Enhanced power enhancements for testing many moment equalities: Beyond the 2- and ∞-norm

Anders Bredahl Kock*
University of Oxford
Department of Economics
10 Manor Rd, Oxford OX1 3UQ
anders.kock@economics.ox.ac.uk

David Preinerstorfer
WU Vienna University of Economics and Business
Institute for Statistics and Mathematics
Welthandelsplatz 1, 1020 Vienna
david.preinerstorfer@wu.ac.at

Abstract

Tests based on the 2- and ∞-norm have received considerable attention in high-dimensional testing problems, as they are powerful against dense and sparse alternatives, respectively. The power enhancement principle of Fan et al. (2015) combines these two norms to construct improved tests that are powerful against both types of alternatives. In the context of testing whether a candidate parameter satisfies a large number of moment equalities, we construct tests that harness the strength of *all* p-norms with $p \in [2, \infty]$. As a result, these tests are consistent against strictly more alternatives than *any* test based on a single p-norm. In particular, our tests are consistent against more alternatives than tests based on the 2- and ∞-norm, which is what most implementations of the power enhancement principle target.

We illustrate our general results in the linear instrumental variable model with many instruments, for which we also provide numerical results and an empirical illustration.

Keywords: High-dimensional testing; power enhancement principle; moment equalities; instrumental variables;

*Anders Bredahl Kock's research was supported by ERC grant number 101124535 – HIDI (UKRI EP/Z002222/1). He is also a member of, and grateful for support from, ACE and CoRE (hosted by Aarhus University).

1 Introduction

In many statistical settings target parameters are identified by moment equalities. This is the case, e.g., in M-, Z-, and GMM-estimation problems or when moment-based quantities, such as mean vectors or covariance matrices, are themselves the targets. Given a sample $\boldsymbol{X}_{1,n}, \ldots, \boldsymbol{X}_{n,n}$ of identically distributed random variables, one then frequently wishes to test whether or not a candidate parameter $\boldsymbol{\beta}_n^*$ satisfies said moment conditions in the sense that $\mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = \mathbf{0}_d$. Denote the scaled empirical counterpart of the population moments $\mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)$ by

$$\boldsymbol{H}_n(\boldsymbol{\beta}_n^*) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_n(\boldsymbol{X}_{i,n}, \boldsymbol{\beta}_n^*),$$

and let $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_n^*)$ be an estimator of the covariance matrix $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ of $\boldsymbol{H}_n(\boldsymbol{\beta}_n^*)$. It is then natural to base a test of $\mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = \mathbf{0}_d$ on the distance of $\hat{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*)$ from the origin. To measure this distance, one most commonly makes use of a p-norm, which then requires the choice of the exponent p. Most tests currently used in practice are based on $p = 2$ or $p = \infty$. It is well-understood in the statistics literature that in settings where d is large relative to n (which is the focus of the present article) tests based on $p = 2$ are relatively powerful against "dense" alternatives, whereas tests based on $p = \infty$ are relatively powerful against "sparse" alternatives, cf., e.g., Ingster and Suslina (2003). To construct a test that is simultaneously powerful against dense as well as many sparse alternatives, Fan et al. (2015) combined tests based on the 2- and ∞-norm via their *power enhancement principle*. The idea of combining the 2- and ∞-norm based tests or a *finite* number of p-norm based tests in order to construct a more powerful test has since gained considerable popularity in many subfields of modern statistics, cf. Xu et al. (2016), Yang and Pan (2017), He et al. (2021), Yu et al. (2023), Yu et al. (2024a), Li et al. (2024b) [testing high-dimensional means and covariance matrices], Tang et al. (2018) [conditional marginal regression], Kock and Preinerstorfer (2019) [testing in LAN models], Liu et al. (2019) [genetic pathway testing], Jammalamadaka et al. (2020) [tests for uniformity on the sphere], Feng et al. (2022), Juodis and Reese (2022) [tests for cross-sectional independence in high-dimensional panel data models], Zhang et al. (2022), Li et al. (2024a) [change point detection and structural breaks], Fan et al. (2024) [testing with many instrumental variables], Ge et al. (2024), Yu et al. (2024b) [asset pricing and finance].

Despite the success of the power enhancement principle and related combination procedures, sparse and dense alternatives are merely two (conceptually useful) "endpoints" between which a continuum of "semi-sparse" alternatives exist. This continuum of structures is mirrored by a continuum of p-norms, $p \in [2, \infty]$, between the two extremes $p = 2$ and $p = \infty$ that most currently used tests are based on.[1] In the clean but restrictive testbed of the Gaussian sequence model, Kock and Preinerstorfer (2023) showed that there exist semi-sparse alternatives against which tests based on the 2- and ∞-norm are inconsistent, but against which tests based on any $p \in (2, \infty)$ are consistent. Thus, it is important to harness the power from *all* — infinitely many — p-norms, $p \in [2, \infty]$, which Kock and Preinerstorfer (2023) exploited to construct a test that is "dominant" — in the Gaussian sequence model — in the following sense: If there exists *some* p for which the corresponding p-norm based test is consistent against a given alternative, then so is their test.

In the present paper, we study size and power properties of p-norm based tests with $p \in [2, \infty]$ in Section 3, including a characterization of their consistency properties. Then, in Section 4, we use these results together with the principle of Kock and Preinerstorfer (2023) to construct a test $\psi_n(\boldsymbol{\beta}_n^*)$ that simultaneously dominates all p-norm based tests with $p \in [2, \infty]$ in terms of consistency for testing $H_{0,n} : \mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = \mathbf{0}_d$. In contrast to applications of the classic power enhancement of Fan et al. (2015), which combines two test statistics, $\psi_n(\boldsymbol{\beta}_n^*)$ is based on combining a number of p-norm based test statistics that increases *unboundedly* in n. For any sequence of alternatives we prove in Theorem 4.1 that if there exists some $p \in [2, \infty]$ such that the corresponding p-norm based test is consistent, then so is $\psi_n(\boldsymbol{\beta}_n^*)$. In particular, $\psi_n(\boldsymbol{\beta}_n^*)$

[1] The results in Kock and Preinerstorfer (2023) show that already in the Gaussian sequence model tests based on p-norms with $p \leq 2$ are dominated (in terms of consistency) by the 2-norm based test. For simplicity, we therefore focus on $p \in [2, \infty]$ in the present paper.

1. is consistent against dense alternatives whenever a test based on the 2-norm is consistent.

2. is consistent against sparse alternatives whenever a test based on the ∞-norm is consistent.

3. is consistent against many semi-sparse alternatives that tests based on $p = 2$, $p = \infty$ or the power enhancement principle are inconsistent against (and may even have asymptotic power equaling their asymptotic size). This power gain can be of practical relevance as there is often no reason to believe that alternatives (or, more relevant, the non-centrality parameter in (6) below) are exactly sparse or dense. The gain comes from harnessing the strengths of p-norms beyond 2 and ∞, cf. also the discussion in Section 4.

We illustrate our results in the running example of the linear instrumental variable model with many instruments, which we recall in Section 2.1.3. Here our general construction yields a test that is powerful irrespective of whether the first-stage is sparse, dense, or semi-sparse. Numerical results and an empirical illustration can be found in Sections 6 and 7, respectively.

2 The problem

We observe realizations of identically distributed (but not necessarily independent) random variables $\boldsymbol{X}_{1,n}, \ldots, \boldsymbol{X}_{n,n}$, each of which is defined on an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and takes its values in the measurable space $(\mathcal{X}_n, \mathcal{A}_n)$. We wish to test whether a candidate parameter $\boldsymbol{\beta}_n^*$ in the parameter space $\mathbf{B}_n \neq \emptyset$ satisfies the *moment equalities*

$$\mathbb{E}h_{j,n}(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = 0, \qquad j = 1, \ldots, d(n), \tag{1}$$

for $h_{j,n} : \mathcal{X}_n \times \mathbf{B}_n \to \mathbb{R}$ such that $h_{j,n}(\cdot, \boldsymbol{\beta}_n) : \mathcal{X}_n \to \mathbb{R}$ is measurable for every $n \in \mathbb{N}$, every $\boldsymbol{\beta}_n \in \mathbf{B}_n$, and every $j = 1, \ldots, d$; additionally, without further mentioning, we maintain the assumption that

$$\mathbb{E}h_{j,n}^2(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n) < \infty. \tag{2}$$

For our general results, no structure needs to be imposed on the parameter space $\mathbf{B}_n$. Of course, the complexity of constructing confidence sets for the parameters via test inversion depends on the size and structure of $\mathbf{B}_n$.

For $h_n = (h_{1,n}, \ldots, h_{d,n})'$ and $\mathbf{0}_d$ denoting the $d \times 1$ vector of zeros, the requirement in (1) is more conveniently expressed as $\mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = \mathbf{0}_d$. We thus consider the testing problem

$$H_{0,n} : \mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = \mathbf{0}_d \quad \text{against} \quad H_{1,n} : \mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) \neq \mathbf{0}_d. \tag{3}$$

In this paper we are interested asymptotic size and power properties of tests when $d = d(n) \to \infty$ *as the sample size* $n \to \infty$. That is, there are "many" moment equalities. To conserve on notation, we write d instead of $d(n)$ whenever this causes no confusion.

Finally, we stress that since the tests to be studied build on plugging in a candidate $\boldsymbol{\beta}_n^*$, no assumptions need to be made concerning the identification of the parameter(s) satisfying (1). Thus, the tests are trivially robust to (weak) identification problems.

2.1 Examples

The true population parameters in M- and Z-estimation problems, such as (non-linear) least squares and maximum likelihood satisfy first-order conditions of the form (1). Let us illustrate how testing mean vectors, testing covariance matrices, and inference in the presence of many weak instruments fit within our general testing framework. In Section K of the supplementary appendix we also provide details for treatment effect testing examples.

2.1.1 Testing high-dimensional mean vectors

Given data $\boldsymbol{X}_{i,n}$ with values in $\mathbb{R}^d$, one can test whether its mean vector equals a given $\boldsymbol{\beta}_n^*$ or not by choosing $h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = \boldsymbol{X}_{1,n} - \boldsymbol{\beta}_n^*$. Clearly, $\mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = \mathbf{0}_d$ if and only if $\mathbb{E}\boldsymbol{X}_{1,n} = \boldsymbol{\beta}_n^*$. In this example, $\mathcal{X}_n = \mathbf{B}_n = \mathbb{R}^d$. Confer Huang et al. (2022) for a recent overview of tests on high-dimensional means.

2.1.2 Testing high-dimensional covariance matrices

Given data $\boldsymbol{X}_{i,n}$ with values in $\mathbb{R}^d$ and mean zero,[2] one can test whether the covariance matrix of $\boldsymbol{X}_{1,n}$ equals the covariance matrix $\boldsymbol{\beta}_n^*$ by choosing

$$h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = \text{vech}\left(\boldsymbol{X}_{1,n}\boldsymbol{X}_{1,n}' - \boldsymbol{\beta}_n^*\right),$$

where vech$(\cdot)$ is the standard half-vectorization operator. Clearly, $\mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = \mathbf{0}_{d(d+1)/2}$ if and only if $\mathbb{E}\boldsymbol{X}_{1,n}\boldsymbol{X}_{1,n}' = \boldsymbol{\beta}_n^*$. In this example, $\mathcal{X}_n = \mathbb{R}^d$ and one can choose, e.g., $\mathbf{B}_n$ as the set of $d \times d$ covariance matrices.

2.1.3 Inference in the presence of many (weak) instruments

Consider the classic linear instrumental variable (IV) setting in which $\boldsymbol{X}_{i,n} = (y_{i,n}, \mathbf{Y}_{i,n}', \boldsymbol{z}_{i,n}')'$ for $y_{i,n} \in \mathbb{R}$ an outcome of interest, $\mathbf{Y}_{i,n} \in \mathbb{R}^k$ a vector of endogenous explanatory variables and $\boldsymbol{z}_{i,n} \in \mathbb{R}^d$ a vector of instruments. Thus, $\mathcal{X}_n = \mathbb{R}^{1+k+d}$ and $\mathbf{B}_n = \mathbb{R}^k$. Because the OLS estimator is biased when explanatory variables are endogenous, one here bases inference on the vector of instruments by working with the moment equalities obtained via $h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = (y_{1,n} - \mathbf{Y}_{1,n}'\boldsymbol{\beta}_n^*)\boldsymbol{z}_{1,n}$. One then wishes to test whether or not a candidate parameter vector $\boldsymbol{\beta}_n^*$ lies in the identified set, i.e.,

$$H_{0,n} : \mathbb{E}\left[(y_{1,n} - \mathbf{Y}_{1,n}'\boldsymbol{\beta}_n^*)\boldsymbol{z}_{1,n}\right] = \mathbf{0}_d \qquad \text{against} \qquad H_{1,n} : \mathbb{E}\left[(y_{1,n} - \mathbf{Y}_{1,n}'\boldsymbol{\beta}_n^*)\boldsymbol{z}_{1,n}\right] \neq \mathbf{0}_d.$$

Note that the number of moment equalities equals the number of instruments d. This can be large even for k fixed. Inference on $\boldsymbol{\beta}_n^*$ in the presence of many (weak) instruments has received considerable recent attention as witnessed by, e.g., the works of Andrews and Stock (2007), Anatolyev and Gospodinov (2011), Belloni et al. (2012), Tchuente and

[2] It is without loss of generality to assume $\mathbb{E}\boldsymbol{X}_{1,n} = \mathbf{0}_d$ since $\boldsymbol{Y}_{i,n} = (\boldsymbol{X}_{2i,n} - \boldsymbol{X}_{2i-1,n})/\sqrt{2}$, $i = 1, \ldots, \lfloor n/2 \rfloor$, has mean zero and the same covariance matrix as $\boldsymbol{X}_{1,n}$.

Carrasco (2016), Kaffo and Wang (2017), Crudu et al. (2021), Mikusheva (2021), Matsushita and Otsu (2021), Mikusheva and Sun (2022), Matsushita and Otsu (2022), Dovì et al. (2024), Boot and Ligtenberg (2023). There are several important practical reasons for this interest. First, when identification is weak, researchers may seek to obtain more precise inference by using a large number of instruments in order to capture more of the exogenous variation in the endogenous covariates. Second, approaches such as the granular IV approach of Gabaix and Koijen (2020), the saturation approach of Blandhol et al. (2022) or Mendelian Randomization (Davey Smith and Ebrahim (2003)), the latter using genetic variation as instruments, can lead to situations where the number of instruments is large compared to the sample size. The same is true for technical instruments such as transformations or interactions or empirical strategies such as "judge designs," cf. Mikusheva and Sun (2024).

3 Characterization of consistency properties of p-norm based tests

Recall from the introduction that given $\boldsymbol{\beta}_n^* \in \mathbf{B}_n$, we write

$$\boldsymbol{H}_n(\boldsymbol{\beta}_n^*) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_n(\boldsymbol{X}_{i,n}, \boldsymbol{\beta}_n^*),$$

and let $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_n^*)$ be a positive semidefinite and symmetric estimator of the covariance matrix $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ of $\boldsymbol{H}_n(\boldsymbol{\beta}_n^*)$, which is guaranteed to exist by (2).

Remark 3.1. Already under independent sampling, the choice of $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_n^*)$ may depend on instance-specific additional structural information on $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ (e.g., bandedness, sparsity or factor structure), the incorporation of which can improve the estimator. We are particularly interested in a setting where no structural information on $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ is available and leave the specific choice of estimator unspecified in the following. In Section 3.4 we highlight how the precision of the chosen $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_n^*)$ influences the allowed growth rate of d and discuss precision guarantees available in the literature.

A common way of measuring the empirical evidence against the null of $\mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = \mathbf{0}_d$ is based on the the Euclidean norm of $\hat{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*)$, that is, one rejects $H_{0,n}$ null whenever

$$S_{n,2}(\boldsymbol{\beta}_n^*) := \sqrt{\boldsymbol{H}_n'(\boldsymbol{\beta}_n^*)\hat{\boldsymbol{\Sigma}}_n^{-1}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*)} = \left\|\hat{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*)\right\|_2 \tag{4}$$

exceeds a critical value $\kappa_{n,2}$ chosen to ensure that the resulting test has a desired (asymptotic) size $\alpha \in (0,1)$, and where $\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^d x_i^2}$ for $\boldsymbol{x} \in \mathbb{R}^d$.[3] For $\boldsymbol{x} = (x_1, \ldots, x_d)' \in \mathbb{R}^d$ and $p \in [2, \infty]$, define the p-norm

$$\|\boldsymbol{x}\|_p := \begin{cases} \left(\sum_{i=1}^d |x_i|^p\right)^{\frac{1}{p}} & \text{if } p < \infty, \\ \max_{i=1,\ldots,d} |x_i| & \text{else} \end{cases}$$

and introduce, analogously to $S_{n,2}(\boldsymbol{\beta}_n^*)$ in (4) but based on the p-norm, the family of test statistics

$$S_{n,p}(\boldsymbol{\beta}_n^*) := \left\|\hat{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*)\right\|_p, \qquad p \in [2, \infty]. \tag{5}$$

For sequences of critical values $(\kappa_{n,p})_{n\in\mathbb{N}}$ that guarantee a desired asymptotic size $\alpha \in (0,1)$, we shall now study consistency properties of the *p-norm based* tests $\mathbb{1}(S_{n,p}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,p})$ and coverage properties of the associated confidence sets

$$\left\{\boldsymbol{\beta}_n \in \mathbf{B}_n : S_{n,p}(\boldsymbol{\beta}_n) \leq \kappa_{n,p}\right\},$$

when $d \to \infty$ as $n \to \infty$. Define

$$\mathbf{B}^* := \bigtimes_{n=1}^{\infty} \mathbf{B}_n,$$

i.e., the set of possible *sequences* of parameters $\boldsymbol{\beta}_n$ for $n \in \mathbb{N}$, along which power or coverage properties of tests for the sequence of testing problems in (3) will be studied in

[3] In order to avoid taking a stance on $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_n^*)$ potentially not being invertible, we denote by $\boldsymbol{A}^{-1}$ the Moore-Penrose pseudoinverse for any matrix $\boldsymbol{A}$ and define $\boldsymbol{A}^{-1/2} := \left(\boldsymbol{A}^{1/2}\right)^{-1}$ in case $\boldsymbol{A}$ is symmetric and positive semidefinite, and where $\boldsymbol{A}^{1/2}$ denotes the unique symmetric positive semidefinite square root of $\boldsymbol{A}$. Recall that the Moore-Penrose inverse is identical to the regular matrix inverse whenever the latter exists.

the current setup. In particular, we write $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*, \ldots)$. Let

$$\mathbf{B}^{(0)} = \left\{\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots) \in \mathbf{B}^* : \mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n) = \mathbf{0}_d \text{ for every } n \in \mathbb{N}\right\} \subseteq \mathbf{B}^*$$

be the set of sequences of parameters satisfying $H_{0,n}$ in (3) for every $n \in \mathbb{N}$.

As we shall see, the scaled deviations from the null hypothesis

$$\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*) = \sqrt{n}\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)\mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) \in \mathbb{R}^d, \tag{6}$$

play the role of a noncentrality parameter in characterizing the consistency properties of tests based on $S_{n,p}(\boldsymbol{\beta}_n^*)$.

3.1 Main assumption

To present the assumption under which our results are obtained, let $\boldsymbol{Z}_d \sim \mathsf{N}_d(\mathbf{0}_d, \mathbf{I}_d)$,

$$\mathcal{C}_n = \left\{C \subseteq \mathbb{R}^{d(n)} : C \text{ is convex and Borel measurable}\right\},$$

and $\|\boldsymbol{A}\|_2$ be the spectral norm of the matrix $\boldsymbol{A}$. The assumption is then as follows.

Assumption 3.1. Let $\boldsymbol{\beta}^* \in \mathbf{B}^*$, assume that $d \to \infty$, that the eigenvalues of $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ are (uniformly) bounded away from zero and from above, and that

1. $\|\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_n^*) - \boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)\|_2 = O_{\mathbb{P}}(a_n)$ with $d^{3/4}a_n \to 0$, and
2. $\sup_{C \in \mathcal{C}_n} \left|\mathbb{P}\left(\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*) \in C\right) - \mathbb{P}\left(\boldsymbol{Z}_d + \boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*) \in C\right)\right| \to 0$.

Note that Assumption 3.1 does not impose independence. Essentially, Assumption 3.1 requires an upper bound on the estimation error of the covariance matrix, $\|\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_n^*) - \boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)\|_2$, and a Gaussian approximation to hold.[4] In Section 3.4 below we discuss catalogues of sufficient conditions for both of these two requirements. In particular,

[4] We impose the Gaussian approximation to hold over convex sets as we study all p-norm based tests simultaneously. If one is only interested in a single p, it may be enough for the approximation to hold over p-norm balls with arbitrary centres and radii for this given p — potentially allowing for d to grow faster; cf. also Theorem B.1 in Appendix B.

Lemma 3.4 essentially shows that under i.i.d. sampling and bounded fourth moments Assumption 3.1 holds if $d/n^{2/5} \to 0$ (up to logarithmic factors in n). We also discuss how this growth rate of d cannot be improved by much under these moment assumptions even if one uses a test based on the ∞-norm.

Finally, in Section 3.4.2 we state and discuss an alternative assumption not imposing that the smallest eigenvalue of $\mathbf{\Sigma}_n(\boldsymbol{\beta}_n^*)$ is (uniformly) bounded away from zero and under which our results can be established with minor adjustments of our arguments.

3.2 Size and power properties of p-norm based tests: $p \in [2, \infty)$

We first consider the case of $p \in [2, \infty)$ as the characterization of their power has a common structure. The case of $p = \infty$ is then covered in Section 3.3.

We denote the cdf of a standard normal distribution by Φ, let $\sigma_p^2 := \mathbb{V}ar(|Z|^p)$ with $Z \sim \mathsf{N}_1(0, 1)$, and define the functions

$$\lambda_p(x) = \mathbb{E}|Z + x|^p \qquad \text{and} \qquad g_p(x) = x^2 \vee |x|^p, \qquad x \in \mathbb{R},\ p \in [2, \infty). \tag{7}$$

Theorem 3.1. *Let $p \in [2, \infty)$, $\alpha \in (0, 1)$, and Assumption 3.1 be satisfied.*

1. *Size control: Suppose $\boldsymbol{\beta}^* \in \mathbf{B}^{(0)}$. A sequence of real numbers $(\kappa_{n,p})_{n \in \mathbb{N}}$ satisfies*

$$\mathbb{P}\left(S_{n,p}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,p}\right) \to \alpha, \tag{8}$$

   if and only if $\kappa_{n,p} = \kappa_{n,p}(\alpha) = \left[(\Phi^{-1}(1-\alpha) + o(1))d^{1/2}\sigma_p + d\lambda_p(0)\right]^{1/p}$.

2. *Local power: If*

$$\frac{1}{\sqrt{d}} \sum_{i=1}^{d} [\lambda_p(\theta_{i,n}(\boldsymbol{\beta}_n^*)) - \lambda_p(0)] \to c \in [0, \infty),$$

   then, for $(\kappa_{n,p})_{n \in \mathbb{N}}$ satisfying (8),

$$\mathbb{P}\left(S_{n,p}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,p}\right) \to 1 - \Phi(\Phi^{-1}(1-\alpha) - c/\sigma_p).$$

3. Consistency: For $(\kappa_{n,p})_{n\in\mathbb{N}}$ satisfying (8), it holds that

$$\mathbb{P}\left(S_{n,p}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,p}\right) \to 1 \quad \iff \quad \frac{\sum_{i=1}^d [\lambda_p(\theta_{i,n}(\boldsymbol{\beta}_n^*)) - \lambda_p(0)]}{\sqrt{d}} \to \infty \tag{9}$$

$$\iff \quad \frac{\sum_{i=1}^d g_p(\theta_{i,n}(\boldsymbol{\beta}_n^*))}{\sqrt{d}} \to \infty. \tag{10}$$

Part 1. of Theorem 3.1 characterizes sequences of critical values that yield asymptotic size control. For any $\alpha \in (0,1)$, a canonical choice is $\kappa_{n,p} = [\Phi^{-1}(1-\alpha)d^{1/2}\sigma_p + d\lambda_p(0)]^{1/p}$.

Part 2. provides results on local asymptotic power and Part 3. provides a complete characterization of the alternatives that a p-norm based test is consistent against. Apart from when $p = 2$, neither of these can be expressed in terms of only the p-norm of the (scaled) deviation from the null $\|\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)\|_p$; instead they depend on the asymptotic behaviour of

$$d^{-1/2}\sum_{i=1}^d [\lambda_p(\theta_{i,n}(\boldsymbol{\beta}_n^*)) - \lambda_p(0)].$$

Concerning consistency, the divergence of the latter is equivalent to the divergence of

$$\frac{\sum_{i=1}^d g_p(\theta_{i,n}(\boldsymbol{\beta}_n^*))}{\sqrt{d}} = \frac{\sum_{i=1}^d \left[\theta_{i,n}^2(\boldsymbol{\beta}_n^*) \vee |\theta_{i,n}(\boldsymbol{\beta}_n^*)|^p\right]}{\sqrt{d}},$$

which is somewhat easier to interpret and highlights that even for $p \neq 2$ the 2-norm enters the characterization of the consistency of a p-norm based test.

Concerning the consistency of a p-norm based test (with asymptotic size in $(0,1)$), (10) implies that a sufficient condition is that $d^{-1/2}\|\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)\|_2^2 \to \infty$ *or* $d^{-1/2}\|\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)\|_p^p \to \infty$, cf. also the previous display. What is more, observing that

$$g_p(x) \leq g_q(x) \qquad \text{for} \qquad 2 \leq p < q < \infty \text{ and all } x \in \mathbb{R},$$

Part 3. of Theorem 3.1 shows that

$$\mathbb{P}\left(S_{n,p}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,p}\right) \to 1 \qquad \text{implies} \qquad \mathbb{P}\left(S_{n,q}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,q}\right) \to 1 \tag{11}$$

when $(\kappa_{n,p})_{n\in\mathbb{N}}$ and $(\kappa_{n,q})_{n\in\mathbb{N}}$ are chosen such that the corresponding tests have asymptotic sizes in $(0,1)$. In words, *any violation of the moment conditions* $\mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = \mathbf{0}_d$ *that a test based on the p-norm is consistent against, a test based on the q-norm will also be consistent against*, $2 \leq p < q < \infty$. Thus, the q-norm based test always weakly dominates the p-norm based test in terms of consistency. This domination is *strict* if and only if there exists a $\boldsymbol{\beta}^* \in \mathbf{B}^*$ such that $\liminf_{n\to\infty} \mathbb{P}\left(S_{n,p}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,p}\right) < 1$ but $\mathbb{P}\left(S_{n,q}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,q}\right) \to 1$, i.e., if and only if (cf. (10))

$$\liminf_{n\to\infty} \frac{\sum_{i=1}^d g_p(\theta_{i,n}(\boldsymbol{\beta}_n^*))}{\sqrt{d}} < \infty \qquad \text{and} \qquad \frac{\sum_{i=1}^d g_q(\theta_{i,n}(\boldsymbol{\beta}_n^*))}{\sqrt{d}} \to \infty. \tag{12}$$

By Theorem 3.4 in Kock and Preinerstorfer (2023) the $\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)$ that satisfy (12) are necessarily approximately sparse and unbalanced. Furthermore, if the limit inferior to the left in (12) is zero, then there exists a subsequence along which the p-norm based test has asymptotic power equal to its size, while the q-norm based test is consistent — a substantial difference in power.

The consequences of Theorem 3.1 for the relative size of confidence sets based on p-norm based tests are discussed in Section I.1 of the appendix.

Theorem B.1 in Section B of the supplementary appendix states a version of Theorem 3.1 under milder but more high-level assumptions.

3.2.1 Proof idea of Theorem 3.1

A substantial part of the proof of Theorem 3.1 follows from the following lemma, which is a restatement of Lemma C.3 in the supplementary appendix.

Lemma 3.2. *Let Assumption 3.1 be satisfied. Then, if* $\|\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)\|_2^2/d$ *is bounded, it holds that*

$$\sup_{C\in\mathcal{C}_n} \left|\mathbb{P}\left(\hat{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*) \in C\right) - \mathbb{P}\left(\boldsymbol{Z}_d + \boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*) \in C\right)\right| \to 0 \tag{13}$$

Lemma 3.2 asserts that the Gaussian approximation over convex sets for the high-dimensional sums in Part 2. of Assumption 3.1 remains valid when $\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)$ is replaced by $\hat{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^*)$. To show that this replacement is possible, we use Part 1. of Assumption 3.1 to show that $\hat{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*)$ is sufficiently close to $\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*)$. This, in turn, allows us to use the Gaussian anti-concentration inequalities in Lemma 2.6 of Bentkus (2003) to establish (13).

For $r \in (0, \infty)$, let $\mathbb{B}_p(r) = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_p < r\}$, which is convex for $p \in [1, \infty]$. By Lemma 3.2

$$\begin{aligned}
\mathbb{P}\left(S_{n,p}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,p}\right) &= 1 - \mathbb{P}\left(S_{n,p}(\boldsymbol{\beta}_n^*) < \kappa_{n,p}\right) \\
&= 1 - \mathbb{P}\left(\hat{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*) \in \mathbb{B}_p(\kappa_{n,p})\right) \\
&\approx 1 - \mathbb{P}\left(\boldsymbol{Z}_d + \boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*) \in \mathbb{B}_p(\kappa_{n,p})\right) \\
&= \mathbb{P}\left(\left\|\boldsymbol{Z}_d + \boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)\right\|_p \geq \kappa_{n,p}\right)
\end{aligned}$$

Therefore, to establish the asymptotic properties of $\mathbb{P}\left(S_{n,p}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,p}\right)$ in Theorem 3.1 it suffices to study the rejection properties of p-norm based tests in the Gaussian sequence model. That is, Lemma 3.2 essentially allows us to deduce the asymptotic properties of $S_{n,p}(\boldsymbol{\beta}_n^*) = \|\hat{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*)\|_p$ in Theorem 3.1 from the ones of p-norms of the high-dimensional Gaussian mean shift $\boldsymbol{Z}_d + \boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)$ in the "limit experiment", which was analyzed in Kock and Preinerstorfer (2023). Note that Lemma 3.2 relies on $d^{-1}\sum_{i=1}^d g_2(\theta_{i,n}(\boldsymbol{\beta}_n^*)) = d^{-1}\|\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)\|_2^2$ being bounded. In regimes wherein this is not the case, arguments tailored to those settings are used to establish Part 3. of Theorem 3.1.

Lemma 3.2 may be of independent interest beyond the scope of this paper as it allows one to deduce probabilistic properties of the scaled "central statistic" $\hat{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*)$ from those of $\boldsymbol{Z}_d + \boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)$ even when $d \to \infty$.

3.3 Size and power properties of tests based on the supremum-norm: $p = \infty$

Tests based on the supremum-norm have received considerable attention due to the work on high-dimensional Gaussian approximations over hyperrectangles

$$\mathcal{H}_n = \left\{\prod_{j=1}^d [a_j, b_j] \cap \mathbb{R} : \; -\infty \leq a_j \leq b_j \leq \infty, \; j = 1, \ldots, d\right\}.$$

The following theorem, the proof of which also builds on Lemma 3.2, characterizes which alternatives tests based on the supremum-norm are consistent against.

Theorem 3.3. *Let $\alpha \in (0,1)$ and suppose Assumption 3.1 is satisfied. Then, the following holds.*

a) Size control: Suppose $\boldsymbol{\beta}^* \in \mathbf{B}^{(0)}$. A sequence of real numbers $(\kappa_{n,\infty})_{n\in\mathbb{N}}$ satisfies

$$\mathbb{P}\left(S_{n,\infty}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,\infty}\right) \to \alpha \tag{14}$$

if and only if $\kappa_{n,\infty} = \kappa_{n,\infty}(\alpha) = \sqrt{2\log(d)} - \frac{\log\log(d)+\log(4\pi)}{2\sqrt{2\log(d)}} - \frac{\log(-\log(1-\alpha)/2)+o(1)}{\sqrt{2\log(d)}}$.

b) Consistency: For $(\kappa_{n,\infty})_{n\in\mathbb{N}}$ as in (14)

$$\mathbb{P}\left(S_{n,\infty}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,\infty}\right) \to 1 \quad \iff \quad \sum_{i=1}^{d} \frac{\overline{\Phi}\left(\mathfrak{c}_d - |\theta_{i,d}(\boldsymbol{\beta}_n^*)|\right)}{\Phi\left(\mathfrak{c}_d - |\theta_{i,d}(\boldsymbol{\beta}_n^*)|\right)} \to \infty,$$

where $\mathfrak{c}_d := \sqrt{2\log(d)} - \frac{\log\log(d)}{2\sqrt{2\log(d)}}$ for $d \geq 2$ and $\overline{\Phi} = 1 - \Phi$ (and one may set $\mathfrak{c}_1 := 0$).

Well-known *sufficient*, yet not necessary, conditions for consistency of the supremum-norm based test, such as $\|\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)\|_\infty - \sqrt{2\log(d)} \to \infty$, are special cases of Theorem 3.3. In particular, the ∞-norm based test is consistent against the sparse alternative $\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*) = (\sqrt{3\log(d)}, 0, \ldots, 0)'$, whereas no p-norm based test with $p \in [2,\infty)$ is consistent against that alternative by (10) of Theorem 3.1.

Concerning dense alternatives of the form $\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*) = (c_n, \ldots, c_n)'$, it follows that the supremum-norm based test is consistent if and only if $\sqrt{\log(d)}|c_n| \to \infty$, cf. Appendix B.1 of Kock and Preinerstorfer (2023). Thus, the supremum-norm based test is *not* consistent if $c_n = 1/\sqrt{\log(d)}$ whereas by (10) of Theorem 3.1 every p-norm based test with $p \in [2,\infty)$ *is* consistent. Hence, i) the monotonicity in (11) does not extend to $q = \infty$ and ii) no single p-norm based test dominates all others in terms of its consistency properties.

The exact characterization of consistency in Theorem 3.3 sheds new light on the frequently used supremum-norm based test for testing moment equalities. It is more informative than the currently available power analysis of the supremum-norm based test through a maximin-lens. The latter provides the sufficient condition for consistency $\|\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)\|_\infty - \sqrt{2\log(d)} \to \infty$, but is silent about the alternatives not satisfying this condition that the supremum-norm based test is nevertheless consistent against. For example, if $\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*) = (1, \ldots, 1)'$, then by the observation on dense alternatives in the previous paragraph the supremum-norm based test *is* consistent, yet $\|\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)\|_\infty - \sqrt{2\log(d)} \not\to \infty$.

3.4 Discussion of Assumption 3.1 and an alternative assumption

3.4.1 Discussion of Assumption 3.1

Concerning Part 1. of Assumption 3.1, it is well-known that if the vectors $h_n(\boldsymbol{X}_{i,n}, \boldsymbol{\beta}_n^*)$ are i.i.d. and appropriately sub-Gaussian, then the empirical covariance matrix $\hat{\boldsymbol{\Sigma}}_{n,\text{emp}}(\boldsymbol{\beta}_n^*)$ satisfies $\|\hat{\boldsymbol{\Sigma}}_{n,\text{emp}}(\boldsymbol{\beta}_n^*) - \boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)\|_2 = O_{\mathbb{P}}(\sqrt{d/n})$, cf., e.g., Vershynin (2012) or Koltchinskii and Lounici (2017). If, in addition, $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ possesses further structure, this can be utilised to construct estimators with even better performance guarantees. For example, Bickel and Levina (2008) showed that if $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ is sparse, $\hat{\boldsymbol{\Sigma}}_{n,\text{emp}}(\boldsymbol{\beta}_n^*)$ can be thresholded to yield an estimator satisfying $\|\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_n^*) - \boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)\|_2 = O_{\mathbb{P}}(\sqrt{\log(d)/n})$. The overviews in Fan et al. (2016) and Cai et al. (2016) list further estimators utilising structural properties of $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ and their performance guarantees.

One is often not willing to assume that the vector $h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)$ is sub-Gaussian. In particular, much effort has recently been devoted to the construction of covariance matrix estimators with the sub-Gaussian performance guarantee $\|\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_n^*) - \boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)\|_2 = O_{\mathbb{P}}(\sqrt{d/n})$ even when the entries of $h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)$ only possess *four* moments. Although the sample covariance matrix does not work well under such heavy tails,[5] the existence of estimators satisfying $\|\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_n^*) - \boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)\|_2 = O_{\mathbb{P}}(\sqrt{d/n})$ — without imposing structure on $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ — has been established in Abdalla and Zhivotovskiy (2024) and Oliveira and Rico (2024), cf. also Mendelson and Zhivotovskiy (2020) for bounds containing an extra factor $\sqrt{\log(d)}$.[6] These results are proven under the assumption that the kurtosis of all

[5]This is in analogy to the sub-optimality of the sample average under heavy tails in the one-dimensional mean estimation problem, cf. the overview in Lugosi and Mendelson (2019).

[6]As acknowledged by Abdalla and Zhivotovskiy (2024) and Oliveira and Rico (2024), their estimators of the population covariance matrix are targeting the best possible statistical performance guarantees at the expense of not being practical to implement. "User-friendly" estimators, losing logarithmic factors

one-dimensional marginals of $h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)$ is bounded (in particular only four moments need to exist):

Assumption 3.2. Denote $\boldsymbol{\mu}_n(\boldsymbol{\beta}_n^*) = \mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)$ and let $\langle\cdot,\cdot\rangle$ be the standard inner product in $\mathbb{R}^d$. There exists a real number $L(\boldsymbol{\beta}^*) \geq 1$, such that for all $t \in \mathbb{R}^d$ and every $n \in \mathbb{N}$

$$\left(\mathbb{E}\langle[h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) - \boldsymbol{\mu}_n(\boldsymbol{\beta}_n^*)], t\rangle^4\right)^{\frac{1}{4}} \leq L(\boldsymbol{\beta}^*)\left(\mathbb{E}\langle[h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) - \boldsymbol{\mu}_n(\boldsymbol{\beta}_n^*)], t\rangle^2\right)^{\frac{1}{2}}.$$

As pointed out in Mendelson and Zhivotovskiy (2020), Assumption 3.2 is satisfied if, e.g., $h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) - \boldsymbol{\mu}_n(\boldsymbol{\beta}_n^*)$ follows a multivariate t-distribution with $\nu > 4$ degrees of freedom. This example is one of rather heavy tails as only moments strictly lower than ν exist.

Concerning Part 2. of Assumption 3.1, note that by translation invariance of convex sets (and denoting $\boldsymbol{\mu}_n(\boldsymbol{\beta}_n^*) = \mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)$) this is equivalent to

$$\sup_{C\in\mathcal{C}_n}\left|\mathbb{P}\left(n^{-1/2}\sum_{i=1}^n \boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)[h_n(\boldsymbol{X}_{i,n}, \boldsymbol{\beta}_n^*) - \boldsymbol{\mu}_n(\boldsymbol{\beta}_n^*)] \in C\right) - \mathbb{P}\left(\boldsymbol{Z}_d \in C\right)\right| \to 0; \quad (15)$$

that is (scaled) partial sums in $\mathbb{R}^d$ with mean $\mathbf{0}_d$ and identity covariance matrix $\mathbf{I}_d$ should obey a Gaussian approximation. In case of i.i.d. sampling and $\max_{j=1,\ldots,d}\mathbb{E}|h_{j,n}(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)|^3$ being (uniformly) bounded from above, it follows from Theorem 1.1 in Bentkus (2003) that (15) is true if $d/n^{2/7} \to 0$. Furthermore, by Theorem 2.1 in Fang and Koike (2024), the convergence in (15) holds under i.i.d. sampling and Assumption 3.2 (or, slightly weaker, $\max_{j=1,\ldots,d}\mathbb{E}|h_{j,n}(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)|^4$ being (uniformly) bounded) if, up to logarithmic factors in n, $d/n^{2/5} \to 0$.

The above discussion results in the following statement.

Lemma 3.4. *Let $\boldsymbol{\beta}^* \in \mathbf{B}^*$, let $d \to \infty$, and suppose that the eigenvalues of $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ are (uniformly) bounded away from zero and from above. Furthermore, let $\boldsymbol{X}_{1,n}, \ldots, \boldsymbol{X}_{n,n}$ be i.i.d. for each $n \in \mathbb{N}$. Let Assumption 3.2 be satisfied. Suppose $\frac{d}{n^{2/5}}[\log(n)]^{2/5} \to 0$ and one uses the estimator $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_n^*)$ from Abdalla and Zhivotovskiy (2024) (with their $\eta = 0$,*

in d in the performance guarantees, have been proposed in Ke et al. (2019).

$\delta = 1/d$ and $p = 4$) or Oliveira and Rico (2024) (with their $\eta = 0$, $\alpha = 1/d$ and $p = 4$) applied to the auxiliary sample

$$\left(h_n(\boldsymbol{X}_{2,n}, \boldsymbol{\beta}_n^*) - h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)\right)/\sqrt{2}, \ldots, \left(h_n(\boldsymbol{X}_{2\lfloor n/2\rfloor,n}, \boldsymbol{\beta}_n^*) - h_n(\boldsymbol{X}_{2\lfloor n/2\rfloor-1,n}, \boldsymbol{\beta}_n^*)\right)/\sqrt{2}.$$

Then Assumption 3.1 is satisfied.

Under the assumptions of Lemma 3.4, Theorems 3.1 and 3.3 provide a characterization of the consistency properties of p-norm based tests for all $p \in [2, \infty]$. In particular, even though the $h_{j,n}(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)$ only possess four moments, the characterization also holds for p-norm based tests with $p > 4$; that is the consistency properties of p-norm based tests are characterized even when the pth moments of the entries of $h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)$ need not exist.

Remark 3.2. Many recent tests in high-dimensional testing problems are based on the supremum-norm such that Gaussian approximations over the class of hyperrectangles $\mathcal{H}_n \subseteq \mathcal{C}_n$ suffice. This may allow d to increase faster than $n^{2/5}$ when the entries of $h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)$ have only four moments. However, even for the smaller class of sets $\mathcal{H}_n$, there exist distributions with bounded fourth moments such that the Gaussian approximation breaks down if $d = n^{1+\zeta}$ for some $\zeta > 0$, cf. Zhang and Wu (2017) and Kock and Preinerstorfer (2024) for precise formulations. In particular, it may then happen that $\mathbb{E}(h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)) = 0$ but that the distribution of $h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)$ is such that for all $\alpha \in (0, 1)$ the rejection frequency of the supremum-norm based test using Gaussian size α critical values tends to *one*. Thus, not even supremum-norm based tests are guaranteed to control size uniformly over distributions with bounded fourth moments if d grows faster than n.

In Section I.3 in the appendix we show how a sample split can be used to accommodate any growth rate of d (at the price of a potential loss of power but without sacrificing asymptotic size).

3.4.2 Alternative to Assumption 3.1

Note that $S_{n,p}(\boldsymbol{\beta}_n^*)$ in (5) and subsequent statistics only depend on $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}_n^*)$ via $\hat{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^*)$. Furthermore, inspection of the proof of Theorem 3.1 and subsequent theorems shows that Part 1. of Assumption 3.1 and the smallest eigenvalue of $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ being uniformly bounded away from zero are only used to show that $\|\hat{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^*) - \boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)\|_2 = O_{\mathbb{P}}(a_n)$ with $d^{3/4}a_n \to 0$. Therefore, throughout the paper one can replace that part of Assumption 3.1 by the existence of an estimator $\widehat{\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)}$ of $\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)$ satisfying $\|\widehat{\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)} - \boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)\|_2 = O_{\mathbb{P}}(a_n)$ with $d^{3/4}a_n \to 0$. Specifically, Assumption 3.1 can be replaced by:

Assumption 3.3 (Alternative to Assumption 3.1). Let $\boldsymbol{\beta}^* \in \mathbf{B}^*$, assume that $d \to \infty$, that $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ is positive definite with eigenvalues (uniformly) bounded from above, and

1. $\|\widehat{\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)} - \boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)\|_2 = O_{\mathbb{P}}(a_n)$ with $d^{3/4}a_n \to 0$, and
2. $\sup_{C\in\mathcal{C}_n}\left|\mathbb{P}\left(\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*) \in C\right) - \mathbb{P}\left(\boldsymbol{Z}_d + \boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*) \in C\right)\right| \to 0.$

We now provide a primitive condition for Part 1. of Assumption 3.3 (Part 2. is identical to Part 2. of Assumption 3.1). Write $\boldsymbol{\Omega}_n(\boldsymbol{\beta}_n^*) = \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\beta}_n^*)$ for the precision matrix, let $\hat{\boldsymbol{\Omega}}_n(\boldsymbol{\beta}_n^*)$ be a symmetric positive semidefinite estimator of $\boldsymbol{\Omega}_n(\boldsymbol{\beta}_n^*)$, and set $\widehat{\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)} = \hat{\boldsymbol{\Omega}}_n^{1/2}(\boldsymbol{\beta}_n^*)$. Since $\|\boldsymbol{\Sigma}_n^{1/2}(\boldsymbol{\beta}_n^*)\|_2$ is (uniformly) bounded from above and

$$\|\widehat{\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)} - \boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)\|_2 = \|\hat{\boldsymbol{\Omega}}_n^{1/2}(\boldsymbol{\beta}_n^*) - \boldsymbol{\Omega}_n^{1/2}(\boldsymbol{\beta}_n^*)\|_2 \leq \|\boldsymbol{\Sigma}_n^{1/2}(\boldsymbol{\beta}_n^*)\|_2\,\|\hat{\boldsymbol{\Omega}}_n(\boldsymbol{\beta}_n^*) - \boldsymbol{\Omega}_n(\boldsymbol{\beta}_n^*)\|_2,$$

the inequality following from, e.g., (7.2.13) in Horn and Johnson (1985), it follows that Part 1. of Assumption 3.3 is satisfied if $\|\hat{\boldsymbol{\Omega}}_n(\boldsymbol{\beta}_n^*) - \boldsymbol{\Omega}_n(\boldsymbol{\beta}_n^*)\|_2 = O_{\mathbb{P}}(a_n)$ with $d^{3/4}a_n \to 0$. Estimators $\hat{\boldsymbol{\Omega}}_n(\boldsymbol{\beta}_n^*)$ obeying this convergence rate can be obtained from the expanding catalogue of precision matrix estimators, cf., e.g., Cai et al. (2011); Fan et al. (2016); Cheng et al. (2025).

4 A test dominating all p-norm based tests

The discussion following Theorem 3.3 revealed that no single p-norm based test is "best" in terms of consistency. This makes choosing a single p to base a test on very difficult. Even if one *knows* that the deviation from the null $\mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)$ is sparse, which *could* suggest using a test based on the supremum-norm, sparsity need not be inherited by the noncentrality parameter

$$\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*) = \sqrt{n}\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)\mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)$$

due to the presence of $\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)$. Therefore, one often does not know anything about the structure of $\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)$, which is the actual quantity entering the characterization of the power of p-norm based tests in Theorems 3.1 and 3.3. This underscores the importance of having a test that is powerful irrespective of the unknown structure of $\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)$.

One could hope that combining tests based on the two "endpoints" $p = 2$ and $p = \infty$ of the interval $[2, \infty]$ results in a test that is consistent whenever *some* p-norm based test for $p \in [2, \infty]$ is consistent. However, Corollary 4.2 in Kock and Preinerstorfer (2023) shows that already in the Gaussian sequence model this is not the case: There exist (semi-sparse) alternatives against which tests based on $p = 2$ and $p = \infty$ are inconsistent, but against which a test based on any single $p \in (2, \infty)$ *is* consistent. To construct a test that is consistent whenever a test based on some $p \in [2, \infty]$ is consistent, we use a construction similar in spirit to the one employed in the Gaussian sequence model in Kock and Preinerstorfer (2023). A crucial difference is that the present construction explicitly includes the supremum-norm based test rather than using a p_n-norm based test with sufficiently quickly increasing p_n to dominate the former. The present construction is more convenient as it does not require the analysis of tests for moving p_n, which allows us to reduce the number of approximation steps in the analysis.

4.1 The dominant test

Let $\alpha \in (0, 1)$ and α_2, α_I, and α_∞ be non-negative with $\alpha_2 + \alpha_I + \alpha_\infty = \alpha$ and let $\kappa_{n,2}$ and $\kappa_{n,\infty}$ satisfy

$$\mathbb{P}\left(\|\boldsymbol{Z}_d\|_2 \geq \kappa_{n,2}\right) = \alpha_2 \quad \text{and} \quad \mathbb{P}\left(\|\boldsymbol{Z}_d\|_\infty \geq \kappa_{n,\infty}\right) = \alpha_\infty.$$

Furthermore, let p_n be a strictly increasing and unbounded sequence in $(2,\infty)$ and let m_n be a non-decreasing and unbounded sequence in $\mathbb{N}$. Fix an array

$$\mathcal{A} = \left\{\alpha_{n,p_j} \in (0,1) : n \in \mathbb{N},\ j = 1,\dots,m_n\right\}$$

such that

$$\sum_{j=1}^{m_n} \alpha_{n,p_j} = \alpha_I \text{ for every } n \in \mathbb{N} \quad \text{and} \quad \lim_{n\to\infty} \alpha_{n,p_j} > 0 \text{ for every } j \in \mathbb{N},$$

where the conditions implicitly impose the existence of the respective limits. For every $n \in \mathbb{N}$ and every $j = 1,\dots,m_n$, choose $\kappa_{n,p_j} > 0$ and $c_n \in (0,1]$ such that

$$\mathbb{P}\left(\|\boldsymbol{Z}_d\|_{p_j} \geq \kappa_{n,p_j}\right) = \alpha_{n,p_j} \qquad \text{and} \qquad \mathbb{P}\left(\max_{p\in\mathfrak{P}_n} \kappa_{n,p}^{-1}\|\boldsymbol{Z}_d\|_p \geq c_n\right) = \alpha, \tag{16}$$

where $\mathfrak{P}_n = \{2, p_1, \dots, p_{m_n}, \infty\}$. Define the test $\psi_n(\boldsymbol{\beta}_n^*)$ as

$$\psi_n(\boldsymbol{\beta}_n^*) := \mathbb{1}\left\{\max_{p\in\mathfrak{P}_n} \kappa_{n,p}^{-1} S_{n,p}(\boldsymbol{\beta}_n^*) \geq c_n\right\}. \tag{17}$$

Thus, $\psi_n(\boldsymbol{\beta}_n^*)$ rejects $H_{0,n} : \mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = \mathbf{0}_d$ if for any p-norm based tests with $p \in \mathfrak{P}_n$ it is the case that $S_{n,p}(\boldsymbol{\beta}_n^*)$ exceeds $c_n\kappa_{n,p}$. The tests based on $p = 2$ and $p = \infty$ are included in the construction of $\psi_n(\boldsymbol{\beta}_n^*)$ to make it powerful against dense and sparse alternatives, respectively. The tests based on $p \in \{p_1,\dots,p_{m_n}\}$ are included to cover the (semi-sparse) alternatives that neither tests based on the 2- or supremum-norm are consistent against. The quantities α_2, α_I, and α_∞ control how much size is distributed to the components of $\psi_n(\boldsymbol{\beta}_n^*)$ targeting dense, semi-sparse and sparse alternatives, respectively while the array $\mathcal{A}$ distributes the size α_I to the p-norm based tests with $p \in \{p_1,\dots,p_{m_n}\}$. To implement $\psi_n(\boldsymbol{\beta}_n^*)$, we determine the $\kappa_{n,p}$ for each $p \in \mathfrak{P}_n$ by making 100,000 draws of $\boldsymbol{Z}_d$ and set $\kappa_{n,p}$ equal to the empirical $(1-\alpha_{n,p})$-quantiles of $\|\boldsymbol{Z}_d\|_p$. Having obtained the $\kappa_{n,p}$, we obtain c_n as the empirical $(1-\alpha)$-quantile of $\max_{p\in\mathfrak{P}_n} \kappa_{n,p}^{-1}\|\boldsymbol{Z}_d\|_p$. Note that the role of c_n is to ensure that $\psi_n(\boldsymbol{\beta}_n^*)$ has asymptotic size α, cf. Part 1 of Theorem 4.1

below. Instead of simulating c_n, one can also choose $c_n = 1$. The latter will generally result in $\psi_n(\boldsymbol{\beta}_n^*)$ having asymptotic size not larger but potentially smaller than α, i.e., a conservative asymptotic level α test.

Theorem 4.1. *Let $\alpha \in (0,1)$ and suppose Assumption 3.1 is satisfied. Then, the following holds.*

1. *Size control: If $\boldsymbol{\beta}^* \in \mathbf{B}^{(0)}$, then $\lim_{n\to\infty} \mathbb{E}\psi_n(\boldsymbol{\beta}_n^*) = \alpha$.*
2. *Dominance: For any $p \in [2,\infty]$, $\bar{\alpha}_p \in (0,1)$, and $(\bar{\kappa}_{n,p}(\bar{\alpha}_p))_{n\in\mathbb{N}}$ given by (8) in case $p \in [2,\infty)$ and by (14) in case $p = \infty$ it holds that*

$$\mathbb{P}\left(S_{n,p}(\boldsymbol{\beta}_n^*) \geq \bar{\kappa}_{n,p}(\bar{\alpha}_p)\right) \to 1 \qquad \textit{implies} \qquad \mathbb{E}\psi_n(\boldsymbol{\beta}_n^*) \to 1.$$

Part 1 of Theorem 4.1 shows that the critical values $\kappa_{n,p}$, $p \in \mathfrak{P}_n$, and the c_n that guarantee exact size $\alpha \in (0,1)$ under Gaussianity in (16) also guarantee that $\psi_n(\boldsymbol{\beta}_n^*)$ has asymptotic size α. This may not be obvious ex-ante as the construction of $\psi_n(\boldsymbol{\beta}_n^*)$ is based on combining $|\mathfrak{P}_n| = (m_n + 2) \to \infty$ tests. However, by Lemma 3.2, this becomes a trivial consequence of writing non-rejection of $\psi_n(\boldsymbol{\beta}_n^*)$ as the event

$$\left\{\hat{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}_n^*)\boldsymbol{H}_n(\boldsymbol{\beta}_n^*) \in \bigcap_{p\in\mathfrak{P}_n} \mathbb{B}_p(c_n\kappa_{n,p})\right\}$$

where for $r \in (0,\infty)$ we set $\mathbb{B}_p(r) = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_p < r\}$ and observe that $\bigcap_{p\in\mathfrak{P}_n} \mathbb{B}_p(c_n\kappa_{n,p})$ is convex.

Part 2 shows that $\psi_n(\boldsymbol{\beta}_n^*)$ is consistent against *every* deviation from the null hypothesis that *some* p-norm based test with $p \in [2,\infty]$ is consistent against (irrespective of the asymptotic size of the latter, as long as it is a number in $(0,1)$). In particular, $\psi_n(\boldsymbol{\beta}_n^*)$ is consistent whenever tests based on the 2- or ∞-norm with asymptotic size in $(0,1)$ are consistent. Recall that the (dense and sparse) alternatives that tests based on these two norms are consistent against are the ones targeted by current applications of the power enhancement principle of Fan et al. (2015), cf. also Kock and Preinerstorfer (2019). In addition, $\psi_n(\boldsymbol{\beta}_n^*)$ is consistent against further alternatives, as it is also consistent as soon as there exists a $p \in (2, \infty)$ such that a test based on this p is consistent (a property that tests based on the power enhancement principle that combine 2- and ∞-norm based tests do not share, cf. Kock and Preinerstorfer (2023)).

Finally, Theorem I.1 in Section I.2 in the appendix makes precise a sense in which there are no alternatives against which a given p-norm based test has substantially higher asymptotic power than a suitably constructed test $\psi_n(\boldsymbol{\beta}_n^*)$.

Remark 4.1. The implementation of $\psi_n(\boldsymbol{\beta}_n^*)$ requires the choice of a set $\mathfrak{P}_n$ of p-norms to use in its construction. The simulations in Section 6 and the application in Section 7 suggest that the performance of $\psi_n(\boldsymbol{\beta}_n^*)$ is not particularly sensitive to the choice of $\mathfrak{P}_n$ as long as it includes p-norms *in addition* to $p = 2$ and $p = \infty$.

Remark 4.2. In Section I.4 of the supplementary appendix we consider, following a suggestion of a referee, variations of $\psi_n(\boldsymbol{\beta}_n)$ that replace $\mathfrak{P}_n$ by the union of $\{2, \infty\}$ with a countable and unbounded subset $\mathfrak{P}'$ of $(2, \infty)$, to obtain the test

$$\psi_n'(\boldsymbol{\beta}_n^*) := \mathbb{1}\left\{\sup_{p \in \mathfrak{P}'} \kappa_{n,p}^{-1} S_{n,p}(\boldsymbol{\beta}_n^*) > c_n\right\}.$$

For example, $\mathfrak{P}'$ could be chosen as the union of $\{2, \infty\}$ with all rational or natural numbers greater than 2. In Section I.4 of the supplementary appendix we establish that $\psi_n'(\boldsymbol{\beta}_n^*)$ has the same size and dominance properties as established for $\psi_n(\boldsymbol{\beta}_n^*)$ in Theorem 4.1. However, $\psi_n'(\boldsymbol{\beta}_n^*)$ is more challenging to compute than $\psi_n(\boldsymbol{\beta}_n^*)$, as $\psi_n'(\boldsymbol{\beta}_n^*)$ involves a supremum over an infinite set.

5 Many IVs and dominating the Anderson-Rubin test and sup-score type tests

Recall that in the context of the linear IV model in Section 2.1.3 one has

$$h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*) = (y_{1,n} - \mathbf{Y}_{1,n}'\boldsymbol{\beta}_n^*)\boldsymbol{z}_{1,n}.$$

The theory developed in the previous sections has (under suitable moment assumptions) the following consequences for testing whether $\boldsymbol{\beta}_n^*$ satisfies $\mathbb{E}(y_{1,n} - \boldsymbol{Y}_{1,n}'\boldsymbol{\beta}_n^*)\boldsymbol{z}_{1,n} = \mathbf{0}_d$:

1. Choosing $p = 2$ results in a classic weak identification robust Anderson-Rubin (AR) test. Thus, this AR test is dominated in terms of consistency by any p-norm based test with $p \in (2, \infty)$, cf. the relationship in (11). There is no ranking between tests based on $p \in [2, \infty)$ and the sup-score type test corresponding to $p = \infty$; see Belloni et al. (2012) for the definition of the original sup-score statistic.

2. For simplicity, let $\mathbf{B}_n = \mathbb{R}$ and assume that $\boldsymbol{\beta}_n$ satisfies $\mathbb{E}(y_{1,n} - \boldsymbol{\beta}_n\boldsymbol{Y}_{1,n})\boldsymbol{z}_{1,n} = \mathbf{0}_d$. Even if only one instrument (say) is relevant in the sense of $\mathbb{E}\boldsymbol{z}_{1,n}\boldsymbol{Y}_{1,n} = (a_n, 0, \dots, 0)'$ for some $a_n \neq 0$ such that the moments

$$\mathbb{E}(y_{1,n} - \boldsymbol{\beta}_n^*\boldsymbol{Y}_{1,n})\boldsymbol{z}_{1,n} = \mathbb{E}\boldsymbol{z}_{1,n}\boldsymbol{Y}_{1,n}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_n^*) = \big(a_n(\boldsymbol{\beta}_n - \boldsymbol{\beta}_n^*), 0, \dots, 0\big)'$$

only differ from zero in one entry, this sparsity need not be inherited by

$$\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*) = \sqrt{n}[\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)]^{-1/2}\big(a_n(\boldsymbol{\beta}_n - \boldsymbol{\beta}_n^*), 0, \dots, 0\big)'$$

unless one is willing to impose structure on the covariance matrix $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ of $(y_{1,n} - \boldsymbol{\beta}_n^*\boldsymbol{Y}_{1,n})\boldsymbol{z}_{1,n}$. Thus, as it is $\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)$ that determines the power properties of p-norm based tests, one cannot advise on which p to use solely based on the number of instruments that one suspects to be relevant. Similarly, one should be cautious basing advice on the sparsity structure of $\boldsymbol{\pi}_n$ in the "first-stage" $\boldsymbol{Y}_{1,n} = \boldsymbol{\pi}_n'\boldsymbol{z}_{1,n} + \nu_{1,n}$: Since $\boldsymbol{\pi}_n = [\mathbb{E}(\boldsymbol{z}_{1,n}\boldsymbol{z}_{1,n}')]^{-1}\mathbb{E}\boldsymbol{z}_{1,n}\boldsymbol{Y}_{1,n}$ (assuming that $\mathbb{E}(\boldsymbol{z}_{1,n}\boldsymbol{z}_{1,n}')$ is invertible) one can write $\mathbb{E}\boldsymbol{z}_{1,n}\boldsymbol{Y}_{1,n} = \mathbb{E}(\boldsymbol{z}_{1,n}\boldsymbol{z}_{1,n}')\boldsymbol{\pi}_n$ and hence

$$\begin{aligned}
\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*) &= \sqrt{n}[\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)]^{-1/2}\mathbb{E}(y_{1,n} - \boldsymbol{\beta}_n^*\boldsymbol{Y}_{1,n})\boldsymbol{z}_{1,n} \\
&= \sqrt{n}[\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)]^{-1/2}\mathbb{E}\boldsymbol{z}_{1,n}\boldsymbol{Y}_{1,n}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_n^*) \\
&= \sqrt{n}[\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)]^{-1/2}\mathbb{E}(\boldsymbol{z}_{1,n}\boldsymbol{z}_{1,n}')\boldsymbol{\pi}_n \cdot (\boldsymbol{\beta}_n - \boldsymbol{\beta}_n^*),
\end{aligned}$$

cf. the first equality in the penultimate display. Again the sparsity pattern of the first-stage $\boldsymbol{\pi}_n$ may not carry over to $\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)$; the latter determining the power properties of p-norm based tests.

3. In light of the previous point, it is difficult to give advice on which single p-norm to base a test on even if one is willing to impose assumptions on the number of relevant instruments. Thus, it is useful that the test $\psi_n(\boldsymbol{\beta}_n^*)$ in (17) is consistent whenever some p-norm based test (including the AR- or sup-score type tests) is consistent. Therefore, $\psi_n(\boldsymbol{\beta}_n^*)$ is consistent against strictly more alternatives than any test based on a single p. In this sense, $\psi_n(\boldsymbol{\beta}_n^*)$ does not rely on knowing whether $\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)$ is sparse or not.

6 Numerical results

In Section 6.1 we investigate numerically the properties of our test $\psi_n(\boldsymbol{\beta}_n^*)$ for hypothesis testing on a parameter in the presence of many instruments, cf. Sections 2.1.3 and 5. To investigate the effect of very large d without unduly increasing the computational burden, Section J in the appendix considers directly the limiting Gaussian testing problem which by Lemma 3.2 approximates the rejection frequencies of $\psi_n(\boldsymbol{\beta}_n^*)$ and p-norm based tests in many moment equality testing problems.

6.1 Testing in the presence of many instruments

We generate data from the following standard two-equation model with one endogenous regressor ($\mathbf{B}_n = \mathbb{R}$ for all $n \in \mathbb{N}$), but many instruments d:

$$y_{i,n} = \boldsymbol{Y}_{i,n}\boldsymbol{\beta}_n + u_{i,n}, \qquad \text{“Second stage”} \tag{18}$$

$$\boldsymbol{Y}_{i,n} = \boldsymbol{\pi}_n'\boldsymbol{z}_{i,n} + \nu_{i,n}, \qquad i = 1, \ldots, n. \qquad \text{“First stage”} \tag{19}$$

The variables are named and interpreted as in Section 2.1.3. All variables are i.i.d. across $i = 1, \ldots, n$ and we consider a setting where the instruments $\boldsymbol{z}_{i,n}$ and error terms have heavy tails in the sense that the d independent entries of $\boldsymbol{z}_{1,n}$ follow a $t(5)$-distribution and

$$\begin{pmatrix} u_{1,n} \\ \nu_{1,n} \end{pmatrix} \sim t_5\left(\mathbf{0}_2, \boldsymbol{\Xi}\right), \qquad \text{where} \quad \boldsymbol{\Xi} = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}.$$

In addition, $\boldsymbol{z}_{1,n} \perp\!\!\!\perp (u_{1,n}, \nu_{1,n})$. Thus, the "true" parameter $\boldsymbol{\beta}_n$ satisfies $\mathbb{E}(y_{1,n}-\boldsymbol{Y}_{1,n}\boldsymbol{\beta}_n)\boldsymbol{z}_{1,n} = \mathbb{E}u_{1,n}\boldsymbol{z}_{1,n} = \mathbf{0}_d$ and we test whether the candidate parameter $\boldsymbol{\beta}_n^*$ satisfies

$$H_{0,n} : \mathbb{E}(y_{1,n} - \boldsymbol{Y}_{1,n}\boldsymbol{\beta}_n^*)\boldsymbol{z}_{1,n} = \mathbf{0}_d \qquad \text{vs.} \qquad H_{1,n} : \mathbb{E}(y_{1,n} - \boldsymbol{Y}_{1,n}\boldsymbol{\beta}_n^*)\boldsymbol{z}_{1,n} \neq \mathbf{0}_d,$$

which falls within our general testing framework. Throughout we use $\boldsymbol{\beta}_n^* = 0$ and gauge the size and power of the tests considered by generating data for a range of $\boldsymbol{\beta}_n$. We consider $(n, d) \in \left\{(5{,}000, 100), (25{,}000, 500), (100{,}000, 1{,}000)\right\}$.

The number of "relevant" instruments equals the number of non-zero entries of $\boldsymbol{\pi}_n \in \mathbb{R}^d$, i.e., the sparsity/denseness of the "first stage". For each pair (n, d) we consider the sparsest and densest possible first stage, namely $\boldsymbol{\pi}_n = (1, 0, \ldots, 0)'$ and $\boldsymbol{\pi}_n = (1, \ldots, 1)'$. As the latter first stage is clearly more informative, we adjust the magnitude of the deviation of $\boldsymbol{\beta}_n$ from $\boldsymbol{\beta}_n^* = 0$ according to the "strength" of $\boldsymbol{\pi}_n$ to get non-trivial power curves.[7] Finally, we consider a setting of a semi-sparse first stage where a moderate number of instruments is relevant. Here $\boldsymbol{\pi}_n = (\boldsymbol{\iota}_{d_{R,n}}, 0, \ldots, 0)'$ where $\boldsymbol{\iota}_{d_{R,n}}$ is a (row) vector of $d_{R,n}$ ones and $d_{R,n} = 4, 7, 12$, for $n = 5{,}000, 25{,}000$ and $100{,}000$, respectively. We also note that a simple calculation reveals that in the present setting the number of non-zero entries of $\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*) = \sqrt{n}\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)\mathbb{E}h_n(\boldsymbol{X}_{1,n}, \boldsymbol{\beta}_n^*)$ equals that of $\boldsymbol{\pi}_n$ as $\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_n^*)$ is block-diagonal. Thus, when we vary the sparsity of $\boldsymbol{\pi}_n$, we identically vary that of $\boldsymbol{\theta}_n(\boldsymbol{\beta}_n^*)$.

Throughout, we estimate $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_n^*)$ by the sample covariance matrix of the vectors $(y_{i,n} - \boldsymbol{Y}_{i,n}\boldsymbol{\beta}_n^*)\boldsymbol{z}_{i,n}$, $i = 1, \ldots, n$. We also experimented with the median-of-means estimator in Section 3.4 of Ke et al. (2019), but (although being more precise for heavy-tailed distributions and allowing for larger d) this did not improve the performance of the resulting tests in the settings considered.

[7]Alternatively, one can adjust the magnitude of the entries of $\boldsymbol{\pi}_n$, i.e. change the strength of the individual instruments, with the number of non-zero entries of $\boldsymbol{\pi}_n$.

In order to investigate the sensitivity of $\psi_n(\boldsymbol{\beta}_n^*)$ in (17) to the choice of $\mathfrak{P}_n$, we implement it with $\mathfrak{P}_n^{(1)} = \{2, 3, 5, 10, \infty\}$, $\mathfrak{P}_n^{(2)} = \{2, 10, \infty\}$, $\mathfrak{P}_n^{(3)} = \{2, 3, 4, 5, 6, 7, 8, 9, 10, \infty\}$, and $\mathfrak{P}_n^{(4)} = \{2, \infty\}$. Here $\mathfrak{P}_n^{(2)}$ is included to investigate the effect of only including few (i.e., one) p-norm in addition to 2 and ∞ and $\mathfrak{P}_n^{(3)}$ is included to investigate the effect of including many additional p-norms. Finally, $\mathfrak{P}_n^{(4)}$ has been included to illustrate that combining only the 2- and ∞-norm leads to a power loss against semi-sparse alternatives. For each $j \in \{1, 2, 3, 4\}$, we use $\alpha = 0.05$, $\alpha_2 = \alpha_\infty = \alpha_{n,p} = \alpha/|\mathfrak{P}_n^j|$ for $p \in \mathfrak{P}_n^{(j)}$ and all n considered.[8] We also implemented the power enhancement based test of Fan et al. (2015), which combines the 2- and ∞-norm only, exactly as described in that reference (cf. pages 1503–1506).[9] The number of replications is 1,000 for $n = 5{,}000$ and $n = 25{,}000$, but 500 for $n = 100{,}000$. Figure 1 plots by how much the power of $\psi_n(\boldsymbol{\beta}_n^*)$ implemented with $\mathfrak{P}_n^{(1)}$ *exceeds* that of each of the other tests considered. Thus, when comparing to the 2-norm based test we plot the power difference $\mathbb{E}\psi_n(\boldsymbol{\beta}_n^*) - \mathbb{P}\left(S_{n,2}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,2}\right)$ and similarly for the other tests considered. Positive numbers mean that $\psi_n(\boldsymbol{\beta}_n^*)$ implemented with $\mathfrak{P}_n^{(1)}$ is superior and vice versa for negative numbers. The raw power function can be found in Figure I.1 in Section I.5 of the appendix. Note that these reveal that the supremum-norm based test often has slightly lower size than the remaining tests. Thus, the (large) power deficits of this test when the first stage is dense or semi-sparse should be interpreted in this light.[10] Figure 1 reveals:

- For each pair (n, d) (i.e., in each row) the test $\psi_n(\boldsymbol{\beta}_n^*)$ has power similar to that based on the supremum-norm when there is one relevant instrument (sparse first stage). The power can be much higher than that of the 2-norm and PE based test. When all instruments are relevant (dense first stage), $\psi_n(\boldsymbol{\beta}_n^*)$ has power only marginally

[8]For $j = 4$ this means that $\alpha_2 = \alpha_\infty = \alpha/2$ and $\alpha_I = 0$.

[9]For $n = 5{,}000$ we also implement the jacknifed AR test of Mikusheva and Sun (2022). For larger sample sizes the implementation time became prohibitive for the jacknifed AR test. Unreported simulations revealed that for sufficiently small n (relative to d) it had power similar to that of the 2-norm based test, but it controlled size better than the p-norm based tests that we study. This is in line with the theoretical results on size control of the jackknifed AR test in Mikusheva and Sun (2022).

[10]Section J in the supplementary appendix carries out simulations in the context of the Gaussian sequence model. There the critical values can be simulated to ensure that all tests considered have size exactly 0.05 (up to simulation error). These simulations confirm all the findings of the present section. Thus, the power differences in the present section do not seem to be due to (small) differences in the size of the tests.

lower than the 2-norm and PE based ones (which have identical power). The power of $\psi_n(\boldsymbol{\beta}_n^*)$ can be much higher than that of the supremum-norm based test. $\psi_n(\boldsymbol{\beta}_n^*)$ is more powerful than 2-norm, ∞-norm and PE based tests when the first-stage is semi-sparse. Throughout it does not matter which $\mathfrak{P}_n^{(j)}$ one implements $\psi_n(\boldsymbol{\beta}_n^*)$ with as long as one includes p-norms *additional* to 2 and ∞: As predicted by our theoretical results, choosing $\mathfrak{P}_n^{(4)} = \{2, \infty\}$ leads to a power loss against semi-sparse alternatives (but still outperforms the PE based test).

- As n and d increase (i.e., in each column) all the above power differences become more pronounced. For $(n, d) = (100{,}000, 1{,}000)$ the gain of $\psi_n(\boldsymbol{\beta}_n)$ over the 2-norm and PE based test is up to 0.67 and 0.52, respectively, for the sparse first stage. The gain is up to 0.83 over the ∞-norm based test for the dense first stage. For the semi-sparse first stage and $(n, d) = (100{,}000, 1{,}000)$, the power advantage of $\psi_n(\boldsymbol{\beta}_n^*)$ is 0.22 over the 2-norm and PE based test (which have identical power in this setting) whereas the power gain over the supremum-norm based test is 0.20.

In Section J of the appendix we shall see that the power gains from using $\psi_n(\boldsymbol{\beta}_n^*)$ instead of a 2- or supremum-norm based test (or a power enhancement based combination of these) become even larger in higher dimensions.

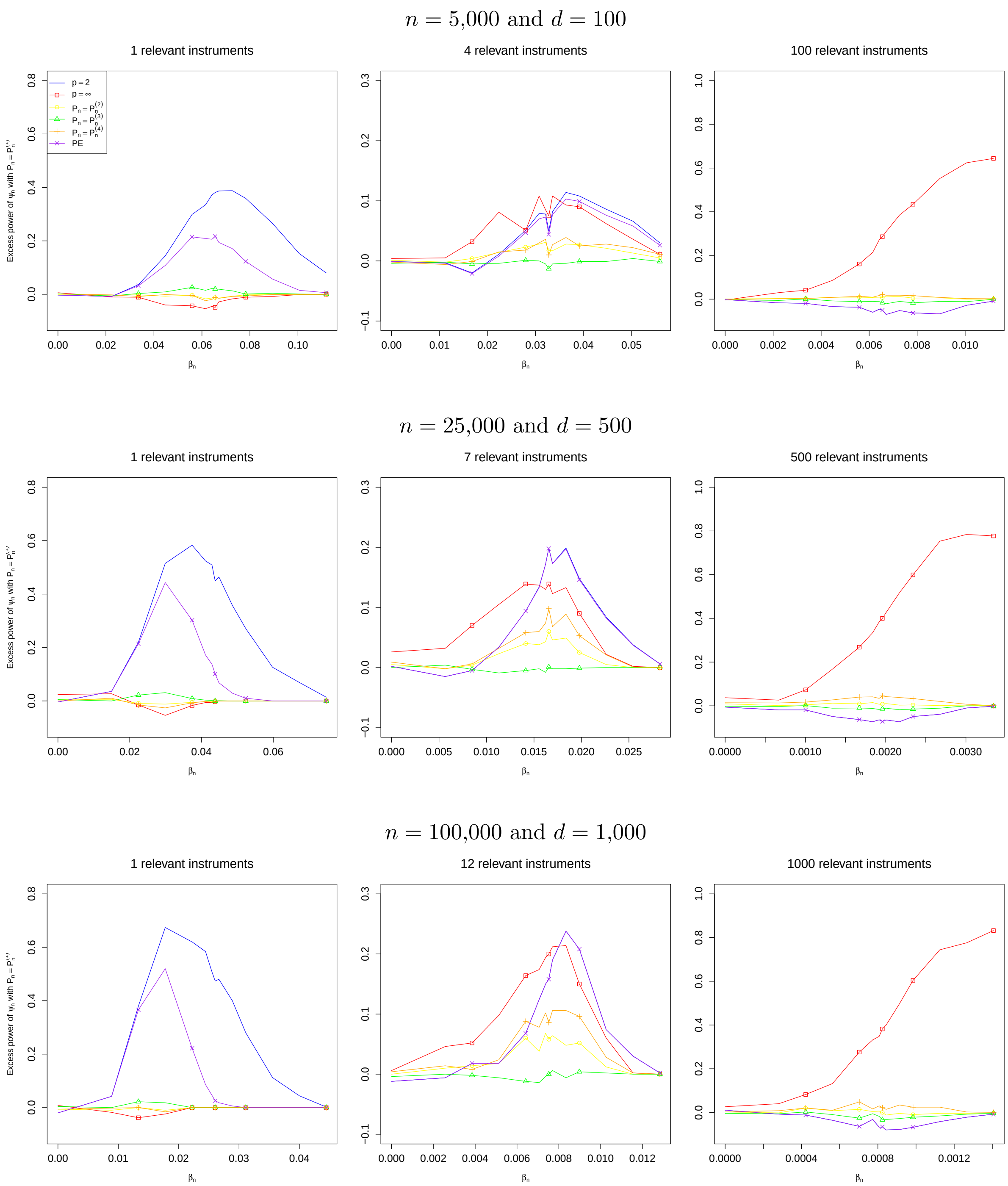

Figure 1: Excess power of $\psi_n(\boldsymbol{\beta}_n^*)$ implemented with $\mathfrak{P}_n^{(1)}$ over each of the other tests studied for $(n, d) = (5{,}000, 100)$ [first row], $(25{,}000, 500)$ [second row], $(100{,}000, 1{,}000)$ [third row] for sparse [first column], semi-sparse [second column] and dense [third column] alternatives. PE is the power enhancement principle of Fan et al. (2015). Full implementation details are given in the body text.

7 Application: Returns to schooling

In this section we illustrate the tests studied in the context of the the two-equation IV model in (18) and (19) with $y_{i,n}$ being the logarithm of earnings and $\boldsymbol{Y}_{i,n}$ being years of schooling. Since, e.g., innate ability/intelligence or motivation are difficult to measure and control for, these are captured by the error term $u_{i,n}$. Furthermore, educational attainment $\boldsymbol{Y}_{i,n}$ is likely correlated with these and hence correlated with $u_{i,n}$. Thus, regressing $y_{i,n}$ on $\boldsymbol{Y}_{i,n}$ by OLS results in a biased estimator. This endogeneity problem is often tackled by seeking instruments $\boldsymbol{z}_{i,n}$ that correlate with educational attainment $\boldsymbol{Y}_{i,n}$ but do not otherwise affect wages $y_{i,n}$. In the context of (19) this implies $\boldsymbol{\pi}_n \neq \mathbf{0}_d$.

Using the U.S. National Longitudinal Survey of Young Men (NLSYM) as in Card (1995), we utilize that growing up nearby a 2- or 4- year college (two binary variables) is positively correlated with educational attainment and plausibly uncorrelated with unobserved ability factors such as intelligence. Furthermore, following again Card (1995) we use that the educational attainment of the father and mother is likely related to one's own educational attainment. For each parent we construct (orthogonalized) monomials of orders one to four and interact these to create $4 \times 4 = 16$ instruments. Finally, we use whether or not the family had a library card at home when the individual was 14 as an instrument. Interacting all these instrumental variables creates a total of $2 \cdot 2 \cdot 4 \cdot 4 \cdot 2 = 128$ instruments. Dropping observations with missing values results in a sample size of $n = 2{,}033$. These dimensions are of the same order of magnitude as in the setting $(n, d) = (5{,}000, 100)$ studied for this two-equation model in the simulations in Section 6 (where all tests achieved the desired size).

We construct confidence intervals with nominal coverage 95% for the effect $\boldsymbol{\beta}_n$ of schooling $\boldsymbol{Y}_{i,n}$ on the logarithm of wages $y_{i,n}$ by test inversion as detailed for p-norm based tests in Section I.1 of the appendix.[11] To this end, the tests are implemented exactly as in the simulations in Section 6.1. The confidence intervals can be found in Figure 2 and we note that in this application:

1. The confidence intervals based on all implementations of $\psi_n(\boldsymbol{\beta}_n^*)$ are about as short

[11] The analogue constructions based on $\psi_n(\boldsymbol{\beta}_n^*)$ and the PE based test are obvious.

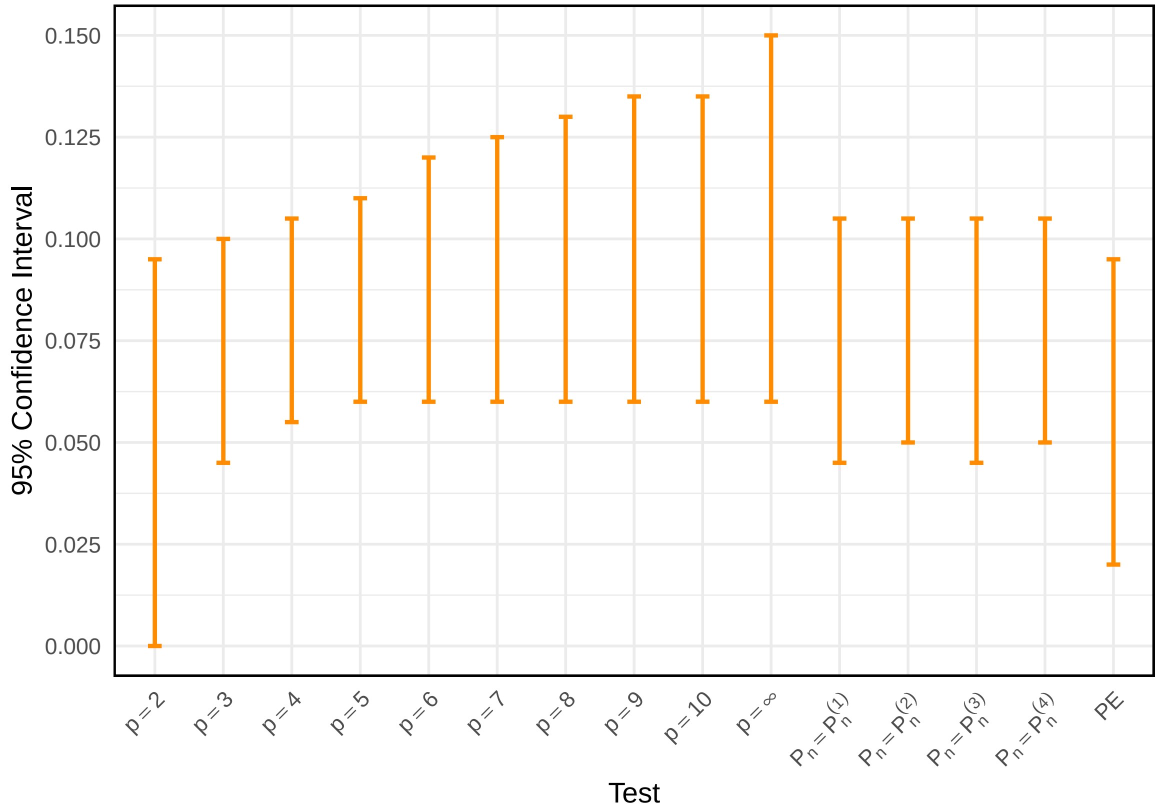

Figure 2: Confidence intervals for $\boldsymbol{\beta}_n$ based on test inversion as outlined in Section I.1 of the appendix. $p = j$ for $j \in \{2, \ldots, 10, \infty\}$ corresponds to the confidence interval based on inverting $\mathbb{1}\left\{S_{n,j}(\boldsymbol{\beta}_n^*) \geq \kappa_{n,j}\right\}$ and $P_n = P_n^{(k)}$ for $k \in \{1, \ldots, 4\}$ corresponds to the confidence interval based on inverting $\psi_n(\boldsymbol{\beta}_n)$ implemented with $\mathfrak{P}_n = \mathfrak{P}_n^{(k)}$. Finally, PE is the confidence interval resulting from inverting the test based on the power enhancement principle in Fan et al. (2015).

as those based on the 4- and 5-norm based tests which are the individual p-norms resulting in the shortest confidence intervals.

2. In accordance with the simulations in Section 6.1, the size of the confidence intervals constructed via $\psi_n(\boldsymbol{\beta}_n)$ are not particularly sensitive to the choice of $\mathfrak{P}_n$.

3. The 2- and ∞-norm based tests result in the longest confidence intervals. This also explains why the confidence interval based on the PE principle, which combines these norms only, is longer than those combining additional p-norms.

Overall, this exhibits an example where enhanced power enhancements incorporating p-norms targeting semi-sparse alternatives lead to a substantial improvement.

References

Abdalla, P. and N. Zhivotovskiy (2024): "Covariance estimation: Optimal dimension-free guarantees for adversarial corruption and heavy tails," *Journal of the European Mathematical Society (forthcoming).*

Anatolyev, S. and N. Gospodinov (2011): "Specification testing in models with many instruments," *Econometric Theory*, 27, 427–441.

Andrews, D. W. and J. H. Stock (2007): "Testing with many weak instruments," *Journal of Econometrics*, 138, 24–46.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012): "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, 80, 2369–2429.

Bentkus, V. (2003): "On the dependence of the Berry–Esseen bound on dimension," *Journal of Statistical Planning and Inference*, 113, 385–402.

Bickel, P. and E. Levina (2008): "Covariance regularization by thresholding," *Annals of Statistics*, 36, 2577–2604.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2022): "When is tsls actually late?" Tech. rep., National Bureau of Economic Research.

Boot, T. and J. W. Ligtenberg (2023): "Identification- and many instrument-robust inference via invariant moment conditions," *arXiv preprint arXiv:2303.07822*.

Cai, T., W. Liu, and X. Luo (2011): "A constrained ℓ_1 minimization approach to sparse precision matrix estimation," *Journal of the American Statistical Association*, 106, 594–607.

Cai, T., Z. Ren, and H. Zhou (2016): "Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation," *Electronic Journal of Statistics*, 10, 1–59.

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

Cheng, C., Y. Ke, and W. Zhang (2025): "Large Precision Matrix Estimation with Unknown Group Structure," *Journal of the American Statistical Association*, 1–24.

Crudu, F., G. Mellace, and Z. Sándor (2021): "Inference in instrumental variable models with heteroskedasticity and many instruments," *Econometric Theory*, 37, 281–310.

Davey Smith, G. and S. Ebrahim (2003): ""Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease?" *International journal of epidemiology*, 32, 1–22.

Dovì, M.-S., A. B. Kock, and S. Mavroeidis (2024): "A Ridge-regularised jackknifed Anderson-Rubin test," *Journal of Business & Economic Statistics*, 42, 1083–1094.

Fan, J., Y. Liao, and H. Liu (2016): "An overview of the estimation of large covariance and precision matrices," *The Econometrics Journal*, 19, C1–C32.

Fan, J., Y. Liao, and J. Yao (2015): "Power enhancement in high-dimensional cross-sectional tests," *Econometrica*, 83, 1497–1541.

Fan, Q., Z. Guo, and M. Ziwei (2024): "A heteroskedasticity-robust overidentifying restriction test with high-dimensional covariates," *Journal of Business & Economic Statistics (to appear)*.

Fang, X. and Y. Koike (2024): "Large-dimensional central limit theorem with fourth-moment error bounds on convex sets and balls," *Annals of Applied Probability*, 34, 2065–2106.

Feng, L., T. Jiang, B. Liu, and W. Xiong (2022): "Max-sum tests for cross-sectional independence of high-dimensional panel data," *Annals of Statistics*, 50, 1124 – 1143.

Gabaix, X. and R. S. Koijen (2020): "Granular instrumental variables," Tech. rep., National Bureau of Economic Research.

Ge, S., S. Li, and O. Linton (2024): "Dynamic peer groups of arbitrage characteristics," *Journal of Business & Economic Statistics*, 42, 367–390.

He, Y., G. Xu, C. Wu, and W. Pan (2021): "Asymptotically independent U-statistics in high-dimensional testing," *Annals of Statistics*, 49, 154–181.

Horn, R. A. and C. R. Johnson (1985): *Matrix analysis*, Cambridge University Press.

Huang, Y., C. Li, R. Li, and S. Yang (2022): "An overview of tests on high-dimensional means," *Journal of Multivariate Analysis*, 188, 104813.

Ingster, Y. and I. Suslina (2003): *Nonparametric goodness-of-fit testing under Gaussian models*, vol. 169, Springer Science & Business Media.

Jammalamadaka, S. R., S. Meintanis, and T. Verdebout (2020): "On Sobolev tests of uniformity on the circle with an extension to the sphere," *Bernoulli*, 26, 2226–2252.

Juodis, A. and S. Reese (2022): "The incidental parameters problem in testing for remaining cross-section correlation," *Journal of Business & Economic Statistics*, 40, 1191–1203.

Kaffo, M. and W. Wang (2017): "On bootstrap validity for specification testing with many weak instruments," *Economics Letters*, 157, 107–111.

Ke, Y., S. Minsker, Z. Ren, Q. Sun, and W.-X. Zhou (2019): "User-friendly covariance estimation for heavy-tailed distributions," *Statistical Science*, 34, 454–471.

Kock, A. B. and D. Preinerstorfer (2019): "Power in High-Dimensional Testing Problems," *Econometrica*, 87, 1055–1069.

——— (2023): "Consistency of p-norm based tests in high dimensions: Characterization, monotonicity, domination," *Bernoulli*, 29, 2544–2573.

——— (2024): "A remark on moment-dependent phase transitions in high-dimensional Gaussian approximations," *Statistics & Probability Letters*.

Koltchinskii, V. and K. Lounici (2017): "Concentration inequalities and moment bounds for sample covariance operators," *Bernoulli*, 110–133.

Li, D., R. Li, and H. L. Shang (2024a): "Detection and estimation of structural breaks in high-dimensional functional time series," *The Annals of Statistics*, 52, 1716–1740.

Li, D., L. Xue, H. Yang, and X. Yu (2024b): "Power-Enhanced Two-Sample Mean Tests for High-Dimensional Compositional Data with Application to Microbiome Data Analysis," *arXiv preprint arXiv:2405.02551*.

Liu, Y., W. Sun, A. P. Reiner, C. Kooperberg, and Q. He (2019): "Statistical inference of genetic pathway analysis in high dimensions," *Biometrika*, 106, 651–664.

Lugosi, G. and S. Mendelson (2019): "Mean estimation and regression under heavy-tailed distributions: A survey," *Foundations of Computational Mathematics*, 19, 1145–1190.

Matsushita, Y. and T. Otsu (2021): "Second-order refinements for t-ratios with many instruments," *Journal of Econometrics*, 232, 346–366.

——— (2022): "A jackknife Lagrange multiplier test with many weak instruments," *Econometric Theory*.

Mendelson, S. and N. Zhivotovskiy (2020): "Robust covariance estimation under $L\_4 - L\_2$ norm equivalence," *Annals of Statistics*, 48, 1648–1664.

Mikusheva, A. (2021): "Many weak instruments in time series econometrics," .

Mikusheva, A. and L. Sun (2022): "Inference with many weak instruments," *The Review of Economic Studies*, 89, 2663–2686.

——— (2024): "Weak identification with many instruments," *Econometrics Journal*.

Oliveira, R. and Z. Rico (2024): "Improved covariance estimation: optimal robustness and sub-gaussian guarantees under heavy tails," *Annals of Statistics*, 52, 1953–1977.

Tang, Y., H. J. Wang, and E. Barut (2018): "Testing for the presence of significant covariates through conditional marginal regression," *Biometrika*, 105, 57–71.

Tchuente, G. and M. Carrasco (2016): "Regularization based Anderson Rubin tests for many instruments," .

Vershynin, R. (2012): "How close is the sample covariance matrix to the actual covariance matrix?" *Journal of Theoretical Probability*, 25, 655–686.

Xu, G., L. Lin, P. Wei, and W. Pan (2016): "An adaptive two-sample test for high-dimensional means," *Biometrika*, 103, 609–624.

Yang, Q. and G. Pan (2017): "Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices," *Journal of the American Statistical Association*, 112, 188–200.

Yu, X., D. Li, and L. Xue (2024a): "Fisher's combined probability test for high-dimensional covariance matrices," *Journal of the American Statistical Association*, 119, 511–524.

Yu, X., D. Li, L. Xue, and R. Li (2023): "Power-enhanced simultaneous test of high-dimensional mean vectors and covariance matrices with application to gene-set testing," *Journal of the American Statistical Association*, 118, 2548–2561.

Yu, X., J. Yao, and L. Xue (2024b): "Power enhancement for testing multi-factor asset pricing models via Fisher's method," *Journal of Econometrics*, 239.

Zhang, D. and W. B. Wu (2017): "Gaussian approximation for high dimensional time series," *Annals of Statistics*, 45, 1895–1919.

Zhang, Y., R. Wang, and X. Shao (2022): "Adaptive inference for change points in high-dimensional data," *Journal of the American Statistical Association*, 117, 1751–1752.