Locally regular and efficient tests in potentially non-regular semiparametric models

Adam Lee*

February 7, 2024

Abstract

This paper considers hypothesis testing in semiparametric models which may be non – regular. I introduce a notion of local regularity for (sequences of) tests and show that C(α) – style tests are locally regular under mild conditions, including in cases where locally regular estimators do not exist, such as models which are (semiparametrically) weakly identified. I characterise the appropriate limit experiment in which to study local (asymptotic) optimality of tests in the non – regular case, permitting the generalisation of classical power bounds to this case. I give conditions under which these generalised power bounds are attained by the proposed C(α) – style tests. Two examples are worked out in detail. The finite sample performance of the proposed tests is evaluated in a simulation study and an empirical application.

JEL classification: C10, C12, C14, C21, C39

Keywords: Hypothesis testing, local asymptotics, uniformity, semiparametric models, weak identification, boundary, regularisation, single-index, instrumental variables.

---

*BI Norwegian Business School, adam.lee@bi.no. Previous versions of this paper were titled "Robust and Efficient Inference For Non-Regular Semiparametric Models". I have benefitted from discussions with and comments / questions from Majid Al-Sadoon, Isiah Andrews, Christian Brownlees, Bjarni G. Einarsson, Juan Carlos Escanciano, Kirill Evdokimov, Lukas Hoesch, Geert Mesters, Vladislav Morozov, Jonas Moss, Whitney Newey, Katerina Petrova, Francesco Ravazzolo, Barbara Rossi, André B. M. Souza, Emil Aas Stoltenberg, Philipp Tiozzo and participants at various conferences and seminars. All errors are my own.

1 Introduction

It is often considered desirable that estimators are "locally regular" in that they converge to the same limiting distribution along sequences of "local alternatives" which cannot be consistently distinguished from the true parameter, even asymptotically.[1] Unfortunately, there are many semiparametric models in which locally regular estimators do not exist.[2] One necessary condition is given by Chamberlain (1986), who shows that if the efficient information for a scalar parameter is 0, then no locally regular estimator of that parameter exists. This result can be extended to singularity of the efficient information matrix implying the non-existence of locally regular estimators of Euclidean parameters. I refer to such models as "non-regular".

In this paper, I demonstrate that the situation for testing is different. After defining an appropriate notion of local regularity for tests, I exhibit a broad class of tests – based on the C(α) idea of Neyman (1959, 1979) – which have this property.

One key consequence of this result is that it provides a method to construct tests in a general class of semiparametric models which do not (asymptotically) over-reject under the a semiparametric generalisation of weak identification asymptotics.[3] In addition to the well-studied case where weak identification is due to potential identification failure at certain values of a finite dimensional nuisance parameter, the results in this paper also apply to the case where identification failure is due to the value of an infinite dimensional nuisance parameter.

The approach used to construct such test statistics leads to tests which also behave well in other irregular settings: for example, these tests continue to provide good inference when nuisance functions have been estimated using regularisation or under shape restrictions.[4]

Achieving this local regularity does come at the expense of (local asymptotic) power. I characterise the appropriate limit experiment in which to study (local asymptotic) optimality of tests in the case where the efficient information matrix may be singular: the finite sample experiments converge weakly to an experiment

[1] Precise definitions will be given below. See Bickel, Klaassen, Ritov, and Wellner (1998); van der Vaart (2002), for example, for textbook treatments.

[2] See e.g. Ritov and Bickel (1990); Newey (1990) for some examples.

[3] The notion of semiparametric weak identification asymptotics used in this paper is essentially that of Kaji (2021); see also Andrews and Mikusheva (2022). The only difference is that I work directly with local asymptotic normality [LAN] (as opposed to differentiability in quadratic mean [DQM], which implies LAN in the i.i.d. case). This allows the theory to apply equally to non-i.i.d. models.

[4] See Chetverikov, Santos, and Shaikh (2018) for a recent review of the use of shape constraints in econometrics.

which can be "matched" by a Gaussian shift on the quotient of the original (local) parameter space under an induced covariance function. This permits the generalisation of a number of classical power bounds to this setting. Moreover, I show that the locally regular C(α) tests proposed in this paper acheive these power bounds under certain conditions. These conditions are weaker than those in the literature.[5]

Following the theoretical development, I provide two worked out examples which I use to conduct an extensive simulation study into the finite sample performance of the proposed tests. I consider (i) a single index model and (ii) an instrumental variables model with a nonparametric first stage. In each case, the parameter of interest may fail to be identified depending on the value of an infinite dimensional nuisance parameter, leading to possible weak identification issues. Nevertheless, the tests proposed in this paper display good finite sample performance in each model, including in weakly identified cases. Additionally, in the single index model, I investigate the behaviour of the proposed test when the link function is estimated under a monotonicity constraint which may be close to binding and compare it to a Wald test with the same asymptotic power function. I find that plugging in the monotonicity constrained estimator results in lower power for the Wald test, but not the locally regular C(α) test.

This paper is connected to three main strands of the literature: the first is that concerned with general results on estimation and testing in semiparametric models. Much of this is now textbook material: see e.g. Newey (1990); Choi et al. (1996); Bickel et al. (1998); van der Vaart (1998, 2002). The second is the literature on C(α) - style tests. Such tests were introduced by Neyman (1959, 1979) and have seen many useful applications, most recently as a way to handle machine learning or otherwise high dimensional first steps (see e.g. Chernozhukov, Hansen, and Spindler, 2015; Bravo, Escanciano, and Van Keilegom, 2020; Chernozhukov, Escanciano, Ichimura, Newey, and Robins, 2022). In this paper, the structure which ensures the good performance of these tests in such settings is used for a different purpose – to create tests which remain robust in non-regular settings. Lastly, the literature on robust testing in non – regular or otherwise non – standard settings is closely related to this paper (e.g. Andrews and Guggenberger, 2009; Romano and Shaikh, 2012; Elliott, Müller, and Watson, 2015; McCloskey, 2017). In particular, the locally regular tests derived in this paper are particularly

[5]In particular, the attainment result is well known if either (a) the observations are i.i.d. (cf. van der Vaart, 1998, Chapter 25) or (b) the information operator (as defined in Choi, Hall, and Schick, 1996, p. 846) is boundedly invertible (Choi et al., 1996). The result given in this paper does not require either of these conditions.

useful in cases of weak identification and therefore this paper is closely related to the literature on weak identification robust inference in econometrics (e.g. Staiger and Stock, 1997; Dufour, 1997; Stock and Wright, 2000; Kleibergen, 2005; Andrews and Cheng, 2012; Andrews and Mikusheva, 2015, 2016). In particular this paper is most closely related to the recent work on semiparametric weak identification (Kaji, 2021; Andrews and Mikusheva, 2022) and extends the notion of semiparametric weak identification considered there to non – i.i.d. models.[6]

2 Locally regular testing

The goal considered throughout this paper is to construct hypothesis tests of the form $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ in the sequence of models $\mathcal{P}_n = \{P_{n,\gamma} : \gamma \in \Gamma\}$ where $\gamma = (\theta, \eta) \in \Gamma = \Theta \times \mathcal{H}$ for some open $\Theta \subset \mathbb{R}^{d_\theta}$ and $\mathcal{H}$ an arbitrary set. Each $\mathcal{P}_n$ consists of probability measures on a measurable space $(\mathcal{W}_n, \mathcal{B}(\mathcal{W}_n))$ and is dominated by a σ-finite measure ν_n.[7]

In this section I will define local regularity for testing and heuristically describe how such tests can be constructed, with technical details deferred until the following section. I then explain how this concept can be useful to derive robust testing procedures in two common non-standard inference problems.

2.1 Defining local regularity for tests

Local regularity for estimators To motivate the definition of local regularity for tests, I first recall the definition of local regularity of estimators. Suppose that $\hat{\theta}_n$ is a sequence of estimators of θ and $P_{n,\gamma,h} = P_{n,(\theta+\tau/\sqrt{n},\eta_n(b))}$ a sequence of local alternatives to $P_{n,\gamma}$ for some $h = (\tau, b) \in H := \mathbb{R}^{d_\theta} \times B$.[8] The estimator sequence $\hat{\theta}_n$ is then called "locally regular" at γ if

$$\sqrt{n}\left(\hat{\theta}_n - \theta - \frac{\tau}{\sqrt{n}}\right) \overset{P_{n,\gamma,h}}{\rightsquigarrow} \mathcal{L}_\gamma, \qquad h \in H. \tag{1}$$

[6]Failure of local identification and singularity of the information matrix are closely linked in parametric cases, see Rothenberg (1971). In the semiparametric case, parameters may be nonparametrically identified but nevertheless have a singular efficient information matrix. The relationship between the efficient information matrix and identification is studied in detail by Escanciano (2022).

[7]Typically the index n is sample size and $\mathcal{W}_n$ is the space in which a sample of size n takes its values. This will be the situation considered in Section 3.4 as well as in the examples treated in Section 4.

[8]The general definition of a "local alternative" is given in the following section. In the parametric case one typically takes $\eta_n(b)$ to be of the form $\eta + b/\sqrt{n}$.

for some law $\mathcal{L}_\gamma$ which – as indicated by the notation – may depend on γ but not on h.[9]

Local regularity for tests Motivated by (1), I now define a notion of local regularity for tests of H_0 against H_1.

DEFINITION 2.1: *A sequence of tests* $\phi_n : \mathcal{W}_n \to [0,1]$ *of the hypothesis* $H_0 : \theta = \theta_0$ *against* $H_1 : \theta \neq \theta_0$ *is locally regular if*

$$\pi_n(\tau, b) \coloneqq P_{n,\gamma,h}\phi_n \to \pi_\gamma(\tau), \qquad h \in H. \tag{2}$$

In words, the finite sample (local) power function of the test, π_n converges under each $P_{n,\gamma,h}$ to a function π_γ which depends only on τ. This requirement is in the same spirit as (1): the parameter b which describes local deviations from η does not affect the limit.[10] If a sequence of tests does *not* satisfy (2.1) I shall call it locally non – regular.

In addition to (2), if one is interested in the hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$, typically one additionally wants to ensure that $\pi_\gamma(0) \leq \alpha$, for α a given significance level, i.e. that the test does not asymptotically over – reject.[11]

Local regularity of test sequences as in Definition 2.1 is a pointwise concept. It is also of interest to consider a version of local regularity which holds uniformly over certain subsets.

DEFINITION 2.2: *A sequence of tests* $\phi_n : \mathcal{W}_n \to [0,1]$ *of the hypothesis* $H_0 : \theta = \theta_0$ *against* $H_1 : \theta \neq \theta_0$ *is locally uniformly regular on* $K \subset H$ *if* (2) *holds uniformly on* K

$$\sup_{(\tau,b)\in K} |\pi_n(\tau, b) - \pi_\gamma(\tau)| \to 0. \tag{3}$$

In the case where H is a (pseudo-)metric space and K is a compact set, to go from the pointwise convergence in (2) to the uniform convergence in (3) it is

[9] Ideally the law $\mathcal{L}_\gamma$ is the semiparametric efficiency bound for locally regular estimators given by the Hájek – Le Cam convolution Theorem (see e.g. van der Vaart, 1998, Theorem 25.20 & Lemma 25.25). In this case the estimator sequence $\hat{\theta}_n$ is usually called "best regular".

[10] τ, describing the local deviation from θ *does* affect the limit: (1) may be re-written as

$$\sqrt{n}(\hat{\theta}_n - \theta) \overset{P_{n,\gamma,h}}{\rightsquigarrow} \mathcal{L}_{\gamma,\tau}$$

where $\mathcal{L}_{\gamma,\tau}$ is the law of $X - \tau$ for $X \sim \mathcal{L}_\gamma$.

[11] For other testing problems, other restrictions may be preferable. For example, if θ is scalar and the testing problem is $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$, a more natural requirement would be that $\pi_\gamma(\tau) \leq \alpha$ for all $\tau \leq 0$.

necessary and sufficient to show that the sequence of functions π_n is asymptotically equicontinuous on K.[12]

Directly establishing asymptotic equicontinuity of the power functions $\pi_n(\tau, b)$ may be complicated in many cases. It is, however, often possible to show stronger results which immediately imply this property. For instance, if one can show that the functions $h \mapsto P_{n,\gamma,h}$ are asymptotically equicontinuous in total variation, the required asymptotic equicontinuity of the power functions follows immediately. Despite being (much) stronger, this requirement can often be relatively straightforward to demonstrate. For example, in the classical case of a parametric model for i.i.d. data, this asymptotic equicontinuity in total variation follows from the differentiability in quadratic mean condition typically used to demonstrate local asymptotic normality (cf. e.g. Theorem 7.2 in van der Vaart (1998) and Theorem 80.13 in Strasser (1985)).

A class of locally regular tests To construct tests of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ which have property (2) (or (3)), I use a generalisation of the class of C(α) tests introduced by Neyman (1959, 1979) to characterise optimal tests in regular parametric models.

I will heuristically outline the construction of C(α) tests in such parametric models to build intuition for the theoretical development in the following section.[13] Thus, suppose temporarily that $\mathcal{H} \subset \mathbb{R}^{d_\eta}$ and the observed data $(W_1, \ldots, W_n)$ is drawn i.i.d. from a parametric density p_γ. Let $\dot{\nu}_\gamma$ be the score functions for the (now finite dimensional) nuisance parameter η, i.e. the partial derivatives of the log likelihood for an observation W: $\dot{\nu}_\gamma := \nabla_\eta \log p_\gamma$. Let $f_\gamma = f_{\theta,\eta}$ be a vector of d_θ moment conditions which are mean-zero under the null hypothesis, i.e. $\mathbb{E}_{\theta_0,\eta} f_{\theta_0,\eta}(W) = 0$.

Such regular parametric models are typically locally asymptotically normal ("LAN"): the log-likelihood ratio admits a local quadratic approximation:

$$\log \frac{q_{\gamma,h}}{p_{\theta,\eta}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tau' \dot{\ell}_\gamma(W_i) + b'\dot{\nu}_\gamma(W_i) - \mathbb{E}_\gamma \left[\tau'\dot{\ell}_\gamma(W_i) + b'\dot{\nu}_\gamma(W_i)\right]^2 + o_{P_\gamma}(1), \quad (4)$$

for $q_{\gamma,h} := p_{\theta+\tau/\sqrt{n},\eta+b/\sqrt{n}}$ where $(\tau, b) \in \mathbb{R}^{d_\theta + d_\eta}$ and $\dot{\ell}_\gamma$ the score functions for θ: $\dot{\ell}_\gamma := \nabla_\theta \log p_\gamma$.

[12] The same is true if the requirement that K be compact is replaced with the requirement that K be totally bounded. See e.g. Davidson (2021), p. 123, for the definition of asymptotic equicontinuity.

[13] That is, I will avoid discussing the required regularity conditions for this construction; such details are given in full for the general case treated in the subsequent section.

Under LAN, the asymptotic distribution of scaled sums of the moment condition f_γ under the local alternative $(\theta + \tau/\sqrt{n}, \eta + b/\sqrt{n})$ is given by Le Cam's third Lemma (e.g. van der Vaart, 1998, Example 6.5):

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_\gamma(W_i) \overset{P_{n,\gamma,h}}{\rightsquigarrow} \mathcal{N}\left(\mathbb{E}_\gamma[f_\gamma(W)\dot{\ell}_\gamma(W)']\tau + \mathbb{E}[f_\gamma(W)\dot{\nu}_\gamma(W)']b,\ \mathbb{E}_\gamma[f_\gamma(W)f_\gamma(W)']\right).$$

In consequence, the limiting distribution will depend on b if and only if $\mathbb{E}[f_\gamma(W)\dot{v}_\gamma(W)] \neq 0$.

In order to ensure this covariance is zero, a C(α) test is not based on the original moment condition f_γ, but rather on the orthogonal projection

$$g_\gamma \coloneqq f_\gamma - \mathbb{E}_\gamma\left[f_\gamma(W)\dot{\nu}_\gamma(W)'\right]\mathbb{E}_\gamma\left[\dot{\nu}(W)\dot{\nu}_\gamma(W)'\right]^{-1}\dot{\nu}_\gamma = \Pi\left[f_\gamma \middle| \mathrm{Span}(\dot{\nu}_\gamma)^\perp\right], \tag{5}$$

where $\Pi[\cdot|\mathcal{S}]$ is the orthogonal projection onto the closed subspace $\mathcal{S} \subset L_2(P_\gamma)$.[14] Then, as $\mathbb{E}[g_\gamma(W)\dot{\nu}_\gamma(W)'] = 0$ by construction, by the same argument as above

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_\gamma(W_i) \overset{P_{n,\gamma,h}}{\rightsquigarrow} \mathcal{N}\left(\mathbb{E}_\gamma[g_\gamma(W)\dot{\ell}_\gamma(W)']\tau,\ \mathbb{E}_\gamma[g_\gamma(W)g_\gamma(W)']\right). \tag{6}$$

In practice, for the test to be feasible one must replace the unknown nuisance parameters η with an estimator (which may be estimated under the null). Typically one also estimates the (pseudo-)inverse of $V_\gamma \coloneqq \mathbb{E}_\gamma[g_\gamma(W)g_\gamma(W)']$ and weights the estimated components $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{g}_{n,\theta}(W_i)$ by this matrix in a quadratic form, forming a feasible test statistic $\hat{S}_{n,\theta}$. Provided these estimators are sufficiently accurate, $\hat{S}_{n,\theta}$ will converge to a χ^2 distribution under $P_{n,\gamma,h}$. As such score – type test statistics based on the moment conditions g_γ will have asymptotic distributions free of b. One can then choose an appropriate critical value c such that the test of the form $\mathbf{1}\{\hat{S}_{n,\theta} > c\}$ does not over-reject under any $P_{n,\gamma,h}$ consistent with H_0.

Tests constructed in this manner will be locally regular tests in the sense of definition 2.1. The (rigorous) extension of this argument to the semiparametric case, with possible singularity of the variance matrix V_γ (as may happen, for example, in cases with potential identification failure or in underidentified models) is given in Section 3.

[14]See equation (5) and the surrounding discussion in Neyman (1979).

The non-iid case The heuristic derivation of the class of C(α) tests given above imposed that the researcher observed a random sample, i.e. that the model $\mathcal{P}_n$ consists of n-fold product measures $P_{n,\gamma} := P_\gamma^n$. This is not necessary for the local regularity of C(α) tests, and is not imposed in the general theory discussed in Section 3. It does, however, routinely simplify expressions and the demonstration of the required regularity conditions. Given this and its central role as a benchmark case in statistics and econometrics, the simplifications available in the i.i.d. case are explicitly discussed in Section 3.4.

Power optimality As noted above, Neyman (1959, 1979) initially developed the C(α) test in order to discuss testing optimality. In regular parametric models, the conclusion is (up to regularity conditions) that C(α) tests based on $\dot{\ell}_\gamma$, the scores for θ, attain the (local asymptotic) power bounds for various classes of testing problems. This is known to also be true in (regular) semiparametric models if (a) the data is i.i.d. (cf. Chapter 25 in van der Vaart, 1998) or (b) the information operator (as defined in Choi et al., 1996, p. 846) is boundedly invertible (Choi et al., 1996). In Section 3, I show that this result holds without requiring either (a) or (b) and, moreover, persists in the non – regular case, where the efficient information matrix may be rank deficient.[15]

2.2 Robust testing in non-standard problems

I now explain how the ideas just described can be used to derive tests which are well behaved in the face of two commonly encountered non-standard inference problems in econometrics.[16] I provide examples in each case, for which locally regular C(α) style tests will be explicitly developed in Section 4.

Weak identification As noted in the introduction, in many models there are values of γ such that estimators satisfying (1) do not exist. Points γ, where the parameter of interest is un- or under-identified provide an important class of examples.[17] As is well known, even if θ is identified at γ, finite sample inference may still be poor if γ is too close to a point of identification failure relative to

[15] This more general result contains the classical regular case as a special case.

[16] I note that, whilst widely applicable, the approach developed in this paper does not apply to all types of non-standard inference problems encountered in econometrics. For instance, AR(1) models with a local-to-unity root are locally asymptotically quadratic (LAQ) rather than LAN (Jansson, 2008). The results in this paper are derived for LAN models and hence do not apply in this case.

[17] For such cases it is not possible to estimate θ consistently, let alone regularly.

the amount of information contained in the sample. Such weak identification concerns have been widely studied in models where the part of η which causes the potential identification failure is finite dimensional (e.g. Andrews and Cheng, 2012; Andrews and Mikusheva, 2015). There are also many examples where this may occur due to the value of *infinite-dimensional* nuisance parameters. Kaji (2021) considers estimation in weakly identified semiparametric models, whilst Andrews and Mikusheva (2022) consider semiparametric weak identification in GMM models. Both use a differentiability in quadratic mean (DQM) condition to define "weak identification embeddings" in i.i.d. models. As such DQM - type conditions are less straightforward to work with in non - i.i.d. cases, in this paper I use an analogous notion – based directly on the LAN expansion – which broadens the applicability of this class of semiparametric weak identification sequences. The key property of these sequences is that they are local (i.e. contiguous) alternatives to a point of identification failure.

I now give three examples of semiparametric models where the parameter of interest θ may be un- or under- identified depending on the value of an infinite dimensional nuisance parameter.

Example 2.1 (Single – index model): *Suppose that the researcher observes n i.i.d. copies of $W = (Y, X)$ where*

$$Y = f(X_1 + X_2'\theta) + \epsilon, \qquad \mathbb{E}[\epsilon|X] = 0,$$

and where f belongs to some set of continuously differentiable functions $\mathscr{F}$. The description of the model is completed by a parameter ζ which describes the distribution of (X, ϵ). If f is flat, i.e. $f' = 0$, then the parameter θ is unidentified.

Example 2.2 (Instrumental variables): *Suppose the researcher observes n i.i.d. copies of $W = (Y, X, Z)$ where*

$$Y = X'\theta + Z_1'\beta + \epsilon, \qquad \mathbb{E}[\epsilon|Z] = 0, \qquad Z = (Z_1', Z_2')'.$$

If the k-th component of $\pi(Z) := \mathbb{E}[X|Z]$ is zero, θ_k is unidentified.

Example 2.3 (Independent components supply & demand model): *Suppose the researcher observes n i.i.d. copies of $W = Y \in \mathbb{R}^2$, where*

$$Y = A(\theta, \sigma)^{-1}\epsilon, \qquad \mathbb{E}\epsilon = 0, \quad \mathrm{Var}(\epsilon) = I, \quad \epsilon_1 \perp\!\!\!\perp \epsilon_2,$$

with

$$A(\theta,\sigma)=\begin{bmatrix}\sigma_1^{-1} & 0\\ 0 & \sigma_2^{-1}\end{bmatrix}\begin{bmatrix}1 & -\theta_1\\ 1 & -\theta_2\end{bmatrix},\quad \sigma\in(0,\infty)^2,\quad \theta\in(-\infty,0)\times(0,\infty).$$

This simple equilibrium supply & demand model is completed by the density functions ζ_1,ζ_2 of ϵ_1,ϵ_2 respectively. If no more than one ϵ_k has a standard Gaussian distribution, θ is identified (Comon, 1994). If $\epsilon\sim\mathcal{N}(0,I)$, θ is underidentified.

In each of the above examples, no locally regular estimator exists. For Examples 2.1 and 2.2, locally regular C(α) tests which satisfy $\pi(0)\leq\alpha$ are developed in Section 4. The local regularity of these tests ensures that they do not asymptotically over-reject under semiparametric weak identification parameter sequences. The simulation exercises reported in the same section demonstrate that this approximation provides a good guide to the finite sample reality: unlike many alternative procedures, these tests exhibit finite sample rejection frequencies close to the nominal level under the null, even in very weakly identified settings.

Each of these examples are in the i.i.d. case, though as emphasised above, this is not a requirement. For example, locally regular C(α) tests for the potentially un- / under- identified parameter in a structural vector autoregressive model built on top of an ICA model similar to Example 2.3 were developed in Hoesch, Lee, and Mesters (2024).[18]

Parameters close to the boundary A second non-standard inference problem which has attracted substantial attention in statistics & econometrics is inference when a finite - dimensional nuisance parameters 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 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 (e.g. van der Vaart, 1998, Theorem 6.6).

In consequence, using the "standard" normal approximations which usually obtain for extremum estimators (cf. Newey and McFadden, 1994) can lead to either misleading inference or poor power. In the literature there are examples of

[18] Lee and Mesters (2024) provided locally regular C(α) tests for the potentially un- / under-identified parameter in linear simultaneous equations models built on an ICA model similar to Example 2.3.

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 orthogonalisation (5), plugging in any $\sqrt{n}$ - consistent estimator $\hat{\eta}_n$ of η is 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 (6).

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. Consider the following example, based on the single index model of Example 2.1.

Example 2.4: *Recall the model of Example 2.1:*

$$Y = f(X_1 + X_2'\theta) + \epsilon, \qquad \mathbb{E}[\epsilon|X] = 0,$$

and now suppose that f belongs to some subset of continuously differentiable functions $\mathscr{F}$ which also satisfy a shape restriction. For instance, $\mathscr{F}$ may contain only monotonically increasing functions or convex functions.

Analogously to in 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.

In the simulation study of Section 4, this phenomenon is explored in the context of Example 2.4. The locally regular C(α) test with f estimated under a monotonicity restriction demonstrates good performance, including when the imposed restriction is close to binding. In contrast, a Wald test based on a non-linear least squares estimator (as in Ichimura (1993)) delivers conservative inference when such a restricted estimator of f is used.

3 Main results

This section establishes the main theoretical results of the paper. These are first established under high-level assumptions, which allows the results to be stated in a manner which applies to many situations, including cases with dependent or non-identically distributed data. Section 3.4 considers simplifications which are valid in the benchmark case of i.i.d. data and a "smooth" statistical model.

3.1 The setting

Local asymptotic normality I now formalise and generalise the required LAN condition as in (4). Let $H_\gamma = \mathbb{R}^{d_\theta} \times B_\gamma$ be a subset of a linear space containing 0, and suppose that $\{P_{n,\gamma,h} : h \in H_\gamma\} \subset \mathcal{P}_n$ are such that $P_{n,\gamma} = P_{n,\gamma,0}$. A typical element of H_γ will be written as $h = (\tau, b) \in \mathbb{R}^{d_\theta} \times B_\gamma$.[19] The measures $P_{n,\gamma,h}$ should be interpreted as local perturbations of the measure $P_{n,\gamma}$ in a direction $h \in H_\gamma$.

The null hypothesis $H_0 : \theta = \theta_0$ corresponds to the set of perturbations $H_{\gamma,0} \coloneqq \{(0, b) : b \in B_\gamma\}$ and the alternative $H_1 : \theta \neq \theta_0$ to $H_{\gamma,1} \coloneqq \{h = (\tau, b) : 0 \neq \tau \in \mathbb{R}^{d_\theta}, b \in B_\gamma\}$. As such, $P_{n,\gamma,h}$ for $h \in H_{\gamma,0}$ will be referred to as local perturbations consistent with the null hypothesis, whilst $P_{n,\gamma,h}$ for $h \in H_{\gamma,1}$ are local alternatives. To frame this another way, I consider tests of $\tau = 0$ against $\tau \neq 0$ in the local models $\{P_{n,\gamma,h} : h \in H_\gamma\}$.

The key technical condition under which the theory in this paper is developed is local asymptotic normality (see e.g. van der Vaart, 1998, Chapter 7 or Le Cam and Yang, 2000, Chapter 6). Define the log-likelihood ratios

$$L_{n,\gamma}(h) \coloneqq \log \frac{p_{n,\gamma,h}}{p_{n,\gamma,0}}, \qquad \text{where } \; p_{n,\gamma,h} \coloneqq \frac{\mathrm{d}P_{n,\gamma,h}}{\mathrm{d}\nu_n}, \; \text{for } h \in H_\gamma. \tag{7}$$

Assumption 3.1 (Local asymptotic normality): *$L_{n,\gamma}(h)$ satisfies*

$$L_{n,\gamma}(h) = \Delta_{n,\gamma} h - \frac{1}{2}\|\Delta_{n,\gamma} h\|^2 + R_{n,\gamma}(h), \tag{8}$$

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

$$\Delta_{n,\gamma} h \overset{P_{n,\gamma}}{\rightsquigarrow} \mathcal{N}(0, \sigma_\gamma(h)), \qquad \sigma_\gamma(h) \coloneqq \lim_{n\to\infty} \|\Delta_{n,\gamma} h\|^2.$$

$\Delta_{n,\gamma}$ is the *score operator* (cf. van der Vaart, 1998, p. 371). It produces score functions (or "scores") from "directions" $h \in H_\gamma$. As such, the LAN expansion (8) requires that the log-likelihoods are approximately equal to the score less half of its variance. This approximation and the asymptotic normality of the scores leads to contiguity.

[19] In most examples, H_γ will be a linear space. The more general situation as considered here is nevertheless important to allow for, for example, Euclidean nuisance parameters subject to boundary constraints. In such a setting, if the constraint is binding at η, then η can only be perturbed in certain directions if $P_{n,\gamma,h}$ is to remain within the model.

Remark 3.1: *Assumption 3.1 ensures that the pairs of sequences* $(P_{n,\gamma})_{n\in\mathbb{N}}$ *and* $(P_{n,\gamma,h})_{n\in\mathbb{N}}$ *are mutually contiguous for any* $h \in H_\gamma$ *(see e.g. van der Vaart, 1998, Example 6.5).*

Remark 3.2: *If* H_γ *is (pseudo-)metrised one may consider a uniform version of Assumption 3.1, i.e. uniform local asymptotic normality. Such a version is given in Assumption S2.1 and – as shown by Proposition S2.2 – is equivalent to Assumption 3.1 plus asymptotic equicontinuity on compact sets of* $h \mapsto \Delta_{n,\gamma}h$ *(in* $L_2(P_{n,\gamma})$*) and* $h \mapsto P_{n,\gamma,h}$ *(in total variation). The latter equicontinuity condition, in particular, is of interest regarding local uniform regularity and hence local uniformity of size control; cf. Corollaries 3.3, 3.4 and Lemma 3.1 below.*

In the case where the index n corresponds to an increasing sample of i.i.d. data, LAN is satisfied under a pathwise differentiability in quadratic mean condition (e.g. van der Vaart, 1998, Lemma 25.14); this situation will be discussed further in Section 3.4. Whilst LAN is particularly straightforward to establish in this "smooth i.i.d." case, it also holds in other settings and sufficient conditions for LAN applicable to various settings exist in the literature.[20]

C(α) – style test statistics The C(α) – style test statistics proposed in this paper are feasible versions of a quadratic form of d_θ-moment conditions $g_{n,\gamma} \in L_2(P_{n,\gamma})$. In particular, the statistic will be a quadratic form of estimators of the moment conditions weighted by an estimator of the Moore-Penrose pseudo-inverse of their variance matrix. To derive the limiting distribution of the statistic, I impose a high-level joint convergence requirement on the scores and moment functions.

Assumption 3.2 (Joint convergence): *For* d_θ*-dimensional moment conditions* $g_{n,\gamma} \in L_2(P_{n,\gamma})$,

$$\left(\Delta_{n,\gamma}h,\ g_{n,\gamma}'\right)' \overset{P_{n,\gamma}}{\rightsquigarrow} \mathcal{N}\left(0, \Sigma_\gamma(h)\right), \quad \text{for each } h \in H_\gamma,$$

where, for $h = (\tau, b)$,

$$\Sigma_\gamma(h) := \begin{bmatrix} \sigma_\gamma(h) & \tau'\Sigma_{\gamma,21}' \\ \Sigma_{\gamma,21}\tau & V_\gamma \end{bmatrix} = \lim_{n\to\infty} \begin{bmatrix} \|\Delta_{n,\gamma}h\|^2 & \left\langle \Delta_{n,\gamma}(\tau,0)\,,\, g_{n,\gamma}' \right\rangle \\ \langle g_{n,\gamma}\,,\, \Delta_{n,\gamma}(\tau,0)\rangle & \left\langle g_{n,\gamma}\,,\, g_{n,\gamma}' \right\rangle \end{bmatrix}.$$

Built-in to Assumption 3.2 is a requirement of asymptotic orthogonality of the

[20] Cf. McNeney and Wellner (2000), Lemma 1 in Swensen (1985), Chapter 2 in Taniguchi and Kakizawa (2000) and Section 74 of Strasser (1985).

moment functions and scores for the nuisance parameters η. This generalises the explicit orthogonal projection construction discussed around equations (5) – (6).

REMARK 3.3: *For Assumption 3.2 to hold it is necessary that the $g_{n,\gamma}$ are approximately zero mean: since $(g_{n,\gamma})_{n\in\mathbb{N}}$ is uniformly integrable, $P_{n,\gamma}g_{n,\gamma} = o(1)$. It is also necessary that the $g_{n,\gamma}$ satisfy an approximate orthogonality property with the scores for nuisance parameters. In particular, as $([\Delta_{n,\gamma}h]g_{n,\gamma})_{n\in\mathbb{N}}$ is uniformly integrable for each $h = (\tau, b) \in H_\gamma$,*

$$\lim_{n\to\infty} \left\langle \Delta_{n,\gamma}h\, ,\, g'_{n,\gamma} \right\rangle = \tau'\Sigma'_{\gamma,21} = \lim_{n\to\infty} \left\langle \Delta_{n,\gamma}(\tau,0)\, ,\, g'_{n,\gamma} \right\rangle ,$$

and so

$$\left\langle \Delta_{n,\gamma}(0,b)\, ,\, g'_{n,\gamma} \right\rangle = \left\langle \Delta_{n,\gamma}h\, ,\, g'_{n,\gamma} \right\rangle - \left\langle \Delta_{n,\gamma}(\tau,0)\, ,\, g'_{n,\gamma} \right\rangle = o(1). \tag{9}$$

Given any d_θ moment conditions $f_{n,\gamma} \in L_2^0(P_{n,\gamma})$ moment conditions which satisfy an exact version of the orthogonality condition (9) may be obtained as

$$g_{n,\gamma} := \Pi \left[f_{n,\gamma} \middle| \{\Delta_{n,\gamma}(0,b) : b \in B_\gamma\}^\perp \right]. \tag{10}$$

An important special case of this construction is with $f_{n,\gamma}$ equal to the score function for θ. That is, $f_{n,\gamma} = \dot{\ell}_{n,\gamma}$, a vector of functions in $L_2^0(P_{n,\gamma})$ such that $\tau'\dot{\ell}_{n,\gamma} = \Delta_{n,\gamma}(\tau, 0)$ for each $\tau \in \mathbb{R}^{d_\theta}$. In this case, the function $g_{n,\gamma} = \tilde{\ell}_{n,\gamma} := \Pi \left[\dot{\ell}_{n,\gamma} \middle| \{\Delta_{n,\gamma}(0,b) : b \in B_\gamma\}^\perp \right]$ is often called the *efficient score function.*[21] This yields the optimal choice of moment conditions satisfying Assumption 3.2 in the context of power optimality, as shown in Section 3.3.5 below

In order to construct a C(α) – style statistic, I additionally assume that the researcher can estimate $g_{n,\gamma}$ and the pseudo-inverse of V_γ consistently, given θ.

ASSUMPTION 3.3 (Consistent estimation): *$\hat{g}_{n,\theta}$, $\hat{\Lambda}_{n,\theta}$, $\hat{r}_{n,\theta} \in \{0, 1, \ldots, d_\theta\}$ satisfy*

(i) $\hat{g}_{n,\theta} - g_{n,\gamma} \xrightarrow{P_{n,\gamma}} 0$;
(ii) $\hat{\Lambda}_{n,\theta} \xrightarrow{P_{n,\gamma}} V_\gamma^\dagger$;
(iii) If $r := \operatorname{rank}(V_\gamma) \geq 1$, *then* $\hat{r}_{n,\theta} \xrightarrow{P_{n,\gamma}} r$; *if* $r = 0$, *then* $\operatorname{rank}(\hat{\Lambda}_{n,\theta}) \xrightarrow{P_{n,\gamma}} 0$.

Verification of Assumption 3.3 (i) typically proceeds by model specific arguments.[22] One generally applicable approach to obtain an estimator which satisfies

[21] This terminology is used in, for example, Bickel et al. (1998); van der Vaart (1998, 2002). In some other works (e.g. Choi et al., 1996) "effective" is used in place of "efficient".
[22] See pp. 395 – 396 of van der Vaart (1998) for a heuristic discussion of how this condition may

Assumption 3.3(ii) is to take an initial estimator which is consistent for V_γ, threshold its eigenvalues at an appropriate rate and then take the pseudo-inverse.[23] If one uses the estimator $\hat{\Lambda}_{n,\theta} := \hat{V}_{n,\theta}^{\dagger}$ where $\hat{V}_{n,\theta} \xrightarrow{P_{n,\gamma}} V_\gamma$ and $\hat{r}_{n,\theta} := \operatorname{rank}(\hat{V}_{n,\theta})$ then condition (ii) holds if and only if condition (iii) holds (Andrews, 1987, Theorem 2). Nevertheless, as emphasised by the notation, it is not necessary that the estimate $\hat{\Lambda}_{n,\theta}$ be the pseudo-inverse of an initial estimate.

Given the estimators of Assumption 3.3, the C(α) - style test statistic is

$$\hat{S}_{n,\theta} := \hat{g}_{n,\theta}' \hat{\Lambda}_{n,\theta} \hat{g}_{n,\theta}. \tag{11}$$

The "$\hat{S}$ test" will be the C(α) – style test $\psi_{n,\vartheta}$ of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ at level α, defined as

$$\psi_{n,\theta_0} := \mathbf{1}\left\{\hat{S}_{n,\theta_0} > c_n\right\}, \tag{12}$$

where c_n is the $1 - \alpha$ quantile of a $\chi^2_{\hat{r}_n}$ random variable.

3.2 Local regularity

Under the assumptions given so far we have the following result for the asymptotic distribution of the moment conditions $g_{n,\gamma}$ and test statistic.[24]

Proposition 3.1: *Under Assumptions 3.1 and 3.2, for $h = (\tau, b) \in H_\gamma$*

$$g_{n,\gamma} \overset{P_{n,\gamma,h}}{\rightsquigarrow} \mathcal{N}\left(\Sigma_{\gamma,21}\tau, V_\gamma\right).$$

If Assumption 3.3 also holds, then additionally

$$\hat{g}_{n,\theta} \overset{P_{n,\gamma,h}}{\rightsquigarrow} \mathcal{N}\left(\Sigma_{\gamma,21}\tau, V_\gamma\right) \qquad \text{and} \qquad \hat{S}_{n,\theta} \overset{P_{n,\gamma,h}}{\rightsquigarrow} \chi^2_r\left(\tau'\Sigma_{\gamma,21}' V_\gamma \Sigma_{\gamma,21}\tau\right),$$

with $r = \operatorname{rank}(V_\gamma)$.

Pointwise local regularity Based on the preceding proposition standard arguments allow the derivation of the asymptotic rejection probabilities based on the tests $\psi_{n,\theta}$. In particular, Theorem 3.1 demonstrates that the proposed C(α) -

be satisfied based on a Taylor expansion in the case where the estimand is the efficient score function and the observations are i.i.d.. Example 25.61 in van der Vaart (1998) further points out that this condition should be particularly simple to verify in the special case where the dependence on η is linear and the model appropriately convex.

[23] Full details of this approach are given in Appendix section S2.1. Other regularisation schemes are also possible; cf. Lütkepohl and Burda (1997); Dufour and Valéry (2016)

[24] $\chi^2_r(\nu)$ denotes the non-central χ^2 distribution with r degrees of freedom and non-centrality ν.

style test is locally regular in the sense of Definition 2.1.

Theorem 3.1: *Suppose that Assumptions 3.1, 3.2 and 3.3 hold and* $h = (\tau, b) \in H_\gamma$*. Then, if* $r \geq 1$

$$\lim_{n\to\infty} P_{n,\gamma,h}\psi_{n,\theta} = 1 - \mathrm{P}\left(\chi_r^2\left(a\right) \leq c_r\right), \quad a = \tau'\Sigma_{\gamma,21}' V_\gamma^\dagger \Sigma_{\gamma,21}\tau,$$

where c_r *is the* $1-\alpha$ *quantile of the* χ_r^2 *distribution. If, instead,* $r = 0$*, then*

$$\lim_{n\to\infty} P_{n,\gamma,h}\psi_{n,\theta} = 0.$$

Remark 3.4: *Theorem 3.1 shows that the test sequence* $(\psi_{n,\theta})_{n\in\mathbb{N}}$ *is locally regular in the sense of Definition 2 as its local power function* $\pi_n(\tau, b) := P_{n,\gamma,h}\psi_{n,\theta}$ *satisfies*

$$\pi_n(\tau, b) \to \pi(\tau) := \begin{cases} 1 - \mathrm{P}\left(\chi_r^2\left(a\right) \leq c_r\right) & \text{if } r \geq 1 \\ 0 & \text{otherwise} \end{cases}, \qquad a = \tau'\Sigma_{\gamma,21}' V_\gamma^\dagger \Sigma_{\gamma,21}\tau.$$

Since a in Remark 3.4 is equal to zero when $\tau = 0$, i.e. when $h \in H_{\gamma,0}$, the sequence of tests $\psi_{n,\theta}$ is asymptotically of level α under any local perturbation consistent with the null hypothesis. Inverting the C(α) - style test yields confidence sets with analogous coverage properties.

Corollary 3.1: *Suppose that Assumptions 3.1, 3.2 and 3.3 hold. Then* $\psi_{n,\theta}$ *is of asymptotic level* α *for the hypothesis* $H_0 : h \in H_{\gamma,0}$ *against* $H_1 : h \in H_{\gamma,1}$*. In particular, if* $r \geq 1$

$$\lim_{n\to\infty} \pi_n(0, b) = \lim_{n\to\infty} P_{n,\gamma,h}\psi_{n,\theta} = \alpha, \qquad h = (0, b) \in H_{\gamma,0}.$$

If, instead, $r = 0$*, then*

$$\lim_{n\to\infty} \pi_n(0, b) = \lim_{n\to\infty} P_{n,\gamma,h}\psi_{n,\theta} = 0, \qquad h = (0, b) \in H_{\gamma,0}.$$

Corollary 3.2: *Suppose that Assumptions 3.1, 3.2 and 3.3 hold. Define*

$$C_n := \left\{\vartheta \in \Theta : \hat{S}_{n,\vartheta} \leq c_n\right\}.$$

Then, if $r \geq 1$*,*

$$\lim_{n\to\infty} P_{n,\gamma,h}(\theta \in C_n) = 1 - \alpha, \qquad h = (0, b) \in H_{\gamma,0}.$$

If, instead, $r = 0$, then

$$\lim_{n\to\infty} P_{n,\gamma,h}(\theta \in C_n) = 1, \qquad h = (0, b) \in H_{\gamma,0}.$$

Uniform local regularity The limit results in the foregoing section are pointwise in h. These can be extended to limits which hold (locally) uniformly (i.e. uniformly over some subsets K of H_γ) under various conditions. I provide explicit versions of such results for the testing case; analogous results hold for confidence sets.

In order to state these results, some additional structure on H_γ (or $H_{\gamma,0}$ if a uniform version of Corollary 3.1 is all that is desired) is required. One straightforward approach to this is to place a measure structure on H_γ whence uniformity except for on a "small" subset of H_γ holds automatically by Egorov's Theorem, provided $h = (\tau, b) \mapsto \pi_n(\tau, b)$ is measurable. See Appendix section S2.6 for details.

An alternative approach, detailed below, is to work with a (pseudo-)metric structure on H_γ (or $H_{\gamma,0}$). As pointwise convergence of the finite sample (local) power functions is given by Remark 3.4, to "upgrade" this to uniform convergence on compact (or totally bounded) subsets it is (necessary and) sufficient that the (local) power functions are asymptotically equicontinuous.

Corollary 3.3: *Suppose that the conditions of Theorem 3.1 hold and that (H_γ, d) is a pseudometric space. If the functions $h = (\tau, b) \mapsto \pi_n(\tau, b)$ are asymptotically equicontinuous on a compact subset $K \subset H_\gamma$ then*

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

i.e. the sequence of tests $(\psi_{n,\theta})_{n\in\mathbb{N}}$ is locally uniformly regular on K as in Definition 2.2.

This can be specialised to a result concerning only the rejection probability under the null as follows.

Corollary 3.4: *Suppose that the conditions of Corollary 3.1 hold and that $(H_{\gamma,0}, d)$ is a pseudometric space. If the functions $h = (\tau, b) \mapsto \pi_n(\tau, b)$ are asymptotically equicontinuous on a compact subset $K \subset H_{\gamma,0}$ then*

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

I now give two sufficient conditions for the asymptotic equicontinuity requirements of Corollaries 3.3 and 3.4. The first is a trivially sufficient condition but concerns only the measures $P_{n,\gamma,h}$ and is tightly connected with the uniform strengthening of the LAN condition in Assumption 3.1 discussed in Remark 3.2 and Appendix section S2.3.

Lemma 3.1: *If (H_γ, d) (resp. $(H_{\gamma,0}, d)$) is a pseudometric space and $(h \mapsto P_{n,\gamma,h})_{n\in\mathbb{N}}$ is asymptotically equicontinuous in total variation on a subset $K \subset H_\gamma$ (resp. $K \subset H_{\gamma,0}$), then $(h \mapsto P_{n,\gamma,h}\psi_{n,\theta})_{n\in\mathbb{N}}$ is asymptotically equicontinuous on K.*

Remark 3.5: *Application of Lemma 3.1 to Corollary 3.3 (or Corollary 3.4) requires (asymptotic) equicontinuity in total variation of the functions $h \mapsto P_{n,\gamma,h}$ on compact subsets $K \subset H_\gamma$. This holds for any compact K under the uniform LAN (ULAN) condition in Assumption S2.1, as shown in Proposition S2.2.*

In the parametric i.i.d. case LAN is often verified by establishing a differentiability in quadratic mean condition, e.g. equation (7.1) in van der Vaart (1998). As established by Theorem 7.2 of van der Vaart (1998), this suffices to ensure that the remainder $R_{n,\gamma}$ in the LAN expansion (8) satisfies $R_{n,\gamma}(h_n) \to 0$ for any $h_n \to h$. This is sufficient for the ULAN expansion in Assumption S2.1 to hold (with the usual Euclidean metric). Provided the remainder of this condition holds, the asymptotic equicontinuity required by Lemma 3.1 holds for any compact $K \subset H_\gamma$ (Proposition S2.2).

Despite the close link this compact asymptotic equicontinuity in total variation requirement has with the ULAN condition, it is stronger than necessary for the results in Corollaries 3.3, 3.4 and may require a relatively strong pseudometric on H_γ. The following Lemma provides a weaker condition at the expense of a more complicated statement.

Lemma 3.2: *Suppose the conditions of Theorem 3.1 hold and that (H_γ, d) (resp. $(H_{\gamma,0}, d)$) is a pseudometric space. Let δ be metrise weak convergence on the space of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and let $Q_{n,\gamma,h}$ be the pushforward measure of $P_{n,\gamma,h}$ under $\hat{S}_{n,\theta}$. Suppose that on a subset $K \subset H_\gamma$ (resp. $K \subset H_{\gamma,0}$),*

(i) $(h \mapsto Q_{n,\gamma,h})_{n\in\mathbb{N}}$ is asymptotically equicontinuous in δ;
(ii) $(h \mapsto P_{n,\gamma,h}(\hat{r}_{n,\theta} = r))_{n\in\mathbb{N}}$ is asymptotically equicontinuous;
(iii) $(h \mapsto P_{n,\gamma,h}(\hat{\Lambda}_{n,\theta} = 0))_{n\in\mathbb{N}}$ is asymptotically equicontinuous;

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

Remark 3.6: *In Lemma 3.2, Conditions (i) and (ii) are required only in the case where $r \geq 1$ whilst Condition (iii) is required only in the case where $r = 0$. This is evident from inspection of its proof.*

The case with $r = 0$ Throughout the results in this section a distinction has been made between the case where $r = \operatorname{rank}(\tilde{\mathcal{I}}_\gamma) \geq 1$ and where $r = 0$. This distinction is not artificial: it appears in the proofs of the above results, which require different arguments for the case with $r = 0$, and will resurface in the subsequent section on power bounds. As will be presented there, when the moment conditions $g_{n,\gamma}$ are chosen optimally and $r = 0$ the model contains no information (asymptotically) about deviations from the null in *any* direction τ. This is in contrast to the "intermediate" case where $0 < r < d_\theta$. In such a case, whilst no locally regular estimator of θ can exist (Chamberlain, 1986), the model does contain information about deviations from the null in certain directions τ, which can be exploited by C(α) - style tests of the proposed form.

3.3 Local asymptotic power bounds

The preceding section established the local regularity of C(α) - style tests based on moment functions $g_{n,\gamma}$ satisfying certain asymptotic orthogonality conditions. Thus far, nothing has been said about the choice of $g_{n,\gamma}$ beyond these orthogonality requirements.

The choice of the functions $g_{n,\gamma}$ is fundamentally what determines the attainable power of the corresponding test. As such, they ought to be chosen such that the resulting test has good power against alternatives of interest. One natural choice is the efficient score function:

$$\tilde{\ell}_{n,\gamma} := \Pi\left[\dot{\ell}_{n,\gamma}\middle|\{\Delta_{n,\gamma}(0,b) : b \in B_\gamma\}^\perp\right] \tag{13}$$

where $\dot{\ell}_{n,\gamma}$ are elements of $L_2^0(P_{n,\gamma})$ such that $\Delta_{n,\gamma}(\tau, 0) = \dot{\ell}_{n,\gamma}'\tau$ for each $\tau \in \mathbb{R}^{d_\theta}$.[25] It is well known that tests based on the efficient score function have certain optimality properties in regular models when the observations are (a) i.i.d. (cf. Section 25.6 van der Vaart, 1998) or (b) when the information operator is boundedly invertible Choi et al. (1996). I show below that this result holds without requiring (a) or (b). Moreover, a generalised version of this result continues to hold for the class of non-regular models considered in this paper.

[25] That is, if $e_1, \ldots, e_{d_\theta}$ are the canonical basis vectors in $\mathbb{R}^{d_\theta}$, the i-th element of $\dot{\ell}_{n,\gamma}$ is $\Delta_{n,\gamma}(e_i, 0)$.

In particular, under conditions on the limit variance matrix V_γ which appears in Assumption 3.2, I show that,

(i) If $\theta \in \mathbb{R}$, $\psi_{n,\theta}$ achieves the local asymptotic power bound for uniformly most powerful asymptotically unbiased tests;
(ii) If $\theta \in \mathbb{R}^{d_\theta}$ and $1 \leq r \leq d_\theta$, $\psi_{n,\theta}$ posseses a local asymptotic maximin optimality property and a local asymptotic minimal regret property.

(i) is essentially classical as $d_\theta = 1$ implies that $r \in \{0, 1\}$ and there is no "intermediate" case between the regular and fully degenerate cases. This case is included partially for completeness given the importance of testing a scalar parameter in practical applications and partially because the required conditions are weaker than in other treatments in the literature, as noted above.

(ii) generalises the classical results on (local asymptotic) maximin optimality and minimal regret to encompass also the non – regular case. In the (regular) case where $r = d_\theta$, the classical results are recovered. The results here establish that the test is a good choice if the researcher does not have particular alternatives in mind against which they wish to direct power.[26]

I also formally demonstrate the unsurprising result that if $r = 0$, then no test with correct asymptotic size has non-trivial asymptotic power against any sequence of local alternatives.

The results in this section are derived using the limits of experiments framework of Le Cam (e.g. Strasser, 1985; Le Cam, 1986; Le Cam and Yang, 2000; van der Vaart, 1998). In particular, I show that the local experiments consisting of the measures $P_{n,\gamma,h}$ for $h \in H_\gamma$ converge weakly to a limit experiment which has a close relationship to a Gaussian shift experiment on the Hilbert space formed by taking the quotient of H_γ under the seminorm induced by the variance function $\sigma(h)$. The relation between these experiments is sufficiently tight that power bounds derived in the latter transfer to the former.[27]

The inner – product structure For this development the space H_γ is required to be linear and I will therefore assume that B_γ (and hence H_γ) is a linear space.[28]

[26] If the researcher does have particular alternatives in mind, tests can be constructed which direct power towards these alternatives (cf. Bickel, Ritov, and Stoker, 2006).

[27] That the local experiments do not converge to the mentioned Gaussian shift experiment is essentially a purely technical point: the Gaussian shift experiment is defined on a different parameter space to the local experiments, whilst (weak) convergence of experiments in the sense of Le Cam (1986) is defined for experiments with the same parameter space.

[28] In the preceding sections H_γ was required only to be a subset of a linear space containing the zero vector.

Then, under LAN, there exists a positive semi-definite symmetric bilinear form $\langle \cdot \, , \, \cdot \rangle_\gamma$ on $H_\gamma = \mathbb{R}^{d_\theta} \times B_\gamma$ such that

$$\sigma_\gamma(h) = \langle h \, , \, h \rangle_\gamma \ .$$

This can be seen as a by-product of the following Lemma.

LEMMA 3.3: *Suppose Assumption 3.1 holds and B_γ is a linear space. Let Δ_γ be the square integrable stochastic process defined on H_γ such that*

$$\Delta_{n,\gamma} h \overset{P_{n,\gamma}}{\rightsquigarrow} \Delta_\gamma h.$$

Then Δ_γ is a mean-zero Gaussian linear process with covariance kernel K_γ, where

$$K_\gamma(h, g) \coloneqq \lim_{n\to\infty} P_{n,\gamma} \left[\Delta_{n,\gamma} h \Delta_{n,\gamma} g\right].$$

For $h, g \in H_\gamma$, setting $\langle h \, , \, g \rangle_\gamma \coloneqq K_\gamma(h, g)$, where K_γ is the covariance kernel of Δ_γ yields a positive semi-definite symmetric bilinear form. Let $\| \cdot \|_\gamma$ denote the seminorm induced by $\langle \cdot \, , \, \cdot \rangle_\gamma$ on H_γ and note that with this definition $\|h\|_\gamma^2 = \sigma_\gamma(h)$.

REMARK 3.7: *Suppose that $\langle \cdot \, , \, \cdot \rangle$ is an inner product on $H_\gamma = \mathbb{R}^{d_\theta} \times B_\gamma$. The existence of the positive semi-definite symmetric bilinear form $\langle \cdot \, , \, \cdot \rangle_\gamma$ is equivalent to the existence of a bounded, self-adjoint, positive semi-definite linear operator B_γ such that $\langle h \, , \, h \rangle_\gamma = \langle h \, , \, \mathsf{B}_\gamma h \rangle$ for $h \in H_\gamma$ (cf. Choi et al., 1996, p. 845).*

With this established, define $\mathbb{H}_\gamma$ as the quotient of H_γ by the subspace on which the semi-norm $\| \cdot \|_\gamma$ vanishes:

$$\mathbb{H}_\gamma \coloneqq H_\gamma \, / \, \{h \in H_\gamma : \|h\|_\gamma = 0\}. \tag{14}$$

which is an inner product space when equipped with the natural inner product induced by $\langle \cdot \, , \, \cdot \rangle_\gamma$, which I also denote by $\langle \cdot \, , \, \cdot \rangle_\gamma$. Elements of the quotient space $\mathbb{H}_\gamma$ are not elements of H_γ but sets of such elements ("cosets"). To emphasise this distinction, often the coset corresponding to a representative element $h \in H_\gamma$ is denoted by $[h]$, a convention which is followed here. Further details on this construction are given in section S2.2 of the supplementary material.[29]

[29] Analogous comments apply to the related space $\mathbb{H}_{\gamma,1}$, defined below. In both cases, to avoid an excess of parentheses / brackets, if $h = (\tau, b)$ I will write either $[h]$ or $[\tau, b]$, rather than $[(\tau, b)]$.

The limit experiment The weak limit of the sequence of experiments consisting of the measures $P_{n,\gamma,h}$ can be obtained by standard results on weak convergence of experiments.

Proposition 3.2: *Suppose that Assumption 3.1 holds and that B_γ is a linear space and define the sequence of experiments*

$$\mathscr{E}_{n,\gamma} := (\mathcal{W}_n, \mathcal{B}(\mathcal{W}_n), (P_{n,\gamma,h} : h \in H_\gamma)) .$$

Let Δ_γ be the Gaussian process defined in Lemma 3.3 and let $(\Omega, \mathcal{F}, \mathrm{P})$ be the probability space on which it is defined. Define the experiment $\mathscr{E}_\gamma := (\Omega, \mathcal{F}, (P_{\gamma,h} : h \in H_\gamma))$ according to

$$P_{\gamma,0} := \mathrm{P}, \qquad \frac{\mathrm{d}P_{\gamma,h}}{\mathrm{d}P_{\gamma,0}} = \exp\left(\Delta_\gamma h - \frac{1}{2}\|h\|_\gamma^2\right).$$

Then $\mathscr{E}_{n,\gamma}$ converges weakly to $\mathscr{E}_\gamma$.

Under the additional assumption that $\mathbb{H}_\gamma$ is separable, this limiting experiment is, at least for the purpose of testing, essentially equivalent to a Gaussian shift on $(\mathbb{H}_\gamma, \langle\cdot\,,\,\cdot\rangle_\gamma)$, in the sense given by the Proposition below.

Assumption 3.4: *B_γ is a linear space and $\mathbb{H}_\gamma$ as defined in (14) is separable.*

Proposition 3.3: *Suppose that Assumptions 3.1 and 3.4 hold. If $\mathscr{E}_\gamma$ is the experiment defined in Proposition 3.2, there is a Gaussian shift experiment $\mathscr{G}_\gamma := (\Omega, \mathcal{F}, (G_{[h]} : [h] \in \mathbb{H}_\gamma))$ such that*

$$d_{TV}(P_{\gamma,h}, G_{[h]}) = 0, \quad h \in H_\gamma.$$

The efficient information matrix I next define the *efficient information matrix*, $\tilde{\mathcal{I}}_\gamma$, via an orthogonal projection in the Hilbert space $\overline{\mathbb{H}_\gamma}$, the completion of $\mathbb{H}_\gamma$. $\tilde{\mathcal{I}}_\gamma$ determines the power bounds for tests of $\tau = 0$. In the i.i.d. case $\tilde{\mathcal{I}}_\gamma$ is the covariance matrix of the efficient score function for a single observation (as shown in Section 3.4 below) and thus this definition coincides with the usual one (cf. van der Vaart, 1998, Section 25.4).

Let $\|\tau\| := \inf_{b \in B_\gamma} \|(\tau, b)\|_\gamma$, which defines a semi-norm on $\mathbb{R}^{d_\theta}$. Equipping the quotient $\mathbb{H}_{\gamma,1} := \mathbb{R}^{d_\theta} / \{\tau \in \mathbb{R}^{d_\theta} : \|\tau\| = 0\}$ with the natural norm induced by $\|\cdot\|$ (which I will also denote by $\|\cdot\|$) turns it into a normed space.[30] Define

[30]See Section S2.2 for further details on this construction.

the linear map $\pi_1 : \mathbb{H}_\gamma \to \mathbb{H}_{\gamma,1}$ as $\pi_1([\tau, b]) := [\tau]$. π_1 is continuous: suppose that $[h_n] = [\tau_n, b_n] \to [0]$. Then $\pi_1[h_n] = [\tau_n] \to [0] = \pi_1[0]$, since $\|[\tau_n]\| = \inf_{b \in B_\gamma} \|(\tau_n, b)\|_\gamma \to 0$. As such there is a unique continuous extension of π_1 to $\overline{\mathbb{H}_\gamma}$, which will henceforth also be called π_1.

Since π_1 is continuous, its kernel $\ker \pi_1 = \pi_1^{-1}(\{[0]\})$ is closed. Let Π be the orthogonal projection onto $\ker \pi_1$ and define $\Pi^\perp := I - \Pi$, the orthogonal projection onto $[\ker \pi_1]^\perp$. By the Pythagorean theorem

$$\|[h]\|_\gamma^2 = \|\Pi^\perp[h]\|_\gamma^2 + \|\Pi[h]\|_\gamma^2.$$

Let e_i be the i-th canonical basis vector in $\mathbb{R}^{d_\theta}$ and define the *efficient information matrix* $\tilde{\mathcal{I}}_\gamma$ as the $d_\theta \times d_\theta$ matrix with i, j-th entry $\tilde{\mathcal{I}}_{\gamma,ij}$ given by

$$\tilde{\mathcal{I}}_{\gamma,ij} = \left\langle \Pi^\perp[e_i, 0]\,,\, \Pi^\perp[e_j, 0] \right\rangle_\gamma .$$

The next Lemma records the relationship between $\tilde{\mathcal{I}}_\gamma$ and (a) the subspace $\{\tau \in \mathbb{R}^{d_\theta} : \|\tau\| = 0\}$, (b) the norm on $\mathbb{H}_{\gamma,1}$.

LEMMA 3.4: *Under Assumption 3.4,* $\|\tau\|^2 = \tau' \tilde{\mathcal{I}}_\gamma \tau$ *and* $\ker \tilde{\mathcal{I}}_\gamma = \{\tau \in \mathbb{R}^{d_\theta} : \|\tau\| = 0\}$.

An alternative expression for $\tilde{\mathcal{I}}_\gamma$, based on the limiting Gaussian process of Lemma 3.3, is given in the following Lemma.

LEMMA 3.5: *Suppose that Assumption 3.1 holds, B_γ is a linear space and suppose that the Gaussian process Δ_γ as defined in Lemma 3.3 is defined on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Let $\mathcal{T} := \{\Delta_\gamma(h) : h = (0, b) \in H_\gamma\} \subset L_2(\mathrm{P})$. For e_i the i-th canonical basis vector in $\mathbb{R}^{d_\theta}$, let $\tilde{\Delta}_\gamma(e_i, 0) := \Pi\left[\Delta_\gamma(e_i, 0) \middle| \mathcal{T}^\perp\right]$. Then,*

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

With this setup local asymptotic power bounds can be readily obtained via known results for the Gaussian shift experiment $\mathscr{G}_\gamma$.

3.3.1 Two-sided tests of a scalar parameter

The following Theorem records the power bound for (locally asymptotically) unbiased two-sided tests of a scalar θ. As previously mentioned, in the case where $d_\theta = 1$, the matrix $\tilde{\mathcal{I}}_\gamma$ has rank either 0 or 1 and there is no intermediate case. Theorem 3.2 handles both cases simultaneously.

THEOREM 3.2: *Suppose that assumptions 3.1 and 3.4 hold and* $d_\theta = 1$. *Let* $\phi_n : \mathcal{W}_n \to [0,1]$ *be a sequence of locally asymptotically unbiased level* α *tests of* $H_0 : \tau = 0$ *against* $H_1 : \tau \neq 0$. *That is,*

$$\limsup_{n\to\infty} P_{n,\gamma,h}\phi_n \le \alpha, \qquad h \in H_{\gamma,0}$$

and

$$\liminf_{n\to\infty} P_{n,\gamma,h}\phi_n \ge \alpha, \qquad h \in H_{\gamma,1}.$$

Then, for any $h \in H_\gamma$,

$$\limsup_{n\to\infty} P_{n,\gamma,h}\phi_n \le 1 - \Phi\left(z_{\alpha/2} - \tilde{\mathcal{I}}_\gamma^{1/2}\tau\right) + 1 - \Phi\left(z_{\alpha/2} + \tilde{\mathcal{I}}_\gamma^{1/2}\tau\right),$$

where z_α *is the* $1-\alpha$ *quantile and* Φ *the CDF of the standard normal distribution.*

That the two-sided power bound of Theorem 3.2 is achieved by the test $\psi_{n,\theta}$ provided $\Sigma_{\gamma,21}V_\gamma^\dagger\Sigma_{\gamma,21} = \tilde{\mathcal{I}}_\gamma$ and $r = 1$ is an immediate consequence of Theorem 3.1.[31]

COROLLARY 3.5: *Suppose that assumptions 3.1, 3.2 and 3.3 hold with* $\Sigma_{\gamma,21}V_\gamma^\dagger\Sigma_{\gamma,21}' = \tilde{\mathcal{I}}_\gamma$ *and* $r = 1$. *Then, for* $h \in H_\gamma$,

$$\lim_{n\to\infty} P_{n,\gamma,h}\psi_{n,\theta} = 1 - \Phi\left(z_{\alpha/2} - \tilde{\mathcal{I}}_\gamma^{1/2}\tau\right) + 1 - \Phi\left(z_{\alpha/2} + \tilde{\mathcal{I}}_\gamma^{1/2}\tau\right).$$

3.3.2 Tests for multivariate parameters

Unlike in the scalar case, when $d_\theta > 1$ there is a truly intermediate case where: $0 < \operatorname{rank}(\tilde{\mathcal{I}}_\gamma) < d_\theta$. Here I permit $0 < \operatorname{rank}(\tilde{\mathcal{I}}_\gamma) \le d_\theta$ and establish two results each of which contains the corresponding full rank case as a special case ($\operatorname{rank}(\tilde{\mathcal{I}}_\gamma) = d_\theta$). The first is a maximin power bound for potentially non − regular models, which shows that the local asymptotic maximin power over $h = (\tau, b)$ with $\tau'\tilde{\mathcal{I}}_\gamma\tau \ge a$ of any test of $\tau = 0$ which is asymptotically of level α is bounded above by $1 - P(\chi_r^2(a) \le c_r)$. In the case that $\tau \in \ker \tilde{\mathcal{I}}_\gamma$, one has that $\tau'\tilde{\mathcal{I}}_\gamma\tau = 0$ and the result demonstrates that no test can have non-trivial local asymptotic maximin power against such alternatives. Following this, I establish a related result: the most stringent test (in the sense of Wald, 1943) in the limit experiment has the same power function as the maximin test, and no sequence of asymptotic level α tests can correspond to a test in the limit experiment with smaller regret (as

[31] Typically the moment conditions $g_{n,\gamma}$ will be chosen such that $\Sigma_{\gamma,21} = V_\gamma = \tilde{\mathcal{I}}_\gamma$ in order to achieve this equality. See Section 3.3.5 below.

defined in equation (18) below). I give conditions under which the test sequence $\psi_{n,\theta}$ attains these power bounds.

Maximin optimal testing

THEOREM 3.3: *Suppose that assumptions 3.1 and 3.4 hold and* $r := \operatorname{rank}(\tilde{\mathcal{I}}_\gamma) \geq 1$. *Let* $\phi_n : \mathcal{W}_n \to [0,1]$ *be a sequence of tests such that for each* $h = (0, b) \in H_{\gamma,0}$

$$\limsup_{n\to\infty} P_{n,\gamma,h}\phi_n \leq \alpha$$

Let c_r *the* $1-\alpha$ *quantile of a* χ^2_r *random variable. Then, if* $a \geq 0$,

$$\limsup_{n\to\infty} \inf \left\{ P_{n,\gamma,h}\phi_n : h = (\tau, b) \in H_\gamma,\ \tau'\tilde{\mathcal{I}}_\gamma\tau \geq a \right\} \leq 1 - \mathrm{P}(\chi^2_r(a) \leq c_r). \quad (15)$$

Similarly to the two-sided case, that the power bound on the right hand side of (15) is achieved by the test $\psi_{n,\theta}$ provided $\Sigma_{\gamma,21}V_\gamma^\dagger\Sigma_{\gamma,21} = \tilde{\mathcal{I}}_\gamma$ and $\operatorname{rank}(V_\gamma) = \operatorname{rank}(\tilde{\mathcal{I}}_\gamma) = r \geq 1$ is a consequence of Theorem 3.1.[32] In order that the test be asymptotically maximin over a compact subset K_a of $\{h = (\tau, b) \in H_\gamma : \tau'\tilde{\mathcal{I}}_\gamma\tau \geq a\}$, with $a = \inf\{\tau'\tilde{\mathcal{I}}_\gamma\tau = a : h \in K_a\}$, some uniformity (and hence additional structure) is required. In particular, I suppose that (H_γ, d) is a pseudometric space for some pseudometric d. I emphasise that this pseudometric need not be related to $\|\cdot\|_\gamma$ as defined just preceding Remark 3.7:

REMARK 3.8: *The pseudometric* d *placed on* H_γ *in the second part of Corollary 3.6 and in Lemmas 3.1, 3.6 need not be related to the seminorm* $\|\cdot\|_\gamma$.

COROLLARY 3.6: *Suppose that assumptions 3.1, 3.2 and 3.3 hold with* $\Sigma_{\gamma,21}V_\gamma^\dagger\Sigma'_{\gamma,21} = \tilde{\mathcal{I}}_\gamma$ *and* $r = \operatorname{rank}(\tilde{\mathcal{I}}_\gamma) = \operatorname{rank}(V_\gamma) \geq 1$. *Then for* $h = (\tau, b) \in H_\gamma$

$$\lim_{n\to\infty} P_{n,\gamma,h}\psi_{n,\theta} = 1 - \mathrm{P}\left(\chi^2_r(a) \leq c_r\right), \quad a = \tau'\tilde{\mathcal{I}}_\gamma\tau. \quad (16)$$

Additionally, suppose that (H_γ, d) *is a metric space and let* K_a *be a compact subset of* $\{h = (\tau, b) \in H_\gamma : \tau'\tilde{\mathcal{I}}_\gamma\tau \geq a\}$ *such that* $a = \inf\{\tau'\tilde{\mathcal{I}}_\gamma\tau : h = (\tau, b) \in K_a\}$. *If the functions* $h \mapsto P_{n,\gamma,h}\psi_{n,\theta}$ *are asymptotically equicontinuous on* K_a,

$$\lim_{n\to\infty} \inf_{h\in K_a} P_{n,\gamma,h}\psi_{n,\theta} = 1 - \mathrm{P}\left(\chi^2_r(a) \leq c_r\right). \quad (17)$$

A sufficient condition for the asymptotic equicontinuity required for the second

[32] See footnote 31.

part of Corollary 3.6 based on an asymptotic equicontinuity in total variation requirement was given as Lemma 3.1 in the previous section. A version based on the weaker conditions used in Lemma 3.2 is given below, adapted to the present context (cf. Remark 3.6).

LEMMA 3.6: *Suppose the conditions of the first part of Corollary 3.6 hold and that* (H_γ, d) *is a pseudometric space. Let* δ *be any metric on the space of probability measures which metrises weak convergence and let* $Q_{n,\gamma,h}$ *be the pushforward measure of* $P_{n,\gamma,h}$ *under* $\hat{S}_{n,\theta}$*. Suppose that on a compact subset* $K \subset H_\gamma$*,*

(i) the functions $h \mapsto Q_{n,\gamma,h}$ *are asymptotically equicontinuous in* δ*;*
(ii) the functions $h \mapsto P_{n,\gamma,h}(\hat{r}_{n,\theta} = r)$ *are asymptotically equicontinuous,*

then $h \mapsto P_{n,\gamma,h}\psi_{n,\theta}$ *are asymptotically equicontinuous on* K*.*

Most stringent tests The last power optimality concept I consider is based on the concept of stringency (due to Wald, 1943) and delivers a similar message.[33]

Let $\mathcal{C}$ be the class of all tests of level α for the hypothesis $K_0 : \tau = 0$ against $K_1 : \tau \neq 0$ in the experiment $\mathscr{E}_\gamma$. That is, if $\phi \in \mathcal{C}$ then $P_{\gamma,h}\phi \leq \alpha$ for all $h = (0, b) \in H_{\gamma,0}$. Define $\pi^\star(h) := \sup_{\phi \in \mathcal{C}} P_{\gamma,h}\phi$ for all $h \in H_{\gamma,1}$ and define the *regret* of a test $\phi \in \mathcal{C}$ as

$$R(\phi) := \sup \{\pi^\star(h) - P_{\gamma,h}\phi : h \in H_{\gamma,1}\}. \tag{18}$$

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

THEOREM 3.4: *Suppose that Assumptions 3.1 and 3.4 hold and* $r = \operatorname{rank}(\tilde{\mathcal{I}}_\gamma) \geq 1$*. The most stringent level* α *test of* $K_0 : \tau = 0$ *against* $K_1 : \tau \neq 0$ *in* $\mathscr{E}_\gamma$*,* ψ*, has power function*

$$\pi(h) := P_{\gamma,h}\psi = 1 - \mathrm{P}(\chi^2_r(a) \leq c_r), \qquad a = \tau'\tilde{\mathcal{I}}_\gamma\tau,\ h = (\tau, b). \tag{19}$$

The first part of Corollary 3.6 provides conditions under which this is the asymptotic power of $\psi_{n,\theta}$ under the sequence of local alternatives $P_{n,\gamma,h}$. The following Proposition demonstrates that if $\pi_n : H_\gamma \to [0, 1]$ is a sequence of power functions corresponding to tests in the experiments $\mathscr{E}_{n,\gamma}$ of asymptotic size α, then

[33] The development here is based on Section 9, Chapter 11 in Le Cam (1986); in particular compare Theorem 3.4 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.

each cluster point of π_n corresponds to a test ϕ in the limit experiment $\mathscr{E}_\gamma$ whose regret is bounded below by that of the most stringent test, ψ.[34]

Proposition 3.4: *Suppose that Assumptions 3.1 and 3.4 hold and that* $r = \operatorname{rank}(\tilde{\mathcal{I}}_\gamma) \geq 1$*. Let* $\phi_n : \mathcal{W}_n \to [0,1]$ *be a sequence of tests such that for each* $h = (0, b) \in H_{\gamma,0}$*,*

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

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

3.3.3 The degenerate case

Finally, I record a negative, if unsurprising, result. If the efficient information matrix $\tilde{\mathcal{I}}_\gamma$ is zero, no test with correct asymptotic size has non – trivial asymptotic power against any sequence of local alternatives.

Proposition 3.5: *Suppose that assumptions 3.1 and 3.4 hold and* $r := \operatorname{rank}(\tilde{\mathcal{I}}_\gamma) = 0$*. Let* $\phi_n : \mathcal{W}_n \to [0,1]$ *be a sequence of tests such that for each* $h = (0, b) \in H_{\gamma,0}$

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

Then for $h \in H_\gamma$*,*

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

3.3.4 Discussion of the power bounds

There are a number of important aspects to highlight regarding the interpretation of the power bounds obtained in the preceding subsections.

Optimality in multivariate testing problems Just as in the classical finite – dimensional case, the maximin optimality and stringency results presented across Theorems 3.3, 3.4, Corollary 3.6 and Proposition 3.4 should not be taken in an absolute sense. Nevertheless, they seem reasonable if the researcher does not have directions against which they wish to direct power a priori. If there are alternatives of particular interest, then one could construct a locally regular test by utilising

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

the same moment conditions $g_{n,\gamma}$ but weighting them differently, similarly to as in Bickel et al. (2006).

The intermediate case with $1 \leq \operatorname{rank}(\tilde{\mathcal{I}}_\gamma) < d_\theta$ A key benefit of these multivariate power results is that they apply equally to cases where the efficient information matrix has reduced rank. Such a scenario can occur for various reasons. Firstly the model may simply not identify all parameters of interest θ (e.g. underidentification). Secondly some (but not all) of the elements of θ may be weakly identified (e.g. weak underidentification). The power results above apply in either of these cases.

There are a number of other papers which provide inference results in similarly rank deficient settings, including Rotnitzky, Cox, Bottai, and Robins (2000); Han and McCloskey (2019); Andrews and Guggenberger (2019); Amengual, Bei, and Sentana (2023). Unlike the present paper, none of these papers consider optimal testing in this setting.

Alternative approximations In the case where $\operatorname{rank}(\tilde{\mathcal{I}}_\gamma) = 0$, Proposition 3.5 reveals that the LAN approximation in 3.1 is, in a certain sense, the wrong approximation: it does not provide any useful way of (asymptotically) comparing tests, whilst other approximations might provide valuable comparisons.

Alternative approximations have been explored in, for example, the IV model (e.g. Moreira, 2009) and semiparametric GMM models by Andrews and Mikusheva (2022, 2023). For example, in the IV case Moreira (2009) considers alternatives which are at a fixed distance from the true parameter, rather than in a shrinking $\sqrt{n}$-neighbourhood. Whether such an approach can be generalised to general semiparametric models is an interesting question for future research.

3.3.5 Attaining the power bounds

I now demonstrate that provided that $g_{n,\gamma}$ is equal to the efficient score function, $\tilde{\ell}_{n,\gamma}$ as defined in (13) up to an error which vanishes in mean – square, the $C(\alpha)$ – style test based on $g_{n,\gamma}$ attains the power bounds established in Section 3.3. This result is well known in two special cases: (i) the regular i.i.d. case (cf. Section 25.6 in van der Vaart (1998); Corollary 3.8 below) and (ii) when the information operator B_γ in Remark 3.7 is positive – definite with the information operator for η, $\mathsf{B}_{\gamma,22}$, boundedly invertible (Choi et al., 1996). Here I provide a general version of the result which does not require (i) or (ii). In particular, I show that

$\Sigma_{\gamma,21}V_\gamma^\dagger\Sigma_{\gamma,21} = \tilde{\mathcal{I}}_\gamma$, which suffices given Theorem 3.1 and the power bounds in Theorems 3.2, 3.3 and 3.4.

THEOREM 3.5: *Suppose that Assumptions 3.1, 3.2, 3.3 and 3.4 hold and that* $g_{n,\gamma} = \tilde{\ell}_{n,\gamma}$. *Then* $\Sigma_{\gamma,21} = V_\gamma = \tilde{\mathcal{I}}_\gamma$, *hence* $\Sigma_{\gamma,21}V_\gamma^\dagger\Sigma_{\gamma,21} = \tilde{\mathcal{I}}_\gamma$.

REMARK 3.9: *More generally, if Assumptions 3.1, 3.2, 3.3, 3.4 hold and* $\lim_{n\to\infty} P_{n,\gamma}\left\|g_{n,\gamma} - \tilde{\ell}_{n,\gamma}\right\|^2 = 0$, *then*

$$\lim_{n\to\infty}\left\langle \dot{\ell}_{n,\gamma}\,,\, g'_{n,\gamma}\right\rangle = \Sigma_{\gamma,21} = \tilde{\mathcal{I}}_\gamma = V_\gamma = \lim_{n\to\infty}\left\langle g_{n,\gamma}\,,\, g'_{n,\gamma}\right\rangle,$$

hence $\Sigma_{\gamma,21}V_\gamma^\dagger\Sigma_{\gamma,21} = \tilde{\mathcal{I}}_\gamma$.

3.4 The smooth i.i.d. case

In this section we provide lower level conditions which are sufficient for some of the high level conditions in the in the benchmark case for semiparametric theory: where our data observations are i.i.d. and the model is "smooth". The discussion here is intended to (a) demonstrate that the results of the foregoing section apply to a large range of semiparametric models that are used in practice and (b) facilitate the application of these results. It should not be interpreted to suggest that the results of the foregoing section do not apply to situations with dependent or non-identically distributed data. They often do, though sufficient conditions may be more complex to verify.

ASSUMPTION 3.5 (Product measures): *Suppose that* $W^{(n)} = (W_1, \ldots, W_n) \in \prod_{i=1}^n \mathcal{W} = \mathcal{W}_n$ *and that each of the probability measures* $P_{n,\gamma,h}$ *is product measure:* $P_{n,\gamma,h} = P_{\gamma,h}^n$. *Each measure in* $\{P_\gamma : \gamma \in \Gamma\}$ *is dominated by a σ-finite measure* ν.

In this i.i.d. setting, it is well known that quadratic mean differentiability of the square root of the density $p_\gamma = \frac{dP_\gamma}{d\nu}$ is sufficient for LAN. In particular, if

$$\lim_{n\to\infty}\int\left[\sqrt{n}\left(\sqrt{p_{\gamma,h_n}} - \sqrt{p_\gamma}\right) - \frac{1}{2}A_\gamma h\sqrt{p_\gamma}\right]^2 = 0, \tag{20}$$

for a measurable $A_\gamma h : \mathcal{W} \to \mathbb{R}$, then with $\Delta_{n,\gamma}h := \frac{1}{\sqrt{n}}\sum_{i=1}^n [A_\gamma h](W_i)$ the remainder term $R_{n,\gamma}$ in the LAN expansion satisfies $R_{n,\gamma}(h_n) \xrightarrow{P_\gamma} 0$ (see e.g. van der Vaart and Wellner, 1996, Lemma 3.10.11). This can be used to establish either the LAN condition required by Assumption 3.1 by taking $h_n = h$ for each $n \in \mathbb{N}$ or the ULAN condition as in Assumption S2.1 by considering convergent sequences $h_n \to h$. Sufficient conditions for (20) (at least with $h_n = h$) are well known: see e.g. Lemma 7.6 in van der Vaart (1998).

In this case, the variables $A_\gamma h$ typically take the form

$$[A_\gamma h](W_i) = \tau' \dot{\ell}_\gamma(W_i) + [D_\gamma b](W_i), \quad h = (\tau, b) \in H_\gamma, \tag{21}$$

where $\dot{\ell}_\gamma$ is a vector of functions in $L_2^0(P_\gamma)$ (typically the partial derivatives of $\theta \mapsto \log p_\gamma$ at γ) and $D_\gamma : \overline{\text{lin}}\ B_\gamma \to L_2^0(P_\gamma)$ a bounded linear map. Showing that condition (20) holds (with $h_n = h$) is typically the most straightforward way to verify the quadratic approximation to the log likelihood required by Assumption 3.1. If $A_\gamma : \overline{\text{lin}}\ H_\gamma \to L_2(P_\gamma)$ is a bounded linear map, then the remainder of Assumption 3.1 also follows directly.

Lemma 3.7: *Suppose that Assumption 3.5 holds and for each $h \in H_\gamma$ equation (20) holds (with $h_n = h$) with $A_\gamma : \overline{\text{lin}}\ H_\gamma \to L_2(P_\gamma)$ a bounded linear map. Then Assumption 3.1 holds with $P_{n,\gamma,h} = P^n_{\gamma, h/\sqrt{n}}$ and $[\Delta_{n,\gamma} h](W^{(n)}) = \mathbb{G}_n A_\gamma h$.*

When the data are i.i.d., the the joint convergence of $(\Delta_{n,\gamma} h, g'_{n,\gamma})$ as in Assumption 3.2 is particularly straightforward to verify. As noted in the discussion around (10), it can be ensured that the orthogonality condition holds exactly, by performing an orthogonal projection. The convergence required by Assumption 3.2 then follows straightforwardly. As with $[\Delta_{n,\gamma} h](W^{(n)}) = \mathbb{G}_n A_\gamma h$, in the i.i.d. setting typically $g_{n,\gamma}$ will have the form $g_{n,\gamma}(W^{(n)}) = \mathbb{G}_n g_\gamma$.

Lemma 3.8: *Suppose that Assumptions 3.5 and 3.1 hold, with $\Delta_{n,\gamma} h = \frac{1}{\sqrt{n}} \sum_{i=1}^n A_\gamma h$, where $A_\gamma h$ is as in equation (21). Additionally suppose that $g_\gamma \in \{D_\gamma b : b \in B_\gamma\}^\perp \subset L_2^0(P_\gamma)$. Then Assumption 3.2 holds with $g_{n,\gamma}(W^{(n)}) \coloneqq \mathbb{G}_n g_\gamma$.*

Corollary 3.7: *In the setting of Lemma 3.8, if $f_\gamma \in L_2^0(P_\gamma)$ and g_γ is the orthogonal projection of f_γ onto the orthogonal complement of $\{D_\gamma b : b \in B_\gamma\} \subset L_2(P_\gamma)$, then Assumption 3.2 holds with $g_{n,\gamma}(W^{(n)}) \coloneqq \mathbb{G}_n g_\gamma$.*

In this i.i.d. setting, the power bounds of section 3.3 can be attained by choosing g_γ as the efficient score for a single observation $\tilde{\ell}_\gamma$. This follows from the following corollary which shows that $\Sigma_{\gamma,21} = V_\gamma = \tilde{\mathcal{I}}_\gamma$, hence $\Sigma_{\gamma,21} V_\gamma^\dagger \Sigma_{\gamma,21} = \tilde{\mathcal{I}}_\gamma$ as required for the C(α)-style test based on $g_{n,\gamma}$ to obtain the power bounds.

Corollary 3.8: *In the setting of Corollary 3.8, if $f_\gamma = \dot{\ell}_\gamma$, then $\Sigma_{\gamma,21} = V_\gamma = \tilde{\mathcal{I}}_\gamma$.*

4 Examples

I will now illustrate the application of the general results in Section 3 to the single index and IV models (introduced as Examples 2.1 and 2.2 respectively). In each example I construct a locally regular C(α) test and conduct a simulation study to investigate its finite sample performance. As each of these models is well known, in the main text I work under high level conditions to avoid repeating standard regularity conditions; lower level sufficient conditions are given in section S3 of the supplementary material.

4.1 Single index model

Consider the single index (regression) model (SIM) of Example 2.1: the researcher observes n i.i.d. copies of $W = (Y, X)$ such that

$$Y = f(X_1 + X_2'\theta) + \epsilon, \quad \mathbb{E}[\epsilon|X] = 0, \tag{22}$$

for $X = (X_1, X_2) \in \mathbb{R}^K$ a vector of covariates such that $(\epsilon, X) \sim \zeta$ for some density ζ (with respect to some σ-finite measure ν) and an unknown, continuously differentiable link function $f : \mathbb{R} \to \mathbb{R}$, which may be required to satisfy additional shape and smoothness constraints.

Chapter 2 of Horowitz (2009) provides an overview of this model. The efficient score and semiparametric efficiency bound for this model were obtained by Newey and Stoker (1993). Ichimura (1993) demonstrated that θ could be estimated by minimising a semiparametric least squares criterion and that, under homoskedasticity, this yields an efficient estimator. More recently, the estimation of θ subject to shape constraints on f has been considered (e.g. Kuchibhotla, Patra, and Sen, 2023).

I will consider two cases in which potentially non-standard inference problems may arise in this model. Firstly, I will consider the case where θ is potentially weakly identified due to f being close to flat, i.e. $f' \approx 0$.[35] Secondly I will consider the case where inference on θ is conducted with f estimated subject to a monotonicity restriction which is close to binding. As the latter case imposes a different restriction on the potential f functions, I distinguish these cases by refering to the former as "Model A" and the latter as "Model B".

[35] That θ is unidentified if $f' = 0$ is clear from (22); cf. Theorem 2.1 in Horowitz (2009).

Model setup Both models are accomodated simultaneously in the following development. The model parameters are $\gamma = (\theta, \eta)$ where $\eta = (f, \zeta)$ and the density of one observation with respect to a σ-finite measure $\tilde{\nu}$ is

$$p_\gamma(W) \coloneqq \zeta(Y - f(V_\theta), X), \qquad V_\theta \coloneqq X_1 + X_2'\theta. \tag{23}$$

Let P_γ denote the corresponding probability measure. The parameters γ are restricted by the following Asssumption. Let $\mathscr{X}$ be the support of X, $\mathscr{D}$ a convex open set containing $\{x_1 + x_2'\theta : \theta \in \Theta, x \in \mathscr{X}\}$, $C_b^1(\mathscr{D})$ the class of real functions which are bounded and continuously differentiable with bounded derivative on $\mathscr{D}$ and $\mathscr{I}(\mathscr{D})$ the set of monotone increasing functions $f : \mathscr{D} \to \mathbb{R}$.

Assumption 4.1: *The parameters $\gamma = (\theta, f, \zeta) \in \Gamma = \Theta \times \mathscr{F} \times \mathscr{Z}$ where*

(i) Θ is an open subset of $\mathbb{R}^{d_\theta}$;
(ii) $\mathscr{F} = C_b^1(\mathscr{D})$ (Model A) or $\mathscr{F} = C_b^1(\mathscr{D}) \cap \mathscr{I}(\mathscr{D})$ (Model B);
(iii) $\zeta \in \mathscr{Z}$ where

$$\mathscr{Z} \coloneqq \left\{ \zeta \in L_1(\mathbb{R}^{1+K}, \nu) : \zeta \geq 0, \int_{\mathbb{R}\times\mathscr{X}} \zeta \,\mathrm{d}\nu = 1, \text{ if } (\epsilon, X) \sim \zeta \text{ then (24) holds} \right\},$$

with $L_1(A, \nu)$ is the space of ν – integrable functions on A and

$$\mathbb{E}[\epsilon|X] = 0,\ \mathbb{E}[\epsilon^2] < \infty,\ \mathbb{E}[(|\epsilon|^{2+\rho} + |\phi(\epsilon, X)|^{2+\rho} + 1)\|X\|^{2+\rho}] < \infty,\ \mathbb{E}[XX'] \succ 0, \tag{24}$$

for $\phi(\epsilon, X)$ the derivative of $e \mapsto \log \zeta(e, X)$.

Additionally, for each $\gamma \in \Gamma$, p_γ is a probability density with respect to some σ-finite measure $\tilde{\nu}$.

That p_γ is a valid probability density holds automatically (with $\tilde{\nu} = \nu$) when $\epsilon|X$ is continuously distributed, see Appendix section S3.1.2.

Local Asymptotic Normality Consider local perturbations $P_{\gamma+\varphi_n(h)}$ for

$$\varphi_n(h) = \left(\frac{\tau}{\sqrt{n}},\ \varphi_{n,2}(b_1, b_2) \right), \qquad h = (\tau, b_1, b_2) \in H_\gamma = \mathbb{R}^{d_\theta} \times B_{\gamma,1} \times B_{\gamma,2}. \tag{25}$$

$B_{\gamma,1}$ is the set which indexes the perturbations to f and consists of a subset of the continuously differentiable functions $b_1 : \mathscr{D} \to \mathbb{R}$. $B_{\gamma,2}$ indexes the perturbations to ζ and consists of a subset of the functions $b_2 : \mathbb{R}^{1+K} \to \mathbb{R}$ which are continuously differentiable in their first argument and satisfy[36]

$$\mathbb{E}[b_2(\epsilon, X)] = 0,\ \mathbb{E}[\epsilon b_2(\epsilon, X)|X] = 0,\ \mathbb{E}[b_2(\epsilon, X)^2] < \infty \qquad \text{for } (\epsilon, X) \sim \zeta. \quad (26)$$

The precise form of $\varphi_{n,2}$ is left unspecified. It is required only that the resulting local perturbations satisfy the LAN property below.[37]

ASSUMPTION 4.2: *Suppose that $\mathcal{W}_n = \prod_{i=1}^n \mathbb{R}^{1+K}$ and $P_{n,\gamma,h} := P^n_{\gamma+\varphi_n(h)} \ll \nu_n$ for all $\gamma \in \Gamma$ and $h \in H_\gamma$ and are such that Assumption 3.1 holds with*

$$\log \frac{p_{n,\gamma,h}}{p_{n,\gamma,0}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n [A_\gamma h](W_i) - \frac{1}{2}\sigma_\gamma(h) + o_{P_{n,0}}(1), \quad h \in H_\gamma, \qquad (27)$$

where $\sigma_\gamma(h) = \int [A_\gamma h]^2 \, \mathrm{d}P_\gamma$ and $[A_\gamma h](W_i)$ is as in equation (21) with

$$\begin{aligned} \dot{\ell}_\gamma(W) &:= -\phi(Y - f(V_\theta), X) f'(V_\theta) X_2 \\ [D_\gamma b](W) &:= -\phi(Y - f(V_\theta), X) b_1(V_\theta) + b_2(Y - f(V_\theta), X). \end{aligned}$$

The test statistic In order to construct the test, I set $g_{n,\gamma} = \mathbb{G}_n g_\gamma$ for:

$$g_\gamma(W) := \omega(X)(Y - f(V_\theta)) f'(V_\theta) \left(X_2 - \frac{\mathbb{E}[\omega(X) X_2 | V_\theta]}{\mathbb{E}[\omega(X)|V_\theta]} \right), \qquad (28)$$

for a known weighting function $\omega : \mathbb{R}^K \to [\underline{\omega}, \overline{\omega}] \subset (0, \infty)$.[38] To verify the joint convergence conditions in Assumption 3.2 I additionally assume that (under P_γ)

$$\mathbb{E}[\epsilon^2|X] \le C < \infty, \quad \mathbb{E}\left[\epsilon \phi(\epsilon, X)|X\right] = -1, \quad \text{a.s. .} \qquad (29)$$

The latter condition can be shown to hold under Assumption 4.1 and additional regularity conditions.[39]

[36] A heuristic motivation for these restrictions is given in Appendix section S3.1.2.

[37] Specific examples of $\varphi_{n,2}(b)$ and B_γ for which this condition is satisfied are given in Appendix section S3.1.2.

[38] This coincides with the efficient score function $\tilde{\ell}_\gamma$ (as derived by Newey and Stoker, 1993),

$$\tilde{\ell}_\gamma(W) = \tilde{\omega}(X)(Y - f(V_\theta)) f'(V_\theta) \left(X_2 - \frac{\mathbb{E}[\tilde{\omega}(X) X_2 | V_\theta]}{\mathbb{E}[\tilde{\omega}(X)|V_\theta]} \right), \quad \tilde{\omega}(X) := \mathbb{E}[\epsilon^2|X]^{-1},$$

in the (typically infeasible) case with $\omega = \tilde{\omega}$.

[39] In particular, if for all $x \in \mathscr{X}$ with positive marginal density ζ_X one has $\lim_{|e|\to\infty} |e|\zeta(e,x) = 0$, then integrating by parts yields that for almost all such x,

$$\int e\phi(e,x) \frac{\zeta(e,x)}{\zeta_X(x)} \,\mathrm{d}e = \int e \frac{\zeta'(e,x)}{\zeta_X(x)} \,\mathrm{d}e = \zeta_X(x)^{-1} \left[\lim_{e\to\infty} e\zeta(e,x) - \lim_{e\to-\infty} e\zeta(e,x) \right] - \int \frac{\zeta(e,x)}{\zeta_X(x)} \,\mathrm{d}e = -1,$$

Proposition 4.1: *Suppose Assumptions 4.1 and 4.2 hold and (under P_γ) (29) holds. Then Assumption 3.2 holds with $g_{n,\gamma} = \mathbb{G}_n g_\gamma$ for g_γ as in (28).*

To form a feasible $\hat{g}_{n,\theta}$, estimators of f, f', Z_1 and Z_2 are required for $Z_1(V) := \mathbb{E}[\omega(X)X_2|V]$ and $Z_2(V) := \mathbb{E}[\omega(X)|V]$. To keep the notation concise let $Z_3 := f$, $Z_4 := f'$ Define also $Z_0 := Z_1/Z_2$ and correspondingly $\hat{Z}_{0,n,i} := \hat{Z}_{1,n,i}/\hat{Z}_{2,n,i}$. The estimator of $g_{n,\gamma}$ is $\hat{g}_{n,\theta}\left(W^{(n)}\right) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{n,\theta,i}$ with

$$\hat{g}_{n,\theta,i} := \omega(X_i)(Y_i - \hat{f}_{n,i}(V_{\theta,i}))\widehat{f'}_{n,i}(V_{\theta,i})\left(X_{2,i} - \hat{Z}_{1,n,i}(V_{\theta,i})/\hat{Z}_{2,n,i}(V_{\theta,i})\right). \quad (30)$$

Let $\check{V}_{n,\theta} := \frac{1}{n}\sum_{i=1}^n \hat{g}_{n,\theta,i}\hat{g}'_{n,\theta,i}$. If V_γ is known to have full rank then let $\hat{V}_{n,\theta} := \check{V}_{n,\theta}$, $\hat{\Lambda}_{n,\theta} := \hat{V}_{n,\theta}^{-1}$ and $\hat{r}_{n,\theta} = \operatorname{rank}(V_\gamma)$. Else form the estimator $\hat{V}_{n,\theta}$ according to the construction in subsection S2.1 using a truncation rate ν_n. $\hat{\Lambda}_{n,\theta}$ is then taken to be $\hat{V}_{n,\theta}^{\dagger}$ and $\hat{r}_{n,\theta} := \operatorname{rank}(\hat{V}_{n,\theta})$. Under the following condition, these estimators satisfy the conditions of Assumption 3.3.

Assumption 4.3: *Suppose that equation (29) holds (under P_γ), X has compact support, $\mathbb{E}[\epsilon^4] < \infty$, and with P_γ probability approaching one each $\mathsf{R}_{l,n,i} \leq r_n = o(n^{-1/4})$ where*

$$\mathsf{R}_{l,n,i} := \left[\int \left\|\hat{Z}_{l,n,i}(v) - Z_l(v)\right\|^2 \mathrm{d}\mathcal{V}_\gamma(v)\right]^{1/2},$$

where $\mathcal{V}_\gamma$ the law of V_θ under P_γ for $l = 1, \ldots, 4$ and where $\hat{Z}_{l,n,i}(V_{\theta,i})$ is $\sigma(\{V_{\theta,i}\} \cup \mathcal{C}_{n,j})$ measurable where $j = 1$ if $i > \lfloor n/2 \rfloor$ and 2 otherwise, with $\mathcal{C}_{n,1} := \{W_j : j \in \{1, \ldots, \lfloor n/2 \rfloor\}\}$ and $\mathcal{C}_{n,2} := \{W_j : j \in \{\lfloor n/2 \rfloor + 1, \ldots, n\}\}$.

The rate conditions in Assumption 4.3 can be satisfied by e.g. sample – split series estimators under standard smoothness conditions; see e.g. Belloni, Chernozhukov, Chetverikov, and Kato (2015). Appropriate estimators for f and f' may differ between Model A and Model B: in Model B one may wish to impose the restriction that f is monotonically increasing in the estimation. This can be achieved by using, for example, I – Splines (cf. Ramsay, 1988; Meyer, 2008).

Proposition 4.2: *Suppose Assumptions 4.1, 4.2 and 4.3 hold and that either V_γ is known to be full rank or $\hat{V}_{n,\theta}$ is constructed as in subsection S2.1 with truncation rate ν_n such that $r_n = o(\nu_n)$. Then Assumption 3.3 holds with $V_\gamma := P_\gamma[g_\gamma g_\gamma']$.*

---

where ζ' denotes the derivative of ζ with respect to its first argument.

A consequence of Assumption 4.2 and Propositions 4.1 and 4.2 is that the test $\psi_{n,\theta}$ formed as in (12) is locally regular by Theorem 3.1 (cf. Remark 3.4). See Appendix section S3.1.3 for a discussion of uniform local regularity in this model.

4.1.1 Simulation studies

I examine the finite sample performance of the proposed test in a simulation study. The simulation designs focus on two non-standard cases: (i) where $f' \approx 0$ and hence θ weakly identified and (ii) where f is estimated subject to a monotonicity constraint which is close to binding. I take $K = 1$ and test the hypothesis that $H_0 : \theta = \theta_0$ at a nominal level of 5%. Each study reports the results of 5000 monte carlo replications with a sample size of $n \in \{400, 600, 800\}$. A number of different choices for the link function f and the distribution ζ are considered. I report empirical rejection frequencies for the proposed test of H_0 based on $\hat{g}_{n,\theta}$ along with a Wald test in the style of Ichimura (1993). Finite sample power curves for the test proposed in this paper are also reported.

Design 1: Weak identification I set $\theta_0 = 1$ and consider two different classes of link function f. The first has $f(v) = f_j(v) = 5\exp(-v^2/2c_j^2)$, whilst the second sets $f(v) = f_j(v) = 25\,(1 + \exp(-v/c_j))^{-1}$. The values of c_j considered are recorded in Table 4. In each case, as c_j increases, the derivative of f flattens out as depicted in Figures S2 and S4 respectively.[40]

For each link function I consider various possible distributions for ζ. X is taken to be either $X = (Z_1, Z_2)$ or $X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$, where each Z_k is independently drawn from a $U(-1, 1)$. The error term is drawn as $\epsilon \sim \mathcal{N}(0, 1)$. Results for other error distributions, including heteroskedastic designs, are qualitatively similar and are presented in section S4.1 of the supplementary material.

I compute the $\hat{S}$ test based on $\hat{g}_{n,\theta}$ as in (30), with $\omega(X) = 1$. The functions f, f' and Z_1 are estimated via smoothing splines using the base R function `smooth.spline` with 20 knots.[41] The truncation parameter ν is set to 10^{-4}. I additionally compute a Wald test in the style of Ichimura (1993), using the same non-parametric estimates as for $\hat{g}_{n,\theta}$.[42]

[40] Figures S1 and S3 depict the functions themselves.

[41] In this setting $Z_2(V_\theta) = 1$ is known.

[42] That is, given an estimate $\hat{f}$, $\hat{\theta} = \arg\min_{\theta \in \Theta_\star} \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{f}(V_{\theta,i}))^2$, for some compact $\Theta_\star$. I take $\Theta_\star = [-10, 10]$. The variance matrix is computed as $\hat{\sigma}^2 / \frac{1}{n}\sum_{i=1}^n \left(\hat{f}'(V_{\hat{\theta},i})\left[X_2 - \hat{Z}(V_{\hat{\theta},i})\right]\right)^2$, for $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{f}(V_{\hat{\theta},i}))^2$.

The finite sample empirical rejection frequencies for both of these procedures are recorded in Table 5 for the case with exponential f_j and Table 6 for the case with logistic f_j. The $\hat{S}$ test display empirical rejection frequncies close to the nominal 5% for all simulation designs considered. In contrast, the Wald test based on the Ichimura (1993) – style estimator over – rejects in most of the simulation designs considered. As n increases, the size of the Wald test reduces and approaches the nominal level, though in many designs the rejection rate remains substantially above the nominal level at $n = 800$.

Figures 1 & 2 contain power plots of the $\hat{S}$ test, which show the expected shape given the power results in section 3.3. For particularly flat index functions there is very identifying information and hence very little power available. As the index function moves away from the point of identification failure, the available power increases, which is reflected in the increased power provided by the proposed test.

Design 2: Monotonicity constraint I set $\theta_0 = 0$ 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 plotted in Figure S5.[43] These flat sections may cause any monotonicity constraints to bind in the estimation of f. I explore the effect this has on the rejection frequencies of the $\hat{S}$ test based on (30) with $\omega(X) = 1$ and the Ichimura (1993) – style Wald test. Both tests are computed with f, f' estimated by 9 monotonic I – splines, whilst Z_1 is estimated using 6 cubic B – splines. As the efficient information is always positive in this simulation design, $\nu = 0$.

Similar to in Design 1, X is taken to be either $X = (Z_1, Z_2)$ or $X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$, where each Z_k is independently drawn from a $U(-3/2, 3/2)$. The error term is drawn as $\epsilon \sim \mathcal{N}(0, 1)$.

Table 7 displays the empirical rejection frequencies attained by the $\hat{S}$ test and the Wald test. The $\hat{S}$ test provides rejection rates close to the nominal level of 5% in each simulation design considered. The Wald test displays both over- and under-rejection depending on the exact simulation design. In particular, for the logistic function f_1, the Wald test over-rejects (similarly as to the strongly identified cases in Design 1). For the two double logistic functions, the rejection rate is lower: it slightly exceeds the nominal level for f_2 and under-rejects for f_3, likely due to the (close to) binding monotonicity constraint. Results for other error distributions, including heteroskedastic designs, are qualitatively similar and are presented in section S4.1 of the supplementary material.

[43]Formal definitions can be found in section S4.1.2.

The 3 panels of Figure 3 depict the finite-sample power curves for the $\hat{S}$ and Wald tests with $X = (Z_1, Z_2)$, with $f = f_1, f_2, f_3$ respectively. In each panel, the $\hat{S}$ typically appears to provide higher power.[44] This is particularly true for $f = f_3$, where the under-rejection of the Wald test under the null observed in Table 7 persists under the alternative, yielding an under-powered test relative to the $\hat{S}$ test. The results for the case with $X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$ are qualitatively similar, see Figure S7 in the Supplementary material.

4.2 IV model

Consider the instrumental variables model of Example 2.2: n i.i.d. copies of $W = (Y, X, Z)$ are observed where

$$Y = X'\theta + Z_1'\beta + \epsilon, \qquad \mathbb{E}[\epsilon|Z] = 0, \qquad Z = (Z_1', Z_2')'. \tag{31}$$

Let $K := d_\theta + d_Z + 1$, i.e. the dimension of W. Define $\pi(Z) := \mathbb{E}[X|Z]$ and $\upsilon = X - \mathbb{E}[X|Z] = X - \pi(Z)$ such that $\mathbb{E}[\upsilon|Z] = 0$. This yields the following two equation model:

$$\begin{aligned} Y &= X'\theta + Z_1'\beta + \epsilon \\ X &= \pi(Z) + \upsilon \end{aligned}, \qquad \mathbb{E}[U|Z] = 0, \quad U := (\epsilon, \upsilon')', \quad \alpha := (\theta', \beta')'. \tag{32}$$

If $\pi(Z) = 0$ the instruments Z provide no information about θ. More generally, $\pi(Z)$ being rank deficient a.s. can cause underidentification of θ. Note that lack of identification or weak identification in this model can be very different from the model with an assumed linear first stage as there are many data configurations in which $\mathbb{E}[X|Z]$ provides substantial identifying information about θ whilst the linear projection of X onto the columns of Z may be uniformative. In such situations, tests which can exploit such non-linear identifying information can provide substantially more power than tests which (implicitly) use a linear first stage. This is illustrated in the simulation study and the empirical applications below.[45]

In the special case where $\pi(Z) = \varpi f(Z) = \sum_{j=1}^{J} \varpi_j f_j(Z)$ for a known function f, such non-linear effects could also be captured by classical weak-instrument robust statistics, such as the LM test (Kleibergen, 2002), by replacing the instru-

[44] This is not true for alternatives around the null for $f = f_1$. However, this is not a like-for-like comparison, as the Wald test over-rejects when $f = f_1$, see Table 7.

[45] This does not contradict the optimality results that have been derived for, e.g., the AR test (e.g. Moreira, 2009; Chernozhukov, Hansen, and Jansson, 2009) as these results are derived under an imposed linear first-stage.

ments Z with $f(Z)$. In practice f is generally not known but π may be non-parametrically estimated by using an approximation of the form $\varpi_n f^{(n)}(Z) = \sum_{j=1}^{J_n} \varpi_{n,j} f_j(Z)$ with $(f_1, \ldots, f_{J_n})$ an increasing number of basis functions. The test developed below is a LM type test based on an orthogonalised score statistic in which such a non-parametric estimate of π is plugged-in. The orthogonalisation ensures that neither (i) weak identification nor (ii) the plugged-in nonparametric estimator causes the resulting sequence of tests to be (locally) non – regular.

Model setup Let ζ denote the density of $\xi := (\epsilon, \upsilon', Z')$ with respect to a σ - finite measure ν. The parameters of the IV model are $\gamma = (\theta, \eta)$ with the nuisance parameters collected in $\eta = (\beta, \pi, \zeta)$ and the density of one observation given by

$$p_\gamma(W) = \zeta(Y - X'\theta - Z_1'\beta, X - \pi(Z), Z), \tag{33}$$

with respect to a σ - finite measure $\tilde{\nu}$ and P_γ denotes the corresponding measure. The model parameters are restricted by the following assumption.

Assumption 4.4: *The parameters $\gamma = (\theta, \beta, \pi, \zeta) \in \Gamma = \Theta \times \mathcal{B} \times \mathscr{P} \times \mathscr{Z}$ where*

(i) Θ is an open subset of $\mathbb{R}^{d_\theta}$ and $\mathcal{B}$ is an open subset of $\mathbb{R}^{d_\beta}$;
(ii) $\mathscr{Z}$ is a subset of the set of density functions on $\mathbb{R}^K$ with respect to ν;
(iii) For $(\pi, \zeta) \in \mathscr{P} \times \mathscr{Z}$, if $\xi := (U', Z')'$, then

$$\mathbb{E}[U|Z] = 0, \qquad \mathbb{E}\|\xi\|^4 < \infty, \qquad \mathbb{E}\|\pi(Z)\|^4 < \infty, \qquad \mathbb{E}\|\phi(\xi)\|^4 < \infty,$$

where $\phi_1 := \nabla_\epsilon \log \zeta(\epsilon, \upsilon, Z)$, $\phi_2 := \nabla_\upsilon \log \zeta(\epsilon, \upsilon, Z)$ and $\phi := (\phi_1, \phi_2')'$.

Additionally, for each $\gamma \in \Gamma$, p_γ is a probability density with respect to some σ-finite measure $\tilde{\nu}$.

Assumption 4.4 essentially imposes only that certain moments exist and the IV conditional mean restriction. That p_γ is a valid probability density holds automatically (with $\nu = \tilde{\nu}$) when $U|Z$ is continuously distributed; see Appendix section S3.2.2 for a discussion.

Local Asymptotic Normality Consider local perturbations of the form $P_{\gamma+\varphi_n(h)}$ for

$$\varphi_n(h) := \left(\frac{\tau}{\sqrt{n}}, \frac{b_0}{\sqrt{n}}, \varphi_{n,1}(b_1), \varphi_{n,2}(b_2)\right), \quad h = (\tau, b) \in H_\gamma := \mathbb{R}^{d_\theta} \times B_\gamma, \tag{34}$$

with $B_\gamma := \mathbb{R}^{d_\beta} \times B_{\gamma,1} \times B_{\gamma,2}$. $B_{\gamma,1}$ is a subset of the bounded functions $b_1 : \mathbb{R}^{d_Z} \to \mathbb{R}^{d_\theta}$ and $B_{\gamma,2}$ a subset of the functions $b_2 : \mathbb{R}^K \to \mathbb{R}$ which are bounded and continuously differentiable in its first $1 + d_\theta$ components with bounded derivative and such that[46]

$$\mathbb{E}\left[b_2(U, Z)\right] = 0, \quad \mathbb{E}\left[U b_2(U, Z)|Z\right] = 0, \qquad \text{for } (U', Z')' \sim \zeta. \tag{35}$$

The precise form of $\varphi_{n,1}$ and $\varphi_{n,2}$ is left unspecified. It is required only that the resulting local perturbations satisfy the LAN property below.[47]

Assumption 4.5: *Suppose that $\mathcal{W}_n = \prod_{i=1}^n \mathbb{R}^K$ and $P_{n,\gamma,h} := P^n_{\gamma+\varphi_n(h)}$ for all $\gamma \in \Gamma$ and $h \in H_\gamma$ and are such that Assumption 3.1 holds with*

$$\log \frac{p_{n,\gamma,h}}{p_{n,\gamma,0}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n [A_\gamma h](W_i) - \frac{1}{2}\sigma_\gamma(h) + o_{P_{n,0}}(1), \quad h \in H_\gamma, \tag{36}$$

where $\sigma_\gamma(h) = \int [A_\gamma h]^2 \,\mathrm{d}P_\gamma$ and $[A_\gamma h](W_i)$ has the form given in equation (21) with

$$\begin{aligned}
\dot{\ell}_\gamma(W) &:= -\phi_1(Y - X'\theta - Z_1'\beta, X - \pi(Z), Z)X_1 \\
[D_\gamma b](W) &:= -\phi(Y - X'\theta - Z_1'\beta, X - \pi(Z), Z)' \left[\begin{smallmatrix} b_0' Z_1 & b_1(Z) \end{smallmatrix}\right] + b_2(Y - X'\theta - Z_1'\beta, X - \pi(Z), Z).
\end{aligned}$$

The test statistic g_γ will be set equal to the efficient score function for θ, $\tilde{\ell}_\gamma$, under homoskedasticity. I first give the efficient score for θ in the following Lemma (without imposing homoskedasticity), for which I introduce some additional notation. Let $J(Z) := \mathbb{E}[UU'|Z]$ (under γ) and $E = (E_1, E_2')' = J(Z)^{-1}U$ where the partitioning is conformal with $U = (\epsilon, v')' = (Y - X'\theta - Z_1'\beta, X' - \pi(Z)')'$. I also assume the following conditions hold a.s.:[48]

$$\begin{aligned}
0 < c \le \lambda_{\min}(J(Z)) \le \lambda_{\max}(J(Z)) \le C < \infty, \\
\mathbb{E}\left[\phi(\epsilon, v, Z)U'|Z\right] = -I, \qquad \mathbb{E}[\phi_1(\epsilon, v, Z)vU'] = 0.
\end{aligned} \tag{37}$$

Lemma 4.1: *If Assumptions 4.4, 4.5 and equation (37) hold, then for $\mathbb{E}$ taken*

[46] A heuristic motivation for these restrictions is given in Appendix section S3.2.2.

[47] Examples of $\varphi_{n,1}, \varphi_{n,2}$ and $B_{\gamma,1}, B_{\gamma,2}$ for which this condition is satisfied are given in Appendix section S3.2.2.

[48] Under Assumption 4.4 the second two equations hold provided for each $i = 1, \ldots, 1 + d_\alpha$, $\lim_{|u_i| \to \infty} |u_i| \zeta(u, z) = 0$; see Lemma S3.2. All expectations in (37) are taken under P_γ and the $i - j$ indexing in the definition of q_1 is conformal with $U = (\epsilon, v')'$.

under P_γ,

$$\tilde{l}_\gamma(W) = \begin{bmatrix} \tilde{l}_{\gamma,1}(W) \\ \tilde{l}_{\gamma,2}(W) \end{bmatrix} = \begin{bmatrix} \pi(Z) \\ Z_1 \end{bmatrix} \left[E_1 - \mathbb{E}[E_1 E_2'|Z]\mathbb{E}[E_2 E_2'|Z]^{-1} E_2\right]$$

is the efficient score for (θ, β). *In consequence,*

$$\begin{aligned} \tilde{\ell}_\gamma(W) &= \tilde{l}_{\gamma,1}(W) - \mathbb{E}\left[\tilde{l}_{\gamma,1}\tilde{l}_{\gamma,2}'\right]\mathbb{E}\left[\tilde{l}_{\gamma,2}\tilde{l}_{\gamma,2}'\right]^{-1}\tilde{l}_{\gamma,2}(W) \\ &= q_1(J(Z))(Y - X'\theta - Z_1'\beta)\left[\pi(Z) - \mathbb{E}[XZ_1']\mathbb{E}[Z_1 Z_1']^{-1} Z_1\right] \ , \end{aligned} \tag{38}$$

is the efficient score for θ, *where* $q_1(J) := (J_{1,1})^{-1}$.

Define $\bar{J} := \mathbb{E}[J(Z)] = \mathbb{E}[UU']$ and use in in place of $J(Z)$ in (38) to form

$$\bar{\ell}_\gamma(W) := q_1(\bar{J})(Y - X'\theta - Z_1'\beta)\left[\pi(Z) - \mathbb{E}[XZ_1']\mathbb{E}[Z_1 Z_1']^{-1} Z_1\right]. \tag{39}$$

This function also belongs to the orthocomplement of $\{D_\gamma b : b \in B_\gamma\}$ and, moreover, when $J(Z) = \bar{J}$ a.s., coincides with the efficient score function. These facts are shown in the next Lemma.

Lemma 4.2: *Suppose that Assumptions 4.4, 4.5 and equation* (37) *hold. Then,* $\bar{\ell}_\gamma$ *as defined in* (39) *belongs to the orthocomplement of* $\{D_\gamma b : b \in B_\gamma\}$ *(in* $L_2(P_\gamma)$*). If* $J(Z) = \bar{J}$ *a.s., then* $\bar{\ell}_\gamma = \tilde{\ell}_\gamma$ *a.s..*

Assumption 3.2 is satisfied with $g_{n,\gamma} = \mathbb{G}_n \bar{\ell}_\gamma$.

Proposition 4.3: *Suppose that Assumptions 4.4, 4.5 and equation* (37) *hold. Then Assumption 3.2 is satisfied with* $g_{n,\gamma} = \mathbb{G}_n \bar{\ell}_\gamma$.

Suppose that $\hat{\beta}_n$ and $\hat{\pi}_{n,i}(Z_i)$ are estimators of β and $\pi(Z_i)$ respectively. Let the i-th residual in based on these estimators be $\hat{U}_{n,i}$, that is:

$$\hat{U}_{n,i} := \begin{bmatrix} Y_i - X_i'\theta - Z_{1,i}'\hat{\beta}_n \\ X_i - \hat{\pi}_n(Z_i) \end{bmatrix}, \tag{40}$$

and set $\hat{J}_n := \frac{1}{n}\sum_{i=1}^n \hat{U}_{n,i}\hat{U}_{n,i}'$ Then put

$$\hat{g}_{n,\theta,i} := q_1(\hat{J}_n)\left(Y_i - X_i'\theta - Z_{1,i}'\hat{\beta}_n\right)\left[\hat{\pi}_n(Z_i) - \left[\frac{1}{n}\sum_{i=1}^n X_i Z_{1,i}'\right]\left[\frac{1}{n}\sum_{i=1}^n Z_{1,i}Z_{1,i}'\right]^{-1} Z_i\right], \tag{41}$$

and

$$\check{V}_{n,\theta} := \frac{1}{n}\sum_{i=1}^{n} \hat{g}_{n,\theta,i}\hat{g}'_{n,\theta,i}. \tag{42}$$

If $V_\gamma := \mathbb{E}[\bar{\ell}_\gamma \bar{\ell}'_\gamma]$ is known to have full rank, put $\hat{V}_{n,\theta} := \check{V}_{n,\theta}$, $\hat{\Lambda}_{n,\theta} := \hat{V}_{n,\theta}^{-1}$ and $\hat{r}_{n,\theta} = \operatorname{rank}(V_\gamma)$. Else form the estimator $\hat{V}_{n,\theta}$ according to the construction in subsection S2.1 using a truncation rate ν_n. $\hat{\Lambda}_{n,\theta}$ is then taken to be $\hat{V}_{n,\theta}^{\dagger}$ and $\hat{r}_{n,\theta} := \operatorname{rank}(\hat{V}_{n,\theta})$.

Assumption 4.6: *Suppose that, given θ, $\hat{\beta}_n$ and $\hat{\pi}_{n,i}(Z_i)$ are estimators such that $\sqrt{n}(\hat{\beta}_n - \beta) = O_{P_{n,\gamma}}(1)$, $\hat{\beta}_n$ takes values in $\mathscr{S}_n := \{CZ/\sqrt{n} : Z \in \mathbb{Z}^{d_\beta}\}$ for some $d_\beta \times d_\beta$ matrix C and and with $P_{n,\gamma}$ – probability approching one,*

$$\left[\int \|\hat{\pi}_{n,i}(z) - \pi(z)\|^2 \, \mathrm{d}\zeta_Z(z)\right]^{1/2} \leq \delta_n = o(1), \tag{43}$$

where ζ_Z is the marginal distribution of Z and $\hat{\pi}_{n,i}(Z_i)$ is $\sigma(\{Z_i\} \cup \mathcal{C}_{n,j})$ – measurable where $j = 1$ if $i > \lfloor n/2 \rfloor$ and 2 otherwise, with $\mathcal{C}_{n,1} := \{W_j : j \in \{1, \ldots, \lfloor n/2 \rfloor\}\}$ and $\mathcal{C}_{n,2} := \{W_j : j \in \{\lfloor n/2 \rfloor + 1, \ldots, n\}\}$.

Suppose also that

$$\mathbb{E}\left[\epsilon^4 \|\pi(Z)\|^4\right] < \infty, \qquad \mathbb{E}\left[\epsilon^4 \|Z_1\|^4\right] < \infty, \qquad \mathbb{E}[Z_1 Z_1'] \text{ is nonsingular,}$$

and that either V_γ is full rank, or the truncation rate ν_n satisfies $\delta_n^2 + n^{-1/2} = o(\nu_n)$.

Assumption 4.6 merits some commentary. Firstly the discretisation of $\hat{\beta}_n$ is a technical device due to Le Cam (1960) which permits the proof of Proposition 4.4 below to go through under weaker conditions. This can always be arranged given a $\sqrt{n}$ – consistent initial estimator, by replacing its value with the closest point in the set $\mathscr{S}_n$. Secondly, due to the structure of the estimands $\bar{\ell}_\gamma(W_i)$, the rate δ_n in (43) need only converge to zero. There is no requirement that, for example, $\delta_n = o(n^{-1/4})$.

Proposition 4.4: *Suppose that Assumptions 4.4, 4.5, 4.6 hold along with equation (37). Then Assumption 3.3 holds with $\hat{g}_{n,\theta} := \frac{1}{\sqrt{n}}\sum_{i=1}^n g_{n,\theta,i}$ for $g_{n,\theta,i}$ in (41), $g_{n,\gamma} := \frac{1}{\sqrt{n}}\sum_{i=1}^n \bar{\ell}_\gamma(W_i)$ and $\hat{\Lambda}_{n,\theta}$ defined below equation (42).*

A consequence of Assumption 4.5 and Propositions 4.3 and 4.4 is that the test $\psi_{n,\theta}$ formed as in (12) is locally regular by Theorem 3.1 (cf. Remark 3.4). See Appendix section S3.2.3 for a discussion of uniform local regularity in this model.

4.2.1 Simulation study

I explore the quality of the asymptotic approximation developed above in the finite sample setting in various simulation designs. In all simulation settings considered below I consider $n \in \{200, 400, 600\}$ and empirical rejection frequencies are computed based on 5000 simulated data sets (Design 1) or 2500 simulated data sets (Design 2). Two simulation designs are reported here (some additional simulation results for these designs are reported in section S4.2.1 of the supplementary material) and two additional designs covering (i) heteroskedasticity and (ii) overidentified models are reported in section S4.2.2 of the supplementary material.

Design 1: Univariate, just identified The first case considered is a setting with $d_\theta = 1$, $Z_1 = 1$, Z_2 univariate and homoskedastic errors. I consider $\pi(Z) = \pi(Z_2)$ equal to the exponential, logistic and linear functions detailed in Table 8 and plotted in Figures S8 – S10. For each function type there are 3 considered functions, indexed by $j = 1, \ldots, 3$. The higher j, the weaker the identification of θ (for given values of all other parameters). I draw (ϵ, υ) from a multivariate normal distribution with unit variances and covariance 0.95. Z_2 is drawn as an independent standard normal random variable.

I consider the results of applying the test developed in the previous section with π estimated by (i) OLS, (ii) series regression with Legendre polynomials. The truncation parameter ν is set to 0.1. I consider both setting the number of polynomials, k, to 6 and selecting k via information criteria. I additionally consider a number of alternative testing approaches: the Anderson and Rubin (1949) test (AR), a TSLS Wald test (both implicitly using a linear first stage) and GMM Wald and LM tests using $k = 6$ Legendre polynomials.[49]

The empirical rejection frequencies under the null for each of these tests are reported in Table 9. The results indicate that the empirical null rejection probability of the $\hat{S}$ tests proposed in this paper is well controlled in all scenarios. When θ is well identified (i.e. low j), the $\hat{S}$ tests generally are very close to the nominal 5% level, whilst in cases of high j they typically reject between 0 and 5% of the time.[50] The exception to this finding is when $\pi(Z_2)$ is estimated by OLS and the

[49] In the homoskedastic just identified setting (with a linear first stage) the AR test is numerically equivalent to the LM test (Kleibergen, 2002) and CLR test (Moreira, 2003). Moreover, if π is truly linear, all of these tests (and therefore the AR in particular) are uniformly most powerful unbiased as demonstrated by Moreira (2009).

[50] A null rejection probability of 0 is for weakly identified cases is in accordance with the theoretical predictions of Theorem 3.1, corresponding to the conservative case observed when $r = 0$.

true π is exponential, where the empirical rejection frequency is always close to zero, as this estimator performs very poorly in this setting.

As expected 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 two GMM tests which display even greater levels of over-rejection.

Given the results in Table 9, I consider the power of the $\hat{S}$ tests proposed above and the AR test. These are plotted for the exponential, logistic and linear design in figures 4 – 6. These power plots demonstrate that for the exponential π design, there is very limited power available via the AR test. This is not suprising given the linear projections in the definition of the AR statistic. For similar reasons the $\hat{S}$ test with π estimated via OLS shows even lower power and never exceeds the nominal 5% level. The $\hat{S}$ tests based on Legendre polynomials by contrast display substantial power in the $j = 1, 2$ designs. This is as expected, as – unlike the AR test – these tests are able to exploit the non-linear identifying information provided by Z.

In the logistic π case, no test appears to provide non trivial power for the $j = 3$ case (at least over the interval considered). For the $j = 1, 2$ cases, the AR and $\hat{S}$ tests based on either OLS or Legendre polynomials display similar power.

Finally, in the linear π case the (optimal) AR test dominates. In the well identified case ($j = 1$) the $\hat{S}$ test based on OLS estimates of π matches its power curve, though it performs less well when there is less identifying information. The other $\hat{S}$ tests continue to provide reasonable power in the cases where identification is stronger.

Design 2: Bivariate, just identified I now consider the case where $d_\theta = 2$, and Z_2 is bivariate. $\pi(Z) = \pi(Z_2)$ will be taken as $\pi(Z_2) = (\pi_1(Z_{2,1}), \pi_2(Z_{2,2}))'$ with each π_i $(i = 1, 2)$ being one of the exponential, logistic or linear functions considered in Design 1.[51] The Z_2 are drawn from a zero - mean, multivariate normal distribution with covariance matrix $\text{Var}(Z_2) = \left[\begin{smallmatrix} 1 & 0.4 \\ 0.4 & 1 \end{smallmatrix}\right]$. The (homoskedastic) error terms ϵ, υ are drawn (independently) from a zero-mean multivariate normal such that each has variance 1 and the covariances are $\text{Cov}(\epsilon, \upsilon_i) = 0.9$ and $\text{Cov}(\upsilon_1, \upsilon_2) = 0.7$. As in Design 1, $Z_1 = 1$.

The tests considered are similar to as in Design 1. In particular, I consider the

[51] This separation is assumed unknown to the researcher and is not imposed in the estimation of π.

$\hat{S}$ tests with π estimated by (i) OLS and (ii) tensor product Legendre polynomials. I consider both fixing the number of polynomials at $k = 3$ in each of the univariate series which form the tensor product basis and choosing $k \in \{3, 4, 5, 6, 7\}$ using information criteria. ν is set to 0.1. I additionally consider the AR test, a TSLS Wald test and GMM Wald and LM tests using the tensor product basis as instruments (along with Z_1).

The results are shown in Tables 10 & 11. The first table has $\pi_1 = \pi_2$ whilst the second table caries the form of π_1 only, with π_2 always remaining linear. The results are qualitatively similar across the two tables: the $\hat{S}$ test with Legendre polynomials appears to always control the null rejection probability of the test close to the nominal 5% level, as does the AR test. The $\hat{S}$ test with OLS estimates typically underreject. The TSLS Wald test and two GMM tests considered overreject when identification is weaker (higher j).

Figures 7 – 15 show power surfaces for the AR test and $\hat{S}$ tests computed with π estimated using $k = 3$ Legendre polynomials. These are plotted for cases where π_1 and π_2 have the same form and where (i) both π_1 and π_2 have $j = 1$, corresponding to a strongly identified setting; (ii) where π_1 has $j = 1$ and π_2 has $j = 3$, in which θ_2 is weakly identified and (iii) where both π_i have $j = 3$, i.e. θ is weakly identified.

Figures 7 – 9 show the case with exponential π. For this design the AR test is unable to provide non-trivial power regardless of the identification strength (j). In the strongly identified case (i) the $\hat{S}$ test provides good power in all directions. In case (ii), where θ_2 is weakly identified, the $\hat{S}$ test continues to provide good power against violations of the null in the first co-ordinate but (as expected) only trivial power in the second co-ordinate. In the weakly identified case (iii), neither test is able to provide reasonable power against the considered alternatives.

For the logistic and linear cases depicted in Figures 10 – 12 and 13 – 15 respectively, all tests display good power in all directions in the strongly identified case (i) and good power against violations of the null in the first co-ordinate but only trivial power against violations in the second co-ordinate in the case (ii) θ_2 weakly identified. The AR test seems to provide marginally higher power in these cases, but the difference is minor with both tests displaying similar power surfaces. In case (iii) where θ is weakly identified, the $\hat{S}$ test displays only trivial power. The same is true of the AR test in the logistic design; it is able to provide some power in one corner of the plot for the linear design.

Some additional simulation results for this design are reported in section S4.2.1 in the supplementary material. In particular, these results highlight that choosing the number of polynomial basis functions k by AIC yields power surfaces which are typically very similar to those in Figures 7 – 15.

Additional simulation results Section S4 contains a discussion of two further simulation designs, which I briefly summarise here. Design 3 replicates Design 1 with the addition of heteroskedastic errors. Briefly, the results are qualitiatively the same as found in Design 1 with the $\hat{S}$ tests well controlling the empirical rejection frequency in each scenario. As in the homoskedastic case, the $\hat{S}$ test provides substantially higher power than the AR test when π is exponential, and competitive power with the AR test in the logistic and the linear cases.

Design 4 considers an over-identified model. Specifically, the base setup is as in Design 1 (with $\rho = 0.95$ and Gaussian errors), however Z_2 is bivarate mean-zero normal with $\mathrm{Var}(Z_2) = [\begin{smallmatrix} 1 & 0.4 \\ 0.4 & 1 \end{smallmatrix}]$ and $\pi(Z_2) = (\pi_1(Z_{2,1}) + \pi_2(Z_{2,2}))/2$ where π_1 and π_2 have one of the exponential, logistic or linear forms of Table 8.[52] For the $\hat{S}$ tests, π is estimated in the same manner as for Design 2: using series regressions on tensor product bases formed of Legendre polynomials. A version with π estimated using OLS on Z_2 (and a constant) is also reported, along with GMM Wald and LM tests using these tensor product bases as instruments, the TSLS Wald test and AR, LM (Kleibergen, 2002) and CLR (Moreira, 2003) tests.[53]

The results show that both the "usual" weak instrument robust tests (AR, LM and CLR) and the $\hat{S}$ tests are able to well control the null rejection frequency, unlike the TSLS Wald and GMM based tests. In terms of power, the results are similar to the other designs: in the exponential case the $\hat{S}$ tests with nonparametrically estimated π are the only tests able to provide non-trivial power regardless of identification strength. In the logistic and linear cases, the $\hat{S}$ tests typically provide comparable power to the LM and CLR tests, with the latter two tests performing slightly better in the linear cases.

5 Empirical applications

In this section I apply the test developed in section 4.2 for the IV model (Example 2.2) to two classic instrumental variables studies.

[52]This functional form is treated as unknown and not imposed in the estimation of π.

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

5.1 Institutions and economic performance

Acemoglu, Johnson, and Robinson (2001) investigate the effect of better intitutions (i.e. better property rights, less distortionary policies) on economic performance. They use a measure of the average protection against expropriation risk as a measure of the strength of insitutions. Due to potential endogeneity concerns, they instrument this variable by the mortality rates of European colonial settlers. In the notation of Example 2.2, Y is the 1995 log gdp per capita on a ppp basis, X is a measure of the risk of government expropriation of private foreign investment, Z_1 contains a constant and any exogenous variables in the given specifications and Z_2 is the log of a measure of European settler mortality. I refer to the original article Acemoglu et al. (2001) for more detail on the data and a discussion of the literature.

Table 4 in Acemoglu et al. (2001) provides their two-stage least squares estimates for a variety of specifications along with the associated (Wald) confidence intervals; for details of the specifications, see Acemoglu et al. (2001). Table 1 below replicates these results (for specifications (1) – (8)), along with: (i) the first stage F statistic; (ii) OLS point estimates and the associated (Wald) confidence intervals; (iii) weak instrument robust Anderson and Rubin (1949) (hereafter AR) confidence intervals and (iv) weak instrument robust confidence intervals formed by inverting the $\hat{S}$ test developed in Section 4.2. This is implemented using

$$\hat{\pi}(Z) = \hat{\alpha}_1'\tilde{Z}_1 + \hat{\alpha}_2' p_k(Z_2),$$

with p_k the first k Legendre orthogonal polynomials, $\tilde{Z}_1$ is Z_1 excluding the constant term.[54] and $\hat{\alpha}$ is the OLS estimate in the regression of X on $(\tilde{Z}_1, p_k(Z_2))$. k is chosen via AIC, the truncation parameter $\nu = 0.001$ and the test is inverted over an evenly spaced grid of 1000 points between -1 and 3.

As can be see in Table 1, the first stage F statistic for many of the considered specifications is below the rule-of-thumb cutoff of 10 proposed by Staiger and Stock (1997) and so there is suggestive evidence that the instruments may be weak in these specifications.[55]

Overall the $\hat{S}$ based confidence intervals do not change the conclusions of Acemoglu et al. (2001) – all the confidence intervals exclude zero and and suggest similarly sized effects as found by Acemoglu et al. (2001). The results do however

[54] $p_{k,1}(z) = 1$.

[55] This possibility is considered by Albouy (2012); Acemoglu, Johnson, and Robinson (2012) who also report AR confidence intervals for some of these specifications.

demonstrate that the ability of the $\hat{S}$ test to exploit potential non-linearities in the relationship between the endogenous variable and instrument can result in (test inversion) confidence intervals which are much shorter than the AR confidence intervals (whilst remaining robust to weak identification). Table 2 shows the lengths of the OLS, 2SLS and $\hat{S}$ confidence intervals as a fraction of the length of the AR confidence interval for specifications (1) – (7) (specification (8) is excluded due to the AR confidence interval having infinite length). In all specifications inverting the $\hat{S}$ test produces a shorter confidence interval than inverting the AR test: the reduction in length ranges from around a 75% reduction in specification (7) to essentially no reduction in specification (5).[56]

Table 1: Point estimates and confidence intervals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
n	64	64	60	60	37	37	64	64
F	22.95	13.09	8.65	7.83	30.54	21.61	6.23	3.46
k	3	3	4	4	1	1	3	3
Point estimates								
OLS	0.52	0.47	0.49	0.47	0.48	0.47	0.42	0.4
2SLS	0.94	1	1.28	1.21	0.58	0.58	0.98	1.11
Confidence intervals								
OLS	$[0.4, 0.64]$	$[0.34, 0.6]$	$[0.33, 0.64]$	$[0.32, 0.62]$	$[0.35, 0.61]$	$[0.32, 0.61]$	$[0.31, 0.54]$	$[0.28, 0.52]$
2SLS	$[0.63, 1.26]$	$[0.55, 1.44]$	$[0.56, 2]$	$[0.5, 1.92]$	$[0.38, 0.78]$	$[0.34, 0.81]$	$[0.38, 1.58]$	$[0.18, 2.04]$
AR	$[0.7, 1.43]$	$[0.68, 1.88]$	$[0.81, 3.33]$	$[0.75, 3.44]$	$[0.39, 0.82]$	$[0.35, 0.89]$	$[0.6, 3.59]$	$(-\infty, -9.24] \cup [0.59, \infty)$
S	$[0.58, 1.05]$	$[0.47, 0.97]$	$[0.26, 1.12]$	$[0.02, 0.97]$	$[0.41, 0.84]$	$[0.36, 0.85]$	$[0.32, 1.09]$	$[0.26, 0.97]$

Notes: F is the first stage F statistic; k is the number of polynomials in Z_2 chosen by AIC.

Table 2: Length of confidence interval relative to AR

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
OLS	0.335	0.213	0.121	0.110	0.611	0.535	0.077
2SLS	0.857	0.736	0.571	0.526	0.922	0.888	0.401
AR	1.000	1.000	1.000	1.000	1.000	1.000	1.000
S	0.647	0.416	0.344	0.354	0.992	0.917	0.256

5.2 Returns to schooling

I revisit the problem of estimating the returns to schooling using IV methods. In particular I use the data from the original Card (1995) article: 1976 wage data and education data from the 1966 NLS cohort. See Card (1995) for details on the

[56] Specifications (5) & (6) yield very similar confidence intervals to the AR intervals as in these specifications AIC chooses $k = 1$.

dataset and Card (2001) for a review of the literature. I set Y to be the log wage in 1976, X the years of education, Z_2 is an indicator for growing up near a $4-$year college interacted with father's education, and Z_1 includes an intercept and controls for race, experience, SMSA and region.

Table 3 below provides (i) OLS and (ii) two-stage least squares estimates along with their associated (Wald) confidence intervals, (iii) weak instrument robust AR confidence intervals and (iv) weak instrument robust confidence intervals formed by inverting the $\hat{S}$ test developed in Section 4.2. This is implemented using

$$\hat{\pi}(Z) = \hat{\alpha}_1' \tilde{Z}_1 + \hat{\alpha}_2' p_k(Z_2),$$

with p_k the first k Legendre orthogonal polynomials, $\tilde{Z}_1$ is Z_1 excluding the constant term.[57] and $\hat{\alpha}$ is the OLS estimate in the regression of X on $(\tilde{Z}_1, p_k(Z_2))$. AIC chooses $k = 2$, the truncation parameter is set to $\nu = 0.001$ and the test is inverted over an evenly spaced grid of 1000 points between -0.2 and 0.2.[58]

Table 3: Point estimates and confidence intervals

Method	Point Estimate	Confidence Interval	Relative Length
OLS	0.076	$[0.069, 0.084]$	0.178
2SLS	0.085	$[0.043, 0.128]$	0.975
AR		$[0.042, 0.13]$	1.000
S		$[0.042, 0.118]$	0.868

The confidence interval obtained by inverting the $\hat{S}$ test is similar but slightly shorter than the AR interval, achieving an approximate 15% reduction in length.

6 Conclusion

In this paper I introduced a notion of local regularity for (sequences of) tests analogous to the notion of local regularity for estimators widely used in semiparametric estimation theory. I established that $C(\alpha)-$style tests are locally regular under mild conditions, including in non – regular cases where locally regular estimators do not exist. Such non – regular cases include, for example, semiparametric weak identification asymptotics.

I additionally generalised the classical local asymptotic power bounds for locally asymptotically normal semiparametric models to the case where the efficient

[57] $p_{k,1}(z) = 1$.

[58] The first stage $F-$statistic in this model is 79.141, far exceeding the rule-of-thumb cutoff of 10 suggested in Staiger and Stock (1997).

information matrix has positive, but potentially deficient, rank. As such, these results also apply in cases of underidentification (or weak underidentification). Moreover, I demonstrated that, for a certain choice of moment function, the C(α) – style test attains these power bounds. This improves on known results as it does not require the data to be i.i.d. nor the information operator to be boundedly invertible.

Three examples are developed in detail and the approach is validated in simulation studies which demonstrate that the asymptotic theory provides an accurate approximation to the finite sample performance of the considered C(α) tests. In particular, it is shown that in a single-index model, the considered C(α) test is robust to plugging in a shape constrained estimator of the link function, whilst in a IV model, the considered C(α) test remains robust under semiparametric weak identification asymptotics and can have substantially higher power than classical weak instrument robust tests when the relationship between the endogenous variables and instruments is non-linear. This latter point is also highlighted in two empirical examples using the IV model in which inverting the C(α) test can lead to substantially shorter confidence intervals than the Anderson and Rubin (1949) test.

References

Acemoglu, D., Johnson, S., and Robinson, J. A. (2001), "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91, 1369–1401.

— (2012), "The Colonial Origins of Comparative Development: An Empirical Investigation: Reply," *American Economic Review*, 102, 3077–3110.

Albouy, D. Y. (2012), "The Colonial Origins of Comparative Development: An Empirical Investigation: Comment," *American Economic Review*, 102, 3059–76.

Amengual, D., Bei, X., and Sentana, E. (2023), "Hypothesis Tests with a Repeatedly Singular Information Matrix," Working paper.

Anderson, T. W. and Rubin, H. (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, 20, 46 – 63.

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.

— (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 Cheng, X. (2012), "Estimation and Inference With Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2153–2211.

Andrews, D. W. K. and Guggenberger, P. (2009), "Hybrid and Size-Corrected Subsampling Methods," *Econometrica*, 77, 721–762.

— (2010), "Asymptotic Size and a Problem with Subsampling and with the m out of n Bootstrap," *Econometric Theory*, 26, 426–468.

— (2019), "Identification- and singularity-robust inference for moment condition models," *Quantitative Economics*, 10, 1703–1746.

Andrews, I. and Mikusheva, A. (2015), "Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models," *Quantitative Economics*, 6, 123–152.

— (2016), "Conditional Inference With a Functional Nuisance Parameter," *Econometrica*, 84, 1571–1612.

— (2022), "Optimal Decision Rules for Weak GMM," *Econometrica*, 90, 715–748.

— (2023), "GMM is Inadmissible Under Weak Identification," .

Belloni, A., Chernozhukov, V., Chetverikov, D., and Kato, K. (2015), "Some new asymptotic theory for least squares series: Pointwise and uniform results," *Journal of Econometrics*, 186, 345–366.

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.

Bickel, P. J., Ritov, Y., and Stoker, T. M. (2006), "Tailor-made tests for goodness of fit to semiparametric hypotheses," *The Annals of Statistics*, 34, 721 – 741.

Bravo, F., Escanciano, J. C., and Van Keilegom, I. (2020), "Two-step semiparametric empirical likelihood inference," *Ann. Statist.*, 48, 1–26.

Card, D. (1995), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, eds. Christofides, L. N., Grant, E. K., and Swidinsky, R., University of Toronto Press, pp. 201–222.

— (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69, 1127–1160.

Chamberlain, G. (1986), "Asymptotic efficiency in semi-parametric models with censoring," *Journal of Econometrics*, 32, 189–218.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. M. (2022), "Locally Robust Semiparametric Estimation," *Econometrica*, 90, 1501–1535.

Chernozhukov, V., Hansen, C., and Jansson, M. (2009), "ADMISSIBLE INVARIANT SIMILAR TESTS FOR INSTRUMENTAL VARIABLES REGRESSION," *Econometric Theory*, 25, 806—818.

Chernozhukov, V., Hansen, C., and Spindler, M. (2015), "Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach," *Annual Review of Economics*, 7, 649–688.

Chetverikov, D., Santos, A., and Shaikh, A. M. (2018), "The Econometrics of Shape Restrictions," *Annual Review of Economics*, 10, 31–63.

Choi, S., Hall, W. J., and Schick, A. (1996), "Asymptotically uniformly most powerful tests in parametric and semiparametric models," *Ann. Statist.*, 24, 841–861.

Comon, P. (1994), "Independent component analysis, A new concept?" *Signal Processing*, 36, 287–314.

Conway, J. B. (1985), *A Course in Functional Analysis*, New York, NY, USA: Springer.

Davidson, J. (2021), *Stochastic limit theory*, Oxford University Press, 2nd ed.

Dufour, J.-M. (1997), "Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1387.

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

Elliott, G., Müller, U. K., and Watson, M. W. (2015), "Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811.

Escanciano, J. C. (2022), "Semiparametric Identification and Fisher Information," *Econometric Theory*, 38, 301–338.

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

Han, S. and McCloskey, A. (2019), "Estimation and inference with a (nearly) singular Jacobian," *Quantitative Economics*, 10, 1019–1068.

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

Horowitz, J. L. (2009), *Semiparametric and Nonparametric Methods in Econometrics*, Springer-Verlag New York.

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 Tracts in Mathematics, Cambridge University Press.

Jansson, M. (2008), "Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis," *Econometrica*, 76, 1103–1142.

Kaji, T. (2021), "Theory of Weak Identification in Semiparametric Models," *Econometrica*, 89, 733–763.

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

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

— (2005), "Testing Parameters in GMM Without Assuming that They Are Identified," *Econometrica*, 73, 1103–1123.

Kuchibhotla, A. K., Patra, R. K., and Sen, B. (2023), "Semiparametric Efficiency in Convexity Constrained Single-Index Model," *Journal of the American Statistical Association*, 118, 272–286.

Le Cam, L. M. (1960), *Locally Asymptotically Normal Families of Distributions: Certain Approximations to Families of Distributions and Their Use in the Theory of Estimation and Testing Hypotheses*, University of California Berkeley, Calif: University of California publications in statistics, University of California Press.

— (1986), *Asymptotic Methods in Statistical Decision Theory*, Springer Series in Statistics, New York, NY, USA: Springer.

Le Cam, L. M. and Yang, G. L. (2000), *Asypmtotics in Statistics: Some Basic Concepts*, New York, NY, USA: Springer, 2nd ed.

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

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

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

McCloskey, A. (2017), "Bonferroni-based size-correction for nonstandard testing problems," *Journal of Econometrics*, 200, 17–35.

McNeney, B. and Wellner, J. A. (2000), "Application of convolution theorems in semiparametric models with non-i.i.d. data," *Journal of Statistical Planning and Inference*, 91, 441 – 480.

Meyer, M. C. (2008), "Inference Using Shape-Restricted Regression Splines," *The Annals of Applied Statistics*, 2, 1013–1033.

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

— (2009), "Tests with correct size when instruments can be arbitrarily weak," *Journal of Econometrics*, 152, 131–140.

Newey, W. K. (1990), "Semiparametric efficiency bounds," *Journal of Applied Econometrics*, 5, 99–135.

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.

Newey, W. K. and Stoker, T. M. (1993), "Efficiency of Weighted Average Derivative Estimators and Index Models," *Econometrica*, 61, 1199–1223.

Neyman, J. (1959), "Optimal Asymptotic Tests of Composite Statistical Hypotheses," in *Probability and Statistics, the Harald Cramér Volume*, ed. Grenander, U., New York, USA: Wiley.

— (1979), "C(α) Tests and Their Use," *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 41, 1–21.

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.

Ritov, Y. and Bickel, P. J. (1990), "Achieving Information Bounds in Non and Semiparametric Models," *The Annals of Statistics*, 18, 925 – 938.

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

Romano, J. P. and Shaikh, A. M. (2012), "On the uniform asymptotic validity of subsampling and the bootstrap," *The Annals of Statistics*, 40, 2798 – 2822.

Rothenberg, T. J. (1971), "Identification in Parametric Models," *Econometrica*, 39, 577–591.

Rotnitzky, A., Cox, D. R., Bottai, M., and Robins, J. (2000), "Likelihood-Based Inference with Singular Information Matrix," *Bernoulli*, 6, 243–284.

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

Staiger, D. and Stock, J. H. (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Stock, J. H. and Wright, J. H. (2000), "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

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

Swensen, A. R. (1985), "The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend," *Journal of Multivariate Analysis*, 16, 54–70.

Taniguchi, M. and Kakizawa, Y. (2000), *Asymptotic Theory of Statistical Inference for Time Series*, Springer Series in Statistics, Springer New York.

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.

Wald, A. (1943), "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54, 426–482.

A Proofs of the main results

Proof of Proposition 3.1. Combination of Assumptions 3.1 and 3.2 yields

$$\left(g'_{n,\theta,\lambda}, L_{n,\gamma}(h)\right) \overset{P_{n,\gamma,0}}{\rightsquigarrow} \mathcal{N}\left(\begin{pmatrix} 0 \\ -\frac{1}{2}\sigma_\gamma(h) \end{pmatrix}, \begin{pmatrix} V_\gamma & \tau'\Sigma'_{\gamma,21} \\ \Sigma_{\gamma,21}\tau & \sigma_\gamma(h) \end{pmatrix}\right).$$

Consequently, by Le Cam's third Lemma (e.g. van der Vaart, 1998, Example 6.7)

$$g_{n,\gamma} \overset{P_{n,\gamma,h}}{\rightsquigarrow} Z_\tau \mathcal{N}\left(\Sigma_{\gamma,21}\tau, V_\gamma\right).$$

The second claim follows by combining the preceding display with Assumption 3.3(i), Remark 3.1 and Slutsky's Theorem. Combining the second claim with Assumption 3.3(ii), Slutsky's Theorem and the continuous mapping Theorem, the asymptotic distribution of $\hat{S}_{n,\theta}$ under $P_{n,\gamma,h}$ is the law of

$$Z'V_\gamma^\dagger Z, \qquad Z \sim \mathcal{N}\left(\Sigma_{\gamma,21}\tau, V_\gamma\right).$$

That this law is the indicated non-central χ^2_r distribution follows from Theorem 9.2.3 of Rao and Mitra (1971). □

Proof of Theorem 3.1. Since $\hat{r}_n \xrightarrow{P_{n,\gamma}} r$ we have that $P_{n,\gamma}\{c_n = c_\alpha\} \to 1$. Therefore, by Proposition 3.1, Remark 3.1 and Slutsky's Theorem that $\hat{S}_{n,\theta} - c_n \rightsquigarrow S - c$ under $P_{n,\gamma,h}$ where $S \sim \chi^2_r(a)$. Since the χ^2_r distribution is continuous, by the Portmanteau Theorem,

$$\lim_{n\to\infty} P_{n,\gamma,h}\psi_{n,\theta} = \lim_{n\to\infty} P_{n,\gamma,h}\left(\hat{S}_{n,\theta} > c_n\right) = L\{S - c > 0\} = 1 - \mathrm{P}(\chi^2_r(a) \le c_\alpha),$$

for L the law of S.

For the second result, note that if $\operatorname{rank}(\hat{\Lambda}_{n,\theta}) \xrightarrow{P_{n,\gamma}} r = 0$, then $P_{n,\gamma}(R_n) \to 1$ for $R_n := \{\hat{\Lambda}_{n,\theta} = 0\}$. On this set $\hat{S}_{n,\theta} = 0$ and so $\psi_{n,\theta} = 0$. By Remark 3.1, $P_{n,\gamma,h}(R_n) \to 1$ and therefore $P_{n,\gamma,h}\psi_{n,\theta} \le 1 - P_{n,\gamma,h}(R_n) \to 0$. □

Proof of Corollary 3.1. Since $\tau = 0$, $\chi^2_r(a) = \chi^2_r(0) = \chi^2_r$, and hence $\mathrm{P}(\chi^2_r(a) \le c_\alpha) = 1 - \alpha$. Apply Theorem 3.1 along the stated sequences. □

Proof of Corollary 3.2. Note that $P_{n,\gamma,h_n}(\theta \in C_n) = 1 - P_{n,\gamma,h_n}\psi_{n,\theta}$ and apply Corollary 3.1. □

Proof of Corollary 3.3. Let $\pi_n(h) := P_{n,\gamma,h}\psi_{n,\theta}$ and $\pi(h) := 1 - \mathrm{P}(\chi^2_r(a) \le c_r)$ if $r \ge 1$ or $\pi(h) := 0$ if $r = 0$. By Remark 3.4, $\pi_n(h) \to \pi(h)$ pointwise in $h \in H_\gamma$. Since the π_n are asymptotically equicontinuous on K, the convergence is also uniform on K. □

Proof of Corollary 3.4. Let $\pi_n(h) := P_{n,\gamma,h}\psi_{n,\theta}$ and $\pi(h) := \alpha$ if $r \ge 1$ or $\pi(h) := 0$ if $r = 0$. By Corollary 3.1, $\pi_n(h) \to \pi(h) =: c$ pointwise in $h \in H_{\gamma,0}$. Since the π_n

are asymptotically equicontinuous on K, the convergence is also uniform on K. Hence,

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

Proof of Lemma 3.1. This follows from the fact that (e.g. Strasser, 1985, Theorem 2.3)

$$|P_{n,\gamma,h}\psi_{n,\theta}-P_{n,\gamma,h'}\psi_{n,\theta}| \le d_{TV}(P_{n,\gamma,h},P_{n,\gamma,h'}). \qquad \square$$

Proof of Lemma 3.2. The proof is given for the case where (H_γ,d) is a pseudo-metric space and $K\subset H_\gamma$. The argument in the case with $H_{\gamma,0}$ replacing H_γ is analogous.

First suppose $r\ge 1$. Then, by asymptotic equicontinuity of $h\mapsto Q_{n,\gamma,h}$ on K, one has for any $h_n\to h$ (through K) that $\delta(Q_{n,\gamma,h_n},Q_{n,\gamma,h})\to 0$ as $n\to\infty$ and by the asymptotic equicontinuity (on K) of $h\mapsto P_{n,\gamma,h}(\hat r_{n,\theta}=r)$ one has that $|P_{n,\gamma,h}(\hat r_{n,\theta}=r)-P_{n,\gamma,h_n}(\hat r_{n,\theta}=r)|\to 0$. Since $\hat r_{n,\theta}\xrightarrow{P_{n,\gamma,h}} r$ (Assumption 3.3 (iii) and Remark 3.1), it follows that also $\hat r_{n,\theta}\xrightarrow{P_{n,\gamma,h_n}} r$. Hence, under P_{n,γ,h_n},

$$\hat S_{n,\theta}-c_n \rightsquigarrow S-c_r,\quad S\sim\chi^2_r(a) \implies P_{n,\gamma,h_n}\varphi_{n,\theta}\to 1-\mathrm{P}(\chi^2_r(a)\le c_r)=:\pi(\tau),$$

by Proposition 3.1 where c_r and a are as in Theorem 3.1. Thus, by Theorem 3.1,

$$|P_{n,\gamma,h_n}\varphi_{n,\theta}-P_{n,\gamma,h}\varphi_{n,\theta}| \le |P_{n,\gamma,h_n}\varphi_{n,\theta}-\pi(\tau)|+|P_{n,\gamma,h}\varphi_{n,\theta}-\pi(\tau)|\to 0,$$

which implies the required asymptotic equicontinuity on K.

In the case where $r=0$, the asymptotic equicontinuity on K of $h\mapsto P_{n,\gamma,h}(\hat\Lambda_{n,\theta}=0)$ implies that for any $h_n\to h$ (through K), $|P_{n,\gamma,h_n}(\hat\Lambda_{n,\theta}=0)-P_{n,\gamma,h}(\hat\Lambda_{n,\theta}=0)|\to 0$. In combination with $\operatorname{rank}(\hat\Lambda_{n,\theta})\xrightarrow{P_{n,\gamma,h}} 0$ (Assumption 3.3 (iii) and Remark 3.1), this implies that $P_{n,\gamma,h_n}(\hat\Lambda_{n,\theta}=0)\to 1$ and thus $P_{n,\gamma,h_n}\varphi_{n,\theta}\to 0$. Thus, by Theorem 3.1

$$|P_{n,\gamma,h_n}\varphi_{n,\theta}-P_{n,\gamma,h}\varphi_{n,\theta}| \le |P_{n,\gamma,h_n}\varphi_{n,\theta}|+|P_{n,\gamma,h}\varphi_{n,\theta}|\to 0,$$

which implies the required asymptotic equicontinuity on K. $\square$

Proof of Lemma 3.3. For $a,b\in\mathbb{R}$ and $h_1,h_2\in H_\gamma$ we have

$$\Delta_{n,\gamma}(a_1h_1+a_2h_2)=a_1\Delta_{n,\gamma}h_1+a_2\Delta_{n,\gamma}h_2$$

and in consequence $\Delta_\gamma(a_1h_1+a_2h_2)=a_1\Delta_\gamma(h_1)+a_2\Delta_\gamma(h_2)$, hence Δ_γ is linear. We next note that K_γ is well-defined and a bilinear, symmetric, positive semidefinite (i.e. covariance) kernel. Firstly, since for $h\in H_\gamma$, we have $\sigma_\gamma(h)=\lim_{n\to\infty}\|\Delta_{n,\gamma}h\|^2$, $(\|\Delta_{n,\gamma}h\|^2)_{n\in\mathbb{N}}$ is bounded and Cauchy (in $\mathbb{R}$). Letting $K_{n,\gamma}(h,g):=P_{n,\gamma}[\Delta_{n,\gamma}h\Delta_{n,\gamma}g]$ and using Cauchy – Schwarz

$$\begin{aligned}|K_{n,\gamma}(h,g)-K_{m,\gamma}(h,g)| &\le |P_{n,\gamma}[(\Delta_{n,\gamma}h-\Delta_{m,\gamma}h)\Delta_{n,\gamma}g]+P_{n,\gamma}[\Delta_{m,\gamma}h(\Delta_{n,\gamma}g-\Delta_{m,\gamma}g)]\\ &\le \|\Delta_{n,\gamma}h-\Delta_{m,\gamma}h\|\|\Delta_{n,\gamma}g\|+\|\Delta_{m,\gamma}h\|\|\Delta_{n,\gamma}g-\Delta_{m,\gamma}g\|,\end{aligned}$$

hence $(K_{n,\gamma}(h,g))_{n\in\mathbb{N}}$ is also Cauchy and thus has a limit in $\mathbb{R}$. Bilinearity of K_γ follows directly from its definition and the linearity of $\Delta_{n,\gamma}$. Symmetry follows from the symmetry of multiplication in $\mathbb{R}$. For positive semi-definiteness, let $h_1,\ldots,h_K \in H_\gamma$, $a \in \mathbb{R}^K$. Since each $\Delta_{n,\gamma}h \in L_2^0(P_{n,\gamma})$, each matrix $\mathcal{K}_n := [K_{n,\gamma}(h_k,h_j)]_{k,j=1}^K$ is a covariance matrix and hence positive semi-definite. Therefore, for each n,

$$\sum_{k=1}^K\sum_{j=1}^K a_k a_j K_{n,\gamma}(h_k,h_j) = a'\mathcal{K}_n a \geq 0,$$

hence the same holds in the limit, i.e. with $K_{n,\gamma}$ and $\mathcal{K}_n$ replaced by K_γ and $\mathcal{K} := [K_\gamma(h_k,h_j)]_{k,j=1}^K$ respectively.

Finally, by Assumption 3.1 and the fact that $K(h,h) = \sigma_\gamma(h)$, $\Delta_\gamma h \sim \mathcal{N}(0, K(h,h))$. Using this along with the linearity of $\Delta_{n,\gamma}$, Δ_γ and the bilinearity of K,

$$\sum_{k=1}^K a_k\Delta_{n,\gamma}(h_k) = \Delta_{n,\gamma}\left(\sum_{k=1}^K a_k h_k\right) \overset{P_{n,\gamma}}{\rightsquigarrow} \Delta_\gamma\left(\sum_{k=1}^K a_k h_k\right) = \sum_{k=1}^K a_k\Delta_\gamma h_k \sim \mathcal{N}\left(0, a'\mathcal{K}a\right),$$

where $\mathcal{K} := [K(h_k,h_j)]_{k,j=1}^K$. Therefore, by the Cramér – Wold Theorem, Δ_γ is a mean-zero Gaussian process with covariance kernel K. ☐

Proof of Proposition 3.2. Remark 3.1 and the transitivity of (mutual) contiguity ensures that the experiments $\mathscr{E}_{n,\gamma}$ are contiguous (cf. Strasser, 1985, Definition 61.1). Hence by Theorem 61.6 of Strasser (1985) it suffices to show that the finite dimensional marginal distributions of $h \mapsto L_{n,\gamma}(h)$ converge (under $P_{n,\gamma}$) to those of $h \mapsto L(h) := \Delta_\gamma h - \frac{1}{2}\|h\|_\gamma^2$. For this it is enough to note that the finite dimensional marginal distributions of $\Delta_{n,\gamma}$ converge to those of Δ_γ (under $P_{n,\gamma}$), as follows by the Cramér – Wold Theorem (as in the proof of Lemma 3.3). ☐

Proof of Proposition 3.3. Let $G_{[0]} := P_{\gamma,0}$. Define a map $Z : \mathbb{H}_\gamma \to L_2(\Omega,\mathcal{F},G_{[0]})$ according to $Z[h] = \Delta_\gamma(h)$ for any arbitrary $h \in \pi_V^{-1}([h])$, where π_V is the quotient map from $H_\gamma \to \mathbb{H}_\gamma$; that this is well defined is noted in footnote 60. By the definition of $\langle\cdot\,,\,\cdot\rangle_\gamma$ on $\mathbb{H}_\gamma$ (cf. subsection S2.2) this is a standard Gaussian process for $\mathbb{H}_\gamma$.[59] Let each $G_{[h]}$ be defined such that

$$\frac{\mathrm{d}G_{[h]}}{\mathrm{d}G_{[0]}} = \exp\left(Z[h] - \frac{1}{2}\|[h]\|_\gamma^2\right),$$

and note that by Theorem 69.4 in Strasser (1985), $\mathscr{G}_\gamma$ is a Gaussian shift on $(\mathbb{H}_\gamma, \langle\cdot\,,\,\cdot\rangle_\gamma)$.

For any $h \in H_\gamma$ we have that $Z[h] = \Delta_\gamma g$ for some $g \in \pi_V^{-1}([h])$ and $\Delta_\gamma h = \Delta_\gamma g$ $P_{\gamma,0}$–almost surely by footnote 60. Since $\|h\|_\gamma = \|[h]\|_\gamma$ (cf. subsection S2.2) one therefore has that

$$\frac{\mathrm{d}G_{[h]}}{\mathrm{d}G_{[0]}} = \exp\left(Z[h] - \frac{1}{2}\|[h]\|_\gamma^2\right) = \exp\left(\Delta_\gamma h - \frac{1}{2}\|h\|_\gamma^2\right) = \frac{\mathrm{d}P_{\gamma,h}}{\mathrm{d}P_{\gamma_0}}, \quad P_{\gamma,0} - \text{a.s.}.$$

[59] More formally, it is the restriction to $\mathbb{H}_\gamma$ of a standard Gaussian process for $\overline{\mathbb{H}_\gamma}$ (cf. Definition 68.3 in Strasser, 1985).

By construction, each $P_{\gamma,h}$ and $G_{[h]}$ is absolutely continuous with respect to $P_{\gamma,0} = G_{[0]}$. Hence, by Lemma 2.4 in Strasser (1985),

$$d_{TV}(P_{\gamma,h}, G_{[h]}) = \frac{1}{2}\int \left|\frac{\mathrm{d}P_{\gamma,h}}{\mathrm{d}P_{\gamma 0}} - \frac{\mathrm{d}G_{[h]}}{\mathrm{d}P_{\gamma 0}}\right| \mathrm{d}P_{\gamma,0} = \frac{1}{2}\int \left|\frac{\mathrm{d}P_{\gamma,h}}{\mathrm{d}P_{\gamma 0}} - \frac{\mathrm{d}G_{[h]}}{\mathrm{d}G_{[0]}}\right| \mathrm{d}P_{\gamma,0} = 0.$$

□

Proof of Lemma 3.4. We have that

$$\begin{aligned}
\langle [(\tau,b)]\,,\,([t,g])\rangle_\gamma &= \langle [(\tau,0)] + [(0,b)]\,,\,[(t,0)] + [(0,g)]\rangle_\gamma \\
&= \left\langle \Pi^\perp[(\tau,0)] + \Pi[(\tau,0)] + [(0,b)]\,,\,\Pi^\perp[(t,0)] + \Pi[(t,0)] + [(0,g)]\right\rangle_\gamma \\
&= \left\langle \Pi^\perp[(\tau,0)]\,,\,\Pi^\perp[(t,0)]\right\rangle_\gamma + \langle \Pi[(\tau,0)] + [(0,b)]\,,\,\Pi[(t,0)] + [(0,g)]\rangle_\gamma \\
&= \tau'\tilde{\mathcal{I}}_\gamma t + \langle \Pi[(\tau,0)] + [(0,b)]\,,\,\Pi[(t,0)] + [(0,g)]\rangle_\gamma\,.
\end{aligned} \tag{44}$$

From this and the fact that $\Pi[(\tau,0)] \in \ker \pi_1$, we obtain

$$\|[\tau]\|^2 = \inf_{b\in B_\gamma} \|[(\tau,b)]\|_\gamma^2 = \inf_{[h]\in\ker\pi_1} \|[(\tau,0)] - [h]\|_\gamma^2 = \tau'\tilde{\mathcal{I}}_\gamma\tau + \inf_{[h]\in\ker\pi_1} \|\Pi[(\tau,0)] - [h]\|_\gamma^2 = \tau'\tilde{\mathcal{I}}_\gamma\tau.$$

Hence, if $\|\tau\| = \|[\tau]\| = 0$, it follows that $\tau'\tilde{\mathcal{I}}_\gamma\tau = 0$. Hence $\tilde{\mathcal{I}}_\gamma^{1/2}\tau = 0$. But then $\tilde{\mathcal{I}}_\gamma\tau = \tilde{\mathcal{I}}_\gamma^{1/2}\tilde{\mathcal{I}}_\gamma^{1/2}\tau = 0$. Conversely suppose that $\tau \in \ker \tilde{\mathcal{I}}_\gamma$. Then, $\|\tau\|^2 = \tau'\tilde{\mathcal{I}}_\gamma\tau = 0$ and hence $\|\tau\| = 0$. □

Proof of Lemma 3.5. Let $V := \{h \in H_\gamma : \|h\|_\gamma = 0\}$ and let $\pi_V : H_\gamma \to \mathbb{H}_\gamma = H_\gamma / V$ be the canonical projection. For $h \in H_\gamma$ we write $[h] := \pi_V(h)$. Define $Z : \overline{\mathbb{H}_\gamma} \to L_2(\mathrm{P})$ as $Z[h] = \Delta_\gamma(h)$, which is a mean-zero linear Gaussian process with covariance kernel $K([h],[g]) = K_\gamma(h,g) = \langle [h]\,,\,[g]\rangle_\gamma$.[60] $Y := \operatorname{ran} Z \subset L_2(\mathrm{P})$ is evidently a linear space and

$$\langle [h]\,,\,[g]\rangle := \mathbb{E}\left[Z[h]Z[g]\right] = K([h],[g]) = \langle [h]\,,\,[g]\rangle_\gamma\,, \quad [h],[g] \in \overline{\mathbb{H}_\gamma}.$$

The above display and the completeness of $\overline{\mathbb{H}_\gamma}$ yield the closedness of $\operatorname{ran} Z$. Hence Z is a Hilbert space isomorphism from $\overline{\mathbb{H}_\gamma}$ to Y. As such it is bijective, with $Z^* = Z^{-1}$. We have

$$\mathcal{T} = \{\Delta_\gamma(h) : h = (0,b) \in H_\gamma\} = \{Z[h] : h = (0,b) \in H_\gamma\} = \{Z[h] : [h] \in \ker \pi_1'\},$$

where π_1' denotes the restriction of π_1 to $\mathbb{H}_\gamma$. We next note 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 $[h] \in \ker \pi_1'$,

$$\langle [g]\,,\,[h]\rangle_\gamma = \langle Z[g]\,,\,Z[h]\rangle_{L_2(P)} = 0.$$

Since $\ker \pi_1 = \operatorname{cl}\ker \pi_1'$ by Lemma S2.1, for any $[h] \in \ker \pi_1$ there is a sequence in $\ker \pi_1'$ which converges to $[h]$ and the result follows by taking limits. For the

[60] This is well-defined: for any other $g \in H_\gamma$ with $\pi_V(g) = [h]$, one has $\Delta_\gamma(g) = \Delta_\gamma(h) + \Delta_\gamma(v)$ where $\|v\|_\gamma = K(v,v)_\gamma = 0$ and hence $\Delta_\gamma(v) = 0$ a.s..

other inclusion note that a corollary of Lemma S2.1 is that $(\ker \pi_1)^\perp = (\ker \pi_1')^\perp$. Hence, if $[g] \in (\ker \pi_1)^\perp$, for any $[h] \in \ker \pi_1'$ we have

$$\langle Z[g]\,,\,Z[h]\rangle_{L_2(P)} = \langle [g]\,,\,[h]\rangle = 0.$$

Let Q denote the orthogonal projection on $(\ker \pi_1)^\perp \subset \overline{\mathbb{H}_\gamma}$ and R that on $\mathcal{T}^\perp \subset Y$. Since Z is a Hilbert space isomorphism one has

$$R\Delta_\gamma(h) = RZ[h] = ZQZ^*Z[h] = ZQZ^{-1}Z[h] = ZQ[h], \quad [h] \in \overline{\mathbb{H}_\gamma}.$$

Hence for $[h] = [e_i, 0]$, $RZ[e_i, 0] = ZQ[e_i, 0]$ implying

$$\mathbb{E}\left[\tilde{\Delta}_\gamma(e_i, 0)\tilde{\Delta}_\gamma(e_j, 0)\right] = \left\langle \Pi^\perp[e_i, 0]\,,\,\Pi^\perp[e_j, 0]\right\rangle_\gamma = \tilde{\mathcal{I}}_{\gamma,ij}. \qquad \square$$

Proof of Theorem 3.2. Define the bounded linear map $T : \overline{\mathbb{H}_\gamma} \to \mathbb{R}$ as

$$T[h] := \left\langle \Pi^\perp[1, 0]\,,\,\Pi^\perp[h]\right\rangle_\gamma = \left\langle \Pi^\perp[1, 0]\,,\,[h]\right\rangle_\gamma.$$

The equality in the preceding display follows from the idempotentcy and self-adjointness of orthogonal projections. The boundedness of T follows from the Cauchy - Schwarz inequality as orthogonal projections have norm one. For any $[h] = [\tau, b] \in \mathbb{H}_\gamma$,

$$T[h] = \left\langle \Pi^\perp[1, 0]\,,\,\Pi^\perp[\tau, b]\right\rangle_\gamma = \tilde{\mathcal{I}}_\gamma \tau. \tag{45}$$

Assume first that $\sqrt{\tilde{\mathcal{I}}_\gamma} \neq 0$, hence $\operatorname{rank}(\tilde{\mathcal{I}}_\gamma) = 1$. Then, let $[u] = \frac{\Pi^\perp[1,0]}{\sqrt{\tilde{\mathcal{I}}_\gamma}}$ and note that $\|[u]\| = 1$ and for any $[h] \in \ker T$, $\langle [u]\,,\,[h]\rangle_\gamma = \frac{1}{\sqrt{\tilde{\mathcal{I}}_\gamma}} T[h] = 0$, hence $[u] \in [\ker T]^\perp$ and $T[u] = \sqrt{\tilde{\mathcal{I}}_\gamma} > 0$. Lemma 71.5 in Strasser (1985) ensures that any unbiased level α test ϕ of $K_0 : T[h] = 0$ against $K_1 : T[h] \neq 0$ in the (restricted) Gaussian shift $\mathscr{G}_\gamma$ satisfies

$$G_{[h]}\phi \leq 1 - \Phi\left(z_{\alpha/2} - \tilde{\mathcal{I}}_\gamma^{1/2}\tau\right) + 1 - \Phi\left(z_{\alpha/2} + \tilde{\mathcal{I}}_\gamma^{1/2}\tau\right) \tag{46}$$

By Proposition 3.2, $\mathscr{E}_{n,\gamma} \rightsquigarrow \mathscr{E}_\gamma$ where the latter is dominated. Let $\pi_n(h) := P_{n,\gamma,h}\phi_n$ and fix an arbitrary $h^\star$ for which the bound will be shown. There is a subsequence $(n_m)_{m\in\mathbb{N}}$ along which $\lim_{m\to\infty} \pi_{n_m}(h^\star) = \limsup_{n\to\infty} \pi_n(h^\star)$. Since $[0, 1]^{H_\gamma}$ is compact in the product topology, there is a subnet $(n_{m(s)})_{s\in S}$ of $(n_m)_{m\in\mathbb{N}}$ and a function $\pi : H_\gamma \to [0, 1]$ such that $\lim_{s\in S} \pi_{n_{m(s)}}(h) = \pi(h)$ for all $h \in H_\gamma$. By our hypotheses and equation (45) for any h_0 such that $[h_0] \in \ker T \cap \mathbb{H}_\gamma$ and any h_1 such that $[h_1] \in \mathbb{H}_\gamma \setminus (\ker T \cap \mathbb{H}_\gamma)$

$$\pi(h_0) = \lim_{s\in S} \pi_{n_{m(s)}}(h_0) \leq \alpha \leq \lim_{s\in S} \pi_{n_{m(s)}}(h_1) = \pi(h_1).$$

Theorem 7.1 in van der Vaart (1991) ensures the existence of a test ϕ in $\mathscr{E}_\gamma$ with power function π. The above display and Proposition 3.3 ensures that ϕ is an unbiased and level α test of $\ker T \cap \mathbb{H}_\gamma$ against $\mathbb{H}_\gamma \setminus (\ker T \cap \mathbb{H}_\gamma)$ in $\mathscr{G}_\gamma$. Therefore by Proposition 3.3 again

$$\limsup_{n\to\infty} P_{n,\gamma,h^\star}\phi_n = \lim_{m\to\infty} \pi_{n_m}(h^\star) = \pi(h^\star) = P_{\gamma,h^\star}\phi = G_{[h^\star]}\phi,$$

and combination with (46) proves the result for this case.

To complete the proof suppose that $\tilde{\mathcal{I}}_\gamma = 0$. Then, $\operatorname{rank}(\tilde{\mathcal{I}}_\gamma) = 0$ and

$$1 - \Phi\left(z_{\alpha/2} - \tilde{\mathcal{I}}_\gamma^{1/2}\tau\right) + 1 - \Phi\left(z_{\alpha/2} + \tilde{\mathcal{I}}_\gamma^{1/2}\tau\right) = \alpha.$$

That this provides an upper bound for $\limsup_{n\to\infty} P_{n,\gamma,h}\phi_n$ follows from the assumption that the test is asymptotically of level α and Proposition 3.5. $\square$

Proof of Corollary 3.5. This follows from Theorem 3.1: if $\Sigma_{\gamma,21}V_\gamma^\dagger\Sigma'_{\gamma,s21} = \tilde{\mathcal{I}}_\gamma$ and $r = 1$,

$$\lim_{n\to\infty} P_{n,\gamma,h_n}\psi_{n,\theta} = 1 - \mathrm{P}(\chi_1^2(a) > c_\alpha) = 1 - \mathrm{P}(Z^2 > c_\alpha), \quad Z \sim \mathcal{N}\left(\tilde{\mathcal{I}}_\gamma^{1/2}\tau, 1\right).$$

Elementary manipulations show that

$$1 - \mathrm{P}(Z^2 > c_\alpha) = 1 - \Phi\left(z_{\alpha/2} - \tilde{\mathcal{I}}_\gamma^{1/2}\tau\right) + 1 - \Phi\left(z_{\alpha/2} + \tilde{\mathcal{I}}_\gamma^{1/2}\tau\right). \qquad \square$$

Proof of Theorem 3.3. π_1 is valued in $\mathbb{H}_{\gamma,1} = \mathbb{R}^{d_\theta} / \ker\tilde{\mathcal{I}}_\gamma$, by Lemma 3.4. It is also surjective: for any $[\tau] \in \mathbb{R}^{d_\theta} / \ker\tilde{\mathcal{I}}_\gamma$ let $t \in \pi_{\ker\tilde{\mathcal{I}}_\gamma}^{-1}(\{[\tau]\})$ where $\pi_{\ker\tilde{\mathcal{I}}_\gamma}$ is the quotient map between $\mathbb{R}^{d_\theta}$ and $\mathbb{H}_{\gamma,1}$. Then $\pi_1[(t,0)] = [t] = [\tau]$. Therefore, since $\mathbb{H}_{\gamma,1} = \mathbb{R}^{d_\theta} / \ker\tilde{\mathcal{I}}_\gamma \approx \operatorname{ran}\tilde{\mathcal{I}}_\gamma$ (e.g. Roman, 2005, Theorem 3.5), $\dim\operatorname{ran}\pi_1 = r$. The codimension of $\ker\pi_1$ is also r since $\overline{\mathbb{H}_\gamma} / \ker\pi_1 \approx \operatorname{ran}\pi_1$ (e.g. Roman, 2005, Theorem 3.5). By linearity and $[(0,b)] \in \ker\pi_1$, $\Pi[(\tau,b)] = \Pi[(\tau,0)] + [(0,b)]$. This, along with Lemma 3.4 and e.g. Theorem 2.6 in Conway (1985), yields $\|[(\tau,b)] - \Pi[(\tau,b)]\|_\gamma^2 = \|[(\tau,0)] - \Pi[(\tau,0)]\|_\gamma^2 = \|[\tau]\|^2 = \tau'\tilde{\mathcal{I}}_\gamma\tau$.

Define the sets

$$M_a := \left\{[(\tau,b)] \in \mathbb{H}_\gamma : \tau'\tilde{\mathcal{I}}_\gamma\tau = a\right\}, \qquad \bar{M}_a := \left\{[(\tau,b)] \in \overline{\mathbb{H}_\gamma} : \tau'\tilde{\mathcal{I}}_\gamma\tau = a\right\}.$$

Suppose that ϕ is a test on $\mathscr{G}_\gamma$ such that $G_{[0]}\phi \le \alpha$, where $G_{[h]}$ denotes the Gaussian measure under $[h] \in \overline{\mathbb{H}_\gamma}$. We first consider the case where $a > 0$. ϕ is a level α test of $K_0 : \{[0]\}$ against $K_1 : [\ker\pi_1]^\perp \setminus \{[0]\}$ in the restriction of the standard Gaussian shift experiment on $[\ker\pi_1]^\perp$.[61] Then by Theorem 30.2 in Strasser (1985)

$$\inf_{[h]\in\bar{M}_a} G_{[h]}\phi \le \inf_{[h]\in\bar{M}_a\cap[\ker\pi_1]^\perp} G_{[h]}\phi \le 1 - \mathrm{P}(\chi_r^2(a) \le c_r).$$

We claim that $\operatorname{cl} M_a = \overline{M}_a$. The function $q : \overline{\mathbb{H}_\gamma} \to \mathbb{R}$ given by $q([(\tau,b)]) =$

[61]Note that $[\ker\pi_1]^\perp$ is a finite dimensional linear subspace (of $\overline{\mathbb{H}_\gamma}$) since it is isomorphic to $\overline{\mathbb{H}_\gamma} / \ker\pi_1$ which is of dimension r.

$\|[(\tau,b)]-\Pi[(\tau,b)]\|_\gamma^2 = \tau'\tilde{\mathcal{I}}_\gamma\tau$ is continuous. Hence, if $M_a \ni [(\tau_n,b_n)] \to [(\tau,b)]$, $\tau'\tilde{\mathcal{I}}_\gamma\tau = a$ and hence $\operatorname{cl} M_a \subset \overline{M}_a$. For the converse, let $[(\tau,b)] \in \overline{M}_a$ and suppose that $([(\tau_n,b_n)])_{n\in\mathbb{N}} \subset \mathbb{H}_\gamma$ converges to $[(\tau,b)]$. Since q is continuous, $a_n := q([(\tau_n,b_n)]) > 0$ for all large enough n and $a_n \to a$. Let $[(\tilde{\tau}_n,\tilde{b}_n)] := \frac{\sqrt{a}}{\sqrt{a_n}}[(\tau_n,b_n)]$ and observe that

$$\tilde{\tau}_n'\tilde{\mathcal{I}}_\gamma\tilde{\tau}_n = \|[(\tilde{\tau}_n,\tilde{b}_n)]-\Pi[(\tilde{\tau}_n,\tilde{b}_n)]\|_\gamma^2 = \frac{a}{a_n}\|[(\tau_n,b_n)]-\Pi[(\tau_n,b_n)]\|_\gamma^2 = \frac{a}{a_n}q([(\tau_n,b_n)]) = a,$$

and $[(\tilde{\tau}_n,\tilde{b}_n)] \to [(\tau,b)]$. Thus $\overline{M}_a \subset \operatorname{cl} M_a$. Hence $\operatorname{cl} M_a = \overline{M}_a$. Since $[h] \mapsto G_{[h]}\phi$ is continuous, by the preceding display we have:[62]

$$\inf_{[h]\in M_a} G_{[h]}\phi = \inf_{[h]\in \bar{M}_a} G_{[h]}\phi \le 1 - \mathrm{P}(\chi_r^2(a) \le c_r). \tag{47}$$

For the case where $a = 0$, note that $[0] \in M_0$ and therefore, since ϕ is of level α,

$$\inf_{[h]\in M_0} G_{[h]}\phi \le G_{[0]}\phi \le \alpha = 1 - \mathrm{P}(\chi_r^2(0) \le c_r). \tag{48}$$

Fix $a \ge 0$ and let $\mathcal{R} := 1 - \mathrm{P}(\chi_r^2(a) \le c_r)$. Let $\pi_n(h) := P_{n,\gamma,h}\phi_n$ and suppose that for some $\varepsilon > 0$,

$$\limsup_{n\to\infty} \inf\left\{P_{n,\gamma,h}\phi_n : h = (\tau,b) \in H_\gamma,\ \tau'\tilde{\mathcal{I}}_\gamma\tau = a\right\} \ge \mathcal{R} + \varepsilon.$$

It immediately follows that

$$\limsup_{n\to\infty} \inf\left\{\pi_n(h) : h = (\tau,b) \in \tilde{H}_\gamma,\ \tau'\tilde{\mathcal{I}}_\gamma\tau = a\right\} \ge \mathcal{R} + \varepsilon,$$

and so, for some subsequence $(n_m)_{m\in\mathbb{N}}$,

$$\lim_{m\to\infty} \inf\left\{\pi_{n_m}(h) : h = (\tau,b) \in \tilde{H}_\gamma,\ \tau'\tilde{\mathcal{I}}_\gamma\tau = a\right\} \ge \mathcal{R} + \varepsilon.$$

Since $[0,1]^{H_\gamma}$ is compact in the product topology, there is a subnet $(n_{m(s)})_{s\in S}$ of $(n_m)_{m\in\mathbb{N}}$ and a function $\pi : H_\gamma \to [0,1]$ such that $\lim_{s\in S}\pi_{n_{m(s)}}(h) = \pi(h)$ for all $h \in H_\gamma$. Take any h such that $[h] \in M_a$ and note that the preceding display implies that

$$\pi(h) = \lim_{s\in S}\pi_{n_{m(s)}}(h) \ge \lim_{s\in S}\inf\left\{\pi_{n_{m(s)}}(h) : h = (\tau,b) \in \tilde{H}_\gamma,\ \tau'\tilde{\mathcal{I}}_\gamma\tau = a\right\} \ge \mathcal{R} + \varepsilon. \tag{49}$$

By Proposition 3.2, $\mathscr{E}_{n,\gamma} \rightsquigarrow \mathscr{E}_\gamma$ where the latter is dominated and hence by Theorem 7.1 in van der Vaart (1991) there is a test ϕ in $\mathscr{E}_\gamma$ with power function π. Now consider the restriction of $\mathscr{G}_\gamma$ to $[\ker \pi_1]^\perp$. By hypothesis, Corollary S2.1

[62]The continuity of the indicated map follows from the fact that Gaussian shift experiments are continuous in the total variation norm.

and Proposition 3.3

$$G_{[0]}\phi = P_{\gamma,0}\phi = \pi(0) = \lim_{s\in S} \pi_{n_m(s)}(0) \le \limsup \pi_n(0) \le \alpha.$$

Hence ϕ is a test of level α of K_0 against K_1 in this experiment, and

$$\inf_{[h]\in M_a} G_{[h]}\phi = \inf_{h:[h]\in M_a} P_{\gamma,h}\phi = \inf_{h:[h]\in M_a} \pi(h) \ge \mathcal{R} + \varepsilon,$$

by (49) and Proposition 3.3, but this contradicts either (47) if $a > 0$ or (48) if $a = 0$. □

Proof of Corollary 3.6. Equation (16) follows from Theorem 3.1, since $r > 0$ and

$$a = \tau\Sigma'_{\gamma,21} V_\gamma^\dagger \Sigma_{\gamma,21}\tau = \tau\tilde{\mathcal{I}}_\gamma\tau.$$

For the second part, let $f_n(h) := P_{n,\gamma,h}\psi_{n,\theta}$. By the pointwise convergence given in equation (16) and the asymptotic equicontinuity assumption,

$$\lim_{n\to\infty} f_n(h) = 1 - \mathrm{P}\left(\chi_r^2\left(\tau'\tilde{\mathcal{I}}_\gamma\tau\right) \le c_r\right) =: f(h),$$

uniformly on K_a. We may conclude that for any sequence $h_n \to h \in K_a$,

$$\lim_{n\to\infty} f_n(h_n) = f(h) \ge f_\star := 1 - \mathrm{P}\left(\chi_r^2\left(a\right) \le c_r\right). \tag{50}$$

In light of the preceding lower bound, if (17) does not hold, there must exist a sequence $h_n \in K_a$ such that $\limsup_{n\to\infty} f_n(h_n) < f_\star$. Since K_a is compact, there is a subsequence $h_{n_m} \to h \in K_a$. Embed this into a full sequence as $h_m^* := h_{n_1}$ for $m = 1, \ldots, n_1$ and $h_m^* := h_{n_k}$ for $n_k \le m < n_{k+1}$. Then $f_{n_m}(h_{n_m})$ is a subsequence of $f_m(h_m^*)$ and $h_m^* \to h$, so by (50)

$$\lim_{m\to\infty} f_{n_m}(h_{n_m}) = \lim_{m\to\infty} f_m(h_m^*) = f(h) \ge f_\star,$$

which contradicts $\limsup_{n\to\infty} f_n(h_n) < f_\star$. □

Proof of Lemma 3.6. By asymptotic equicontinuity of $h \mapsto Q_{n,\gamma,h}$ on K, one has for any $h_n \to h$ (through K) that $\delta(Q_{n,\gamma,h_n}, Q_{n,\gamma,h}) \to 0$ as $n \to \infty$ and by the asymptotic equicontinuity of $h \mapsto P_{n,\gamma,h}(\hat{r}_{n,\theta} = r)$ (on K) one has that $|P_{n,\gamma,h}(\hat{r}_{n,\theta} = r) - P_{n,\gamma,h_n}(\hat{r}_{n,\theta} = r)| \to 0$. Since $\hat{r}_{n,\theta} \xrightarrow{P_{n,\gamma,h}} r$ (Assumption 3.3 (iii) and Remark 3.1), it follows that also $\hat{r}_{n,\theta} \xrightarrow{P_{n,\gamma,h_n}} r$. Hence, under P_{n,γ,h_n},

$$\hat{S}_{n,\theta} - c_n \rightsquigarrow S - c_r, \quad S \sim \chi_r^2(a) \implies P_{n,\gamma,h_n}\psi_{n,\theta} \to 1 - \mathrm{P}(\chi_r^2(a) \le c_r) =: \pi(a),$$

by Proposition 3.1 where $a = \tau'\tilde{\mathcal{I}}_\gamma\tau$, $h = (\tau, b)$. Therefore,

$$|P_{n,\gamma,h_n}\psi_{n,\theta} - P_{n,\gamma,h}\psi_{n,\theta}| \le |P_{n,\gamma,h_n}\psi_{n,\theta} - \pi(a)| + |P_{n,\gamma,h}\psi_{n,\theta} - \pi(a)| \to 0,$$

by Corollary 3.6 (cf. equation (16)), implying the required asymptotic equicontinuity on K. $\square$

Proof of Theorem 3.4. Denote by $\overline{\mathscr{G}}_\gamma$ the Gaussian shift on $\overline{\mathbb{H}_\gamma}$. We will first consider minimising the regret

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

where $\tilde{\mathcal{C}}$ is the class of level–α tests of $[h] \in \ker \pi_1$ against $[h] \notin \ker \pi_1$ in $\mathscr{G}_\gamma$.

Consider the Neyman – Pearson test, $\psi^\star$, of $[g] \in \ker \pi_1$ against $[g] + [h]$, where $[h] \in [\ker \pi_1]^\perp$ in $\overline{\mathscr{G}}_\gamma$. This test rejects when

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

for a k chosen such that the test is of level α and where Z is the central process of $\overline{\mathscr{G}}_\gamma$ (and thus a standard Gaussian process under $G_{[0]}$). The choice of k does not depend on the $[g] \in \ker \pi_1$ and the power of this test depends only on $\|\Pi^\perp[h]\|_\gamma^2 = \tau'\tilde{\mathcal{I}}_\gamma\tau$ where $\pi_1[h] = [\tau]$.

Now let $[h] \in \mathbb{H}_\gamma \setminus \ker \pi_1$ 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.[63] Thus $\psi^\star \in \tilde{\mathcal{C}}$ and

$$\tilde{\pi}^\star([h]) \coloneqq \sup_{\phi \in \tilde{\mathcal{C}}} G_{[h]}\phi = G_{[h]}\psi^\star. \tag{51}$$

For $i = 1, \ldots, d_\theta$, let $u_i \coloneqq \Pi^\perp[(e_i, 0)]$ and let $X \coloneqq (Zu_1, \ldots, Zu_{d_\theta})'$. Let ψ be the test which rejects when

$$\left(X'\tilde{\mathcal{I}}_\gamma^\dagger X\right)^2 > c_r,$$

for c_r the $1-\alpha$ quantile of a χ_r^2 random variable. By Theorem 69.10 in Strasser (1985)

$$X \sim \mathcal{N}(\tilde{\mathcal{I}}_\gamma\tau, \tilde{\mathcal{I}}_\gamma) \qquad \text{under } G_{[h]}, \quad \text{where } [\tau] = \pi_1[h].$$

Therefore, by e.g. Theorem 9.2.3 in Rao and Mitra (1971),

$$G_{[h]}\psi = 1 - \mathrm{P}(\chi_r^2(a) \le c_r), \quad a = \tau'\tilde{\mathcal{I}}_\gamma\tau = \|\Pi^\perp[h]\|_\gamma^2.$$

As both $G_{[h]}\psi^\star$ and $G_{[h]}\psi$ depend only on $\|\Pi^\perp[h]\|_\gamma^2$ the same is true of $G_{[h]}\psi^\star - G_{[h]}\psi$. Fix a $\varepsilon > 0$ and suppose that for some test $\phi \in \tilde{\mathcal{C}}$, $\tilde{R}(\phi) < \tilde{R}(\psi) - 2\varepsilon$. There is an $a > 0$ such that

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

[63] 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).

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

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

hence

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

which contradicts Theorem 30.2 in Strasser (1985).

To complete the proof, it suffices to show that a test $\varphi : \Omega \to [0,1]$ is in $\mathcal{C}$ if and only if it is in $\tilde{\mathcal{C}}$ and $R(\varphi) = \tilde{R}(\varphi)$. The first part follows from the observation $h \in H_{\gamma,0}$ if and only if $[h] \in \ker \pi_1$ and Proposition 3.3 which implies that for any $h \in H_\gamma$, $P_{\gamma,h}\varphi = G_{[h]}\varphi$. For the second part, (51), Proposition 3.3 and the first part together imply that $\tilde{\pi}^\star([h]) = \pi^\star(h)$ for all $h \in H_\gamma$. Therefore,

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

Since $h \in H_{\gamma,1}$ if and only if $[h] \in \mathbb{H}_\gamma \setminus \ker \pi_1$, one therefore has

$$\tilde{R}(\varphi) = \sup\{\tilde{\pi}^\star([h]) - G_{[h]}\varphi : [h] \in \mathbb{H}_\gamma \setminus \ker \pi_1\} = \sup\{\pi^\star(h) - P_{\gamma,h}\varphi : h \in H_{\gamma,1}\} = R(\varphi).$$

□

Proof of Proposition 3.4. Let π be a cluster point of π_n. Then $\pi = \lim_{s\in S} \pi_{n(s)}$ for some subnet $(\pi_{n(s)})_{s\in S}$ of $(\pi_n)_{n\in\mathbb{N}}$. That is, $\lim_{s\in S} \pi_{n(s)}(h) = \pi(h)$ for each $h \in H_\gamma$. Hence by Proposition 3.2 and Theorem 7.1 in van der Vaart (1991) there is a test ϕ in $\mathscr{E}_\gamma$ such that $P_{\gamma,h}\phi = \pi(h)$. It follows from the hypothesis that ϕ is of level α. The result follows from Theorem 3.4. □

Proof of Proposition 3.5. Let $h \in H_\gamma$ be the element along which the conclusion is to be shown. By equation (44), if $r = 0$ then $\|[h] - \Pi[h]\|_\gamma = 0$, i.e. $[h] = \Pi[h] \in \ker \pi_1$. Therefore there is a $h^* := (0,b) \in H_{\gamma,0}$ such that $\|h - h^*\|_\gamma = 0$. By Corollary S2.1 and the hypothesis, it follows that

$$\limsup P_{n,\gamma,h}\phi_n \leq \limsup_{n\to\infty} P_{n,\gamma,h^*}\phi_n + \limsup_{n\to\infty} |P_{n,\gamma,h^*}\phi_n - P_{n,\gamma,h}\phi_n| \leq \alpha. \qquad \square$$

Proof of Theorem 3.5. As orthogonal projection operators are idempotent and self-adjoint,

$$\tilde{\mathcal{I}}_n := P_n[\tilde{\ell}_n \tilde{\ell}_n'] = P_n[\dot{\ell}_n \tilde{\ell}_n'].$$

Define $K_n(h,g) := P_{n,\gamma}[\Delta_{n,\gamma}h \Delta_{n,\gamma}g]$ and let G_n be a zero-mean Gaussian process with covariance kernel K_n. We will first show that

$$\tilde{\mathcal{I}}_{n,ij} = \mathbb{E}\left[\mathsf{G}_n(e_i,0)\tilde{\mathsf{G}}_n(e_j,0)\right], \tag{52}$$

where

$$\tilde{\mathsf{G}}_n(e_j,0) := \mathsf{G}_n(e_j,0) - \mathbb{E}\left[\mathsf{G}_n(e_j,0)|\{\mathsf{G}_n(0,b) : b \in B\}\right]. \tag{53}$$

For (52), let Z_n be defined as $Z_n(\Delta_{n,\gamma}h) := \mathsf{G}_n h$. Z_n is evidently linear. It is also bounded as $\|Z_n(\Delta_{n,\gamma}h)\| = \|\mathsf{G}_n h\| = \sqrt{K_n(h,h)} = \|\Delta_{n,\gamma}h\|$. Hence we may extend this map by continuity to a bounded linear map $Z_n : \mathrm{cl}\{\Delta_{n,\gamma}h : h \in H\} \to \mathrm{cl}\{\mathsf{G}_n h : h \in H\}$. Z_n is surjective: let $G = \lim_{m\to\infty} \mathsf{G}_n h_m$. Then, setting $D = \lim_{m\to\infty} \Delta_{n,\gamma}h_m \in \mathrm{cl}\{\Delta_{n,\gamma}h : h \in H\}$, $G = Z_n(D)$ by continuity. Finally, for any $h, g \in H_\gamma$ we have

$$P_n\left[\Delta_{n,\gamma}h\Delta_{n,\gamma}g\right] = K_n(h,g) = \mathbb{E}\left[\mathsf{G}_n h \mathsf{G}_n g\right] = \mathbb{E}\left[Z_n(\Delta_{n,\gamma}h)Z_n(\Delta_{n,\gamma}g)\right],$$

and the same holds for elements in the closure by continuity. Hence Z_n is a Hilbert space isomorphism. Let R be the orthogonal projection onto $\{\mathsf{G}_n(0,b) : b \in B\}^\perp$ and Q the orthogonal projection onto $\{\Delta_{n,\gamma}(0,b) : b \in B\}^\perp$. For $h \in H_\gamma$ one has

$$R\mathsf{G}_n h = RZ_n(\Delta_{n,\gamma}h) = Z_n Q Z_n^{-1} Z_n(\Delta_{n,\gamma}h) = Z_n Q \Delta_{n,\gamma}h.$$

This extends to elements in the closure by continuity. Hence

$$\tilde{\mathcal{I}}_{n,ij} = P_n\left[\Delta_{n,\gamma}(e_i,0)Q\Delta_{n,\gamma}(e_j,0)\right] = \mathbb{E}\left[\mathsf{G}_n(e_i,0)R\mathsf{G}_n(e_j,0)\right]. \tag{54}$$

By Theorem 9.1 in Janson (1997),

$$\mathbb{E}\left[\mathsf{G}_n(e_j,0)|\{\mathsf{G}_n(0,b) : b \in B\}\right] = \Pi\left[\mathsf{G}_n(e_j,0)|\mathrm{cl}\left\{\mathsf{G}_n(0,b) : b \in B\right\}\right], \tag{55}$$

and so $\tilde{G}_n(e_j,0) = R\mathsf{G}_n(e_j,0)$. Combining this with (54) yields (52). Now put $\mathscr{G}_n := \sigma(\{\mathsf{G}_n(0,b) : b \in B\})$, $\mathscr{G}_n := \sigma(\{\mathsf{G}(0,b) : b \in B\})$ and define

$$X_n := \left(\mathsf{G}_n(e_i,0), \mathbb{E}[\mathsf{G}_n(e_j,0)|\mathscr{G}_n]\right), \qquad X := \left(\mathsf{G}(e_i,0), \mathbb{E}[\mathsf{G}(e_j,0)|\mathscr{G}]\right).$$

By (55) and $K_n(h,h) \to K_\gamma(h,h)$ (by Assumption 3.1, cf. Lemma 3.3), $(X_n)_{n\in\mathbb{N}}$ is a sequence of Gaussian random variables with bounded second moment and so uniformly square integrable. Given this and Lemma 3.5, to complete the proof it suffices to show that $X_n \rightsquigarrow X$. In the present setting this follows from Theorem S2.2 in the supplementary material. □

Proof of Lemma 3.8. We have that $P_\gamma^n\left(\Delta_{n,\gamma}h, g'_{n,\gamma}\right) = 0$ and since the observations are i.i.d.,

$$\begin{aligned}
P_\gamma^n\left(\Delta_{n,\gamma}h, g'_{n,\gamma}\right)'\left(\Delta_{n,\gamma}h, g'_{n,\gamma}\right) &= P_\gamma^n\begin{bmatrix}[\Delta_{n,\gamma}h]^2 & [\Delta_{n,\gamma}h]g'_{n,\gamma}\\ g_{n,\gamma}[\Delta_{n,\gamma}h] & g_{n,\gamma}g'_{n,\gamma}\end{bmatrix}\\
&= P_\gamma\begin{bmatrix}[A_\gamma h]^2 & [\Delta_{n,\gamma}h]g'_{\theta,\lambda}\\ g_{\theta,\lambda}[\Delta_{n,\gamma}h] & g_{\theta,\lambda}g'_{\theta,\lambda}\end{bmatrix}\\
&= P_\gamma\begin{bmatrix}[A_\gamma h]^2 & \tau'\dot{\ell}_\gamma g'_{\theta,\lambda}\\ g_{\theta,\lambda}\dot{\ell}'_\gamma\tau & g_{\theta,\lambda}g'_{\theta,\lambda}\end{bmatrix}\\
&=: \Sigma_\gamma(h),
\end{aligned}$$

where the second to last equality is due to $g_\gamma \in \{D_\gamma b : b \in B_\eta\}^\perp$. Since the observations are i.i.d., for each $h \in H_\gamma$, the weak convergence

$$\left(\Delta_{n,\gamma}h, g'_{n,\gamma}\right) \overset{P_\gamma^n}{\rightsquigarrow} \mathcal{N}(0, \Sigma_\gamma(h)),$$

follows by the central limit theorem. □

Proof of Corollary 3.7. By construction $g_\gamma \in \{D_\gamma b : b \in B_\eta\}^\perp$. Apply Lemma 3.8. □

Proof of Corollary 3.8. By properties of orthogonal projections $P_\gamma[\dot{\ell}_\gamma g'_{\theta,\lambda}] = P_\gamma[g_{\theta,\lambda} g'_{\theta,\lambda}]$ (e.g. Theorem 12.14 in Rudin, 1991). Hence, given the expression for $\Sigma_\gamma(h)$ in the proof of Lemma 3.8, we need show only that $P_\gamma[g_{\theta,\lambda} g'_{\theta,\lambda}] = \tilde{\mathcal{I}}_\gamma$. For any $h_i = (\tau_i, b_i) \in H_\gamma$, $i = 1, 2$, by the i.i.d. assumption

$$P_{n,\gamma}[\Delta_{n,\gamma} h_1 \Delta_{n,\gamma} h_2] = P_\gamma \left[A_\gamma h_1 A_\gamma h_2\right],$$

which is constant in n. In conjunction with Lemma 3.3, this implies that for each $n \in \mathbb{N}$,

$$P_\gamma \left[A_\gamma h_1 A_\gamma h_2\right] = \mathrm{P}[\Delta_\gamma h_1 \Delta_\gamma h_2].$$

Let $X = \operatorname{cl} \operatorname{ran} A_\gamma \subset L_2(P_\gamma)$ and $Y = \operatorname{cl} \operatorname{ran} \Delta_\gamma \subset L_2(\mathrm{P})$; these are Hilbert spaces when equipped with the inner products induced by the left and right hand side of the preceding display respectively.[64] Define the map $U : \operatorname{ran} A_\gamma \to \operatorname{ran} \Delta_\gamma$ by $U A_\gamma h := \Delta_\gamma h$ for $h \in H_\gamma$. This is evidently a bounded, linear, surjective isometry. It can therefore be uniquely extended to a bounded, linear, surjective isometry $U : X \to Y$, i.e. U is a Hilbert space isomorphism between X and Y. Hence, if R is the orthogonal projection onto $\mathcal{T}^\perp$ (defined in Lemma 3.5) and Q that onto $\{D_\gamma b : b \in B_\eta\}^\perp = \{A_\gamma h : h = (0, b) \in H_\gamma\}^\perp$, one has

$$R \Delta_\gamma h = R U A_\gamma h = U Q U^* U A_\gamma h = U Q A_\gamma h,$$

which implies the required conclusion since the i-th element of $\dot{\ell}_\gamma$ is $A_\gamma(e_i, 0)$.[65] □

B Tables & figures

B.1 Single index model

Table 4: Index functions used in the simulation exeriments

name	expression	c_1	c_2	c_3
Exponential	$f_j(v) = 5\exp(-v^2/2c_j^2)$	1.25	2	4
Logistic	$f_j(v) = 25(1 + \exp(-v/c_j))^{-1}$	0.75	3	12

[64] Where, as usual, we identify a.s. equal functions.

[65] Cf. the last step in the proof of Lemma 3.5.

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

	$X=(Z_1,Z_2)$			$X=(Z_1, 0.2Z_1+0.4Z_2+0.8)$		
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

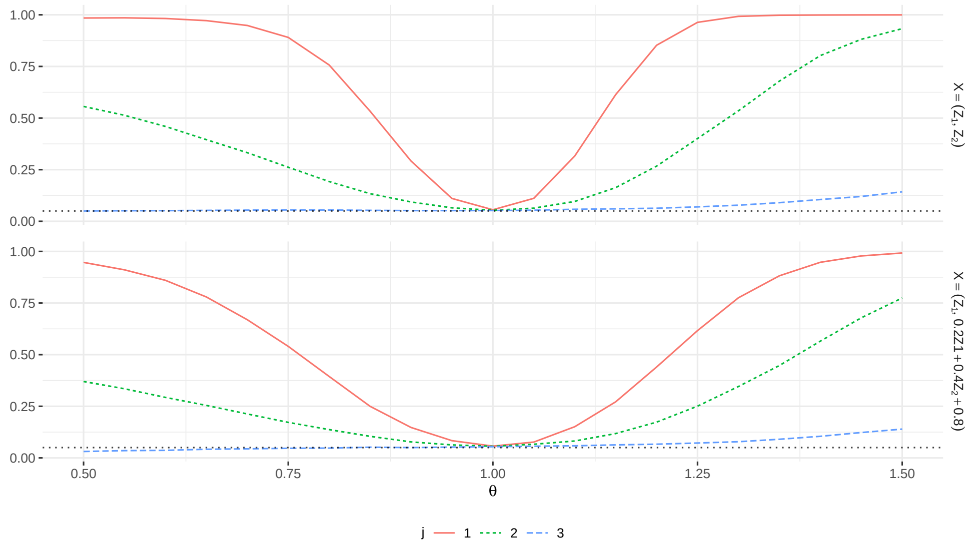
Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-1,1)$ are independently drawn.

Table 6: ERF (%) $\epsilon \sim \mathcal{N}(0,1),\ f_j(v) = 25\,(1+\exp(-v/c_j))^{-1}$

	$X=(Z_1,Z_2)$			$X=(Z_1, 0.2Z_1+0.4Z_2+0.8)$		
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

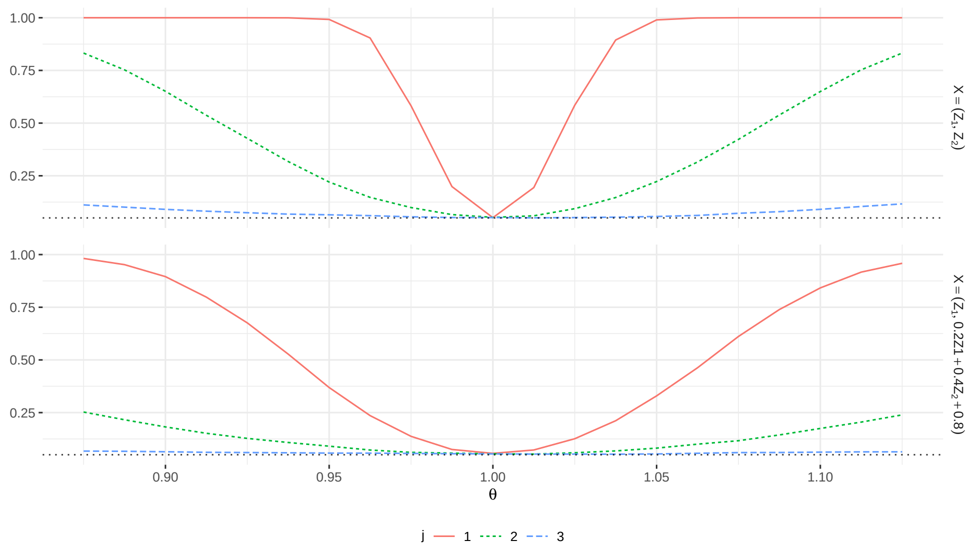
Notes: Based on 5000 Monte carlo replications. The $Z_k \sim U(-1,1)$ are independently drawn.

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



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

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



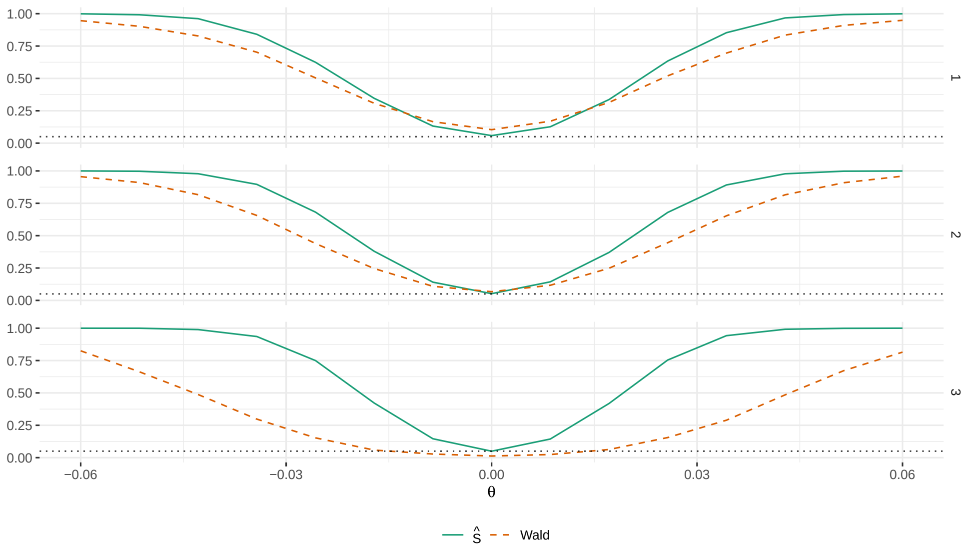
Based on 5000 Monte carlo replications. The $Z_k \sim U(-1,1)$ are independently drawn.

Table 7: ERF (%) $\epsilon \sim \mathcal{N}(0,1)$, 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	6.26	5.90	5.82	6.62	6.60	6.52
600	5.82	4.98	5.30	6.46	5.34	5.06
800	5.82	5.20	4.98	5.98	5.60	5.36
Wald						
400	13.18	8.54	3.88	13.78	8.28	4.18
600	10.42	6.66	2.32	12.22	7.22	2.50
800	10.40	6.72	1.26	11.94	6.78	1.26

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

Figure 3: ERF (%) $\epsilon \sim \mathcal{N}(0,1)$, index function as in (S35), $X = (Z_1, Z_2)$



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

B.2 IV model

Table 8: Index functions used in the simulation exeriments

name	expression	c_1	c_2	c_3
Exponential	$\pi_j(z) = 5\exp(-z^2/2c_j^2)$	1.250	2.000	4.00
Logistic	$\pi_j(z) = 25(1+\exp(-z/c_j))^{-1}$	0.296	2.667	24.00
Linear	$\pi_j(z) = c_j z$	1.000	0.300	0.09

The c_j for the logistic functions correspond to $c_j = 72/3^{(7-2j)}$.

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

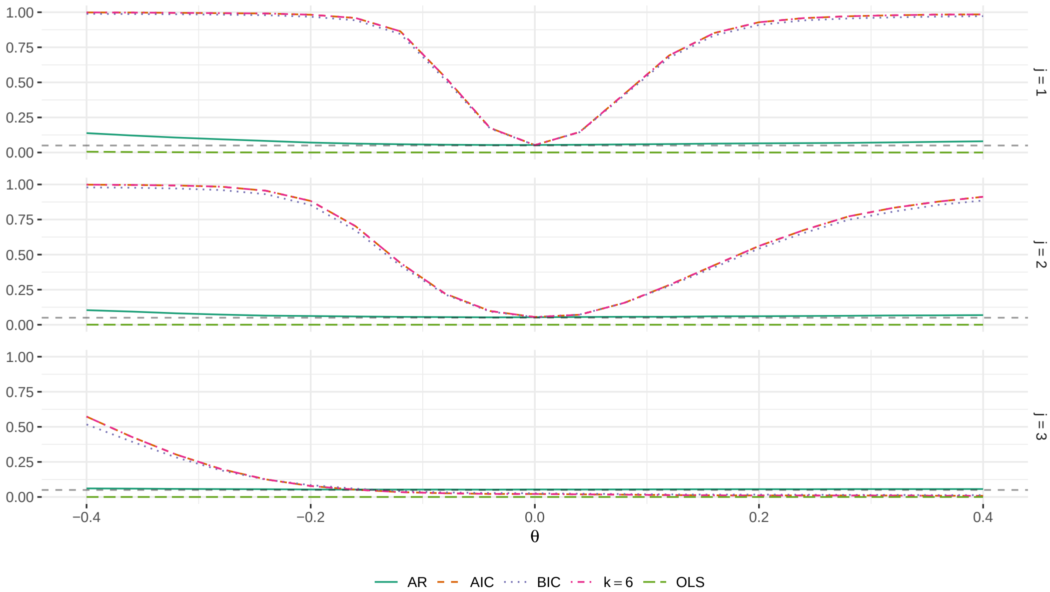


Table 9: Empirical rejection frequencies, IV, Design 1

		$\hat{S}$				AR	TSLS W	GMM W	GMM LM
n	j	OLS	$k=6$	AIC	BIC				
Exponential									
200	1	0.74	5.78	5.78	5.70	4.98	4.94	8.04	5.56
200	2	0.24	6.16	6.16	6.20	4.98	12.20	13.58	6.10
200	3	0.04	6.14	6.14	6.38	4.98	30.36	58.44	19.48
400	1	0.12	5.18	5.18	5.08	5.30	4.86	6.24	5.32
400	2	0.02	5.74	5.74	5.60	5.30	11.36	8.66	5.30
400	3	0.00	3.18	3.18	3.72	5.30	30.72	35.98	12.92
600	1	0.04	5.30	5.30	5.30	5.36	5.40	6.00	5.32
600	2	0.00	5.50	5.50	5.40	5.36	12.18	7.54	5.70
600	3	0.00	2.14	2.14	2.64	5.36	30.92	28.02	10.94
Logistic									
200	1	4.74	5.00	5.00	4.52	4.98	4.94	4.92	5.20
200	2	4.74	4.74	4.74	4.78	4.98	4.98	5.64	5.36
200	3	2.24	6.80	6.80	6.92	4.98	7.76	40.42	15.24
400	1	5.18	4.86	4.60	4.54	5.30	5.36	5.10	4.98
400	2	5.18	5.04	5.04	5.00	5.30	5.10	5.50	5.24
400	3	2.28	4.26	4.26	4.62	5.30	5.92	23.88	10.52
600	1	5.34	5.46	5.30	4.76	5.36	5.28	5.14	5.60
600	2	5.34	5.46	5.46	5.44	5.36	5.54	6.00	5.48
600	3	2.68	3.68	3.68	3.92	5.36	6.30	18.82	9.42
Linear									
200	1	4.74	5.32	5.32	5.50	4.98	4.64	8.56	5.60
200	2	2.28	7.04	7.04	7.06	4.98	6.88	33.46	12.68
200	3	0.14	4.66	4.66	5.28	4.98	15.12	89.54	46.52
400	1	5.18	5.32	5.32	5.32	5.30	5.16	6.52	5.44
400	2	2.36	5.08	5.08	5.18	5.30	5.42	19.74	8.88
400	3	0.02	1.12	1.12	2.08	5.30	10.92	78.72	37.06
600	1	5.34	5.58	5.58	5.62	5.36	5.42	7.00	6.18
600	2	2.86	4.30	4.30	4.62	5.36	6.08	15.52	8.58
600	3	0.00	0.28	0.28	0.66	5.36	10.68	69.00	31.44

Notes: Based on 5000 Monte carlo replications. All $\hat{S}$ tests except "OLS" use Legendre polynomials to estimate π.

Figure 5: $\pi_j(z) = 25(1 + \exp(-z/c_j))^{-1}$

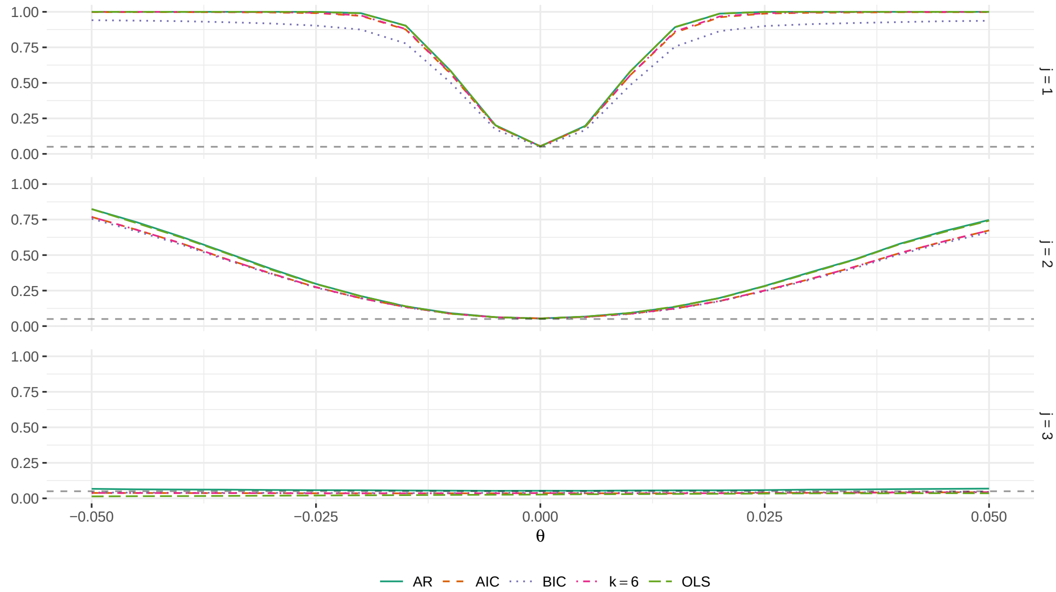


Figure 6: $\pi_j(z) = c_j z$

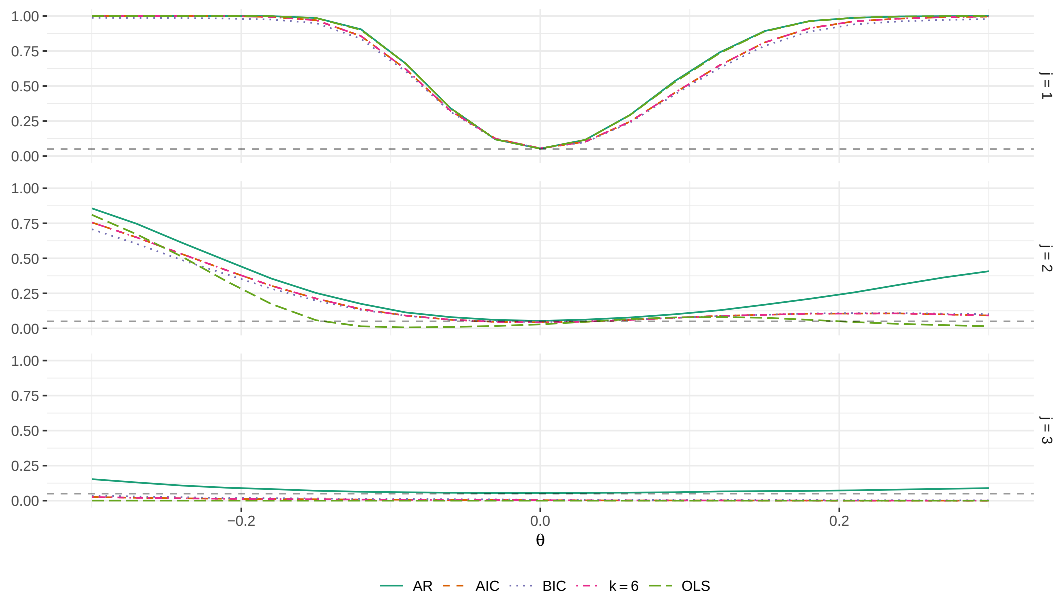


Table 10: Empirical rejection frequencies, IV, Design 2, $\pi_{1,j} = \pi_{2,j}$

		$\hat{S}$				AR	TSLS W	GMM W	GMM LM
n	j	OLS	$k=3$	AIC	BIC				
Exponential									
200	1	3.62	4.88	4.80	4.90	5.28	3.94	21.40	9.32
200	2	2.28	5.44	5.40	5.14	5.28	13.16	49.92	17.62
200	3	0.92	5.50	5.46	5.06	5.28	36.94	99.40	76.08
400	1	0.64	4.82	4.82	4.64	5.20	3.84	13.22	7.16
400	2	0.14	4.96	4.96	4.82	5.20	12.12	27.84	12.14
400	3	0.02	6.48	6.48	5.68	5.20	35.00	95.40	59.42
600	1	0.22	4.64	4.64	4.80	5.34	4.04	9.48	6.24
600	2	0.00	5.04	5.04	5.08	5.34	13.04	19.12	9.08
600	3	0.00	6.34	6.34	6.24	5.34	36.68	88.00	48.02
Logistic									
200	1	5.06	4.56	1.86	2.42	5.28	5.40	6.22	5.46
200	2	5.06	4.52	4.56	4.56	5.28	5.50	9.06	5.92
200	3	4.56	5.16	5.14	4.84	5.28	8.72	95.36	60.26
400	1	5.00	5.14	3.34	3.12	5.20	5.30	5.98	5.58
400	2	5.00	4.78	4.78	4.80	5.20	5.40	6.58	5.64
400	3	4.52	6.16	6.16	5.80	5.20	7.42	79.14	42.14
600	1	5.04	4.40	3.64	3.32	5.34	5.30	5.24	5.16
600	2	5.04	4.82	4.82	4.54	5.34	5.10	6.40	5.20
600	3	4.12	5.94	5.94	5.62	5.34	6.94	63.94	32.82
Linear									
200	1	5.06	4.92	4.90	4.68	5.28	5.28	20.86	9.42
200	2	4.14	5.02	4.98	4.80	5.28	8.28	90.42	49.82
200	3	2.38	5.74	5.78	5.34	5.28	15.28	99.98	95.46
400	1	5.00	5.04	5.04	4.68	5.20	5.30	12.34	7.38
400	2	4.30	5.78	5.78	5.60	5.20	6.94	68.50	34.12
400	3	0.34	6.92	6.92	6.20	5.20	12.16	99.96	91.84
600	1	5.04	4.98	4.98	4.72	5.34	5.22	10.08	6.88
600	2	3.86	5.58	5.58	5.30	5.34	6.52	53.24	26.44
600	3	0.00	6.04	6.04	5.58	5.34	11.34	99.78	88.42

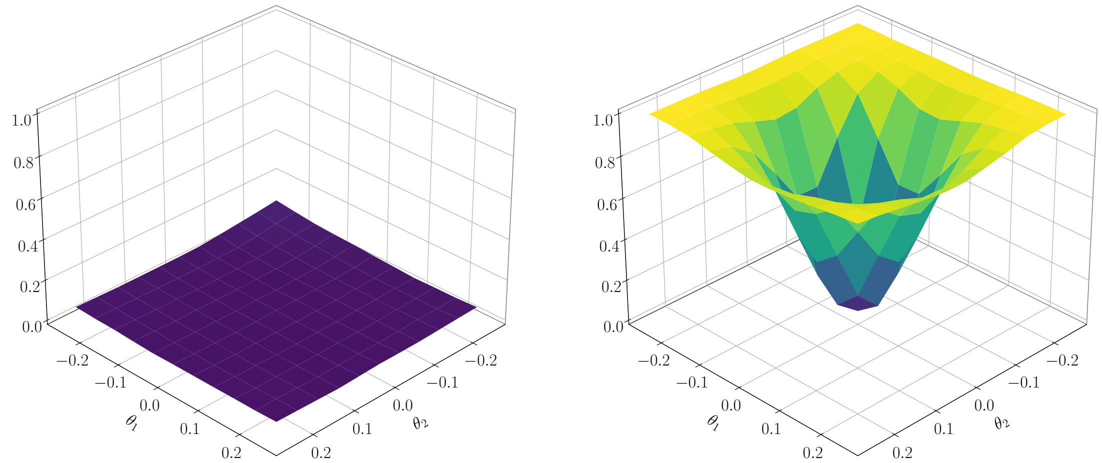
Notes: Based on 2500 Monte carlo replications. $k = 3$ indicates that each univariate series forming the tensor series has $k = 3$.

Table 11: Empirical rejection frequencies, IV, Design 2, $\pi_{2,j} = c_j Z_{2,2}$

		$\hat{S}$				AR	TSLS W	GMM W	GMM LM
n	j	OLS	$k=3$	AIC	BIC				
Exponential									
200	1	4.88	4.74	4.68	4.66	5.28	5.10	24.60	11.38
200	2	4.14	5.54	5.44	4.90	5.28	12.00	85.52	44.02
200	3	1.40	5.24	5.22	4.94	5.28	29.20	99.94	92.38
400	1	5.26	4.98	4.98	4.96	5.20	4.76	15.06	8.42
400	2	3.70	5.62	5.62	5.36	5.20	11.30	61.32	28.94
400	3	0.06	6.56	6.56	5.88	5.20	26.64	99.68	86.80
600	1	4.84	4.64	4.64	4.22	5.34	5.02	11.64	7.10
600	2	3.48	5.64	5.64	5.40	5.34	11.72	47.58	23.32
600	3	0.00	6.40	6.40	5.80	5.34	27.80	99.22	82.16
Logistic									
200	1	5.06	4.80	3.38	3.64	5.28	5.26	17.16	8.88
200	2	5.62	5.54	5.58	5.28	5.28	7.88	78.96	39.44
200	3	4.52	5.30	5.30	5.04	5.28	14.74	99.80	86.88
400	1	5.00	5.22	4.40	4.12	5.20	5.14	10.62	7.22
400	2	5.80	5.90	5.90	5.70	5.20	6.70	54.62	25.14
400	3	4.00	6.28	6.28	5.94	5.20	11.84	99.04	80.50
600	1	5.04	4.68	4.22	3.86	5.34	5.24	9.10	6.42
600	2	5.92	5.26	5.26	4.96	5.34	6.24	42.44	21.60
600	3	3.72	6.32	6.32	5.88	5.34	11.34	97.44	74.64
Linear									
200	1	5.06	4.92	4.90	4.68	5.28	5.28	20.86	9.42
200	2	4.14	5.02	4.98	4.80	5.28	8.28	90.42	49.82
200	3	2.38	5.74	5.78	5.34	5.28	15.28	99.98	95.46
400	1	5.00	5.04	5.04	4.68	5.20	5.30	12.34	7.38
400	2	4.30	5.78	5.78	5.60	5.20	6.94	68.50	34.12
400	3	0.34	6.92	6.92	6.20	5.20	12.16	99.96	91.84
600	1	5.04	4.98	4.98	4.72	5.34	5.22	10.08	6.88
600	2	3.86	5.58	5.58	5.30	5.34	6.52	53.24	26.44
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 $k = 3$.

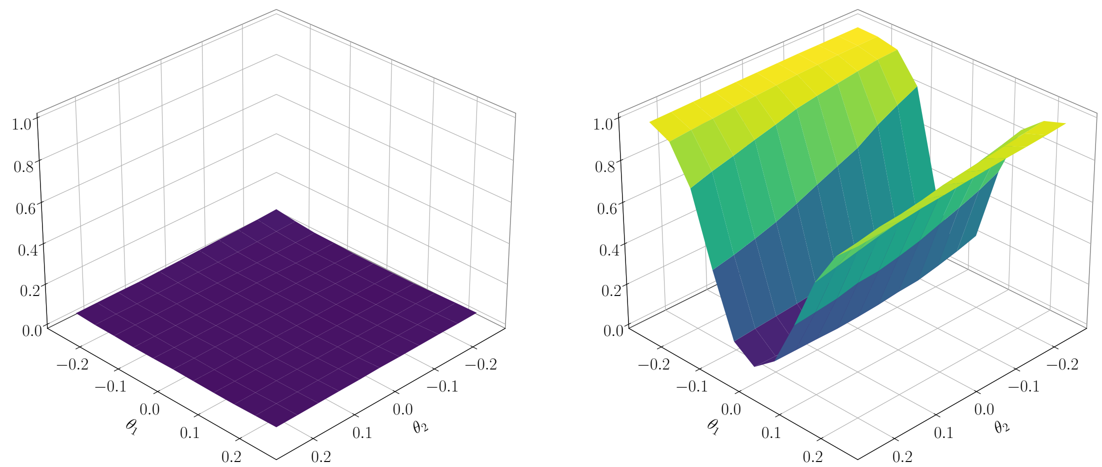
Figure 7: π_i exponential with $j = 1$ $(i = 1, 2)$



(a) AR

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

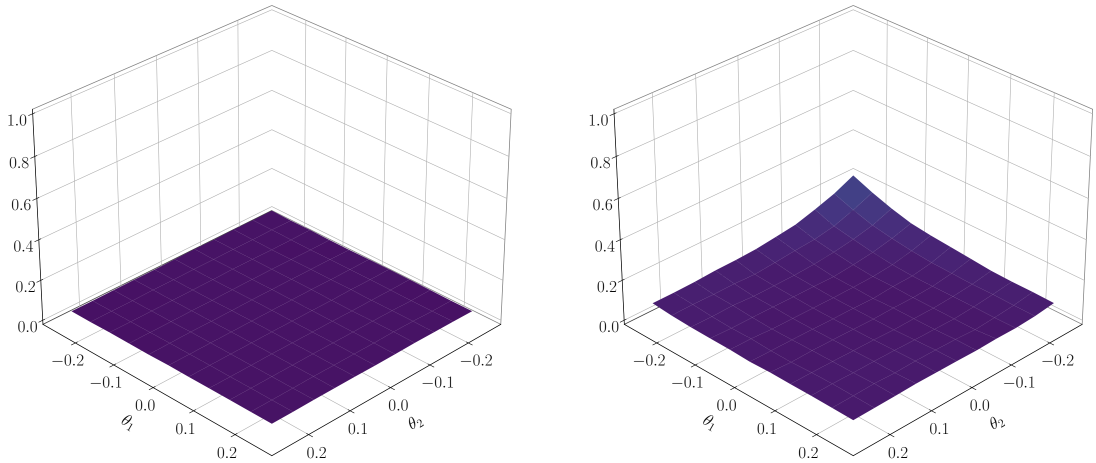
Figure 8: π_1 exponential with $j = 1$, π_2 exponential with $j = 3$



(a) AR

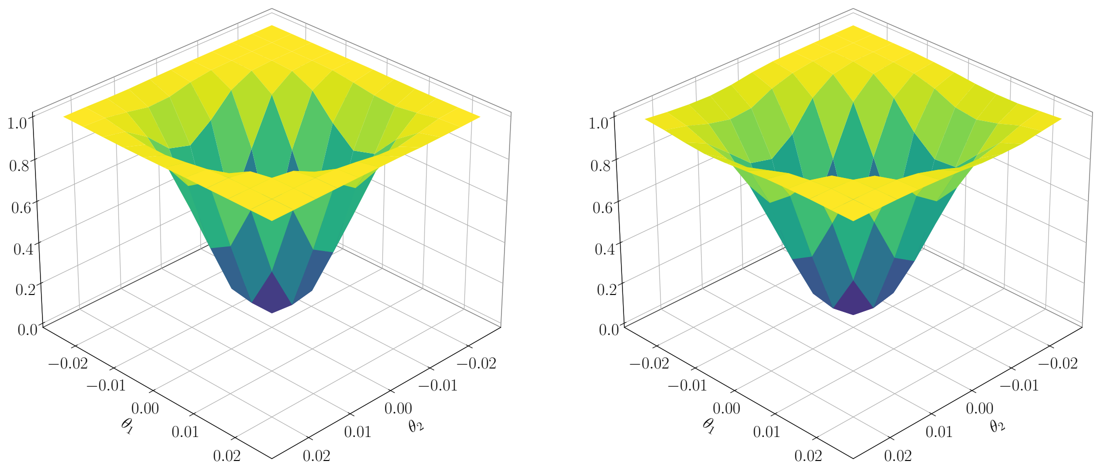
(b) $\hat{S}$ $(k = 3)$

Figure 9: π_1 exponential with $j = 3$, π_2 exponential with $j = 3$



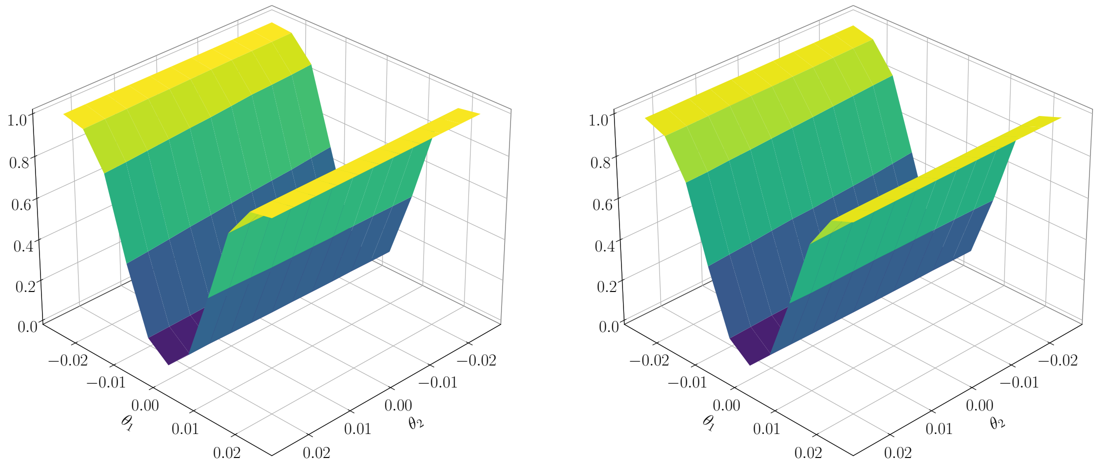
(a) AR (b) $\hat{S}$ $(k = 3)$

Figure 10: π_i logistic with $j = 1$ $(i = 1, 2)$



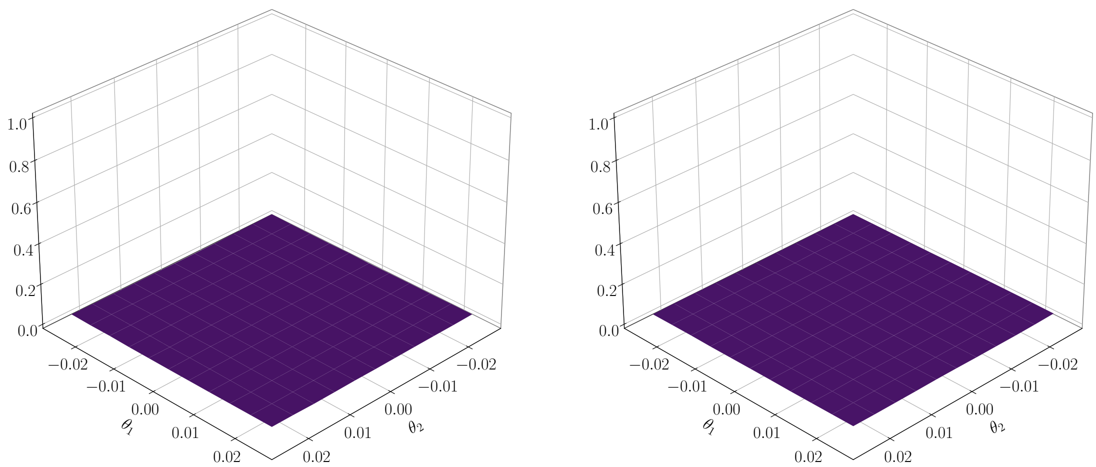
(a) AR (b) $\hat{S}$ $(k = 3)$

Figure 11: π_1 logistic with $j = 1$, π_2 logistic with $j = 3$



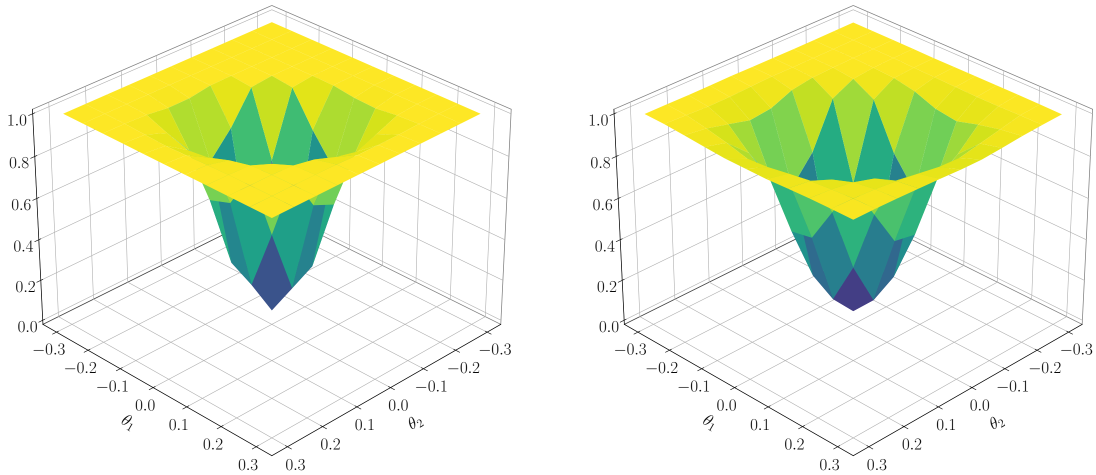
(a) AR (b) $\hat{S}$ $(k = 3)$

Figure 12: π_1 logistic with $j = 3$, π_2 logistic with $j = 3$



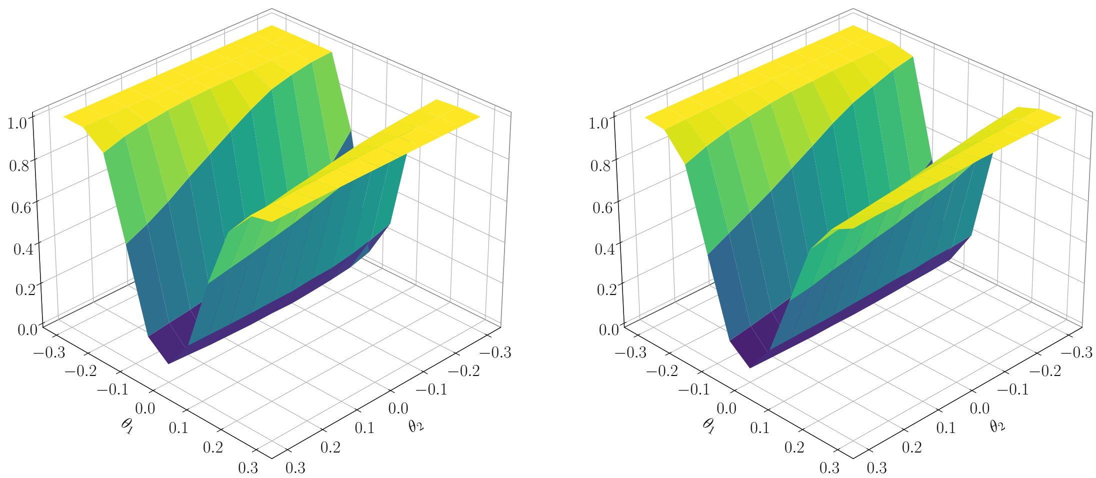
(a) AR (b) $\hat{S}$ $(k = 3)$

Figure 13: π_i linear with $j = 1$ $(i = 1, 2)$



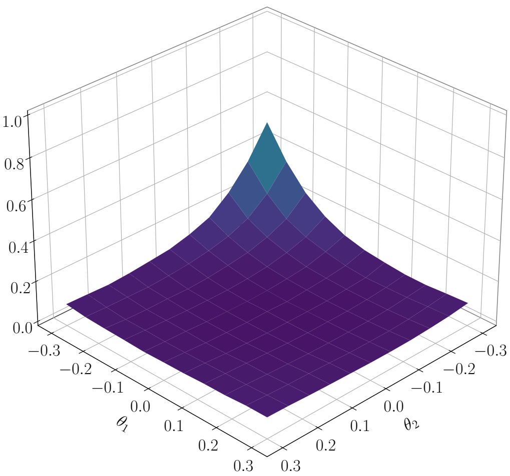
(a) AR (b) $\hat{S}$ $(k = 3)$

Figure 14: π_1 linear with $j = 1$, π_2 linear with $j = 3$

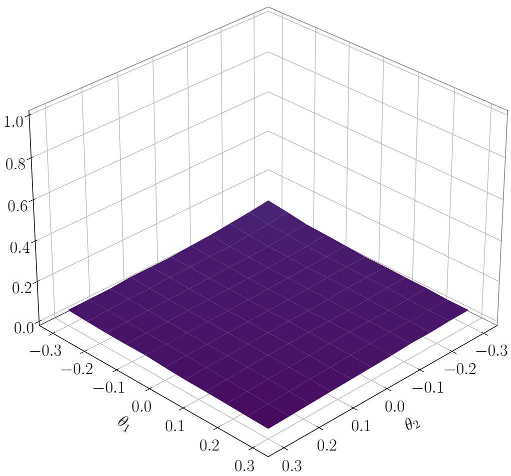


(a) AR (b) $\hat{S}$ $(k = 3)$

Figure 15: π_1 linear with $j = 3$, π_2 linear with $j = 3$



(a) AR



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