{\theta,i}) - b_1(V_{\theta,i})]^2 \to 0$ under ζ and hence – analogously to in (S15) –

$$\mathbb{E}\left[b_2(e_{n,b_{1,n},i}, X_i) - b_2(e_{n,b_1,i}, X_i)\right]^2 \lesssim \|b_{1,n} - b_1\|^2_{L_2(\zeta)} \to 0.$$

Since the data are i.i.d across rows, this verifies (S32). For (S33) by the definition of $B_{\gamma,2}$, under $P_{n,\gamma,\tilde{h}_n}$, $\mathbb{E}b_{2,n}(e_{n,b_{1,n},i}, X_i) = \mathbb{E}b_2(e_{n,b_{1,n},i}, X_i) = 0$ and by (S16),

$$\mathbb{E}\left[\left|b_2(e_{n,b_{1,n},i}, X_i) - b_2(e_{n,b_1,i}, X_i)\right|\right] \lesssim n^{-1/2}\|b_{1,n} - b_1\|_{L_1(\zeta)} = o(n^{-1/2}).$$

As data are i.i.d. across rows, this verifies (S33). Apply Lemma S3.4. □

Asymptotic equicontinuity on class of subsets of H_γ To show asymptotic equicontinuity on subsets of H_γ note that H_γ is itself a linear space and equip it with the seminorm

$$\|h\| := \|(\tau, b_1, b_2)\| := \sqrt{\|\tau\|^2 + \sup_{v\in\mathscr{D}} |b_1(v)|^2 + \|b_2\|_{L_2(\mathbb{R}^{1+K},\varsigma)^2}}. \quad \text{(S17)}$$

Under the pseudometric $d_2(h_1, h_2) := \|h_1 - h_2\|$ induced by this norm, I show asymptotic equicontinuity in total variation over the subsets $K_\gamma^B := \{(\tau, b_1, b_2) \in H_\gamma : \sup_{v\in\mathscr{D}} |b_1'(v)| \leq B\}$.

PROPOSITION S3.3: *Let $H_{\gamma,0} = \mathbb{R}^{d_\theta} \times B_{\gamma,1}^A \times B_{\gamma,2}$ and equip it with the pseudometric d_2 induced by the seminorm (S17). Suppose that Assumptions 4.1, 4.2 hold and that $e \mapsto \sqrt{\varsigma(e,x)}$ is continuously differentiable. Then $h \mapsto P_{n,\gamma,h}$ is asymptotically equicontinuous in total variation on each K_γ^B.*

Proof. Let $h_n = (\tau_n, b_{1,n}, b_{2,n}) \to (\tau, b_1, b_2) = h$ with $h_n, h \in K_\gamma^B$. It suffices to show that $\lim_{n\to\infty} d_{TV}(P_{n,\gamma,h_n}, P_{n,\gamma,h}) = 0$. Let $q_{n,\gamma,\tau,b_1} := p_{(\theta+\tau/\sqrt{n}, f+b_1/\sqrt{n},\varsigma)}$ such that

$$\frac{p_{n,\gamma,h_n}(W)}{p_{n,\gamma,h}(W)} = \frac{q_{n,\gamma,\tau_n,b_{1,n}}(W)}{q_{n,\gamma,\tau,b_1}(W)} \times \frac{1 + b_{2,n}(Y - f(V_{\tilde{\theta}_n}) - b_{1,n}(V_{\theta_n})/\sqrt{n}, X)}{1 + b_2(Y - f(V_{\theta_n}) - b_1(V_{\theta_n})/\sqrt{n}, X)},$$

where $\tilde{\theta}_n := \theta + \tau_n/\sqrt{n}$ and $\theta_n := \theta + \tau/\sqrt{n}$. By Assumption 4.2, Remark 3.1 and Lemma S2.4 it suffices to show that each of

$$\sum_{i=1}^n \log \frac{q_{n,\gamma,\tau_n,b_{1,n}}(W_i)}{q_{n,\gamma,\tau,b_1}(W_i)}, \quad \sum_{i=1}^n \log \frac{1 + b_{2,n}(Y_i - f(V_{\tilde{\theta}_n,i}) - b_{1,n}(V_{\tilde{\theta}_n,i})/\sqrt{n}, X_i)/\sqrt{n}}{1 + b_2(Y_i - f(V_{\theta_n,i}) - b_1(V_{\theta_n,i})/\sqrt{n}, X_i)/\sqrt{n}}, \quad \text{(S18)}$$

are $o_{P_{n,\gamma}}(1)$. For the former we will appeal to Lemma S2.3. Let Q_{n,γ,τ,b_1} be the measure on $\mathcal{W}^n$ corresponding to q^n_{n,γ,τ,b_1} and note that this is equal to $P_{n,\gamma,(\tau,b_1,0)}$. The set Γ is a subset of a linear space by Asssumption 4.1. The set $\mathbb{R}^{d_\theta} \times B_{\gamma,1}$ can be viewed as a normed linear space by equipping it with the norm $\|(\tau, b_1)\| := \|(\tau, b_1, 0)\|$ where the right hand side norm is that in (S17). $\tilde{K}_\gamma^B := \{(\tau, b_1) \in \mathbb{R}^{d_\theta} \times B_{\gamma,1} : \sup_{v\in\mathscr{D}} |b_1'|(v) \leq B\}$ is a subset of this space. Define the linear map $\varphi(\tau, b_1) := (\tau, b_1, 0)$. Similarly to as noted in the proof of Proposition S3.1, each $\gamma + s\varphi(\tau, b_1) \in \Gamma$ if $\mathcal{V}$, the intersection of a neighbourhood of 0 in $\mathbb{R}^{d_\theta} \times B_{\gamma,1}$ and $\tilde{K}_\gamma^B$, is taken small enough, which also ensures that condition (i) of Lemma S2.3 holds given Assumption 4.1. Condition (ii) holds by the Assumption that $e \mapsto \sqrt{\varsigma(e,x)}$ is continuously differentiable and the chain rule. For condition (iii) with $u_{\gamma,\tau,b_1,s}(w) := \frac{\partial \log p_{\gamma+t(\tau,b_1,0)}(w)}{\partial t}_{|t=s}$ we have (cf. equation (S10))

$$\begin{aligned}
u_{\gamma_\star,\tau,b_1,s}(W) = &-\phi(Y - f(V_{\theta+\tilde\tau+s\tau}) - \tilde b_1(V_{\theta+\tilde\tau+s\tau}) - sb_1(V_{\theta+\tilde\tau+s\tau}), X) \\
&\times \left[f'(V_{\theta+\tilde\tau+s\tau})X_2'\tau + \tilde b_1'(V_{\theta+\tilde\tau+s\tau})X_2'\tau + sb_1'(V_{\theta+\tilde\tau+s\tau})X_2'\tau + b_1(V_{\theta+\tilde\tau+s\tau})\right],
\end{aligned}$$

for $\gamma_\star = \gamma + (\tilde\tau, \tilde b_1, 0)$, $(\tilde\tau, \tilde b_1, 0) \in \mathcal{V}$. Under $P_{\gamma_\star+s\varphi(\tau,b_1)}$,

$$\left(Y - f(V_{\theta+\tilde\tau+s\tau}) - \tilde b_1(V_{\theta+\tilde\tau+s\tau}) - sb_1(V_{\theta+\tilde\tau+s\tau}),\ X\right) \sim \zeta.$$

By the definition of $\mathcal{V}$ and $f \in C_b^1(\mathscr{D})$, $|f'| \leq F$, $|\tilde b_1'| \leq B$ and $|b_1'| \leq B$. Hence, for $(\epsilon, X) \sim \zeta$,

$$\begin{aligned}
P_{\gamma_\star+s\varphi(h)}\left[u^2_{\gamma_\star,\tau,b_1,s}\right] &\leq 2\mathbb{E}\left[\phi(\epsilon,X)^2(2B+F)^2(X_2'\tau)^2\right] + 2\mathbb{E}\left[\phi(\epsilon,X)^2 b_1(V_{\theta+\tilde\tau+s\tau})^2\right] \\
&\leq 2(2B+F)^2\|\tau\|^2\mathbb{E}\left[\phi(\epsilon,X)^2\|X_2\|^2\right] + 2\mathbb{E}\left[\phi(\epsilon,X)^2\right]\sup_{v\in\mathscr{D}}|b_1(v)|^2 \\
&\lesssim \|(\tau,b_1)\|^2
\end{aligned}$$

by Assumption 4.1. By Lemma S2.3 one then has that $\{(\tau,b_1) \mapsto Q_{n,\gamma,\tau,b_1} : n \geq N\}$ is equicontinuous in total variation on each $\tilde K_\gamma^B$ for some $N \in \mathbb{N}$. By the definition of the seminorm in (S12), if $h_n \to h$, then $\|(\tau_n, b_{1,n}) - (\tau, b_1)\| \to 0$ and so $d_{TV}(Q_{n,\gamma,\tau,b_1}, Q_{n,\gamma,\tau,b_{1,n}}) \to 0$. By Lemma 2.4 in Strasser (1985)

$$\lim_{n\to\infty}\int\left|\frac{q^n_{n,\gamma,\tau_n,b_{1,n}}}{p^n_{n,\gamma}} - \frac{q^n_{n,\gamma,\tau,b_1}}{p^n_{n,\gamma}}\right| \mathrm{d}P_{n,\gamma} = 0 \implies \prod_{i=1}^n\frac{q_{n,\gamma,\tau_n,b_{1,n}}(W_i)}{p_{n,\gamma}(W_i)} - \prod_{i=1}^n\frac{q_{n,\gamma,\tau,b_1}(W_i)}{p_{n,\gamma}(W_i)} = o_{P_{n,\gamma}}(1),$$

and hence

$$\sum_{i=1}^n\left[\log q_{n,\gamma,\tau_n,b_{1,n}}(W_i) - \log q_{n,\gamma,\tau,b_1}(W_i)\right] = \sum_{i=1}^n \log\frac{q_{n,\gamma,\tau_n,b_{1,n}}(W_i)}{q_{n,\gamma,\tau,b_1}(W_i)} = o_{P_{n,\gamma}}(1)$$

by the continuous mapping theorem, which establishes the required condition for the first term in (S18). For the second term, let $\tilde h_n := (\tau_n, b_{1,n}, 0)$. Since $d_{TV}(P_{n,\gamma,(\tau,b_1,0)}, P_{n,\gamma,\tilde h_n}) \to 0$, $P_{n,\gamma,(\tau,b_1,0)} \triangleleft \triangleright P_{n,\gamma,\tilde h_n}$. Combined with Remark 3.1 and Assumption 4.2 this reveals that it suffices to show that that the sum on the right hand side of (S18) converges to zero in $P_{n,\gamma,\tilde h_n}$ – probability. For this, we verify the conditions of Lemma S3.4, which we will apply with the arrays formed by $(e_{n,\tau_n,b_{1,n},i}, X_i)$ and $(e_{n,\tau,b_1,i}, X_i)$, where $e_{n,\tau,b_1,i} := Y_i - f(V_{\theta+n^{-1/2}\tau,i}) -$

$b_1(V_{\theta+n^{-1/2}\tau,i})/\sqrt{n}$. Under $P_{n,\gamma,\tilde{h}_n}$, $(e_{n,\tau_n,b_{1,n},i}, X_i) \sim \zeta$ and so

$$\mathbb{E}\left[b_{2,n}(e_{n,\tau_n,b_{1,n},i}, X_i) - b_2(e_{n,\tau_n,b_{1,n},i}, X_i)\right]^2 \leq \|b_{2,n} - b_2\|^2_{L_2(\zeta)} \to 0 \tag{S19}$$

Note that

$$e_{n,\tau_n,b_{1,n},i} - e_{n,\tau,b_1,i} = \frac{b_1(V_{\theta+n^{-1/2}\tau,i}) - b_{1,n}(V_{\theta+n^{-1/2}\tau_n,i})}{\sqrt{n}} + f(V_{\theta+n^{-1/2}\tau,i}) - f(V_{\theta+n^{-1/2}\tau_n,i}).$$

and since b_2 is continuously differentiable in the first argument with bounded derivative, $f \in C^1_b(\mathscr{D})$, and b_1 has its derivative uniformly bounded by B,

$$\begin{aligned} &|b_2(e_{n,\tau_n,b_{1,n},i}, X_i) - b_2(e_{n,\tau,b_1,i}, X_i)| \\ &\qquad \lesssim \frac{\|X_2\|\|\tau - \tau_n\|}{n} + \frac{\sup_{v\in\mathscr{D}} |b_1(v) - b_{1,n}(v)| + \|X_2\|\|\tau - \tau_n\|}{\sqrt{n}} \end{aligned} \tag{S20}$$

By the choice of norm and Assumption 4.1 the expected value of the square of the right hand side converges to zero, hence

$$\mathbb{E}\left[b_2(e_{n,\tau_n,b_{1,n},i}, X_i) - b_2(e_{n,\tau,b_1,i}, X_i)\right]^2 \to 0.$$

Since the data are i.i.d across rows, this verifies (S32). For (S33) by the definition of $B_{\gamma,2}$, under $P_{n,\gamma,\tilde{h}_n}$, $\mathbb{E}b_{2,n}(e_{n,\tau_n,b_{1,n},i}, X_i) = \mathbb{E}b_2(e_{n,\tau_n,b_{1,n},i}, X_i) = 0$ and by (S20) and Assumption 4.1

$$\mathbb{E}\left[\left|b_2(e_{n,b_{1,n},i}, X_i) - b_2(e_{n,b_1,i}, X_i)\right|\right] \lesssim n^{-1/2}\left[\sup_{v\in\mathscr{D}} |b_1(v) - b_{1,n}(v)| + \|\tau - \tau_n\|\right] = o(n^{-1/2}).$$

As data are i.i.d. across rows, this verifies (S33). Apply Lemma S3.4. □

S3.2 IV model with non-parametric first stage

S3.2.1 Proofs of results in the main text

Proof of Lemma 4.1. We first note that $\mathbb{E}[E_iE_j|Z] = [J(Z)^{-1}]_{i,j}$ for $i, j \in \{1, 2\}$ where $J(Z)$ is nonsingular by equation (37). The same equation combined with Proposition 2.8.4 in Bernstein (2009) also implies that $q_1(J(Z))$ exists and is positive. Letting $\xi = \xi_\gamma(W) = (Y - X'\theta - Z_1'\beta, X - \pi(Z), Z)$ we have

$$\dot{\imath}_\gamma(W) := \left(\dot{\imath}_{\gamma,1}(W)', \dot{\imath}_{\gamma,2}(W)'\right)' = -\phi_1(\xi)(X', Z_1')'$$
$$[D_{\gamma,1}b_1](W) := -\phi_2(\xi)'b_1(Z), \quad [D_{\gamma,2}b_2](W) := b_2(\xi).$$

We first project $\dot{l}_\gamma$ and $D_{\gamma,1}b_1$ onto the orthocomplement of $\{D_{\gamma,2}b_2 : b_2 \in B_{\gamma,2}\}$. By Corollary S2.2 and Proposition A.3.5 in Bickel et al. (1998) these projections are, respectively:

$$\breve{l}_\gamma(W) = \mathbb{E}\left[-\phi_1(\xi)[X', Z_1']'U'|Z\right]\mathbb{E}\left[UU'|Z\right]^{-1}U$$
$$[\breve{D}_{\gamma,1}b_1](W) = \mathbb{E}\left[-b_1(Z)'\phi_2(\xi)U'|Z\right]\mathbb{E}\left[UU'|Z\right]^{-1}U$$

for $U = (\epsilon, v')' = (Y - X'\theta - Z_1'\beta, X' - \pi(Z)')'$. Denoting $K := d_\beta$, and evaluating the first conditional expectation using (37) we obtain:

$$\breve{l}_\gamma(W) = \begin{bmatrix}\pi(Z)\\ Z_1\end{bmatrix}\begin{bmatrix}1 & 0_K'\end{bmatrix} J(Z)^{-1}U = \pi(Z)V_1 \ ,$$
$$[\breve{D}_{\gamma,1}b_1](W) = b_1(Z)'\begin{bmatrix}0_K & I_K\end{bmatrix} J(Z)^{-1}U = b_1(Z)'V_2 \ .$$

By Propositions S2.3 and S2.4 $\tilde{l}_\gamma$ can now be found by projecting $\breve{l}_\gamma$ onto the orthocomplement of $\{\breve{D}_{\gamma,1}b_1 : b_1 \in B_{\gamma,1}\} = \{b_1(Z)'V_2 : b_1 \in B_{\gamma,1}\}$. That this projection is $\tilde{l}_\gamma$ follows from the observation that (a)

$$\begin{aligned}
\mathbb{E}\left[\tilde{l}_\gamma(W)b_1(Z)'V_2\right] &= \mathbb{E}\left[[\pi(Z)', Z_1']'\left[V_1 - \mathbb{E}[V_1V_2'|Z]\mathbb{E}[V_2V_2'|Z]^{-1}V_2\right]V_2'b_1(Z)\right]\\
&= \mathbb{E}\left[[\pi(Z)', Z_1']'\,\mathbb{E}\left[V_1V_2' - \mathbb{E}[V_1V_2'|Z]\mathbb{E}[V_2V_2'|Z]^{-1}V_2V_2'|Z\right]b_1(Z)\right]\\
&= \mathbb{E}\left[[\pi(Z)', Z_1']'\left[\mathbb{E}[V_1V_2'|Z] - \mathbb{E}[V_1V_2'|Z]\mathbb{E}[V_2V_2'|Z]^{-1}\mathbb{E}[V_2V_2'|Z]\right]b_1(Z)\right]\\
&= 0,
\end{aligned}$$

and (b) the components of $[\pi(Z)', Z_1']'\,\mathbb{E}[V_1V_2'|Z]\mathbb{E}[V_2V_2'|Z]^{-1}V_2$ are in in $\mathrm{cl}\{b_1(Z)'V_2 : b_1 \in B_{\gamma,1}\}$. This follows as $\mathbb{E}\|\,[\pi(Z)', Z_1']'\,\mathbb{E}[V_1V_2'|Z]\mathbb{E}[V_2V_2'|Z]^{-1}V_2\|^2 < \infty$ by equation (37) and Assumption 4.4, and any such component may be arbitrarily well approximated by $b_{1,n}(Z)'V_2$ for bounded measurable functions $b_{1,n} \in B_{\gamma,1}$ since the set of such functions is dense in L_2. The first equality in the final claim follows from Example A.2.1 in Bickel et al. (1998). For the second equality note that with $Q(Z) := J(Z)^{-1}$

$$\begin{aligned}
E_1 - Q(Z)_{1,2}Q(Z)_{2,2}^{-1}E_2 &= \left[\begin{bmatrix}1 & 0_K'\end{bmatrix} - Q(Z)_{1,2}Q(Z)_{2,2}^{-1}\begin{bmatrix}0_K & I_K\end{bmatrix}\right]Q(Z)U\\
&= \left[\begin{bmatrix}Q(Z)_{1,1} & Q(Z)_{1,2}\end{bmatrix} - Q(Z)_{1,2}Q(Z)_{2,2}^{-1}\begin{bmatrix}Q(Z)_{2,1} & Q(Z)_{2,2}\end{bmatrix}\right]U\\
&= \begin{bmatrix}Q(Z)_{1,1} - Q(Z)_{1,2}Q(Z)_{2,2}^{-1}Q(Z)_{2,1} & 0_K\end{bmatrix}U.
\end{aligned}$$

The result then follows from the the block matrix inversion formula (e.g. Proposition 2.8.7 in Bernstein (2009)) and direct calculation. ☐

Proof of Lemma 4.2. For the first part of the Lemma observe that by the moment conditions in Assumption 4.4 and the bounds on $J(Z)$ in (37), $\bar{\ell}_\gamma \in L_2(P_\gamma)$. Moreover, by the definition of $B_{\gamma,2}$, with $M \coloneqq \mathbb{E}[XZ_1']\mathbb{E}[Z_1Z_1']^{-1}$,

$$\mathbb{E}\left[\bar{\ell}_\gamma(W)b_2(\xi)\right] = q_1(\bar{J})\mathbb{E}\left[(\pi(Z) - MZ_1)\mathbb{E}\left[\epsilon b_2(U,Z)|Z\right]\right] = 0;$$

by equation (37):

$$\mathbb{E}\left[\bar{\ell}_\gamma(W)\phi_2(\xi)'b_1(Z)\right] = q_1(\bar{J})\mathbb{E}\left[(\pi(Z) - MZ_1)\mathbb{E}\left[\epsilon\phi_2(\epsilon,v,Z)'|Z\right]b_1(Z)\right] = 0;$$

and since $\mathbb{E}[v|Z] = 0$ by Assumption 4.4,

$$\begin{aligned}
\mathbb{E}\left[\bar{\ell}_\gamma(W)\phi_1(\xi)b_0'Z_1\right] &= q_1(\bar{J})\mathbb{E}\left[\mathbb{E}\left[\epsilon\phi_1(\epsilon,v,Z)|Z\right](\pi(Z)Z_1' - MZ_1Z_1')\right]b_0 \\
&= -q_1(\bar{J})\left[\mathbb{E}[\pi(Z)Z_1'] - \mathbb{E}[XZ_1']\mathbb{E}[Z_1Z_1']^{-1}\mathbb{E}[Z_1Z_1']\right]b_0 \\
&= -q_1(\bar{J})\left[\mathbb{E}[\pi(Z)Z_1'] - \mathbb{E}[\pi(Z)Z_1']\mathbb{E}[Z_1Z_1']^{-1}\mathbb{E}[Z_1Z_1']\right]b_0 \\
&= 0.
\end{aligned}$$

The the second claim in the Lemma follows immediately if $J(Z) = \bar{J}$ a.s.. ☐

Proof of Proposition 4.3. Assumptions 4.4, 4.5 and Lemma 4.2 (with equation (37)) establish that the conditions of Lemma 3.8 are satisfied. ☐

Proof of Proposition 4.4. Let $g_\gamma \coloneqq \bar{\ell}_\gamma$ for consistency of notation. Let $\beta_n = \beta + b_{n,0}/\sqrt{n}$ with $b_{n,0} \to b_0 \in \mathbb{R}^{d_\beta}$. Let $\breve{U}_{n,i}$, $\breve{J}_n$, $\breve{E}_{n,i}$, $\breve{g}_{n,\theta,i}$, $\breve{V}_{n,\theta}$, $\breve{\Lambda}_{n,\theta}$ and $\breve{r}_{n,\theta}$ be formed analogously to $\hat{U}_{n,i}$, $\hat{J}_n$, $\hat{E}_{n,i}$, $\hat{g}_{n,\theta,i}$, $\hat{V}_{n,\theta}$, $\hat{\Lambda}_{n,\theta}$ and $\hat{r}_{n,\theta}$ (as defined in and around equations (40) – (42)) with β_n in place of $\hat{\beta}_n$. By Assumption 4.6, $\hat{\beta}_n \in \mathscr{S}_n$. It suffices to show that Assumption 3.3 holds for $\breve{g}_{n,\theta} \coloneqq \frac{1}{\sqrt{n}}\sum_{i=1}^n \breve{g}_{n,\theta,i}$, $\breve{\Lambda}_{n,\theta}$ and $\breve{r}_{n,\theta}$ (e.g. Hoesch, Lee, and Mesters, 2024, Lemma S3.1).

For Assumption 3.3 part (i), by Lemma S3.1,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \left[\breve{g}_{n,\theta,i} - g_\gamma(W_i)\right] = \sum_{l=1}^4 \mathsf{R}_{n,l} = o_{P_{n,\gamma}}(1).$$

For Assumption 3.3 parts (ii) and (iii) we first establish the rate of convergence of

$\check{V}_{n,\theta}$. We have

$$\check{V}_{n,\theta} - V_\gamma = \frac{1}{n}\sum_{i=1}^{n}\left[\breve{g}_{n,\theta,i}\breve{g}_{n,\theta,i}' - g_\gamma(W_i)g_\gamma(W_i)'\right] + \frac{1}{\sqrt{n}}\mathbb{G}g_\gamma g_\gamma'. \tag{S21}$$

For the first right hand side term, we have by Cauchy — Schwarz,

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\left[\breve{g}_{n,\theta,i}\breve{g}_{n,\theta,i}' - g_\gamma(W_i)g_\gamma(W_i)'\right]\right\| \lesssim \left[\sum_{l=1}^{4}\mathsf{S}_{n,l}\right]\left[\frac{1}{n}\sum_{i=1}^{n}\|\breve{g}_{n,\theta,i}\|^2 + \frac{1}{n}\sum_{i=1}^{n}\|g_\gamma(W_i)\|^2\right].$$

Under Assumption 4.4, $\mathbb{E}\|g_\gamma(W_i)\|^2 < \infty$, hence $\frac{1}{n}\sum_{i=1}^{n}\|g_\gamma(W_i)\|^2 = O_{P_{n,\gamma}}(1)$ by Markov's inequality. By Lemma S3.1 $\frac{1}{n}\sum_{i=1}^{n}\|\breve{g}_{n,\theta,i} - g_\gamma(W_i)\|^2 \lesssim \sum_{l=1}^{4}\mathsf{S}_{n,l} = O_{P_{n,\gamma}}(\delta_n^2 + n^{-1})$. Using this gives

$$\frac{1}{n}\sum_{i=1}^{n}\|\breve{g}_{n,\theta,i}\|^2 \lesssim \frac{1}{n}\sum_{i=1}^{n}\|\breve{g}_{n,\theta,i} - g_\gamma(W_i)\|^2 + \frac{1}{n}\sum_{i=1}^{n}\|g_\gamma(W_i)\|^2 = O_{P_{n,\gamma}}(1),$$

hence the first right hand side term in (S21) is $O_{P_{n,\gamma}}(\delta_n^2 + n^{-1})$ as $\sum_{l=1}^{4}\mathsf{S}_{n,l} = O_{P_{n,\gamma}}(\delta_n^2 + n^{-1})$ by Lemma S3.1.

For the second right hand side term, by Assumption 4.6 $\mathbb{E}\|g_\gamma(W)\|^4 < \infty$. Hence by the central limit theorem $\mathbb{G}g_\gamma g_\gamma' = O_{P_{n,\gamma}}(1)$ and so the second right hand side term in (S21) is $O_{P_{n,\gamma}}(n^{-1/2})$. Putting these pieces together yields $\|\check{V}_{n,\theta} - V_\gamma\| = O_{P_{n,\gamma}}(\delta_n^2 + n^{-1/2})$. The proof is then complete by the choice of rate in Assumption 4.6 and Proposition S2.1. □

S3.2.2 The LAN condition

Here I provide examples of local perturbations $P_{n,\gamma,h}$ and lower level conditions under which the LAN condition in Assumption 4.5 holds. Let

$$\varphi_{n,1}(b_1) := b_1/\sqrt{n}, \quad \varphi_{n,2}(b_2) := \zeta b_2/\sqrt{n}, \qquad (b_1, b_2) \in B_{\gamma,1} \times B_{\gamma,2}, \tag{S22}$$

where $B_{\gamma,1}$ is the space bounded functions $b_1 : \mathbb{R}^{d_Z} \to \mathbb{R}^{d_\theta}$ and $B_{\gamma,2}$ the space of bounded functions $b_2 : \mathbb{R}^K \to \mathbb{R}$ which are continuously differentiable in their first $1 + d_\theta$ components with bounded derivative and such that (35) hold.

Proposition S3.4: *Let $H_\gamma := \mathbb{R}^{d_\theta} \times B_\gamma$. Then, if Assumption 4.4 holds, $u \mapsto \sqrt{\zeta(u,z)}$ is continuously differentiable and $p_{n,\gamma,h} = p^n_{\gamma+\varphi_n(h)}$ with p_γ as in (33) and φ_n as in (34) & (S22), Assumption* **??** *holds.*

Proof. We being by noting that due to the definition of $B_{\gamma,1}$ and $B_{\gamma,2}$, for all large enough n each $p_{\gamma+\varphi_n(h)}$ is a valid density and $\gamma + \varphi_n(h) \in \Gamma$. Given the product construction of $P_{n,\gamma,h}$ Assumption 3.5 is satisfied and to apply Lemma 3.7 it remains to verify differentiability in quadratic mean as in (20) (with $h_n = h$).

Let $q_{\gamma,\tau,b,t} := p_{(\theta,\eta)+t(\tau,(b_0,b_1,b_2\zeta))}$, $t \in [0,\infty)$ and abbreviate $q_\gamma := q_{\gamma,0,0,0}$. Analogously to above, for all small enough τ, b and t, this is a probability density and $(\theta,\eta) + t(\tau,(b_0,b_1,b_2\zeta)) \in \Gamma$. Letting

$$\dot{l}_\gamma(W) := -\phi(Y - X'\theta - Z_1'\beta, X - \pi(Z), Z)\begin{bmatrix} X \\ Z_1 \end{bmatrix},$$

it suffices to show that

$$\int \left[q_{\gamma,\tau,b,t}^{1/2} - q_\gamma^{1/2} - \frac{t}{2}\left((\tau',b_0')\dot{l}_\gamma - \phi' b_1 + b_2\right) q_\gamma^{1/2}\right]^2 \mathrm{d}\nu = o(t^2) \quad \text{as } t \downarrow 0. \tag{S23}$$

For this we will verify the conditions of Lemma 7.6 in van der Vaart (1998). That $t \mapsto \sqrt{q_{\gamma,\tau,b,t}}$ is continuously differentiable follows from the assumed continuous differentiability of $(e,v) \mapsto \sqrt{\zeta(e,v,z)}$. Denote by ι_s the derivative of $t \mapsto \log q_{\gamma,\tau,b,t}$ at $t = s$. Under $q_{\gamma,\tau,b,s}$, $\iota_s(W)$ has the same law as

$$\begin{aligned} E_s := &-\phi_1(\epsilon,\upsilon,Z)[X',Z_1'](\tau',b_0')' - \phi_2(\epsilon,\upsilon,Z)'b_1(Z) \\ &+ \frac{b_2(\epsilon,\upsilon,Z) - sb_{2,1}(\epsilon,\upsilon,Z)[X',Z_1'](\tau',b_0') - sb_{2,2}(\epsilon,\upsilon,Z)'b_1(Z)}{1 + sb_2(\epsilon,\upsilon,Z)}, \end{aligned}$$

where $b_{2,i}$ indicates the derivative of $(e,v) \mapsto b_2(e,v,z)$ in the i-th argument. It suffices to show that $\mathbb{E}E_{s_n}^2 \to \mathbb{E}E_s^2$ for all $s_n \to s$ in a neighbourhood of zero. In particular, take a neighbourhood $\mathcal{U} := [0,\delta)$ such that $1 + sb_2(\epsilon,\upsilon,Z)$ is bounded below (as b_2 is bounded). That $E_{s_n}^2 \to E_s^2$ pointwise is evident, hence it suffices to demonstrate that $(E_{s_n}^2)_{n\in\mathbb{N}}$ is uniformly integrable. We do so by exhibiting a dominating function. Let

$$\overline{E^2} := C\left[\phi_1(\epsilon,\upsilon,Z)^2[\|X\|^2 + \|Z\|^2] + \|\phi_2(\epsilon,\upsilon,Z)\|^2 + \|X\|^2 + \|Z\|^2 + 1\right]$$

for some positive constant C. Provided C is taken large enough, by the moment conditions in Assumption 4.4 and the boundedness of b_1, b_2, $E_s^2 \leq \overline{E^2}$ a.s. and $\mathbb{E}\overline{E^2} < \infty$. Thus $(E_{s_n}^2)_{n\in\mathbb{N}}$ is uniformly integrable, concluding the demonstration that the conditions of Lemma 7.6 in van der Vaart (1998) are satisfied and hence, (S23) holds. □

Motivation for the conditions on b_2 Equation (35) imposes two conditions on b_2, i.e. the score corresponding to the density function ζ. The first simply requires b_2 to be mean-zero, which is a requirement of *any* score function (cf. Assumption 3.1 or Lemma 1.7 in van der Vaart, 2002). The second requires that $\mathbb{E}[Ub_2(U, Z)|Z] = 0$. Here I will heuristically argue that "well-behaved" parametric submodels lead to scores with this property. This argument is essentially identical to that for the single index model given in Section S3.1.2. Let $\zeta_\beta(u, z)$ denote a parametric family of density functions of (U, Z) with respect to $\nu_1 \otimes \nu_2$ such that the marginal density of Z, $\iota(z) = \int \zeta_\beta(u, z)\,\mathrm{d}\nu_1(u)$, does not depend on β. Provided the parametric family is sufficiently well – behaved, scores for β in the model $\{\zeta_\beta : \beta \in \mathcal{B}\}$ for $\mathcal{B}$ some open set, have the form $\nabla_\beta \log \zeta_\beta = \varphi_\beta$. The conditional expectation of this score can be written (Z-a.s.)

$$\begin{aligned}\int u\varphi_\beta(u, Z)\frac{\zeta_\beta(u, Z)}{\int \zeta_\beta(u, Z)\,\mathrm{d}\nu_1(u)}\,\mathrm{d}\nu_1(u) &= \int u\frac{\nabla_\beta\zeta_\beta(u, Z)}{\zeta_\beta(u, Z)}\frac{\zeta_\beta(u, Z)}{\int \zeta_\beta(u, Z)\,\mathrm{d}\nu_1(u)}\,\mathrm{d}\nu_1(u)\\ &= \int u\frac{\nabla_\beta\zeta_\beta(u, Z)}{\int \zeta_\beta(u, Z)\,\mathrm{d}\nu_1(u)}\,\mathrm{d}\nu_1(u).\end{aligned}$$

Provided the derivatives exist, since $\nabla_\beta u = 0$ and $\nabla_\beta \iota(z) = 0$,

$$\nabla_\beta u\frac{\zeta_\beta(u, z)}{\int \zeta_\beta(u, z)\,\mathrm{d}\nu_1(u)} = u\frac{\nabla_\beta\zeta_\beta(u, z)}{\int \zeta_\beta(u, z)\,\mathrm{d}\nu_1(u)}.$$

Additionally (Z-a.s.)

$$\mathbb{E}[\epsilon|Z] = \int u\frac{\zeta_\beta(u, Z)}{\int \zeta_\beta(u, Z)\,\mathrm{d}\nu_1(u)}\,\mathrm{d}\nu_1(u) = 0 \quad \Longrightarrow \quad \nabla_\beta \int u\frac{\zeta_\beta(u, Z)}{\int \zeta_\beta(u, Z)\,\mathrm{d}\nu_1(u)}\,\mathrm{d}\nu_1(u) = 0.$$

If the last derivative can be taken inside the integral, combination of these displays yields

$$\int u\varphi_\beta(u, Z)\frac{\zeta_\beta(u, Z)}{\int \zeta_\beta(u, Z)\,\mathrm{d}\nu_1(u)}\,\mathrm{d}\nu_1(u) = \int \nabla_\beta u\frac{\zeta_\beta(u, Z)}{\int \zeta_\beta(u, Z)\,\mathrm{d}\nu_1(u)}\,\mathrm{d}\nu_1(u) = 0.$$

Thus any score φ_β in such a well-behaved parametric submodel must satisfy the property imposed on b_2.

Ensuring p_γ is a probability density The discussion here is similar to the case of the single index model treated in section S3.1.2. In particular, similarly to in that case, p_γ is a valid probability density with respect to $\tilde{\nu} = \nu$ if ν is invariant with respect to

$$F_\gamma(y, x, z) \coloneqq (y - x'\theta - z_1'\beta, x - \pi(z), z)$$

for each $\gamma \in \Gamma$, i.e. $[\nu \circ F_\gamma^{-1}] = \nu$. In this case evidently $p_\gamma \geq 0$ by (33) and by the invariance

$$\int p_\gamma \,\mathrm{d}\nu = \int \zeta \circ F_\gamma \,\mathrm{d}\nu = \int \zeta \,\mathrm{d}[\nu \circ F_\gamma^{-1}] = \int \zeta \,\mathrm{d}\nu = 1.$$

This invariance holds, for example, in the important special case where $U|Z$ is continuously distributed. Suppose that $U|Z$ has (conditional) density $\zeta_{1,z}$ with respect to Lebesgue measure λ and ζ_2 is the marginal density of Z with respect to some dominating measure ν_2. Then $\zeta(u, z) = \zeta_{1,z}(u)\zeta_2(z)$ is a density with respect to $\nu \coloneqq \lambda \otimes \nu_2$. Let $A = A_1 \times A_2 \subset \mathbb{R}^{1+d_\theta} \times \mathbb{R}^{d_Z}$ be a measurable rectangle and define

$$G_{\gamma,z}(y, x) = \begin{bmatrix} y - x'\theta - z_1'\beta \\ x - \pi(z) \end{bmatrix} \qquad \Longrightarrow \qquad J_{G_{\gamma,z}}(y, x) = \begin{bmatrix} 1 & -\theta' \\ 0 & I \end{bmatrix}.$$

Then $(u, z) \in F_\gamma(A)$ if and only if $z \in A_2$ and $u \in G_{\gamma,z}(A_1)$. By Tonelli's theorem

$$\nu(F_\gamma(A)) = \int \mathbf{1}_{A_2}(z) \left[\int \mathbf{1}_{G_{\gamma,z}(A_1)}(u) \,\mathrm{d}\lambda(u) \right] \mathrm{d}\nu_2(z) = \int_{A_2} \lambda(G_{\gamma,z}(A_1)) \,\mathrm{d}\nu_2(z).$$

Since $|\det J_{G_{\gamma,z}}(z)| = 1$, by change of variables

$$\lambda(G_{\gamma,z}(A_1)) = \int_{G_{\gamma,z}(A_1)} \mathrm{d}\lambda = \int_{A_1} |\det J_{G_{\gamma,z}}(z)| \,\mathrm{d}\lambda = \int_{A_1} \mathrm{d}\lambda = \lambda(A_1),$$

for each z. Hence,

$$[\nu \circ F_\gamma^{-1}](A) = \nu(F_\gamma(A)) = \int_{A_2} \lambda(G_{\gamma,z}(A_1)) \,\mathrm{d}\nu_2(z) = \int_{A_2} \lambda(A_1) \,\mathrm{d}\nu_2 = \lambda(A_1) \times \nu_2(A_2) = \nu(A).$$

Since the measurable rectangles form a separating class, $\nu \circ F_\gamma^{-1} = \nu$.

S3.2.3 Uniform local regularity

Here I discuss strengthening the local regularity of the proposed C(α) test as discussed in Section 4.2 to the uniform local regularity of Definition 2.2. I place a (pseudo-)metric structure on $H_\gamma \coloneqq \mathbb{R}^{d_\theta} \times B_\gamma$ as defined in Proposition S3.4. In particular, H_γ can be viewed as a linear subspace of[10]

$$\mathbb{R}^{d_\theta} \times \mathbb{R}^{d_\beta} \times \prod_{k=1}^{d_\theta} L_2(\mathbb{R}^K, \phi_{1+k}^2 \zeta) \times L_2(\mathbb{R}^{1+d_\theta}, \zeta), \tag{S24}$$

since we may identify each $b_1 \in B_{\gamma,1}$ with a $\tilde{b}_1 : \mathbb{R}^K \to \mathbb{R}^{d_\theta}$ according to $\tilde{b}_1(u,z) := b_1(z)$ and each such $\tilde{b}_1$ has components $\tilde{b}_{1,k} \in L_2(\mathbb{R}^K, \phi_{1+k}^2 \zeta)$ since $b_{1,k}$ is bounded and hence under Assumption 4.4

$$\int \tilde{b}_{1,k}(u,z)^2 \phi_{1+k}(u,z)^2 \zeta(u,z)\, \mathrm{d}\nu = \int b_{1,k}(z)^2 \phi_{1+k}(u,z)^2 \zeta(u,z)\, \mathrm{d}\nu < \infty.$$

Equip H_γ with the corresponding norm

$$\|h\| = \sqrt{\|(\tau', b_0')'\|_2^2 + \sum_{k=1}^{d_\theta} \int b_{1,k}(z)^2 \phi_{1+k}(u,z)^2 \zeta(u,z)\, \mathrm{d}\nu + \int b_2(u,z)^2 \zeta(u,z)\, \mathrm{d}\nu}. \tag{S25}$$

I will additionally assume that

$$\mathbb{E}\left[\phi_{1+k}(U,Z)^2 | Z\right] \geq c > 0, \qquad (U,Z) \sim \zeta. \tag{S26}$$

This is a mild condition which ensures that the $L_2(\mathbb{R}^K, \phi^2\zeta)$ norm is stronger than the $L_2(\mathbb{R}^K, \zeta)$ norm on considered subspace (cf. the discussion following equation (S13)). H_γ can then be considered a (pseudo-)metric space under $d(h_1, h_2) = \|h_1 - h_2\|$. With this structure I obtain the asymptotic equicontinuity in total variation required by Lemma 3.1.

PROPOSITION S3.5: *Suppose that Assumptions 4.4, 4.5 and equation (S26) hold, $u \mapsto \sqrt{\zeta(u,z)}$ is continuously differentiable and $P_{n,\gamma,h}$ is as in Proposition S3.4. Then $h \mapsto P_{n,\gamma,h}$ is asymptotically equicontinuous in total variation on H_γ.*

Proof. Let $h_n \to h$. It suffices to show that $\lim_{n\to\infty} d_{TV}(P_{n,\gamma,h_n}, P_{n,\gamma,h}) = 0$. Let

$$\xi_{n,\gamma,h} := (Y - X'[\theta + \tau/\sqrt{n}] - Z_1'[\beta + b_0/\sqrt{n}], X - \pi(Z) - b_1(Z)/\sqrt{n}, Z),$$

and $q_{n,\gamma,h} := \zeta(\xi_{n,\gamma,h})$, so that we have

$$\frac{p_{n,\gamma,h_n}(W)}{p_{n,\gamma,h}(W)} = \frac{q_{n,\gamma,h_n}(W)}{q_{n,\gamma,h}(W)} \times \frac{1 + b_{2,n}(\xi_{n,\gamma,h_n})/\sqrt{n}}{1 + b_2(\xi_{n,\gamma,h})/\sqrt{n}}.$$

[10] Cf. footnote 8.

By Proposition S3.4, Remark 3.1 and Lemma S2.4 it therefore suffices to show that

$$\sum_{i=1}^{n} \log \frac{q_{n,\gamma,h_n}(W_i)}{q_{n,\gamma,h}(W_i)} = o_{P_{n,\gamma}}(1), \quad \sum_{i=1}^{n} \log \frac{1 + b_{2,n}(\xi_{n,\gamma,h_n,i})/\sqrt{n}}{1 + b_2(\xi_{n,\gamma,h,i})/\sqrt{n}} = o_{P_{n,\gamma}}(1). \quad \text{(S27)}$$

Each $q_{n,\gamma,h}$ is a probability density function; let $Q_{n,\gamma,h}$ be the probability measure corresponding to $q^n_{n,\gamma,h}$. We now show that for some $N \in \mathbb{N}$, $\{h \mapsto Q_{n,\gamma,h} : n \geq N\}$ is equicontinuous in total variation on H_γ by verifying the conditions of Lemma S2.3. The set Γ is a subset of a linear space by Assumption 4.4; let H_γ is a (semi-)normed linear space as discused immediately prior to the statement of Proposition S3.5 with the seminorm given by (S25). Note that $Q_{n,\gamma,h} = P_{n,\gamma,(\tau,b_0,b_1,0)}$ for $h = (\tau, b)$ and let $\varphi(h) := (\tau, b_0, b_1, 0)$. Similarly to as noted in Proposition S3.4, each $\gamma + s\varphi(h) \in \Gamma$ for all $h \in \mathcal{V}$ when the latter is taken as a sufficiently small neighbourhood of zero 0 in H_γ. In conjunction with the fact that $Q_{n,\gamma,h}$ is the n–fold product of $P_{\gamma+(\tau,b_0,b_1,0)/\sqrt{n}}$, where P_γ has density (33), this ensures that condition (i) is satisfied. Condition (ii) holds as

$$\sqrt{q_{\gamma+t(\tau,b_0,b_1,0)}(W)} = \sqrt{\zeta(Y - X'[\theta + t\tau] - Z_1'[\beta + tb_0], X - \pi(Z) - tb_1(Z), Z)},$$

is continuously differentiable in t on $[0,1]$ for each W, all $\gamma \in \Gamma$ and $(\tau, b_0, b_1, 0) \in \mathcal{V}$ by the chain rule. For (iii), letting $u_{\gamma,h,s}(w) := \frac{\partial \log q_{\gamma+t(\tau,b_0,b_1,0)}(w)}{\partial t}_{|t=s}$ we have

$$u_{\gamma_\star,h,s}(W) = -\phi(Y - X'[\theta + \tilde{\tau} + s\tau] - Z_1'[\beta + \tilde{b}_0 + sb_0], X - \pi(Z) - \tilde{b}_1 - sb_1(Z), Z)' \begin{bmatrix} X'\tau + Z_1'b_0 \\ b_1(Z) \end{bmatrix},$$

for any $\gamma_\star = \gamma + (\tilde{\tau}, \tilde{b}_0, \tilde{b}_1, 0)$, $(\tilde{\tau}, \tilde{b}_0, \tilde{b}_1, 0) \in \mathcal{V}$. Under $Q_{\gamma_\star+s\varphi(h)}$,

$$(Y - X'[\theta + \tilde{\tau} + s\tau] - Z_1'[\beta + \tilde{b}_0 + sb_0], X - \pi(Z) - \tilde{b}_1(Z) - sb_1(Z), Z) \sim \zeta.$$

Hence with $(\epsilon, \upsilon, Z) \sim \zeta$, by Assumption 4.4 and equation (S25),

$$Q_{\gamma_\star+s\varphi(h)}[u^2_{\gamma_\star,h,s}] \leq \mathbb{E}\left[\phi_1(\epsilon,\upsilon,Z)^2\right]\left[\|\tau\|^2 + \|b_0\|^2\right] + \sum_{k=1}^{d_\theta} \mathbb{E}\left[\phi_{1+k}(\epsilon,\upsilon,Z)^2 b_{1,k}(Z)^2\right] \lesssim \|h\|^2.$$

By Lemma S2.3 $\{h \mapsto Q_{n,\gamma,h} : n \geq N\}$ is equicontinuous in total variation on H_γ. Hence (e.g. Strasser, 1985, Lemma 2.4)

$$\int \left| \frac{q^n_{n,\gamma,h_n}}{p^n_\gamma} - \frac{q^n_{n,\gamma,h}}{p^n_\gamma} \right| \mathrm{d}P_{n,\gamma} = \int \left| q^n_{n,\gamma,h_n} - q^n_{n,\gamma,h} \right| \mathrm{d}\tilde{\nu}^n \to 0,$$

which implies

$$\prod_{i=1}^{n} \frac{q_{n,\gamma,h_n}(W_i)}{p_\gamma(W_i)} - \prod_{i=1}^{n} \frac{q_{n,\gamma,h}(W_i)}{p_\gamma(W_i)} = o_{P_{n,\gamma}}(1),$$

and so

$$\sum_{i=1}^{n} [\log q_{n,\gamma,h_n}(W_i) - \log q_{n,\gamma,h}(W_i)] = \sum_{i=1}^{n} \log \frac{q_{n,\gamma,h_n}(W_i)}{q_{n,\gamma,h}(W_i)} = o_{P_{n,\gamma}}(1)$$

by the continuous mapping theorem, which establishes the first condition in (S27).

For the second condition in (S27), let $\tilde{h} := (\tau, b_0, b_1, 0)$ and $\tilde{h}_n := (\tau_n, b_{0_n}, b_{1,n}, 0)$. A consequence of the equicontinuity in total variation of $\{h \mapsto Q_{n,\gamma,h} : n \geq N\}$ is $d_{TV}(P_{n,\gamma,\tilde{h}}, P_{n,\gamma,\tilde{h}_n}) \to 0$, hence $P_{n,\gamma,\tilde{h}} \triangleleft \triangleright P_{n,\gamma,\tilde{h}_n}$. Combined with Remark 3.1 and Assumption 4.5 this reveals that it suffices to show that that the sum on the right hand side of (S27) converges to zero in $P_{n,\gamma,\tilde{h}_n}$ – probability. For this, we verify the conditions of Lemma S3.4, which we will apply with the arrays formed by $\xi_{n,\gamma,h_n,i}$ and $\xi_{n,\gamma,h,i}$. Under $P_{n,\gamma,\tilde{h}_n}$, $\xi_{n,\gamma,h_n,i} \sim \zeta$ and so

$$\mathbb{E}\left[b_{2,n}(\xi_{n,\gamma,h_n,i}) - b_2(\xi_{n,\gamma,h_n,i})\right]^2 \leq \|b_{2,n} - b_2\|^2_{L_2(\zeta)} \to 0. \tag{S28}$$

Note that

$$\xi_{n,\gamma,h_n,i} - \xi_{n,\gamma,h,i} = n^{-1/2}(X_i'[\tau - \tau_n] + Z_{1,i}'[b_0 - b_{0,n}], b_1(Z_i) - b_{1,n}(Z_i), 0)$$

and since b_2 is continuously differentiable in its first two arguments, with bounded derivative,

$$|b_2(\xi_{n,\gamma,h_n,i}) - b_2(\xi_{n,\gamma,h_n,i})| \lesssim \frac{|X_i'[\tau - \tau_n]| + |Z_{1,i}'[b_0 - b_{0,n}]| + |b_1(Z_i) - b_{1,n}(Z_i)|}{\sqrt{n}}. \tag{S29}$$

By equation (S26), $\mathbb{E}[b_{1,n}(Z_i) - b_1(Z_i)]^2 \to 0$ under ζ. Since $\|X_i\|^2$ and $\|Z_{1,i}\|^2$ have finite $L_2(\zeta)$ norm – analogously to in (S28) –

$$\mathbb{E}\left[b_2(\xi_{n,\gamma,h_n,i}) - b_2(\xi_{n,\gamma,h,i})\right]^2 \lesssim (\|\tau - \tau_n\|_2^2 + \|b_0 - b_{0,n}\|_2^2 + \|b_{1,n} - b_1\|^2_{L_2(\zeta)}) \to 0.$$

As the data are i.i.d across rows, this verifies (S32). By the definition of $B_{\gamma,2}$, under $P_{n,\gamma,\tilde{h}_n}$, $\mathbb{E}b_{2,n}(\xi_{n,\gamma,h_n,i}) = \mathbb{E}b_2(\xi_{n,\gamma,h_n,i}) = 0$ and by (S29) and the fact that

$\|X_i\|$ and $\|Z_{1,i}\|$ have finite $L_1(\zeta)$ norm,

$$\mathbb{E}\left[|b_2(\xi_{n,\gamma,h_n,i}) - b_2(\xi_{n,\gamma,h_n,i})|\right] \lesssim n^{-1/2}(\|(\tau, b_0) - (\tau_n, b_{0,n})\|_1 + \|b_{1,n} - b_1\|_{L_1(\zeta)})$$
$$= o(n^{-1/2}).$$

As data are i.i.d. across rows, this verifies (S33). Apply Lemma S3.4. $\square$

S3.2.4 Supporting lemmas

LEMMA S3.1: *In the setting of Proposition 4.4:*

(i) $\|\hat{M}_n - M\| = O_{P_{n,\gamma}}(n^{-1/2})$;

(ii) $\frac{1}{n}\sum_{i=1}^n \|\breve{U}_{n,i} - U_i\|^2 = O_{P_{n,\gamma}}(\delta_n^2 + n^{-1})$;

(iii) $\|\breve{J}_n - \bar{J}\| = O_{P_{n,\gamma}}(\delta_n^2 + n^{-1/2})$;

(iv) $\|\breve{J}_n^{-1} - \bar{J}^{-1}\| = O_{P_{n,\gamma}}(\delta_n^2 + n^{-1/2})$;

(v) $\mathsf{R}_{n,1} = \frac{1}{\sqrt{n}}\sum_{i=1}^n q(\breve{J}_n)\breve{U}_{n,i}\left[M - \hat{M}_n\right] Z_{1,i} = o_{P_{n,\gamma}}(1)$;

(vi) $\mathsf{R}_{n,2} = \frac{1}{\sqrt{n}}\sum_{i=1}^n q(\breve{J}_n)\breve{U}_{n,i}\left[\hat{\pi}_{n,i}(Z_i) - \pi(Z_i)\right] = o_{P_{n,\gamma}}(1)$;

(vii) $\mathsf{R}_{n,3} = \frac{1}{\sqrt{n}}\sum_{i=1}^n q(\breve{J}_n)(\breve{U}_{n,i} - U_i)f(Z_i) = o_{P_{n,\gamma}}(1)$;

(viii) $\mathsf{R}_{n,4} = \frac{1}{\sqrt{n}}\sum_{i=1}^n (q(\breve{J}_n) - q(\bar{J})U_i f(Z_i) = o_{P_{n,\gamma}}(1)$;

(ix) $\mathsf{S}_{n,1} = \frac{1}{n}\sum_{i=1}^n \|q(\breve{J}_n)\breve{U}_{n,i}\left[M - \hat{M}_n\right] Z_{1,i}\|^2 = O_{P_{n,\gamma}}(n^{-1})$;

(x) $\mathsf{S}_{n,2} = \frac{1}{n}\sum_{i=1}^n \|q(\breve{J}_n)\breve{U}_{n,i}\left[\hat{\pi}_{n,i}(Z_i) - \pi(Z_i)\right]\|^2 = O_{P_{n,\gamma}}(\delta_n^2)$;

(xi) $\mathsf{S}_{n,3} = \frac{1}{n}\sum_{i=1}^n \|q(\breve{J}_n)(\breve{U}_{n,i} - U_i)f(Z_i)\|^2 = O_{P_{n,\gamma}}(n^{-1})$;

(xii) $\mathsf{S}_{n,4} = \frac{1}{n}\sum_{i=1}^n \|(q(\breve{J}_n) - q(\bar{J})U_i f(Z_i)\|^2 = O_{P_{n,\gamma}}(\delta_n^4 + n^{-1})$;

with $q(J) = e_1' J^{-1} - J_{1,2}^{-1}[J_{2,2}^{-1}]^{-1}[0_K', I_K]J^{-1}$, $\hat{M}_n := \left[\frac{1}{n}\sum_{i=1}^n X_i Z_{1,i}'\right]\left[\frac{1}{n}\sum_{i=1}^n Z_{1,i}Z_{1,i}'\right]^{-1}$, $M := \mathbb{E}[XZ_1']\,\mathbb{E}[Z_1Z_1']^{-1}$, $f(Z_i) := \pi(Z_i) - MZ_{1,i}$ *and* $\hat{f}_{n,i}(Z_i) := \hat{\pi}_{n,i}(Z_i) - \hat{M}_n Z_{1,i}$.

Proof. First observe that

$$q(J) = \left[\, J_{1,1}^{-1} \; J_{1,2}^{-1} \,\right] - J_{1,2}^{-1}[J_{2,2}^{-1}]^{-1}\left[\, J_{2,1}^{-1} \; J_{2,2}^{-1} \,\right] = \left[\, J_{1,1}^{-1} - J_{1,2}^{-1}[J_{2,2}^{-1}]^{-1}J_{2,1}^{-1} \; 0' \,\right] = q_1(J)e_1' = \left[\, (J_{1,1})^{-1} \; 0' \,\right],$$

with $q_1(J) := J_{1,1}^{-1} - J_{1,2}^{-1}[J_{2,2}^{-1}]^{-1}J_{2,1}^{-1} = (J_{1,1})^{-1}$ (cf. Proposition 2.8.7 in Bernstein (2009)). Abbreviate $\tilde{\pi}_{n,i}(Z_i) := \hat{\pi}_{n,i}(Z_i) - \pi(Z_i)$. By Propositions S3.4 & S3.5, Remark 3.1, Lemma 2.15 and Remark 18.3 (both) in Strasser (1985), $P_{n,\gamma} \triangleleft \triangleright P_{n,\gamma,(0,b_{0,n},0,0)}$, which we will henceforth abbreviate to $P_{n,\gamma,b_{0,n}}$.

(i) By the moment conditions in Assumption 4.4, $\mathbb{E}\|XZ_1'\|^2 < \infty$ and $\mathbb{E}\|Z_1Z_1'\|^2 < \infty$ (under $P_{n,\gamma}$). Since the samples are i.i.d., then by the central limit theorem

$$\left\| \left[\frac{1}{n}\sum_{i=1}^n X_i Z_{1,i}'\right] \left[\frac{1}{n}\sum_{i=1}^n Z_{1,i} Z_{1,i}'\right]^{-1} - M \right\| = O_{P_{n,\gamma}}(n^{-1/2}).$$

(ii) $\|\breve{U}_{n,i} - U_i\|^2 \lesssim (\breve{U}_{n,i,1} - U_{i,1})^2 + \|\breve{U}_{n,i,2} - U_{i,2}\|^2$, so we will handle the sample means of the two right hand side terms separately. The first is

$$\frac{1}{n}\sum_{i=1}^n (\breve{U}_{n,i,1} - U_{i,1})^2 = (\beta_n - \beta)\frac{1}{n}\sum_{i=1}^n Z_{1,i}Z_{1,i}'(\beta_n - \beta) = O_{P_{n,\gamma}}(n^{-1}).$$

The second is

$$\frac{1}{n}\sum_{i=1}^n \|\breve{U}_{n,i,2} - U_{i,2}\|^2 = \frac{1}{n}\sum_{i=1}^n \|\hat{\pi}_{n,i}(Z_i) - \pi(Z_i)\|^2 = O_{P_{n,\gamma}}(n^{-1})$$

Let F_n be the sets on which (43) holds and fix any $\varepsilon > 0$. Then for all large enough $n \in \mathbb{N}$ and K large enough, by Markov's inequality

$$\begin{aligned} P_{n,\gamma}\left(\frac{1}{n}\sum_{i=1}^n \|\tilde{\pi}_{n,i}(Z_i)\|^2 > K\delta_n^2\right) &\leq P_{n,\gamma}\left(\mathbf{1}_{F_n}\frac{1}{n}\sum_{i=1}^n \|\tilde{\pi}_{n,i}(Z_i)\|^2 > K\delta_n^2\right) + P_{n,\gamma}F_n^{\complement} \\ &\leq K^{-1}\delta_n^{-2}\left[\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\mathbf{1}_{F_{n,i}}\mathbb{E}\left[\|\tilde{\pi}_{n,i}(Z_i)\|^2 | \mathcal{C}_{n,j}\right]\right]\right] + \varepsilon/2 \\ &< \varepsilon. \end{aligned}$$

where the expectation is under $P_{n,\gamma}$ and $F_{n,i} \supset F_n$ is the $\mathcal{C}_{n,j}$ – measurable set on which the bound (43) holds for index i.

(iii) One has

$$\|\breve{J}_n - \bar{J}\| \leq \frac{1}{n}\sum_{i=1}^n \|\breve{U}_{n,i}\| \left\|\breve{U}_{n,i} - U_i\right\| + \frac{1}{n}\sum_{i=1}^n \left\|\breve{U}_{n,i} - U_i\right\| \|U_i\| + \left\|\frac{1}{n}\sum_{i=1}^n U_iU_i' - \mathbb{E}[U_iU_i']\right\|.$$

The last right hand side term is $O_{P_{n,\gamma}}(n^{-1/2})$ by the CLT given the moment conditions and sampling assumption in Assumption 4.4. Using these same assumptions with Markov's inequality, the Cauchy – Schwarz inequality and (ii) yields that the other two right hand side terms are $O_{P_{n,\gamma}}(\delta_n^2 + n^{-1})$.

(iv) This follows from (iii) and the identity $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$, noting that (iii) also implies that $\|\breve{J}_n^{-1}\| = O_{P_{n,\gamma}}(1)$.

(v) Note that q is continuous and $\|q(\bar{J})\|$ is finite by equation (37). We have

$$\mathsf{R}'_{n,1} = q(\breve{J}_n)\left[\frac{1}{n}\sum_{i=1}^{n} Z'_{1,i}U_i + \frac{1}{n}\sum_{i=1}^{n} Z'_{1,i}(\breve{U}_{n,i} - U_i)\right]\sqrt{n}[M - \hat{M}_n]'.$$

$q(\breve{J}_n)$ is $O_{P_{n,\gamma}}(1)$ by continuity and (iii); $\sqrt{n}\|M - \hat{M}_n\| = O_{P_{n,\gamma}}(1)$ by (i); $\frac{1}{n}\sum_{i=1}^{n} Z'_{1,i}U_i$ converges to zero in probability by the weak law of large numbers; $\frac{1}{n}\sum_{i=1}^{n} Z'_{1,i}(\breve{U}_{n,i} - U_i)$ also converges to zero in probability, by the moment conditions in Assumption 4.4 along with (ii) and the Cauchy – Schwarz inequality.

(vi) We have

$$\begin{aligned}
\mathsf{R}_{n,2} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} q(\breve{J}_n)\breve{U}_{n,i}\left[\hat{\pi}_{n,i}(Z_i) - \pi(Z_i)\right] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} q_1(\breve{J}_n)\breve{\epsilon}_{n,i}\left[\hat{\pi}_{n,i}(Z_i) - \pi(Z_i)\right] \\
&= q_1(\hat{J}_n)\left[\frac{1}{n}\sum_{i=1}^{n}\tilde{\pi}_{n,i}(Z_i)Z'_{1,i}\right]\sqrt{n}(\beta - \beta_n) + q_1(\hat{J}_n)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\pi}_{n,i}(Z_i)\epsilon_i.
\end{aligned} \tag{S30}$$

By Cauchy – Schwarz, Assumption 4.4 and the argument in (ii),

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\tilde{\pi}_{n,i}(Z_i)Z'_{1,i}\right\| \le \left[\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\pi}_{n,i}(Z_i)\|^2\right]^{1/2}\left[\frac{1}{n}\sum_{i=1}^{n}\|Z_{1,i}\|^2\right]^{1/2} = o_{P_{n,\gamma}}(1).$$

Since by (iii) $q_1(\breve{J}_n) = O_{P_{n,\gamma}}(1)$ and the same holds for $\sqrt{n}(\beta - \beta_n)$, the first term on the last line in (S30) is $o_{P_{n,\gamma}}(1)$. By (iii) $q_1(\breve{J}_n) = O_{P_{n,\gamma}}(1)$. We may split the (scaled) sum into

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{m_n}\tilde{\pi}_{n,i}(Z_i)\epsilon_i + \frac{1}{\sqrt{n}}\sum_{i=m_n+1}^{n}\tilde{\pi}_{n,i}(Z_i)\epsilon_i,$$

where $m_n = \lfloor n/2 \rfloor$. We show the first term is $o_{P_{n,\gamma}}(1)$; the argument for the second term is analogous. By Assumptions 4.4 and 4.6, with $F_{n,i} \in \sigma(\mathcal{C}_{n,j})$ as in (ii), $F_n^1 := \cap_{i=1}^{m_n} F_{n,i} \in \sigma(\mathcal{C}_{n,2})$ and expectations under $P_{n,\gamma}$, for $i \neq k$ with $i, k \le m_n$,

$$\mathbb{E}\left[\mathbf{1}_{F_n^1}\tilde{\pi}_{n,i}(Z_i)\epsilon_i\epsilon_k\tilde{\pi}_{n,k}(Z_k)\epsilon_i \;\middle|\; Z_i, \mathcal{C}_{n,2}\right] = \mathbb{E}\left[\tilde{\pi}_{n,i}(Z_i)\mathbb{E}[\epsilon_i|Z_i]\mathbf{1}_{F_n^1}\epsilon_k\tilde{\pi}_{n,k}(Z_k)\right] = 0$$

and for $i \le m_n$, by equation (37), for some constant $C \in (0, \infty)$,

$$\begin{aligned}\mathbb{E}\left[\epsilon_i^2 \|\tilde{\pi}_{n,i}(Z_i)\|^2 \mathbf{1}_{F_n^1}\right] &= \mathbb{E}\left[\mathbb{E}\left[\epsilon_i^2 | Z_i, \mathcal{C}_{n,2}\right] \mathbf{1}_{F_n^1} \|\tilde{\pi}_{n,i}(Z_i)\|^2\right] \\ &\le C\mathbb{E}\left[\mathbf{1}_{F_n^1}\mathbb{E}\left[\|\tilde{\pi}_{n,i}(Z_i)\|^2 | \mathcal{C}_{n,2}\right]\right] \\ &\le C\delta_n^2.\end{aligned}$$

Hence, by Markov's inequality,

$$\begin{aligned}P_{n,\gamma}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{m_n}\tilde{\pi}_{n,i}(Z_i)\epsilon_i\right| > \varepsilon\right) &\le P_{n,\gamma}\left(\mathbf{1}_{F_n^1}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{m_n}\tilde{\pi}_{n,i}(Z_i)\epsilon_i\right| > \varepsilon\right) + P_{n,\gamma}F_n^{\complement} \\ &\le \frac{1}{n}\sum_{i=1}^{m_n}\mathbb{E}\left[\mathbf{1}_{F_n^1}\mathbb{E}\left[\|\tilde{\pi}_{n,i}(Z_i)\|^2 | \mathcal{C}_{n,2}\right]\right] + o(1) \\ &\le m_n C\delta_n / n + o(1) \\ &= o(1).\end{aligned}$$

(vii) We have

$$\begin{aligned}\mathsf{R}_{n,3} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} q(\breve{J}_n)(\breve{U}_{n,i} - U_i)f(Z_i) \\ &= q_1(\breve{J}_n)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\breve{\epsilon}_{n,i} - \epsilon_i)f(Z_i) \\ &= q_1(\breve{J}_n)\left[\frac{1}{n}\sum_{i=1}^{n} f(Z_i)Z_{1,i}'\right]\sqrt{n}(\beta - \beta_n) \\ &= o_{P_{n,\gamma}}(1),\end{aligned}$$

by (iii), Assumption 4.4 and the CLT, as

$$\mathbb{E}\left[f(Z)Z_1\right] = \mathbb{E}\left[\pi(Z)Z_1' - \mathbb{E}[\pi(Z)Z_1']\mathbb{E}[Z_1Z_1']^{-1}Z_1Z_1'\right] = 0.$$

(viii) Since $\mathbb{E}[Uf(Z)] = \mathbb{E}[\mathbb{E}[U|Z]f(Z)] = 0$ and $q(\breve{J}_n) \xrightarrow{P_{n,\gamma}} q(\bar{J})$ by (iii), the moment conditions in Assumption 4.4 allow the application of the CLT to yield

$$\mathsf{R}_{n,4} = (q(\breve{J}_n) - q(\bar{J}))\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i f(Z_i) = o_{P_{n,\gamma}}(1).$$

(ix) Note that

$$\breve{\epsilon}_{n,i}^2 - \epsilon_i^2 = (\breve{\epsilon}_{n,i} - \epsilon_i)(\breve{\epsilon}_{n,i} + \epsilon_i) = 2(\beta - \beta_n)'Z_{1,i}\epsilon_i + (\beta - \beta_n)'Z_{1,i}Z_{1,i}'(\beta - \beta_n).$$

Therefore

$$\frac{1}{n}\sum_{i=1}^{n}\breve{\epsilon}_{n,i}^{2}\|Z_{1,i}\|^{2} = \frac{1}{n}\sum_{i=1}^{n}\epsilon_{i}^{2}\|Z_{1,i}\|^{2} + 2\|\beta-\beta_{n}\|\frac{1}{n}\sum_{i=1}^{n}\|Z_{1,i}\|^{3}|\epsilon_{i}| + \|\beta-\beta_{n}\|^{2}\frac{1}{n}\sum_{i=1}^{n}\|Z_{1,i}\|^{4}$$
$$= O_{P_{n,\gamma}}(1),$$

by the moment conditions in Assumption 4.4 combined with Hölder's inequality. Therefore by (i) and (iii) along with the continuity of q,

$$\frac{1}{n}\sum_{i=1}^{n}\|q(\breve{J}_{n})\breve{U}_{n,i}\left[M-\hat{M}_{n}\right]Z_{1,i}\|^{2} \le q_{1}(\breve{J}_{n})\left\|M-\hat{M}_{n}\right\|^{2}\frac{1}{n}\sum_{i=1}^{n}\breve{\epsilon}_{n,i}^{2}\|Z_{1,i}\|^{2} = O_{P_{n,\gamma}}(n^{-1}).$$

(x) We have

$$\mathsf{S}_{n,1} = \frac{1}{n}\sum_{i=1}^{n}\|q(\breve{J}_{n})\breve{U}_{n,i}\left[\hat{\pi}_{n,i}(Z_{i})-\pi(Z_{i})\right]\|^{2} = q_{1}(\breve{J}_{n})^{2}\frac{1}{n}\sum_{i=1}^{n}\breve{\epsilon}_{n,i}^{2}\|\hat{\pi}_{n,i}(Z_{i})-\pi(Z_{i})\|^{2}.$$

By (iii) along with the continuity of q, and the fact that $\hat{\beta}_{n}$ is $\sqrt{n}$ – consistent, it will suffice to show that $\frac{1}{n}\sum_{i=1}^{n}\breve{\epsilon}_{n,i}^{2}\|\hat{\pi}_{n,i}(Z_{i})-\pi(Z_{i})\|^{2} = O_{P_{n,\gamma}}(\delta_{n}^{2})$. For this it suffices to show that

$$\frac{1}{n}\sum_{i=1}^{m_{n}}\|\hat{\pi}_{n,i}(Z_{i})-\pi(Z_{i})\|^{2}\breve{\epsilon}_{n,i}^{2} = O_{P_{n,\gamma}}(\delta_{n}^{2}), \quad \frac{1}{n}\sum_{i=m_{n}+1}^{n}\|\hat{\pi}_{n,i}(Z_{i})-\pi(Z_{i})\|^{2}\breve{\epsilon}_{n,i}^{2} = O_{P_{n,\gamma}}(\delta_{n}^{2}).$$

We will show the first; the second follows analogously. By contiguity it suffices to show the conclusion under $P_{n,\gamma,b_{0,n}}$. In particular, by Markov's inequality, for all large enough $n \in \mathbb{N}$ and large enough K

$$\begin{aligned} P_{n,\gamma,b_{0,n}}&\left(\frac{1}{n}\sum_{i=1}^{m_{n}}\breve{\epsilon}_{n,i}^{2}\|\tilde{\pi}_{n,i}(Z_{i})\|^{2} > K\delta_{n}^{2}\right) \\ &\le P_{n,\gamma,b_{0,n}}\left(\mathbf{1}_{F_{n}^{1}}\frac{1}{n}\sum_{i=1}^{m_{n}}\breve{\epsilon}_{n,i}^{2}\|\tilde{\pi}_{n,i}(Z_{i})\|^{2} > K\delta_{n}^{2}\right) + P_{n,\gamma,b_{0,n}}F_{n}^{\complement} \\ &\le K^{-1}\delta_{n}^{-2}\left[\frac{1}{n}\sum_{i=1}^{m_{n}}\mathbb{E}\left[\mathbf{1}_{F_{n}^{1}}\mathbb{E}\left[\|\tilde{\pi}_{n,i}(Z_{i})\|^{2}\breve{\epsilon}_{n,i}^{2}|\mathcal{C}_{n,2}\right]\right]\right] + \varepsilon/2, \end{aligned}$$

as $P_{n,\gamma,b_{0,n}}F_{n}^{\complement} \to 0$ by contiguity. Note that the distribution of $(\breve{\epsilon}_{n,i}, Z_{i})|\mathcal{C}_{n,2}$ under $P_{n,\gamma,b_{0,n}}$ is that of $(\epsilon_{i}, Z_{i})|\mathcal{C}_{n,2}$ under $P_{n,\gamma}$ given the independence of $\mathcal{C}_{n,2}$ from each W_{i} with $1 \le i \le m_{n}$ ensured (under either measure) by the product structure. Therefore, under $P_{n,\gamma,b_{0,n}}$, $\mathbb{E}[\breve{\epsilon}_{n,i}^{2}|Z_{i},\mathcal{C}_{n,2}] \le C$ by equation

(37) and hence

$$\mathbb{E}\left[\mathbf{1}_{F_n^1}\mathbb{E}\left[\|\tilde{\pi}_{n,i}(Z_i)\|^2\breve{\epsilon}_{n,i}^2|\mathcal{C}_{n,2}\right]\right] \lesssim \mathbb{E}\left[\mathbf{1}_{F_n^1}\mathbb{E}\left[\|\tilde{\pi}_{n,i}(Z_i)\|^2|\mathcal{C}_{n,2}\right]\right] \leq \delta_n^2,$$

where the last inequality follows by the definition of F_n^1. Combing the two preceding displays yields

$$\frac{1}{n}\sum_{i=1}^{m_n}\|\hat{\pi}_{n,i}(Z_i)-\pi(Z_i)\|^2\breve{\epsilon}_{n,i}^2 = O_{P_{n,\gamma,b_{0,n}}}(\delta_n^2) = O_{P_{n,\gamma}}(\delta_n^2),$$

with the last equality following by contiguity.

(xi) Since $\|f(Z_i)\|^2 \leq 2\left[\|\pi(Z_i)\|^2 + \|M\|^2\|Z_{1,i}\|^2\right]$, by the moment conditions in Assumption 4.4 and Cauchy – Schwarz,

$$\frac{1}{n}\sum_{i=1}^{n}\|Z_{1,i}\|^2\|f(Z_i)\|^2 = O_{P_{n,\gamma}}(1).$$

Therefore by the fact that q is continuous and (iii)

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\|q(\breve{J}_n)(\breve{U}_{n,i}-U_i)f(Z_i)\|^2 &\leq q(\breve{J}_n)^2\frac{1}{n}\sum_{i=1}^{n}\|\breve{\epsilon}_{n,i}-\epsilon_i\|^2\|f(Z_i)\|^2\\ &= q(\breve{J}_n)^2\|\beta-\beta_n\|^2\frac{1}{n}\sum_{i=1}^{n}\|Z_{1,i}\|^2\|f(Z_i)\|^2\\ &= O_{P_{n,\gamma}}(n^{-1}).\end{aligned}$$

(xii) Noting the bound on $\|f(Z_i)\|^2$ in the previous item by the moment conditions in Assumption 4.4 and Cauchy – Schwarz,

$$\frac{1}{n}\sum_{i=1}^{n}\|\epsilon_i\|^2\|f(Z_i)\|^2 = O_{P_{n,\gamma}}(1).$$

Since $q_1(J)$ is locally Lipschitz at any non-singular J, combining the above with (iii) yields

$$\frac{1}{n}\sum_{i=1}^{n}\|(q(\breve{J}_n)-q(\bar{J})U_if(Z_i)\|^2 \leq (q_1(\breve{J}_n)-q_1(\bar{J}))^2\frac{1}{n}\sum_{i=1}^{n}\|U_i\|^2\|f(Z_i)\|^2 = O_{P_{n,\gamma}}(\delta_n^4+n^{-1}).$$

□

LEMMA S3.2: *Suppose that Assumption 4.4 holds and that for each $i = 1, \ldots, 1+$*

$d_\alpha =: K$

$$\lim_{|u_i|\to\infty} |u_i|\zeta(u,z) = 0, \qquad \textit{(pointwise) } \nu\text{–a.e..}$$

Then, the second two conditions in (37) *hold.*

Proof. First note that all the integrals exist by the moment conditions in Assumption 4.4. To evaluate them we integrate by parts. To simplify the notation we use a different notation from the main text: $\tilde{\zeta}_i$ and $\tilde{\phi}_i$ will denote the derivative and logarithmic derivative of ζ with respect to u_i. That is, $\tilde{\phi}_1 = \phi_1$ and $\tilde{\phi}_i = \phi_{2,i-1}$ for $i \geq 2$ and similarly for ζ. Then for the first condition

$$\int u_i\tilde{\phi}_i(u,z)\zeta(u,z)\,\mathrm{d}\nu(u,z) = \left[\lim_{u_i\to\infty} u_i\zeta(u,z) - \lim_{u_i\to-\infty} u_i\zeta(u,z)\right] - \int \zeta(u,z)\,\mathrm{d}\nu(u,z) = -1.$$

For $j \neq i$,

$$\int u_i\tilde{\phi}_j(u,z)\zeta(u,z)\,\mathrm{d}\nu(u,z) = \int\int u_i \int \tilde{\zeta}_j(u,z)\,\mathrm{d}u_j\,\mathrm{d}(u_1,\ldots,u_{j-1},u_{j+1},\ldots,u_K)\,\mathrm{d}\nu_Z(z) = 0,$$

where the equality follows because we may integrate by parts to find

$$\int \tilde{\zeta}_j(u,z)\,\mathrm{d}u_j = \left[\lim_{u_i\to\infty} \zeta(u,z) - \lim_{u_i\to-\infty} \zeta(u,z)\right] - \int 0\times\zeta(u,z)\,\mathrm{d}u_j = 0, \quad \text{(S31)}$$

where the limits are zero as ζ is a density function with respect to $\lambda^K \times \nu_Z$.

For the second condition let $j \in \{2,\ldots,K\}$. Firstly we have

$$\begin{aligned}
\int u_1u_j\phi_1(u,z)\zeta(u,z)\,\mathrm{d}\nu(u,z) &= \int\int u_j \int u_1\tilde{\zeta}_1(u,z)\,\mathrm{d}u_1\,\mathrm{d}(u_2,\ldots,u_K)\,\mathrm{d}\nu_Z(z) \\
&= -\int\int u_j \int \zeta(u,z)\,\mathrm{d}u_1\,\mathrm{d}(u_2,\ldots,u_K)\,\mathrm{d}\nu_Z(z) \\
&= -\int u_j\zeta(u,z)\,\mathrm{d}\nu(u,z) \\
&= 0
\end{aligned}$$

since $\mathbb{E}[v] = 0$ and

$$\int u_1\tilde{\zeta}_1(u,z)\,\mathrm{d}u_1 = \left[\lim_{u_1\to\infty} u_1\zeta(u,z) - \lim_{u_1\to-\infty} u_1\zeta(u,z)\right] - \int \zeta(u,z)\,\mathrm{d}u_1 = -\int \zeta(u,z)\,\mathrm{d}u_1.$$

Secondly suppose that also $i \in \{2, \ldots, K\}$ and note that using (S31):

$$\int u_i u_j \tilde{\phi}_1(u,z)\zeta(u,z)\,\mathrm{d}\nu(u,z) = \int\int u_i u_j \int \tilde{\zeta}_1(u,z)\,\mathrm{d}u_1\,\mathrm{d}(u_2,\ldots,u_K)\,\mathrm{d}\nu_Z(z) = 0.$$

□

S3.3 Some technical tools

Lemma S3.3: *Let $(m_n)_{n\in\mathbb{N}}$ be an increasing sequence of natural numbers such that $m_n \le n$, $(Y_{n,i})_{n\in\mathbb{N},1\le i\le m_n}$ a triangular array of random vectors and $\mathcal{C}_n$ a collection of random variables. Suppose that with probability approaching one either*

(i) $\mathbb{E}\left[\|Y_{n,i}\||\mathcal{C}_n\right] \le \delta_n n^{-1/2}$ *for some* $\delta_n \to 0$ *and all* $i \le m_n$*; or*

(ii) For each component $Y_{n,i,s}$ *of* $Y_{n,i}$ *and any* $i \ne j \le m_n$, $\mathbb{E}[Y_{n,i,s}Y_{n,j,s}|\mathcal{C}_n] = 0$ *almost surely and* $\mathbb{E}[Y_{n,i,s}^2|\mathcal{C}_n] \le \delta_n$ *for some* $\delta_n \to 0$ *and all* $i \le m_n$.

Then $\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n} Y_{n,i}$ *converges to zero in probability.*

Proof. If condition (i) holds, $\mathbb{E}\left\|m_n^{-1/2}\sum_{i=1}^{m_n} Y_{n,i}\right\| \le \delta_n m_n^{1/2} n^{-1/2} \to 0$. If condition (ii) holds, $\mathbb{E}\left(m_n^{-1/2}\sum_{i=1}^{m_n} Y_{n,i,s}\right)^2 = m_n^{-1}\sum_{i=1}^{m_n}\mathbb{E}Y_{n,i,s}^2 \le \delta_n \to 0$ for each component $Y_{n,i,s}$ of $Y_{n,i}$. In either case the claim then follows by Markov's inequality. □

Lemma S3.4: *Suppose that* $(Y_{n,i})_{n\in\mathbb{N},1\le i\le n}$ *and* $(Z_{n,i})_{n\in\mathbb{N},1\le i\le n}$ *are triangular arrays of random vectors, independent along rows and* $(b_n)_{n\in\mathbb{N}}$ *a sequence of functions such that each* $\mathbb{E}[b_n(Y_{n,i})^2]$ *exists and* b *a bounded function such that as* $n \to \infty$,

$$\max_{1\le i\le n}\mathbb{E}\left[(b_n(Y_{n,i}) - b(Y_{n,i}))^2\right] = o(1), \qquad \max_{1\le i\le n}\mathbb{E}\left[(b(Y_{n,i}) - b(Z_{n,i}))^2\right] = o(1) \tag{S32}$$

and

$$\max_{1\le i\le n}\mathbb{E}\left[b_n(Y_{n,i}) - b(Z_{n,i})\right] = o(n^{-1/2}). \tag{S33}$$

Then, with P_n *the law of* $(Y_{n,1}, Z_{n,1}, \ldots, Y_{n,n}, Z_{n,n})$,

$$l_n := \sum_{i=1}^{n}\log\frac{1 + b_n(Y_{n,i})/\sqrt{n}}{1 + b(Z_{n,i})/\sqrt{n}} = o_{P_n}(1).$$

Proof. (S32) implies that $(b_n(Y_{n,i}))_{n\in\mathbb{N},1\le i\le n}$ is uniformly square P_n – integrable.

The Lindeberg condition therefore holds for $b_n(Y_{n,i})/\sqrt{n}$:

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

for any $\delta > 0$, which implies that (e.g. Gut, 2005, Remark 7.2.4):

$$\max_{1\leq i\leq n}\frac{|b_n(Y_{n,i})|}{\sqrt{n}}\xrightarrow{P_n} 0. \tag{S34}$$

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

$$E_n := \left\{\max_{1\leq i\leq n}|b_n(Y_{n,i})|/\sqrt{n}\leq\varepsilon\right\},\quad F_n := \left\{\max_{1\leq i\leq n}|b(Z_{n,i})|/\sqrt{n}\leq\varepsilon\right\}.$$

Since b is bounded, $P_nF_n \to 1$; $P_nE_n \to 1$ follows from (S34). Hence $P_nF_n \cap E_n \to 1$. On $E_n \cap F_n$ we can perform a two-term Taylor expansion of $\log(1+x)$ to obtain

$$\begin{aligned}\log(1+&b_n(Y_{n,i})/\sqrt{n}) - \log(1+b(Z_{n,i})/\sqrt{n})\\ &= \frac{b_n(Y_{n,i})}{\sqrt{n}} - \frac{1}{2}\frac{b_n(Y_{n,i})^2}{n} - \frac{b(Z_{n,i})}{\sqrt{n}} + \frac{1}{2}\frac{b(Z_{n,i})^2}{n}\\ &\quad + R\left(\frac{b_n(Y_{n,i})}{\sqrt{n}}\right) - R\left(\frac{b(Z_{n,i})}{\sqrt{n}}\right),\end{aligned}$$

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

$$\begin{aligned}l_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}b_n(Y_{n,i}) - b(Z_{n,i}) - \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}[b_n(Y_{n,i})^2 - b(Z_{n,i})^2]\\ + \sum_{i=1}^{n}R\left(\frac{b_n(Y_{n,i})}{\sqrt{n}}\right) - R\left(\frac{b(Z_{n,i})}{\sqrt{n}}\right).\end{aligned}$$

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

$$\begin{aligned}\sum_{i=1}^{n}\left|R\left(\frac{b_n(Y_{n,i})}{\sqrt{n}}\right)\right| &\leq \sum_{i=1}^{n}\left|\frac{b_n(Y_{n,i})}{\sqrt{n}}\right|\left|\frac{b_n(Y_{n,i})^2}{n}\right|\\ &\leq \max_{1\leq i\leq n}\frac{|b_n(Y_{n,i})|}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n}b_n(Y_{n,i})^2,\end{aligned}$$

by Markov's inequality, the fact that $b_n(Y_{n,i})$ is uniformly L_2 – bounded and (S34), the right hand side term converges to zero in P_n – probability. With $\bar{b}$ an upper bound for b,

$$\sum_{i=1}^{n} \left| R\left(\frac{b(Z_{n,i})}{\sqrt{n}}\right) \right| \leq n^{-1/2} \frac{1}{n} \sum_{i=1}^{n} \bar{b} \to 0,$$

hence the left hand side in the display above is $o_{P_n}(1)$. Thus,

$$l_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b_n(Y_{n,i}) - b(Z_{n,i}) - \frac{1}{2}\frac{1}{n} \sum_{i=1}^{n} [b_n(Y_{n,i})^2 - b(Z_{n,i})^2] + o_{P_n}(1),$$

and it remains to show that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} b_n(Y_{n,i}) - b(Z_{n,i})$ and $\frac{1}{n} \sum_{i=1}^{n} [b_n(Y_{n,i})^2 - b(Z_{n,i})^2]$ also converge to zero in probability. For the second of these we have

$$\begin{aligned}
&P_n \left(\left| \frac{1}{n} \sum_{i=1}^{n} [b_n(Y_{n,i})^2 - b(Z_{n,i})^2] \right| > \varepsilon \right) \\
&\leq \varepsilon^{-1} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[b_n(Y_{n,i})^2 - b(Z_{n,i})^2 \right] \\
&\leq \varepsilon^{-1} \max_{1 \leq i \leq n} \mathbb{E}\left[b_n(Y_{n,i})^2 - b(Z_{n,i})^2 \right] \\
&= \varepsilon^{-1} \max_{1 \leq i \leq n} \mathbb{E}\left[(b_n(Y_{n,i}) - b(Z_{n,i}))\,(b_n(Y_{n,i}) + b(Z_{n,i})) \right] \\
&\to 0,
\end{aligned}$$

by Markov's inequality, the Cauchy–Schwarz inequality, (S32), the fact that b is bounded and $b_n(Y_{n,i})$ is (uniformly) L_2 bounded. For the remaining term, we start by noting that by (S33) it suffices to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{b}_n(Y_{n,i}) - \tilde{b}(Z_{n,i}) \xrightarrow{P_n} 0,$$

for $\tilde{b}_n(Y_{n,i}) := b_n(Y_{n,i}) - \mathbb{E}[b_n(Y_{n,i})]$ and $\tilde{b}(Z_{n,i}) := b(Z_{n,i}) - \mathbb{E}[b(Z_{n,i})]$. By (S32), (S33), the row-wise independence and Markov's inequality:

$$P_n \left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{b}_n(Y_{n,i}) - \tilde{b}(Z_{n,i}) \right| > \varepsilon \right) \leq \frac{1}{\varepsilon^2} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\left(\tilde{b}_n(Y_{n,i}) - \tilde{b}(Z_{n,i}) \right)^2 \right] \to 0.$$

This completes the proof that $l_n \xrightarrow{P_n} 0$, as required. □

S4 Additional simulation details & results

S4.1 Single index model

S4.1.1 Design 1

The f_j functions used in simulation Design 1 are plotted in Figures S1 – S4. Tables S1 – S8 display the empirical rejection frequencies for 4 alternative specifications for the distribution of ϵ: (1) $\epsilon|\xi \sim \sqrt{5}(-1)^{\xi}\,\mathrm{Beta}(2,3)$ with $\xi \sim \mathrm{Bernoulli}(1/2)$; (2) $\epsilon = \xi/\sqrt{3/2}$ with $\xi \sim t(6)$; (3) $\epsilon \sim \mathcal{N}(0, \log(2 + (X_1 + X_2\theta)^2))$; (4) $\epsilon \sim \mathcal{N}(0, 1 + \sin(X_1)^2)$.[11] These distributions are referred to as η_i for $i = 1, 2, 3, 4$ in what follows, with $\mathcal{N}(0,1)$ being denoted η_0. As in the $\mathcal{N}(0,1)$ case considered in the main text, the $\hat{S}$ test is close to the nominal 5% level in all simluation designs considered. In contrast, the Wald test is often oversized, with the degree of over-rejection varying across the simulation designs.

Tables S9 and S10 consider the sensitivity of the empirical rejection frequency of the $\hat{S}$ test to $\boldsymbol{\nu}$. In both cases, the test displays under-rejection for sufficiently large $\boldsymbol{\nu}$, but is otherwise relatively insensitive to this parameter.

S4.1.2 Design 2

The f_j functions used in simulation Design 2 are defined as follows. Let $b(x) \coloneqq \mathbf{1}\{x > 0\}\exp(-1/x)$, a bump function and form the smooth transition function $a(x) \coloneqq b(x)/(b(x) + b(1-x))$. Then let $g(v; a, b) \coloneqq 1/(1 + \exp(-(x-b)/a))$, a logistic function. The double logistic functions used are then defined as

$$
\begin{aligned}
f_1(v) &\coloneqq 8g(v, 0.25, 0)\ ;\\
f_2(v) &\coloneqq 4\big[\mathbf{1}\{4v \le -1\}g(4v, 0.4, -3)\\
&\qquad + \mathbf{1}\{1- < 4v \le 1\}a((4v+1)/2)(1 + g(1, 0.4, 3) - g(-1, 0.4, -3))\\
&\qquad + \mathbf{1}\{4v > 1\}(1 + g(4v, 0.4, 3))\big]\ ;\\
f_3(v) &\coloneqq 4\big[\mathbf{1}\{3v \le -1\}g(3v, 0.2, -3)\\
&\qquad + \mathbf{1}\{1- < 3v \le 1\}a((3v+1)/2)(1 + g(1, 0.2, 3) - g(-1, 0.2, -3))\\
&\qquad + \mathbf{1}\{3v > 1\}(1 + g(2v, 0.2, 3))\big]\ .
\end{aligned}
\tag{S35}
$$

These functions are plotted in Figure S5 and their derivatives are plotted in Figure S6.

[11] The second distribution for ϵ is based on the corresponding simulation design in Kuchibhotla and Patra (2020).

S4.2 IV model

Here I report (i) some additional results from the simulation designs in the main text and (ii) the results of some additional simulation designs, extending the simulation study in section 4.2.1 of the main text. In all designs, I consider $n \in \{200, 400, 600\}$. Empirical rejection frequencies are computed based on 5000 simulated data sets in Designs 1, 3, 4 and 2500 simulated data sets in Design 2. The π_j functions used in the simulations are plotted in Figures S8 – S10.

S4.2.1 Additional results from simulation designs 1 & 2

Design 1: Univariate, just identified Table S15 records the empirical rejection frequencies of the $\hat{S}$ tests in Design 1 with $k = 3$ fixed and k chosen by AIC / BIC as ν is varied. These results demonstrate that the choice of ν plays a limited role in most of the simulation designs. For the weakly identified designs a larger ν seems to be necessary to avoid some overejection. For the non-linear designs, the degree of overrejection observed here when ν is chosen too small is limited and remains substantially below that of the TSLS Wald and GMM tests (as recorded in Table 9) in all cases. For the linear designs the degree of overrejection is slightly larger and similar to that shown by the TSLS Wald test (and subtantially below that shown by the GMM tests).

Design 2: Bivariate, just identified Tables S16 and S17 report the empirical rejection frequencies of the considered tests when (i) π_1 and π_2 have the same form, but with j fixed at 3 for π_2 and (ii) where π_2 is always linear with $j = 3$. These tables show qualitative the same patterns as Table 10 in the main text: the $\hat{S}$ tests based on Legendre polynomials and AR test always show rejection frequencies close to the nominal level. The $\hat{S}$ test with OLS estimates is either close to the nominal level or underrejects, depending on the particular sub-design. The TSLS Wald test tends to overreject as do the GMM tests.

Tables S18, S19, S20 and S21 correspond, respectively, to the designs of Tables 10, 11, S16 and S17 and show that in these designs, the empirical rejection frequencies are relatively insensitive to the choice of ν.

Figures S11 – S19 display power surfaces in the settings of Figures 7 – 15 with the number of polynomials k chosen by information criteria. As noted in the main text, choosing k by AIC yields power surfaces which are similar to those with $k = 3$ fixed; choosing k by BIC tends to provide slightly lower power.

Finally, figures S20 – S25 display power surfaces where π_1 is either exponential or logistic and π_2 is linear. These show qualitatively similar behaviour to the power surfaces in the main text. One notable case is the strongly identified case with π_1 exponential and π_2 linear (Figure S20) here the AR test is able to detect violations of the null in θ_2 but not θ_1 – as the identfying information pertaining to the latter cannot be captured by a linear first stage – whilst the $\hat{S}$ test can detect violations in any direction.

S4.2.2 Additional simulation designs

Design 3: Univariate, just identified, heteroskedastic This design is the same as Design 1 in the main text with the addition of heteroskedastic errors. In particular, $d_\theta = 1$, $Z_1 = 1$ and $Z_2 \sim N(0, 1)$ is univariate. The considered $\pi(Z) = \pi(Z_2)$ are the exponential, logistic and linear functions detailed in Table 8 and plotted in Figures S8 – S10, for $j = 1, 2, 3$. I draw $(\tilde{\epsilon}, \tilde{\upsilon})$ from a multivariate normal distribution with unit variances and covariance 0.95 and set

$$\begin{bmatrix} \epsilon \\ \upsilon \end{bmatrix} = \begin{bmatrix} \sqrt{1+\sin(Z_2)^2} & 0 \\ 0 & \sqrt{1+\cos(Z_2)^2} \end{bmatrix} \begin{bmatrix} \tilde{\epsilon} \\ \tilde{\upsilon} \end{bmatrix}.$$

The same tests are considered as in Design 1 in the main text, with the AR and TSLS tests adapted to account for heteroskedasticity (following Andrews, Stock, and Sun (2019) for the AR test) and the GMM tests using a (feasible) efficient weighting matrix.

The empirical rejection frequencies under the null for each of these tests are reported in Table S22 with $\nu = 0.1$, which reveal that similar patterns hold in this heteroskedastic design: the null rejection probability of the $\hat{S}$ tests is well controlled in all scenarios, with some conservativeness for high j and rejection rates close to the nominal 5% level for lower j. The behaviour of the other tests is also similar to as in Design 1: the AR test always yields a rejection frequency of around 5%, whilst the TSLS Wald test generally provides a rejection frequency close to the nominal level when j is low, but begins to overreject as j increases. The same pattern is seen for the four GMM tests which exhibit substantial over-rejection in weakly identified settings (i.e. high j). The sensitivity of the rejection frequencies to the choice of ν is examined in Table S23, which reveals qualitatively the same behaviour as in Design 1.

The power of the $\hat{S}$ tests and the AR test in this design is examined in figures S26 – S28. The results are similar to the homoskedastic case. In particular, in the case with π exponential the AR test provides very little power across all j, whilst the $\hat{S}$ tests provide substantial power for $j = 1, 2$. For logistic π, the AR test and those $\hat{S}$ tests provide comparable power. Finally for the linear π simulations, the AR test provides the most power. In the case $j = 1$, the OLS based $\hat{S}$ tests provides comparable power to the AR test with the Legendre polynomial based test not far behind; both $\hat{S}$ tests perform slightly worse when $j = 2$.

Design 4: Univariate, over identified This design is based on Design 1 in the main text with the difference that Z_2 is bivariate whilst X remains scalar. In particular, $d_\theta = 1$, $Z_1 = 1$ and $Z_2 \sim N(0, \text{Var}(Z_2))$ where $\text{Var}(Z_2) = \left[\begin{smallmatrix} 1 & 0.4 \\ 0.4 & 1 \end{smallmatrix}\right]$. I draw (ϵ, υ) from a multivariate normal distribution with unit variances and covariance 0.95. π is formed by taking the mean of two of exponential, logistic and linear functions,[12] one evaluated at $Z_{2,1}$ and the other at $Z_{2,2}$.

The $\hat{S}$ tests considered estimate π by either (i) OLS or (ii) series regressions using tensor product bases formed of Legendre polynomials, both with $\nu = 0.1$. I additionally report the results of GMM Wald and LM tests using these tensor product bases as instruments and the TSLS Wald test and AR, LM (Kleibergen, 2002) and CLR (Moreira, 2003) tests.[13]

Tables S24 reports the empirical rejection frequencies of the $\hat{S}$ tests and the other considered tests. Similar to the other designs, the $\hat{S}$ tests have empirical rejection frequencies which do not substantially exceed the nominal 5% level in any of the considered cases. When OLS is used to estimate π, the $\hat{S}$ displays some underrejection for the weakly identified designs and where π_j is exponential. The other tests behave as one would expect: each of the "usual" weak-instrument robust tests (AR, LM & CLR) have rejection rates which are close to the nominal 5% level across all simulation designs. As in other designs, the TSLS Wald and GMM based tests display rejection frequencies close to the nominal 5% level in (some of the) strongly identified cases, but overreject substantially as identification weakens.

Table S25 investigates the sensitivity of the (Legendre polynomial based) $\hat{S}$ tests to the choice of ν. For the weakly identified designs a larger ν seems to be necessary to avoid some slight overejection. Nevertheless, the overrejection found when ν is too small is minor in each case and substantially below that observed for the TSLS Wald or GMM based tests.

The power of the $\hat{S}$, AR, LM and CLR tests is plotted in figures S29 – S31.

[12] as detailed in Table 8 and plotted in Figures S8 – S10.

[13] The CLR test is implemented using the p-value approximation given by Andrews, Moreira, and Stock (2007).

Similar to in Design 1, in the exponential case only the $\hat{S}$ tests (with π estimated non-parametrically) provide non-trivial power. These tests appear to have better finite sample performace either for k fixed at 3 or when k is chosen by AIC. For the logistic case, the $\hat{S}$ test based on Legendre polynomials or OLS are competitive with the CLR and LM tests and provide more power than the AR test when k is fixed at 3; when k is chosen by AIC the power declines slightly; using BIC causes a substantial power decline. In the linear case, unsurprisingly the LM and CLR tests provide the most power, with the AR test providing slightly more than the Legendre based $\hat{S}$ test.

S5 Tables and Figures

Figure S1: Index functions $f_j(v) = 5\exp(-v^2/2c_j^2)$

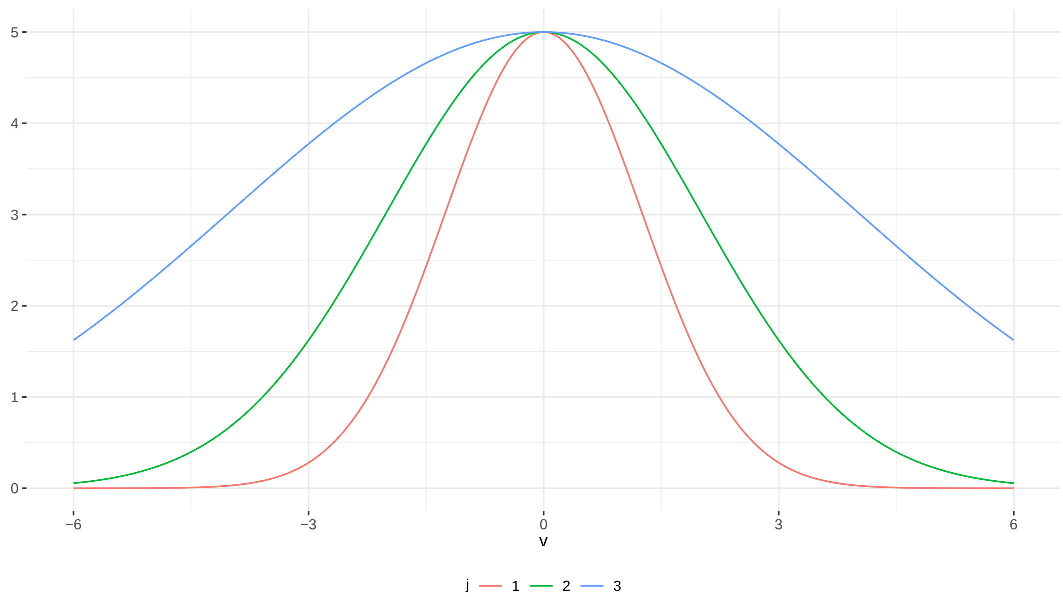


Figure S2: Derivatives $f_j'(v)$ of index functions $f_j(v) = 5\exp(-v^2/2c_j^2)$

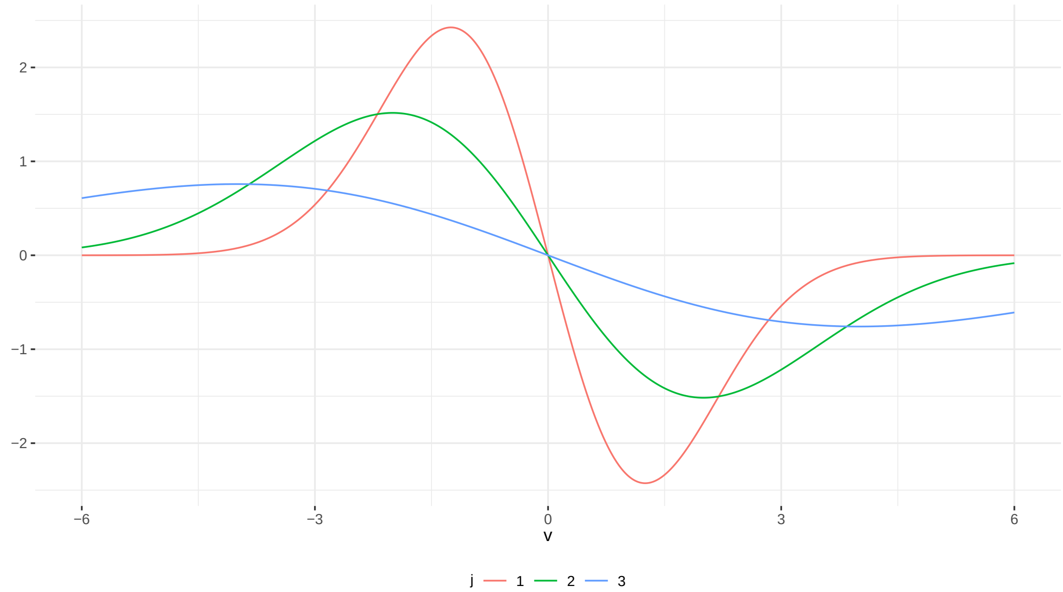


Figure S3: Index functions $f_j(v) = 25\,(1 + \exp(-v/c_j))^{-1}$

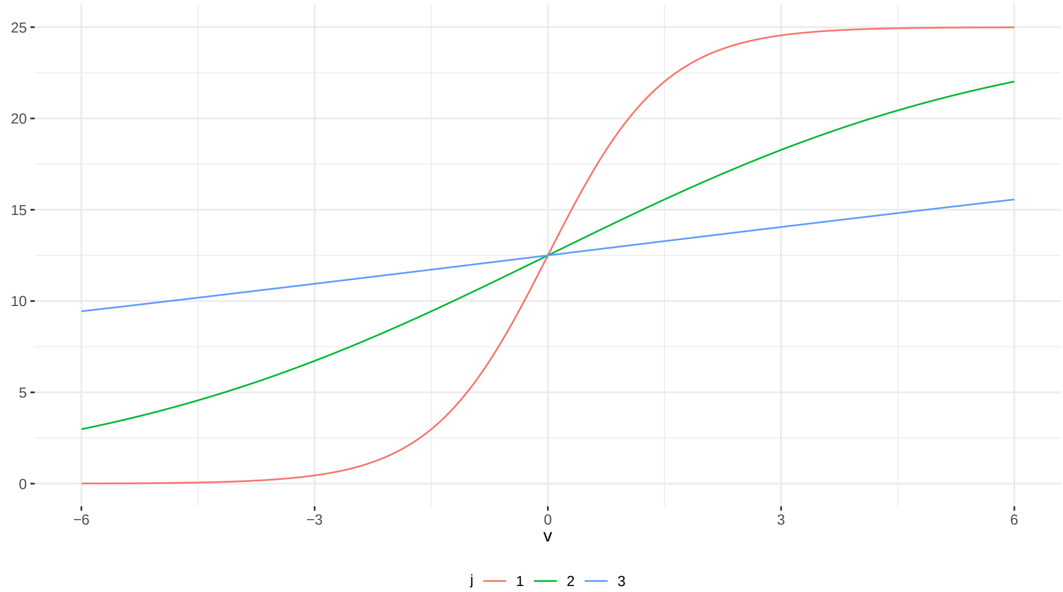


Figure S4: Derivatives $f_j'(v)$ of index functions $f_j(v) = 25\,(1 + \exp(-v/c_j))^{-1}$

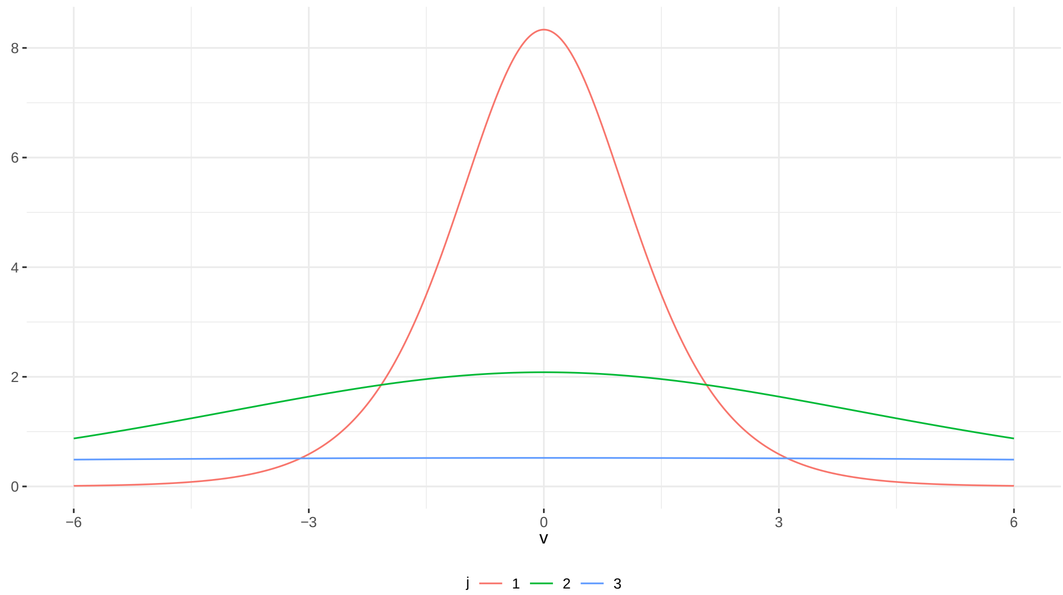


Figure S5: Double logistic index functions as in (S35)

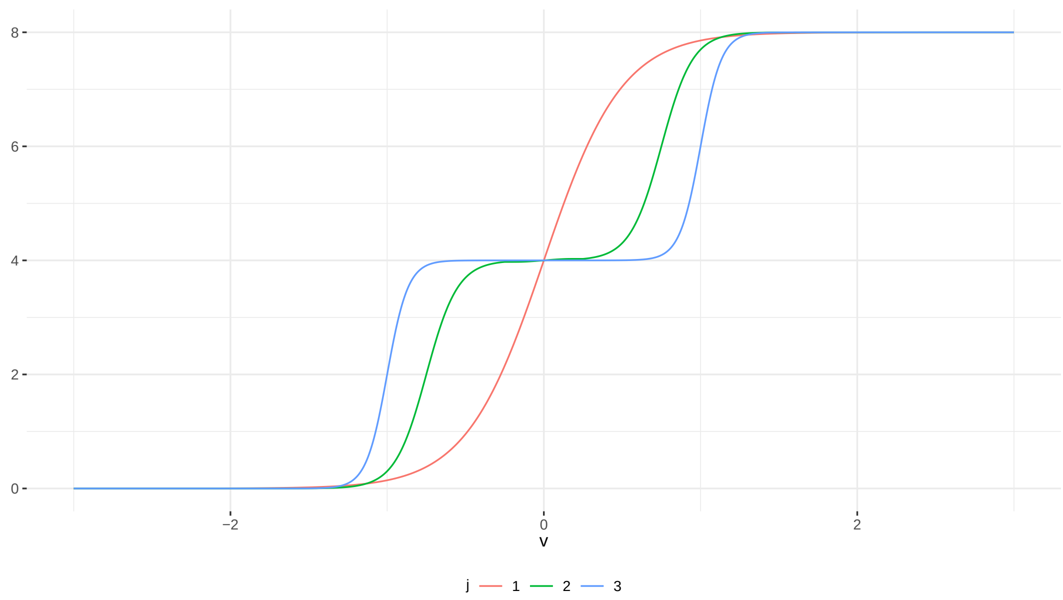


Figure S6: Derivatives of double logistic index functions as in (S35)

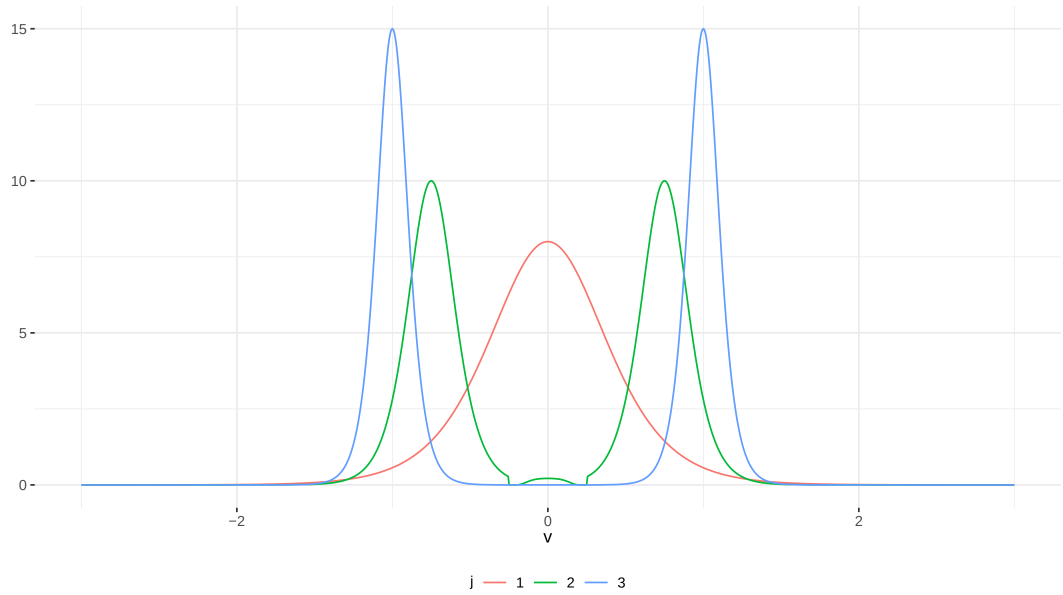


Table S1: ERF (%) $\epsilon|\xi \sim \sqrt{5}(-1)^{\xi}\,\mathrm{Beta}(2,3)$, $\xi \sim \mathrm{Bernoulli}(1/2)$, $f_j(v) = 5\exp(-v^2/2c_j^2)$

	$X = (Z_1, Z_2)$			$X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$		
n	f_1	f_2	f_3	f_1	f_2	f_3
$\hat{S}$						
400	5.94	5.66	5.08	5.46	5.20	4.98
600	4.90	5.06	4.34	5.64	5.40	4.90
800	5.56	5.82	5.26	5.16	4.88	4.50
Wald						
400	14.24	22.14	13.82	14.36	18.70	13.78
600	11.24	22.30	14.24	12.04	16.74	12.16
800	11.14	20.24	14.92	8.98	14.34	11.12

Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-1,1)$ are independently drawn.

Table S2: ERF (%) $\epsilon|\xi \sim \sqrt{5}(-1)^{\xi}\,\mathrm{Beta}(2,3)$, $\xi \sim$ Bernoulli(1/2), $f_j(v) = 25\,(1+\exp(-v/c_j))^{-1}$

	$X=(Z_1,Z_2)$			$X=(Z_1, 0.2Z_1+0.4Z_2+0.8)$		
n	f_1	f_2	f_3	f_1	f_2	f_3
$\hat{S}$						
400	4.92	5.02	4.86	6.26	5.94	5.62
600	4.56	4.40	4.74	4.90	4.96	4.90
800	4.40	4.52	4.56	4.90	4.42	4.46
Wald						
400	6.94	12.62	13.06	8.12	10.86	9.72
600	5.78	10.62	15.90	6.08	8.48	9.18
800	5.70	8.64	17.84	5.56	7.92	11.08

Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-1,1)$ are independently drawn.

Table S3: ERF (%) $\epsilon = \xi/\sqrt{3/2}$, $\xi \sim t(6)$, $f_j(v) = 5\exp(-v^2/2c_j^2)$

	$X=(Z_1,Z_2)$			$X=(Z_1, 0.2Z_1+0.4Z_2+0.8)$		
n	f_1	f_2	f_3	f_1	f_2	f_3
$\hat{S}$						
400	5.86	5.58	5.64	5.30	5.14	4.64
600	5.60	5.50	5.36	5.76	5.66	5.14
800	5.58	5.32	5.42	5.70	5.62	5.56
Wald						
400	14.64	23.16	13.80	14.52	18.72	13.22
600	12.18	23.34	13.94	11.84	16.92	12.78
800	11.26	19.70	14.44	10.70	16.44	11.48

Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-1,1)$ are independently drawn.

Table S4: ERF (%) $\epsilon = \xi/\sqrt{3/2}$, $\xi \sim t(6)$, $f_j(v) = 25\,(1+\exp(-v/c_j))^{-1}$

	$X = (Z_1, Z_2)$			$X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$		
n	f_1	f_2	f_3	f_1	f_2	f_3
$\hat{S}$						
400	5.32	5.16	5.16	6.02	5.52	5.32
600	5.32	5.44	5.34	5.62	5.36	5.26
800	5.12	5.26	5.20	5.66	5.42	5.40
Wald						
400	7.36	12.94	12.90	8.40	11.56	9.10
600	6.54	10.74	15.76	6.64	9.36	9.30
800	5.88	9.24	16.92	6.42	8.46	12.38

Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-1,1)$ are independently drawn.

Table S5: ERF (%) $\epsilon \sim \mathcal{N}(0, \log(2 + (X_1 + X_2\theta)^2))$, $f_j(v) = 5\exp(-v^2/2c_j^2)$

	$X = (Z_1, Z_2)$			$X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$		
n	f_1	f_2	f_3	f_1	f_2	f_3
$\hat{S}$						
400	5.84	5.70	5.30	5.36	5.26	5.12
600	5.38	5.36	5.20	5.66	5.32	4.86
800	5.52	5.50	5.50	5.74	6.06	5.70
Wald						
400	18.46	30.50	14.08	19.42	23.14	15.64
600	15.12	29.20	15.06	14.90	21.86	15.46
800	13.60	24.72	17.08	14.02	21.70	14.98

Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-1,1)$ are independently drawn.

Table S6: ERF (%) $\epsilon \sim \mathcal{N}(0, \log(2 + (X_1 + X_2\theta)^2)),\ f_j(v) = 25\,(1 + \exp(-v/c_j))^{-1}$

	$X = (Z_1, Z_2)$			$X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$		
n	f_1	f_2	f_3	f_1	f_2	f_3
$\hat{S}$						
400	5.36	5.30	5.26	6.20	5.68	5.38
600	5.40	5.34	5.42	5.50	5.54	5.52
800	5.28	5.26	5.26	5.82	5.26	5.46
Wald						
400	5.88	13.52	14.90	6.64	12.94	10.84
600	4.76	10.36	20.26	5.30	10.70	11.40
800	4.30	8.82	21.86	4.56	9.64	14.28

Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-1, 1)$ are independently drawn.

Table S7: ERF (%) $\epsilon \sim \mathcal{N}(0, 1 + \sin(X_1)^2),\ f_j(v) = 5\exp(-v^2/2c_j^2)$

	$X = (Z_1, Z_2)$			$X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$		
n	f_1	f_2	f_3	f_1	f_2	f_3
$\hat{S}$						
400	5.76	5.42	5.74	5.38	5.42	5.06
600	5.50	5.54	5.20	5.82	5.48	5.40
800	5.60	5.30	5.50	5.62	5.82	5.40
Wald						
400	18.48	22.82	14.28	17.72	20.36	14.58
600	14.22	27.06	14.96	14.46	19.36	14.54
800	13.70	25.38	14.38	12.06	19.52	13.50

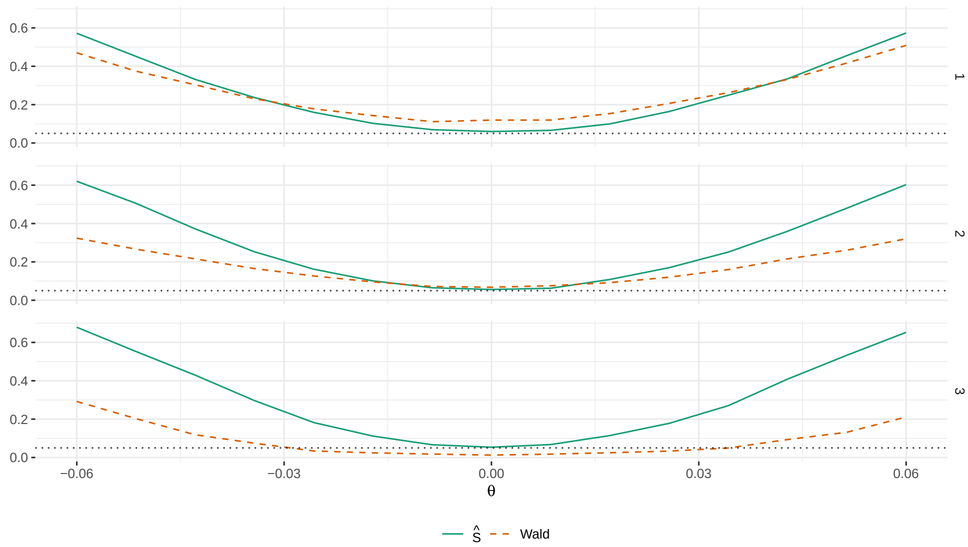
Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-1, 1)$ are independently drawn.

Table S8: ERF (%) $\epsilon \sim \mathcal{N}(0, 1 + \sin(X_1)^2)$, $f_j(v) = 25\,(1 + \exp(-v/c_j))^{-1}$

	$X = (Z_1, Z_2)$			$X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$		
n	f_1	f_2	f_3	f_1	f_2	f_3
$\hat{S}$						
400	5.28	5.10	5.18	6.06	5.68	5.24
600	5.38	5.26	5.50	5.44	5.32	5.24
800	5.02	5.10	5.04	5.66	5.42	5.46
Wald						
400	8.48	14.88	12.14	8.68	13.12	10.00
600	7.90	13.10	14.10	6.80	11.60	9.90
800	7.40	11.60	16.98	6.16	9.54	11.56

Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-1, 1)$ are independently drawn.

Figure S7: ERF (%) $\epsilon \sim \mathcal{N}(0, 1)$, index function as in (S35), $X = (Z_1, 0.2Z_2 + 0.4Z_2 + 0.8)$



Based on 5000 Monte carlo replications. The $Z_k \sim U(-3/2, 3/2)$ are independently drawn.

Table S9: ERF (%) $f_j(v) = 5\exp(-v^2/2c_j^2)$

		$X = (Z_1, Z_2)$					$X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$				
n	j	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0
$\epsilon \sim \eta_0$											
400	1	5.86	5.86	5.86	5.86	5.86	4.06	5.30	5.30	5.30	5.30
400	2	1.32	5.58	5.58	5.58	5.58	0.34	5.14	5.14	5.14	5.14
400	3	0.38	5.12	5.64	5.64	5.64	0.16	4.30	4.64	4.64	4.64
600	1	5.60	5.60	5.60	5.60	5.60	3.88	5.76	5.76	5.76	5.76
600	2	0.94	5.50	5.50	5.50	5.50	0.28	5.66	5.66	5.66	5.66
600	3	0.44	5.08	5.36	5.36	5.36	0.06	4.96	5.14	5.14	5.14
800	1	5.58	5.58	5.58	5.58	5.58	3.74	5.70	5.70	5.70	5.70
800	2	0.76	5.32	5.32	5.32	5.32	0.18	5.62	5.62	5.62	5.62
800	3	0.34	5.20	5.42	5.42	5.42	0.08	5.42	5.56	5.56	5.56
$\epsilon \sim \eta_1$											
400	1	5.94	5.94	5.94	5.94	5.94	3.76	5.46	5.46	5.46	5.46
400	2	1.38	5.66	5.66	5.66	5.66	0.10	5.20	5.20	5.20	5.20
400	3	0.38	4.66	5.08	5.08	5.08	0.04	4.68	4.98	4.98	4.98
600	1	4.90	4.90	4.90	4.90	4.90	3.90	5.64	5.64	5.64	5.64
600	2	1.08	5.06	5.06	5.06	5.06	0.14	5.40	5.40	5.40	5.40
600	3	0.38	4.12	4.34	4.34	4.34	0.06	4.78	4.90	4.90	4.90
800	1	5.56	5.56	5.56	5.56	5.56	4.08	5.16	5.16	5.16	5.16
800	2	0.78	5.82	5.82	5.82	5.82	0.12	4.88	4.88	4.88	4.88
800	3	0.16	5.02	5.26	5.26	5.26	0.06	4.42	4.50	4.50	4.50
$\epsilon \sim \eta_2$											
400	1	5.18	5.18	5.18	5.18	5.18	4.06	6.04	6.04	6.04	6.04
400	2	1.30	5.14	5.14	5.14	5.14	0.50	5.86	5.86	5.86	5.86
400	3	0.44	4.34	4.56	4.58	4.58	0.22	5.28	5.64	5.64	5.64
600	1	5.44	5.44	5.44	5.44	5.44	3.70	5.50	5.50	5.50	5.50
600	2	0.92	5.42	5.42	5.42	5.42	0.28	5.38	5.38	5.38	5.38
600	3	0.34	4.84	5.14	5.14	5.14	0.14	5.10	5.24	5.24	5.24
800	1	5.32	5.32	5.32	5.32	5.32	3.14	5.10	5.10	5.10	5.10
800	2	0.74	5.62	5.62	5.62	5.62	0.22	5.14	5.14	5.14	5.14
800	3	0.26	4.86	5.12	5.12	5.12	0.14	5.40	5.58	5.58	5.58
$\epsilon \sim \eta_3$											
400	1	5.84	5.84	5.84	5.84	5.84	5.08	5.36	5.36	5.36	5.36
400	2	1.62	5.70	5.70	5.70	5.70	0.64	5.26	5.26	5.26	5.26
400	3	0.56	4.94	5.30	5.30	5.30	0.32	4.86	5.12	5.12	5.12
600	1	5.38	5.38	5.38	5.38	5.38	5.30	5.66	5.66	5.66	5.66
600	2	1.26	5.36	5.36	5.36	5.36	0.46	5.32	5.32	5.32	5.32
600	3	0.40	4.94	5.20	5.20	5.20	0.14	4.74	4.86	4.86	4.86
800	1	5.52	5.52	5.52	5.52	5.52	5.58	5.74	5.74	5.74	5.74
800	2	0.94	5.50	5.50	5.50	5.50	0.50	6.06	6.06	6.06	6.06
800	3	0.36	5.36	5.50	5.50	5.50	0.14	5.50	5.70	5.70	5.70
$\epsilon \sim \eta_4$											
400	1	5.76	5.76	5.76	5.76	5.76	5.18	5.38	5.38	5.38	5.38
400	2	2.22	5.42	5.42	5.42	5.42	0.64	5.42	5.42	5.42	5.42
400	3	0.70	5.28	5.74	5.74	5.74	0.32	4.80	5.06	5.06	5.06
600	1	5.50	5.50	5.50	5.50	5.50	5.52	5.82	5.82	5.82	5.82
600	2	2.00	5.54	5.54	5.54	5.54	0.46	5.48	5.48	5.48	5.48
600	3	0.70	5.02	5.20	5.20	5.20	0.14	5.32	5.40	5.40	5.40
800	1	5.60	5.60	5.60	5.60	5.60	5.46	5.62	5.62	5.62	5.62
800	2	1.66	5.30	5.30	5.30	5.30	0.52	5.82	5.82	5.82	5.82
800	3	0.38	5.34	5.50	5.50	5.50	0.18	5.28	5.40	5.40	5.40

Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-1, 1)$ are independently drawn.

Table S10: ERF (%) $f_j(v) = 25\,(1+\exp(-v/c_j))^{-1}$

n	j	$X=(Z_1,Z_2)$ 10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	$X=(Z_1, 0.2Z_1+0.4Z_2+0.8)$ 10^{-1}	10^{-3}	10^{-5}	10^{-7}	0
$\epsilon \sim \eta_0$											
400	1	5.32	5.32	5.32	5.32	5.32	6.02	6.02	6.02	6.02	6.02
400	2	5.16	5.16	5.16	5.16	5.16	5.38	5.52	5.52	5.52	5.52
400	3	0.50	5.16	5.16	5.16	5.16	0.22	5.32	5.32	5.32	5.32
600	1	5.32	5.32	5.32	5.32	5.32	5.62	5.62	5.62	5.62	5.62
600	2	5.44	5.44	5.44	5.44	5.44	5.32	5.36	5.36	5.36	5.36
600	3	0.28	5.34	5.34	5.34	5.34	0.06	5.26	5.26	5.26	5.26
800	1	5.12	5.12	5.12	5.12	5.12	5.66	5.66	5.66	5.66	5.66
800	2	5.26	5.26	5.26	5.26	5.26	5.40	5.42	5.42	5.42	5.42
800	3	0.32	5.20	5.20	5.20	5.20	0.08	5.40	5.40	5.40	5.40
$\epsilon \sim \eta_1$											
400	1	4.92	4.92	4.92	4.92	4.92	6.26	6.26	6.26	6.26	6.26
400	2	5.02	5.02	5.02	5.02	5.02	5.88	5.94	5.94	5.94	5.94
400	3	0.40	4.86	4.86	4.86	4.86	0.10	5.62	5.62	5.62	5.62
600	1	4.56	4.56	4.56	4.56	4.56	4.90	4.90	4.90	4.90	4.90
600	2	4.40	4.40	4.40	4.40	4.40	4.94	4.96	4.96	4.96	4.96
600	3	0.50	4.74	4.74	4.74	4.74	0.08	4.90	4.90	4.90	4.90
800	1	4.40	4.40	4.40	4.40	4.40	4.90	4.90	4.90	4.90	4.90
800	2	4.52	4.52	4.52	4.52	4.52	4.42	4.42	4.42	4.42	4.42
800	3	0.22	4.56	4.56	4.56	4.56	0.08	4.46	4.46	4.46	4.46
$\epsilon \sim \eta_2$											
400	1	5.06	5.06	5.06	5.06	5.06	6.24	6.24	6.24	6.24	6.24
400	2	5.14	5.14	5.14	5.14	5.14	5.56	5.72	5.72	5.72	5.72
400	3	0.66	5.34	5.34	5.34	5.34	0.20	5.62	5.62	5.62	5.62
600	1	5.44	5.44	5.44	5.44	5.44	5.70	5.70	5.70	5.70	5.70
600	2	5.38	5.38	5.38	5.38	5.38	5.32	5.54	5.54	5.54	5.54
600	3	0.46	5.38	5.38	5.38	5.38	0.14	5.62	5.62	5.62	5.62
800	1	5.44	5.44	5.44	5.44	5.44	5.82	5.82	5.82	5.82	5.82
800	2	4.96	4.96	4.96	4.96	4.96	5.56	5.60	5.60	5.60	5.60
800	3	0.26	4.98	4.98	4.98	4.98	0.16	5.68	5.68	5.68	5.68
$\epsilon \sim \eta_3$											
400	1	5.36	5.36	5.36	5.36	5.36	6.20	6.20	6.20	6.20	6.20
400	2	5.30	5.30	5.30	5.30	5.30	5.56	5.68	5.68	5.68	5.68
400	3	0.72	5.26	5.26	5.26	5.26	0.40	5.38	5.38	5.38	5.38
600	1	5.40	5.40	5.40	5.40	5.40	5.50	5.50	5.50	5.50	5.50
600	2	5.34	5.34	5.34	5.34	5.34	5.52	5.54	5.54	5.54	5.54
600	3	0.58	5.42	5.42	5.42	5.42	0.22	5.52	5.52	5.52	5.52
800	1	5.28	5.28	5.28	5.28	5.28	5.82	5.82	5.82	5.82	5.82
800	2	5.26	5.26	5.26	5.26	5.26	5.24	5.26	5.26	5.26	5.26
800	3	0.52	5.26	5.26	5.26	5.26	0.14	5.46	5.46	5.46	5.46
$\epsilon \sim \eta_4$											
400	1	5.28	5.28	5.28	5.28	5.28	6.06	6.06	6.06	6.06	6.06
400	2	5.10	5.10	5.10	5.10	5.10	5.68	5.68	5.68	5.68	5.68
400	3	1.22	5.18	5.18	5.18	5.18	0.32	5.24	5.24	5.24	5.24
600	1	5.38	5.38	5.38	5.38	5.38	5.44	5.44	5.44	5.44	5.44
600	2	5.26	5.26	5.26	5.26	5.26	5.32	5.32	5.32	5.32	5.32
600	3	0.88	5.50	5.50	5.50	5.50	0.16	5.24	5.24	5.24	5.24
800	1	5.02	5.02	5.02	5.02	5.02	5.66	5.66	5.66	5.66	5.66
800	2	5.10	5.10	5.10	5.10	5.10	5.42	5.42	5.42	5.42	5.42
800	3	0.76	5.04	5.04	5.04	5.04	0.22	5.46	5.46	5.46	5.46

Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-1,1)$ are independently drawn.

Table S11: ERF (%) $\epsilon|\xi \sim \sqrt{5}(-1)^{\xi}\,\text{Beta}(2,3)$, index function as in (S35)

	$X = (Z_1, Z_2)$			$X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$		
n	f_1	f_2	f_3	f_1	f_2	f_3
$\hat{S}$						
400	6.40	6.22	5.50	7.12	6.42	5.54
600	5.86	5.52	5.26	5.92	5.86	5.26
800	5.60	5.22	4.90	5.26	5.72	5.08
Wald						
400	13.32	7.22	3.66	14.28	7.96	3.76
600	11.70	7.34	2.66	12.46	7.62	2.12
800	10.22	6.36	1.14	9.84	7.04	1.14

Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-3/2, 3/2)$ are independently drawn.

Table S12: ERF (%) $\epsilon = \xi/\sqrt{3/2}$, $\xi \sim t(6)$,, index function as in (S35)

	$X = (Z_1, Z_2)$			$X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$		
n	f_1	f_2	f_3	f_1	f_2	f_3
$\hat{S}$						
400	6.32	6.34	5.88	6.48	5.56	5.16
600	6.12	5.50	5.08	6.26	5.52	5.24
800	5.76	5.56	4.38	5.52	5.32	4.86
Wald						
400	13.12	8.52	3.68	13.62	7.94	3.62
600	11.20	6.82	2.50	11.84	7.18	2.30
800	11.28	6.30	1.64	10.72	7.12	1.14

Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-3/2, 3/2)$ are independently drawn.

Table S13: ERF (%) $\epsilon \sim \mathcal{N}(0, \log(2 + (X_1 + X_2\theta)^2))$, index function as in (S35)

	$X = (Z_1, Z_2)$			$X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$		
n	f_1	f_2	f_3	f_1	f_2	f_3
$\hat{S}$						
400	6.74	6.42	6.28	6.74	6.32	5.78
600	6.32	5.60	5.74	5.56	5.94	5.42
800	5.64	5.64	4.80	6.20	5.28	5.18
Wald						
400	11.36	9.58	5.74	10.96	9.22	5.64
600	9.38	7.20	2.92	8.96	7.84	3.00
800	7.46	7.04	1.80	8.78	7.10	1.76

Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-3/2, 3/2)$ are independently drawn.

Table S14: ERF (%) $\epsilon \sim \mathcal{N}(0, \log(2 + (X_1 + X_2\theta)^2))$, index function as in (S35)

	$X = (Z_1, Z_2)$			$X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$		
n	f_1	f_2	f_3	f_1	f_2	f_3
$\hat{S}$						
400	6.42	5.72	6.54	6.40	6.02	5.92
600	6.30	5.50	5.46	5.28	5.36	5.92
800	5.56	5.56	5.26	5.88	5.64	5.18
Wald						
400	10.50	8.80	5.86	12.02	9.80	5.64
600	9.22	7.56	3.48	9.66	7.62	3.46
800	7.80	6.58	2.38	8.00	7.50	2.62

Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-3/2, 3/2)$ are independently drawn.

Figure S8: $\pi_j(z) = 5\exp(-z^2/2c_j^2)$

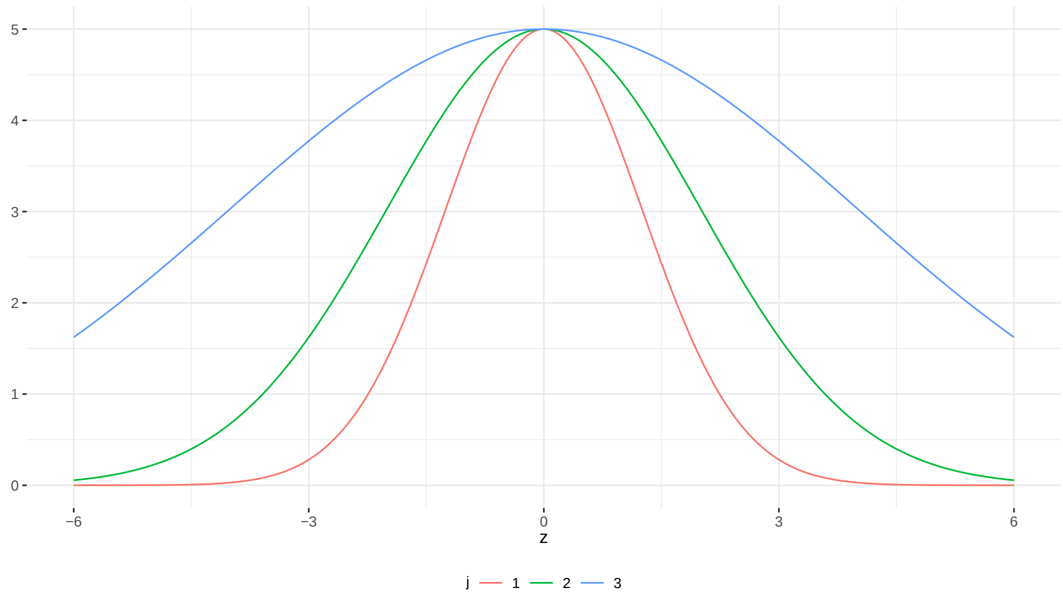


Figure S9: $\pi_j(z) = 25(1 + \exp(-z/c_j))^{-1}$

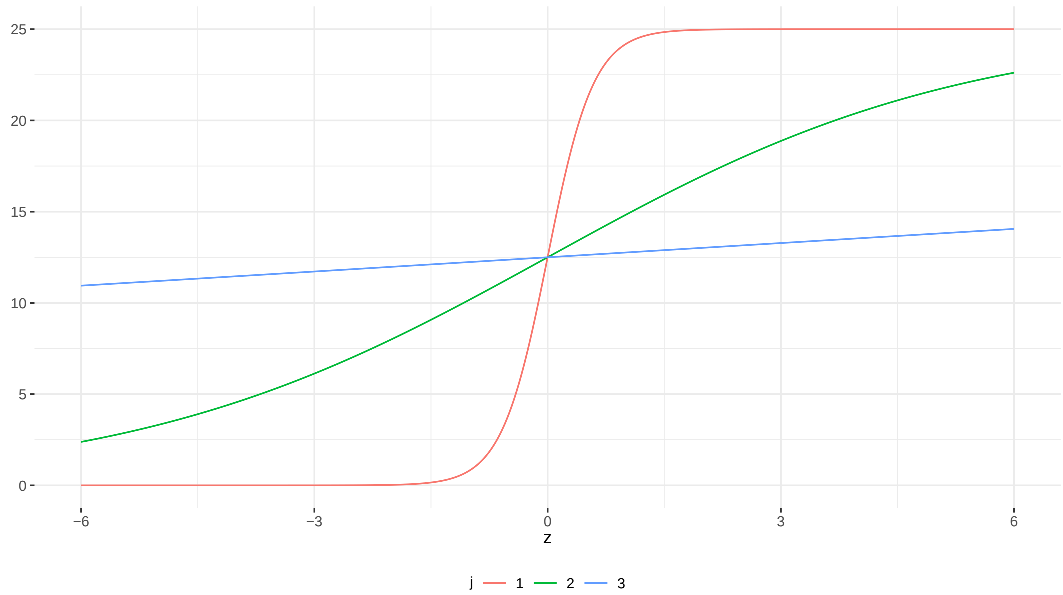


Figure S10: $\pi_j(z) = c_j z$

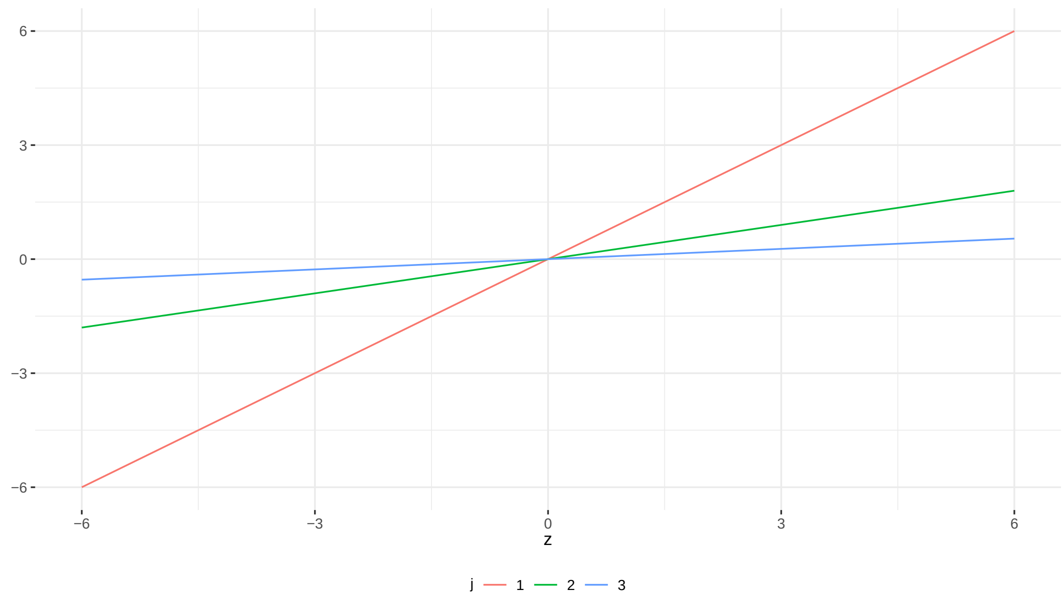


Table S15: Empirical rejection frequencies, Design 1

		k = 6					AIC					BIC				
n	j	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0
Exponential																
200	1	5.78	5.78	5.78	5.78	5.78	5.78	5.78	5.78	5.78	5.78	5.70	5.70	5.70	5.70	5.70
200	2	6.16	6.16	6.16	6.16	6.16	6.16	6.16	6.16	6.16	6.16	6.20	6.20	6.20	6.20	6.20
200	3	6.14	9.32	9.32	9.32	9.32	6.14	9.32	9.32	9.32	9.32	6.38	9.00	9.00	9.00	9.00
400	1	5.18	5.18	5.18	5.18	5.18	5.18	5.18	5.18	5.18	5.18	5.08	5.08	5.08	5.08	5.08
400	2	5.74	5.74	5.74	5.74	5.74	5.74	5.74	5.74	5.74	5.74	5.60	5.60	5.60	5.60	5.60
400	3	3.18	8.70	8.70	8.70	8.70	3.18	8.70	8.70	8.70	8.70	3.72	8.40	8.40	8.40	8.40
600	1	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30
600	2	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.40	5.40	5.40	5.40	5.40
600	3	2.14	7.92	7.92	7.92	7.92	2.14	7.92	7.92	7.92	7.92	2.64	7.86	7.86	7.86	7.86
Logistic																
200	1	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	4.52	4.52	4.52	4.52	4.52
200	2	4.74	4.74	4.74	4.74	4.74	4.74	4.74	4.74	4.74	4.74	4.78	4.78	4.78	4.78	4.78
200	3	6.80	9.20	9.20	9.20	9.20	6.80	9.20	9.20	9.20	9.20	6.92	8.64	8.64	8.64	8.64
400	1	4.86	4.86	4.86	4.86	4.86	4.60	4.60	4.60	4.60	4.60	4.54	4.54	4.54	4.54	4.54
400	2	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.00	5.00	5.00	5.00	5.00
400	3	4.26	8.58	8.58	8.58	8.58	4.26	8.58	8.58	8.58	8.58	4.62	7.96	7.96	7.96	7.96
600	1	5.46	5.46	5.46	5.46	5.46	5.30	5.30	5.30	5.30	5.30	4.76	4.76	4.76	4.76	4.76
600	2	5.46	5.46	5.46	5.46	5.46	5.46	5.46	5.46	5.46	5.46	5.44	5.44	5.44	5.44	5.44
600	3	3.68	8.08	8.08	8.08	8.08	3.68	8.08	8.08	8.08	8.08	3.92	7.62	7.62	7.62	7.62
Linear																
200	1	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.50	5.50	5.50	5.50	5.50
200	2	7.04	8.56	8.56	8.56	8.56	7.04	8.56	8.56	8.56	8.56	7.06	8.14	8.14	8.14	8.14
200	3	4.66	11.38	11.38	11.38	11.38	4.66	11.38	11.38	11.38	11.38	5.28	10.60	10.60	10.60	10.60
400	1	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32
400	2	5.08	7.98	7.98	7.98	7.98	5.08	7.98	7.98	7.98	7.98	5.18	7.48	7.48	7.48	7.48
400	3	1.12	12.14	12.14	12.14	12.14	1.12	12.14	12.14	12.14	12.14	2.08	11.34	11.34	11.34	11.34
600	1	5.58	5.58	5.58	5.58	5.58	5.58	5.58	5.58	5.58	5.58	5.62	5.62	5.62	5.62	5.62
600	2	4.30	7.44	7.44	7.44	7.44	4.30	7.44	7.44	7.44	7.44	4.62	7.26	7.26	7.26	7.26
600	3	0.28	11.18	11.18	11.18	11.18	0.28	11.18	11.18	11.18	11.18	0.66	10.48	10.48	10.48	10.48

Notes: Based on 5000 Monte carlo replications. All tests use Legendre polynomials to estimate π.

Table S16: Empirical rejection frequencies, IV, Design 2, π_1, π_2 same form with $j = 5$ for π_2

		$\hat{S}$				AR	TSLS W	GMM W	GMM LM
n	j	OLS	$k = 3$	AIC	BIC				
Exponential									
200	1	2.62	5.70	5.58	5.30	5.28	21.00	93.88	54.52
200	2	1.74	5.76	5.68	5.26	5.28	25.84	95.02	56.78
200	3	0.92	5.50	5.46	5.06	5.28	36.94	99.40	76.08
400	1	0.24	5.78	5.78	5.22	5.20	20.02	78.92	39.90
400	2	0.06	5.68	5.68	5.62	5.20	24.14	81.04	40.68
400	3	0.02	6.48	6.48	5.68	5.20	35.00	95.40	59.42
600	1	0.08	5.82	5.82	5.22	5.34	20.96	68.12	31.50
600	2	0.00	5.80	5.80	5.46	5.34	25.24	69.98	32.28
600	3	0.00	6.34	6.34	6.24	5.34	36.68	88.00	48.02
Logistic									
200	1	5.78	5.54	3.98	4.02	5.28	8.88	87.08	47.96
200	2	5.68	5.62	5.66	5.36	5.28	8.48	85.84	47.24
200	3	4.56	5.16	5.14	4.84	5.28	8.72	95.36	60.26
400	1	5.92	6.02	5.56	4.96	5.20	7.38	66.74	32.40
400	2	5.78	5.92	5.92	5.68	5.20	7.24	65.46	31.32
400	3	4.52	6.16	6.16	5.80	5.20	7.42	79.14	42.14
600	1	5.92	5.40	5.00	4.14	5.34	7.04	53.42	27.08
600	2	5.84	5.26	5.26	4.98	5.34	6.88	52.02	26.38
600	3	4.12	5.94	5.94	5.62	5.34	6.94	63.94	32.82
Linear									
200	1	5.08	5.34	5.30	4.96	5.28	17.48	99.54	83.06
200	2	4.24	5.18	5.16	4.92	5.28	15.28	99.74	85.92
200	3	2.38	5.74	5.78	5.34	5.28	15.28	99.98	95.46
400	1	5.04	6.10	6.10	5.98	5.20	12.72	98.56	77.96
400	2	3.74	6.30	6.30	5.96	5.20	11.96	98.86	79.30
400	3	0.34	6.92	6.92	6.20	5.20	12.16	99.96	91.84
600	1	5.14	5.24	5.24	4.94	5.34	12.22	96.84	74.44
600	2	3.72	6.24	6.24	5.74	5.34	11.42	97.12	74.04
600	3	0.00	6.04	6.04	5.58	5.34	11.34	99.78	88.42

Notes: Based on 2500 Monte carlo replications. $k = 3$ indicates that each univariate series forming the tensor series has $k = 3$.

Table S17: Empirical rejection frequencies, IV, Design 2, $\pi_2 = c_3 Z_{2,2}$

		$\hat{S}$				AR	TSLS W	GMM W	GMM LM
n	j	OLS	$k=3$	AIC	BIC				
Exponential									
200	1	2.74	5.82	5.68	5.34	5.28	13.08	99.66	83.68
200	2	2.22	5.94	5.84	5.32	5.28	17.98	99.70	85.30
200	3	1.40	5.24	5.22	4.94	5.28	29.20	99.94	92.38
400	1	0.40	6.02	6.02	5.64	5.20	9.38	98.66	78.10
400	2	0.10	5.94	5.94	5.58	5.20	14.16	98.92	79.46
400	3	0.06	6.56	6.56	5.88	5.20	26.64	99.68	86.80
600	1	0.10	5.72	5.72	5.22	5.34	8.96	96.94	74.46
600	2	0.00	5.82	5.82	5.50	5.34	14.40	97.46	75.12
600	3	0.00	6.40	6.40	5.80	5.34	27.80	99.22	82.16
Logistic									
200	1	5.16	5.60	4.42	4.64	5.28	18.22	99.62	82.86
200	2	5.10	5.62	5.56	5.40	5.28	17.94	99.62	83.70
200	3	4.52	5.30	5.30	5.04	5.28	14.74	99.80	86.88
400	1	5.00	6.16	5.94	5.12	5.20	13.22	98.72	78.72
400	2	5.02	6.16	6.16	5.94	5.20	13.04	98.70	79.14
400	3	4.00	6.28	6.28	5.94	5.20	11.84	99.04	80.50
600	1	5.20	5.60	4.72	4.12	5.34	12.52	97.32	74.84
600	2	5.14	5.46	5.46	5.14	5.34	12.36	97.18	75.16
600	3	3.72	6.32	6.32	5.88	5.34	11.34	97.44	74.64
Linear									
200	1	5.08	5.34	5.30	4.96	5.28	17.48	99.54	83.06
200	2	4.24	5.18	5.16	4.92	5.28	15.28	99.74	85.92
200	3	2.38	5.74	5.78	5.34	5.28	15.28	99.98	95.46
400	1	5.04	6.10	6.10	5.98	5.20	12.72	98.56	77.96
400	2	3.74	6.30	6.30	5.96	5.20	11.96	98.86	79.30
400	3	0.34	6.92	6.92	6.20	5.20	12.16	99.96	91.84
600	1	5.14	5.24	5.24	4.94	5.34	12.22	96.84	74.44
600	2	3.72	6.24	6.24	5.74	5.34	11.42	97.12	74.04
600	3	0.00	6.04	6.04	5.58	5.34	11.34	99.78	88.42

Notes: Based on 2500 Monte carlo replications. $k = 3$ indicates that each univariate series forming the tensor series has $k = 3$.

Table S18: Empirical rejection frequencies, IV, Design 2, $\pi_{1,j} = \pi_{2,j}$

		k = 3					AIC					BIC				
n	j	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0
Exponential																
200	1	4.88	4.88	4.88	4.88	4.88	4.80	4.80	4.80	4.80	4.80	4.90	4.90	4.90	4.90	4.90
200	2	5.44	5.44	5.44	5.44	5.44	5.40	5.40	5.40	5.40	5.40	5.14	5.14	5.14	5.14	5.14
200	3	5.50	5.08	5.08	5.08	5.08	5.46	5.04	5.04	5.04	5.04	5.06	4.76	4.76	4.76	4.76
400	1	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.64	4.64	4.64	4.64	4.64
400	2	4.96	4.96	4.96	4.96	4.96	4.96	4.96	4.96	4.96	4.96	4.82	4.82	4.82	4.82	4.82
400	3	6.48	5.88	5.88	5.88	5.88	6.48	5.88	5.88	5.88	5.88	5.68	5.44	5.44	5.44	5.44
600	1	4.64	4.64	4.64	4.64	4.64	4.64	4.64	4.64	4.64	4.64	4.80	4.80	4.80	4.80	4.80
600	2	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.08	5.08	5.08	5.08	5.08
600	3	6.34	6.28	6.28	6.28	6.28	6.34	6.28	6.28	6.28	6.28	6.24	6.10	6.10	6.10	6.10
Logistic																
200	1	4.56	4.56	4.56	4.56	4.56	1.86	1.86	1.86	1.86	1.86	2.42	2.42	2.42	2.42	2.42
200	2	4.52	4.52	4.52	4.52	4.52	4.56	4.56	4.56	4.56	4.56	4.56	4.56	4.56	4.56	4.56
200	3	5.16	4.98	4.98	4.98	4.98	5.14	4.96	4.96	4.96	4.96	4.84	4.70	4.70	4.70	4.70
400	1	5.14	5.14	5.14	5.14	5.14	3.34	3.34	3.34	3.34	3.34	3.12	3.12	3.12	3.12	3.12
400	2	4.78	4.78	4.78	4.78	4.78	4.78	4.78	4.78	4.78	4.78	4.80	4.80	4.80	4.80	4.80
400	3	6.16	5.76	5.76	5.76	5.76	6.16	5.76	5.76	5.76	5.76	5.80	5.72	5.72	5.72	5.72
600	1	4.40	4.40	4.40	4.40	4.40	3.64	3.64	3.64	3.64	3.64	3.32	3.32	3.32	3.32	3.32
600	2	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.54	4.54	4.54	4.54	4.54
600	3	5.94	5.46	5.46	5.46	5.46	5.94	5.46	5.46	5.46	5.46	5.62	5.10	5.10	5.10	5.10
Linear																
200	1	4.92	4.92	4.92	4.92	4.92	4.90	4.90	4.90	4.90	4.90	4.68	4.68	4.68	4.68	4.68
200	2	5.02	4.90	4.90	4.90	4.90	4.98	4.86	4.86	4.86	4.86	4.80	4.72	4.72	4.72	4.72
200	3	5.74	5.20	5.20	5.20	5.20	5.78	5.20	5.20	5.20	5.20	5.34	5.06	5.06	5.06	5.06
400	1	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	4.68	4.68	4.68	4.68	4.68
400	2	5.78	5.60	5.60	5.60	5.60	5.78	5.60	5.60	5.60	5.60	5.60	5.52	5.52	5.52	5.52
400	3	6.92	6.08	6.08	6.08	6.08	6.92	6.08	6.08	6.08	6.08	6.20	5.70	5.70	5.70	5.70
600	1	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.72	4.72	4.72	4.72	4.72
600	2	5.58	5.42	5.42	5.42	5.42	5.58	5.42	5.42	5.42	5.42	5.30	5.10	5.10	5.10	5.10
600	3	6.04	6.52	6.52	6.52	6.52	6.04	6.52	6.52	6.52	6.52	5.58	5.92	5.92	5.92	5.92

Notes: Based on 2500 Monte carlo replications. $k = 3$ indicates that each univariate series forming the tensor series has $k = 3$.

Table S19: Empirical rejection frequencies, IV, Design 2, $\pi_{2,j} = c_j Z_{2,2}$

		k = 3					AIC					BIC				
n	j	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0
Exponential																
200	1	4.74	4.74	4.74	4.74	4.74	4.68	4.68	4.68	4.68	4.68	4.66	4.66	4.66	4.66	4.66
200	2	5.54	5.54	5.54	5.54	5.54	5.44	5.44	5.44	5.44	5.44	4.90	4.90	4.90	4.90	4.90
200	3	5.24	5.02	5.02	5.02	5.02	5.22	5.00	5.00	5.00	5.00	4.94	4.78	4.78	4.78	4.78
400	1	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.96	4.96	4.96	4.96	4.96
400	2	5.62	5.70	5.70	5.70	5.70	5.62	5.70	5.70	5.70	5.70	5.36	5.42	5.42	5.42	5.42
400	3	6.56	6.52	6.52	6.52	6.52	6.56	6.52	6.52	6.52	6.52	5.88	5.86	5.86	5.86	5.86
600	1	4.64	4.64	4.64	4.64	4.64	4.64	4.64	4.64	4.64	4.64	4.22	4.22	4.22	4.22	4.22
600	2	5.64	5.70	5.70	5.70	5.70	5.64	5.70	5.70	5.70	5.70	5.40	5.48	5.48	5.48	5.48
600	3	6.40	6.02	6.02	6.02	6.02	6.40	6.02	6.02	6.02	6.02	5.80	6.08	6.08	6.08	6.08
Logistic																
200	1	4.80	4.80	4.80	4.80	4.80	3.38	3.38	3.38	3.38	3.38	3.64	3.64	3.64	3.64	3.64
200	2	5.54	5.56	5.56	5.56	5.56	5.58	5.60	5.60	5.60	5.60	5.28	5.30	5.30	5.30	5.30
200	3	5.30	5.04	5.04	5.04	5.04	5.30	5.04	5.04	5.04	5.04	5.04	4.92	4.92	4.92	4.92
400	1	5.22	5.22	5.22	5.22	5.22	4.40	4.40	4.40	4.40	4.40	4.12	4.12	4.12	4.12	4.12
400	2	5.90	6.02	6.02	6.02	6.02	5.90	6.02	6.02	6.02	6.02	5.70	5.82	5.82	5.82	5.82
400	3	6.28	6.26	6.26	6.26	6.26	6.28	6.26	6.26	6.26	6.26	5.94	6.10	6.10	6.10	6.10
600	1	4.68	4.68	4.68	4.68	4.68	4.22	4.22	4.22	4.22	4.22	3.86	3.86	3.86	3.86	3.86
600	2	5.26	5.48	5.48	5.48	5.48	5.26	5.48	5.48	5.48	5.48	4.96	5.08	5.08	5.08	5.08
600	3	6.32	5.92	5.92	5.92	5.92	6.32	5.92	5.92	5.92	5.92	5.88	5.42	5.42	5.42	5.42
Linear																
200	1	4.92	4.92	4.92	4.92	4.92	4.90	4.90	4.90	4.90	4.90	4.68	4.68	4.68	4.68	4.68
200	2	5.02	4.90	4.90	4.90	4.90	4.98	4.86	4.86	4.86	4.86	4.80	4.72	4.72	4.72	4.72
200	3	5.74	5.20	5.20	5.20	5.20	5.78	5.20	5.20	5.20	5.20	5.34	5.06	5.06	5.06	5.06
400	1	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	4.68	4.68	4.68	4.68	4.68
400	2	5.78	5.60	5.60	5.60	5.60	5.78	5.60	5.60	5.60	5.60	5.60	5.52	5.52	5.52	5.52
400	3	6.92	6.08	6.08	6.08	6.08	6.92	6.08	6.08	6.08	6.08	6.20	5.70	5.70	5.70	5.70
600	1	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.72	4.72	4.72	4.72	4.72
600	2	5.58	5.42	5.42	5.42	5.42	5.58	5.42	5.42	5.42	5.42	5.30	5.10	5.10	5.10	5.10
600	3	6.04	6.52	6.52	6.52	6.52	6.04	6.52	6.52	6.52	6.52	5.58	5.92	5.92	5.92	5.92

Notes: Based on 2500 Monte carlo replications. $k = 3$ indicates that each univariate series forming the tensor series has $k = 3$.

Table S20: Empirical rejection frequencies, IV, Design 2, π_1, π_2 same form with $j = 3$ for π_2

		k = 3					AIC					BIC				
n	j	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0
Exponential																
200	1	5.70	5.70	5.70	5.70	5.70	5.58	5.58	5.58	5.58	5.58	5.30	5.30	5.30	5.30	5.30
200	2	5.76	5.74	5.74	5.74	5.74	5.68	5.66	5.66	5.66	5.66	5.26	5.24	5.24	5.24	5.24
200	3	5.50	5.08	5.08	5.08	5.08	5.46	5.04	5.04	5.04	5.04	5.06	4.76	4.76	4.76	4.76
400	1	5.78	5.88	5.88	5.88	5.88	5.78	5.88	5.88	5.88	5.88	5.22	5.36	5.36	5.36	5.36
400	2	5.68	5.78	5.78	5.78	5.78	5.68	5.78	5.78	5.78	5.78	5.62	5.66	5.66	5.66	5.66
400	3	6.48	5.88	5.88	5.88	5.88	6.48	5.88	5.88	5.88	5.88	5.68	5.44	5.44	5.44	5.44
600	1	5.82	5.90	5.90	5.90	5.90	5.82	5.90	5.90	5.90	5.90	5.22	5.24	5.24	5.24	5.24
600	2	5.80	6.10	6.10	6.10	6.10	5.80	6.10	6.10	6.10	6.10	5.46	5.70	5.70	5.70	5.70
600	3	6.34	6.28	6.28	6.28	6.28	6.34	6.28	6.28	6.28	6.28	6.24	6.10	6.10	6.10	6.10
Logistic																
200	1	5.54	5.54	5.54	5.54	5.54	3.98	4.00	4.00	4.00	4.00	4.02	4.04	4.04	4.04	4.04
200	2	5.62	5.64	5.64	5.64	5.64	5.66	5.68	5.68	5.68	5.68	5.36	5.38	5.38	5.38	5.38
200	3	5.16	4.98	4.98	4.98	4.98	5.14	4.96	4.96	4.96	4.96	4.84	4.70	4.70	4.70	4.70
400	1	6.02	6.12	6.12	6.12	6.12	5.56	5.62	5.62	5.62	5.62	4.96	5.10	5.10	5.10	5.10
400	2	5.92	6.10	6.10	6.10	6.10	5.92	6.10	6.10	6.10	6.10	5.68	5.88	5.88	5.88	5.88
400	3	6.16	5.76	5.76	5.76	5.76	6.16	5.76	5.76	5.76	5.76	5.80	5.72	5.72	5.72	5.72
600	1	5.40	5.64	5.64	5.64	5.64	5.00	5.16	5.16	5.16	5.16	4.14	4.36	4.36	4.36	4.36
600	2	5.26	5.64	5.64	5.64	5.64	5.26	5.64	5.64	5.64	5.64	4.98	5.20	5.20	5.20	5.20
600	3	5.94	5.46	5.46	5.46	5.46	5.94	5.46	5.46	5.46	5.46	5.62	5.10	5.10	5.10	5.10
Linear																
200	1	5.34	5.30	5.30	5.30	5.30	5.30	5.28	5.28	5.28	5.28	4.96	4.96	4.96	4.96	4.96
200	2	5.18	4.96	4.96	4.96	4.96	5.16	4.94	4.94	4.94	4.94	4.92	4.90	4.90	4.90	4.90
200	3	5.74	5.20	5.20	5.20	5.20	5.78	5.20	5.20	5.20	5.20	5.34	5.06	5.06	5.06	5.06
400	1	6.10	6.26	6.26	6.26	6.26	6.10	6.26	6.26	6.26	6.26	5.98	6.10	6.10	6.10	6.10
400	2	6.30	6.34	6.34	6.34	6.34	6.30	6.34	6.34	6.34	6.34	5.96	6.18	6.18	6.18	6.18
400	3	6.92	6.08	6.08	6.08	6.08	6.92	6.08	6.08	6.08	6.08	6.20	5.70	5.70	5.70	5.70
600	1	5.24	6.14	6.14	6.14	6.14	5.24	6.14	6.14	6.14	6.14	4.94	5.58	5.58	5.58	5.58
600	2	6.24	5.98	5.98	5.98	5.98	6.24	5.98	5.98	5.98	5.98	5.74	5.42	5.42	5.42	5.42
600	3	6.04	6.52	6.52	6.52	6.52	6.04	6.52	6.52	6.52	6.52	5.58	5.92	5.92	5.92	5.92

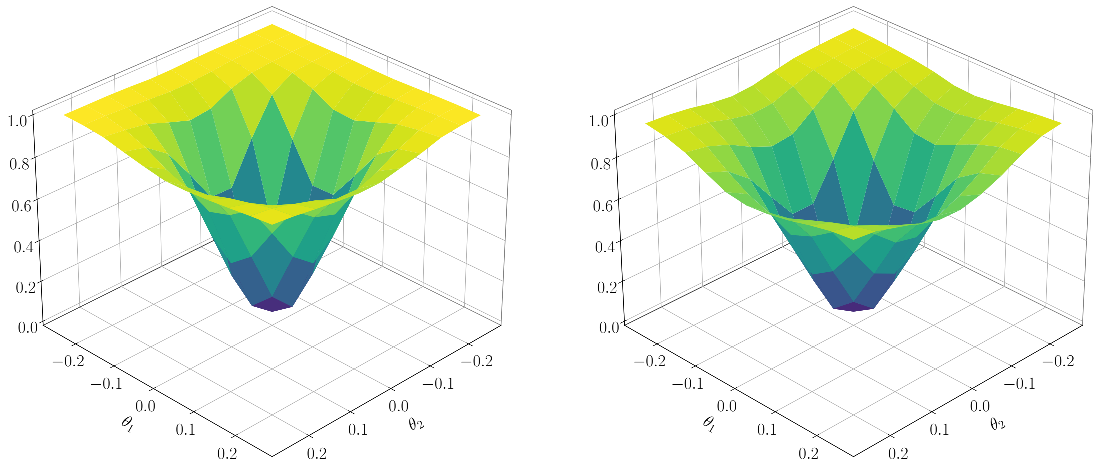
Notes: Based on 2500 Monte carlo replications. $k = 3$ indicates that each univariate series forming the tensor series has $k = 3$.

Table S21: Empirical rejection frequencies, IV, Design 2, $\pi_2 = c_5 Z_{2,2}$

		k = 3					AIC					BIC				
n	j	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0
Exponential																
200	1	5.82	5.80	5.80	5.80	5.80	5.68	5.66	5.66	5.66	5.66	5.34	5.34	5.34	5.34	5.34
200	2	5.94	5.92	5.92	5.92	5.92	5.84	5.82	5.82	5.82	5.82	5.32	5.28	5.28	5.28	5.28
200	3	5.24	5.02	5.02	5.02	5.02	5.22	5.00	5.00	5.00	5.00	4.94	4.78	4.78	4.78	4.78
400	1	6.02	6.38	6.38	6.38	6.38	6.02	6.38	6.38	6.38	6.38	5.64	5.80	5.80	5.80	5.80
400	2	5.94	6.26	6.26	6.26	6.26	5.94	6.26	6.26	6.26	6.26	5.58	5.94	5.94	5.94	5.94
400	3	6.56	6.52	6.52	6.52	6.52	6.56	6.52	6.52	6.52	6.52	5.88	5.86	5.86	5.86	5.86
600	1	5.72	6.68	6.68	6.68	6.68	5.72	6.68	6.68	6.68	6.68	5.22	5.82	5.82	5.82	5.82
600	2	5.82	6.74	6.74	6.74	6.74	5.82	6.74	6.74	6.74	6.74	5.50	6.00	6.00	6.00	6.00
600	3	6.40	6.02	6.02	6.02	6.02	6.40	6.02	6.02	6.02	6.02	5.80	6.08	6.08	6.08	6.08
Logistic																
200	1	5.60	5.58	5.58	5.58	5.58	4.42	4.44	4.44	4.44	4.44	4.64	4.64	4.64	4.64	4.64
200	2	5.62	5.58	5.58	5.58	5.58	5.56	5.54	5.54	5.54	5.54	5.40	5.36	5.36	5.36	5.36
200	3	5.30	5.04	5.04	5.04	5.04	5.30	5.04	5.04	5.04	5.04	5.04	4.92	4.92	4.92	4.92
400	1	6.16	6.48	6.48	6.48	6.48	5.94	6.26	6.26	6.26	6.26	5.12	5.44	5.44	5.44	5.44
400	2	6.16	6.30	6.30	6.30	6.30	6.16	6.30	6.30	6.30	6.30	5.94	6.06	6.06	6.06	6.06
400	3	6.28	6.26	6.26	6.26	6.26	6.28	6.26	6.26	6.26	6.26	5.94	6.10	6.10	6.10	6.10
600	1	5.60	6.26	6.26	6.26	6.26	4.72	5.72	5.72	5.72	5.72	4.12	4.92	4.92	4.92	4.92
600	2	5.46	6.16	6.16	6.16	6.16	5.46	6.16	6.16	6.16	6.16	5.14	5.66	5.66	5.66	5.66
600	3	6.32	5.92	5.92	5.92	5.92	6.32	5.92	5.92	5.92	5.92	5.88	5.42	5.42	5.42	5.42
Linear																
200	1	5.34	5.30	5.30	5.30	5.30	5.30	5.28	5.28	5.28	5.28	4.96	4.96	4.96	4.96	4.96
200	2	5.18	4.96	4.96	4.96	4.96	5.16	4.94	4.94	4.94	4.94	4.92	4.90	4.90	4.90	4.90
200	3	5.74	5.20	5.20	5.20	5.20	5.78	5.20	5.20	5.20	5.20	5.34	5.06	5.06	5.06	5.06
400	1	6.10	6.26	6.26	6.26	6.26	6.10	6.26	6.26	6.26	6.26	5.98	6.10	6.10	6.10	6.10
400	2	6.30	6.34	6.34	6.34	6.34	6.30	6.34	6.34	6.34	6.34	5.96	6.18	6.18	6.18	6.18
400	3	6.92	6.08	6.08	6.08	6.08	6.92	6.08	6.08	6.08	6.08	6.20	5.70	5.70	5.70	5.70
600	1	5.24	6.14	6.14	6.14	6.14	5.24	6.14	6.14	6.14	6.14	4.94	5.58	5.58	5.58	5.58
600	2	6.24	5.98	5.98	5.98	5.98	6.24	5.98	5.98	5.98	5.98	5.74	5.42	5.42	5.42	5.42
600	3	6.04	6.52	6.52	6.52	6.52	6.04	6.52	6.52	6.52	6.52	5.58	5.92	5.92	5.92	5.92

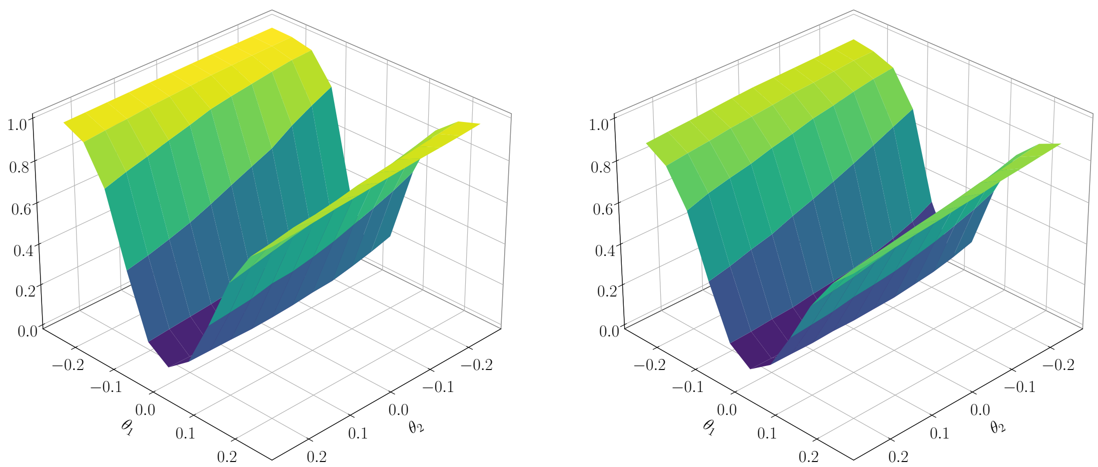
Notes: Based on 2500 Monte carlo replications. $k = 3$ indicates that each univariate series forming the tensor series has $k = 3$.

Figure S11: π_i exponential with $j = 1$ $(i = 1, 2)$



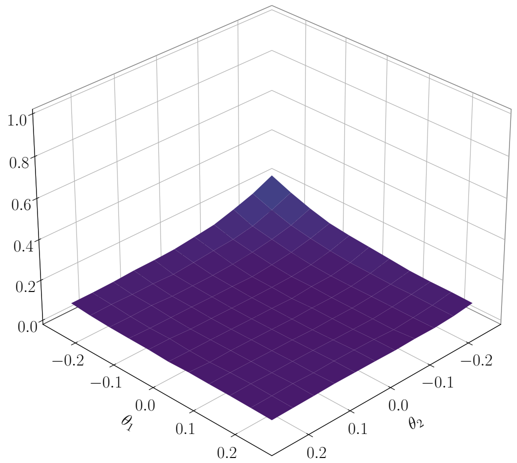
(a) $\hat{S}$ (AIC) (b) $\hat{S}$ (BIC)

Figure S12: π_1 exponential with $j = 1$, π_2 exponential with $j = 3$

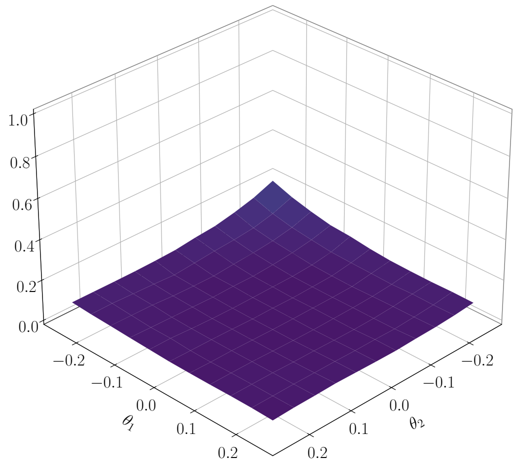


(a) $\hat{S}$ (AIC) (b) $\hat{S}$ (BIC)

Figure S13: π_1 exponential with $j = 3$, π_2 exponential with $j = 3$

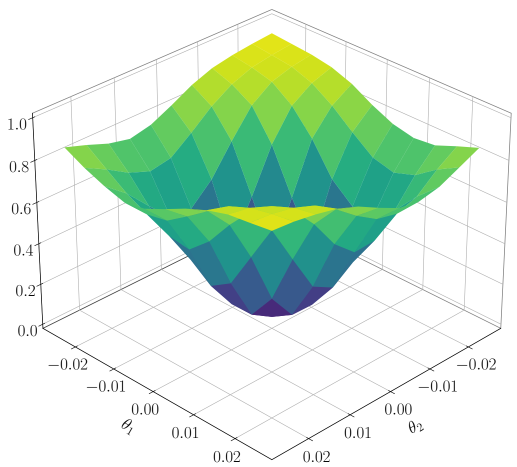


(a) $\hat{S}$ (AIC)

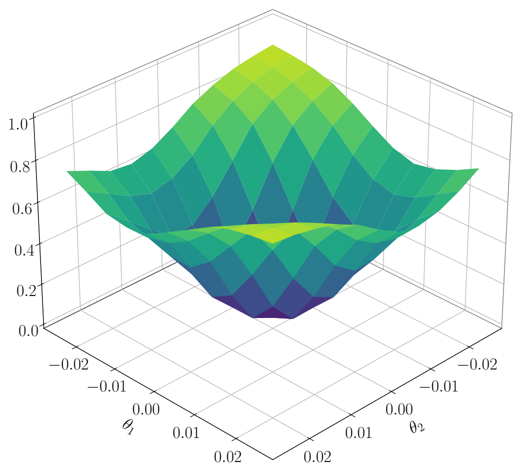


(b) $\hat{S}$ (BIC)

Figure S14: π_i logistic with $j = 1$ $(i = 1, 2)$

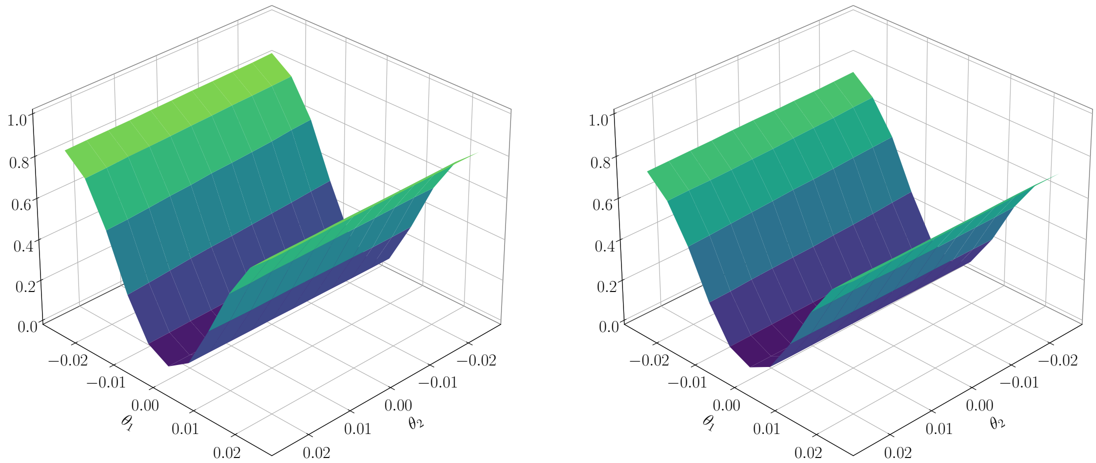


(a) $\hat{S}$ (AIC)



(b) $\hat{S}$ (BIC)

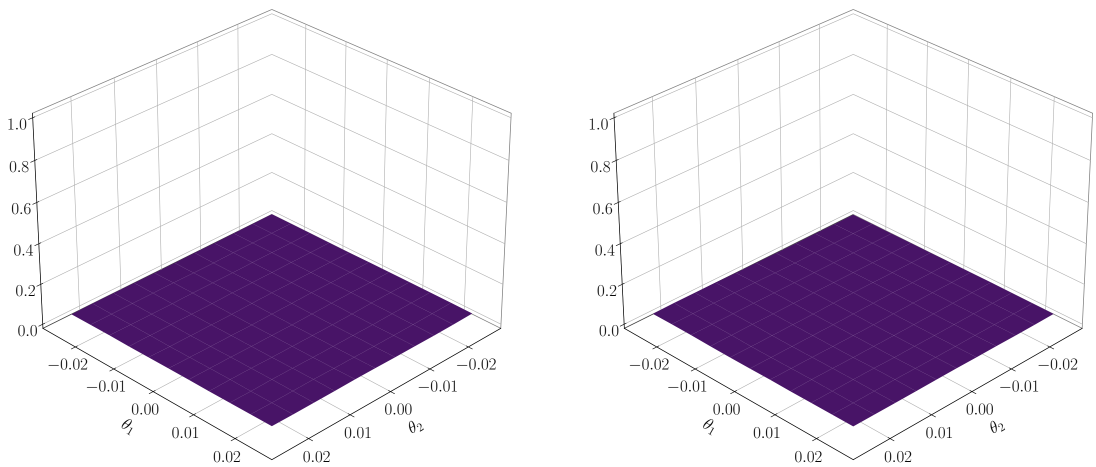
Figure S15: π_1 logistic with $j = 1$, π_2 logistic with $j = 3$



(a) $\hat{S}$ (AIC)

(b) $\hat{S}$ (BIC)

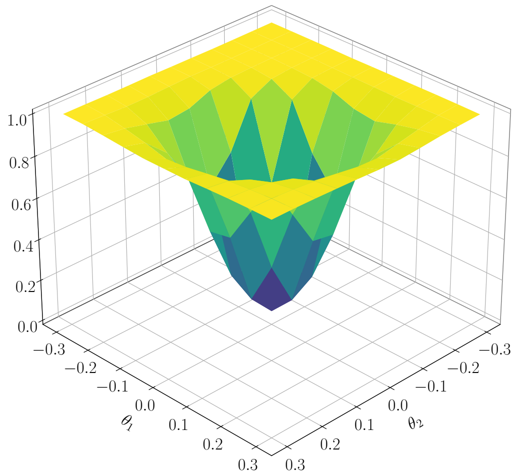
Figure S16: π_1 logistic with $j = 3$, π_2 logistic with $j = 3$



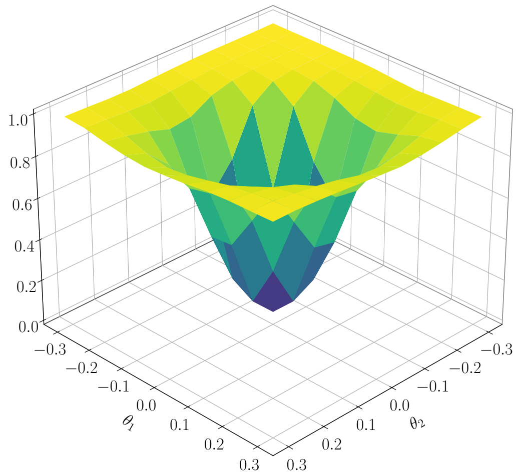
(a) $\hat{S}$ (AIC)

(b) $\hat{S}$ (BIC)

Figure S17: π_i linear with $j = 1$ $(i = 1, 2)$

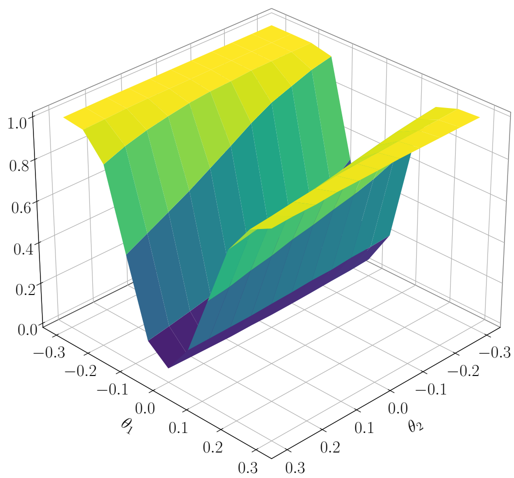


(a) $\hat{S}$ (AIC)

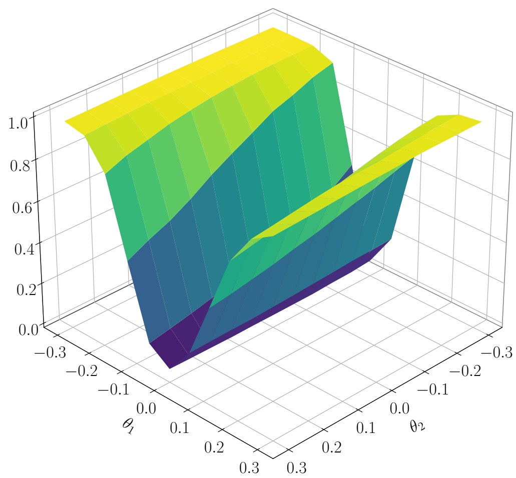


(b) $\hat{S}$ (BIC)

Figure S18: π_1 linear with $j = 1$, π_2 linear with $j = 3$

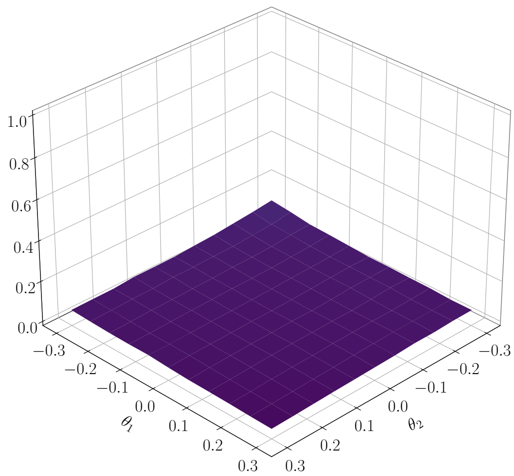


(a) $\hat{S}$ (AIC)

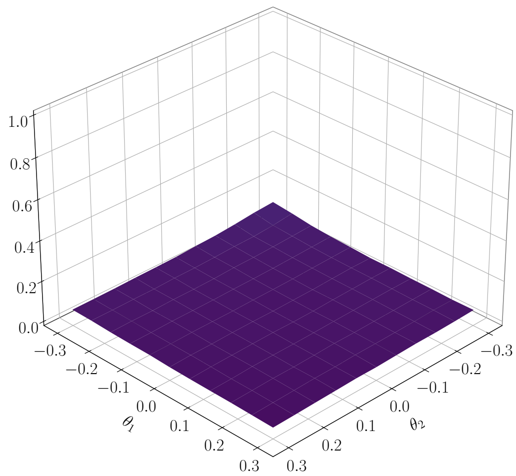


(b) $\hat{S}$ (BIC)

Figure S19: π_1 linear with $j = 3$, π_2 linear with $j = 3$

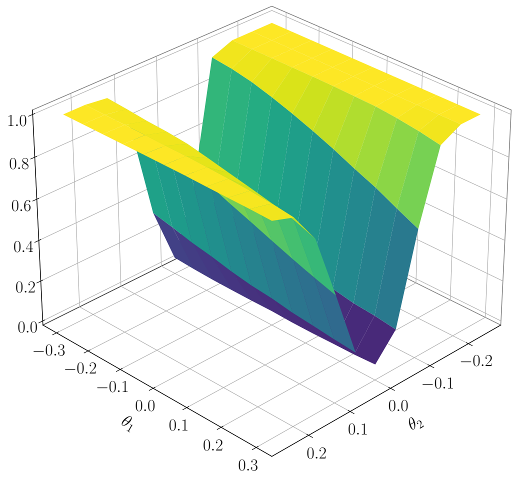


(a) $\hat{S}$ (AIC)

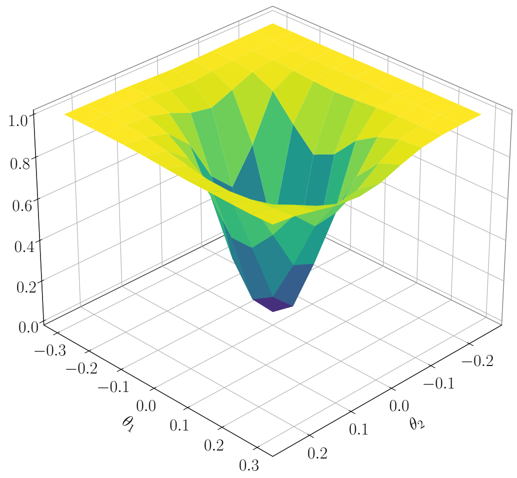


(b) $\hat{S}$ (BIC)

Figure S20: π_1 exponential with $j = 1$, π_2 linear with $j = 1$

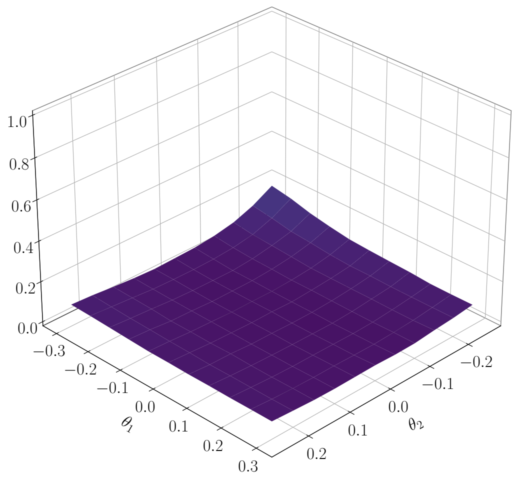


(a) AR

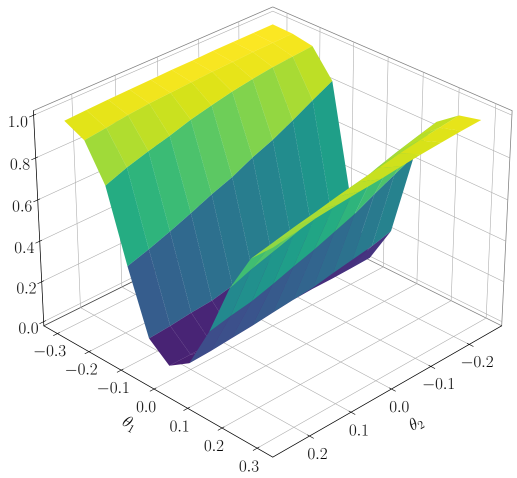


(b) $\hat{S}$ $(k = 3)$

Figure S21: π_1 exponential with $j = 1$, π_2 linear with $j = 3$

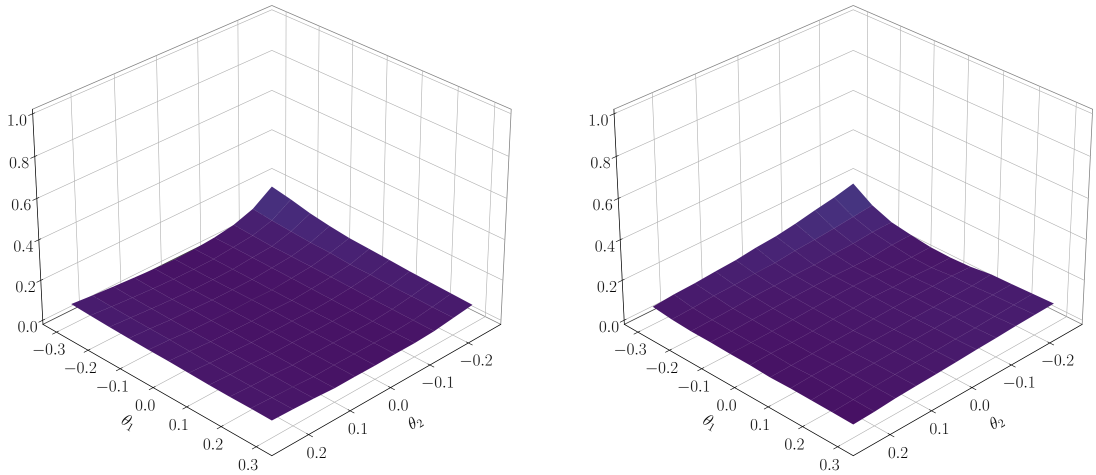


(a) AR



(b) $\hat{S}$ $(k = 3)$

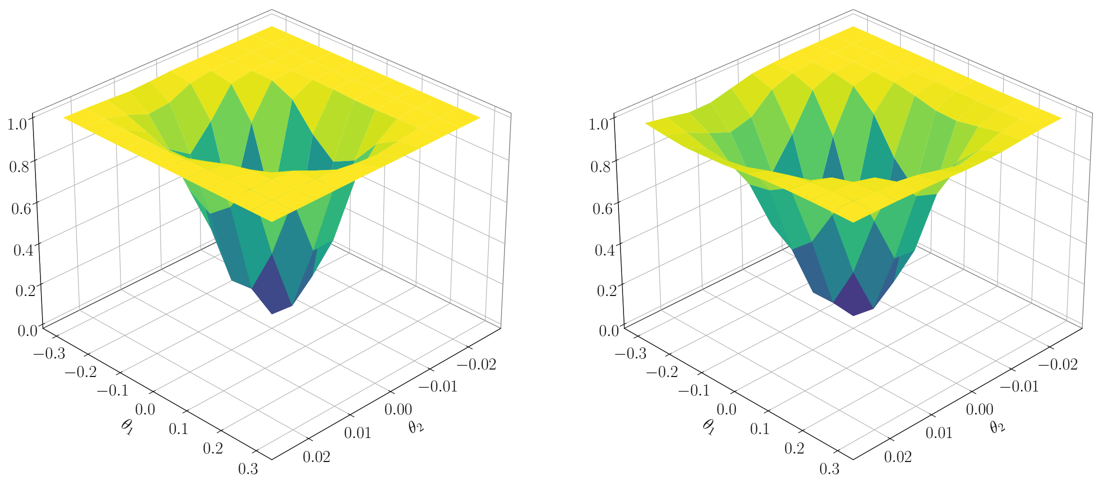
Figure S22: π_1 exponential with $j = 3$, π_2 linear with $j = 3$



(a) AR

(b) $\hat{S}$ $(k = 3)$

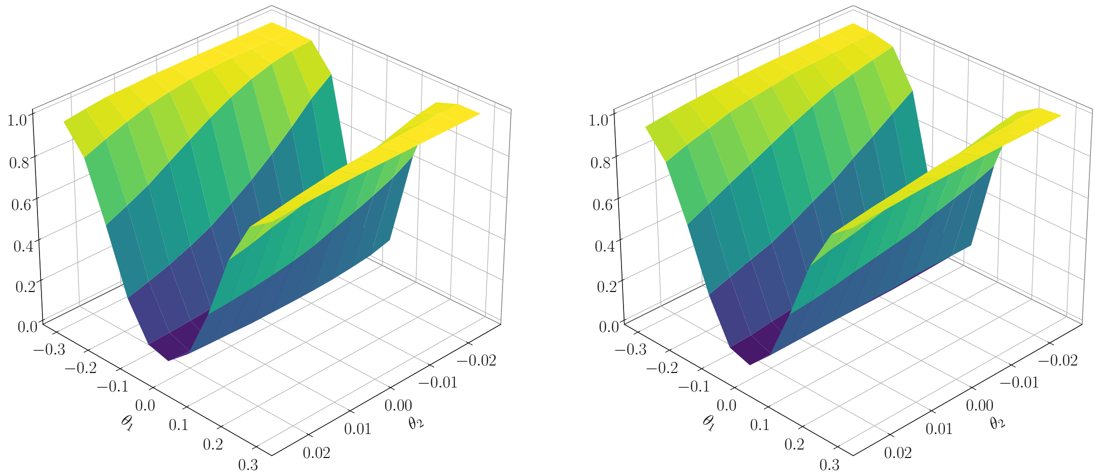
Figure S23: π_1 logistic with $j = 1$, π_2 linear with $j = 1$



(a) AR

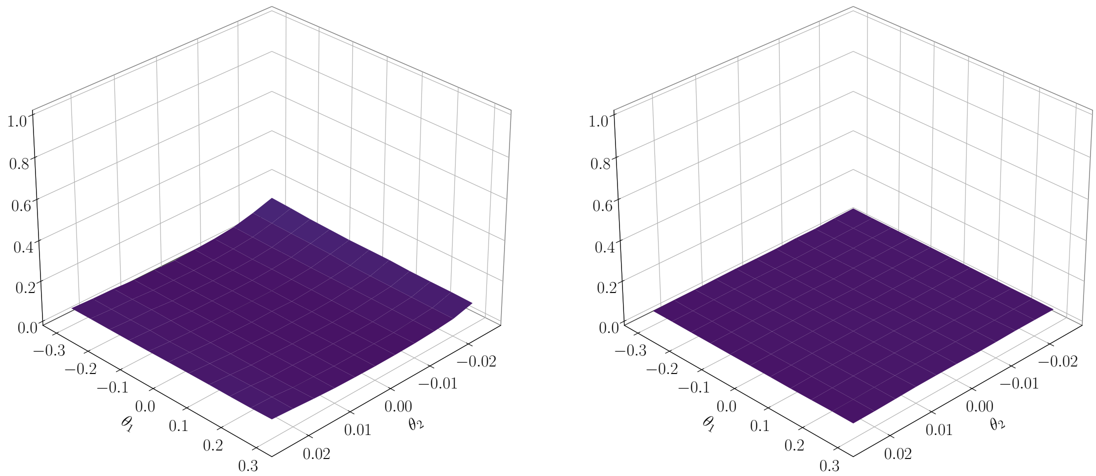
(b) $\hat{S}$ $(k = 3)$

Figure S24: π_1 logistic with $j = 1$, π_2 linear with $j = 3$



(a) AR (b) $\hat{S}$ $(k = 3)$

Figure S25: π_1 logistic with $j = 3$, π_2 linear with $j = 3$



(a) AR (b) $\hat{S}$ $(k = 3)$

Table S22: Empirical rejection frequencies, Design 3

		$\hat{S}$				AR	TSLS W	GMM W	GMM LM
n	j	OLS	$k=6$	AIC	BIC				
Exponential									
200	1	0.78	5.64	5.64	5.60	5.54	1.24	20.10	5.22
200	2	0.28	6.28	6.28	6.22	5.54	4.62	31.56	6.10
200	3	0.04	6.30	6.30	6.32	5.54	23.32	71.74	22.60
400	1	0.06	5.12	5.12	5.06	5.46	1.48	13.74	5.20
400	2	0.02	5.52	5.52	5.62	5.46	4.32	20.96	5.50
400	3	0.00	2.68	2.68	3.20	5.46	22.78	53.20	14.82
600	1	0.02	5.14	5.14	5.08	5.78	1.44	10.98	5.34
600	2	0.00	5.80	5.80	5.84	5.78	4.78	16.76	5.60
600	3	0.00	1.42	1.42	1.94	5.78	22.50	43.78	12.54
Logistic									
200	1	4.78	5.00	4.84	4.82	5.54	5.36	9.10	4.64
200	2	4.78	4.84	4.84	4.76	5.54	5.16	16.50	5.04
200	3	2.22	6.88	6.88	6.86	5.54	8.36	59.20	19.24
400	1	5.24	5.06	4.94	4.62	5.46	5.52	6.56	4.48
400	2	5.24	4.88	4.88	4.88	5.46	5.50	11.38	5.08
400	3	2.02	3.84	3.84	4.26	5.46	6.48	40.42	13.50
600	1	5.52	5.42	5.28	4.92	5.78	5.76	6.56	5.40
600	2	5.52	5.30	5.30	5.30	5.78	5.70	10.12	5.62
600	3	2.10	3.16	3.16	3.46	5.78	6.76	31.58	11.78
Linear									
200	1	4.78	5.50	5.50	5.50	5.54	4.98	21.00	5.96
200	2	2.28	7.04	7.04	7.04	5.54	7.62	52.44	15.88
200	3	0.16	5.30	5.30	5.66	5.54	16.36	93.76	48.02
400	1	5.24	5.26	5.26	5.24	5.46	5.28	14.04	5.48
400	2	2.24	4.70	4.70	4.92	5.46	6.02	35.14	11.64
400	3	0.02	1.28	1.28	2.12	5.46	11.96	88.48	41.48
600	1	5.52	5.70	5.70	5.74	5.78	5.86	12.52	6.06
600	2	2.94	3.98	3.98	4.40	5.78	6.34	27.20	10.52
600	3	0.00	0.40	0.40	0.84	5.78	11.32	83.64	37.48

Notes: Based on 5000 Monte carlo replications.

Table S23: Empirical rejection frequencies, Design 3

		k = 6					AIC					BIC				
n	j	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0
Exponential																
200	1	5.64	5.64	5.64	5.64	5.64	5.64	5.64	5.64	5.64	5.64	5.60	5.60	5.60	5.60	5.60
200	2	6.28	6.30	6.30	6.30	6.30	6.28	6.30	6.30	6.30	6.30	6.22	6.24	6.24	6.24	6.24
200	3	6.30	9.68	9.68	9.68	9.68	6.30	9.68	9.68	9.68	9.68	6.32	9.34	9.34	9.34	9.34
400	1	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.06	5.06	5.06	5.06	5.06
400	2	5.52	5.52	5.52	5.52	5.52	5.52	5.52	5.52	5.52	5.52	5.62	5.62	5.62	5.62	5.62
400	3	2.68	9.34	9.34	9.34	9.34	2.68	9.34	9.34	9.34	9.34	3.20	9.16	9.16	9.16	9.16
600	1	5.14	5.14	5.14	5.14	5.14	5.14	5.14	5.14	5.14	5.14	5.08	5.08	5.08	5.08	5.08
600	2	5.80	5.80	5.80	5.80	5.80	5.80	5.80	5.80	5.80	5.80	5.84	5.84	5.84	5.84	5.84
600	3	1.42	8.54	8.54	8.54	8.54	1.42	8.54	8.54	8.54	8.54	1.94	8.42	8.42	8.42	8.42
Logistic																
200	1	5.00	5.00	5.00	5.00	5.00	4.84	4.84	4.84	4.84	4.84	4.82	4.82	4.82	4.82	4.82
200	2	4.84	4.84	4.84	4.84	4.84	4.84	4.84	4.84	4.84	4.84	4.76	4.76	4.76	4.76	4.76
200	3	6.88	9.58	9.58	9.58	9.58	6.88	9.58	9.58	9.58	9.58	6.86	9.06	9.06	9.06	9.06
400	1	5.06	5.06	5.06	5.06	5.06	4.94	4.94	4.94	4.94	4.94	4.62	4.62	4.62	4.62	4.62
400	2	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88
400	3	3.84	8.94	8.94	8.94	8.94	3.84	8.94	8.94	8.94	8.94	4.26	8.60	8.60	8.60	8.60
600	1	5.42	5.42	5.42	5.42	5.42	5.28	5.28	5.28	5.28	5.28	4.92	4.92	4.92	4.92	4.92
600	2	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30	5.30
600	3	3.16	8.54	8.54	8.54	8.54	3.16	8.54	8.54	8.54	8.54	3.46	8.22	8.22	8.22	8.22
Linear																
200	1	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50
200	2	7.04	9.00	9.00	9.00	9.00	7.04	9.00	9.00	9.00	9.00	7.04	8.56	8.56	8.56	8.56
200	3	5.30	11.34	11.34	11.34	11.34	5.30	11.34	11.34	11.34	11.34	5.66	10.76	10.76	10.76	10.76
400	1	5.26	5.26	5.26	5.26	5.26	5.26	5.26	5.26	5.26	5.26	5.24	5.24	5.24	5.24	5.24
400	2	4.70	8.42	8.42	8.42	8.42	4.70	8.42	8.42	8.42	8.42	4.92	8.10	8.10	8.10	8.10
400	3	1.28	12.10	12.10	12.10	12.10	1.28	12.10	12.10	12.10	12.10	2.12	11.48	11.48	11.48	11.48
600	1	5.70	5.70	5.70	5.70	5.70	5.70	5.70	5.70	5.70	5.70	5.74	5.74	5.74	5.74	5.74
600	2	3.98	7.84	7.84	7.84	7.84	3.98	7.84	7.84	7.84	7.84	4.40	7.72	7.72	7.72	7.72
600	3	0.40	10.90	10.90	10.90	10.90	0.40	10.90	10.90	10.90	10.90	0.84	10.38	10.38	10.38	10.38

Notes: Based on 5000 Monte carlo replications.

Figure S26: Design 3, $\pi_j(z) = 5\exp(-z^2/2c_j^2)$

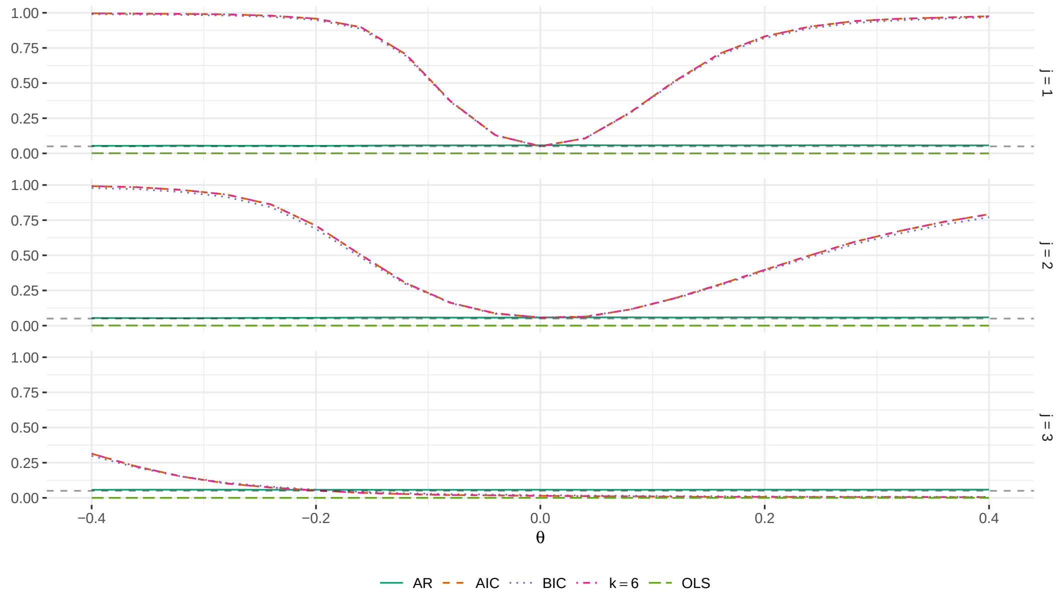


Figure S27: Design 3, $\pi_j(z) = 25(1 + \exp(-z/c_j))^{-1}$

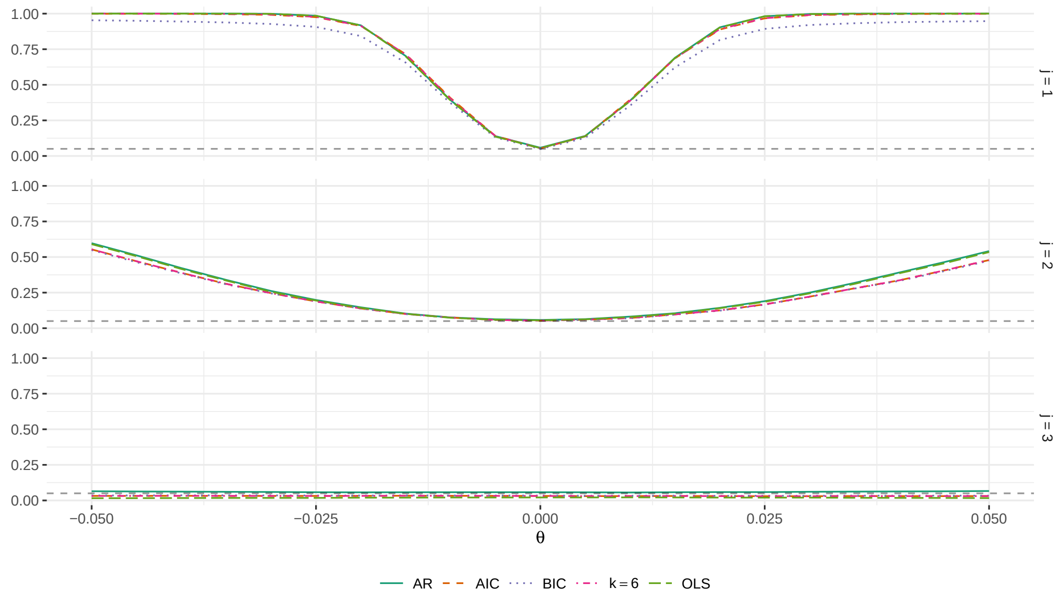


Figure S28: Design 3, $\pi_j(z) = c_j z$

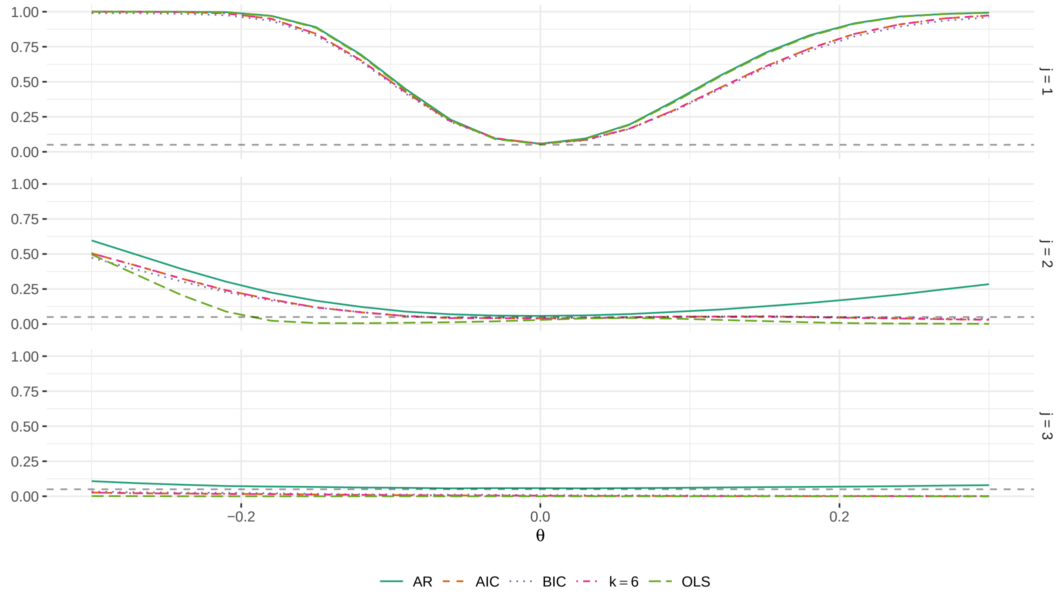


Figure S29: Design 4, $\pi_j(z) = 5\exp(-z^2/2c_j^2)$

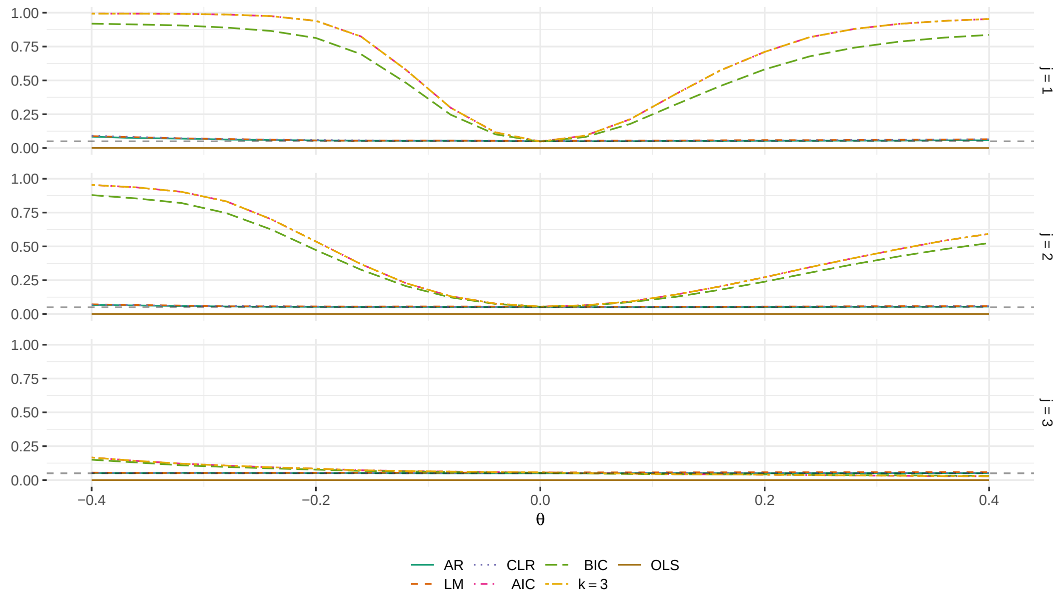


Table S24: Empirical rejection frequencies, IV, Design 4, $\pi_1 = \pi_2$

		$\hat{S}$				AR	LM	CLR	TSLS W	GMM W	GMM LM
n	j	OLS	$k=3$	AIC	BIC						
Exponential											
200	1	1.02	4.80	4.80	4.48	4.84	4.98	5.06	18.12	26.14	12.80
200	2	0.30	5.98	5.96	5.56	4.84	5.18	5.02	35.52	59.94	27.04
200	3	0.04	6.38	6.32	6.00	4.84	5.12	5.00	60.74	99.80	85.28
400	1	0.02	5.42	5.42	5.12	5.38	5.12	5.54	17.70	14.98	9.20
400	2	0.00	6.00	6.00	6.04	5.38	5.24	5.52	35.98	35.12	17.64
400	3	0.00	7.22	7.22	6.94	5.38	5.30	5.54	61.34	97.02	75.28
600	1	0.00	4.82	4.82	4.70	5.04	5.26	5.36	17.48	11.92	8.46
600	2	0.00	5.50	5.50	5.46	5.04	5.54	5.26	35.34	25.74	14.42
600	3	0.00	5.70	5.70	5.44	5.04	5.54	5.32	61.38	92.20	67.24
Logistic											
200	1	5.40	4.60	1.32	2.78	4.84	5.40	6.86	5.44	6.10	5.72
200	2	5.42	5.32	5.26	5.00	4.84	5.36	6.72	5.62	9.64	7.14
200	3	3.58	6.32	6.26	5.62	4.84	5.48	6.60	13.60	97.44	77.14
400	1	5.12	5.22	4.28	2.70	5.38	5.30	6.52	5.18	5.52	5.36
400	2	4.88	5.48	5.48	5.14	5.38	5.34	6.60	5.04	7.20	5.80
400	3	1.92	7.38	7.38	6.86	5.38	5.16	6.24	9.82	86.10	60.82
600	1	5.10	4.58	4.36	2.28	5.04	5.36	6.50	5.28	5.36	5.50
600	2	5.06	4.98	4.98	4.94	5.04	5.38	6.54	5.22	6.70	5.72
600	3	1.16	6.22	6.22	5.80	5.04	5.38	6.36	8.58	74.12	49.12
Linear											
200	1	5.32	5.68	5.62	5.10	4.84	5.44	6.74	5.86	25.98	14.34
200	2	3.80	6.26	6.20	5.62	4.84	5.42	6.72	11.74	94.78	69.16
200	3	0.14	6.22	6.10	5.66	4.84	5.26	5.92	42.12	100.00	97.30
400	1	4.74	5.92	5.92	5.32	5.38	5.32	6.48	5.48	15.60	9.72
400	2	2.94	7.18	7.18	6.72	5.38	5.22	6.30	8.92	78.00	51.38
400	3	0.00	7.40	7.40	6.78	5.38	5.04	5.72	30.70	99.88	95.96
600	1	5.02	5.32	5.32	5.26	5.04	5.40	6.56	5.46	12.44	8.62
600	2	2.64	6.44	6.44	5.98	5.04	5.42	6.38	7.92	63.18	40.16
600	3	0.00	4.52	4.52	4.32	5.04	5.26	6.02	25.00	99.52	94.34

Notes: Based on 5000 Monte carlo replications.

Table S25: Empirical rejection frequencies, IV, Design 4, $\pi_1 = \pi_2$, $\hat{S}$ tests

		k = 3					AIC					BIC				
n	j	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0	10^{-1}	10^{-3}	10^{-5}	10^{-7}	0
Exponential																
200	1	4.80	4.80	4.80	4.80	4.80	4.80	4.80	4.80	4.80	4.80	4.48	4.48	4.48	4.48	4.48
200	2	5.98	5.98	5.98	5.98	5.98	5.96	5.96	5.96	5.96	5.96	5.56	5.56	5.56	5.56	5.56
200	3	6.38	6.38	6.38	6.38	6.38	6.32	6.32	6.32	6.32	6.32	6.00	6.00	6.00	6.00	6.00
400	1	5.42	5.42	5.42	5.42	5.42	5.42	5.42	5.42	5.42	5.42	5.12	5.12	5.12	5.12	5.12
400	2	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.04	6.04	6.04	6.04	6.04
400	3	7.22	7.92	7.92	7.92	7.92	7.22	7.92	7.92	7.92	7.92	6.94	7.50	7.50	7.50	7.50
600	1	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.82	4.70	4.70	4.70	4.70	4.70
600	2	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.50	5.46	5.46	5.46	5.46	5.46
600	3	5.70	7.72	7.72	7.72	7.72	5.70	7.72	7.72	7.72	7.72	5.44	7.08	7.08	7.08	7.08
Logistic																
200	1	4.60	4.60	4.60	4.60	4.60	1.32	1.32	1.32	1.32	1.32	2.78	2.78	2.78	2.78	2.78
200	2	5.32	5.32	5.32	5.32	5.32	5.26	5.26	5.26	5.26	5.26	5.00	5.00	5.00	5.00	5.00
200	3	6.32	6.34	6.34	6.34	6.34	6.26	6.28	6.28	6.28	6.28	5.62	5.64	5.64	5.64	5.64
400	1	5.22	5.22	5.22	5.22	5.22	4.28	4.28	4.28	4.28	4.28	2.70	2.70	2.70	2.70	2.70
400	2	5.48	5.48	5.48	5.48	5.48	5.48	5.48	5.48	5.48	5.48	5.14	5.14	5.14	5.14	5.14
400	3	7.38	7.54	7.54	7.54	7.54	7.38	7.54	7.54	7.54	7.54	6.86	7.02	7.02	7.02	7.02
600	1	4.58	4.58	4.58	4.58	4.58	4.36	4.36	4.36	4.36	4.36	2.28	2.28	2.28	2.28	2.28
600	2	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.98	4.94	4.94	4.94	4.94	4.94
600	3	6.22	7.24	7.24	7.24	7.24	6.22	7.24	7.24	7.24	7.24	5.80	6.72	6.72	6.72	6.72
Linear																
200	1	5.68	5.68	5.68	5.68	5.68	5.62	5.62	5.62	5.62	5.62	5.10	5.10	5.10	5.10	5.10
200	2	6.26	6.28	6.28	6.28	6.28	6.20	6.22	6.22	6.22	6.22	5.62	5.64	5.64	5.64	5.64
200	3	6.22	6.22	6.22	6.22	6.22	6.10	6.10	6.10	6.10	6.10	5.66	5.66	5.66	5.66	5.66
400	1	5.92	5.92	5.92	5.92	5.92	5.92	5.92	5.92	5.92	5.92	5.32	5.32	5.32	5.32	5.32
400	2	7.18	7.30	7.30	7.30	7.30	7.18	7.30	7.30	7.30	7.30	6.72	6.84	6.84	6.84	6.84
400	3	7.40	8.30	8.30	8.30	8.30	7.40	8.30	8.30	8.30	8.30	6.78	7.60	7.60	7.60	7.60
600	1	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.32	5.26	5.26	5.26	5.26	5.26
600	2	6.44	7.20	7.20	7.20	7.20	6.44	7.20	7.20	7.20	7.20	5.98	6.64	6.64	6.64	6.64
600	3	4.52	7.56	7.56	7.56	7.56	4.52	7.56	7.56	7.56	7.56	4.32	6.84	6.84	6.84	6.84

Notes: Based on 5000 Monte carlo replications.

Figure S30: Design 4, $\pi_j(z) = 25(1 + \exp(-z/c_j))^{-1}$

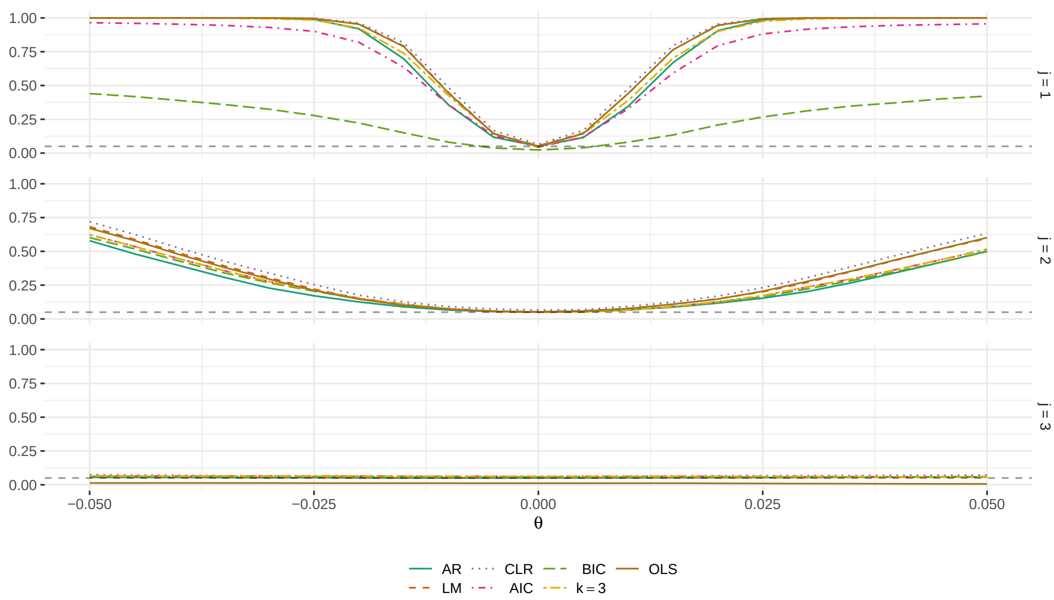
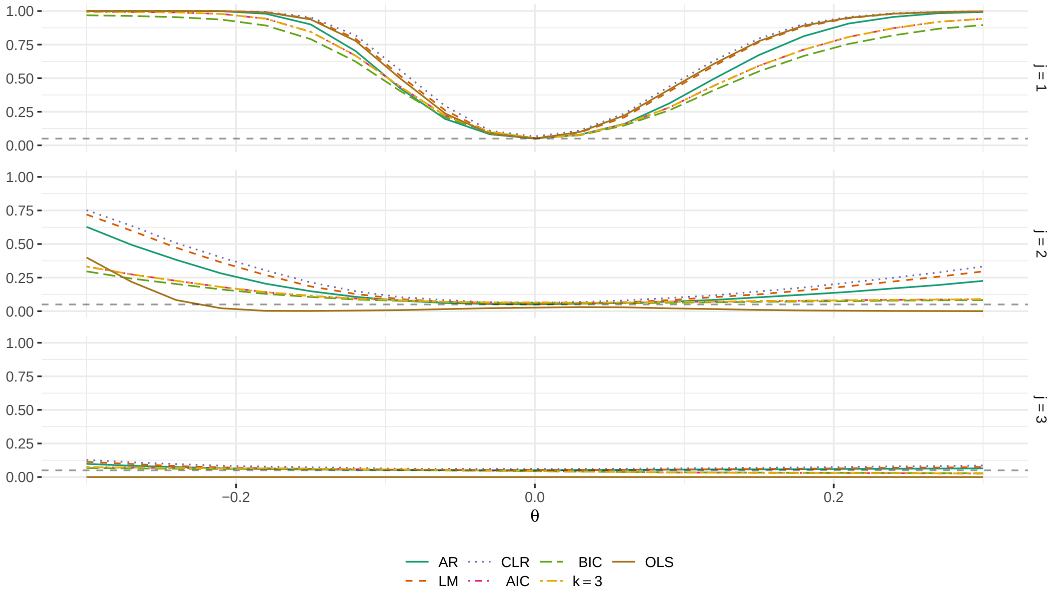


Figure S31: Design 4, $\pi_j(z) = c_j z$



References

Andrews, D. W., Moreira, M. J., and Stock, J. H. (2007), "Performance of conditional Wald tests in IV regression with weak instruments," *Journal of Econometrics*, 139, 116–132.

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

Andrews, I., Stock, J. H., and Sun, L. (2019), "Weak Instruments in Instrumental Variables Regression: Theory and Practice," *Annual Review of Economics*, 11, 727–753.

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

Bernstein, D. S. (2009), *Matrix mathematics: theory, facts, and formulas*, Princeton, NJ, USA: Princeton University Press, 2nd ed.

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

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

Dudley, R. M. (2002), *Real Analysis and Probability*, Cambridge, UK: Cambridge University Press.

Dufour, J.-M. and Valéry, P. (2016), "Rank-robust Regularized Wald-type tests," Working paper.

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

Hoesch, L., Lee, A., and Mesters, G. (2024), "Supplement to "Locally Robust Inference for Non-Gaussian SVAR Models"," *Quantitative Economics Supplementary Material*, forthcoming.

Janson, S. (1997), *Gaussian Hilbert Spaces*, Cambridge Tracts in Mathematics, Cambridge University Press.

Kagan, A. M., Linnik, I. U. V., and Rao, C. R. (1973), *Characterization Problems in Mathematical Statistics*, Wiley.

Kleibergen, F. (2002), "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803.

Kuchibhotla, A. K. and Patra, R. K. (2020), "Efficient estimation in single index models through smoothing splines," *Bernoulli*, 26, 1587 – 1618.

Lee, A. and Mesters, G. (2024a), "Locally Robust Inference for Non-Gaussian Linear Simultaneous Equations Models," *Journal of Econometrics*, 240, 105647.

— (2024b), "Supplement to 'Locally Robust Inference for Non-Gaussian Linear Simultaneous Equations Models'," *Journal of Econometrics Supplementary Material*, 240.

Lütkepohl, H. and Burda, M. M. (1997), "Modified wald tests under nonregular conditions," *Journal of Econometrics*, 78, 315–332.

Moreira, M. J. (2003), "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048.

Newey, W. 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*, eds. Barnett, W. A., Powell, J., and Tauchen, G. E., Cambridge University Press, International Symposia in Economic Theory and Econometrics.

Roman, S. (2005), *Advanced Linear Algebra*, Graduate Texts in Mathematics, Springer.

Royden, H. and Fitzpatrick, P. (2010), *Real Analysis*, Pearson Prentice Hall.

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

Serfozo, R. (1982), "Convergence of Lebesgue Integrals with Varying Measures," *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 44, 380–402.

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

van der Vaart, A. W. (1998), *Asymptotic Statistics*, New York, NY, USA: Cambridge University Press.

— (2002), "Semiparametric Statistics," in *Lectures on Probability Theory and Statistics: Ecole d'Eté de Probabilités de Saint-Flour XXIX - 1999*, ed. Bernard, P., Springer.