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

Adam Lee*

August 25, 2024

This Supplementary Material contains the following sections:

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

S1 Notation

$x := y$ 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. k-times continuously differentiable functions belong to $\mathcal{C}^k$. The standard basis vectors in $\mathbb{R}^K$ are $e_1, \ldots, e_K$. $M^\dagger$ is the Moore – Penrose pseudoinverse of M. $Pf := \int f \, \mathrm{d}P$, $\mathbb{P}_n f := \frac{1}{n}\sum_{i=1}^n f(Y_i)$ and $\mathbb{G}_n f := \sqrt{n}(\mathbb{P}_n - P)f$. For sequences 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 $Q_n \triangleleft P_n$ and $P_n \triangleleft Q_n$. $X \perp\!\!\!\perp Y$ indicates that random vectors X and Y are independent; $X \simeq Y$ indicates that they have the same distribution. $a \lesssim b$ means that $a \le Cb$ for some constant $C \in (0, \infty)$; C may change from line to line. If X is a subset of a topological space, $\operatorname{cl} X$ means the (topological) 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 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 $d_{TV}(P, Q)$. d_2 is the Mallows-2 metric (e.g. Bickel, Klaassen, Ritov, and Wellner, 1998, Appendix 6). $\overset{P_n}{\rightsquigarrow}$ denotes weak convergence under the sequence of measures $(P_n)_{n\in\mathbb{N}}$. If the sequence of measures is clear from context, we write just $\rightsquigarrow$.

S2 Additional results & discussion

S2.1 Inference under shape constraints

A non-standard inference problem which has attracted substantial attention in statistics & econometrics is inference when (finite-dimensional) nuisance parameters η may be at, or close to, the boundary. See, amongst others, Geyer (1994); Andrews (1999, 2001); Ketz (2018). In this scenario, as explained in detail in the aforementioned papers, the limiting distributions of extremum estimators are non-normal when the true parameter is at the boundary of the parameter space. In otherwise regular models, the same true when the true parameter is "close" to this boundary, i.e. along local (contiguous) alternatives to such a boundary point, by virtue of Le Cam's third Lemma.

In consequence, the normal approximations which usually obtain for extremum estimators (cf. Newey and McFadden, 1994) can lead to either misleading inference or poor power. The literature contains examples of boundary problems where "standard" tests over-reject (e.g. Andrews and Guggenberger, 2010) as well as examples where they are conservative and exhibit poor power (e.g. Ketz, 2018).

Under regularity conditions, boundary - constrained estimators of the nuisance parameters typically remain $\sqrt{n}$ - consistent (albeit not asymptotically normal). Due to the approximate orthogonalisation (10), plugging in any $\sqrt{n}$ - consistent estimator $\hat{\eta}_n$ of η is typically sufficient to ensure that the resulting feasible moment function (i.e. $\hat{g}_{n,\theta} = g_{\theta,\hat{\eta}_n}$) achieves the same normal limit as in Proposition 1.

In the semiparametric setting a natural generalisation of this boundary - constrained phenomenon is that of inference when nuisance functions are estimated under shape restrictions which may be close to binding.

Example 1 (Single-index model, continued): *Suppose that the class $\mathscr{F}$ of permitted link functions f in equation (1) imposes a shape restriction. For instance, $\mathscr{F}$ may contain only monotonically increasing functions or convex functions.*

Analogously to the parametric case, plugging in nuisance functions estimated under shape constraints causes no problems for C(α) style tests, which retain the same asymptotic distribution whether or not the constraints are (close to) binding. This is explored in simulation (based on Example 1) in Section S5.1.

Note that the power results of Section 3 typically do not apply to models with shape-constraints as – like in the parametric boundary case – the set B of possible perturbations to η will typically be a (linear) cone rather than a linear space.

S2.2 Uniform Local Asymptotic Normality

H is assumed to be a subset (containing 0) of a linear space equipped with a pseudometric.[1]

Assumption S1 (Uniform local asymptotic normality): *Equation (5) holds and $R_n(h_n) \xrightarrow{P_n} 0$ for any $h_n \to h$ in H. Additionally, for each $h_n \to h$ in H, $(\Delta_n h_n)_{n\in\mathbb{N}}$ is uniformly square P_n-integrable and $(\Delta_n h_n, \Delta_n h)' \overset{P_n}{\rightsquigarrow} \mathcal{N}\left(0, \sigma(h)\left[\begin{smallmatrix}1 & 1\\ 1 & 1\end{smallmatrix}\right]\right)$ where $\sigma(h) := \lim_{n\to\infty} \|\Delta_n h\|^2$.*

Remark S1: *The joint convergence of $(\Delta_n h_n, \Delta_n h)'$ in Assumption S1 is needed because H is not required to be linear. If H is a linear space this follows from d_2 convergence of (the law of) $\Delta_n h_n$ to $\mathcal{N}(0, \sigma(h))$ and the Cramér – Wold Theorem.*

[1] Proposition S1 below is an adaptation of Theorem 80.13 in Strasser (1985).

Remark S2: *If $(\Delta_n)_{n\in\mathbb{N}}$ is asymptotically equicontinuous on compact subsets $K \subset H$, then $h_n \to h$ in H implies $\|\Delta_n(h_n - h)\| \to 0$. In consequence $(\Delta_n h)_{n\in\mathbb{N}}$ being uniformly square P_n-integrable and $\Delta_n h \overset{P_n}{\rightsquigarrow} \mathcal{N}(0, \sigma(h))$ for each $h \in H$, suffices for $(\Delta_n h_n)_{n\in\mathbb{N}}$ being uniformly square P_n-integrable and for any $h_n \to h \in H$*

$$(\Delta_n h_n,\ \Delta_n h)' = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} (\Delta_n h_n - \Delta_n h,\ \Delta_n h)' \overset{P_n}{\rightsquigarrow} \mathcal{N}\left(0, \sigma(h) \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]\right).$$

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

Proposition S1: *Assumption S1 is equivalent to Assumption 1 plus asymptotic equicontinuity on compact subsets $K \subset H$ of $(\Delta_n)_{n\in\mathbb{N}}$ and $(h \mapsto P_{n,h})_{n\in\mathbb{N}}$ (in d_{TV}).*

Proof. Suppose Assumption 1 and the asymptotic equicontinuity conditions hold. Let $h_n \to h$ in H. By asymptotic equicontinuity of $(h \mapsto P_{n,h})_{n\in\mathbb{N}}$,

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

In combination with (compact) asymptotic equicontinuity of $(\Delta_n)_{n\in\mathbb{N}}$, this implies $R_n(h_n) - R_n(h) = o_{P_n}(1)$. That $(\Delta_n h_n)_{n\in\mathbb{N}}$ is uniformly square P_n-integrable and the joint weak convergence under P_n follows from the argument in Remark S2. Conversely, suppose Assumption S1 holds and let $h_n \to h$ in H. Then $\Delta_n(h_n - h) \xrightarrow{P_n} 0$ and so $\|\Delta_n(h_n - h)\|^2 \to 0$ by uniform square integrability (Serfozo, 1982, Theorem 2.7). $d_{TV}(P_{n,h_n}, P_{n,h}) \to 0$ holds by Lemma S6 since

$$L_n(h_n) - L_n(h) = \Delta_n h_n - \frac{1}{2}\|\Delta_n h_n\|^2 + R_n(h_n) - \left[\Delta_n h - \frac{1}{2}\|\Delta_n h\|^2 + R_n(h)\right],$$

and $R_n(h_n) = o_{P_n}(1)$, $R_n(h) = o_{P_n}(1)$, $\|\Delta_n(h_n - h)\|^2 \to 0$. $\square$

In the i.i.d. case, Lemma 4 recorded sufficient conditions for LAN (Assumption 1). Similar sufficient conditions are available for ULAN (Assumption S1).

Lemma S1: *Suppose that Assumption 5 holds and for each $h_n \to h$ in H, equation (24) holds with $A : \overline{\mathrm{lin}}\ H \to L_2(P)$ a bounded linear map. Then Assumption S1 holds with $P_{n,h} = P^n_{h/\sqrt{n}}$ and $[\Delta_n h](W^{(n)}) = \mathbb{G}_n Ah$.*

Proof. That $R_n(h_n) \xrightarrow{P_n} 0$ in (5) and that $Ah \in L_2^0(P)$ follows from (24) by e.g. Lemma 3.10.11 in van der Vaart and Wellner (1996). This immediately implies that $\Delta_n h$ is uniformly square integrable by the i.i.d assumption and that $[\Delta_n h](W^{(n)}) = \mathbb{G}_n Ah \rightsquigarrow \mathcal{N}(0, \sigma(h))$ for $\sigma(h) := \int (Ah)^2 \, \mathrm{d}P$ by the central limit theorem. In view of Remark S2 it remains to show that Δ_n is asymptotically equicontinuous on compact subsets $K \subset H$. This follows since A is bounded: for $h_n \to h$, $\|\Delta_n(h_n - h)\| = \|\mathbb{G}_n A(h_n - h)\| = \|A(h_n - h)\| \leq \|A\|\|h_n - h\| \to 0$. $\square$

S2.3 Additional results on uniform local regularity

S2.3.1 Asymptotic equicontinuity of power functions

Lemma S2: *Suppose the conditions of Theorem 1 hold and that (H, d) is a pseudo-metric space. Let δ metrise weak convergence on the space of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and let $Q_{n,h} = P_{n,h} \circ \hat{S}_{n,\theta_0}^{-1}$. Suppose that on a subset $K \subset H$,*

(i) $(h \mapsto Q_{n,h})_{n\in\mathbb{N}}$ is asymptotically equicontinuous in δ;

(ii) $(h \mapsto P_{n,h}(\hat{r}_{n,\theta_0} = r))_{n\in\mathbb{N}}$ is asymptotically equicontinuous;

(iii) $(h \mapsto P_{n,h}(\hat{\Lambda}_{n,\theta_0} = 0))_{n\in\mathbb{N}}$ is asymptotically equicontinuous;

then $(h \mapsto P_{n,h}\psi_{n,\theta_0})_{n\in\mathbb{N}}$ is asymptotically equicontinuous on K.

Proof. First suppose $r \geq 1$. By asymptotic equicontinuity of $h \mapsto Q_{n,h}$ and $h \mapsto P_{n,h}(\hat{r}_{n,\theta_0} = r)$ on K, for any $h_n \to h$ (through K), $\delta(Q_{n,h_n}, Q_{n,h}) \to 0$ and $|P_{n,h}(\hat{r}_{n,\theta_0} = r) - P_{n,h_n}(\hat{r}_{n,\theta_0} = r)| \to 0$. Since $\hat{r}_{n,\theta_0} \xrightarrow{P_{n,h}} r$ (Assumption 3 (iii) and Remark 1), $\hat{r}_{n,\theta_0} \xrightarrow{P_{n,h_n}} r$. Hence, under P_{n,h_n},

$$\hat{S}_{n,\theta_0} - c_n \rightsquigarrow S - c_r, \quad S \sim \chi^2_r(a) \implies P_{n,h_n}\psi_{n,\theta_0} \to 1 - \mathrm{P}(\chi^2_r(a) \leq c_r) =: \pi(\tau),$$

by Proposition 1 where c_r and a are as in Theorem 1. Thus, by Theorem 1,

$$|P_{n,h_n}\psi_{n,\theta_0} - P_{n,h}\psi_{n,\theta_0}| \leq |P_{n,h_n}\psi_{n,\theta_0} - \pi(\tau)| + |P_{n,h}\psi_{n,\theta_0} - \pi(\tau)| \to 0.$$

In the case where $r = 0$, the asymptotic equicontinuity on K of $h \mapsto P_{n,h}(\hat{\Lambda}_{n,\theta_0} = 0)$ implies that if $h_n \to h$ (through K), $|P_{n,h_n}(\hat{\Lambda}_{n,\theta_0} = 0) - P_{n,h}(\hat{\Lambda}_{n,\theta_0} = 0)| \to 0$. In combination with $\operatorname{rank}(\hat{\Lambda}_{n,\theta_0}) \xrightarrow{P_{n,h}} 0$ (Assumption 3 (iii) and Remark 1), this implies that $P_{n,h_n}(\hat{\Lambda}_{n,\theta_0} = 0) \to 1$ and thus $P_{n,h_n}\psi_{n,\theta_0} \to 0$. Thus, by Theorem 1

$$|P_{n,h_n}\psi_{n,\theta_0} - P_{n,h}\psi_{n,\theta_0}| \leq |P_{n,h_n}\psi_{n,\theta_0}| + |P_{n,h}\psi_{n,\theta_0}| \to 0. \qquad \square$$

Remark S3: *In Lemma S2, Conditions (i) and (ii) are required only in the case where $r \geq 1$ whilst Condition (iii) is required only in the case where $r = 0$.*

S2.3.2 Uniform results under a measure structure

Let π and π_n be as defined in Theorem 1 and Corollary 1 respectively.

Corollary S1: *Suppose the conditions of Theorem 1 hold, $(H, \mathcal{S}, Q)$ 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 $Q(H \setminus K) < \varepsilon$ and*

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

Proof. The pointwise convergence is the result of Theorem 1. π is measurable as the pointwise limit of measurable functions. By Egorov's theorem, $\pi_n(h) \to \pi(h)$ uniformly on a K satisfying the given requirements. □

Sufficient conditions for the measurability requirement in Corollary S1 are: H is a topological space, $\mathcal{S}$ its Borel σ-algebra and each π_n is continuous.

S2.4 Alternative representations of $\tilde{\mathcal{I}}$

Lemma S3: *Suppose Assumption 1 holds, B is a linear space and let $(\Omega, \mathcal{F}, \mathrm{P})$ be the probability space on which the Gaussian process Δ of Lemma 2 is defined. If $\mathcal{T} := \{\Delta(h) : h = (0, b) \in H\} \subset L_2(\mathrm{P})$ and $\tilde{\Delta}(e_i, 0) := \Pi\left[\Delta(e_i, 0)\middle|\mathcal{T}^{\perp}\right]$, then,*

$$\mathbb{E}\left[\Delta(e_i, 0)\tilde{\Delta}(e_j, 0)\right] = \mathbb{E}\left[\tilde{\Delta}(e_i, 0)\tilde{\Delta}(e_j, 0)\right] = \tilde{\mathcal{I}}_{ij}.$$

Proof. Define $Z : \overline{\mathbb{H}} \to L_2(\mathrm{P})$ as $Z[h] = \Delta(h)$. Z is a mean-zero linear Gaussian process with covariance kernel $K([h], [g]) = K(h, g) = \langle [h], [g]\rangle_K$ (see footnote 50). $Y := \operatorname{ran} Z \subset L_2(\mathrm{P})$ is a Hilbert space since for $[h], [g] \in \overline{\mathbb{H}}$, $\mathbb{E}[Z[h]Z[g]] = K([h], [g]) = \langle [h], [g]\rangle_K$, which along with the completeness of $\overline{\mathbb{H}}$ yields the closedness of $\operatorname{ran} Z$. Hence Z is a Hilbert space isomorphism from $\overline{\mathbb{H}}$ to Y. If $\pi_1' := \pi|_{\mathbb{H}}$,

$$\mathcal{T} = \{\Delta(h) : h = (0, b) \in H\} = \{Z[h] : h = (0, b) \in H\} = \{Z[h] : [h] \in \ker \pi_1'\}.$$

We next show that $\mathcal{T}^{\perp} = \{Z[h] : [h] \in (\ker \pi_1)^{\perp}\}$. For the first inclusion suppose that $Z[g] \in \mathcal{T}^{\perp}$. Then, for any $[h] \in \ker \pi_1'$,

$$\langle [g], [h]\rangle_K = \langle Z[g], Z[h]\rangle_{L_2(\mathrm{P})} = 0, \tag{S1}$$

and the inclusion follows by taking limits as $\ker \pi_1 = \operatorname{cl} \ker \pi_1'$ by Lemma S5. For the other inclusion note that a corollary of Lemma S5 is that $(\ker \pi_1)^{\perp} = (\ker \pi_1')^{\perp}$.

Hence, if $[g] \in (\ker \pi_1)^{\perp}$, for any $[h] \in \ker \pi_1'$ (S1) holds. Finally, let Q denote the orthogonal projection on $(\ker \pi_1)^{\perp} \subset \overline{\mathbb{H}}$ and R that on $\mathcal{T}^{\perp} \subset Y$. Then for $[h] \in \overline{\mathbb{H}}$, $R\Delta(h) = RZ[h] = ZQZ^*Z[h] = ZQZ^{-1}Z[h] = ZQ[h]$, since Z is a Hilbert space isomorphism. Hence $RZ[e_i, 0] = ZQ[e_i, 0]$ implying $\mathbb{E}\left[\tilde{\Delta}(e_i, 0)\tilde{\Delta}(e_j, 0)\right] = \left\langle \Pi^{\perp}[e_i, 0], \Pi^{\perp}[e_j, 0]\right\rangle_K = \tilde{\mathcal{I}}_{ij}$. $\square$

In the i.i.d. setting the efficient information matrix $\tilde{\mathcal{I}}$ coincides with the variance matrix of the efficient score function $\tilde{\ell} = \Pi[\dot{\ell}|\{Db : b \in B\}^{\perp}]$.

Lemma S4: *If Assumptions 1 and 5 hold and B is a linear space then $\tilde{\mathcal{I}} = \int \tilde{\ell}\tilde{\ell}' \,\mathrm{d}P$.*

Proof. For $h_1, h_2 \in H$, by the i.i.d. assumption and Lemma 2, $P_n[\Delta_n h_1 \Delta_n h_2] = P[Ah_1 Ah_2] = \mathrm{P}[\Delta h_1 \Delta h_2]$. $X := \operatorname{cl}\operatorname{ran} A \subset L_2(P)$ and $Y := \operatorname{cl}\operatorname{ran} \Delta \subset L_2(\mathrm{P})$ are Hilbert spaces when equipped with the inner products given by $(h_1, h_2) \mapsto P[Ah_1 Ah_2]$ and $(h_1, h_2) \mapsto \mathrm{P}[\Delta h_1 \Delta h_2]$ respectively. Define $U : \operatorname{ran} A \to \operatorname{ran} \Delta$ by $UAh := \Delta h$ for $h \in H$. U is a bounded, linear, surjective isometry and can be uniquely extended to a Hilbert space isomorphism $U : X \to Y$. Let $R := \Pi\left[\cdot|\mathcal{T}^{\perp}\right]$ ($\mathcal{T}^{\perp}$ as in Lemma S3) and $Q := \Pi\left[\cdot|\{Db : b \in B\}^{\perp}\right]$. Then $R\Delta h = RUAh = UQU^*UAh = UQAh$, which implies the conclusion as $e_i'\dot{\ell}_\gamma = A(e_i, 0)$. $\square$

S2.5 Most stringent tests

Here I consider most stringent tests; this delivers a similar message to the maximin analysis in the main text.[2] Let $\mathcal{C}$ be the class of all tests of level α for the hypothesis $\mathrm{K}_0 : h \in H_0$ against $\mathrm{K}_1 : h \in H_1$ in the experiment $\mathscr{E}$. Define $\pi^{\star}(h) := \sup_{\phi \in \mathcal{C}} P_h \phi$ for all $h \in H_1$. The *regret* of a test $\phi \in \mathcal{C}$ is

$$R(\phi) := \sup\left\{\pi^{\star}(h) - P_h\phi : h \in H_1\right\}. \tag{S2}$$

A test $\phi \in \mathcal{C}$ is called *most stringent at level α* if it minimises $R(\phi)$ over $\mathcal{C}$.

Theorem S1: *Suppose Assumptions 1 and 4 hold and $r = \operatorname{rank}(\tilde{\mathcal{I}}) \geq 1$. The most stringent level α test of $\mathrm{K}_0 : h \in H_0$ against $\mathrm{K}_1 : h \in H_1$ in $\mathscr{E}$ has power function*

$$\pi(h) = 1 - \mathrm{P}(\chi_r^2(a) \leq c_r), \qquad a = \tau'\tilde{\mathcal{I}}\tau, \quad h = (\tau, b) \in H. \tag{S3}$$

[2] The development here is based on Section 9, Chapter 11 in Le Cam (1986); in particular compare Theorem S1 with Corollary 2 of (Le Cam, 1986, Section 9, Chapter 11) which treats the case of a Gausian shift experiment indexed by a Euclidean space.

Proof. Denote by $\overline{\mathscr{G}}$ the Gaussian shift on $\overline{\mathbb{H}}$ and $\tilde{R}$ the regret

$$\tilde{R}(\phi) := \sup\left\{\tilde{\pi}^\star([h]) - G_{[h]}\phi : [h] \in \mathbb{H} \setminus \ker \pi_1\right\}, \quad \tilde{\pi}^\star([h]) := \sup\left\{G_{[h]}\varphi : \varphi \in \tilde{\mathcal{C}}\right\},$$

where $\tilde{\mathcal{C}}$ is the class of level–α tests of $\ker \pi_1$ against $\mathbb{H} \setminus \ker \pi_1$ in $\mathscr{G}$. The Neyman – Pearson test, $\psi^\star$, of $[g] \in \ker \pi_1$ against $[g]+[h]$, with $[h] \in [\ker \pi_1]^\perp$ rejects when

$$\exp\left(Z[g+h] - Z[g] - \frac{1}{2}\|[g+h]\|_K^2 + \frac{1}{2}\|[g]\|_K^2\right) = \exp\left(Z\Pi^\perp[h] - \frac{1}{2}\|\Pi^\perp[h]\|_K^2\right)$$

exceeds a k chosen such that the test is of level α, for Z the central process of $\overline{\mathscr{G}}$. k does not depend on the $[g] \in \ker \pi_1$ and the power of this test depends only on $\|\Pi^\perp[h]\|_K^2 = \tau'\tilde{\mathcal{I}}\tau$ where $\pi_1[h] = [\tau]$. Now let $[h] \in \mathbb{H} \setminus \ker \pi_1$ and consider testing $K_1 : [h]$ against $K_0 : [h] \in \ker \pi_1$. One has $[h] = [g] + \Pi^\perp[h]$ where $[g] = \Pi[h] \in \ker \pi_1$. By the preceding observations, $\psi^\star$ is a most powerful level-α test for this hypothesis.[3] Thus $\psi^\star \in \tilde{\mathcal{C}}$ and

$$\tilde{\pi}^\star([h]) := \sup_{\phi \in \tilde{\mathcal{C}}} G_{[h]}\phi = G_{[h]}\psi^\star. \tag{S4}$$

For $i = 1, \ldots, d_\theta$, let $u_i := \Pi^\perp[(e_i, 0)]$ and let $X := (Zu_1, \ldots, Zu_{d_\theta})'$. Let ψ be the test which rejects when $(X'\tilde{\mathcal{I}}^\dagger X)^2 > c_r$, for c_r the $1-\alpha$ quantile of the χ_r^2. By Theorem 69.10 in Strasser (1985) and Theorem 9.2.3 in Rao and Mitra (1971), $G_{[h]}\psi = 1 - \mathrm{P}(\chi_r^2(a) \leq c_r)$, $a = \tau'\tilde{\mathcal{I}}\tau = \|\Pi^\perp[h]\|_K^2$ where $[\tau] = \pi_1[h]$. $G_{[h]}\psi^\star - G_{[h]}\psi$ depends only on $\|\Pi^\perp[h]\|_K^2$. Fix a $\varepsilon > 0$ and suppose that for some $\phi \in \tilde{\mathcal{C}}$, $\tilde{R}(\phi) < \tilde{R}(\psi) - 2\varepsilon$. There is an $a > 0$ such that

$$\sup\left\{G_{[g+h]}\psi^\star - G_{[g+h]}\psi : [h] \in [\ker \pi_1]^\perp, \|[h]\|_K^2 = a\right\} \geq \tilde{R}(\psi) - \varepsilon, \text{ for all } [g] \in \ker \pi_1.$$

In consequence, for all $[g] \in \ker \pi_1$, all $[h] \in S_a := \{[h] \in [\ker \pi_1]^\perp : \|[h]\|_K^2 = a\}$,

$$G_{[g+h]}\psi^\star - G_{[g+h]}\phi \leq \tilde{R}(\psi) - 2\varepsilon \leq G_{[g+h]}\psi^\star - G_{[g+h]}\psi - \varepsilon,$$

which produces a contradiction to Theorem 30.2 in Strasser (1985):

$$\inf_{[h] \in S_a} G_{[h]}\phi \geq \inf_{[h] \in S_a} G_{[h]}\psi + \varepsilon = 1 - \mathrm{P}(\chi_r^2(a) \leq c_r) + \varepsilon.$$

[3] Suppose there were another level α test ϕ of K_0 against K_1, with strictly higher power than $\psi^\star$. Then, this would also be a test of level α for $[g]$ against $\Pi^\perp[h]$. But this would contradict the Neyman – Pearson Lemma (e.g. Lehmann and Romano, 2005, Theorem 3.2.1).

To complete the proof, it suffices to show that $\varphi : \Omega \to [0,1]$ is in $\mathcal{C}$ if and only if $\varphi \in \tilde{\mathcal{C}}$ and $R(\varphi) = \tilde{R}(\varphi)$. The first part follows from $h \in H_0$ if and only if $[h] \in \ker \pi_1$ and Proposition 3. For the second part, (S4), Proposition 3 and the first part together imply that $\tilde{\pi}^\star([h]) = \pi^\star(h)$ for all $h \in H$. Therefore,

$$\tilde{\pi}^\star([h]) - G_{[h]}\varphi = \tilde{\pi}^\star(\pi_V(h)) - G_{\pi_V(h)}\varphi = \pi^\star(h) - P_h\varphi, \quad h \in H.$$

As $h \in H_1 \iff [h] \in \mathbb{H} \setminus \ker \pi_1$, $\tilde{R}(\varphi) = \sup\{\pi^\star(h) - P_h\varphi : h \in H_1\} = R(\varphi)$. $\square$

The first part of Corollary 3 provides conditions under which (S3) is the asymptotic power function of ψ_{n,θ_0} under $P_{n,h}$. The following Proposition demonstrates that if $\pi_n : H \to [0,1]$ is a sequence of power functions corresponding to tests in the experiments $\mathscr{E}_n$ of asymptotic size α, then each cluster point of π_n corresponds to a test ϕ in the limit experiment $\mathscr{E}$ whose regret is bounded below by that of the most stringent test, ψ.[4]

Proposition S2: *Suppose Assumptions 1 and 4 hold and that* $r = \operatorname{rank}(\tilde{\mathcal{I}}) \geq 1$. *Let* $\phi_n : \mathcal{W}_n \to [0,1]$ *be a sequence of tests such that for each* $h = (0,b) \in H$,

$$\limsup_{n\to\infty} P_{n,h}\phi_n \leq \alpha. \tag{S5}$$

For each $h \in H$, *let* $\pi_n(h) := P_{n,h}\phi_n$. *If* π *is a cluster point of* π_n *(with respect to the topology of pointwise convergence on* $[0,1]^H$*), then* π *is the power function of a test* ϕ *in* $\mathscr{E}$ *and* $R(\phi) \geq R(\psi)$.

Proof. By (S5), Proposition 2 and Theorem 7.1 in van der Vaart (1991) there is a level α test ϕ in $\mathscr{E}$ with $P_h\phi = \pi(h)$. Apply Theorem S1. $\square$

S3 Technicalities

Lemma S5: *Suppose Assumption 1 holds and* B *is a linear space. Let* π_1' *denote the restriction of* π_1 *to* $\mathbb{H}$. *Then, the closure of* $\ker \pi_1'$ *in* $\overline{\mathbb{H}}$ *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 $\ker \pi_1 = \{[h] \in \overline{\mathbb{H}} : [h] = [0,b]\} \subset \operatorname{cl} \ker \pi_1' = \operatorname{cl}\{[h] \in \mathbb{H} : [h] = [0,b]\}$. Let $[h] = [0,b] \in \ker \pi_1$. There is a sequence $\mathbb{H} \ni [h_n] = [t_n, b_n] \to [h]$. Using (44)

$$\|[h_n] - [h]\|_K = t_n'\tilde{\mathcal{I}}t_n + \|t_n'\boldsymbol{e} + [0,b_n] - [0,b]\|_K, \quad \boldsymbol{e} := (\Pi[e_1,0],\ldots,\Pi[e_{d_\theta},0])'$$

[4]The space of functions from H to $[0,1]$ is equipped with the topology of pointwise convergence.

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 $[0, \check{b}_{j,n}] \in \mathbb{H}$ such that $\left\|[0, \check{b}_{j,n}] - \Pi[e_j, 0]\right\|_K \leq 1/(n|t_{n,j}|)$. Putting $[0, \tilde{b}_n] \coloneqq t_n' \check{\boldsymbol{e}}_n + [0, b_n]$,

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

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

Lemma S6: *Suppose that for $h_n, g \in H$, $P_{n,g} \triangleleft P_n$ and $L_n(h_n) - L_n(g) = o_{P_n}(1)$. Then $d_{TV}(P_{n,h_n}, P_{n,g}) \to 0$.*

Proof. By the contiguity $L_n(h_n) - L_n(g) = o_{P_{n,g}}(1)$. Apply Lemma S3.3 in Hoesch, Lee, and Mesters (2024). □

Corollary S2: *Suppose that Assumption 1 holds and H is a linear space equipped with the semi-norm $\|\cdot\|_K$. If $h, g \in H$ satisfy $\|h - g\|_K = 0$, $d_{TV}(P_{n,h}, P_{n,g}) \to 0$.*

Proof. By Assumption 1, the reverse triangle inequality and $\sigma(h - g) = \|h - g\|_K$ we have that $L_n(h) - L_n(g) = o_{P_n}(1)$. Apply Lemma S6 with $h_n = h$. □

Lemma S7: *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*

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

Proof. Suppose $b \in B$. Then $\mathbb{E}[b(U, X)UZ] = \mathbb{E}\left[\mathbb{E}[b(U, X)U|X]Z\right] = 0$. Conversely suppose $b \in L_2(\Omega, \sigma(U, X), \mathrm{P})$ and $\mathbb{E}[b(U, X)UZ] = 0$ for Z any bounded $\sigma(X)$ – measurable random variable. By Proposition A.3.1 in Bickel et al. (1998), $\mathbb{E}[b(U, X)U|X] = 0$ a.s. whence $b \in \operatorname{cl} B$ by Lemma C.7 in Newey (1991) □

Theorem S2: *Let H be a Hilbert space. Let $h_n, h \in H$, and L_n, L closed (proper) linear subspaces of H. Let $g_n \coloneqq \Pi(h_n|L_n)$ and $g \coloneqq \Pi(h|L)$. If (i) $h_n \to h$ and (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 be the orthogonal projection onto L_n and Π that onto L. First suppose $h_n = h$ $(n \in \mathbb{N})$. As $(g_n)_{n \in \mathbb{N}}$ is bounded, any subsequence contains a weakly convergent subsequence, say $g_{n_k} \rightharpoonup g^\star$. By self-adjointness and idempotency (SAI)

$$\langle g_{n_k}, g_{n_k} \rangle = \langle \Pi_{n_k} h, \Pi_{n_k} h \rangle = \langle h, \Pi_{n_k} h \rangle \to \langle h, g^\star \rangle. \tag{S6}$$

Let $f \in L$. By hypothesis there are $(f_n)_{n\in\mathbb{N}}$ with $f_n \to f$ and $f_n \in L_n$ for $n \geq N_1$. So $f_{n_k} \to f$ and $f_{n_k} \in L_{n_k}$ for $k \geq K_1$. Since $h - \Pi_{n_k} h \rightharpoonup h - g^\star$, 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, $\langle h - g^\star \, , \, f\rangle = \lim_{k\to\infty} \langle h - g_{n_k} \, , \, f_{n_k}\rangle = 0$. Hence $g^\star = \Pi h = g$. By SAI of Π and (S6), $\lim_{k\to\infty} \langle g_{n_k} \, , \, g_{n_k}\rangle = \langle h \, , \, \Pi h\rangle = \langle \Pi h \, , \, \Pi h\rangle = \langle g \, , \, g\rangle$ and hence $g_{n_k} \to g$ by the Radon – Riesz Theorem. As the initial subsequence was arbitrary, $g_n \to g$. To complete the proof, for $h_n \to h$ an arbitrary convergent sequence, $\|g_n - g\| \leq \|h_n - h\| + \|\Pi_n h - \Pi h\|$. The first RHS term is $o(1)$ by assumption; the second by the case with $h_n = h$. □

Theorem S3: *Let H be a linear space and $B \subset H$ a linear subspace of H. Suppose that G_n is a Gaussian process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with index set H and covariance kernel K_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 let $\mathscr{G}_n := \sigma(\{\mathsf{G}_n f : f \in B\})$ and $\mathscr{G}_n := \sigma(\{\mathsf{G} f : f \in B\})$. Then,*

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

Proof. $\mathrm{cl}\{\mathsf{G}b : b \in B\}$ is a separable Hilbert space and so has an orthonormal basis, which may be taken to be formed of $\mathsf{G}b_j$, $(b_j)_{j\in\mathbb{N}} \subset B$. Let $\mathcal{G}_n := \sigma(\{\mathsf{G}_n b_i : i \in \mathbb{N}\})$, $\mathcal{G} := \sigma(\{\mathsf{G}b_i : i \in \mathbb{N}\})$, $B_m := (b_1, \ldots, b_m)$, $\mathcal{G}_n^m := \sigma(\{\mathsf{G}_n b : b \in B_m\})$, $\mathcal{G}^m := \sigma(\{\mathsf{G}b : b \in B_m\})$, $X_n^m := (\mathsf{G}_n h, \mathbb{E}[\mathsf{G}_n g|\mathcal{G}_n^m])$ and $X^m := (\mathsf{G}h, \mathbb{E}[\mathsf{G}g|\mathcal{G}^m])$. Now let $Z_n^m := (\mathsf{G}_n h, \mathsf{G}_n g, \mathsf{G}_n b_1, \ldots, \mathsf{G}_n b_m)' \sim \mathcal{N}(0, \Sigma_n^m)$ and $Z^m := (\mathsf{G}_n h, \mathsf{G} g, \mathsf{G}_n b_1, \ldots, \mathsf{G}_n b_m)' \sim \mathcal{N}(0, \Sigma^m)$. Partition Σ^m conformally with $Z_1^m = \mathsf{G}h$, $Z_2^m = \mathsf{G}g$ and $Z_3^m = (\mathsf{G}_n b_1, \ldots, \mathsf{G}_n b_m)'$ and similarly for Σ_n^m, Z_n^m. Then

$$\begin{aligned}
X_n^m &= (\mathsf{G}_n h, \, \mathbb{E}[\mathsf{G}_n g|\mathcal{G}_n^m]) = \left(Z_{n,1}^m, \, Z_{n,2}^m - [\Sigma_n^m]_{2,3}[\Sigma_n^m]_{3,3}^{-1} Z_{n,3}^m\right) \\
X^m &= (\mathsf{G}h, \, \mathbb{E}[\mathsf{G}g|\mathcal{G}^m]) = \left(Z_1^m, \, Z_2^m - [\Sigma^m]_{2,3}[\Sigma^m]_{3,3}^{-1} Z_3^m\right) .
\end{aligned}$$

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 the inverses in the preceding displays exist for all sufficiently large n since $\{\mathsf{G}b_i : i \in \mathbb{N}\}$ 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,

$$X_n^m \rightsquigarrow X^m. \tag{S7}$$

Let Π^m be the orthogonal projection onto $S_m := \mathrm{Span}\{\mathsf{G}b : b \in B_m\}$. 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). $S_m \subset S_{m+1}$ and $S := \mathrm{cl}\{\mathsf{G}b : b \in B\} = \mathrm{cl}\cup_{m\in\mathbb{N}} S_m$, so by Theorem S2 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

$$X^m \rightsquigarrow X. \tag{S8}$$

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 $Y := \mathbb{E}[\mathsf{G}h|\mathcal{G}]$. As $Y_n \in \mathrm{cl}\{\mathsf{G}_n b : b \in B\}$ and $Y_n^m \in \{\mathsf{G}_n b : b \in B_m\}$ (Janson, 1997, Theorem 9.1), $Y_n - Y_n^m \sim \mathcal{N}(0, \sigma_{n,m}^2)$ where $\sigma_{n,m}^2 := \mathrm{Var}(Y_n - Y_n^m)$. As $Y_n = \mathbb{E}[\mathsf{G}_n h|\mathscr{G}_n]$ and $Y_n^m = \mathbb{E}[\mathsf{G}_n h|\mathscr{G}_n^m]$ (Janson, 1997, Theorem 9.1),

$$\mathbb{P}\left(\|X_n - X_n^m\| > \varepsilon\right) = \mathbb{P}\left(|Y_n - Y_n^m| > \varepsilon\right) \leq C\exp\left(-\frac{\varepsilon^2}{\sigma_{n,m}^2}\right). \tag{S9}$$

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 the separable metric space $(\mathbb{R}^\infty, d)$ 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 Skorohod's representation Theorem (e.g. Billingsley, 1999, Theorem 6.7) there are random elements $\tilde{F}_n$ and $\tilde{F}$ defined on a common probability space such that $\tilde{F}_n \to \tilde{F}$ surely, $\mathcal{L}(\tilde{F}) = \mathcal{L}(F)$ and $\mathcal{L}(\tilde{F}_n) = \mathcal{L}(F_n)$. Thus $\tilde{F}_n$ and $\tilde{F}$ are Gaussian processes. As $\mathrm{Cov}(\tilde{F}_{n,i}, \tilde{F}_{n,j}) = K_n(f_{i-1}, f_{j-1}) \to K(f_{i-1}, f_{j-1}) = \mathrm{Cov}(\tilde{F}_i, \tilde{F}_j)$ each $(\tilde{F}_{n,i})_{n\in\mathbb{N}}$ is uniformly square integrable. As $(\mathbb{R}^\infty, d)$ has the topology of pointwise convergence 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

$$\begin{aligned}
&\tilde{Y}_n^m := \mathbb{E}[\tilde{F}_{n,1}|\{\tilde{F}_{n,i} : 2 \leq i \leq m\}] \sim Y_n^m, \quad && \tilde{Y}_n := \mathbb{E}[\tilde{F}_{n,1}|\{\tilde{F}_{n,i} : i \in \mathbb{N}, i \neq 1\}] \sim Y_n,\\
&\tilde{Y}^m := \mathbb{E}[\tilde{F}_1|\{\tilde{F}_i : 2 \leq i \leq m\}] \sim Y^m, && \tilde{Y} := \mathbb{E}[\tilde{F}_1|\{\tilde{F}_i : i \in \mathbb{N}, i \neq 1\}] \sim Y.
\end{aligned}$$

Define $\tilde{S}_n^m := \mathrm{Span}\{\tilde{F}_{n,i} : 2 \leq i \leq m\}$, $\tilde{S}_n := \mathrm{cl}\,\mathrm{Span}\{\tilde{F}_{n,i} : i \in \mathbb{N}, i \neq 1\}$, $\tilde{S}^m := \mathrm{Span}\{\tilde{F}_i : 2 \leq i \leq m\}$ and $\tilde{S} := \mathrm{cl}\,\mathrm{Span}\{\tilde{F}_i : i \in \mathbb{N}, i \neq 1\}$. Then $\tilde{Y}_n^m = \Pi[\tilde{F}_{n,1}|S_n^m]$, $\tilde{Y}_n = \Pi[\tilde{F}_{n,1}|S_n]$, $\tilde{Y}^m = \Pi[\tilde{F}_1|S^m]$, $\tilde{Y} = \Pi[\tilde{F}_1|S]$ by Theorem 9.1 in Janson (1997). We will apply Theorem S2 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. By Theorem S2,

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

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

To see that $\lim_{m\to\infty} \sigma_m^2 = 0$ 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 (with m in place of n) hold. Hence $Y^m \xrightarrow{L_2} Y$ and so $\sigma_m^2 = \operatorname{Var}(Y - Y^m) \to 0$. In conjunction with (S9),

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\|X_n - X_n^m\| > \varepsilon\right) \leq \lim_{m\to\infty} \limsup_{n\to\infty} C \exp\left(-\frac{\varepsilon^2}{\sigma_{n,m}^2}\right) = 0. \tag{S10}$$

The result now follows from Theorem 3.2 in Billingsley (1999).[5] □

Lemma S8: *Let $(m_n)_{n\in\mathbb{N}}$ be an increasing sequence of natural numbers such that $m_n \leq n$, $(Y_{n,i})_{n\in\mathbb{N},1\leq i\leq m_n}$ a triangular array of random vectors and $\mathcal{C}_n$ a collection of random variables. Suppose that with probability approaching one either*

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

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

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

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

S4 Additional details for the examples

S4.1 Single index model

S4.1.1 Proofs of results in the main text

Proof of Proposition 5. As is easy to verify, each component of g_n belongs to $L_2^0(P_n)$. For any $b \in B$, $\mathbb{E}[\epsilon b_2(\epsilon, X)|X] = 0$ by (28). Plugging in for Db and using this allows the conclusion that $\mathbb{E}\left[g(W)[Db](W)\right] = 0$. Apply Lemma 5. □

Proof of Proposition 6. For part (i) of Assumption 3 note that for some $a_j \in$

[5]Equations (S7), (S8) and (S10) verify the required hypotheses.

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

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

We verify that one of (i) or (ii) of Lemma S8 is satisfied with $Y_{n,i} = R_{l,n,i}$ for $i = 1, \ldots, m_n := \lfloor n/2 \rfloor$ or $i = m_n + 1, \ldots, n$. Suppose that $1 \le i \le m_n$ and let $\mathcal{C}_n = \mathcal{C}_{n,2}$ (the case with $m_n + 1 \le i \le n$ and $\mathcal{C}_n = \mathcal{C}_{n,1}$ is analogous). Each $\hat{Z}_{k,n,i}(V_{\theta,i})$ is $\sigma(V_{\theta,i}, \mathcal{C}_n)$ – measurable for $k = 0,1,2,3,4$. $f(V_{\theta,i})$, $f'(V_{\theta,i})$, $\omega(X_i)$ and $X_{2,i} - Z_0(V_{\theta,i})$ are bounded uniformly in i and there are events E_n with $P_{n,0}E_n \to 1$ on which $\mathsf{R}_{l,n,i} \le r_n$, $\hat{f}_{n,i}(V_{\theta,i})$, $\widehat{f'}_{n,i}(V_{\theta,i})$, $\hat{Z}_{1,n,i}(V_{\theta,i})$ are bounded above uniformly in i and $\hat{Z}_{2,n,i}(V_{\theta,i})$ is bounded above and below uniformly in i, for all large enough $n \in \mathbb{N}$. On these sets,

$$
\mathbb{E}\left[\left\|\hat{Z}_{0,n,i}(V_{\theta,i}) - Z_0(V_{\theta,i})\right\|^2 \middle| \mathcal{C}_n\right] \lesssim r_n^2. \tag{S11}
$$

For $l = 1,2,3$, the first part of condition (ii) follows by the law of iterated expectations and independence since $\mathbb{E}[\omega(X_i)(X_{2,i} - Z_0(V_{\theta,i}))|V_{\theta,i}] = 0$ ($l = 1$) and $\mathbb{E}[\epsilon_i|X_i] = 0$ ($l = 2,3$). The second part follows by the uniform boundedness noted above, $R_{l,n,i} \le r_n$ on E_n along with equations (30) and (S11).

$l = 4$: By the uniform boundedness and the 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 RHS 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 and the Cauchy – Schwarz inequality, $\mathbb{E}[\|R_{4,n,i}\||\mathcal{C}_n] \lesssim \mathbb{E}\left[\left|\hat{f}_{n,i}(V_{\theta,i}) - f(V_{\theta,i})\right|\left\|\hat{Z}_{0,n,i}(V_{\theta,i}) - Z_0(V_{\theta,i})\right\|\middle|\mathcal{C}_n\right]$. For a $C > 0$, the RHS is upper bounded by $Cr_n\mathsf{R}_{3,n,i} = o(n^{-1/2})$ on E_n by (S11).

For parts (ii) and (iii) of Asssumption 3, we show that $\|\check{V}_{n,\theta} - V\| = o_{P_{n,0}}(\nu_n)$, which suffices by Proposition S1 of Lee and Mesters (2024). For $\check{V} := \mathbb{P}_n gg'$,

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

$\mathbb{E}(g_l g_k)^2 < \infty$ by $\mathbb{E}[\epsilon^4] < \infty$ and Assumption 8. Hence $\frac{1}{\sqrt{n}}\mathbb{G}_n[gg'] = O_{P_{n,0}}(n^{-1/2})$

by the CLT. For the other term,

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

For $l=1,2,3$ we established that if $1\le i\le m_n$ and $\mathcal{C}_n=\mathcal{C}_{n,2}$ then $\mathbb{E}[R^2_{l,n,i,k}|\mathcal{C}_n]\lesssim r_n^2$ on E_n. We show this also holds for $l=4,5$. (The case with $m_n+1\le i\le n$ with $\mathcal{C}_n=\mathcal{C}_{n,1}$ is once again analogous). For $l\in\{4,5\}$, $\mathbb{E}\left[R^2_{l,n,i,k}|\mathcal{C}_n\right]\lesssim \mathbb{E}\left[\left(\hat{f}_{n,i}(V_{\theta,i})-f(V_{\theta,i})\right)^2\middle|\mathcal{C}_n\right]$, by the uniform boundedness (for all large enough n) and the RHS is bounded above by r_n^2 on E_n. By Markov's inequality, $\frac{1}{n}\left[\sum_{i=1}^{m_n}R^2_{l,n,i,k}+\sum_{i=m_n+1}^{n}R^2_{l,n,i,k}\right]=O_{P^n}(r_n^2)$ for $l=1,\ldots,5$ hence the same is true of $\frac{1}{n}\sum_{i=1}^{n}\|\hat{g}_{n,\theta,i,}-g(W_i)\|^2$. Therefore, $\|\check{V}_{n,\theta}-\check{V}\|_2=O_{P^n}(r_n)$ as $\left\|\check{V}_{n,\theta}-\check{V}\right\|_2^2$ is bounded above by a multiple of

$$\frac{1}{n}\sum_{i=1}^{n}\|\hat{g}_{n,\theta,i}\|^2\,\frac{1}{n}\sum_{i=1}^{n}\|\hat{g}_{n,\theta,i}-g(W_i)\|^2+\frac{1}{n}\sum_{i=1}^{n}\|\hat{g}_{n,\theta,i}-g(W_i)\|^2\,\frac{1}{n}\sum_{i=1}^{n}\|g(W_i)\|^2.$$

□

S4.1.2 The LAN condition

Here I provide examples of local perturbations $P_{n,h}$ and lower level conditions under which the LAN condition in Assumption 7 holds. Let φ_n be as in equation (3) with $B_1 := C_b^1(\mathscr{D})$ and B_2 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 (28) holds.

Proposition S3: *Suppose Assumption 6 holds, $\mathcal{W}_n=\prod_{i=1}^{n}\mathbb{R}^{1+K}$, $e\mapsto\sqrt{\zeta(e,x)}\in\mathcal{C}^1$, and $p_{n,h}=p^n_{\gamma+\varphi_n(h)}$ with p_γ as in (2). Then Assumption 7 holds.*

Proof. Define $\gamma_t(h):=\gamma+t(\tau,b_1,b_2\zeta)$ for $h=(\tau,b_1,b_2)$ and $t\in[0,\infty)$. It is easy to verify that $P_{\gamma_t(h)}\in\{P_\gamma:\gamma\in\Gamma\}$ for all small enough t. This ensures the required domination in Assumption 5 given Assumption 6. Next note that $t\mapsto\sqrt{p_{\gamma_t(h)}}$ is continuously differentiable everywhere since it is a composition of continuously differentiable functions for t small enough that $(1+tb_2)$ is bounded away from zero. This ensures that $q_t(W):=\frac{\mathrm{d}\log p_{\gamma_s(h)(W)}}{\mathrm{d}s}|_{s=t}$ is defined for small enough t. Writing $v_t := V_{\theta+t\tau}$ and $e_t := Y - f(v_t) - tb_1(v_t)$ this has the form

$$q_t(W) := -\phi(e_t, X)[f'(v_t)X_2'\tau + tb_1'(v_t)X_2'\tau + b_1(v_t)] \\ + \frac{b_2(e_t, X) - tb_2'(e_t, X)[f'(v_t)X_2'\tau + tb_1'(v_t)X_2'\tau + b_1(v_t)]}{1 + tb_2(e_t, X)}, \tag{S12}$$

which is a composition of continuous functions. By boundedness of f', b_1, b_1', b_2, b_2', $(1+tb_2)^{-1}$ and equation (26), $\int |q_t(W)|^{2+\rho} \,\mathrm{d}P_{\gamma_t(h)} \le C\mathbb{E}\left[(|\phi(\epsilon, X)|^{2+\rho} + 1) \|X\|^{2+\rho}\right] < \infty$ for a positive constant $C < \infty$ and a $\rho > 0$. 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 $\int q_{t_n}(W)^2 p_{\gamma_{t_n}(h)}(W) \,\mathrm{d}\lambda \to \int q_t(W)^2 p_{\gamma_t(h)}(W) \,\mathrm{d}\lambda$. Applying Lemma 1.8 in van der Vaart (2002) demonstrates that equation (24) holds, with Ah as in (25). Lemma 1.7 of van der Vaart (2002) ensures that $Ah \in L_2^0(P)$. The form of Ah reveals that it is a linear map on H. It is bounded:

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

where $C_1, C_2 \in (0, \infty)$ are positive constants. Apply Lemma 4. □

S4.2 IV model with non-parametric first stage

S4.2.1 Proofs of results in the main text

Proof of Lemma 6. $J(Z)$ is nonsingular by (37). By Proposition 2.8.4 in Bernstein (2009), $J(Z)^{-1}_{1,1} = \mathbb{E}[\epsilon^2|Z]^{-1}$ exists and is positive. Define $dotl(W) := \left(\dot{l}_1(W)', \dot{l}_2(W)'\right)' = -\phi_1(\xi)(X', Z_1')'$, $[D_1 b_1](W) := -\phi_2(\xi)'b_1(Z)$ and $[D_2 b_2](W) := b_2(\xi)$, where $\xi = (Y - X'\theta - Z_1'\beta, X - \pi(Z), Z)$. By Lemma S7 and Proposition A.3.5 in Bickel et al. (1998), with $\mathcal{T}_2 := \{[D_2 b_2](W) : b_2\}$,

$$\begin{aligned}
\breve{l}_\gamma(W) &:= \Pi\left[\dot{l}(W)|\mathcal{T}_2^\perp\right] = \mathbb{E}\left[-\phi_1(\xi)[X', Z_1']'U'|Z\right]\mathbb{E}\left[UU'|Z\right]^{-1}U, \\
[\breve{D}_1 b_1](W) &:= \Pi\left[[\dot{D}_1 b_1](W)|\mathcal{T}_2^\perp\right] = \mathbb{E}\left[-b_1(Z)'\phi_2(\xi)U'|Z\right]\mathbb{E}\left[UU'|Z\right]^{-1}U.
\end{aligned}$$

Let $K := d_\beta$, and evaluating the conditional expectations using (37) we obtain:

$$\left.\begin{aligned}
\breve{l}_\gamma(W) &= \begin{bmatrix}\pi(Z)\\ Z_1\end{bmatrix}\begin{bmatrix}1 & 0_K'\end{bmatrix} J(Z)^{-1}U = \pi(Z)E_1\ , \\
[\breve{D}_{\gamma,1} b_1](W) &= b_1(Z)'\begin{bmatrix}0_K & I_K\end{bmatrix} J(Z)^{-1}U = b_1(Z)'E_2\ .
\end{aligned}\right\}, \quad E := J(Z)^{-1}U.$$

The projection of $\breve{l}(W)$ onto the orthocomplement of $\{\breve{D}_1 b_1 : b_1 \in B_1\}$ is equal to $\tilde{l}(W) := \Pi\left[\dot{l}(W)|\{[D_1 b_1](W) + D_2 b_2(W) : b \in B\}^{\perp}\right]$ by Proposition A.2.4 in Bickel et al. (1998). The components of $[\pi(Z)', Z_1']' \, \mathbb{E}[E_1 E_2'|Z]\mathbb{E}[E_2 E_2'|Z]^{-1} E_2$ belong to $\mathrm{cl}\{b_1(Z)'V_2 : b_1 \in B_1\}$ as B_1 is dense in L_2 and by iterated expectations

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

Hence $\tilde{l}(W)$, the efficient score for (θ, β), has the form

$$\tilde{l}(W) = \begin{bmatrix} \tilde{l}_1(W) \\ \tilde{l}_2(W) \end{bmatrix} = \begin{bmatrix} \pi(Z) \\ Z_1 \end{bmatrix} \left[E_1 - \mathbb{E}[E_1 E_2'|Z]\mathbb{E}[E_2 E_2'|Z]^{-1} E_2\right]. \tag{S13}$$

$\tilde{\ell}(W) = \tilde{l}_1(W) - \mathbb{E}[\tilde{l}_1(W)\tilde{l}_2(W)']\mathbb{E}[\tilde{l}_2(W)\tilde{l}_2(W)']^{-1}\tilde{l}_2(W)$ by Example A.2.1 in Bickel et al. (1998). To calculate this note that with $Q(Z) := J(Z)^{-1} = \mathbb{E}[EE'|Z]$

$$E_1 - Q(Z)_{1,2} Q(Z)_{2,2}^{-1} E_2 = \left[Q(Z)_{1,1} - Q(Z)_{1,2} Q(Z)_{2,2}^{-1} Q(Z)_{2,1}\right] U_1 = \mathbb{E}[\epsilon^2|Z]^{-1}\epsilon$$

from which the result follows by direct calculation. □

Proof of Lemma 7. The second claim follows from the expressions in (38) & (39). For the first, by Assumption 9 and (37), $\mathbb{E}[\|g(W)\|^2] < \infty$. By (35)

$$\mathbb{E}\left[g(W) b_2(\xi)\right] = \mathbb{E}[\epsilon^2]^{-1}\mathbb{E}\left[(\pi(Z) - M Z_1)\mathbb{E}\left[\epsilon b_2(U, Z)|Z\right]\right] = 0,$$

where $M := \mathbb{E}[X Z_1']\mathbb{E}[Z_1 Z_1']^{-1}$ and $\xi = (Y - X'\theta - Z_1'\beta, X - \pi(Z), Z)$. By (37)

$$\mathbb{E}\left[g(W)\phi_2(\xi)' b_1(Z)\right] = \mathbb{E}[\epsilon^2]^{-1}\mathbb{E}\left[(\pi(Z) - M Z_1)\mathbb{E}\left[\epsilon\phi_2(\epsilon, v, Z)'|Z\right] b_1(Z)\right] = 0.$$

Lastly also $\mathbb{E}\left[g(W)\phi_1(\xi) b_0' Z_1\right] = 0$ as by (37) and $\mathbb{E}[v|Z] = 0$,

$$\mathbb{E}\left[g(W)\phi_1(\xi) b_0' Z_1\right] = -\mathbb{E}[\epsilon^2]^{-1}\left[\mathbb{E}[\pi(Z) Z_1'] - \mathbb{E}[\pi(Z) Z_1']\mathbb{E}[Z_1 Z_1']^{-1}\mathbb{E}[Z_1 Z_1']\right] b_0. \quad \square$$

Proof of Proposition 7. Assumptions 9, 10, equation (37) and Lemma 7 verify the conditions required to apply Lemma 5. □

Proof of Proposition 8. Let $\beta_n = \beta + b_{n,0}/\sqrt{n}$ with $b_{n,0} \to b_0 \in \mathbb{R}^{d_\beta}$. Let $\breve{\epsilon}_{n,i}$, $\breve{g}_{n,\theta,i}$, $\breve{V}_{n,\theta}$, $\breve{\Lambda}_{n,\theta}$ and $\breve{r}_{n,\theta}$ be formed analogously to $\hat{\epsilon}_{n,i}$, $\hat{g}_{n,\theta,i}$, $\hat{V}_{n,\theta}$, $\hat{\Lambda}_{n,\theta}$ and $\hat{r}_{n,\theta}$ with β_n in place of $\hat{\beta}_n$. As $\hat{\beta}_n \in \mathscr{S}_n$, by Lemma S3.1 in Hoesch et al. (2024) it suffices to show that Assumption 3 holds for $\breve{g}_{n,\theta} := \frac{1}{\sqrt{n}}\sum_{i=1}^n \breve{g}_{n,\theta,i}$, $\breve{\Lambda}_{n,\theta}$ and $\breve{r}_{n,\theta}$. For Assumption 3 part (i), by Lemma S9, $\frac{1}{\sqrt{n}}\sum_{i=1}^n \left[g_{n,\theta,i} - g(W_i)\right] = \sum_{l=1}^4 \mathsf{R}_{n,l} = o_{P_{n,0}}(1)$. For

Assumption 3 parts (ii) and (iii), note that

$$\check{V}_{n,\theta} - V = \frac{1}{n}\sum_{i=1}^{n}\left[g_{n,\theta,i}g_{n,\theta,i}' - g(W_i)g(W_i)'\right] + \frac{1}{\sqrt{n}}\mathbb{G}_n g g'. \tag{S14}$$

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

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

As $\mathbb{E}\|g(W_i)\|^2 < \infty$ under Assumption 9 , $\frac{1}{n}\sum_{i=1}^{n}\|g(W_i)\|^2 = O_{P_{n,0}}(1)$. By Lemma S9 $\frac{1}{n}\sum_{i=1}^{n}\|g_{n,\theta,i} - g(W_i)\|^2 \lesssim \sum_{l=1}^{4}\mathsf{S}_{n,l} = O_{P_{n,0}}(\delta_n^2 + n^{-1})$, hence by the preceding display, the first RHS term in (S14) is $O_{P_{n,0}}(\delta_n^2 + n^{-1})$. By Assumption 11 $\mathbb{E}\|g(W)\|^4 < \infty$. Hence $\mathbb{G}_n g_\gamma g_\gamma' = O_{P_{n,0}}(1)$ by the CLT and so the second RHS term in (S14) is $O_{P_{n,0}}(n^{-1/2})$. The result now follows by the condition on ν_n and Proposition S1 of Lee and Mesters (2024). □

Lemma S9: *In the setting of Proposition 8, with* $\breve{s}_n := \frac{1}{n}\sum_{i=1}^{n}\breve{\epsilon}_{n,i}^2$,

(i) $\|\hat{M}_n - M\| = O_{P_{n,0}}(n^{-1/2})$ *where* $\hat{M}_n := \left[\frac{1}{n}\sum_{i=1}^{n}X_iZ_{1,i}'\right]\left[\frac{1}{n}\sum_{i=1}^{n}Z_{1,i}Z_{1,i}'\right]^{-1}$;

(ii) $\frac{1}{n}\sum_{i=1}^{n}|\breve{\epsilon}_{n,i} - \epsilon_{n,i}|^2 = O_{P_{n,0}}(n^{-1})$ *and* $|\breve{s}_n^{-1} - \mathbb{E}[\epsilon^2]^{-1}| = O_{P_{n,0}}(n^{-1/2})$;

(iii) $\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\pi}_{n,i}(Z_i)\|^2 x_i^2 = O_{P_{n,0}}(\delta_n^2)$ *for* $x_i \in \{1, \epsilon_i\}$;

(iv) $\mathsf{R}_{n,1} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\breve{s}_n^{-1}\breve{\epsilon}_{n,i}[M - \hat{M}_n]Z_{1,i} = o_{P_{n,0}}(1)$;

(v) $\mathsf{R}_{n,2} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\breve{s}_n^{-1}\breve{\epsilon}_{n,i}\left[\hat{\pi}_{n,i}(Z_i) - \pi(Z_i)\right] = o_{P_{n,0}}(1)$;

(vi) $\mathsf{R}_{n,3} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\breve{s}_n^{-1}(\breve{\epsilon}_{n,i} - \epsilon_i)f(Z_i) = o_{P_{n,0}}(1)$, *where* $f(Z_i) := \pi(Z_i) - MZ_{1,i}$;

(vii) $\mathsf{R}_{n,4} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\breve{s}_n^{-1} - \mathbb{E}[\epsilon^2]^{-1})\epsilon_i f(Z_i) = o_{P_{n,0}}(1)$;

(viii) $\mathsf{S}_{n,1} = \frac{1}{n}\sum_{i=1}^{n}\|\breve{s}_n^{-1}\breve{\epsilon}_{n,i}\left[M - \hat{M}_n\right]Z_{1,i}\|^2 = O_{P_{n,0}}(n^{-1})$;

(ix) $\mathsf{S}_{n,2} = \frac{1}{n}\sum_{i=1}^{n}\|\breve{s}_n^{-1}\breve{\epsilon}_{n,i}\left[\hat{\pi}_{n,i}(Z_i) - \pi(Z_i)\right]\|^2 = O_{P_{n,0}}(\delta_n^2)$;

(x) $\mathsf{S}_{n,3} = \frac{1}{n}\sum_{i=1}^{n}\|\breve{s}_n^{-1}(\breve{\epsilon}_{n,i} - \epsilon_i)f(Z_i)\|^2 = O_{P_{n,0}}(n^{-1})$;

(xi) $\mathsf{S}_{n,4} = \frac{1}{n}\sum_{i=1}^{n}\|(\breve{s}_n^{-1} - \mathbb{E}[\epsilon^2]^{-1})\epsilon_i f(Z_i)\|^2 = O_{P_{n,0}}(n^{-1})$.

Proof. Let $\tilde{\pi}_{n,i}(Z_i) := \hat{\pi}_{n,i}(Z_i) - \pi(Z_i)$. By a simplification of the argument in Proposition S4, (S17) holds for $t \mapsto \gamma + t(0, (b_0, 0, 0))$. Then $P_{n,0} \triangleleft\triangleright P_{n,(0,b_{0,n},0,0)} := Q_n$ by Example 6.5, Theorem 7.2 & Lemma 7.6 in van der Vaart (1998).

(i) Follows from the CLT, given the moment conditions in Assumption 9.

(ii) The first holds by standard arguments as $\beta_n - \beta = O(n^{-1/2})$ and $\mathbb{E}\|Z_i\|^2 < \infty$; the second by the CLT and delta method.

(iii) As $P_{n,0}(\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\pi}_{n,i}(Z_i)\|^2 x_i^2 > K\delta_n^2)$ is bounded by $P_{n,0}(\mathbf{1}_{F_n}\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\pi}_{n,i}(Z_i)\|^2 x_i^2 > K\delta_n^2) + P_{n,\gamma}F_n^{\complement}$. By Markov's inequality $\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\pi}_{n,i}(Z_i)\|^2 x_i^2 = O_{P_{n,0}}(\delta_n^2)$ as by (37), (41) $\mathbb{E}[\mathbf{1}_{F_n}\|\tilde{\pi}_{n,i}(Z_i)\|^2 x_i^2] \leq \mathbb{E}[\mathbb{E}[\mathbf{1}_{F_{n,i}}\|\tilde{\pi}_{n,i}(Z_i)\|^2 x_i^2 | Z_i, \mathcal{C}_{n,-i}]] \lesssim \delta_n^2$, where $F_{n,i}$ is the $\sigma(\mathcal{C}_{n,-i})$ – measurable set on which (41) holds for index i.

(iv) $\mathsf{R}'_{n,1} = \breve{s}_n^{-1}\left[\frac{1}{n}\sum_{i=1}^n Z'_{1,i}\epsilon_i + \frac{1}{n}\sum_{i=1}^n Z'_{1,i}(\breve{\epsilon}_{n,i}-\epsilon_i)\right]\sqrt{n}[M-\hat{M}_n]'$. By (i) and (ii) it suffices to note $\frac{1}{n}\sum_{i=1}^n Z'_{1,i}\epsilon_i = o_{P_{n,0}}(1)$ by the WLLN and $\frac{1}{n}\sum_{i=1}^n Z'_{1,i}(\breve{\epsilon}_{n,i}-\epsilon_i) = o_{P_{n,0}}(1)$ by $\mathbb{E}\|Z_i\|^2 < \infty$, (ii) and Cauchy – Schwarz.

(v) $\mathsf{R}_{n,2} = \breve{s}_n^{-1}\sqrt{n}(\beta-\beta_n)'\frac{1}{n}\sum_{i=1}^n Z_{1,i}\tilde{\pi}_{n,i}(Z_i) + \breve{s}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i\tilde{\pi}_{n,i}(Z_i)$. The first RHS term is $o_{P_{n,0}}(1)$ by (iii) and $\mathbb{E}Z_{1,i}^2 < \infty$. For the second, by $\breve{s}_n^{-1} = O_{P_{n,0}}(1)$, Assumption 11 and Markov's inequality it suffices to observe that $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{F_n}\tilde{\pi}_{n,i,k}(Z_i)^2\epsilon_i^2\right] \lesssim \delta_n^2$ by the argument in (iii) and by (42), $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n\sum_{j=1,j\neq i}^n \mathbf{1}_{F_n}\mathbf{1}_{G_n}\tilde{\pi}_{n,i,k}(Z_i)\tilde{\pi}_{n,j,k}(Z_j)'\epsilon_i\epsilon_j\right] \lesssim \delta_n^2 \to 0$.

(vi) $\mathsf{R}_{n,3} = \breve{s}_n^{-1}[\frac{1}{n}\sum_{i=1}^n f(Z_i)Z'_{1,i}]\sqrt{n}(\beta-\beta_n)$, where the bracketed term is $o_{P_{n,0}}(1)$ by the WLLN as $\mathbb{E}[f(Z)Z'_1] = \mathbb{E}[\pi(Z)Z'_1 - \mathbb{E}[\pi(Z)Z'_1]\mathbb{E}[Z_1Z'_1]^{-1}Z_1Z'_1] = 0$ and the remaning factors are $O_{P_{n,0}}(1)$ by (ii) and $\beta-\beta_n = O(n^{-1/2})$.

(vii) As $\mathbb{E}[\epsilon f(Z)] = \mathbb{E}[\mathbb{E}[\epsilon|Z]f(Z)] = 0$ this follows from (ii) and the WLLN.

(viii) $\mathsf{S}_{n,1} \lesssim s_n^{-2}\|M-M_n\|^2[\frac{1}{n}\sum_{i=1}^n \epsilon_i^2\|Z_{1,i}\|^2 + 2\|\beta-\beta_n\||\epsilon_i|\|Z_{1,i}\|^3 + \|\beta-\beta_n\|^2\|Z_{1,i}\|^4]$ hence this follows by (i), (ii) and the moment conditions in Assumption 9.

(ix) By contiguity $Q_nF_n \to 1$. As the (conditional) distribution of $(\breve{\epsilon}_{n,i}, Z_i)|C_{n,-i}$ under Q_n as that of $(\epsilon_i, Z_i)|C_{n,-i}$, under Q_n, $\mathbb{E}[\breve{\epsilon}_{n,i}^2|Z_i, \mathcal{C}_{n,-i}] \leq C$ a.s. by (37). Therefore, under Q_n, $\mathbb{E}\left[\mathbf{1}_{F_n}\frac{1}{n}\sum_{i=1}^n \breve{\epsilon}_{n,i}^2\|\tilde{\pi}_{n,i}(Z_i)\|^2\right] \lesssim \delta_n^2$, similar to in (iii), and hence Markov's inequality implies $\frac{1}{n}\sum_{i=1}^n \breve{\epsilon}_{n,i}^2\|\tilde{\pi}_{n,i}(Z_i)\|^2 = O_{Q_n}(\delta_n^2)$. By contiguity this holds also under $P_{n,0}$.

(x) As $\mathsf{S}_{n,2} \leq \breve{s}_n^{-2}\|\beta-\beta_n\|^2\frac{1}{n}\sum_{i=1}^n \|Z_{1,i}\|^2[\|\pi(Z_i)\|^2 + \|M\|^2\|Z_{1,i}\|^2]$, the result holds by (ii), $\beta-\beta_n = O(n^{-1/2})$ & the moment conditions in Assumption 9.

(xi) Since $\mathsf{S}_{n,4} \leq (\breve{s}_n^{-1} - \mathbb{E}[\epsilon_i^2])^2\frac{1}{n}\sum_{i=1}^n \epsilon_i^2(\|\pi(Z_i)\|^2 + \|M\|^2\|Z_{1,i}\|^2)$, this follows from (ii) and the moment conditions in Assumption 9. □

Condition (42) The condition in equation (42) is natural when $\hat{\pi}_{n,i}$ is a leave-one-out series estimator: $\hat{\pi}_{n,i}(Z_i) := \hat{\alpha}'_{n,i}p_{K_n}(Z_i)$ for $p_{K_n}(Z_i)$ a K_n-vector of functions of Z_i and $\hat{\alpha}_{n,i} = Q_{n,i}^{-1}\frac{1}{n-1}\sum_{j=1,j\neq i}^n p_{K_n}(Z_i)X'_i$ with $Q_{n,i} := \left[\frac{1}{n-1}\sum_{j=1,j\neq i}^n p_{K_n}(Z_i)p_{K_n}(Z_i)'\right]$. Then, with $\tilde{\pi}_{n,i}(Z_i) := \hat{\pi}_{n,i}(Z_i) - \pi(Z_i)$ and $G_n \in \sigma(Z_1,\ldots,Z_n)$,

$$\mathbb{E}\left[\mathbf{1}_{G_n}\tilde{\pi}_{n,i,k}(Z_i)\tilde{\pi}_{n,j,k}(Z_j)\epsilon_i\epsilon_j\right] = -\mathbb{E}\left[\mathbf{1}_{G_n}\tilde{\pi}_{n,i,k}(Z_i)\epsilon_j\mathbb{E}\left[p_{K_n}(Z_j)'\hat{\alpha}_{n,j}e_k\epsilon_i|Z_i,\mathcal{C}_{n,-i}\right]\right]. \tag{S15}$$

as by $\mathbb{E}[\epsilon_i|Z_i] = 0$ and independence $\mathbb{E}[\epsilon_i\pi(Z_i)|Z_i, \mathcal{C}_{n,-i}] = 0$. The RHS of (S15) is

$$-\mathbb{E}\left[\mathbf{1}_{G_n}\tilde{\pi}_{n,i,k}(Z_i)\epsilon_j p_{K_n}(Z_j)'Q_{n,j}^{-1}\frac{1}{n-1}\sum_{l=1,l\neq j}^n p_{K_n}(Z_l)\mathbb{E}\left[X'_l\epsilon_i|Z_i,\mathcal{C}_{n,-i}\right]e_k\right],$$

and as $\mathbb{E}\left[X_l'\epsilon_i|Z_i,\mathcal{C}_{n,-i}\right]=0$ if $l\neq i$, therefore with $\mu(Z_i):=\mathbb{E}[\upsilon_i'\epsilon_i|Z_i]e_k$,

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{G_n}\tilde{\pi}_{n,i,k}(Z_i)\tilde{\pi}_{n,j,k}(Z_j)\epsilon_i\epsilon_j\right] &= -\mathbb{E}\left[\mathbf{1}_{G_n}\tilde{\pi}_{n,i,k}(Z_i)\epsilon_j p_{K_n}(Z_j)'Q_{n,j}^{-1}\frac{1}{n-1}p_{K_n}(Z_i)\mu(Z_i)\right]\\ &= -\frac{1}{n-1}\mathbb{E}\left[\mathbf{1}_{G_n}e_k' p_{K_n}(Z_j)'Q_{n,j}^{-1}p_{K_n}(Z_i)\mu(Z_i)\mathbb{E}\left[\tilde{\pi}_{n,i,k}(Z_i)\epsilon_j|Z_j,\mathcal{C}_{n,-j}\right]\right].\end{aligned}$$

Arguing as before with the roles of i and j interchanged and using (37) yields

$$|\mathbb{E}\left[\mathbf{1}_{G_n}\tilde{\pi}_{n,i}(Z_i)\tilde{\pi}_{n,j}(Z_j)\epsilon_i\epsilon_j\right]|\leq C^2\frac{|\mathbb{E}\left[\mathbf{1}_{G_n}p_{K_n}(Z_j)'Q_{n,j}^{-1}p_{K_n}(Z_i)p_{K_n}(Z_i)'Q_{n,i}^{-1}p_{K_n}(Z_j)\right]|}{(n-1)^2}.$$

Therefore (42) holds if the RHS is bounded above by a constant multiple of δ_n^2/n.

S4.2.2 The LAN condition

Here I provide examples of $P_{n,h}$ and lower level conditions under which the LAN condition in Assumption 10 holds. Let φ_n be as in (34) with

$$\varphi_{n,1}(b_1):=b_1/\sqrt{n},\quad \varphi_{n,2}(b_2):=\zeta b_2/\sqrt{n},\qquad (b_1,b_2)\in B_1\times B_2, \tag{S16}$$

where B_1 is the space of bounded functions $b_1:\mathbb{R}^{d_Z}\to\mathbb{R}^{d_\theta}$ and B_2 the space of bounded functions $b_2:\mathbb{R}^{d_w}\to\mathbb{R}$ which are continuously differentiable in their first $1+d_\theta$ components with bounded derivative and such that (35) hold.

Proposition S4: *If Assumption 9 holds, $\mathcal{W}_n=\prod_{i=1}^n\mathbb{R}^{d_w}$, $u\mapsto\sqrt{\zeta(u,z)}\in\mathcal{C}^1$ and $p_{n,h}=p^n_{\gamma+\varphi_n(h)}$ with p_γ as in (33). Then Assumption 10 holds.*

Proof. For all large enough n each $\gamma+\varphi_n(h)\in\Gamma$. Assumption 5 is satisfied by construction; to apply Lemma 4 it remains to verify (24) (with $h_n=h$). Let $q_{\tau,b,t}:=p_{(\theta,\eta)+t(\tau,(b_0,b_1,b_2\zeta))}$, $t\in[0,\infty)$ and let $q:=q_{0,0,0}$. For all small enough τ,b and t, $\gamma+t(\tau,(b_0,b_1,b_2\zeta))\in\Gamma$. It suffices to show

$$\int\left[q_{\tau,b,t}^{1/2}-q^{1/2}-\frac{t}{2}\left((\tau',b_0')\dot{l}-\phi'b_1+b_2\right)q^{1/2}\right]^2\mathrm{d}\nu=o(t^2)\qquad\text{as } t\downarrow 0, \tag{S17}$$

where $\dot{l}(W):=-\phi(\epsilon(\theta,\beta),\upsilon(\pi),Z)[X',Z_1']'$. Note that $t\mapsto\sqrt{q_{\tau,b,t}}\in\mathcal{C}^1$ follows from $(e,v)\mapsto\sqrt{\zeta(e,v,z)}\in\mathcal{C}^1$. Under $q_{\tau,b,s}$, $\frac{\partial\log q_{\tau,b,t}}{\partial t}|_{t=s}$ 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. Take a neighbourhood of 0, $\mathcal{U} := [0, \delta)$ such that $1 + sb_2(\epsilon, v, Z)$ is bounded below. Let $\overline{E^2} := C\,[\phi_1(\epsilon, v, Z)^2[\|X\|^2 + \|Z\|^2] + \|\phi_2(\epsilon, v, Z)\|^2 + \|X\|^2 + \|Z\|^2 + 1]$ for some positive constant C. Provided C is large enough, by Assumption 9 $E_s^2 \leq \overline{E^2}$ a.s. and $\mathbb{E}\overline{E^2} < \infty$. Therefore, as $E_{s_n}^2 \to E_s^2$ pointwise, $\mathbb{E}E_{s_n}^2 \to \mathbb{E}E_s^2$, which verifies that Lemma 7.6 in van der Vaart (1998) applies, whence (S17) holds. □

S5 Additional simulation details & results

S5.1 Single index model

As discussed in Section S2.1, C(α) tests are not subject to distortions when nuisance parameters are estimated under shape constraints. Here I explore this in simulation, using Example 1 with $\mathscr{F}$ restricted to contain only monotonically increasing functions. I set $H_0 : \theta = 0$ and consider three possible link functions: f_1 is a logistic function, whilst f_2 and f_3 are double logistic functions which include a flat section inbetween two increasing sections. These functions are formally defined in (S18) below and plotted in Figure S3. Each considered link function has flat sections which may cause monotonicity constraints to bind in the estimation of f. I explore the effect this has on the rejection frequencies of the $\psi_{n,\theta}$ test as described on p. 20 and an Ichimura (1993) – style Wald test. Both tests are computed with f, f' estimated by 9 monotonic I – splines (e.g. Ramsay, 1988), whilst Z_1 is estimated using 6 cubic B – splines. As $\tilde{\mathcal{I}} > 0$ in this design, $\boldsymbol{\nu} = 0$. ϵ is drawn from a standard normal and the covariates are drawn as $X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$, where each $Z_k \sim U(-1.5, 1.5)$ is independent.

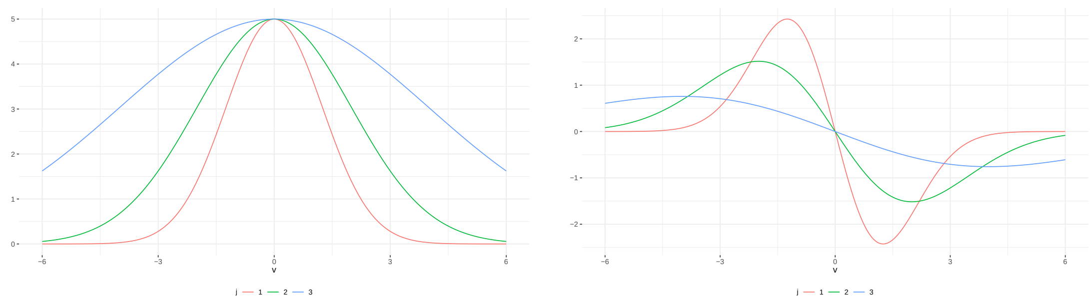
The f_j functions used are as follows. Let $b(x) := \mathbf{1}\{x > 0\} \exp(-1/x)$ (a bump function) and form the smooth transition function $a(x) := b(x)/(b(x) + b(1 - x))$. Then with $g(v; a, b) := 1/(1 + \exp(-(x - b)/a))$, a logistic function, let

$$
\begin{aligned}
f_1(v) &:= 8g(v, 0.25, 0)\ ;\\
f_2(v) &:= 4\big[\mathbf{1}\{4v \leq -1\}g(4v, 0.4, -3) + \mathbf{1}\{4v > 1\}(1 + g(4v, 0.4, 3))\\
&\qquad + \mathbf{1}\{1- < 4v \leq 1\}a((4v + 1)/2)(1 + g(1, 0.4, 3) - g(-1, 0.4, -3))\big]\\
f_3(v) &:= 4\big[\mathbf{1}\{3v \leq -1\}g(3v, 0.2, -3) + \mathbf{1}\{3v > 1\}(1 + g(2v, 0.2, 3))\ ;\\
&\qquad + \mathbf{1}\{1- < 3v \leq 1\}a((3v + 1)/2)(1 + g(1, 0.2, 3) - g(-1, 0.2, -3))\big]\ .
\end{aligned}
\tag{S18}
$$

Table S1 displays the empirical rejection frequencies attained by $\psi_{n,\theta}$ and the Wald test. The former provides rejection rates close to the nominal level of 5% in each simulation design considered. The Wald test displays substantial overrejection in each simulation design. The 3 panels of Figure S4 depict the finite-sample power curves for $f = f_1, f_2, f_3$ respectively. In each panel, the Wald test shows a relatively slow increase in power as θ moves away from 0 with $\psi_{n,\theta}$ providing a much higher rate of increase in power as θ deviates from the null.[6]

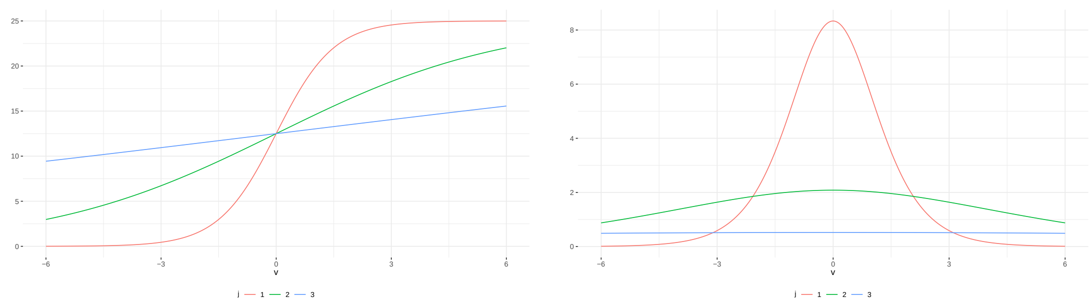
S6 Tables and Figures

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



(a) f (b) f'

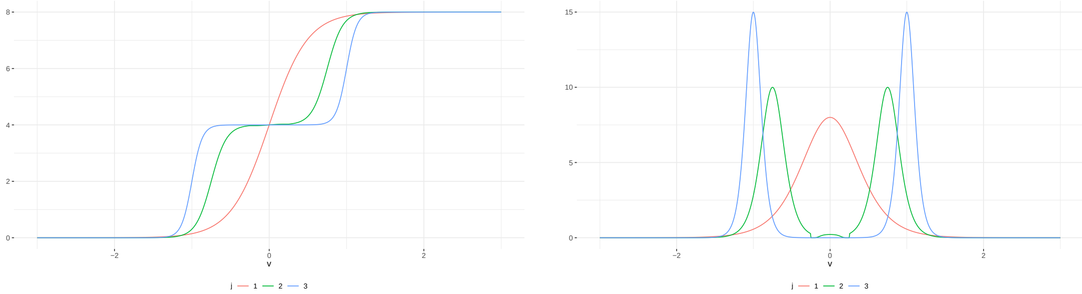
Figure S2: Index functions$f_j(v) = 25\,(1+\exp(-v/c_j))^{-1}$



(a) f (b) f'

[6]The power of the Wald test exceeds that of $\psi_{n,\theta}$ around the null. However, this is not a like-for-like comparison, as the Wald test over-rejects; see Table S1.

Figure S3: Double logistic index functions as in (S18)

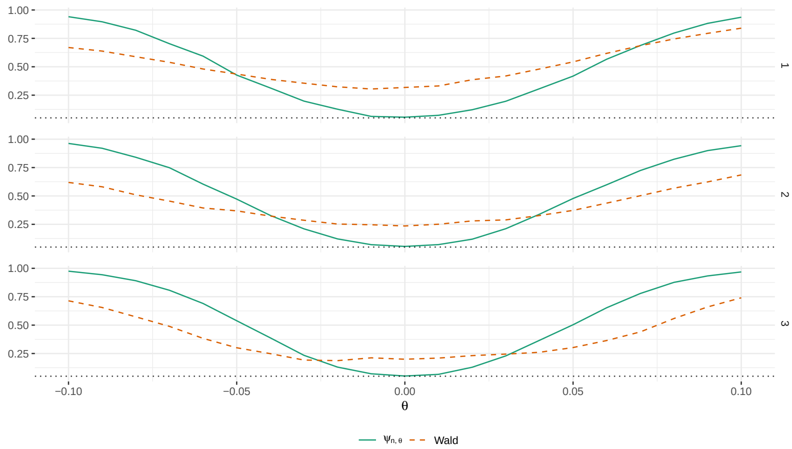


(a) f (b) f'

Table S1: ERF (%), index function as in (S18)

	$\psi_{n,\theta}$			Wald		
n	f_1	f_2	f_3	f_1	f_2	f_3
400	6.60	5.56	5.58	34.92	28.78	26.38
600	6.02	5.58	4.90	34.72	26.42	23.26
800	5.60	5.56	5.20	31.78	23.58	19.98

Figure S4: ERF (%), index function as in (S18)



References

Andrews, D. W. K. (1999), "Estimation When a Parameter is on a Boundary," *Econometrica*, 67, 1341–1383.

— (2001), "Testing When a Parameter is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734.

Andrews, D. W. K. and Guggenberger, P. (2010), "Asymptotic Size and a Problem with Subsampling and with the m out of n Bootstrap," *Econometric Theory*, 26, 426–468.

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.

Geyer, C. J. (1994), "On the asymptotics of constrained M-estimation," *Annals of Statistics*, 22, 1993–2010.

Hoesch, L., Lee, A., and Mesters, G. (2024), "Supplement to "Locally Robust Inference for Non-Gaussian SVAR Models"," *Quantitative Economics Supplementary Material*, 15.

Ichimura, H. (1993), "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models," *Journal of Econometrics*, 58, 71–120.

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

Ketz, P. (2018), "Subvector inference when the true parameter vector may be near or at the boundary," *Journal of Econometrics*, 207, 285–306.

Le Cam, L. M. (1986), *Asymptotic Methods in Statistical Decision Theory*, New York, NY, USA: Springer.

Lee, A. and Mesters, G. (2024), "Supplement to 'Locally Robust Inference for Non-Gaussian Linear Simultaneous Equations Models'," *Journal of Econometrics Supplementary Material*, 240.

Lehmann, E. L. and Romano, J. P. (2005), *Testing Statistical Hypotheses*, New York, NY, USA: Springer, 3rd ed.

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.

Newey, W. K. and McFadden, D. (1994), "Chapter 36 Large sample estimation and hypothesis testing," Elsevier, vol. 4 of *Handbook of Econometrics*, pp. 2111–2245.

Ramsay, J. O. (1988), "Monotone Regression Splines in Action," *Statistical Science*, 3, 425–441.

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

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

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*, W. de Gruyter.

van der Vaart, A. W. (1991), "An Asymptotic Representation Theorem," *International Statistical Review / Revue Internationale de Statistique*, 59, 97–121.

— (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.

van der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York, NY, USA: Springer-Verlag New York, Inc.