# LOCALLY REGULAR AND EFFICIENT TESTS IN NON-REGULAR SEMIPARAMETRIC MODELS

Adam  $Lee^*$ 

December 20, 2024

#### Abstract

This paper considers hypothesis testing in semiparametric models which may be nonregular. I show that  $C(\alpha)$  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 and generalise classical power bounds to this case. I give conditions under which these power bounds are attained by the proposed  $C(\alpha)$  style tests. The application of the theory to a single index model and an instrumental variables model is worked out in detail.

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

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

<sup>\*</sup>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 exhibit the same limiting behaviour under the true parameter as they do under sequences of "local alternatives" which cannot be consistently distinguished from the true parameter.<sup>1</sup> Unfortunately, there are many models in which locally regular estimators do not exist.<sup>2</sup> One necessary condition is given by Chamberlain (1986): if the efficient information for a scalar parameter is 0, then no locally regular estimator of that parameter exists. Similarly, singularity of the efficient information matrix implies the non-existence of locally regular estimators of Euclidean parameters. Models in which this may occur are called "non-regular". Many widely used models are non-regular (at least at certain parameter values): examples include single index models, instrumental variables, errors-in-variables, mixed proportional hazards, discrete choice models and sample selection models.

In this paper, I demonstrate that locally regular *tests* exist in a broad class of non-regular models, despite the non-existence of locally regular estimators. In particular, I show that a class of tests based on the  $C(\alpha)$  idea of Neyman (1959, 1979) are locally regular.

These tests are based on a quadratic form of moment conditions evaluated under the null hypothesis. The key  $[C(\alpha)]$  idea which ensures the local regularity is that the moment conditions must be (asymptotically) orthogonal to the collection of score functions for all nuisance parameters. Such moment conditions can always be constructed from any initial moment conditions by an orthogonal projection.

A key advantage of these  $C(\alpha)$  tests is that they do not (asymptotically) overreject under (semiparametric) weak identification asymptotics, i.e. under local alternatives to a point of identification failure.<sup>3</sup> The local regularity of these tests ensures that if the test is asymptotically of level  $\alpha$  under any fixed parameter consistent with the null, it is also asymptotically of level  $\alpha$  under any sequence of local alternatives consistent with the null, i.e. under (semiparametric) weak identification asymptotics. In addition to the well-studied case where weak identification stems from potential identification failure due to a finite dimensional nuisance parameter, the results in this paper also cover the case where identification failure is due to an infinite dimensional nuisance parameter and thus provide a gener-

<sup>&</sup>lt;sup>1</sup>Precise definitions will be given below. See Bickel, Klaassen, Ritov, and Wellner (1998); van der Vaart (1998), for example, for textbook treatments.

<sup>&</sup>lt;sup>2</sup>See e.g. Chamberlain (1986, 1992); Newey (1990); Ritov and Bickel (1990) for some examples.
<sup>3</sup>The semiparametric weak identification asymptotics used are those of Kaji (2021) (see also Andrews and Mikusheva, 2022), suitably generalised to permit non-i.i.d models.

ally applicable approach to weak identification robust inference in semiparametric models.<sup>4</sup> Even in the case where the identification failure due to a finite dimensional nuisance parameter, the resulting weak identification robust tests appear to be new in the literature.<sup>5</sup> The tests proposed here are derived directly from an asymptotic orthogonality condition. As such they are close in spirit to the identification robust test of Kleibergen (2005) which also requires an orthogonalisation, albeit with respect to different objects and in a different Hilbert space.

Achieving local regularity does *not* come at the expense of (local asymptotic) power. I characterise power bounds for tests in non-regular models and show that the  $C(\alpha)$  tests proposed in this paper achieve these power bounds provided the moment conditions are chosen optimally. These power bounds contain those for regular models as a special case. Moreover, the conditions required for attainment of the power bounds are weaker than those in the literature.<sup>6</sup>

Following the theoretical development, I give details of its application to two examples: (i) a single index model which may be weakly identified when the link function is too flat and (ii) an instrumental variables (IV) model which may be weakly identified when the (nonparametric) first stage is too close to a constant function. Simulation experiments based on these examples demonstrate that the proposed tests enjoy good finite sample performance.

The application to IV may also be of interest for empirical researchers concerned about weak instruments. If the instruments are mean independent of the errors, then the test proposed here is robust to weak identification and can be substantially more powerful than tests assuming a linear first stage. This imposes no cost if the true first stage is (approximately) linear: the power of the proposed test is comparable to optimal tests based on a linear first stage. The practical use of these tests is demonstrated in two IV applications with possibly weak instruments.

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). The second is the liter-

<sup>&</sup>lt;sup>4</sup>These  $C(\alpha)$  tests also behave well in other non-standard settings, such as when nuisance functions are estimated under shape constraints; see Section S2.1 for a discussion.

<sup>&</sup>lt;sup>5</sup>For instance, in the case of homoskedastic linear IV, the test that results from the construction in this paper does not coincide with any of the "usual" weak instrument robust tests (e.g. AR, LM, K, CLR). Demonstration of this is available from the author.

<sup>&</sup>lt;sup>6</sup>In particular, in regular models 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 in this paper does not require either of these conditions.

ature on  $C(\alpha)$  tests. These 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 same structure which ensures good performance in such settings is used for a different purpose – to construct 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 especially useful in cases of weak identification and therefore this paper is closely related to the literature on weak identification robust inference (e.g. Staiger and Stock, 1997; Dufour, 1997; Stock and Wright, 2000; Kleibergen, 2005; Andrews and Cheng, 2012; Andrews and Mikusheva, 2015, 2016). More specifically, 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 therein to non - i.i.d. models.<sup>7</sup>

# 2 Locally regular testing

#### 2.1 The local setup

The goal considered throughout this paper is to construct hypothesis tests of  $H_0: \theta = \theta_0$  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  $\sigma$ -finite measure  $\nu_n$ .<sup>8</sup>

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  $P_{n,\gamma} = P_{n,\gamma,0}$ . Elements of  $H_{\gamma}$  will be written as  $h = (\tau, b) \in \mathbb{R}^{d_{\theta}} \times B_{\gamma}$ .<sup>9</sup> The measures  $P_{n,\gamma,h}$  should be viewed

<sup>&</sup>lt;sup>7</sup>Failure of local identification and singularity of the information matrix are closely linked in parametric models, see Rothenberg (1971). In the semiparametric case, parameters may be identified but nevertheless have a singular efficient information matrix. The relationship between the efficient information matrix and identification is considered by Escanciano (2022).

<sup>&</sup>lt;sup>8</sup>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 is the situation considered in Section 4 as well as in the examples in Section 5.

<sup>&</sup>lt;sup>9</sup>In most examples,  $H_{\gamma}$  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

as local perturbations of the measure  $P_{n,\gamma}$  in a "direction"  $h \in H_{\gamma}$ . These local perturbations can be split in two groups: the perturbations  $P_{n,\gamma,h}$  with  $h \in H_{\gamma,0} :=$  $\{(0,b) : b \in B_{\gamma}\}$  correspond to the null hypothesis  $H_0 : \theta = \theta_0$  and those with  $h \in H_{\gamma,1} := \{h = (\tau, b) : 0 \neq \tau \in \mathbb{R}^{d_{\theta}}, b \in B_{\gamma}\}$  to the alternative  $H_1 : \theta \neq \theta_0$ . 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*. The subsequent analysis is local with the parameter  $\gamma$  being considered fixed at a  $\gamma$  consistent with  $H_0$ . As such, to lighten the notation, dependence on  $\gamma$  will be mostly left implicit: I write  $P_{n,h}$  for  $P_{n,\gamma,h}$ , H for  $H_{\gamma}$ ,  $H_i$  for  $H_{\gamma,i}$  (i = 0, 1) and similarly for other objects. I also use the abbreviation  $P_n := P_{n,0}$ .

I use the single-index model as a running example throughout the paper.<sup>10</sup>

EXAMPLE 1 (Single-index model): Suppose that the researcher observes n i.i.d. copies of  $W = (Y, X_1, X_2) \in \mathbb{R}^{2+K}$  where

$$Y = f(X_1 + X'_2\theta) + \epsilon, \qquad \mathbb{E}[\epsilon|X] = 0, \tag{1}$$

and where f belongs to some set of continuously differentiable functions  $\mathscr{F}$ . The description of the model is completed by  $\zeta \in \mathscr{Z}$ , the density function of  $(\epsilon, X)$  with respect to some  $\sigma$ -finite measure. The model  $\mathcal{P}_n$  consists of the product measures  $P_n = P^n$  where P is the probability measure corresponding to the density

$$p(W) = p_{\gamma}(W) \coloneqq \zeta(\epsilon_{f,\theta}, X), \qquad \epsilon_{f,\theta} \coloneqq Y - f(V_{\theta}), \quad V_{\theta} \coloneqq X_1 + X_2'\theta, \qquad (2)$$

for a  $\gamma = (\theta, f, \zeta) \in \Theta \times \mathscr{F} \times \mathscr{Z} = \Gamma$ . A class of local perturbations to this model are the probability measures  $P_{n,h} = P_h^n$  where  $P_h$  has density  $p_{\gamma+\varphi_n(h)}$  with

$$\varphi_n(h) = (\tau, b_1, b_2\zeta) / \sqrt{n}, \qquad h = (\tau, (b_1, b_2)) \in H \coloneqq \mathbb{R}^{d_\theta} \times (B_1 \times B_2), \qquad (3)$$

where  $B_1$  is a subset of the bounded, continuously differentiable functions with bounded derivative and  $B_2$  is a subset of the bounded functions  $b_2 : \mathbb{R}^{1+K} \to \mathbb{R}$ , continuously differentiable in the first argument with bounded derivative.

boundary constraints. In such a setting, if the constraint is binding at  $\gamma$ , then  $\gamma$  can only be perturbed in certain directions if  $P_{n,\gamma,h}$  is to remain within the model.

 $<sup>^{10}\</sup>mathrm{Technical}$  details for this example are deferred to Sections 5.1 and S4.1.

#### 2.2 Local asymptotic normality

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

$$L_n(h) \coloneqq \log \frac{p_{n,h}}{p_{n,0}}, \qquad \text{where} \quad p_{n,h} \coloneqq \frac{\mathrm{d}P_{n,h}}{\mathrm{d}\nu_n}, \text{ for } h \in H.$$
(4)

ASSUMPTION 1 (LAN): For bounded linear maps  $\Delta_n : \overline{\lim} H \to L^0_2(P_n)$ ,

$$L_{n}(h) = \Delta_{n}h - \frac{1}{2} \|\Delta_{n}h\|^{2} + R_{n}(h), \qquad h \in H$$
(5)

with  $R_n(h) \xrightarrow{P_n} 0$  for all  $h \in H$ . Additionally, for each  $h \in H$ , the law of  $\Delta_n h$  converges to  $\mathcal{N}(0, \sigma(h))$  in the Mallows-2 metric,  $d_2$ .

The requirement that  $\Delta_n h$  converges in  $d_2$  is equivalent to requiring that it converges weakly and  $(\Delta_n h)_{n \in \mathbb{N}}$  is uniformly square  $P_n$ -integrable (e.g. Bickel et al., 1998, Appendix A.6). This implies that  $\sigma(h) = \lim_{n \to \infty} ||\Delta_n h||^2$ .

REMARK 1: Assumption 1 ensures that the sequences  $(P_n)_{n\in\mathbb{N}}$  and  $(P_{n,h})_{n\in\mathbb{N}}$  are mutually contiguous for any  $h \in H$  (see e.g. van der Vaart, 1998, Example 6.5).

REMARK 2: If H is (pseudo-)metrised one may consider a uniform version of Assumption 1, i.e. uniform local asymptotic normality (ULAN). Such a version is given in Assumption S1 and is equivalent to Assumption 1 plus asymptotic equicontinuity on compact sets of  $h \mapsto \Delta_n h$  (in  $L_2(P_{n,0})$ ) and  $h \mapsto P_{n,h}$  (in total variation) (Proposition S1). The latter equicontinuity condition is of interest regarding local uniformity of size control; cf. Corollary 1 and Lemma 1 below.

EXAMPLE 1 (continued): Under regularity conditions, the single-index model satisfies Assumption 1 with

$$\Delta_n h \coloneqq \frac{1}{\sqrt{n}} \sum_{i=1}^n \tau' \dot{\ell}(W_i) + [Db](W_i), \tag{6}$$

where for  $\phi(e, x) \coloneqq \frac{\partial \log \zeta(e, x)}{\partial e}$ ,

$$\dot{\ell}(W) \coloneqq -\phi(\epsilon_{f,\theta}, X) f'(V_{\theta}) X_2, \quad [Db](W) \coloneqq -\phi(\epsilon_{f,\theta}, X) b_1(V_{\theta}) + b_2(\epsilon_{f,\theta}, X).$$
(7)

#### 2.3 Local regularity for tests

DEFINITION 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 asymptotically of level  $\alpha$  and locally regular if

$$\pi_n(\tau, b) \coloneqq P_{n,h}\phi_n \to \pi(\tau), \quad h = (\tau, b) \in H \quad and \quad \pi(0) \le \alpha.$$
(8)

That is, the finite sample (local) power function of the test,  $\pi_n$  converges under each  $P_{n,h}$  to a function  $\pi$  which may depend on  $\tau$  (and, implicitly,  $\gamma$ ) but not on b, the parameter which describes local deviations from the nuisance parameter  $\eta$ .<sup>11</sup> If a sequence of tests does *not* satisfy (8) it is *(locally) non-regular*.

Local regularity of test sequences as in (8) is a pointwise concept. It is also of interest to consider a uniform version.

DEFINITION 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 asymptotically of level  $\alpha$  and locally uniformly regular on  $K \subset H$  if (8) holds uniformly on K.

If H is a (pseudo-)metric space and K is a compact set, for the convergence in (8) to hold uniformly on K it is necessary and sufficient to show that the sequence of functions  $\pi_n$  is asymptotically equicontinuous on K.<sup>12</sup>

Directly working with the power functions  $\pi_n$  to show their asymptotic equicontinuity is complicated in many cases. It is, however, often possible to show results which imply this property. For instance, the functions  $h \mapsto P_{n,h}$  being asymptotically equicontinuous in  $d_{TV}$  implies the required asymptotic equicontinuity of the power functions. Despite being (much) stronger, this often holds.<sup>13</sup>

Weak identification asymptotics and local regularity In many models there are parameter values,  $\gamma$ , at which locally regular estimators do not exist. Points where the parameter of interest,  $\theta$ , is un- or under-identified provide an important class of examples. Moreover, as is well known from the literature on weak identification, even if  $\theta$  is identified at  $\gamma$ , finite sample inference may be poor if  $\gamma$  is too close to a point of identification failure relative to the amount of information contained in the sample. Such behaviour has been widely studied in models where the part of  $\gamma$  causing the identification failure is finite dimensional

<sup>&</sup>lt;sup>11</sup>Cf. the definition of a (locally) regular estimator in e.g. van der Vaart, 1998, p. 365.

<sup>&</sup>lt;sup>12</sup>The same is true if K is totally bounded. See e.g. Davidson (2021), p. 123, for the definition of asymptotic equicontinuity.

 $<sup>^{13}\</sup>mathrm{See}$  the discussion following Remark 4 below.

(e.g. Andrews and Cheng, 2012; Andrews and Mikusheva, 2015).

There are also many examples where weak identification may occur due to the value of *infinite-dimensional* nuisance parameters. Kaji (2021) and Andrews and Mikusheva (2022) use a differentiability in quadratic mean (DQM) condition to define semiparametric weak identification asymptotics in i.i.d. models. In particular, they consider sequences  $P_{n,h}^n$  which satisfy

$$\lim_{n \to \infty} \int \left[ \sqrt{n} \left( \sqrt{p_{n,h}} - \sqrt{p_0} \right) - \frac{1}{2} f \sqrt{p_0} \right]^2 \mathrm{d}\nu_n = 0 \tag{9}$$

for a point  $P_0$  where the parameter of interest is unidentified. In the i.i.d. case, (9) implies the LAN expansion in Assumption 1 with  $\Delta_n h = \frac{1}{\sqrt{n}} \sum_{i=1}^n f(W_i)$  (e.g. van der Vaart, 1998, Lemma 25.14).<sup>14</sup> Working with Assumption 1 in place of (9) broadens the applicability of this class of semiparametric weak identification asymptotics to non-i.i.d. models. It is clear from Definition 1 that a locally regular test sequence will have asymptotic null rejection probability (NRP) which does not exceed the nominal level under weak identification asymptotics  $P_{n,h}$ .<sup>15</sup>

I now give two examples of semiparametric models where the parameter of interest  $\theta$  may be un- or under-identified depending on the value of an infinite dimensional nuisance parameter.<sup>16</sup> The first is the running example.

EXAMPLE 1 (continued): As is clear from the model equation  $Y = f(X_1 + X'_2\theta) + \epsilon$ , if f is flat, i.e. f' = 0, then the parameter  $\theta$  is unidentified. The sequences given in (3) are weak identification asymptotic sequences if f' = 0.

EXAMPLE 2 (IV): Suppose the researcher observes n i.i.d. copies of W = (Y, X, Z),

$$Y = X'\theta + Z'_1\beta + \epsilon, \qquad \mathbb{E}[\epsilon|Z] = 0, \qquad Z = (Z'_1, Z'_2)'.$$

If  $\pi(Z) \coloneqq \mathbb{E}[X|Z]$  is constant,  $\theta$  is unidentified; if some components of  $\pi(Z)$  are constant,  $\theta$  is underidentified.

In Examples 1 and 2, at the points of identification failure, no locally regular estimator exists, however locally regular  $C(\alpha)$  tests are developed in Section 5.<sup>17</sup>

<sup>&</sup>lt;sup>14</sup>If Assumption 1 holds with  $\Delta_n$  having this form, the converse is also true.

<sup>&</sup>lt;sup>15</sup>Of course, a (non-regular) test sequence may have asymptotic NRP which depends on b and yet is bounded by the nominal level under  $P_{n,h}$  for all  $h = (0,b) \in H_0$  and / or have an asymptotic power function which depends on b for  $h = (\tau, b) \in H_1$ . Restricting attention to locally regular test sequences may be justified by the power optimality results of Section 3. <sup>16</sup>A further example is the linear simultaneous equations model in Lee and Mesters (2024a).

<sup>&</sup>lt;sup>17</sup>These examples consider i.i.d. data for simplicity. See Hoesch, Lee, and Mesters (2024) for an

#### 2.4 A class of locally regular tests

To construct locally regular tests of  $H_0$ :  $\theta = \theta_0$  against  $H_1$ :  $\theta \neq \theta_0$ , I use a generalisation of the class of  $C(\alpha)$  tests introduced by Neyman (1959, 1979) to characterise optimal tests in regular parametric models. These tests are a based on a quadratic form of (estimators of) a vector of  $d_{\theta}$  moment conditions  $g_n \in L_2(P_n)$ which satisfy the following requirements.

ASSUMPTION 2 (Joint convergence): For  $g_n \in L_2(P_n)^{d_\theta}$  and each  $h = (\tau, b) \in H$ ,

$$(\Delta_n h, g'_n)' \stackrel{P_n}{\rightsquigarrow} \mathcal{N}(0, \Sigma(h)),$$
  
$$\Sigma(h) \coloneqq \begin{bmatrix} \sigma(h) & \tau' \Sigma'_{21} \\ \Sigma_{21} \tau & V \end{bmatrix} = \lim_{n \to \infty} \begin{bmatrix} \|\Delta_n h\|^2 & \langle \Delta_n(\tau, 0), g'_n \rangle \\ \langle g_n, \Delta_n(\tau, 0) \rangle & \langle g_n, g'_n \rangle \end{bmatrix}.$$

Built-in to Assumption 2 is a requirement of asymptotic orthogonality of  $g_n$ and the scores for the nuisance parameters  $\eta$ . This generalises the analogous condition in Neyman (1959, 1979) and is key to the local regularity of  $C(\alpha)$  tests.

REMARK 3: For Assumption 2 to hold it is necessary that the  $g_n$  are approximately zero mean: since  $(g_n)_{n\in\mathbb{N}}$  is uniformly  $P_n$ -integrable,  $P_ng_n = o(1)$ . It is also necessary that the  $g_n$  satisfy an approximate orthogonality property with the scores for nuisance parameters: as  $([\Delta_n h]g_n)_{n\in\mathbb{N}}$  is uniformly  $P_n$ -integrable for each h = $(\tau, b) \in H$ ,  $\lim_{n\to\infty} \langle \Delta_n h, g'_n \rangle = \tau' \Sigma'_{21} = \lim_{n\to\infty} \langle \Delta_n(\tau, 0), g'_n \rangle$ , and so

$$\langle \Delta_n(0,b), g'_n \rangle = \langle \Delta_n h, g'_n \rangle - \langle \Delta_n(\tau,0), g'_n \rangle = o(1).$$
(10)

Given any  $d_{\theta}$  moment conditions  $f_n \in L_2^0(P_n)$ , moment conditions which satisfy an exact version of the orthogonality condition (10) may be obtained as

$$g_n \coloneqq \Pi \left[ f_n \middle| \{ \Delta_n(0, b) : b \in B \}^\perp \right].$$
(11)

An important special case of this construction is with  $f_n$  the score function for  $\theta$ , i.e.  $f_n = \dot{\ell}_n$  such that  $\tau' \dot{\ell}_n = \Delta_n(\tau, 0)$  for each  $\tau \in \mathbb{R}^{d_\theta}$ . The function

$$g_n = \tilde{\ell}_n \coloneqq \Pi \left[ \dot{\ell}_n \middle| \{ \Delta_n(0, b) : b \in B \}^{\perp} \right],$$
(12)

is called the *efficient score function*. This yields a power optimal choice of moment

example of a locally regular  $C(\alpha)$  test of the form proposed in this paper for the potentially un- / under-identified parameter in a structural vector autoregressive model.

conditions satisfying (10) as shown in Section 3 below.

EXAMPLE 1 (continued): Let  $\omega : \mathbb{R}^K \to [\underline{\omega}, \overline{\omega}] \subset (0, \infty)$ . Then  $g_n \coloneqq \mathbb{G}_n g_n$ 

$$g(W) \coloneqq \boldsymbol{\omega}(X)(Y - f(V_{\theta}))f'(V_{\theta})\left(X_2 - \frac{\mathbb{E}[\boldsymbol{\omega}(X)X_2|V_{\theta}]}{\mathbb{E}[\boldsymbol{\omega}(X)|V_{\theta}]}\right),\tag{13}$$

has components which belong to  $\{\Delta_n(0,b) : b \in B\}^{\perp}$  (where  $\Delta_n$  is as in (6)).<sup>18</sup> Under regularity conditions,  $g_n$  satisfies Assumption 2 (see Section 5.1 below).

To construct the test statistic, I assume that consistent estimators of  $g_n$ ,  $V^{\dagger}$ (the Moore-Penrose pseudo-inverse of V) and  $r \coloneqq \operatorname{rank}(V)$  are available, given  $\theta$ .

ASSUMPTION 3 (Consistent estimation):  $\hat{g}_{n,\theta}$ ,  $\hat{\Lambda}_{n,\theta}$ ,  $\hat{r}_{n,\theta} \in \{0, 1, \dots, d_{\theta}\}$  satisfy

(i)  $\hat{g}_{n,\theta} - g_n \xrightarrow{P_n} 0;$ (ii)  $\hat{\Lambda}_{n,\theta} \xrightarrow{P_n} V^{\dagger};$ (iii) If  $r \ge 1$ , then  $\hat{r}_{n,\theta} \xrightarrow{P_n} r$ ; if r = 0, then  $\operatorname{rank}(\hat{\Lambda}_{n,\theta}) \xrightarrow{P_n} 0.$ 

Verification of Assumption 3(i) typically proceeds by model specific arguments. That  $g_n$  is (asymptotically) orthogonal to  $\{\Delta_n(0,b) : b \in B\}$  often helps in establishing this consistency (cf. Chernozhukov et al., 2022). One generally applicable approach to obtain an estimator which satisfies Assumption 3(ii) is to take an initial estimator which is consistent for V, threshold its eigenvalues at an appropriate rate and then take the pseudo-inverse.<sup>19</sup> If one uses the estimator  $\hat{\Lambda}_{n,\theta} := \hat{V}_{n,\theta}^{\dagger}$ where  $\hat{V}_{n,\theta} \xrightarrow{P_n} V$  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.

EXAMPLE 1 (continued): Given estimators  $\hat{f}_{n,i}$ ,  $\hat{f}'_{n,i}$  of f, f' and  $\hat{Z}_{1,n,i}, \hat{Z}_{2,n,i}$  of  $Z_1 := \mathbb{E}[\boldsymbol{\omega}(X)X_2|V_{\theta}], Z_2 := \mathbb{E}[\boldsymbol{\omega}(X)|V_{\theta}], define <math>\hat{g}_{n,\theta} := \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{g}_{n,\theta,i},$ 

$$\hat{g}_{n,\theta,i} \coloneqq \mathbf{\omega}(X_i)(Y_i - \hat{f}_{n,i}(V_{\theta,i}))\hat{f'}_{n,i}(V_{\theta,i}) \left(X_{2,i} - \hat{Z}_{1,n,i}(V_{\theta,i})/\hat{Z}_{2,n,i}(V_{\theta,i})\right).$$
(14)

18g coincides with the efficient score function  $\tilde{\ell}$  (derived by Newey and Stoker, 1993),

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

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

<sup>19</sup>See Section S5 of Lee and Mesters (2024b) for full details of this approach. Other regularisation schemes are also possible (see e.g. Dufour and Valéry, 2016).

Under regularity conditions (see Section 5.1),  $\hat{g}_{n,\theta}$  satisfies part (i) of Assumption 3, thresholding the eigenvalues of  $\frac{1}{n} \sum_{i=1}^{n} \hat{g}_{n,\theta,i} \hat{g}'_{n,\theta,i}$  at an appropriate rate yields an estimator  $\hat{\Lambda}_{n,\theta}$  which satisfies part (ii) and  $\hat{r}_{n,\theta} \coloneqq \operatorname{rank}(\hat{\Lambda}_{n,\theta})$  satisfies part (iii).

Given the estimators of Assumption 3, the  $C(\alpha)$ -style test statistic is

$$\hat{S}_{n,\theta} \coloneqq \hat{g}'_{n,\theta} \hat{\Lambda}_{n,\theta} \hat{g}_{n,\theta}.$$
(15)

The C( $\alpha$ ) – style test  $\psi_{n,\theta_0}$  of H<sub>0</sub> against H<sub>1</sub> at level  $\alpha$  is:

$$\psi_{n,\theta_0} \coloneqq \mathbf{1} \left\{ \hat{S}_{n,\theta_0} > c_n \right\},\tag{16}$$

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

**Local regularity** Assumptions 1 - 3 suffice for local regularity of  $\psi_{n,\theta_0}$ .

**PROPOSITION 1:** Under Assumptions 1 and 2, for  $h = (\tau, b) \in H$ 

$$g_n \stackrel{P_{n,h}}{\rightsquigarrow} \mathcal{N}\left(\Sigma_{21}\tau, V\right)$$

If Assumption 3 also holds, then additionally

$$\hat{g}_{n,\theta_0} \xrightarrow{P_{n,h}} \mathcal{N}(\Sigma_{21}\tau, V) \quad and \quad \hat{S}_{n,\theta_0} \xrightarrow{P_{n,h}} \chi_r^2(\tau' \Sigma'_{21} V \Sigma_{21}\tau).$$

THEOREM 1: Suppose that Assumptions 1, 2 and 3 hold and  $h = (\tau, b) \in H$ . Then,

$$\lim_{n \to \infty} P_{n,h} \psi_{n,\theta_0} = \pi(\tau) \coloneqq \begin{cases} 1 - \mathcal{P}\left(\chi_r^2 \left(\tau' \Sigma_{21}' V^{\dagger} \Sigma_{21} \tau\right) \le c_r\right) & \text{if } r \ge 1\\ 0 & \text{if } r = 0 \end{cases},$$

where  $c_r$  is the  $1 - \alpha$  quantile of the  $\chi^2_r$  distribution.

Theorem 1 immediately shows that  $\psi_{n,\theta_0}$  is locally regular (cf. (8)). The asymptotic orthogonality in (10) is key to this result. If, instead,  $\lim_{n\to\infty} \langle \Delta_n h, g'_n \rangle = \tau' \Sigma'_{21} + c(b)$  with  $c(b) \neq 0$ , then (by Le Cam's third Lemma) the limiting distribution of  $g_n$  under  $P_{n,h}$  would be  $\mathcal{N}(\Sigma_{21}\tau + c(b), V)$  and hence the limiting power function of the test sequence would not be free of b.

**Uniform local regularity** The local regularity given by 1 may be "upgraded" to local uniform regularity (Definition 2) under various conditions. Here I consider the case where H possesses a (pseudo-)metric structure (e.g. if H is a subset of

a (semi-)normed linear space).<sup>20</sup> In this case, for  $\psi_{n,\theta_0}$  to be locally uniformly regular on a compact (or totally bounded)  $K \subset H$  it is necessary and sufficient that the functions  $h = (\tau, b) \mapsto \pi_n(\tau, b)$  are asymptotically equicontinuous.

COROLLARY 1: Suppose that the conditions of Theorem 1 hold and that (H, d)is a pseudometric space. If the functions  $h = (\tau, b) \mapsto \pi_n(\tau, b) \coloneqq P_{n,h}\psi_{n,\theta_0}$  are asymptotically equicontinuous on a compact (or totally bounded) subset  $K \subset H$ ,

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

I now give a sufficient condition for the asymptotic equicontinuity required by Corollary 1.

LEMMA 1: If (H,d) is a pseudometric space and  $(h \mapsto P_{n,h})_{n \in \mathbb{N}}$  is asymptotically equicontinuous in  $d_{TV}$  on  $K \subset H$ , then  $(h \mapsto P_{n,h}\psi_{n,\theta})_{n \in \mathbb{N}}$  is asymptotically equicontinuous on K.

REMARK 4: Lemma 1 requires asymptotic equicontinuity in total variation of the functions  $(h \mapsto P_{n,h})_{n \in \mathbb{N}}$  on subsets  $K \subset H$ . This holds for any compact K under ULAN (Assumption S1), as shown in Proposition S1.

In the parametric i.i.d. case LAN is often verified by establishing a DQM condition, e.g. equation (7.1) in van der Vaart (1998). This is sufficient for the ULAN expansion in Assumption S1 to hold (e.g. van der Vaart, 1998, Theorem 7.2). Semiparametric generalisations of this result are available (e.g. combine Proposition S1 and Lemma S1).<sup>21</sup>

The condition in Lemma 1 is natural given its link with the ULAN condition. Neverthelesss, it is (much) stronger than necessary for the condition required by Corollary 1; see Lemma S2 for a weaker sufficient condition.

### **3** Power optimality

The preceding section established the local regularity of the tests  $\psi_{n,\theta_0}$  based on (estimates of) moment functions  $g_n$  satisfying certain asymptotic orthogonality conditions. Thus far, nothing has been said about the choice of  $g_n$  beyond these

<sup>&</sup>lt;sup>20</sup>If *H* possesses a (finite) measure structure and the functions  $h = (\tau, b) \mapsto \pi_n(\tau, b)$  are measurable then  $\psi_{n,\theta_0}$  is locally uniformly regular except on a "small" subset of *H* by Egorov's Theorem. See Section S2.3.2 for details.

<sup>&</sup>lt;sup>21</sup>Lee and Mesters (2024a) and Hoesch et al. (2024) verify this asymptotic equicontinuity property in i.i.d. and time series semiparametric examples respectively.

orthogonality requirements. The choice of the functions  $g_n$  determines the 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 (12). It is well known that tests based on the efficient score function have certain optimality properties in regular models when (a) the observations are i.i.d. (cf. Section 25.6 van der Vaart, 1998) or (b) when the information operator for  $\eta$  is boundedly invertible (Choi et al., 1996). I show that this optimality persists in non-regular models and does not require (a) or (b).

The results in this section are derived using the limits of experiments framework of Le Cam (e.g. Le Cam, 1986; van der Vaart, 1998). In particular, I show that the local experiments consisting of the measures  $P_{n,h}$  for  $h \in H$  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 under the seminorm induced by the variance function  $\sigma(h)$ . The connection between these experiments is sufficiently tight that power bounds derived in the latter transfer to the former.<sup>22</sup>

**The limit experiment** For this development H is required to be linear and I will therefore assume that B (hence H) is a linear space. Under LAN, there exists a positive semi-definite symmetric bilinear form  $\langle \cdot, \cdot \rangle_K$  on  $H = \mathbb{R}^{d_{\theta}} \times B$  such that  $\sigma(h) = \langle h, h \rangle_K$ . This can be seen as a by-product of the following Lemma.

LEMMA 2: Suppose Assumption 1 holds and B is a linear space. Let  $\Delta$  be a square integrable stochastic process defined on H such that  $\Delta_n h \stackrel{P_n}{\rightsquigarrow} \Delta h$ . Then  $\Delta$  is a mean-zero Gaussian linear process with covariance kernel K, where

$$K(h,g) \coloneqq \lim_{n \to \infty} P_n \left[ \Delta_n h \Delta_n g \right]$$

For  $h, g \in H$ , setting  $\langle h, g \rangle_K \coloneqq K(h, g)$  gives a positive semi-definite symmetric bilinear form. Let  $\|\cdot\|_K$  denote the seminorm induced by  $\langle \cdot, \cdot \rangle_K$  on H.

REMARK 5: Suppose that  $\langle \cdot, \cdot \rangle_H$  is an inner product on  $H = \mathbb{R}^{d_{\theta}} \times B$ . The existence of the positive semi-definite symmetric bilinear form  $\langle \cdot, \cdot \rangle_K$  is equivalent to the existence of a bounded, self-adjoint, positive semi-definite linear operator B such that  $\langle h, h \rangle_K = \langle h, Bh \rangle_H$  for  $h \in H$  (cf. Choi et al., 1996, p. 845).

<sup>&</sup>lt;sup>22</sup>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.

Define  $\mathbb{H}$  as the quotient of H by the subspace on which  $\|\cdot\|_K$  vanishes:

$$\mathbb{H} := H / \{ h \in H : ||h||_{K} = 0 \}.$$
(17)

which is an inner product space when equipped with the natural inner product induced by  $\langle \cdot, \cdot \rangle_K$ , which I also denote by  $\langle \cdot, \cdot \rangle_K$ . An element of  $\mathbb{H}$  corresponding to representative element  $h \in H$  will be denoted by [h].<sup>23</sup>

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

PROPOSITION 2: Suppose that Assumption 1 holds, that B is a linear space and define the sequence of experiments  $\mathscr{E}_n := (\mathcal{W}_n, \mathcal{B}(\mathcal{W}_n), (P_{n,h} : h \in H))$ . Let  $\Delta$  be the Gaussian process of Lemma 2 and let  $(\Omega, \mathcal{F}, P)$  be the probability space on which it is defined. Define the experiment  $\mathscr{E} := (\Omega, \mathcal{F}, (P_h : h \in H))$  according to

$$P_0 \coloneqq \mathbf{P}; \qquad \frac{\mathrm{d}P_h}{\mathrm{d}P_0} = \exp\left(\Delta h - \frac{1}{2}\|h\|^2\right), \quad h \in H.$$

Then  $\mathscr{E}_n$  converges weakly to  $\mathscr{E}$ .

Under the assumption that  $\mathbb{H}$  is separable, the experiment  $\mathcal{E}$  is equivalent to a Gaussian shift on  $(\mathbb{H}, \langle \cdot, \cdot \rangle_K)$ , in the sense given by Proposition 3 below.

ASSUMPTION 4: B is a linear space and  $\mathbb{H}$  as defined in (17) is separable.

PROPOSITION 3: Suppose Assumptions 1 and 4 hold. If  $\mathscr{E}$  is as in Proposition 2, there is a Gaussian shift experiment  $\mathscr{G} := (\Omega, \mathcal{F}, (G_{[h]} : [h] \in \mathbb{H}))$  such that  $d_{TV}(P_h, G_{[h]}) = 0$  for each  $h \in H$ .

The efficient information matrix Power bounds for tests of  $K_0 : h \in H_0$ against  $K_1 : h \notin H_1$  can be expressed in terms of the *efficient information matrix*,  $\tilde{\mathcal{I}}$ , so named because in the i.i.d. setting it is the covariance matrix of the efficient score function for a single observation. Here I provide an alternative definition of this matrix which applies more generally, and reduces to the classical definition in the i.i.d. case (as shown in Lemma S4).

Let  $\|\tau\| \coloneqq \inf_{b \in B} \|(\tau, b)\|_K$ , which defines a semi-norm on  $\mathbb{R}^{d_{\theta}}$ . Equipping the quotient  $\mathbb{H}_1 \coloneqq \mathbb{R}^{d_{\theta}} / \{\tau \in \mathbb{R}^{d_{\theta}} : \|\tau\| = 0\}$  with the natural norm induced by  $\|\cdot\|$ 

<sup>&</sup>lt;sup>23</sup>Analogous comments apply to the related space  $\mathbb{H}_1$ , 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  $[(\tau, b)]$ .

(which I also denote by  $\|\cdot\|$ ) turns it into a normed space. Define the linear map  $\pi_1 : \mathbb{H} \to \mathbb{H}_1$  as  $\pi_1([\tau, b]) \coloneqq [\tau]$ . As  $\pi_1$  is continuous it may be uniquely extended to a continuous function defined on  $\overline{\mathbb{H}}$ , the completion of  $\mathbb{H}$ ; this extension will also be called  $\pi_1$ . Since  $\pi_1$  is continuous, ker  $\pi_1 \subset \overline{\mathbb{H}}$  is closed. Let  $\Pi$  be the orthogonal projection onto ker  $\pi_1$  and define  $\Pi^{\perp} \coloneqq I - \Pi$ , the orthogonal projection onto  $[\ker \pi_1]^{\perp}$ . Let  $e_i$  be the *i*-th canonical basis vector in  $\mathbb{R}^{d_{\theta}}$  and define the *efficient information matrix*  $\tilde{\mathcal{I}}$  as the  $d_{\theta} \times d_{\theta}$  matrix with *i*, *j*-th entry  $\tilde{\mathcal{I}}_{ij}$  given by<sup>24</sup>

$$\tilde{\mathcal{I}}_{ij} = \left\langle \Pi^{\perp}[e_i, 0], \, \Pi^{\perp}[e_j, 0] \right\rangle_K. \tag{18}$$

LEMMA 3: Under Assumption 4,  $\|\tau\|^2 = \tau' \tilde{\mathcal{I}} \tau$  and  $\ker \tilde{\mathcal{I}} = \{\tau \in \mathbb{R}^{d_{\theta}} : \|\tau\| = 0\}.$ 

#### **3.1** Tests of a scalar parameter

The following Theorem records the power bound for (locally asymptotically) unbiased two-sided tests of a scalar  $\theta$ . In this case the matrix  $\tilde{\mathcal{I}}$  has rank either 0 or 1. Theorem 2 handles both cases simultaneously.

THEOREM 2: Suppose that Assumptions 1 and 4 hold and  $d_{\theta} = 1$ . Let  $\phi_n : \mathcal{W}_n \to [0, 1]$  be a sequence of locally asymptotically unbiased level  $\alpha$  tests of  $K_0 : \tau = 0$  against  $K_1 : \tau \neq 0$ . That is,

 $\limsup_{n \to \infty} P_{n,h} \phi_n \leq \alpha, \quad h \in H_0, \qquad and \qquad \liminf_{n \to \infty} P_{n,h} \phi_n \geq \alpha, \quad h \in H_1.$ 

Then, for any  $h \in H$ ,

$$\limsup_{n \to \infty} P_{n,h} \phi_n \le 1 - \Phi\left(z_{\alpha/2} - \tilde{\mathcal{I}}^{1/2}\tau\right) + 1 - \Phi\left(z_{\alpha/2} + \tilde{\mathcal{I}}^{1/2}\tau\right),\tag{19}$$

where  $z_{\alpha}$  is the  $1-\alpha$  quantile and  $\Phi$  the CDF of the standard normal distribution.

Theorem 1 implies that the power bound of Theorem 2 is achieved by the test  $\psi_{n,\theta_0}$  provided  $\Sigma_{21}V^{\dagger}\Sigma'_{21} = \tilde{\mathcal{I}}$  and r = 1.

COROLLARY 2: Suppose that Assumptions 1, 2 and 3 hold with  $\Sigma_{21}V^{\dagger}\Sigma'_{21} = \tilde{\mathcal{I}}$  and r = 1. Then, for  $h \in H$ ,

$$\lim_{n \to \infty} P_{n,h} \psi_{n,\theta_0} = 1 - \Phi \left( z_{\alpha/2} - \tilde{\mathcal{I}}^{1/2} \tau \right) + 1 - \Phi \left( z_{\alpha/2} + \tilde{\mathcal{I}}^{1/2} \tau \right).$$
(20)

<sup>&</sup>lt;sup>24</sup>Lemma S3 gives an alternative expression for  $\tilde{\mathcal{I}}$  based on the Gaussian process  $\Delta$  of Lemma 2.

#### 3.2 Tests of a multivariate parameter

When  $d_{\theta} > 1$  there is an intermediate case where  $0 < \operatorname{rank}(\tilde{\mathcal{I}}) < d_{\theta}$ . Here I permit  $0 < \operatorname{rank}(\tilde{\mathcal{I}}) \leq d_{\theta}$  and establish a maximin power bound for (potentially) non – regular models, which contains the (regular) full rank case as a special case.<sup>25</sup>

THEOREM 3: Suppose that Assumptions 1 and 4 hold and  $r := \operatorname{rank}(\tilde{\mathcal{I}}) \geq 1$ . Let  $\phi_n : \mathcal{W}_n \to [0, 1]$  be a sequence of tests such that for each  $h = (0, b) \in H_0$ 

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

Let  $c_r$  the  $1 - \alpha$  quantile of a  $\chi_r^2$  random variable. Then, if  $a \ge 0$ ,

$$\limsup_{n \to \infty} \inf \left\{ P_{n,h} \phi_n : h = (\tau, b) \in H, \ \tau' \tilde{\mathcal{I}} \tau \ge a \right\} \le 1 - \mathcal{P}(\chi_r^2(a) \le c_r).$$
(21)

By Theorem 1, the power bound on the right hand side of (21) is achieved by  $\psi_{n,\theta_0}$  provided  $\Sigma_{21}V^{\dagger}\Sigma'_{21} = \tilde{\mathcal{I}}$  and  $\operatorname{rank}(V) = \operatorname{rank}(\tilde{\mathcal{I}}) = r \geq 1$ . In order that the test be asymptotically maximin over a compact subset  $K_a$  of  $\{h = (\tau, b) \in H : \tau'\tilde{\mathcal{I}}\tau \geq a\}$ , with  $a = \inf\{\tau'\tilde{\mathcal{I}}\tau = a : h \in K_a\}$ , some uniformity is required.<sup>26</sup>

COROLLARY 3: Suppose that Assumptions 1, 2 and 3 hold with  $\Sigma_{21}V^{\dagger}\Sigma'_{21} = \tilde{\mathcal{I}}$  and  $r = \operatorname{rank}(\tilde{\mathcal{I}}) = \operatorname{rank}(V) \geq 1$ . Then for  $h = (\tau, b) \in H$ 

$$\lim_{n \to \infty} P_{n,h} \psi_{n,\theta_0} = 1 - \mathcal{P}\left(\chi_r^2\left(a\right) \le c_r\right), \quad a = \tau' \tilde{\mathcal{I}} \tau.$$
(22)

Additionally, suppose that (H, d) is a (pseudo-)metric space and let  $K_a$  be a compact subset of  $\{h = (\tau, b) \in H : \tau' \tilde{\mathcal{I}} \tau \geq a\}$  such that  $a = \inf\{\tau' \tilde{\mathcal{I}} \tau : h = (\tau, b) \in K_a\}$ . If the functions  $h \mapsto P_{n,h}\psi_{n,\theta_0}$  are asymptotically equicontinuous on  $K_a$ ,

$$\lim_{n \to \infty} \inf_{h \in K_a} P_{n,h} \psi_{n,\theta_0} = 1 - \mathcal{P}\left(\chi_r^2\left(a\right) \le c_r\right).$$
(23)

A sufficient condition for the asymptotic equicontinuity required for the second part of Corollary 3 based on an asymptotic equicontinuity in total variation requirement was given as Lemma 1 in the previous section.<sup>27</sup>

<sup>&</sup>lt;sup>25</sup>Section S2.5 shows that 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 asymptotically level  $\alpha$  tests can correspond to a test in the limit experiment with smaller regret.

<sup>&</sup>lt;sup>26</sup>The pseudometric d in Corollary 3 need not be related to the seminorm  $\|\cdot\|_{K}$ .

<sup>&</sup>lt;sup>27</sup>As noted there Lemma S2 provides some weaker sufficient conditions; in the present context, condition (iii) of Lemma S2 is not required (cf. Remark S3).

#### 3.3 The degenerate case

If the efficient information matrix  $\tilde{\mathcal{I}}$  is zero, no test with correct asymptotic size has non – trivial asymptotic power against any sequence of local alternatives.

PROPOSITION 4: Suppose Assumptions 1 and 4 hold and  $r \coloneqq \operatorname{rank}(\mathcal{I}) = 0$ . Let  $\phi_n : \mathcal{W}_n \to [0, 1]$  be a sequence of tests such that  $\limsup_{n \to \infty} P_{n,h}\phi_n \leq \alpha$  for each  $h = (0, b) \in H_0$  Then, for  $h \in H$ ,  $\limsup_{n \to \infty} P_{n,h}\phi_n \leq \alpha$ .

#### 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 multivariate optimality results just presented 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, one could construct a locally regular test by utilising the same moment conditions  $g_n$  but weighting them differently (cf. Bickel, Ritov, and Stoker, 2006).

The intermediate case with  $1 \leq \operatorname{rank}(\tilde{\mathcal{I}}) < d_{\theta}$  A key benefit of the multivariate power results is that they apply equally to non-regular models, i.e. cases where  $\tilde{\mathcal{I}}$  is rank deficient. This scenario can occur for various reasons. Firstly the model may not identify all parameters of interest  $\theta$  (i.e. underidentification). Secondly some of the elements of  $\theta$  may be weakly identified (i.e. 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 (e.g. Rotnitzky, Cox, Bottai, and Robins, 2000; Han and McCloskey, 2019; Andrews and Guggenberger, 2019; Amengual, Bei, and Sentana, 2023); none of these papers consider optimal testing.

Alternative approximations In the case where  $\operatorname{rank}(\tilde{\mathcal{I}}) = 0$ , Proposition 4 reveals that the LAN approximation in Assumption 1 is, in a certain sense, the wrong approximation: it does not provide any useful way of (asymptotically) comparing tests. 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 developed for the class of models considered here is an interesting question for future work.

#### 3.5 Attaining the power bounds

Provided that the  $L_2$  distance between  $g_n$  and  $\tilde{\ell}_n$  (as defined in (12)) vanishes,  $\psi_{n,\theta_0}$ attains the power bounds established in the preceding subsections. For regular models this result is well known in two special cases: (a) the i.i.d. case (cf. Section 25.6 in van der Vaart (1998); Lemma S4) and (b) when the information operator B in Remark 5 is positive-definite with B<sub>22</sub>, the information operator for  $\eta$ , boundedly invertible (Choi et al., 1996). Here I provide a general version of this result which applies to both regular and non-regular models and does not require (a) or (b). In particular, I show that  $\Sigma_{21}V^{\dagger}\Sigma'_{21} = \tilde{\mathcal{I}}$ , which suffices given Theorem 1 and the power bounds in Theorems 2 and 3.

THEOREM 4: Suppose that Assumptions 1, 2, 3 and 4 hold and  $g_n$  is such that  $\lim_{n\to\infty} \int \|g_n - \tilde{\ell}_n\|^2 \, \mathrm{d}P_n = 0$ . Then  $\Sigma_{21} = V = \tilde{\mathcal{I}}$ , hence  $\Sigma_{21} V^{\dagger} \Sigma'_{21} = \tilde{\mathcal{I}}$ .

# 4 The smooth i.i.d. case

In this section I give conditions which are sufficient for some of the foregoing Assumptions in the in the benchmark case for semiparametric theory: where the observations are i.i.d. and the model is "smooth".

ASSUMPTION 5 (Product measures): Suppose  $W^{(n)} = (W_1, \ldots, W_n) \in \prod_{i=1}^n \mathcal{W} = \mathcal{W}_n$  and that each  $P_{n,h}$  is a product measure:  $P_{n,h} = P_h^n$ . Each probability measure in  $\mathcal{P}_n$  is dominated by the n-fold product of a  $\sigma$ -finite measure  $\nu$ .

In the i.i.d. setting, it is well known that quadratic mean differentiability of the square root of the density  $p = \frac{dP}{d\nu}$  of  $P \coloneqq P_0$  is sufficient for LAN. In particular, if

$$\lim_{n \to \infty} \int \left[ \sqrt{n} \left( \sqrt{p_{h_n}} - \sqrt{p} \right) - \frac{1}{2} A h \sqrt{p} \right]^2 = 0, \qquad (24)$$

for a measurable  $Ah : \mathcal{W} \to \mathbb{R}$ , then with  $\Delta_n h \coloneqq \frac{1}{\sqrt{n}} \sum_{i=1}^n [Ah](W_i)$  the remainder term  $R_n$  in the LAN expansion satisfies  $R_n(h_n) \xrightarrow{P} 0$  (e.g. van der Vaart and Wellner, 1996, Lemma 3.10.11). This can be used to establish either the LAN expansion required by Assumption 1 by taking  $h_n = h$  for each  $n \in \mathbb{N}$  or the ULAN expansion as in Assumption S1 by considering sequences  $h_n \to h$ . Sufficient conditions for (24) are well known (e.g. van der Vaart, 1998, Lemma 7.6).

In this case, the scores Ah typically take the form

$$[Ah](W_i) = \tau' \dot{\ell}(W_i) + [Db](W_i), \quad h = (\tau, b) \in H,$$
(25)

where  $\dot{\ell}$  is a  $d_{\theta}$ -vector of functions in  $L_2^0(P)$  (typically the partial derivatives of  $\theta \mapsto \log p_{\gamma}$  at  $\gamma$ ) and  $D : \overline{\lim} B \to L_2^0(P)$  a bounded linear map. Showing that (24) holds (with  $h_n = h$ ) is typically the most straightforward way to verify the LAN expansion required by Assumption 1. If  $A : \overline{\lim} H \to L_2(P)$  is a bounded linear map, then the remainder of Assumption 1 also follows directly.<sup>28</sup>

LEMMA 4: Suppose that Assumption 5 holds and for each  $h \in H$  equation (24) holds (with  $h_n = h$ ) with  $A : \overline{\lim} H \to L_2(P)$  a bounded linear map. Then Assumption 1 holds with  $P_{n,h} = P_{h/\sqrt{n}}^n$  and  $[\Delta_n h](W^{(n)}) = \mathbb{G}_n Ah$ .

When the data are i.i.d., the the joint convergence of  $(\Delta_n h, g'_n)$  as in Assumption 2 is particularly straightforward to verify. As noted in the discussion around (11), the required orthogonality condition can be ensured by performing an orthogonal projection. Assumption 2 then follows straightforwardly. In the i.i.d. setting typically  $g_n$  will have the form  $g_n(W^{(n)}) = \mathbb{G}_n g$ .

LEMMA 5: Suppose that Assumptions 1 and 5 hold, with  $\Delta_n h = \frac{1}{\sqrt{n}} \sum_{i=1}^n Ah$ , where Ah is as in equation (25). Additionally suppose that  $g \in \{Db : b \in B\}^{\perp} \subset L_2^0(P)$ . Then Assumption 2 holds with  $g_n(W^{(n)}) \coloneqq \mathbb{G}_n g$ .

COROLLARY 4: In the setting of Lemma 5, if  $f \in L_2^0(P)$  and g is the orthogonal projection  $g = \prod[f|\{Db : b \in B\}^{\perp}]$  then Assumption 2 holds with  $g_n(W^{(n)}) \coloneqq \mathbb{G}_n g$ .

### 5 Examples

I now illustrate the application of the theoretical results to the single index and IV models and conduct simulation studies to investigate finite sample performance of the proposed approach. In this section I work under high level conditions to avoid repeating standard regularity conditions; lower level sufficient conditions are given in section S4 of the supplementary material.

 $<sup>^{28}</sup>$ A version of Lemma 4 for ULAN (Assumption S1) is Lemma S1 in the supplementary material.

#### 5.1 Single index model

Consider the single index model of Example 1. I now formalise the development given in Section 2. The model parameters are  $\gamma = (\theta, \eta)$  where  $\eta = (f, \zeta)$  and the density of one observation with respect to a  $\sigma$ -finite measure  $\tilde{\nu}$  is  $p_{\gamma}$  as in (2);  $P_{\gamma}$ denotes the corresponding probability measure. The parameters  $\gamma$  are restricted by the following Assumption. Let  $\mathscr{X}$  be the support of X,  $\mathscr{D}$  a convex open set containing  $\{x_1 + x'_2\theta : \theta \in \Theta, x \in \mathscr{X}\}$  and  $C_b^1(\mathscr{D})$  the class of real functions which are bounded and continuously differentiable with bounded derivative on  $\mathscr{D}$ .

Assumption 6: The parameters  $\gamma = (\theta, f, \zeta) \in \Gamma = \Theta \times \mathscr{F} \times \mathscr{Z}$  where  $\Theta$  is an open subset of  $\mathbb{R}^{d_{\theta}}$ ,  $\mathscr{F} = C_b^1(\mathscr{D})$  and  $\zeta \in \mathscr{Z}$ , for

$$\mathscr{Z} \coloneqq \left\{ \zeta \in L_1(\mathbb{R}^{1+K}, \nu) : \zeta \ge 0, \, \int_{\mathbb{R} \times \mathscr{X}} \zeta \, \mathrm{d}\nu = 1, \, \text{if} \, (\epsilon, X) \sim \zeta \, \text{then} \, (26) \, \text{holds} \right\},$$

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

$$\mathbb{E}[\epsilon|X] = 0, \ \mathbb{E}[\epsilon^2] < \infty, \ \mathbb{E}[(|\epsilon|^{2+\rho} + |\phi(\epsilon, X)|^{2+\rho} + 1) \|X\|^{2+\rho}] < \infty, \ \mathbb{E}[XX'] \succ 0,$$
(26)

for  $\phi(\epsilon, X)$  the derivative of  $e \mapsto \log \zeta(e, X)$ . Additionally, for each  $\gamma \in \Gamma$ ,  $p_{\gamma}$  is a probability density with respect to some  $\sigma$ -finite measure  $\tilde{\nu}$ .

That  $p_{\gamma}$  is a valid probability density holds automatically (with  $\tilde{\nu} = \nu$ ) when  $\epsilon | X$  is continuously distributed.

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

$$\varphi_n(h) = \left(\frac{\tau}{\sqrt{n}}, \ \varphi_{n,2}(b_1, b_2)\right), \qquad h = (\tau, b_1, b_2) \in H = \mathbb{R}^{d_\theta} \times B_1 \times B_2.$$
(27)

 $B_1$  is the set which indexes the perturbations to f and consists of a subset of the continuously differentiable functions  $b_1 : \mathscr{D} \to \mathbb{R}$ .  $B_2$  indexes the perturbations to  $\zeta$  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

$$\mathbb{E}[b_2(\epsilon, X)] = 0, \ \mathbb{E}[\epsilon b_2(\epsilon, X)|X] = 0, \ \mathbb{E}[b_2(\epsilon, X)^2] < \infty \quad \text{for } (\epsilon, X) \sim \zeta.$$
(28)

The precise form of  $\varphi_{n,2}$  is left unspecified. It is required only that the local perturbations satisfy the LAN property below.<sup>29</sup>

<sup>&</sup>lt;sup>29</sup>Examples of  $\varphi_{n,2}$  and B for which Assumption 7 holds are given in Section S4.1.2.

ASSUMPTION 7: Suppose that  $\mathcal{W}_n = \prod_{i=1}^n \mathbb{R}^{1+K}$  and  $P_{n,h} \coloneqq P_{\gamma+\varphi_n(h)}^n \ll \nu_n$  for all  $\gamma \in \Gamma$  and  $h \in H$  and are such that Assumption 1 holds with

$$\log \frac{p_{n,h}}{p_{n,0}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Ah](W_i) - \frac{1}{2}\sigma(h) + o_{P_{n,0}}(1), \quad h \in H,$$
(29)

where  $\sigma(h) = \int [Ah]^2 dP$  and A is as in equation (25) with

$$\hat{\ell}(W) \coloneqq -\phi(Y - f(V_{\theta}), X) f'(V_{\theta}) X_2$$
$$[Db](W) \coloneqq -\phi(Y - f(V_{\theta}), X) b_1(V_{\theta}) + b_2(Y - f(V_{\theta}), X).$$

**The moment conditions** The test statistic is based on  $g_n := \mathbb{G}_n g$  for g given in (13). This satisfies Assumption 2 under Assumptions 6 & 7 and (30) below.

**PROPOSITION 5:** Suppose Assumptions 6 & 7 hold and under P,

$$\mathbb{E}[\epsilon^2|X] \le C < \infty, \quad \mathbb{E}\left[\epsilon\phi(\epsilon, X)|X\right] = -1, \quad a.s.$$
(30)

Then Assumption 2 holds with  $g_n := \mathbb{G}_n g$  for g given in (13).

A feasible test To form a feasible test  $g_n$  must be replaced by an estimator  $\hat{g}_{n,\theta}$ . Let this have the form  $\hat{g}_{n,\theta}(W^{(n)}) \coloneqq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{g}_{n,\theta,i}$ , for  $\hat{g}_{n,\theta,i}$  defined as in (14). To keep the notation concise let  $Z_3 \coloneqq f$ ,  $Z_4 \coloneqq f'$ ,  $Z_0 \coloneqq Z_1/Z_2$  and correspondingly  $\hat{Z}_{0,n,i} \coloneqq \hat{Z}_{1,n,i}/\hat{Z}_{2,n,i}$ . Let  $\check{V}_{n,\theta} \coloneqq \frac{1}{n} \sum_{i=1}^{n} \hat{g}_{n,\theta,i} \hat{g}'_{n,\theta,i}$  and If V is known to have full rank then let  $\hat{V}_{n,\theta} \coloneqq \check{V}_{n,\theta}$ ,  $\hat{\Lambda}_{n,\theta} \coloneqq \hat{V}_{n,\theta}^{-1}$  and  $\hat{r}_{n,\theta} = \operatorname{rank}(V)$ . Form the estimator  $\hat{V}_{n,\theta}$  according to the construction in Section S5 of Lee and Mesters (2024b) using a truncation rate  $\mathbf{v}_n$ .  $\hat{\Lambda}_{n,\theta}$  is then taken to be  $\hat{V}_{n,\theta}^{\dagger}$  and  $\hat{r}_{n,\theta} \coloneqq \operatorname{rank}(\hat{V}_{n,\theta})$ . Under the following condition, these estimators satisfy the conditions of Assumption 3.

ASSUMPTION 8: Suppose that equation (30) holds (under P), X has compact support,  $\mathbb{E}[\epsilon^4] < \infty$ , and with P probability approaching one  $\mathsf{R}_{l,n,i} \leq r_n = o(n^{-1/4})$ ,

$$\mathsf{R}_{l,n,i} \coloneqq \left[ \int \left\| \hat{Z}_{l,n,i}(v) - Z_l(v) \right\|^2 \, \mathrm{d}\mathcal{V}(v) \right]^{1/2}, \quad l = 1, \dots 4$$

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

The rate conditions in Assumption 8 can be satisfied by e.g. (sample – split)

series estimators under standard conditions; see e.g. Belloni, Chernozhukov, Chetverikov, and Kato (2015).

PROPOSITION 6: Suppose Assumptions 6, 7 and 8 hold and  $\nu_n$  is such that  $r_n = o(\nu_n)$ . Then Assumption 3 holds with  $V \coloneqq \int gg' dP$ .

A consequence of Assumption 7 and Propositions 5 and 6 is that the test  $\psi_{n,\theta_0}$  formed as in (16) is locally regular by Theorem 1.

Simulation study I take K = 1 and test  $H_0: \theta = \theta_0 = 1$  at a nominal level of 5%. Each study reports the results of 5000 monte carlo replications with a sample size of  $n \in \{400, 600, 800\}$ . I report empirical rejection frequencies for the  $\psi_{n,\theta_0}$  test along with a Wald test based on an estimator in the style of Ichimura (1993).

I consider two different classes of link function. The first sets  $f(v) = f_j(v) = 5 \exp(-v^2/2c_j^2)$  ("exponential"); the second  $f(v) = f_j(v) = 25 (1 + \exp(-v/c_j))^{-1}$  ("logistic"). The values of  $c_j$  considered are recorded in Table 1.

Table 1: Functions used in the simulation experiments

name	expression	$c_1$	$c_2$	$c_3$
Exponential	$f_j(v) = 5\exp(-v^2/2c_j^2)$	1	2	4
Logistic	$f_j(v) = 25(1 + \exp(-v/c_j))^{-1}$	1	4	32

In each case, as  $c_j$  increases, the derivative of f flattens out, moving towards a point with f' = 0, at which  $\theta$  is unidentified.<sup>30</sup> I draw covariates as  $X = (Z_1, 0.2Z_1 + 0.4Z_2 + 0.8)$ , where each  $Z_k \sim U(-1, 1)$  is independent. The error term is drawn either as  $\epsilon = v/\sqrt{3/2}$  with  $v \sim t(6)$  ("homoskedastic") or  $\epsilon \sim \mathcal{N}(0, 1 + \sin(X_1)^2)$  ("heteroskedastic").

I compute the test  $\psi_{n,\theta_0}$  as described on p. 20, with  $\omega(X) = 1$ . The functions f, f' and  $Z_1$  are estimated via sample split smoothing splines.<sup>31</sup> The truncation parameter  $\nu$  is set to  $10^{-3}$ . I additionally compute a Wald test in the style of Ichimura (1993), using the same non-parametric estimates as for  $\hat{g}_{n,\theta}$ .<sup>32</sup>

The empirical rejection frequencies of these procedures are recorded in Table 2:  $\psi_{n,\theta_0}$  rejects at close to the nominal 5% for all simulation designs considered whilst the Wald test over – rejects in most of the simulation designs considered. Figure

<sup>&</sup>lt;sup>30</sup>The functions f and f' are plotted in Figures S1 and S2.

<sup>&</sup>lt;sup>31</sup>I use the base R function smooth.spline with 20 knots. In this setting  $Z_2(V_{\theta}) = 1$  is known. <sup>32</sup>Given  $\hat{f}, \hat{\theta} = \arg\min_{\theta \in \Theta_{\star}} \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{f}(V_{\theta,i}))^2$ , for  $\Theta_{\star} = [-10, 10]$ . The asymptotic variance is estimated by  $\hat{\sigma}^2 / \frac{1}{n} \sum_{i=1}^{n} \left( \hat{f}'(V_{\hat{\theta},i}) \left[ X_2 - \hat{Z}(V_{\hat{\theta},i}) \right] \right)^2$ , for  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{f}(V_{\hat{\theta},i}))^2$ .

1 contains power plots of the  $\psi_{n,\theta_0}$  test (n = 800). For almost flat link functions there is very identifying information and hence very little power available. As the link function moves away from the point of identification failure (f' = 0), the available power increases and is captured by the  $\psi_{n,\theta_0}$  test.

	Exponential							Logistic					
	Hor	noskeda	stic	Hete	Heteroskedastic		Но	Homoskedastic			Heteroskedastic		
n	$f_1$	$f_2$	$f_3$	$f_1$	$f_2$	$f_3$	$f_1$	$f_2$	$f_3$	$f_1$	$f_2$	$f_3$	
$\psi_{n, heta_0}$	)												
400	6.08	5.90	5.30	5.66	5.68	5.12	6.36	5.94	4.80	5.66	6.00	4.68	
600	6.12	5.68	5.26	5.44	4.76	4.32	6.04	5.74	4.10	5.26	5.18	4.62	
800	6.20	6.04	5.28	5.46	5.72	5.22	6.00	5.90	4.26	5.62	5.16	4.46	
Wale	1												
400	13.06	18.94	13.54	15.20	20.32	14.38	8.12	13.60	14.74	8.24	15.52	14.92	
600	10.28	16.32	12.58	10.60	18.92	14.18	7.18	10.74	13.88	6.82	12.60	14.84	
800	10.30	16.64	12.84	9.60	19.00	13.62	6.86	10.68	12.94	6.74	11.26	14.18	

Table 2: ERF (%), Single-index model

Figure 1: Single-index model,  $\psi$  power



### 5.2 IV model

In the IV model of Example 2, n i.i.d. copies of W = (Y, X, Z) are observed where

$$Y = X'\theta + Z'_1\beta + \epsilon, \qquad \mathbb{E}[\epsilon|Z] = 0, \qquad Z = (Z'_1, Z'_2)'.$$
 (31)

Let  $d_W \coloneqq d_\theta + d_Z + 1$ . With  $\pi(Z) \coloneqq \mathbb{E}[X|Z]$  and  $\upsilon = X - \pi(Z)$ ,

$$Y = X'\theta + Z'_1\beta + \epsilon, \qquad \mathbb{E}[U|Z] = 0, \quad U \coloneqq (\epsilon, \upsilon')'.$$
(32)  
$$X = \pi(Z) + \upsilon$$

If  $\pi(Z)$  is constant the instruments Z provide no information about  $\theta$ . Weak identification in this model can be very different from in the IV model with a linear first stage: there are many data configurations in which  $\operatorname{Var}(\mathbb{E}[X|Z])$  may be "large" whilst  $\mathbb{E}[XZ']\mathbb{E}[ZZ']^{-1} \approx 0$ . 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. In this section I develop a  $\psi_{n,\theta_0}$ test which can capture such identifying information whilst remaining robust to weak identification.<sup>33, 34</sup>

Let  $\zeta$  denote the density of  $\xi := (\epsilon, v', Z')$  with respect to a  $\sigma$ -finite measure  $\nu$ . The parameters of the IV model are  $\gamma = (\theta, \eta)$  with the nuisance parameters collected in  $\eta = (\beta, \pi, \zeta)$ . The density of one observation is

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

with respect to a  $\sigma$ -finite measure  $\tilde{\nu}$  and  $P_{\gamma}$  denotes the corresponding measure. The model parameters are restricted as follows.

Assumption 9: The parameters  $\gamma = (\theta, \beta, \pi, \zeta) \in \Gamma = \Theta \times \mathcal{B} \times \mathscr{P} \times \mathscr{Z}$  where

(i)  $\Theta$  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  $\mathbb{R}^{d_W}$  with respect to  $\nu$ ;

(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 
$$\phi_1 \coloneqq \nabla_{\epsilon} \log \zeta(\epsilon, v, Z), \ \phi_2 \coloneqq \nabla_{v} \log \zeta(\epsilon, v, Z) \ and \ \phi \coloneqq (\phi_1, \phi_2')'.$$

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

<sup>&</sup>lt;sup>33</sup>This does not contradict optimality results that are known for, e.g., the AR test (Moreira, 2009; Chernozhukov, Hansen, and Jansson, 2009) as these results assume a linear first stage.

<sup>&</sup>lt;sup>34</sup>An alternative approach to capturing this non-linear identifying information (whilst remaining robust to weak instruments) is to use a large number of transformations of the instruments,  $f_1(Z), \ldots, f_M(Z)$ , in a linear first stage, combined with a testing procedure which remains robust in the presence of many weak instruments. In the simulation study below, I compare the  $\psi_{n,\theta_0}$  test to this approach, using the test of Mikusheva and Sun (2022).

Assumption 9 imposes the existence of certain moments and the (IV) conditional mean restriction. That  $p_{\gamma}$  is a valid probability density holds automatically (with  $\nu = \tilde{\nu}$ ) when U|Z is continuously distributed.

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

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

with  $B := \mathbb{R}^{d_{\beta}} \times B_1 \times B_2$ .  $B_1$  is a subset of the bounded functions  $b_1 : \mathbb{R}^{d_Z} \to \mathbb{R}^{d_{\theta}}$ and  $B_2$  a subset of the functions  $b_2 : \mathbb{R}^{d_W} \to \mathbb{R}$  which are bounded and continuously differentiable in its first  $1 + d_{\theta}$  components with bounded derivative and such that

$$\mathbb{E}\left[b_2(U,Z)\right] = 0, \quad \mathbb{E}\left[Ub_2(U,Z)|Z\right] = 0, \quad \text{for } (U',Z')' \sim \zeta.$$
(35)

The precise forms of  $\varphi_{n,1}, \varphi_{n,2}$  are left unspecified. It is required only that the local perturbations satisfy LAN.<sup>35</sup>

Assumption 10: Suppose that  $\mathcal{W}_n = \prod_{i=1}^n \mathbb{R}^{d_W}$ ,  $P_{n,h} \coloneqq P_{\gamma+\varphi_n(h)}^n \ll \nu_n$  for all  $\gamma \in \Gamma$  and  $h \in H$  and are such that Assumption 1 holds with

$$\log \frac{p_{n,h}}{p_{n,0}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Ah](W_i) - \frac{1}{2}\sigma(h) + o_{P_{n,0}}(1), \quad h \in H,$$
(36)

where  $\sigma(h) = \int [Ah]^2 dP$  and A is as in equation (25) with

$$\ell(W) \coloneqq -\phi_1(\epsilon(\theta, \beta), \upsilon(\pi), Z) X_1$$
$$[Db](W) \coloneqq -\phi(\epsilon(\theta, \beta), \upsilon(\pi), Z)' [b'_0 Z_1 \ b_1(Z)] + b_2(\epsilon(\theta, \beta), \upsilon(\pi), Z),$$

where  $\epsilon(\theta, \beta) \coloneqq Y - X'\theta - Z'_1\beta$  and  $\upsilon(\pi) \coloneqq X - \pi(Z)$ .

The moment conditions The test will be based on moment conditions related to the efficient score function for  $\theta$ ,  $\tilde{\ell}$ . This is given in the following Lemma.<sup>36</sup>

LEMMA 6: Suppose Assumptions 9, 10 hold, for  $J(Z) := \mathbb{E}[UU'|Z]$ ,

$$0 < c \le \lambda_{\min}(J(Z)) \le \lambda_{\max}(J(Z)) \le C < \infty, \qquad \lambda_{\min}(\mathbb{E}[Z_1 Z_1']) > 0,$$
  
$$\mathbb{E}\left[\phi(\epsilon, \upsilon, Z)U'|Z\right] = -I, \qquad \mathbb{E}[\phi_1(\epsilon, \upsilon, Z)\upsilon U'] = 0,$$
(37)

<sup>35</sup>Examples of  $\varphi_{n,1}, \varphi_{n,2}$  and *B* for which Assumption 10 holds are given in Section S4.2.2. <sup>36</sup>The last two conditions in (37) hold if  $\lim_{|u_i|\to\infty} |u_i|\zeta(u,z) = 0$  for  $i = 1, \ldots, d_{\alpha}$ .

and that  $B_1$  is dense in  $L_2$ . Define  $\omega(Z) := \mathbb{E}[\epsilon^2 | Z]^{-1}$ . The efficient score for  $\theta$  is

$$\tilde{\ell}(W) = \boldsymbol{\omega}(Z)(Y - X'\boldsymbol{\theta} - Z_1'\boldsymbol{\beta}) \left[ \pi(Z) - \mathbb{E}[\boldsymbol{\omega}(Z)XZ_1']\mathbb{E}[\boldsymbol{\omega}(Z)Z_1Z_1']^{-1}Z_1 \right].$$
(38)

For simplicity, I will use the moment functions

$$g(W) \coloneqq \mathbb{E}[\epsilon^2]^{-1}(Y - X'\theta - Z'_1\beta) \left[\pi(Z) - \mathbb{E}[XZ'_1]\mathbb{E}[Z_1Z'_1]^{-1}Z_1\right].$$
(39)

g belongs to the orthocomplement of  $\{Db : b \in B\}$  and coincides with the efficient score function when  $J(Z) = \mathbb{E}[UU']$  a.s. (i.e. under homoskedasticity).<sup>37</sup>

LEMMA 7: Suppose that Assumptions 9, 10 and equation (37) hold. Then, the moment conditions  $g \in \{Db : b \in B\}^{\perp}$ . If  $\mathbb{E}[\epsilon^2|Z] = \mathbb{E}[\epsilon^2]$  a.s., then  $g = \tilde{\ell}$  a.s..

PROPOSITION 7: Suppose that Assumptions 9, 10 and equation (37) hold. Then Assumption 2 is satisfied with  $g_n = \mathbb{G}_n g$ .

A feasible test Suppose that  $\hat{\beta}_n$  and  $\hat{\pi}_{n,i}(Z_i)$  are estimators of  $\beta$  and  $\pi(Z_i)$  respectively. Let the *i*-th residual in (31) based on  $\theta = \theta_0$  and  $\hat{\beta}_n$  be  $\hat{\epsilon}_{n,i} := Y_i - X'_i \theta - Z'_{1,i} \hat{\beta}_n$ . Let  $\hat{s}_n := \frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_{n,i}^2$  and define

$$\hat{g}_{n,\theta,i} \coloneqq \hat{s}_n^{-1} \hat{\epsilon}_{n,i} \left[ \hat{\pi}_n(Z_i) - \left[ \frac{1}{n} \sum_{i=1}^n X_i Z_{1,i}' \right] \left[ \frac{1}{n} \sum_{i=1}^n Z_{1,i} Z_{1,i}' \right]^{-1} Z_i \right], \quad (40)$$

and

$$\check{V}_{n,\theta} \coloneqq \frac{1}{n} \sum_{i=1}^{n} \hat{g}_{n,\theta,i} \hat{g}'_{n,\theta,i}.$$
(41)

Based on  $\check{V}_{n,\theta}$ , form  $\hat{V}_{n,\theta}$  according to the construction in Section S5 of Lee and Mesters (2024b) using a truncation rate  $\mathbf{v}_n$ , set  $\hat{\Lambda}_{n,\theta} \coloneqq \hat{V}_{n,\theta}^{\dagger}$  and  $\hat{r}_{n,\theta} \coloneqq \operatorname{rank}(\hat{V}_{n,\theta})$ .

The following assumption provides sufficient high-level conditions on the estimators  $\hat{\beta}_n$  and  $\hat{\pi}_{n,i}(Z_i)$  such that Assumption 3 holds. These conditions are compatible with  $\hat{\pi}_{n,i}$  being a leave-one-out series estimator.<sup>38, 39</sup>

Assumption 11: Suppose that, given  $\theta_0$ , (i)  $\hat{\beta}_n$  is an estimator valued in  $\mathscr{S}_n \coloneqq$ 

<sup>&</sup>lt;sup>37</sup>Nevertheless, homoskedasticity is *not* assumed and the results below hold under heteroskedasticity. For full efficiency one could base the test on (38). This is left for future work.

<sup>&</sup>lt;sup>38</sup>The discretisation of  $\hat{\beta}_n$  is a technical device which permits the proof to go through under weaker conditions (cf. Le Cam and Yang, 2000, Chapter 6). This can be arranged given a  $\sqrt{n}$ – consistent initial estimator, by replacing its value with the closest point in the set  $\mathscr{S}_n$ .

<sup>&</sup>lt;sup>39</sup>See e.g. Belloni et al. (2015) for sufficient conditions for (42) and Section S4.2 for a discussion of (43).

 $\{CZ/\sqrt{n}: Z \in \mathbb{Z}^{d_{\beta}}\}\$  for some  $C \in \mathbb{R}^{d_{\beta} \times d_{\beta}}\$  and satisfying  $\sqrt{n}(\hat{\beta}_{n} - \beta) = O_{P_{n,0}}(1)$ and (ii)  $\hat{\pi}_{n,i}(Z_{i})$  are estimators such that  $\hat{\pi}_{n,i}(Z_{i})$  is  $\sigma(Z_{i}, \mathcal{C}_{n,-i})$  measurable for  $\mathcal{C}_{n,-i} \coloneqq \{W_{j}: j = 1, \ldots, n, j \neq i\}$ , and on events  $F_{n}$  with  $P_{n,0}(F_{n}) \to 1$ ,

$$\left[\int \|\hat{\pi}_{n,i}(z) - \pi(z)\|^2 \, \mathrm{d}\zeta_Z(z)\right]^{1/2} \le \delta_n = o(1),\tag{42}$$

where  $\zeta_Z$  is the marginal distribution of Z and for each  $k = 1, \ldots, d_{\theta}, i \neq j$ ,

$$\mathbb{E}\left[\mathbf{1}_{F_n}\mathbf{1}_{G_n}(\hat{\pi}_{n,i,k}(Z_i) - \pi_k(Z_i))(\hat{\pi}_{n,j,k}(Z_j) - \pi_k(Z_j))'\epsilon_i\epsilon_j\right] \lesssim \delta_n^2/n, \ P_{n,0}(G_n) \to 1.$$
(43)

Suppose also  $\delta_n^2 + n^{-1/2} = o(\mathbf{v}_n)$ , (37) holds and  $\mathbb{E}\left[\epsilon^4 (\|\pi(Z)\| + \|Z_1\|)^4\right] < \infty$ .

There is no requirement on the rate  $\delta_n$  in (42), (43) beyond  $\delta_n = o(1)$ .

PROPOSITION 8: Suppose that Assumptions 9, 10, & 11 hold. Then Assumption 3 holds with  $\hat{g}_{n,\theta} \coloneqq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{g}_{n,\theta,i}, g_n \coloneqq \mathbb{G}_n g$  and  $\hat{\Lambda}_{n,\theta}$  defined below equation (41).

A consequence of Assumption 10 and Propositions 7 and 8 is that the test  $\psi_{n,\theta_0}$  formed as in (16) is locally regular by Theorem 1.

Simulation study I test  $H_0$ :  $\theta = \theta_0 = 0$  at a nominal level of 5%. Each study reports the results of 5000 monte carlo replications with a sample size of  $n \in \{200, 400, 600\}$ .<sup>40</sup> Two simulation designs are considered.

Design 1 is a bivariate, just identified design. Here  $d_{\theta} = 2$  and  $Z_2$  is drawn from a zero-mean multivariate normal distribution with covariance matrix  $\begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}$ . The error terms  $\epsilon, v$  are drawn from a zero-mean multivariate normal such that each has variance 1 and the covariances are  $\text{Cov}(\epsilon, v_i) = 0.9$  and  $\text{Cov}(v_1, v_2) = 0.7$ .  $Z_1 = 1$  with  $\beta = 1$  and  $\pi(Z) = \pi(Z_2) = (\pi_1(Z_{2,1}), \pi_2(Z_{2,2}))'$  with each  $\pi_i$  (i = 1, 2)being one of the exponential or logistic functions  $f_j$  in Table 1. The exponential form is a prototypical function shape for which the linear projection of X on Z will provide essentially no identifying information; for the logistic form this linear projection should perform well.<sup>41, 42</sup>

I consider the  $\psi_{n,\theta_0}$  test developed above, with a leave-one-out series estimator of  $\pi$  based on (tensor product) Legendre polynomials. I consider both fixing the

<sup>&</sup>lt;sup>40</sup>The power surfaces in Design 1 are computed with 2500 replications.

<sup>&</sup>lt;sup>41</sup>These functions are plotted in Figures S1 and S2. The separation  $\pi(Z_2) = (\pi_1(Z_{2,1}), \pi_2(Z_{2,2}))'$  is assumed unknown and is not imposed in the estimation of  $\pi$ .

<sup>&</sup>lt;sup>42</sup>Results for the case where  $\pi(Z_2)$  is linear are very similar to the "approximately linear" logistic case and are therefore unreported.

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.  $\nu$  is set to  $10^{-2}$ . I additionally consider the Anderson and Rubin (1949) (AR) test.<sup>43,44</sup>

The empirical rejection frequencies under the null are shown in Tables 3 and 4. The parameter j controls the level of identification: the larger is j the closer  $\pi_i$  is to a constant function. In each specification all the considered tests reject at close to the nominal level. Power surfaces for the  $\psi_{n,\theta_0}$  and AR tests are shown in figures 2 – 7. As can be seen in these figures, the  $\psi_{n,\theta_0}$  test is able to detect deviations from the null when  $\pi_j$  has the exponential form, unlike the AR test.<sup>45</sup> For the logistic form, the power of the two tests is similar. Unsurprisingly, neither test provides non-trivial power when identification is very weak.

т. . . . -. .

Table 3: Empirical rejection frequencies, IV, Design 1

		Expo	Exponential - Exponential			Logistic - Logistic			Exponential - Logistic				
		AR		$\psi$		AR		$\psi$		AR		$\psi$	
n	j		k = 3	AIC	BIC		k=3	AIC	BIC		k = 3	AIC	BIC
200	1	5.52	5.10	4.00	5.18	5.52	4.88	3.40	4.48	5.52	4.56	3.76	4.68
200	2	5.52	6.24	6.26	6.24	5.52	5.46	5.30	5.46	5.52	5.78	5.86	5.78
200	3	5.52	8.36	7.80	8.36	5.52	8.44	7.98	8.44	5.52	7.92	7.74	7.92
400	1	5.60	4.96	4.48	4.80	5.60	4.96	4.14	4.78	5.60	4.88	4.34	4.80
400	2	5.60	6.12	6.20	6.12	5.60	5.68	5.52	5.68	5.60	6.08	6.02	6.08
400	3	5.60	6.76	6.94	6.76	5.60	7.72	7.60	7.72	5.60	6.46	6.60	6.46
600	1	5.38	5.30	4.90	5.22	5.38	5.20	3.80	5.16	5.38	5.32	4.64	5.04
600	2	5.38	6.08	6.10	6.08	5.38	4.98	4.98	4.98	5.38	5.56	5.76	5.56
600	3	5.38	4.36	4.78	4.36	5.38	5.02	5.50	5.02	5.38	4.40	5.06	4.40

*Notes:* E.g. "Exponential - Logistic" indicates that  $\pi_1, \pi_2$  have the exponential and logistic form in Table 1 respectively, with  $c_i$  corresponding to column j.

Design 2 is a univariate, over identified model with heteroskedastic errors.  $Z_1$ ,  $\beta$  and  $Z_2$  are as in Design 1, and  $\pi(Z_2) = (\pi_1(Z_{2,1}) + \pi_2(Z_{2,2}))/2$  where the  $\pi_i$ have one of the exponential or logistic forms of Table 1.<sup>46</sup> I draw  $(\tilde{\epsilon}, \tilde{v})$  from a zero-mean multivariate normal distribution with unit variances, covariance 0.95 and set  $(\epsilon, \upsilon)' = \begin{bmatrix} \sqrt{1 + \sin(Z_{2,1})^2} & 0\\ 0 & \sqrt{1 + \cos(Z_{2,2})^2} \end{bmatrix} (\tilde{\epsilon}, \tilde{\upsilon})'.$ 

<sup>&</sup>lt;sup>43</sup>The AR test is computed with  $Z_2$  as instruments, after partialling out  $Z_1$ .

<sup>&</sup>lt;sup>44</sup>I do not consider alternative weak instrument robust tests based on a linear first stage (e.g. LM, CLR) in this design as the AR test is known to be optimal when the model is just-identified.

<sup>&</sup>lt;sup>45</sup>One could consider an AR test using e.g. some basis functions  $f_1(Z_2), \ldots, f_K(Z_2)$  however as noted in (Mikusheva and Sun, 2022, p. 2669), the AR statistic is not well behaved for large K. The jackknife AR test of Mikusheva and Sun (2022) applies only to the case where  $d_{\theta} = 1$ . <sup>46</sup>This functional form is treated as unknown and not imposed in the estimation of  $\pi$ .

		Exponential - Exponential			Logistic - Logistic				Exponential - Logistic				
		AR		$\psi$		AR		$\psi$		AR		$\psi$	
n	$j_1 - j_2$		k = 3	AIC	BIC		k = 3	AIC	BIC		k = 3	AIC	BIC
200	1 - 3	5.52	6.36	5.30	6.48	5.52	7.18	6.40	7.12	5.52	6.34	5.40	6.30
200	2 - 3	5.52	6.26	6.44	6.26	5.52	6.94	6.78	6.94	5.52	6.46	6.64	6.46
200	3 - 3	5.52	8.36	7.80	8.36	5.52	8.44	7.98	8.44	5.52	7.92	7.74	7.92
400	1 - 3	5.60	5.70	5.38	5.68	5.60	6.24	5.82	6.20	5.60	6.14	5.54	5.98
400	2 - 3	5.60	6.22	6.48	6.22	5.60	6.24	6.40	6.24	5.60	6.74	6.90	6.74
400	3 - 3	5.60	6.76	6.94	6.76	5.60	7.72	7.60	7.72	5.60	6.46	6.60	6.46
600	1 - 3	5.38	5.66	5.40	5.60	5.38	5.46	4.52	5.34	5.38	5.58	5.26	5.36
600	2 - 3	5.38	6.00	6.10	6.00	5.38	5.40	5.40	5.40	5.38	6.12	6.32	6.12
600	3 - 3	5.38	4.36	4.78	4.36	5.38	5.02	5.50	5.02	5.38	4.40	5.06	4.40

Table 4: Empirical rejection frequencies, IV, Design 1

*Notes:* E.g. "Exponential - Logistic" indicates that  $\pi_1, \pi_2$  have the exponential and logistic form in Table 1 respectively, with  $c_{j_1}$  and  $c_{j_2}$  corresponding to column  $j_1 - j_2$ .

Figure 2: IV Design 1, AR power,  $\pi_i$  exponential







Figure 4: IV Design 1, AR power,  $\pi_i$  logistic



Figure 5: IV Design 1,  $\psi~(k=3)$  power,  $\pi_i$  logistic



Figure 6: IV Design 1, AR power,  $\pi_1$  exponential,  $\pi_2$  logistic





Figure 7: IV Design 1,  $\psi$  (k = 3) power,  $\pi_1$  exponential,  $\pi_2$  logistic

The  $\psi$  tests are computed in the same manner as in Design 1. I also compute the AR, LM and CLR tests based on  $Z_2$  (with  $Z_1$  partialled out) as well as the many weak instrument robust jackknife AR test of Mikusheva and Sun (2022). MS<sub>1</sub> uses  $Z_2$  as instruments; MS<sub>2</sub> uses the (tensor product) of Legendre polynomials used to estimate  $\pi$  as instruments.

The empirical rejection frequencies under the null are shown in Tables 5 – 7. As in Design 1, the parameter j controls the level of identification: the larger is j the closer  $\pi_j$  is to a constant function and hence  $\theta$  unidentified. In each specification all the considered tests reject close to the nominal level; the MS<sub>2</sub> test is somewhat oversized for smaller n. The power of these tests is plotted in Figures 8 – 10; the  $\psi_{n,\theta_0}$  tests are denoted by k = 3, AIC and BIC, corresponding to how  $\pi$  is estimated. For the design with both  $\pi_i$  exponential, the  $\psi_{n,\theta_0}$  test clearly delivers the highest power whenever there is non-trivial power available; of the other tests, only MS<sub>2</sub> delivers non-trivial power in this specification. For the case with both  $\pi_i$  logistic, all tests except MS<sub>2</sub> perform similarly, with MS<sub>2</sub> offering lower power. The same holds for the final specification, where  $\pi_1$  is exponential and  $\pi_2$  logistic.

### 6 Empirical applications

In this section I re-analyse two IV studies with potentially weak instruments by inverting the  $\psi_{n,\theta_0}$  test developed in section 5.2 to construct weak-instrument robust confidence intervals (CIs). In each case the  $\psi_{n,\theta_0}$  test is able to exploit non-linearities to yield substantial reductions in CI length relative to AR CIs.

		AR	LM	CLR	$MS_1$	$MS_2$		$\psi$	
n	j						k = 3	AIC	BIC
$200 \\ 200 \\ 200 \\ 200$	$\begin{array}{c} 1 \\ 2 \\ 3 \end{array}$	$5.68 \\ 5.68 \\ 5.68$	$5.52 \\ 6.08 \\ 6.46$	$5.94 \\ 5.88 \\ 6.16$	$7.56 \\ 7.56 \\ 7.56$	9.24 9.24 9.24	$6.36 \\ 7.62 \\ 7.36$	$5.74 \\ 7.62 \\ 7.44$	$6.36 \\ 7.62 \\ 7.36$
$400 \\ 400 \\ 400$	$\begin{array}{c} 1 \\ 2 \\ 3 \end{array}$	$5.28 \\ 5.28 \\ 5.28$	$5.48 \\ 5.72 \\ 6.06$	$5.26 \\ 5.48 \\ 5.84$	7.18 7.18 7.18	8.78 8.78 8.78	$6.32 \\ 7.98 \\ 3.42$	$5.38 \\ 7.98 \\ 3.68$	$6.36 \\ 7.98 \\ 3.42$
600 600 600	$\begin{array}{c} 1 \\ 2 \\ 3 \end{array}$	$5.86 \\ 5.86 \\ 5.86$	$5.48 \\ 5.50 \\ 5.76$	$\begin{array}{c} 6.04 \\ 5.86 \\ 6.30 \end{array}$	$7.58 \\ 7.58 \\ 7.58 \\ 7.58$	7.92 7.92 7.92	$5.16 \\ 6.28 \\ 1.32$	$5.16 \\ 6.20 \\ 1.74$	$5.14 \\ 6.28 \\ 1.32$

Table 5: Empirical rejection frequencies, IV, Design 2, Exponential - Exponential

*Notes:* The functions  $\pi_i$  have the exponential form in Table 1 with  $c_j$  corresponding to column j.

Table 6: Empirical rejection frequencies, IV, Design 2, Logistic - Logistic

		AR	LM	CLR	$MS_1$	$MS_2$		$\psi$	
n	j						k = 3	AIC	BIC
$200 \\ 200 \\ 200 \\ 200$	$\begin{array}{c} 1\\ 2\\ 3\end{array}$	$5.68 \\ 5.68 \\ 5.68$	$5.78 \\ 5.94 \\ 5.82$	$6.94 \\ 7.22 \\ 6.86$	$7.56 \\ 7.56 \\ 7.56$	$9.24 \\ 9.24 \\ 9.24$	$4.82 \\ 5.76 \\ 7.68$	$4.46 \\ 5.74 \\ 7.78$	$4.80 \\ 5.76 \\ 7.68$
$400 \\ 400 \\ 400$	$\begin{array}{c} 1 \\ 2 \\ 3 \end{array}$	$5.28 \\ 5.28 \\ 5.28$	$4.76 \\ 4.80 \\ 4.94$	$5.98 \\ 6.26 \\ 5.80$	$7.18 \\ 7.18 \\ 7.18$	8.78 8.78 8.78	$\begin{array}{c} 4.32 \\ 4.90 \\ 3.18 \end{array}$	$4.04 \\ 4.98 \\ 3.52$	$\begin{array}{c} 4.18 \\ 4.90 \\ 3.18 \end{array}$
600 600 600	$\begin{array}{c} 1 \\ 2 \\ 3 \end{array}$	$5.86 \\ 5.86 \\ 5.86$	$5.12 \\ 5.20 \\ 5.14$	$6.42 \\ 6.32 \\ 6.28$	$7.58 \\ 7.58 \\ 7.58 \\ 7.58$	7.92 7.92 7.92	$4.84 \\ 5.00 \\ 1.56$	$4.50 \\ 4.90 \\ 1.84$	$4.66 \\ 5.00 \\ 1.56$

*Notes:* The functions  $\pi_i$  have the logistic form in Table 1 with  $c_j$  corresponding to column j.



Figure 8: IV design 2, Power curves, exponential  $\pi_i$ 

		AR	LM	CLR	$MS_1$	$MS_2$		$\psi$	
n	j						k = 3	AIC	BIC
200	1	5.68	6.14	7.12	7.56	9.24	5.28	4.62	5.30
200	2	5.68	5.92	7.18	7.56	9.24	6.90	6.90	6.90
200	3	5.68	5.90	6.60	7.56	9.24	7.16	7.26	7.16
400	1	5.28	4.94	6.44	7.18	8.78	4.74	4.28	4.70
400	2	5.28	5.24	6.54	7.18	8.78	5.52	5.64	5.52
400	3	5.28	5.00	5.66	7.18	8.78	3.10	3.42	3.10
600	1	5.86	5.20	6.30	7.58	7.92	4.86	4.56	5.00
600	2	5.86	5.14	6.48	7.58	7.92	5.62	5.50	5.62
600	3	5.86	5.22	6.24	7.58	7.92	1.20	1.58	1.20

Table 7: Empirical rejection frequencies, IV, Design 2, Exponential - Logistic

Notes:  $\pi_1$ ,  $\pi_2$  have the exponential and logistic form in Table 1 respectively with  $c_{j_i}$  corresponding to column j.





Figure 10: IV Design 2, Power curves, exponential  $\pi_1$ , logistic  $\pi_2$ 



#### 6.1 The effect of skilled immigration on productivity

Hornung (2014) studies the long term effect of skilled immigration on productivity using a natural experiment in which the skilled but religiously persecuted French protestants (Hugenots) fled and settled in Prussia.<sup>47</sup> In the notation of Example 2, Y is log output in textile manufacturing, X is the proportion of Hugenots in each town and  $Z_1$  contains various control variables, see Hornung (2014) for details. Hornung (2014) argues that "by the order of centralized ruling by the king and his agents Huguenots were channeled into Prussian towns in order to compensate for severe population losses during the Thirty Years' War", motivating the instrument  $Z_2$ : the percentage population losses during the war. In particular, three different measurements of this population loss are used and refered to as specifications (1), (2) and (3) hereafter.<sup>48</sup>

As noted in Hornung (2014), this instrument may be weak: in each case the first stage F statistic is "small".<sup>49</sup> I implement the  $\psi_{n,\theta_0}$  test using a leave-one-out series estimator of  $\pi$  of the form

$$\hat{\pi}_i(Z_i) = \hat{\pi}'_i p_K(Z_{2,i}) + \hat{\beta}_i Z_{1,i}, \tag{44}$$

where the *i* subscript on the estimated coefficients indicates they have been estimated on all observations except for the *i*-th.  $p_K$  is a vector of a constant and the first *K* Legendre polynomials. I choose  $K \in \{1, 2, 3, 4\}$  and whether to include  $Z_1$ in the model for  $\pi$  by using BIC: all specifications include  $Z_1$  and K = 4.

Table 8 reports 2SLS estimates of  $\theta$  along with 2SLS (Wald) CIs, AR CIs and CIs found by inverting the  $\psi_{n,\theta_0}$  test.<sup>50</sup> The resulting CIs provide a similar interpretation as that based on the AR CIs: the effect of (skilled) Hugenot immigration was positive on textile output. However, the  $\psi_{n,\theta_0}$  based CIs are smaller than the AR CIs, achieving approximately a 40% - 50% reduction in length.

#### 6.2 The effect of racial segregation on inequality

Ananat (2011) estimates the effect of racial segregation (X) on poverty and inequality (Y, measured respectively by the poverty rate and log gini coefficient for black / white city residents), instrumenting segregation by a "railroad division

<sup>&</sup>lt;sup>47</sup>That the Hugenots were more skilled (on average) than the Prussian population seems to be broadly accepted cf. pp. 85-86, 93-95 in Hornung (2014).

<sup>&</sup>lt;sup>48</sup>Specifications (1) & (2) are those considered in the left and and right hand parts of Table 4 of Hornung (2014); Specification (3) is that considered in the left hand part of Table 5.

 <sup>&</sup>lt;sup>49</sup>It is less than the cutoff of 10 suggested by Staiger and Stock (1997) for homoskedastic IV.
 <sup>50</sup>The inversion is performed over a grid of 5000 equally spaced points from -1 to 7.

	(1)	(2)	(3)						
n	150	150	186						
F	3.668	4.791	5.736						
Point estimate									
2SLS	3.475	3.38	1.671						
Confidence intervals									
2SLS	[1.27,  5.68]	[1.294,  5.467]	[0.032,  3.31]						
$\mathbf{AR}$	[1.427,  6.303]	[1.43,  5.985]	[-0.022,  3.379]						
$\psi$	[1.626,  4.099]	[1.637,  4.073]	[1.136, 3.228]						
Relative length of confidence intervals to AR									
2SLS	0.904	0.916	0.964						
$\psi$	0.507	0.535	0.615						

 Table 8:
 Point estimates and confidence intervals

Notes: F is the first stage F statistic. All confidence intervals have nominal coverage of 95%.

index" (RDI,  $Z_2$ ), "a variation on a Herfindahl index that measures the dispersion of a city's land into subunits" via the layout of railroad tracks.<sup>51</sup> The first and second stages also include an intercept and control for railroad track length ( $Z_1$ ).

The instrument may be weak: I calculate the first stage F statistic to be 2.307.<sup>52</sup> I implement the  $\psi_{n,\theta_0}$  test using a leave-one-out series estimator of  $\pi$  of the form (44). I choose  $K \in \{1, 2, 3, 4\}$  and whether to include  $Z_1$  in the model for  $\pi$  by using BIC: this excludes  $Z_1$  and chooses K = 2.

Table 9 reports 2SLS estimates of  $\theta$  along with 2SLS (Wald) CIs, AR CIs and CIs found by inverting the  $\psi_{n,\theta_0}$  test.<sup>53</sup> The resulting CIs provide a similar interpretation as that based on the AR CIs: racial segregation increases poverty and inequality within the Black community and decreases poverty and inequality within the White community. The  $\psi_{n,\theta_0}$  based CIs are shorter than the AR CIs, achieving a reduction in length varying from 5% to around 38%.

<sup>&</sup>lt;sup>51</sup>Section 3 and Appendix A of Ananat (2011) provides evidence that the choice of railroad placement was not related to local social or economic concerns.

<sup>&</sup>lt;sup>52</sup>Ananat (2011) refers to Column 1 of Table 1 when discussing the first stage F statistic. The values in this table imply a first stage F of  $(0.357/0.088)^2 \approx 16.458$ . This "discrepancy" arises from different default choices of robust covariance estimate in R's sandwich package (HC3; my calculation) and STATA's robust command (HC1; Ananat, 2011).

 $<sup>^{53}\</sup>mathrm{The}$  inversion is performed over a grid of 5000 equally spaced points from -1 to 2.

	Pover	ty rate	Gini coefficient						
	White	Black	White	Black					
Point estimate									
2SLS	0.258	-0.196	0.875	-0.334					
Confidence intervals									
2SLS	[-0.026, 0.543]	[-0.334, -0.058]	[0.277,  1.474]	[-0.558, -0.111]					
AR	[-0.04,  0.598]	[-0.394, -0.075]	[0.319,  1.684]	[-0.674, -0.149]					
$\psi$	[0.072,  0.568]	[-0.376, -0.097]	[0.29,  1.138]	[-0.608, -0.106]					
Relative length of confidence intervals to AR									
2SLS	0.891	0.866	0.877	0.853					
$\psi$	0.775	0.877	0.622	0.955					

Table 9: Point estimates and confidence intervals

Notes: All confidence intervals have nominal coverage of 95%.

# 7 Conclusion

In this paper I establish that  $C(\alpha)$ -style tests are locally regular under mild conditions, including in non-regular cases where locally regular estimators do not exist. As a consequence, these tests do not overreject under semiparametric weak identification asymptotics. Additionally I generalise the classical local asymptotic power bounds for LAN models to the case where the efficient information matrix has positive, but potentially deficient, rank, such that these results also apply in cases of underidentification (or weak underidentification). I show that, if the  $C(\alpha)$ test is based on the efficient score function, it attains these power bounds. This (attainment) result improves on results known in the literature in two ways: (i) it applies also to non-regular models and (ii) it does not require the data to be i.i.d. nor the information operator to be boundedly invertible. A simulation study based on two examples shows that the asymptotic theory provides an accurate approximation to the finite sample performance of the proposed tests.

# References

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

Ananat, E. O. (2011), "The Wrong Side(s) of the Tracks: The Causal Effects

of Racial Segregation on Urban Poverty and Inequality," *American Economic Journal: Applied Economics*, 3, 34–66.

- 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. K. (1987), "Asymptotic Results for Generalized Wald Tests," *Econometric Theory*, 3, 348–358.
- 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.
- (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," ArXiv:econ.EM/2204.12462.
- 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.

- Chamberlain, G. (1986), "Asymptotic efficiency in semi-parametric models with censoring," *Journal of Econometrics*, 32, 189–218.
- (1992), "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 567–596.
- 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.
- Choi, S., Hall, W. J., and Schick, A. (1996), "Asymptotically uniformly most powerful tests in parametric and semiparametric models," Ann. Statist., 24, 841–861.
- 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.
- 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*, 15, 523–570.

- Hornung, E. (2014), "Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia," The American Economic Review, 104, 84–122.
- Ichimura, H. (1993), "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models," *Journal of Econometrics*, 58, 71–120.
- Janson, S. (1997), Gaussian Hilbert Spaces, Cambridge University Press.
- Kaji, T. (2021), "Theory of Weak Identification in Semiparametric Models," *Econometrica*, 89, 733–763.
- Kleibergen, F. (2005), "Testing Parameters in GMM Without Assuming that They Are Identified," *Econometrica*, 73, 1103–1123.
- Le Cam, L. M. (1986), Asymptotic Methods in Statistical Decision Theory, New York, NY, USA: Springer.
- Le Cam, L. M. and Yang, G. L. (2000), Asymptotics in Statistics: Some Basic Concepts, New York, NY, USA: Springer, 2nd ed.
- Lee, A. and Mesters, G. (2024a), "Locally Robust Inference for Non-Gaussian Linear Simultaneous Equations Models," *Journal of Econometrics*, 240, 105647.
- (2024b), "Supplement to 'Locally Robust Inference for Non-Gaussian Linear Simultaneous Equations Models'," *Journal of Econometrics Supplementary Material*, 240.
- McCloskey, A. (2017), "Bonferroni-based size-correction for nonstandard testing problems," *Journal of Econometrics*, 200, 17–35.
- Mikusheva, A. and Sun, L. (2022), "Inference with Many Weak Instruments," *The Review of Economic Studies*, 89, 2663–2686.
- Moreira, M. J. (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 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.
- Ritov, Y. and Bickel, P. J. (1990), "Achieving Information Bounds in Non and Semiparametric Models," *The Annals of Statistics*, 18, 925 – 938.
- 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, W. de Gruyter.
- van der Vaart, A. W. (1991), "An Asymptotic Representation Theorem," International Statistical Review / Revue Internationale de Statistique, 59, 97–121.
- (1998), Asymptotic Statistics, New York, NY, USA: Cambridge University Press.
- 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 1. Combination of Assumptions 1 and 2 yields

$$(g'_n, L_n(h))' \xrightarrow{P_n} \mathcal{N}\left(\begin{pmatrix} 0\\ -\frac{1}{2}\sigma(h) \end{pmatrix}, \begin{pmatrix} V & \tau'\Sigma'_{21}\\ \Sigma_{21}\tau & \sigma(h) \end{pmatrix}\right).$$

By Le Cam's third Lemma  $g_n \stackrel{P_{n,h}}{\leadsto} Z_{\tau} \sim \mathcal{N}(\Sigma_{21}\tau, V)$ . The second claim follows from the first with Assumption 3(i), Remark 1 and Slutsky's Theorem. By the second claim, Assumption 3(ii) and standard arguments,  $\hat{S}_{n,\theta_0} \stackrel{P_{n,h}}{\leadsto} Z'_{\tau} V^{\dagger} Z_{\tau}$ .  $\Box$ 

Proof of Theorem 1. If  $r \geq 1$ , as  $\hat{r}_n \xrightarrow{P_n} r$ ,  $P_n\{c_n = c_\alpha\} \to 1$ . By Proposition 1, Remark 1 and Slutsky's Theorem,  $\hat{S}_{n,\theta_0} - c_n \rightsquigarrow S - c$  under  $P_{n,h}$  where  $S \sim \chi_r^2(a)$ . By the Portmanteau Theorem,

$$\lim_{n \to \infty} P_{n,h} \psi_{n,\theta_0} = \lim_{n \to \infty} P_{n,h} \left( \hat{S}_{n,\theta_0} > c_n \right) = L\{S - c > 0\} = 1 - \mathcal{P}(\chi_r^2(a) \le c_\alpha),$$

for *L* the law of *S*. If r = 0, rank $(\hat{\Lambda}_{n,\theta_0}) \xrightarrow{P_n} 0 \implies P_n R_n \to 1$  for  $R_n \coloneqq \{\hat{\Lambda}_{n,\theta_0} = 0\}$ . On  $R_n \hat{S}_{n,\theta_0} = 0 \implies \psi_{n,\theta_0} = 0$ . By Remark 1,  $P_{n,h}\psi_{n,\theta_0} \le 1 - P_{n,h}R_n \to 0$ .

Proof of Corollary 1. By Theorem 1,  $\pi_n(h) \to \pi(h)$   $(h \in H)$ , pointwise. Since the  $\pi_n$  are asymptotically equicontinuous on K, the convergence is uniform on K.  $\Box$ 

Proof of Lemma 1. Immediate from  $|P_{n,h}\psi_{n,\theta_0} - P_{n,h'}\psi_{n,\theta_0}| \le d_{TV}(P_{n,h}, P_{n,h'})$ .  $\Box$ 

Proof of Lemma 2. For  $a, b \in \mathbb{R}$ ,  $h_1, h_2 \in H$ ,  $\Delta_n(a_1h_1 + a_2h_2) = a_1\Delta_nh_1 + a_2\Delta_nh_2$ and so  $\Delta(a_1h_1 + a_2h_2) = a_1\Delta(h_1) + a_2\Delta(h_2)$ , hence  $\Delta$  is linear. We now establish K is a well-defined covariance kernel. For  $h \in H$ ,  $(\|\Delta_nh\|^2)_{n \in \mathbb{N}}$  is Cauchy. Letting  $K_n(h,g) \coloneqq P_n[\Delta_nh\Delta_ng]$  and using Cauchy – Schwarz

$$|K_n(h,g) - K_m(h,g)| \le ||\Delta_n h - \Delta_m h|| ||\Delta_n g|| + ||\Delta_m h|| ||\Delta_n g - \Delta_m g||,$$

hence  $(K_n(h,g))_{n\in\mathbb{N}}$  is also Cauchy and thus has a limit. Bilinearity and symmetry are straightforward to check. For positive semi-definiteness, let  $h_1, \ldots, h_K \in H$ ,  $a \in \mathbb{R}^K$ . As  $\Delta_n h \in L_2^0(P_n)$ ,  $\mathcal{K}_n \coloneqq [K_n(h_k, h_j)]_{k,j=1}^K$  is a covariance matrix, hence  $\sum_{k=1}^K \sum_{j=1}^K a_k a_j K_n(h_k, h_j) = a' \mathcal{K}_n a \ge 0$  for each  $n \in \mathbb{N}$  and hence the same holds with  $K_n$  and  $\mathcal{K}_n$  replaced by K and  $\mathcal{K} \coloneqq [K(h_k, h_j)]_{k,j=1}^K$ .

By Assumption 1 and the fact that  $K(h, h) = \sigma(h)$ ,  $\Delta h \sim \mathcal{N}(0, K(h, h))$ . That  $\Delta$  is a mean-zero Gaussian process with covariance kernel K then follows from the

Cramér – Wold Theorem as  $\sum_{k=1}^{K} a_k \Delta h_k \sim \mathcal{N}(0, a' \mathcal{K} a)$  and

$$\sum_{k=1}^{K} a_k \Delta_n(h_k) = \Delta_n \left( \sum_{k=1}^{K} a_k h_k \right) \stackrel{P_n}{\rightsquigarrow} \Delta \left( \sum_{k=1}^{K} a_k h_k \right) = \sum_{k=1}^{K} a_k \Delta h_k.$$

Proof of Proposition 2. Remark 1 and the transitivity of (mutual) contiguity ensures that the experiments  $\mathscr{E}_n$  are contiguous. By Theorem 61.6 of Strasser (1985) it suffices to show that the finite dimensional marginal distributions (fdds) of  $L_n$ converge (under  $P_n$ ) to those of L, where  $L(h) \coloneqq \Delta h - \frac{1}{2} ||h||^2$ . This follows as the fdds of  $\Delta_n$  converge to those of  $\Delta$  (under  $P_n$ ), by the Cramér – Wold Theorem.  $\Box$ 

Proof of Proposition 3. Let  $G_{[0]} \coloneqq P_0$ . Define  $Z : \mathbb{H} \to L_2(\Omega, \mathcal{F}, G_{[0]})$  by  $Z[h] \coloneqq \Delta(h)$  for an arbitrary  $h \in \pi_V^{-1}([h])$ , where  $\pi_V$  is the quotient map from  $H \to \mathbb{H}$ . This is a standard Gaussian process for  $\mathbb{H}$ . Define  $G_{[h]}$  by  $\frac{\mathrm{d}G_{[h]}}{\mathrm{d}G_{[0]}} = \exp\left(Z[h] - \frac{1}{2}\|[h]\|_K^2\right)$ .  $\mathscr{G}$  is a Gaussian shift on  $(\mathbb{H}, \langle \cdot, \cdot \rangle_K)$  (Strasser, 1985, Theorem 69.4). For any  $h \in H$  we have that  $Z[h] = \Delta g$  for some  $g \in \pi_V^{-1}([h])$  and  $\Delta h = \Delta g P_0^{-1}$ almost surely. Since  $\|h\|_K = \|[h]\|_K$ ,  $P_0$ -a.s.,  $\frac{\mathrm{d}G_{[h]}}{\mathrm{d}G_{[0]}} = \exp\left(Z[h] - \frac{1}{2}\|[h]\|_K^2\right) = \exp\left(\Delta h - \frac{1}{2}\|h\|_K^2\right) = \frac{\mathrm{d}P_h}{\mathrm{d}P_0}$ . As each  $P_h \ll P_0$  and  $G_{[h]} \ll G_{[0]}$ , and  $P_0 = G_{[0]}$ ,  $d_{TV}(P_h, G_{[h]}) = \frac{1}{2}\int \left|\frac{\mathrm{d}P_h}{\mathrm{d}P_0} - \frac{\mathrm{d}G_{[h]}}{\mathrm{d}P_0}\right| \,\mathrm{d}P_0 = \frac{1}{2}\int \left|\frac{\mathrm{d}P_h}{\mathrm{d}P_0} - \frac{\mathrm{d}G_{[h]}}{\mathrm{d}G_{[0]}}\right| \,\mathrm{d}P_0 = 0$ .

Proof of Lemma 3. By straightforward calculation

$$\langle [\tau, b], [t, g] \rangle_K = \tau' \tilde{\mathcal{I}} t + \langle \Pi[\tau, 0] + [0, b], \Pi[t, 0] + [0, g] \rangle_K$$
 (45)

This and  $\Pi[(\tau, 0)] \in \ker \pi_1 \text{ imply } \|[\tau]\|^2 = \inf_{b \in B} \|[\tau, b]\|_K^2 = \tau' \tilde{\mathcal{I}} \tau + \inf_{[h] \in \ker \pi_1} \|\Pi[\tau, 0] - [h]\|_K^2 = \tau' \tilde{\mathcal{I}}_{\gamma} \tau.$  Hence,  $\|\tau\| = \|[\tau]\| = 0 \implies \tau' \tilde{\mathcal{I}}_{\gamma} \tau = 0 \implies \tilde{\mathcal{I}}_{\gamma}^{1/2} \tau = 0$ , and so  $\tilde{\mathcal{I}}_{\gamma} \tau = 0.$  Conversely  $\tau \in \ker \tilde{\mathcal{I}}_{\gamma} \implies \|\tau\|^2 = 0 \implies \|\tau\|^2 = 0 \implies \|\tau\| = 0.$ 

Proof of Theorem 2. Define the bounded linear map  $T : \overline{\mathbb{H}} \to \mathbb{R}$  according to  $T[h] := \langle \Pi^{\perp}[1,0], \Pi^{\perp}[h] \rangle_{K} = \langle \Pi^{\perp}[1,0], [h] \rangle_{K}$ . For any  $[h] = [\tau, b] \in \mathbb{H}$ ,

 $T[h] = \left\langle \Pi^{\perp}[1,0], \, \Pi^{\perp}[\tau,b] \right\rangle_{K} = \tilde{\mathcal{I}}\tau.$ (46)

If  $\tilde{\mathcal{I}} = 0$ , (19) follows from Proposition 4, so assume  $\tilde{\mathcal{I}} \neq 0$ . Any unbiased level  $\alpha$  test  $\phi$  of T[h] = 0 against  $T[h] \neq 0$  in the (restricted) Gaussian shift  $\mathscr{G}$  satisfies

$$G_{[h]}\phi \le 1 - \Phi\left(z_{\alpha/2} - \tilde{\mathcal{I}}^{1/2}\tau\right) + 1 - \Phi\left(z_{\alpha/2} + \tilde{\mathcal{I}}^{1/2}\tau\right),\tag{47}$$

(Strasser, 1985, Lemma 71.5). By Proposition 2,  $\mathscr{E}_n \rightsquigarrow \mathscr{E}$ ;  $\mathscr{E}$  is dominated. Let  $\pi_n(h) \coloneqq P_{n,h}\phi_n$  and fix an arbitrary  $h^*$ . There is a subsequence  $(\pi_{n_m})_{m \in \mathbb{N}}$  such

that  $\lim_{m\to\infty} \pi_{n_m}(h^*) = \limsup_{n\to\infty} \pi_n(h^*)$ . Since  $[0,1]^H$  is compact in the product topology, there is a subnet  $(\pi_{n_{m(s)}})_{s\in S}$  and a function  $\pi: H \to [0,1]$  such that  $\lim_{s\in S} \pi_{n_{m(s)}}(h) = \pi(h)$  for all  $h \in H$ . By our hypotheses and equation (46) for any  $h_0$  such that  $[h_0] \in \ker T \cap \mathbb{H}$  and any  $h_1$  such that  $[h_1] \in \mathbb{H} \setminus (\ker T \cap \mathbb{H})$ 

$$\pi(h_0) = \lim_{s \in S} \pi_{n_{m(s)}}(h_0) \le \alpha \le \lim_{s \in S} \pi_{n_{m(s)}}(h_1) = \pi(h_1).$$
(48)

There exists a test  $\phi$  in  $\mathscr{E}$  with power function  $\pi$  (van der Vaart, 1991, Theorem 7.1). (48) and Proposition 3 ensure that  $\phi$  is an unbiased, level  $\alpha$  test of ker  $T \cap \mathbb{H}$  against  $\mathbb{H} \setminus (\ker T \cap \mathbb{H})$  in  $\mathscr{G}$ . Conclude by combining (47) and (by Proposition 3)

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

Proof of Corollary 2. By Theorem 1,  $\lim_{n\to\infty} P_{n,h_n}\psi_{n,\theta} = 1 - P(Z^2 > c_{\alpha})$  where  $Z \sim \mathcal{N}\left(\tilde{\mathcal{I}}^{1/2}\tau, 1\right)$ .  $1 - P(Z^2 > c_{\alpha})$  is equal to the RHS of (20).

Proof of Theorem 3. By Lemma 3,  $\mathbb{H}_1 = \mathbb{R}^{d_\theta} / \ker \tilde{\mathcal{I}}$ .  $\pi_1 : \mathbb{H} \to \mathbb{H}_1$  is surjective: for any  $[\tau] \in \mathbb{H}_1$  let  $t \in \pi_{\ker \tilde{\mathcal{I}}}^{-1}(\{[\tau]\})$  where  $\pi_{\ker \tilde{\mathcal{I}}}$  is the quotient map from  $\mathbb{R}^{d_\theta}$  to  $\mathbb{H}_1$ . Then  $\pi_1[t, 0] = [t] = [\tau]$ . It follows that dim ran  $\pi_1 = \operatorname{codim} \ker \pi_1 = r$ . By linearity and  $[0, b] \in \ker \pi_1$ ,  $\Pi[\tau, b] = \Pi[\tau, 0] + [0, b]$ . This with Lemma 3 yields  $\|[\tau, b] - \Pi[\tau, b]\|_K^2 = \|[\tau, 0] - \Pi[\tau, 0]\|_K^2 = \|[\tau]\|^2 = \tau' \tilde{\mathcal{I}} \tau$ . Define the sets

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

It is straightforward to check that  $\operatorname{cl} M_a = \overline{M}_a$ . Suppose that  $\phi$  is a test on  $\mathscr{G}$  with  $G_{[0]}\phi \leq \alpha$ . First suppose a > 0.  $\phi$  is a level  $\alpha$  test of  $K_0$ : {[0]} against  $K_1 : [\ker \pi_1]^{\perp} \setminus \{[0]\}$  in the restriction of the standard Gaussian shift experiment on  $[\ker \pi_1]^{\perp}$ . By Theorem 30.2 in Strasser (1985)

$$\inf_{[h]\in M_a} G_{[h]}\phi = \inf_{[h]\in \bar{M}_a} G_{[h]}\phi \le \inf_{[h]\in \bar{M}_a\cap[\ker\pi_1]^{\perp}} G_{[h]}\phi \le 1 - \mathcal{P}(\chi_r^2(a) \le c_r),$$
(49)

since  $[h] \mapsto G_{[h]}\phi$  is continuous. If, instead, a = 0, note that  $[0] \in M_0$  and so,

$$\inf_{[h]\in M_0} G_{[h]}\phi \le G_{[0]}\phi \le \alpha = 1 - \mathcal{P}(\chi_r^2(0) \le c_r).$$
(50)

Fix  $a \ge 0$  and let  $\mathcal{R} := 1 - \mathbb{P}(\chi_r^2(a) \le c_r)$ . Let  $\pi_n(h) := P_{n,h}\phi_n$  and define  $\beta_n := \inf \left\{ P_{n,h}\phi_n : h = (\tau, b) \in H, \ \tau' \tilde{\mathcal{I}}\tau = a \right\}$ . Suppose that  $\limsup_{n \to \infty} \beta_n \ge \mathcal{R} + \varepsilon$  for some  $\varepsilon > 0$ . Hence, for some subsequence  $(n_m)_{m \in \mathbb{N}}$ ,  $\lim_{m \to \infty} \beta_{n_m} \ge \mathcal{R} + \varepsilon$ . Since

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

$$\pi(h) = \lim_{s \in S} \pi_{n_{m(s)}}(h) \ge \liminf_{s \in S} \inf \left\{ \pi_{n_{m(s)}}(h) : h = (\tau, b) \in H, \ \tau' \tilde{\mathcal{I}} \tau = a \right\} \ge \mathcal{R} + \varepsilon.$$
(51)

By Proposition 2,  $\mathscr{E}_n \rightsquigarrow \mathscr{E}$ ;  $\mathscr{E}$  is dominated. There is a test  $\phi$  in  $\mathscr{E}$  with power function  $\pi$  (van der Vaart, 1991, Theorem 7.1). Consider the restriction of  $\mathscr{G}$  to  $[\ker \pi_1]^{\perp}$ . By hypothesis, Corollary S2 and Proposition 3  $G_{[0]}\phi = P_0\phi = \pi(0) =$  $\lim_{s \in S} \pi_{n_m(s)}(0) \leq \limsup \pi_n(0) \leq \alpha$ , hence  $\phi$  is a test of level  $\alpha$  of  $K_0$  against  $K_1$ in this experiment, and  $\inf_{[h] \in M_a} G_{[h]}\phi = \inf_{h:[h] \in M_a} P_h\phi = \inf_{h:[h] \in M_a} \pi(h) \geq \mathcal{R} + \varepsilon$ , by (51) and Proposition 3, but this contradicts (49) if a > 0 or (50) if a = 0.  $\Box$ 

Proof of Corollary 3. Equation (22) follows from Theorem 1. For equation (23), let  $f_n(h) \coloneqq P_{n,h}\psi_{n,\theta_0}$ . By (22) and the asymptotic equicontinuity,  $\lim_{n\to\infty} f_n(h) = 1 - P\left(\chi_r^2\left(\tau'\tilde{\mathcal{I}}\tau\right) \leq c_r\right) \eqqcolon f(h)$ , uniformly on  $K_a$ . Conclude that if  $h_n \to h \in K_a$ ,

$$\lim_{n \to \infty} f_n(h_n) = f(h) \ge f_\star \coloneqq 1 - \mathcal{P}\left(\chi_r^2(a) \le c_r\right).$$
(52)

If (23) fails there is a sequence  $h_n \in K_a$  with  $\limsup_{n\to\infty} f_n(h_n) < f_{\star}$ . Extract a subsequence  $h_{n_m} \to h \in K_a$ . Let  $h_m^* \coloneqq h_{n_1}$  for  $m = 1, \ldots, n_1$  and  $h_m^* \coloneqq h_{n_k}$  for  $n_k \leq m < n_{k+1}$ .  $f_{n_m}(h_{n_m})$  is a subsequence of  $f_m(h_m^*)$  and  $h_m^* \to h$ , so by (52)  $\lim_{m\to\infty} f_{n_m}(h_{n_m}) = \lim_{m\to\infty} f_m(h_m^*) = f(h) \geq f_{\star} > \limsup_{n\to\infty} f_n(h_n)$ .

Proof of Proposition 4. By (45), r = 0 implies  $||[h] - \Pi[h]||_K = 0$  and so  $[h] = \Pi[h] \in \ker \pi_1$ . Hence there is a  $h^* \in H_0$  with  $||h - h^*||_K = 0$ . By Corollary S2,

$$\limsup_{n \to \infty} P_{n,\gamma,h} \phi_n \le \limsup_{n \to \infty} P_{n,\gamma,h^*} \phi_n + \limsup_{n \to \infty} |P_{n,\gamma,h^*} \phi_n - P_{n,\gamma,h} \phi_n| \le \alpha. \quad \Box$$

Proof of Theorem 4. Since  $\lim_{n\to\infty} P_n[\dot{\ell}_n g'_n] = \lim_{n\to\infty} P_n[\dot{\ell}_n \tilde{\ell}'_n]$  we may assume  $g_n = \tilde{\ell}_n$ . By Theorem 12.14 in Rudin (1991),  $\tilde{\mathcal{I}}_n \coloneqq P_n[\tilde{\ell}_n \tilde{\ell}'_n] = P_n[\dot{\ell}_n \tilde{\ell}'_n]$ . Set  $K_n(h,g) \coloneqq P_n[\Delta_n h \Delta_n g]$  and let  $\mathsf{G}_n$  be a zero-mean Gaussian process with covariance kernel  $K_n$ . There exists a Hilbert space isomorphism,  $Z_n : \operatorname{cl}\{\Delta_n h : h \in H\} \to \operatorname{cl}\{\mathsf{G}_n h : h \in H\}$  (Janson, 1997, Theorem 1.23). Let  $R \coloneqq \Pi \left[ \cdot \left| \{\mathsf{G}_n(0,b) : b \in B\}^{\perp} \right| \right]$ and  $Q \coloneqq \Pi \left[ \cdot \left| \{\Delta_n(0,b) : b \in B\}^{\perp} \right]$ .  $R\mathsf{G}_n h = RZ_n(\Delta_n h) = Z_n QZ_n^{-1}Z_n(\Delta_n h) = Z_n Q\Delta_n h$  for  $h \in H$  and extends to elements in the closure by continuity. Hence

$$\tilde{\mathcal{I}}_{n,ij} = P_n \left[ \Delta_n(e_i, 0) Q \Delta_n(e_j, 0) \right] = \mathbb{E} \left[ \mathsf{G}_n(e_i, 0) R \mathsf{G}_n(e_j, 0) \right].$$
(53)

By Theorem 9.1 in Janson (1997),

$$\mathbb{E}\left[\mathsf{G}_{n}(e_{j},0)|\{\mathsf{G}_{n}(0,b):b\in B\}\right] = \Pi\left[\mathsf{G}_{n}(e_{j},0)|\mathrm{cl}\left\{\mathsf{G}_{n}(0,b):b\in B\right\}\right],\tag{54}$$

and so  $\tilde{\mathsf{G}}_n(e_j, 0) \coloneqq \mathsf{G}_n(e_j, 0) - \mathbb{E}\left[\mathsf{G}_n(e_j, 0) | \{\mathsf{G}_n(0, b) : b \in B\}\right] = R\mathsf{G}_n(e_j, 0)$ . Then,  $\tilde{\mathcal{I}}_{n,ij} = \mathbb{E}\left[\mathsf{G}_n(e_i, 0)\tilde{\mathsf{G}}_n(e_j, 0)\right]$  by (53). Set  $\mathscr{G}_n \coloneqq \sigma(\{\mathsf{G}_n(0, b) : b \in B\}), \mathscr{G}_n \coloneqq \sigma(\{\mathsf{G}(0, b) : b \in B\})$  and define  $X_n \coloneqq (\mathsf{G}_n(e_i, 0), \mathbb{E}[\mathsf{G}_n(e_j, 0)|\mathscr{G}_n])$  and  $X \coloneqq (\mathsf{G}(e_i, 0), \mathbb{E}[\mathsf{G}(e_j, 0)|\mathscr{G}])$ , where  $\mathsf{G} \coloneqq \Delta$ . By (54) and  $K_n(h, h) \to K(h, h)$  (Lemma 2),  $(X_n)_{n \in \mathbb{N}}$  are uniformly square integrable Gaussian random vectors and  $X_n \rightsquigarrow X$ (by Theorem S3). Combine with Lemma S3 and Theorem 9.1 of Janson (1997).  $\Box$ 

Proof of Lemma 4.  $R_n(h) \xrightarrow{P_n} 0$  in (5) and  $Ah \in L_2^0(P)$  follows from (24) (van der Vaart and Wellner, 1996, Lemma 3.10.11). Hence  $\Delta_n h$  is uniformly square integrable (i.i.d) and  $[\Delta_n h](W^{(n)}) = \mathbb{G}_n Ah \rightsquigarrow \mathcal{N}(0, \int (Ah)^2 dP)$  (CLT).

Proof of Lemma 5.  $P^n(\Delta_n h, g'_n) = 0$ . By  $g \in \{Db : b \in B\}^{\perp}$  and Assumption 5, the covariance matrix of  $(\Delta_n h, g'_n)$  (under  $P^n$ ) is  $\Sigma(h) = P\begin{bmatrix} [Ah]^2 & \tau' \dot{\ell}g' \\ g\dot{\ell}'\tau & gg' \end{bmatrix}$ . For each  $h \in H$ , the central limit theorem gives  $(\Delta_n h, g'_n) \xrightarrow{P^n} \mathcal{N}(0, \Sigma(h))$ .

Proof of Corollary 4.  $g \in \{Db : b \in B\}^{\perp}$ . Apply Lemma 5.