

# Consistency of $p$ -norm based tests in high dimensions: characterization, monotonicity, domination\*

Anders Bredahl Kock

University of Oxford

CREATES, Aarhus University

10 Manor Rd, Oxford OX1 3UQ

[anders.kock@economics.ox.ac.uk](mailto:anders.kock@economics.ox.ac.uk)

David Preinerstorfer

ECARES, SBS-EM

Université libre de Bruxelles

50 Ave. F.D. Roosevelt, Brussels

[david.preinerstorfer@ulb.be](mailto:david.preinerstorfer@ulb.be)

First version: March 2021

This version: May 2021

## Abstract

Many commonly used test statistics are based on a norm measuring the evidence against the null hypothesis. To understand how the choice of a norm affects power properties of tests in high dimensions, we study the *consistency sets* of  $p$ -norm based tests in the prototypical framework of sequence models with unrestricted parameter spaces, the null hypothesis being that all observations have zero mean. The consistency set of a test is here defined as the set of all arrays of alternatives the test is consistent against as the dimension of the parameter space diverges. We characterize the consistency sets of  $p$ -norm based tests and find, in particular, that the consistency against an array of alternatives cannot be determined solely in terms of the  $p$ -norm of the alternative. Our characterization also reveals an unexpected monotonicity result: namely that the consistency set is strictly increasing in  $p \in (0, \infty)$ , such that tests based on higher  $p$  strictly dominate those based on lower  $p$  in

---

\*We would like to thank Davy Paindaveine, Michael Wolf, and seminar participants at Turku and Paris for helpful comments and discussions. David Preinerstorfer gratefully acknowledges support from the Program of Concerted Research Actions (ARC) of the Université libre de Bruxelles.

terms of consistency. This monotonicity allows us to construct novel tests that dominate, with respect to their consistency behavior, *all*  $p$ -norm based tests without sacrificing size.

## 1 Introduction

Since the advent of the Big Data era, inference in high-dimensional models has received considerable attention in economics. As illustrated by the works of [Belloni and Chernozhukov \(2011\)](#), [Belloni et al. \(2012, 2014, 2017\)](#), [Farrell \(2015\)](#), [Kock and Callot \(2015\)](#), [Athey \(2017\)](#), [Athey et al. \(2018\)](#), [Onatski and Wang \(2018\)](#) or [Gentzkow et al. \(2019\)](#), many contributions have been made in a wide range of subareas of econometrics spanning regression problems, program evaluation and causal inference, estimation of treatment effects, stationary and non-stationary time series, up to the theory of choice. Giving a complete summary of the literature is hardly possible any longer as there are not many areas in which the effects of dimension has not been studied.

A central first question that arises in the context of large amounts of data is whether the data contain a signal. In more formal terms, one is interested in testing the global null hypothesis of no effects, typically being formulated as the high-dimensional target parameter being null. One very interesting aspect concerning that testing problem is that there is a genuine difference between low- and high-dimensional regimes in terms of the consistency behavior of tests. In regular low-dimensional problems the consistency properties of tests are by now well-understood. In particular, one can typically construct tests that are consistent against any non-contiguous deviation from the null hypothesis, cf. Chapter 6.7 in [Roussas \(1972\)](#) and Chapter 14 in [van der Vaart \(2000\)](#).<sup>1</sup> The situation changes drastically when the dimension of the parameter space diverges with the sample size. Here, the choice of a test statistic typically determines the type of alternatives it is (not) consistent against. This led [Fan et al. \(2015\)](#) to put forward their power enhancement princi-

---

<sup>1</sup>Note that precisely for this reason, the classic Hájek-Le Cam theory has focused on the comparison of tests by means of their power properties against contiguous (“local”) sequences of alternatives, cf. [Le Cam \(1960\)](#) and [Hájek \(1972\)](#).

ple, a combination procedure that combines two tests into a better test, that is, a test that is consistent against more alternatives while having the same asymptotic size. Based on this principle, [Kock and Preinerstorfer \(2019\)](#) established that in generic high-dimensional parametric testing problems the following holds: for *any* test of asymptotic size strictly less than one, the power enhancement principle can be used to construct another test with (i) the same asymptotic size, (ii) uniformly non-inferior power and (iii) which is consistent against strictly more alternatives than the initially given test. Hence, every test has blind spots that can be removed without sacrificing asymptotic size. Put differently, this also implies that there cannot exist a test with “optimal” consistency properties.

Many commonly used tests either directly build on or are related to  $p$ -norms. For example, Wald-type tests are widely used in applied and theoretical work. These tests build on measuring whether a properly normalized Euclidean (i.e., 2-) norm of the parameter vector exceeds a critical value. Furthermore, tests based on the supremum-norm have received considerable attention in recent years due to the particular role attributed to sparse alternatives in high-dimensional testing problems. Additionally, procedures related to the 1-norm have also been widely studied at least since the introduction of the Lasso by [Tibshirani \(1996\)](#). In any case, the consistency properties of a test for a high-dimensional testing problem crucially depend on which norm it is based on. Note that the results in [Kock and Preinerstorfer \(2019\)](#) do not rule out the possibility that there exists a best test *within* the family of all  $p$ -norm based tests.

It is therefore interesting to investigate the whole family of  $p$ -norm based tests in detail and ask:

- Does one obtain more powerful tests by using  $p$ -norms other than the 1-, Euclidean or supremum-norm, e.g., the 4-norm?
- Is it possible to rank  $p$ -norm based tests in terms of their power properties?
- Is there a test that simultaneously dominates all  $p$ -norm based tests?

Despite *many* commonly used tests being based on a  $p$ -norm, the exact consistency

properties of such tests in modern high-dimensional testing problems have not been systematically studied. One goal of the present article is to do so — and the resulting ranking of tests in terms of their consistency properties may be somewhat surprising.

## 1.1 Contributions

In the prototypical context of a sequence model, the present paper studies the consistency properties of  $p$ -norm based tests. To reject the “global-null” of all observations having zero mean, these tests gauge whether the  $p$ -norm of the observation vector exceeds a given critical value. Our contributions are as follows:

- *Characterization:* We characterize the *consistency set* of each  $p$ -norm based test, that is the set of arrays of alternatives that the test is consistent against. Somewhat surprisingly, we find that whether the  $p$ -norm based test is consistent against an array of alternatives cannot be determined solely by the  $p$ -norm of the alternative (which would in general only lead to sufficient but not necessary conditions for consistency). For  $p \in (2, \infty)$  it actually suffices that the  $p$ -norm or the Euclidean norm of the alternative is sufficiently large.
- *Monotonicity:* For  $0 < p < q < \infty$  the  $q$ -norm based test is consistent against any array of alternative that the  $p$ -norm based test is consistent against — and more. This finding sheds light on the consequences of employing tests based on the commonly used 1- and Euclidean norms, as these are strictly dominated in terms of the size of their consistency sets. We also identify the structure of the additional alternatives that a test based on a larger power is consistent against. This strict dominance relation does not extend to  $p = \infty$ .
- *Domination:* Although no single  $p$ -norm based test dominates all other ones in terms of the size of its consistency set, we construct a single test that is consistent against any array of alternatives that any  $p$ -norm based test is consistent against — and more. Thus, it is possible to *simultaneously* dominate all members of the commonly used class of  $p$ -norm based tests. Finally, we

elucidate some consequences and relations to minimax adaptive testing of our findings.

## 1.2 Related Work

Our study of consistency properties of  $p$ -norm based tests is related to the work of [Pinelis \(2010\)](#) and [Pinelis \(2014\)](#), in which the local asymptotic relative efficiency (ARE) of  $p$ -norm based tests to that of 2-norm based tests is studied. In terms of ARE, and in contrast to our findings on consistency sets, there is no unequivocal ranking of  $p$ -norm based tests, but interesting Schur<sup>2</sup>-convexity/concavity properties are exposed. These imply that tests based on norms with large powers are relatively more powerful against sparse alternatives than 2-norm based tests and vice versa for dense alternatives; they also suggest that tests based on powers slightly higher than 2 may be preferable in terms of ARE. Bootstrap approximations of  $p$ -norm based test statistics in high dimensions have recently been studied in [Giessing and Fan \(2020\)](#).

Minimax properties of  $p$ -norm based tests in the high-dimensional Gaussian sequence model are discussed in detail in the classic monograph of [Ingster and Suslina \(2003\)](#) in their Chapter 3. These results concern power properties against restricted alternatives. For unrestricted parameter spaces their results provide sufficient conditions for consistency, which, however, turn out not to be necessary in general, and thus do not lead to a characterization of the consistency set of  $p$ -norm based tests. We revisit the results given in [Ingster and Suslina \(2003\)](#) in Section 6 and establish some new adaptation results in the minimax framework building on our results obtained in earlier sections.

## 2 Framework

### 2.1 Model and testing problem

We consider a sequence model

$$y_{i,d} = \theta_{i,d} + \varepsilon_i, \quad i = 1, \dots, d, \tag{1}$$

where the  $\theta_{i,d} \in \mathbb{R}$  are unknown parameters, the unobserved error terms  $\varepsilon_i$  have mean zero and are i.i.d., and where  $y_{1,d}, \dots, y_{d,d}$  are the observations. *For simplicity of presentation, and since it is the most important case, we shall assume throughout the main text of this article that  $\varepsilon_i \sim \mathbb{N}(0, 1)$ .* All results generalize appropriately to non-normal errors under suitable assumptions concerning the tail of the distribution of  $\varepsilon_i$ . Such results are established in Appendix B.

For notational simplicity, we write  $\mathbf{y}_d = (y_{1,d}, \dots, y_{d,d})$ ,  $\boldsymbol{\varepsilon}_d = (\varepsilon_1, \dots, \varepsilon_d)$ , and  $\boldsymbol{\theta}_d = (\theta_{1,d}, \dots, \theta_{d,d}) \in \mathbb{R}^d$ , upon which we may write (1) equivalently as  $\mathbf{y}_d = \boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d$ .

In the sequence model (1), we study properties of tests for the testing problem

$$H_{0,d} : \boldsymbol{\theta}_d = \mathbf{0}_d \quad \text{against} \quad H_{1,d} : \boldsymbol{\theta}_d \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}, \quad (2)$$

where  $\mathbf{0}_d = (0, \dots, 0) \in \mathbb{R}^d$ . That is, we are concerned with what is sometimes referred to as the “global null” of no effect. Oftentimes, a test for this hypothesis is done in the initial stage of a whole series of potential further investigations that are carried out only if the global null is rejected. It is crucial to use a test with good power properties at the initial stage.

The sequence model (1) is an often used prototypical framework in high-dimensional statistics. On the one hand, the simplicity of the model allows one to focus on asymptotic power properties as the dimension  $d$  diverges to  $\infty$  in a clean framework. On the other hand, the framework assumes away certain complications, such as the estimation of nuisance (e.g., variance) parameters, or the interplay between sample size and the dimensionality of the parameter vector, which we do not address.

We also note, that the results derived in a Gaussian sequence model immediately correspond to results in a Gaussian regression model by a sufficiency argument detailed (for completeness) in the following remark. This is a model of fundamental importance for applications.

*Remark 2.1.* Let  $\mathbf{X}_d \in \mathbb{R}^{n \times d}$  be a matrix of full column rank, in particular implying  $d \leq n$ . In the Gaussian linear regression model  $\mathbf{z}_d = \mathbf{X}_d \boldsymbol{\beta}_d + \mathbf{u}_d$  with  $\mathbf{u}_d \sim \mathbb{N}(\mathbf{0}_d, \mathbf{I}_d)$ , the OLS estimator  $\hat{\boldsymbol{\beta}}_d$  is sufficient for  $\boldsymbol{\beta}_d$  and, denoting  $\mathbf{M}_d := (\mathbf{X}_d' \mathbf{X}_d)^{1/2}$ ,

the same holds for  $\mathbf{M}_d \hat{\boldsymbol{\beta}}_d$ , which satisfies

$$\mathbf{M}_d \hat{\boldsymbol{\beta}}_d := \mathbf{M}_d \boldsymbol{\beta}_d + \mathbf{M}_d^{-1} \mathbf{X}'_d \mathbf{u}_d \quad \text{with} \quad \mathbf{M}_d^{-1} \mathbf{X}'_d \mathbf{u}_d \sim \mathbb{N}(\mathbf{0}_d, \mathbf{I}_d),$$

which is clearly of the form (1). Note also that  $\boldsymbol{\theta}_d := \mathbf{M}_d \boldsymbol{\beta}_d = \mathbf{0}_d$  if and only if  $\boldsymbol{\beta}_d = \mathbf{0}_d$ . That is, this problem is statistically equivalent to testing (2) in a (re-parameterized) Gaussian sequence model.

*Remark 2.2.* On a *conceptual and informal* level, results obtained in Gaussian sequence models carry over much more broadly to any situation where the distribution of a properly standardized estimator is approximately Gaussian. The reasoning follows along similar lines as in Remark 2.1: Consider a situation where an estimator  $\hat{\boldsymbol{\beta}}_d$  for a target parameter  $\boldsymbol{\beta}_d \in \mathbb{R}^d$  is available the distribution of which satisfies

$$\hat{\boldsymbol{\beta}}_d \approx \mathbb{N}(\boldsymbol{\beta}_d, \boldsymbol{\Omega}_d).$$

Suppose further that an invertible estimator  $\hat{\boldsymbol{\Omega}}_d \approx \boldsymbol{\Omega}_d$  is available, such that

$$\hat{\boldsymbol{\Omega}}_d^{-1/2} \hat{\boldsymbol{\beta}}_d \approx \mathbb{N}(\boldsymbol{\Omega}_d^{-1/2} \boldsymbol{\beta}_d, \mathbf{I}_d).$$

Then, testing  $\boldsymbol{\beta}_d = \mathbf{0}$  on the basis of  $\hat{\boldsymbol{\beta}}_d$  and  $\hat{\boldsymbol{\Omega}}_d$  is approximated by testing  $\boldsymbol{\theta}_d := \boldsymbol{\Omega}_d^{-1/2} \boldsymbol{\beta}_d = \mathbf{0}$  in a Gaussian sequence model. Precise sets of conditions under which the above approximation statements hold depend on the interplay of the dimension of the target parameter and sample size and on particularities of the specific setup under consideration.

## 2.2 Tests, asymptotic size, and consistency sets

Suppose that for every  $d \in \mathbb{N}$  we are given a (possibly randomized) test  $\varphi_d$  for the testing problem (2). That is,  $\varphi_d$  is a (Borel measurable) function from  $\mathbb{R}^d$  to  $[0, 1]$ . We denote the set of all such sequences of tests  $\{\varphi_d\}_{d \in \mathbb{N}}$  by  $\mathbb{T}$ . With some abuse of notation, we shall often abbreviate  $\{\varphi_d\}_{d \in \mathbb{N}}$  by  $\varphi_d$ , to which we then also refer to as a sequence of tests or just as a test. It will always be clear from the context whether

a sequence of tests or a test for a fixed  $d$  is meant.

We say that the sequence of tests  $\varphi_d \in \mathbb{T}$  has *asymptotic size*  $\alpha \in [0, 1]$  if and only if

$$\lim_{d \rightarrow \infty} \mathbb{E}(\varphi_d(\boldsymbol{\varepsilon}_d)) = \alpha. \quad (3)$$

The subset of all sequences of tests with asymptotic size  $\alpha$  is denoted by  $\mathbb{T}_\alpha \subseteq \mathbb{T}$ . We will mostly be concerned with  $\alpha \in (0, 1)$ .

We consider an asymptotic framework where the coordinates of the parameter vector  $\boldsymbol{\theta}_d = (\theta_{1,d}, \dots, \theta_{d,d})$  depend on  $d$ . That is, we consider (uniform) power properties along triangular arrays of alternatives  $\boldsymbol{\vartheta} = \{\boldsymbol{\theta}_d : d \in \mathbb{N}\}$ , where  $\boldsymbol{\theta}_d \in \mathbb{R}^d$  for every  $d \in \mathbb{N}$ . The set of all such triangular arrays will be denoted by  $\Theta := \times_{d=1}^{\infty} \mathbb{R}^d$ . As is common, we say that the sequence of tests  $\varphi_d$  is *consistent* against an array  $\boldsymbol{\vartheta} = \{\boldsymbol{\theta}_d : d \in \mathbb{N}\} \in \Theta$  if and only if

$$\lim_{d \rightarrow \infty} \mathbb{E}(\varphi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d)) = 1. \quad (4)$$

To any sequence of tests  $\varphi_d \in \mathbb{T}$  we associate its *consistency set*  $\mathcal{C}(\varphi_d) \subseteq \Theta$ , i.e., the subset of all arrays of alternatives  $\boldsymbol{\vartheta} \in \Theta$  that the sequence of tests  $\varphi_d$  is consistent against. The characterization and comparison of consistency sets is the main focus of the present article.

## 2.3 Optimal consistency properties

For a given  $\alpha \in (0, 1)$ , one can compare the quality of two sequences of tests  $\psi_d \in \mathbb{T}_\alpha$  and  $\varphi_d \in \mathbb{T}_\alpha$  on the basis of their consistency sets.<sup>2</sup> Note that, by definition,  $\mathcal{C}(\psi_d) \supseteq \mathcal{C}(\varphi_d)$  if and only if  $\psi_d$  is consistent against *all* alternatives that  $\varphi_d$  is consistent against. If this is the case,  $\psi_d$  weakly dominates  $\varphi_d$  in terms of consistency, whereas  $\psi_d$  strongly dominates  $\varphi_d$  in terms of consistency if even  $\mathcal{C}(\psi_d) \supsetneq \mathcal{C}(\varphi_d)$ .

In search for an “optimal” test, one would hope that there exists a single sequence of tests  $\psi_d \in \mathbb{T}_\alpha$  that is consistent against all alternatives that some  $\varphi_d \in \mathbb{T}_\alpha$  is

---

<sup>2</sup>In principle, one can also compare two sequences of tests with different asymptotic sizes by comparing their consistency sets, but the comparison is less meaningful in such situations since the consistency sets could depend on the asymptotic size.



consistent against, i.e.,  $\psi_d$  satisfies

$$\mathcal{C}(\psi_d) \supseteq \bigcup \{ \mathcal{C}(\varphi_d) : \varphi_d \in \mathbb{T}_\alpha \}. \quad (5)$$

Such a test  $\psi_d$  (in case it exists) would be “consistency-optimal” in the sense that no other test with the same asymptotic size exists that is consistent against more alternatives.

Whereas the optimality property in (5) is often achieved by the usual candidates of tests in standard finite-dimensional testing problems, the consistency properties of tests are much more delicate in high-dimensional testing problems, where a  $\psi_d \in \mathbb{T}_\alpha$  satisfying (5) typically does *not* exist. This is illustrated in the following result, which can be obtained by a similar reasoning as in Section 1.1 of [Kock and Preinerstorfer \(2019\)](#), to which we refer for further discussions and results in the context of models that are locally asymptotically normal.

**Theorem 2.1.** *Let  $\alpha \in (0, 1)$ . For every  $\psi_d \in \mathbb{T}_\alpha$  there exists a  $\bar{\psi}_d \in \mathbb{T}_\alpha$ , such that:*

1.  $\bar{\psi}_d \geq \psi_d$  holds for every  $d \in \mathbb{N}$ , guaranteeing that  $\bar{\psi}_d$  has uniformly non-inferior asymptotic power compared to  $\psi_d$ ;
2.  $\bar{\psi}_d$  is consistent against an array of alternatives against which  $\psi_d$  has asymptotic power at most  $\alpha$ .

*In particular it holds that  $\mathcal{C}(\psi_d) \subsetneq \mathcal{C}(\bar{\psi}_d)$ .*

Theorem 2.1 establishes that the consistency set of *any* test (of asymptotic size  $\alpha \in (0, 1)$ ) can be strictly improved.<sup>3</sup> Inspection of the proof, which largely builds on an argument in Section 3.4.2 of [Ingster and Suslina \(2003\)](#), shows that the improvement is achieved through the power enhancement principle of [Fan et al. \(2015\)](#). Hence, no consistency-optimal test as fancied in the second paragraph of this section and (5) in particular can exist.

---

<sup>3</sup>Actually one can show that any test can be strictly improved against a highly sparse array of alternatives; cf. Theorem A.1 in Appendix A. In this sense, not even the most stringent sparsity assumptions overturn the conclusion of Theorem 2.1.

The literature mainly offers two solutions in situations where (for a given optimality criterion) no optimal test exists: The first is to restrict one’s attention to a certain subclass of alternatives. Restricting the class of alternatives is an approach that is often employed in high-dimensional and nonparametric problems, e.g., by focusing on sparse alternatives or all alternatives for which a given norm exceeds a certain threshold. One then explores and constructs tests that work well for such alternatives (but do perhaps not work so well for others). This approach then typically studies minimax rates of detection, and minimax consistent tests. We offer more discussion and provide some results in that direction in Section 6, but that is not our main focus.

A different approach is to leave the set of alternatives unrestricted as is, but instead focus on a specific class of tests, and to study optimality for this restricted class of tests. In essence, one fixes a subset  $\mathbb{T}_\alpha^* \subseteq \mathbb{T}_\alpha$  and looks for a test whose consistency set contains the consistency sets of all tests in  $\mathbb{T}_\alpha^*$ ; that is one searches for a  $\psi_d \in \mathbb{T}_\alpha$  that satisfies (5) but with  $\mathbb{T}_\alpha$  replaced by  $\mathbb{T}_\alpha^*$ . Whether or not a test  $\psi_d$  with the desired property exists clearly depends on  $\mathbb{T}_\alpha^*$ . In the present article we explore this question for an important subclass of tests, namely  $p$ -norm based tests.

## 2.4 $p$ -norm based tests

For  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$  and  $p \in (0, \infty]$ , we define

$$\|\mathbf{x}\|_p := \begin{cases} \left(\sum_{i=1}^d |x_i|^p\right)^{\frac{1}{p}} & \text{if } p < \infty, \\ \max_{i=1, \dots, d} |x_i| & \text{else.} \end{cases} \quad (6)$$

For  $p \geq 1$  the function  $\|\cdot\|_p$  defines a norm on  $\mathbb{R}^d$ ; whereas for  $p \in (0, 1)$  it only defines a quasinorm on  $\mathbb{R}^d$ . For the sake of brevity, we will refer to  $\|\cdot\|_p$  as the “ $p$ -norm” also in case  $p \in (0, 1)$ . Given a radius  $r \in [0, \infty]$ , we denote  $\mathbb{B}_p^d(r) := \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y}\|_p \leq r\}$ , i.e., the closed “ball” with respect to  $\|\cdot\|_p$  of radius  $r$  centered at the origin.

It is well known that the likelihood-ratio test for (2) rejects if  $\|\mathbf{y}_d\|_2$ , the Euclidean

norm of the vector of observations  $\mathbf{y}_d$ , exceeds a critical value (chosen to satisfy a given size constraint). More generally, any  $p$ -norm delivers a test for (2), i.e., one rejects the null hypothesis if  $\|\mathbf{y}_d\|_p$  exceeds some critical value. Given  $p \in (0, \infty]$  and a sequence of critical values  $\kappa_d$ , we abbreviate the sequence of tests  $\mathbb{1}\{\|\cdot\|_p \geq \kappa_d\}$  by  $\{p, \kappa_d\}$ , and refer to such a test as a  $p$ -norm based test. Consequently, the consistency set of such a sequence of tests is written as  $\mathcal{C}(\{p, \kappa_d\})$ .

By a classical result of Birnbaum (1955) and Stein (1956), tests that reject if the  $p$ -norm of  $\mathbf{y}_d$  exceeds a given critical value are *admissible* for every  $d \in \mathbb{N}$  and every  $p \in [1, \infty]$ ; the reason is that their acceptance region is convex. Furthermore, the tests just described are all *unbiased*, for every  $d \in \mathbb{N}$ , due to Anderson's theorem, cf. Anderson (1955). Hence, all  $p$ -norm based tests with  $p \in [1, \infty]$  are reasonable from a non-asymptotic point of view.

Following the Neyman-Pearson approach, we shall mostly be interested in the situation where the critical values  $\kappa_d$  are chosen such that the asymptotic size of  $\{p, \kappa_d\}$  is in  $(0, 1)$ , i.e.,

$$\lim_{d \rightarrow \infty} \mathbb{P}(\|\boldsymbol{\varepsilon}_d\|_p \geq \kappa_d) = \alpha \in (0, 1). \quad (7)$$

One has that (7) is true if and only if

$$\kappa_d = \begin{cases} \left[ (\Phi^{-1}(1 - \alpha) + o(1)) \sqrt{d} \sigma_p + d \mu_p \right]^{1/p} & \text{if } p \in (0, \infty), \\ \sqrt{2 \log(d)} - \frac{\log \log(d) + \log(4\pi) + 2 \log(-\log(1 - \alpha)/2) + o(1)}{2\sqrt{2 \log(d)}} & \text{if } p = \infty. \end{cases} \quad (8)$$

Here  $\Phi$  denotes the cdf of the standard normal distribution and for  $p \in (0, \infty)$  we abbreviated  $\sigma_p^2 := \text{Var}(|\varepsilon_1|^p)$  and  $\mu_p := \mathbb{E}(|\varepsilon_1|^p)$ . The equality in (8) follows since for  $p \in (0, \infty)$  the test statistic converges in distribution under the null (upon suitable centering and scaling) to the standard normal distribution, whereas for  $p = \infty$  it converges to a slight modification (only taking care of absolute values) of Gumbel's double exponential distribution (cf. Lemma B.3 and Lemma A.2 for formal statements).

### 3 Consistency sets of $p$ -norm based tests

We shall now derive a characterization of  $\mathcal{C}(\{p, \kappa_d\})$  for sequences  $\kappa_d$  satisfying (7). This characterization will in particular allow us to compare the consistency sets for different  $p \in (0, \infty]$ . The present section is divided into the cases  $p \in (0, \infty)$  and  $p = \infty$ , which lead to fundamentally different results.

Characterizing the consistency set of a  $p$ -norm based test requires us to establish a necessary and sufficient condition for the test to be consistent against an array of alternatives. For  $p$ -norm based tests a set of sufficient conditions is also given in Ingster and Suslina (2003), Sections 3.1.2–3.1.4. However, these conditions are not necessary. That is, there are arrays  $\boldsymbol{\vartheta}$  against which a  $p$ -norm based test is consistent, but which do not satisfy the sufficient condition for consistency there.

#### 3.1 Case $p \in (0, \infty)$

For any  $p \in \mathbb{R}$ , denote by  $g_p : \mathbb{R} \rightarrow \mathbb{R}$  the function

$$g_p(x) := x^2 \mathbf{1}_{[-1,1]}(x) + |x|^p \mathbf{1}_{\mathbb{R} \setminus [-1,1]}(x). \quad (9)$$

In particular,  $g_p(x) = x^2 \wedge |x|^p$  for  $p \in (0, 2)$  and  $g_p(x) = x^2 \vee |x|^p$  for  $p \in [2, \infty)$ , where for two real numbers  $x$  and  $y$  we write  $\min(x, y) = x \wedge y$  and  $\max(x, y) = x \vee y$  for readability.

The consistency set of a  $p$ -norm based test with non-trivial asymptotic size is characterized next.

**Theorem 3.1.** *For  $p \in (0, \infty)$  and  $\{p, \kappa_d\} \in \mathbb{T}_\alpha$ ,  $\alpha \in (0, 1)$ , we have*

$$\boldsymbol{\vartheta} \in \mathcal{C}(\{p, \kappa_d\}) \iff \frac{\sum_{i=1}^d g_p(\theta_{i,d})}{\sqrt{d}} \rightarrow \infty. \quad (10)$$

It is a somewhat surprising aspect of Theorem 3.1 that, apart from the case  $p = 2$ , the consistency set  $\mathcal{C}(\{p, \kappa_d\})$  cannot be entirely characterized in terms of the asymptotic behavior of the  $p$ -norm of the elements of an array  $\boldsymbol{\vartheta} \in \Theta$ . Regardless of  $p$ ,

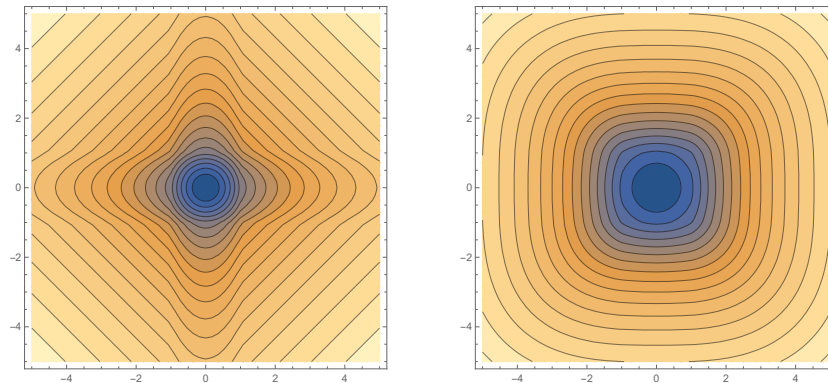


Figure 1: Contour plots of the function characterizing the consistency of the  $p$ -norm based test for  $d = 2$  and  $p = 1$  (left panel) and  $p = 3$  (right panel).

coordinates of  $\theta_d$  that are small in absolute value enter the consistency criterion in (10) via their squares, whereas coordinates with large absolute values enter differently and in dependence on  $p$ . We illustrate this in Figure 1 by showing contour plots of the level sets of the function  $(x_1, x_2) \mapsto g_p(x_1)/\sqrt{2} + g_p(x_2)\sqrt{2}$  for  $p = 1, 3$ , which are genuinely different from the level sets of the corresponding  $p$ -norm in dimension  $d = 2$ . Note that the level sets close to the origin are circular irrespective of the value of  $p$ , i.e., in correspondence to level sets of a Euclidean norm, whereas level sets further away from the origin approach those of the  $p$ -norm the test is based on.

Theorem 3.1 follows from a more general result that also holds in non-Gaussian settings and which is given in Theorem B.4 of Appendix B. All results in the present Section 3.1 carry over to this more general setting, cf. the discussion after the proof of Theorem B.4.

*Remark 3.1.* There is nothing special about the interval  $[-1, 1]$  in the definition of  $g_p$  in Equation (9). In principle, one could replace the interval  $[-1, 1]$  in the definition of  $g_p$  by any interval  $[-M, M]$ , for a fixed  $M > 0$ , and the statement in (10) would still be correct with this re-defined function  $g_p^{(M)}$ , say, in place of  $g_p$ . This follows

immediately from

$$0 < \inf_{x \in \mathbb{R} \setminus \{0\}} \frac{g_p^{(M)}(x)}{g_p(x)} \leq \sup_{x \in \mathbb{R} \setminus \{0\}} \frac{g_p^{(M)}(x)}{g_p(x)} < \infty.$$

We have chosen  $M = 1$  in the formulation of Theorem 3.1 for concreteness and because it is the most convenient choice for later use in the proof of Theorem 3.3. The content of this remark will be instrumental in establishing Theorem 3.4 further below.

We emphasize (also for later use in Section 5) that the condition on the right in Equation (10) does not depend on  $\alpha$ . That is, as long as the sequence of critical values  $\kappa_d$  is chosen such that the asymptotic size of the corresponding test equals  $\alpha \in (0, 1)$  (cf. also (8)), the consistency set does not depend on the concrete value of  $\alpha$ .

An immediate consequence of Theorem 3.1 is the following observation, which we shall make use of later. In particular, it contains an equivalent way of writing (10) in case  $p \in [2, \infty)$  which further clarifies that the consistency of the  $p$ -norm based test against any array of alternatives cannot be settled based on the sequence  $\|\boldsymbol{\theta}_d\|_p$  alone. In fact, it is necessary and sufficient that  $\|\boldsymbol{\theta}_d\|_p^p/\sqrt{d}$  or  $\|\boldsymbol{\theta}_d\|_2^2/\sqrt{d}$  tends to infinity. The condition  $\|\boldsymbol{\theta}_d\|_p^p/\sqrt{d} \rightarrow \infty$ , which is also given in Ingster and Suslina (2003) Corollary 3.6, is only sufficient.

**Corollary 3.2.** *For  $p \in (0, \infty)$  and  $\{p, \kappa_d\} \in \mathbb{T}_\alpha$ ,  $\alpha \in (0, 1)$ , we have:*

1. *If  $p \in [2, \infty)$ , then*

$$\boldsymbol{\vartheta} \in \mathcal{C}(\{p, \kappa_d\}) \Leftrightarrow d^{-1/2} \left( \|\boldsymbol{\theta}_d\|_2^2 \vee \|\boldsymbol{\theta}_d\|_p^p \right) \rightarrow \infty. \quad (11)$$

2. *If  $p \in (0, 2)$ ,*

$$\boldsymbol{\vartheta} \in \mathcal{C}(\{p, \kappa_d\}) \Rightarrow d^{-1/2} \left( \|\boldsymbol{\theta}_d\|_2^2 \wedge \|\boldsymbol{\theta}_d\|_p^p \right) \rightarrow \infty; \quad (12)$$

*but the condition on the right in (12) does not imply the condition to the left.*

We shall now highlight a “monotonicity” result concerning the consistency set of  $p$ -norm based tests, which can be obtained from Theorem 3.1. This result shows that in terms of consistency properties, and choosing among  $p$ -norm based tests with  $p \in (0, \infty)$  and asymptotic size in  $(0, 1)$ , it is best to choose  $p$  “large.”<sup>4</sup>

**Theorem 3.3.** *For sequences of tests  $\{p, \kappa_{d,p}\}$  and  $\{q, \kappa_{d,q}\}$  with asymptotic sizes in  $(0, 1)$ , and  $0 < p < q < \infty$ , we have*

$$\mathcal{C}(\{p, \kappa_{d,p}\}) \subsetneq \mathcal{C}(\{q, \kappa_{d,q}\}).$$

In other words, the larger  $p \in (0, \infty)$ , the larger the consistency set of  $\{p, \kappa_{d,p}\}$ . Somewhat unfortunately, however, the theorem shows — at the same time — that it is impossible to choose  $p$  “large enough.” That tests based on larger powers have larger consistency sets may appear contradictory in light of the inequality  $\|\cdot\|_p \geq \|\cdot\|_q$  for  $1 \leq p \leq q$ , which implies that the test statistic gets *smaller* for large powers. Note, however, that this is compensated by the critical values, cf. (8). We finally emphasize that the asymptotic sizes of the two sequences of tests in Theorem 3.3 do not need to be identical. This is a consequence of the independence of the consistency set of  $\alpha \in (0, 1)$ , cf. Theorem 3.1.

Returning to the discussion in Section 2.3, the monotonicity statement established in Theorem 3.3 immediately raises the question whether there exists a test of asymptotic size  $\alpha \in (0, 1)$  that is consistent against all sequences of alternatives that any  $p$ -norm based test is consistent against; that is:

*Question 1.* Given  $\alpha \in (0, 1)$ , does there exist a  $\psi_d \in \mathbb{T}_\alpha$  such that

$$\mathcal{C}(\{p, \kappa_d\}) \subseteq \mathcal{C}(\psi_d) \tag{13}$$

for every  $p \in (0, \infty)$  and every sequence  $\kappa_d$  such that  $\{p, \kappa_d\} \in \mathbb{T}_\alpha$ ?

---

<sup>4</sup>Because the statement of Theorem 3.3 involves tests based on different powers and the sequences of critical values effecting asymptotic size control depend on the power, we here explicitly indicate this dependence simply to distinguish the two sequences. A similar convention will be applied in the statements below when needed.

*Remark 3.2.* It follows immediately from Theorem 3.3 that the requirement on  $\psi_d$  in (13) of Question 1 can equivalently be replaced by

$$\mathcal{C}(\{p, \kappa_d\}) \subsetneq \mathcal{C}(\psi_d); \quad (14)$$

that is, if a test as in Question 1 exists, then it must not only contain the consistency set of every  $p$ -norm based test ( $p \in (0, \infty)$ ) as a subset, but this set inclusion must actually be strict for every  $p$ . In this sense, if a test  $\psi_d$  as in Question 1 exists, it *strictly* dominates all  $p$ -norm based tests in terms of their consistency behavior.

We will come back to Question 1 in later sections. There, it will also be shown that despite the monotonicity established in Theorem 3.3 for  $p \in (0, \infty)$ , the supremum-norm based (i.e.,  $p = \infty$ ) test does not have the property of  $\psi_d$  sought in Question 1.

In the context of Theorem 3.3 one may ask whether the arrays of alternatives that the  $q$ -norm based test but not the  $p$ -norm based test is consistent against have particular structural properties. Thus, we next study the structure of

$$\mathcal{C}(\{q, \kappa_{d,q}\}) \setminus \mathcal{C}(\{p, \kappa_{d,p}\}) \quad \text{for } p < q. \quad (15)$$

### 3.1.1 Structure of elements of (15)

Let  $2 \leq p < q < \infty$  and fix critical values such that  $\{p, \kappa_{d,p}\}$  and  $\{q, \kappa_{d,q}\}$  have asymptotic sizes in  $(0, 1)$ . Then Corollary 3.2 shows that  $\boldsymbol{\vartheta}$  is an element of the set in (15) if and only if (i) there exists a subsequence along which  $d^{-1/2} \left( \|\boldsymbol{\theta}_d\|_2^2 \vee \|\boldsymbol{\theta}_d\|_p^p \right)$  converges to a real number and (ii) if along any such subsequence it holds that  $d^{-1/2} \|\boldsymbol{\theta}_d\|_q^q \rightarrow \infty$ . A concrete example of such an array is  $\boldsymbol{\vartheta} = \{(d^{\frac{1}{2p}}, 0, \dots, 0) : d \in \mathbb{N}\}$ , i.e., sparse alternatives with a dominating coordinate diverging to  $\infty$  at an appropriate rate. That all elements  $\boldsymbol{\vartheta}$  in the set in (15) are approximately sparse and highly unbalanced, at least along suitably chosen subsequences, will be shown next; the proof makes use of the observation in Remark 3.1.

**Theorem 3.4.** *Let  $0 < p < q < \infty$ , and let  $\{p, \kappa_{d,p}\}$  and  $\{q, \kappa_{d,q}\}$  have asymptotic sizes in  $(0, 1)$ . Then, for every  $\boldsymbol{\vartheta} \in \mathcal{C}(\{q, \kappa_{d,q}\}) \setminus \mathcal{C}(\{p, \kappa_{d,p}\})$  the following holds:*



1. For every  $\delta \in (0, \infty)$  there exists a subsequence  $d'$  of  $d$ , such that

$$\sup_{d'} \frac{\sum_{i=1}^{d'} \mathbb{1}\{|\theta_{i,d'}| > \delta\}}{\sqrt{d'}} < \infty \quad \text{and} \quad \max_{1 \leq i \leq d'} |\theta_{i,d'}| \rightarrow \infty. \quad (16)$$

2. If  $p \geq 2$ , then there exists a subsequence  $d'$  of  $d$ , such that

$$\sup_{\delta \in (0, \infty)} \delta^p \times \sup_{d'} \frac{\sum_{i=1}^{d'} \mathbb{1}\{|\theta_{i,d'}| > \delta\}}{\sqrt{d'}} < \infty \quad \text{and} \quad \max_{1 \leq i \leq d'} |\theta_{i,d'}| \rightarrow \infty. \quad (17)$$

The first part of Theorem 3.4 shows that those arrays  $\boldsymbol{\vartheta} \in \Theta$  that are consistently detected by some  $p$ -norm based test, but only if  $p$  is chosen large enough, are, along a subsequence, (i) “approximately sparse,” in that they have a vanishing fraction of entries larger than a given  $\delta > 0$ , and (ii) “highly unbalanced,” in the sense that

$$\min_{1 \leq i \leq d} |\theta_{i,d}| / \max_{1 \leq i \leq d} |\theta_{i,d}| \approx 0.$$

Furthermore, the second part of Theorem 3.4 shows that in case  $\boldsymbol{\vartheta}$  is not consistently detected by a  $p$ -norm based test with  $p \geq 2$  (but is detected by some  $q$ -norm based test with  $q > p$ ), a bit more can be said concerning (i). Namely that then the fraction of entries larger than a given  $\delta > 0$  decays (at least) at the order  $\delta^{-p}$ , uniformly over a *common* subsequence  $d'$ .

### 3.2 Case $p = \infty$

We shall now present a result that characterizes the consistency set of tests  $\{\infty, \kappa_d\} \in \mathbb{T}_\alpha$ ,  $\alpha \in (0, 1)$ . Structurally, the result is comparable to the one given in Theorem 3.1. Define

$$g_\infty(x) := \begin{cases} w(x) & \text{if } x < 1 \\ e^{-x^2/2}/x & \text{else,} \end{cases} \quad (18)$$

for a fixed continuous function  $w : \mathbb{R} \rightarrow (0, \infty)$  that satisfies  $w(z) \rightarrow \infty$  as  $z \rightarrow -\infty$ ; e.g.,  $w(z) = e^{(z^2 - 2z)/2}$ , in which case  $g_\infty$  is also continuous and strictly decreasing.

The result is as follows, where we recall that we denote the standard Gaussian cdf by  $\Phi$  and set  $\bar{\Phi} := 1 - \Phi$ . Its proof is based on a result applicable beyond the Gaussian setting which can be found in Proposition B.6 in Appendix B.2.

**Theorem 3.5.** *For  $\kappa_d$  such that  $\{\infty, \kappa_d\} \in \mathbb{T}_\alpha$ ,  $\alpha \in (0, 1)$ , we have*

$$\vartheta \in \mathcal{C}(\{\infty, \kappa_d\}) \Leftrightarrow \sum_{i=1}^d \frac{\bar{\Phi}(\mathbf{c}_d - |\theta_{i,d}|)}{\Phi(\mathbf{c}_d - |\theta_{i,d}|)} \rightarrow \infty \Leftrightarrow \sum_{i=1}^d g_\infty(\mathbf{c}_d - |\theta_{i,d}|) \rightarrow \infty, \quad (19)$$

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

First of all, we note that the sequence  $\mathbf{c}_d$  in (19) plays the role of centering. This is in contrast to (10), where the function  $g_p$  is symmetric about 0 and standardization is achieved by multiplying by  $1/\sqrt{d}$ .

In analogy to the discussion after Theorem 3.1 (cf. also Figure 1), we observe that Theorem 3.5 shows that the consistency set of a supremum-norm based test is not characterized exclusively by the supremum-norm of the deviation from the null hypothesis. Figure 2 further clarifies this by showing the contour sets for the function  $(x_1, x_2) \mapsto \bar{\Phi}(\mathbf{c}_2 - |x_1|)/\Phi(\mathbf{c}_2 - |x_1|) + \bar{\Phi}(\mathbf{c}_2 - |x_2|)/\Phi(\mathbf{c}_2 - |x_2|)$ . We note that the contour sets deviate from the ones of a supremum-norm, in particular along the “diagonals” or close to the origin.

Theorem 3.5 also reveals that as long as the critical values  $\kappa_d$  are chosen such that the asymptotic size of a supremum-norm based test is in  $(0, 1)$ , the consistency set remains the same, i.e., does not depend on the concrete asymptotic size. Although this already follows from the first equivalence in Theorem 3.5, the second equivalence statement has the advantage that it is easier to use in order to check whether a particular  $\vartheta \in \Theta$  is in  $\mathcal{C}(\{\infty, \kappa_d\})$ .

Concerning “dense” alternatives, i.e., arrays of the form  $\{\tau_d \boldsymbol{\nu}_d : d \in \mathbb{N}\}$  for a real sequence  $\tau_d$  and  $\boldsymbol{\nu}_d$  denoting the vector of ones of length  $d$ , Theorem 3.5 proves that such alternatives are in  $\mathcal{C}(\{\infty, \kappa_d\})$  if and only if  $\sqrt{\log(d)}|\tau_d| \rightarrow \infty$ ; see Appendix A.5 for details. Furthermore, Theorem 3.5 (together with the condi-

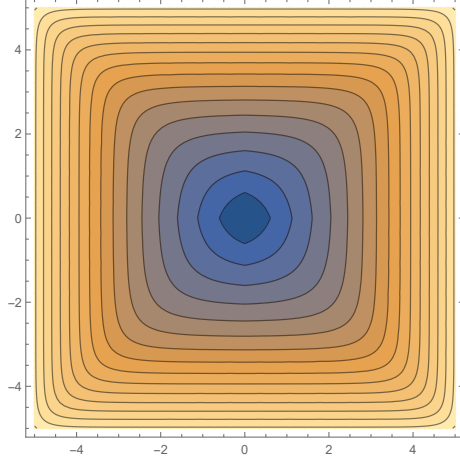


Figure 2: Contour plots of the function characterizing the consistency of the supremum-norm based test for  $d = 2$ .

tions assumed on  $w$ ) immediately implies the well-known fact that  $\boldsymbol{\vartheta} \in \mathcal{C}(\{\infty, \kappa_d\})$  if  $\|\boldsymbol{\theta}_d\|_\infty - \sqrt{2 \log(d)} \rightarrow \infty$ ; see [Ingster and Suslina \(2003, p. 92\)](#).

The discussion in the previous paragraph shows that the supremum-norm based test with asymptotic size  $\alpha$  is consistent against the array  $\{(\sqrt{3 \log(d)}, 0, \dots, 0) : d \in \mathbb{N}\}$ . In contrast, it follows from [Corollary 3.2](#) that this is not the case for any  $p$ -norm based test for  $p \in (0, \infty)$ . This echoes the conventional wisdom that supremum-norm based tests “are more powerful against sparse alternatives.” On the other hand, [Corollary 3.2](#) shows that for  $p \in (0, \infty)$  all  $p$ -norm based tests are consistent against the array  $\{\boldsymbol{\iota}_d / \sqrt{\log(d)} : d \in \mathbb{N}\}$ . The above discussion implies, however, that this is not the case for the supremum-norm based test. Hence, there exists no “ranking” of the consistency sets of the  $p$ -norm based tests for  $p \in (0, \infty)$  and the one of the supremum-norm based test. In particular, the monotonicity result in [Theorem 3.3](#) does *not* extend to  $\infty$ . We summarize this observation:

**Corollary 3.6.** *Let  $p \in (0, \infty)$ . Then, for  $\{p, \kappa_{d,p}\}$  and  $\{\infty, \kappa_{d,\infty}\}$  with asymptotic sizes in  $(0, 1)$ , it holds that*

$$\mathcal{C}(\{p, \kappa_{d,p}\}) \not\subseteq \mathcal{C}(\{\infty, \kappa_{d,\infty}\}) \quad \text{and} \quad \mathcal{C}(\{\infty, \kappa_{d,\infty}\}) \not\subseteq \mathcal{C}(\{p, \kappa_{d,p}\}). \quad (20)$$

Corollary 3.6 implies in particular that supremum-norm based tests do not have the property desired in Question 1.

While the previous section focused on the consistency set for  $p$ -norm based tests with  $p$  fixed, we conclude this section with a result concerning the consistency set of  $p$ -norm based tests with  $p$  growing with  $d$ . The following theorem relates the resulting consistency set to the one of the supremum-norm based test, and will be instrumental in the proof of Theorem 5.1 below.

**Theorem 3.7.** *Let the sequences  $p_d \in (0, \infty)$  and  $\kappa_d$  be such that the sequence of tests  $\{p_d, \kappa_d\} := \mathbb{1}\{\|\cdot\|_{p_d} \geq \kappa_d\}$  is in  $\mathbb{T}_\alpha$ ,  $\alpha \in (0, 1)$ . Under the condition that*

$$\liminf_{d \rightarrow \infty} \frac{p_d}{\log(d)} > 2, \quad (21)$$

*it holds that*

$$\mathcal{C}(\{\infty, \kappa_{d,\infty}\}) \subseteq \mathcal{C}(\{p_d, \kappa_d\}) \quad (22)$$

*for every sequence of critical values  $\kappa_{d,\infty}$  such that  $\{\infty, \kappa_{d,\infty}\}$  has asymptotic size in  $(0, 1)$ .*

The proof of Theorem 3.7 uses the fact that  $\mathcal{C}(\{\infty, \kappa_d\})$  does not depend on the concrete value of the asymptotic size, as long as this size is in  $(0, 1)$ , which we established in Theorem 3.5. Furthermore, it uses a “squeezing-argument” based on results in Bogachev (2006). That article provides asymptotic approximations for  $p_d$ -norm based test statistics under the null and for sequences  $p_d \rightarrow \infty$  (under weak assumptions on the sequences, which necessitate the squeezing argument), cf. also Schlather (2001) and Janßen (2009) for related results. A general version of Theorem 3.7 not assuming Gaussianity is provided in Proposition B.7 and Remark B.4 in Appendix B.2. Although Theorem 3.7 proves that  $p_d$ -norm based tests weakly dominate supremum-norm based tests, the discussion in Section 4.5 of Giessing and Fan (2020) reveals that tests based on a sequence of powers  $p_d$  with an asymptotic behavior as in the theorem just given do not have good power properties against dense alternatives.

## 4 Power enhancements and related combination procedures

To maximize power against sparse alternatives, it is often suggested to use a test based on the supremum-norm. At the other extreme, the typical choice maximizing power against dense alternatives is the likelihood-ratio test. Therefore, one could hope that the supremum-norm based test and the likelihood-ratio test together “suffice” in the sense that whenever some  $p$ -norm based test is consistent against  $\boldsymbol{\vartheta}$ , the supremum-norm based test or the likelihood-ratio test is consistent against  $\boldsymbol{\vartheta}$  as well. That is, one may conjecture that  $\mathcal{C}(\{2, \kappa_{d,2}\}) \cup \mathcal{C}(\{\infty, \kappa_{d,\infty}\})$  contains

$$\bigcup_{p \in (2, \infty)} \mathcal{C}(\{p, \kappa_{d,p}\}); \quad (23)$$

all critical values being chosen so that the corresponding asymptotic sizes are in  $(0, 1)$ . The fact established in Theorem 3.4 that for any  $q \in (2, \infty)$  all elements of the set  $\mathcal{C}(\{q, \kappa_{d,q}\}) \setminus \mathcal{C}(\{2, \kappa_{d,2}\})$  are approximately sparse along a subsequence may make such a conjecture appear even more plausible, since these are the types of alternatives against which the supremum-norm based test has particularly good power properties. Based on Theorems 3.1 and 3.5 we now show, however, that the likelihood-ratio test and the supremum-norm based test do *not* suffice. Specifically, we show that

$$\boldsymbol{\vartheta}^\dagger \in \bigcap_{p \in (2, \infty)} \mathcal{C}(\{p, \kappa_{d,p}\}) \setminus [\mathcal{C}(\{2, \kappa_{d,2}\}) \cup \mathcal{C}(\{\infty, \kappa_{d,\infty}\})], \quad (24)$$

where  $\boldsymbol{\vartheta}^\dagger \in \Theta$  is defined as

$$\boldsymbol{\theta}_d^\dagger := (\tau_d, \dots, \tau_d, 0, \dots, 0) \quad \text{for } \tau_d = \frac{\sqrt{2 \log(d)}}{\log \log(d)}, \quad (25)$$

and where the number of non-zero entries of  $\boldsymbol{\theta}_d$  is  $\lceil \sqrt{d} / \log(d) \rceil$  (at least for  $d$  large enough). The statement is as follows.

**Theorem 4.1.** *For every  $p \in (2, \infty)$  and any sequences of critical values such*

that  $\{2, \kappa_{d,2}\} \in \mathbb{T}_{\alpha_2}$ ,  $\{\infty, \kappa_{d,\infty}\} \in \mathbb{T}_{\alpha_\infty}$ , and  $\{p, \kappa_{d,p}\} \in \mathbb{T}_{\alpha_p}$ , the asymptotic sizes all being in  $(0, 1)$ , it holds that

$$\boldsymbol{\vartheta}^\dagger \in \mathcal{C}(\{p, \kappa_{d,p}\}) \quad \text{but} \quad \boldsymbol{\vartheta}^\dagger \notin \mathcal{C}(\{2, \kappa_{d,2}\}) \cup \mathcal{C}(\{\infty, \kappa_{d,\infty}\}). \quad (26)$$

Furthermore, the following convergences, stronger than the statement on the right in (26), hold

$$\mathbb{P} \left( \|\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d\|_2 \geq \kappa_{d,2} \right) \rightarrow \alpha_2 \quad \text{and} \quad \mathbb{P} \left( \|\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d\|_\infty \geq \kappa_{d,\infty} \right) \rightarrow \alpha_\infty. \quad (27)$$

Theorem 4.1 shows that there exist arrays of alternatives against which (i) all  $p$ -norm based tests are consistent for  $p \in (2, \infty)$ , and (ii) the 2- and supremum-norm based tests have asymptotic power equaling their asymptotic size.

As a corollary to this observation, we can now show that (under reasonable assumptions concerning the asymptotic sizes of the tests involved) it is impossible to obtain a test with the desired properties in Question 1 by “combining” the likelihood-ratio test and the supremum-norm based test into a test  $\varphi_d$  such that  $\mathcal{C}(\varphi_d)$  contains the union in (23). Here we say that a test  $\varphi_d$  is a combination of the likelihood-ratio test and the supremum-norm based test if there exist sequences of critical values  $\kappa_{d,2}$  and  $\kappa_{d,\infty}$  such that

$$\varphi_d \leq \max \left( \{2, \kappa_{d,2}\}_d, \{\infty, \kappa_{d,\infty}\}_d \right) \quad \text{for every } d \in \mathbb{N}. \quad (28)$$

A remarkable special case of such a combination procedure was investigated in Fan et al. (2015), where it was suggested to improve the likelihood-ratio test by the supremum-norm based test in the context of their power enhancement principle. Their construction satisfies Equation (28) for critical values  $\kappa_{d,2}$  such that the corresponding likelihood-ratio test has asymptotic size  $\alpha \in (0, 1)$ , and  $\kappa_{d,\infty}$  such that the corresponding supremum-norm based test has asymptotic size 0. Our result concerning such procedures is as follows.

**Corollary 4.2.** *If  $\{2, \kappa_{d,2}\} \in \mathbb{T}_{\alpha_2}$  and  $\{\infty, \kappa_{d,\infty}\} \in \mathbb{T}_{\alpha_\infty}$  with  $0 \leq \alpha_2 + \alpha_\infty < 1$ , then  $\boldsymbol{\vartheta}^\dagger \notin \mathcal{C}(\varphi_d)$  for every  $\varphi_d \in \mathbb{T}$  that satisfies Equation (28).*

Thus, any combination of the likelihood-ratio test and the supremum-norm based test is inconsistent against  $\vartheta^\dagger$ , and hence cannot be used to answer Question 1 in the affirmative.

## 5 Tests that dominate all $p$ -norm based tests

We here consider Question 1, i.e., whether a test exists that dominates all  $p$ -norm based tests ( $p \in (0, \infty)$ ). The answer turns out to be positive, which we show by constructing sequences of tests  $\psi_d$  that are consistent against any deviation from the null that some  $p$ -norm based test is consistent against. That Question 1 can be answered affirmatively also in the non-Gaussian case is discussed in Appendix B.3.

The idea underlying our construction is related to the power enhancement principle of Fan et al. (2015), in that we intend to “combine” the power of certain  $p$ -norm based tests. Importantly, however, we do not intend to enhance a *single* test by another one, but we intend to combine *all* of them to obtain a better test. To this end, we take the maximum of several test statistics, a way of combining tests that was also used by Spokoiny (1996) (see Equation 3.12 there) to get a minimax-adaptive test for a different testing problem. In principle, this would require us to combine an *uncountable* amount of tests indexed by  $p$  in the non-compact set  $(0, \infty)$ , which seems impossible at first sight. But the monotonicity result in Theorem 3.3 can be used as a “discretization device,” allowing us to get all consistency sets corresponding to powers  $p \in (0, \infty)$  by suitably combining a finite, but in  $d$  increasing, number  $m_d$ , say, of powers  $p_1, \dots, p_{m_d}$ , say, cf. Equation (31) below. Besides the monotonicity result, a crucial aspect exploited in the proof is the independence of the consistency set of  $p$ -norm based tests of their asymptotic size, cf. Theorems 3.1 and 3.5 and the ensuing discussions.

**Theorem 5.1.** *Let  $p_d$  be a strictly increasing and unbounded sequence in  $(0, \infty)$  and let  $m_d$  be a non-decreasing and unbounded sequence in  $\mathbb{N}$ . Choose  $\alpha \in (0, 1)$  and fix an array*

$$\mathcal{A} = \{\alpha_{j,d} \in (0, 1) : d \in \mathbb{N}, j = 1, \dots, m_d\}$$

such that, for  $\mathbb{M} \subseteq \mathbb{N}$  unbounded, it holds that

$$\sum_{j=1}^{m_d} \alpha_{j,d} = \alpha \text{ for every } d \in \mathbb{N}, \quad \lim_{d \rightarrow \infty} \alpha_{m_d,d} > 0, \quad \text{and} \quad \lim_{d \rightarrow \infty} \alpha_{j,d} > 0 \text{ for every } j \in \mathbb{M}, \quad (29)$$

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

$$\mathbb{P}(\|\boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_{j,d}) = \alpha_{j,d} \quad \text{and} \quad \mathbb{E}(\psi_d(\boldsymbol{\varepsilon}_d)) = \alpha, \quad (30)$$

where

$$\psi_d(\cdot) := \mathbb{1} \left\{ \max_{j=1, \dots, m_d} \kappa_{j,d}^{-1} \|\cdot\|_{p_j} \geq c_d \right\}. \quad (31)$$

Then, the following statements hold:

1. The sequence of tests  $\psi_d$  has the property requested in Question 1, that is, it has asymptotic size  $\alpha$  and satisfies

$$\mathcal{C}(\{p, \kappa_d\}) \subseteq \mathcal{C}(\psi_d), \text{ for every } p \in (0, \infty) \text{ and every } \{p, \kappa_d\} \in \mathbb{T}_\alpha. \quad (32)$$

2. Under the additional condition that

$$\liminf_{d \rightarrow \infty} \frac{p_{m_d}}{\log(d)} > 2, \quad (33)$$

it furthermore holds that

$$\mathcal{C}(\{\infty, \kappa_d\}) \subseteq \mathcal{C}(\psi_d), \text{ for every } \{\infty, \kappa_d\} \in \mathbb{T}_\alpha.$$

*Remark 5.1.* Let us note first that for every  $d \in \mathbb{N}$  the acceptance region of the test  $\psi_d$  in (31) is a symmetric convex set if  $p_1 \geq 1$ . It then follows that  $\psi_d$  is admissible and unbiased (cf. Birnbaum (1955) and Stein (1956), and Anderson (1955)).

Second, in case (33) is satisfied, the test  $\psi_d$  does not only have the property of being consistent against every array of alternatives that any  $p$ -norm based test



with  $p \in (0, \infty)$  and asymptotic size in  $(0, 1)$  is consistent against. It moreover also dominates any supremum-norm based test with asymptotic size  $\alpha$ . The proof of this property crucially relies on Theorem 3.7.

Third, by Theorems 3.1 and 3.5, the consistency sets of  $p$ -norm based tests with asymptotic size in  $(0, 1)$  do not depend on the actual value of the asymptotic size. It therefore follows that an apparently stronger (but equivalent) version of Theorem 5.1 is true, in which  $\mathbb{T}_\alpha$  in Parts 1 and 2 is replaced by  $\bigcup_{\tilde{\alpha} \in (0, 1)} \mathbb{T}_{\tilde{\alpha}}$ .

The sequence  $m_d$  regulates the number of norms the test in Equation (31) is based on, whereas the sequence  $p_d$  determines the concrete powers  $p \in (0, \infty)$  used in the construction. The condition in Equation (33) requires that the maximal power used in the test (31) grows sufficiently quickly in  $d$  to guarantee that the supremum-norm based test is dominated by making use of Theorem 3.7.

Intuitively, the role of the array  $\mathcal{A}$  is to regulate the sizes of the “individual” tests involved in the construction in Equation (31). Furthermore,  $c_d$  is a correction term guaranteeing that the test in that display has size exactly equal to  $\alpha$  for every sample size  $d$ . The choice of  $c_d = 1$  would in general only lead to a test of size not greater than  $\alpha$ . Thus, working with a smaller  $c_d$  leads to higher power compared to  $c_d = 1$ , which would correspond to an overly conservative test. The critical values  $\kappa_{j,d}$  and the multiplier  $c_d$  as in Theorem 5.1 can be found by a simple line search. The probabilities that need to be obtained in such computations can be approximated numerically. This is computationally relatively cheap, because the number  $m_d$  of tests involved can be chosen to grow slowly in  $d$ , cf. Example 5.1 below.

*Remark 5.2.* Given sequences  $p_d$  and  $m_d$  as in Theorem 5.1, concrete choices of arrays  $\mathcal{A}$  satisfying  $\sum_{j=1}^{m_d} \alpha_{j,d} = \alpha$  for every  $d \in \mathbb{N}$ , and Equation (29) in Theorem 5.1 with  $\mathbb{M} = \mathbb{N}$ , can easily be obtained from any probability mass function  $\delta_j > 0$ ,  $j \in \mathbb{N}$ , and  $\gamma \in (0, 1)$ , via the transformation (for  $d$  large enough so that  $m_d \geq 2$ )

$$\alpha_{j,d} = \begin{cases} \gamma \alpha \frac{\delta_j}{\sum_{i=1}^{m_d-1} \delta_i} & \text{for } j = 1, \dots, m_d - 1, \\ (1 - \gamma) \alpha & \text{for } j = m_d. \end{cases} \quad (34)$$

*Example 5.1.* A specific example of a test as in Theorem 5.1 is given next. The test we discuss commences with the likelihood-ratio test in the sense that  $p_1 = 2$  (this will be important in Example 5.2 below). Choose  $\mathcal{A}$  as in Remark 5.2 with the geometric probability mass function  $\delta_j = \delta_0(1 - \delta_0)^{j-1}$  for some  $\delta_0 \in (0, 1)$  and  $\gamma \in (0, 1)$ . Choose  $m_d = \lceil 3 \log(d) \rceil + 1$ , and let  $p_d = d + 1$ , so that (33) is satisfied. For critical values  $\kappa_{j,d}$  as defined in (30) (but based on the concrete array, and the concrete sequences  $p_d$  and  $m_d$  just defined) and the corresponding  $c_d \leq 1$ , Theorem 5.1 shows that the corresponding sequence of tests satisfies the property sought for in Question 1 (and also dominates any supremum-norm based test).

The array  $\mathcal{A}$  that the test  $\psi_d$  in (31) is based on also regulates the uniform “closeness” of the asymptotic power function of  $\psi_d$  to the power function of each  $p_j$ -norm based test involved in its construction. The following result quantifies this relation. How the result can be used to mimic the asymptotic power properties of a specific  $p$ -norm based test when working with a test  $\psi_d$  will be discussed subsequently.

**Theorem 5.2.** *In the context of Theorem 5.1, fix  $j \in \mathbb{M}$  and set*

$$\lim_{d \rightarrow \infty} \alpha_{j,d} =: \underline{\alpha}_j > 0. \quad (35)$$

*For any sequence of critical values  $\kappa_d$  such that  $\{p_j, \kappa_d\} \in \mathbb{T}_\alpha$ , the sequence of tests  $\psi_d$  as defined in (31) satisfies*

$$\limsup_{d \rightarrow \infty} \sup_{\boldsymbol{\theta}_d \in \mathbb{R}^d} \left[ \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d) - \mathbb{E}(\psi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d)) \right] \leq \frac{\Phi^{-1}(1 - \underline{\alpha}_j) - \Phi^{-1}(1 - \alpha)}{\sqrt{2\pi}}, \quad (36)$$

and

$$\lim_{d \rightarrow \infty} \sup_{\boldsymbol{\theta}_d \in \mathbb{R}^d} \left[ \mathbb{E}(\psi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d)) - \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d) \right] = 1 - \alpha. \quad (37)$$

Recall from Equation (32) that the asymptotic power of  $\psi_d$  is 1 whenever that of some  $p$ -norm based test is. The inequality in (36) sheds further light on the asymptotic power function of  $\psi_d$  by establishing that it is *nowhere* much below that of  $\{p_j, \kappa_d\}$  for an index  $j \in \mathbb{M}$  such that  $\underline{\alpha}_j \approx \alpha$ , as the upper bound in that inequality

is then approximately 0. Thus, there are no arrays of alternatives for which much is lost by using  $\psi_d$  instead of such a  $\{p_j, \kappa_d\}$ . What is more, Equation (37) (and its proof) shows that much can be gained by using  $\psi_d$ , as arrays of alternatives exist against which  $\psi_d$  is consistent, but against which  $\{p_j, \kappa_d\}$  has asymptotic power equal to its size.

In the following example we illustrate how this reasoning can be incorporated in the construction of  $\psi_d$ .

*Example 5.2.* In case one has reasons to favor a specific  $p$ -norm based test, e.g., the likelihood-ratio test corresponding to  $p = 2$ , but does not want to abandon the idea of using a test  $\psi_d$  that dominates all  $p$ -norm based tests, one can decide on the following compromise, which is possible as a consequence of Theorems 5.1 and 5.2: choose the components in the construction of  $\psi_d$  in such a way that  $\psi_d$  dominates all  $p$ -norm based tests *and* such that the power function of the obtained test  $\psi_d$  is *everywhere* at most slightly smaller than that of the preferred  $p$ -norm based test.

To see how this can be achieved, we focus on the case where  $p = 2$ . Let  $\{2, \kappa_d\}$  be a sequence of likelihood-ratio tests with asymptotic size  $\alpha \in (0, 1)$ . To obtain a test  $\psi_d$  as in Theorem 5.1 and whose power is (asymptotically) nowhere much less than that of  $\{2, \kappa_d\}$ , we can reconsider the test  $\psi_d$  constructed in Example 5.1. Note that  $p_1 = 2$ . Furthermore,  $\underline{\alpha}_1 = \alpha\gamma\delta_0$ , cf. Equation (35). It hence follows from Theorem 5.2 (with  $j = 1$ ) that

$$\limsup_{d \rightarrow \infty} \sup_{\boldsymbol{\theta}_d \in \mathbb{R}^d} \left[ \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_2 \geq \kappa_d) - \mathbb{E}(\psi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d)) \right] \leq \frac{\Phi^{-1}(1 - \alpha\gamma\delta_0) - \Phi^{-1}(1 - \alpha)}{\sqrt{2\pi}}.$$

The upper bound can be made arbitrarily close to 0 by choosing  $\delta_0$  and  $\gamma$  close to 1. At the same time, and in contrast to the likelihood-ratio test  $\{2, \kappa_d\}$ ,  $\psi_d$  has the favorable property of being consistent against every alternative that some  $p$ -norm based test is consistent against.

## 6 Minimax adaptive testing

The main focus of the present article is to study the consistency behavior of  $p$ -norm based tests for the *unrestricted* testing problem (2), and to answer questions like “how do these tests compare” and “can they be dominated” in terms of their consistency properties.

Classical results on  $p$ -norm based tests in the literature on minimax-optimal testing, on the other hand, start with a specific set of alternatives, i.e., complements of  $p$ -norm balls centered at the origin. They then characterize the separation from the null necessary so that uniform consistency is possible, and show that a  $p$ -norm based test can be constructed that is minimax rate optimal against such alternatives. Such results provide a justification for using a  $p$ -norm based test if one cares about power against alternatives in the complement of a  $p$ -norm ball as just described, but they do not answer the questions that we have focused on in previous sections.

In the present section, we want to illustrate that tools similar to the ones used in previous sections can be used to establish adaptivity results in the minimax framework. Essentially, we show that a  $p$ -norm based test is not only minimax optimal against complements of centered  $p$ -norm balls, but is adaptively minimax optimal over all complements of centered  $q$ -norm balls with  $q \leq p$  (assuming that the radii of the balls admit minimax consistent tests), thus extending a result in [Ingster and Suslina \(2003\)](#) for the case where  $p \leq 2$ . This corresponds to the monotonicity result in [Theorem 3.3](#). We also prove an adaptivity result over the whole range of all  $p \in (0, \infty)$  which parallels [Theorem 5.1](#).

Let us first revisit classical results and introduce some notation. Given  $p \in (0, \infty)$  and a radius  $r \in (0, \infty)$ , let  $\mathbb{V}_p^d(r) := \{\boldsymbol{\theta}_d \in \mathbb{R}^d : \|\boldsymbol{\theta}_d\|_p \geq r\}$ . For a sequence of such radii  $r_{p,d}$ ,  $d \in \mathbb{N}$ , we now consider (for every  $d \in \mathbb{N}$ ) the testing problem

$$H_{0,d} : \boldsymbol{\theta}_d = \mathbf{0}_d \quad \text{vs.} \quad H_{1,d} : \boldsymbol{\theta}_d \in \mathbb{V}_p^d(r_{p,d}), \quad (38)$$

and denote the minimal sum of Type 1 and Type 2 errors for this testing problem

by

$$\gamma(r_{p,d}, p) := \inf_{\varphi_d} \left[ \mathbb{E} \varphi_d(\boldsymbol{\varepsilon}_d) + 1 - \inf_{\boldsymbol{\theta}_d \in \mathbb{V}_p^d(r_{p,d})} \mathbb{E} \varphi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d) \right]; \quad (39)$$

here the outer infimum is taken over all Borel measurable functions from  $\mathbb{R}^d$  to  $[0, 1]$ .

Compared to the testing problem (2) studied in previous sections, the alternatives considered in (38) are now separated from the null, the type of separation depending on  $p$ . The main questions concerning the sequence of testing problems (38) (which have long been answered) are: (i) for which sequences of radii does  $\gamma(r_{p,d}, p) \rightarrow 0$ ; and (ii) for which tests this is achieved.

Following Ingster and Suslina (2003), a sequence of radii  $r_{p,d}^*$ ,  $d \in \mathbb{N}$ , is called a sequence of asymptotic *minimax rates/critical radii* for the sequence of testing problems (38) if the following holds for any sequence of radii  $r_{p,d}$  as above and as  $d \rightarrow \infty$ :

$$\begin{aligned} \gamma(r_{p,d}, p) \rightarrow 0 & \quad \text{if and only if} \quad r_{p,d}/r_{p,d}^* \rightarrow \infty \\ \gamma(r_{p,d}, p) \rightarrow 1 & \quad \text{if and only if} \quad r_{p,d}/r_{p,d}^* \rightarrow 0. \end{aligned}$$

Proposition 3.9 (and its proof) in Ingster and Suslina (2003) settles questions (i) and (ii) above as follows: (i) The following sequences constitute sequences of critical radii

$$r_{p,d}^* = \begin{cases} d^{\frac{4-p}{4p}} & \text{if } p \in (0, 2] \\ d^{\frac{1}{2p}} & \text{if } p \in (2, \infty). \end{cases} \quad (40)$$

(ii.a) For every  $p \in [2, \infty)$  and for every sequence of radii  $r_{p,d}$  such that  $r_{p,d}/r_{p,d}^* \rightarrow \infty$ , there exists a sequence of critical values  $\kappa_{p,r_{p,d},d}$ , such that

$$\mathbb{P} \left( \|\boldsymbol{\varepsilon}_d\|_p \geq \kappa_{p,r_{p,d},d} \right) + 1 - \inf_{\boldsymbol{\theta}_d \in \mathbb{V}_p^d(r_{p,d})} \mathbb{P} \left( \|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_p \geq \kappa_{p,r_{p,d},d} \right) \rightarrow 0.$$

That is, there exists a  $p$ -norm based test (the critical values depending on the se-

quence of radii  $r_{p,d}$ ) that is *minimax rate consistent* in the sequence of testing problems (38).<sup>5</sup> (ii.b) For every  $p \in (0, 2)$  and for every sequence of radii  $r_{p,d}$  such that  $r_{p,d}/r_{p,d}^* \rightarrow \infty$ , there exists a 2-norm based test that is minimax rate consistent in the sequence of testing problems (38).

Inspection of the proof of Proposition 3.9 and using Corollary 3.4 in Ingster and Suslina (2003) shows that actually more can be said in case (ii.b): Given a set of radial sequences

$$\{r_{q,d} \in (0, \infty) : q \in (0, 2], d \in \mathbb{N}\} \quad \text{such that} \quad \inf_{q \in (0, 2]} r_{q,d}/r_{q,d}^* \rightarrow \infty,$$

there exists a 2-norm based test that is minimax rate consistent for the sequence of testing problems

$$H_{0,d} : \boldsymbol{\theta}_d = \mathbf{0}_d \quad \text{vs.} \quad H_{1,d} : \boldsymbol{\theta}_d \in \bigcup_{q \in (0, 2]} \mathbb{V}_q^d(r_{q,d}); \quad (41)$$

that is, the 2-norm based test *adapts* to  $q \in (0, 2]$ . The following theorem now shows that such a result actually extends from the 2-norm based test to every  $p \in [2, \infty)$ , which parallels the monotonicity phenomenon in Theorem 3.3. Throughout the remainder of this section, we shall again use the notation  $\sigma_p^2 := \text{Var}(|\varepsilon_1|^p)$  and  $\mu_p := \mathbb{E}(|\varepsilon_1|^p)$ .

**Theorem 6.1.** *Fix  $p \in (2, \infty)$  and suppose a set of radial sequences*

$$\{r_{q,d} \in (0, \infty) : q \in (0, p], d \in \mathbb{N}\} \quad \text{satisfies} \quad r_d := \inf_{q \in (0, p]} r_{q,d}/r_{q,d}^* \rightarrow \infty.$$

*Then, the sequence of tests  $\{p, \kappa_d\}$  with  $\kappa_d := [r_d \sqrt{d\sigma_p^2} + d\mu_p]^{1/p}$  is minimax rate*

---

<sup>5</sup>In light of the minimax rate being  $r_{p,d}^* = d^{\frac{4-p}{4p}}$  for  $p \in (0, 2)$  it is tempting to conjecture that (12) could be replaced by  $\boldsymbol{\vartheta} \in \mathcal{C}(\{p, \kappa_d\})$  being equivalent to  $\frac{\|\boldsymbol{\theta}_d\|_2^2}{\sqrt{d}} \wedge \frac{\|\boldsymbol{\theta}_d\|_p^p}{d^{\frac{4-p}{4}}} \rightarrow \infty$ . This, however, is not the case as can be seen by considering  $\boldsymbol{\theta}_d = (d^{\frac{1}{2p}} \log(d), 0, \dots, 0)$ .

consistent in the sequence of testing problems

$$H_{0,d} : \boldsymbol{\theta}_d = \mathbf{0}_d \quad \text{vs.} \quad H_{1,d} : \boldsymbol{\theta}_d \in \bigcup_{q \in (0,p]} \mathbb{V}_q^d(r_{q,d}); \quad (42)$$

that is, for  $\mathbb{V}_{p,d} := \bigcup_{q \in (0,p]} \mathbb{V}_q^d(r_{q,d})$  it holds that

$$\mathbb{P}(\|\boldsymbol{\varepsilon}_d\|_p \geq \kappa_d) + 1 - \inf_{\boldsymbol{\theta}_d \in \mathbb{V}_{p,d}} \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_p \geq \kappa_d) \rightarrow 0. \quad (43)$$

Given the adaptivity result just obtained, one may ask whether one can construct a *single* test that is minimax rate consistent in any testing problem of the type (42). That is, does there exist a *single* test that is minimax rate consistent in the sequence of testing problems (42) for *every*  $p \in (0, \infty)$  simultaneously? As we shall establish next, this question can be answered affirmatively using a construction related to the one in Theorem 5.1.

**Theorem 6.2.** *Suppose a set of radial sequences*

$$\{r_{q,d} \in (0, \infty) : q \in (0, \infty), d \in \mathbb{N}\} \quad \text{satisfies} \quad r_d := \inf_{q \in (0, \infty)} r_{q,d}/r_{q,d}^* \rightarrow \infty.$$

Let  $p_d$  be a non-decreasing and diverging sequence of natural numbers satisfying

$$p_d \Phi(-r_d) \rightarrow 0 \quad \text{and} \quad \frac{p_d}{\sqrt{d}} (3/2)^{3p_d/2} \rightarrow 0. \quad (44)$$

Then, setting  $\kappa_{j,d} := [r_d \sqrt{d\sigma_j^2} + d\mu_j]^{1/j}$ , the sequence of tests

$$\psi_d^*(\cdot) := \mathbf{1} \left\{ \max_{j=1, \dots, p_d} \kappa_{j,d}^{-1} \|\cdot\|_j \geq 1 \right\} \quad (45)$$

is minimax rate consistent in the sequence of testing problems (42) for every  $p \in (0, \infty)$ ; that is, denoting  $\mathbb{V}_{p,d} := \bigcup_{q \in (0,p]} \mathbb{V}_q^d(r_{q,d})$ , it holds that

$$\mathbb{E}\psi_d^*(\boldsymbol{\varepsilon}_d) + 1 - \inf_{\boldsymbol{\theta}_d \in \mathbb{V}_{p,d}} \mathbb{E}\psi_d^*(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d) \rightarrow 0 \quad \text{for every } p \in (0, \infty). \quad (46)$$

For concreteness, we have decided to base the construction in (45) on all powers from 1 to  $p_d$ . Inspection of the proof shows that this is not crucial, and that one can also achieve the same minimax rate consistency property as in Theorem 6.2 by maximizing over an expanding subset of powers.

## 7 Conclusion

Combining two tests to enhance power is the underlying paradigm of the power enhancement principle recently put forward by Fan et al. (2015). There, it was suggested to combine tests based on the Euclidean and supremum norms. In the present article, we have characterized the consistency sets of  $p$ -norm based tests for all  $p \in (0, \infty]$ . Our characterizations allowed us to reveal an unexpected monotonicity relation of the consistency sets, and, somewhat surprisingly, to asymptotically dominate *all* these tests in terms of set inclusion of consistency sets, but also in a minimax sense. This was achieved by suitably combining a finite number of tests that grows with the dimension of the testing problem.

Even though the Gaussian sequence model is a prototypical framework for high-dimensional inference and results in that setup carry over to many other settings at least on a conceptual level, suitable generalizations of our results continue to hold also in non-Gaussian settings, as we show in detail in Appendix B.

## References

- Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society* 6(2), 170–176.
- Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science* 355(6324), 483–485.
- Athey, S., G. Imbens, and S. Wager (2018). Approximate residual balancing: de-



- biased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society Series B* 80(4), 597–623.
- Batir, N. (2008). Inequalities for the gamma function. *Archiv der Mathematik* 91, 554–563.
- 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(6), 2369–2429.
- Belloni, A. and V. Chernozhukov (2011).  $\ell_1$ -penalized quantile regression in high-dimensional sparse models. *Annals of Statistics* 39(1), 82–130.
- Belloni, A., V. Chernozhukov, I. Fernández-Val, and C. Hansen (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica* 85(1), 233–298.
- Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies* 81(2), 608–650.
- Bingham, N. H., C. M. Goldie, and J. L. Teugels (1987). *Regular Variation*. Cambridge: Cambridge University Press.
- Birnbaum, A. (1955, 03). Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests. *Annals of Mathematical Statistics* 26(1), 21–36.
- Bogachev, L. (2006). Limit laws for norms of iid samples with Weibull tails. *Journal of Theoretical Probability* 19(4), 849–873.
- Fan, J., Y. Liao, and J. Yao (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica* 83(4), 1497–1541.
- Farrell, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *Journal of Econometrics* 189(1), 1–23.

- Feller, W. (1968). *An Introduction to Probability Theory and its Applications. Vol. I*. Third edition. New York: John Wiley & Sons.
- Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica* 87(4), 1307–1340.
- Giessing, A. and J. Fan (2020). Bootstrapping  $\ell_p$ -statistics in high dimensions. *arXiv preprint arXiv:2006.13099*.
- Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, pp. 175–194.
- Ingster, Y. and I. A. Suslina (2003). *Nonparametric Goodness-of-fit Testing under Gaussian Models*. New York: Springer.
- Janßen, A. (2009). Limit laws for power sums and norms of i.i.d. samples. *Probability Theory and Related Fields* 146(3), 515–533.
- Kock, A. B. and L. Callot (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics* 186(2), 325–344.
- Kock, A. B. and D. Preinerstorfer (2019). Power in high-dimensional testing problems. *Econometrica* 87(3), 1055–1069.
- Le Cam, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.* 3, 37–98.
- Leadbetter, M. R., G. Lindgren, and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer: New York.
- Onatski, A. and C. Wang (2018). Alternative asymptotics for cointegration tests in large VARs. *Econometrica* 86(4), 1465–1478.

- Pinelis, I. (2010). Asymptotic efficiency of p-mean tests for means in high dimensions. *arXiv preprint arXiv:1006.0505*.
- Pinelis, I. (2014). Schur 2-concavity properties of Gaussian measures, with applications to hypotheses testing. *Journal of Multivariate Analysis* 124, 384 – 397.
- Roussas, G. G. (1972). *Contiguity of Probability Measures: Some Applications in Statistics*, Volume 63. Cambridge University Press.
- Schlather, M. (2001, 04). Limit distributions of norms of vectors of positive i.i.d. random variables. *Annals of Probability* 29(2), 862–881.
- Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *Annals of Statistics* 24(6), 2477 – 2498.
- Stein, C. (1956, 09). The admissibility of Hotelling’s  $t^2$ -test. *Annals of Mathematical Statistics* 27(3), 616–623.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58(1), 267–288.
- van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.

# Appendix

Appendix **A** contains the proofs of all results in the main body of the paper, whereas Appendix **B** contains results in a more general non-Gaussian framework.

## A Proofs of results in the main part of the paper

Throughout Appendix **A**, all assumptions imposed in the main part of the paper concerning the distribution of the random variables  $\varepsilon_i$  are maintained; that is, the  $\varepsilon_i$  are i.i.d. standard normal.

### A.1 Proof of Theorem 2.1

We establish a slightly stronger result than Theorem 2.1, showing that the sequence of tests  $\bar{\psi}_d$  in Theorem 2.1 can actually be constructed such that the power of the sequence of tests  $\psi_d$  is improved against a 1-sparse vector of alternatives. This is of additional interest, as it implies that the result pertains even in situations where the testing problem is reduced to 1-sparse parameter spaces, i.e., when instead of considering the parameter space  $\Theta = \times_{d=1}^{\infty} \mathbb{R}^d$  one only considers all arrays  $\boldsymbol{\vartheta}$  in

$$\Sigma_1 := \times_{d=1}^{\infty} \bigcup_{i=1}^d \{ \theta \mathbf{e}_i(d) : \theta \in \mathbb{R} \} \subseteq \Theta,$$

where  $\mathbf{e}_i(d)$  denotes the  $i$ -th element of the canonical basis of  $\mathbb{R}^d$ . Note that for every fixed  $d \in \mathbb{N}$ , every element of  $\bigcup_{i=1}^d \{ \theta \mathbf{e}_i(d) : \theta \in \mathbb{R} \}$  has at most one non-zero coordinate, i.e., is 1-sparse.

**Theorem A.1.** *Let  $\alpha \in (0, 1)$ . For every  $\psi_d \in \mathbb{T}_\alpha$  there exists a  $\bar{\psi}_d \in \mathbb{T}_\alpha$ , such that:*

1.  $\bar{\psi}_d \geq \psi_d$  holds for every  $d \in \mathbb{N}$ , guaranteeing that  $\bar{\psi}_d$  has uniformly non-inferior asymptotic power compared to  $\psi_d$ ;

2.  $\bar{\psi}_d$  is consistent against an array  $\boldsymbol{\vartheta} \in \Sigma_1$  against which  $\psi_d$  has asymptotic power at most  $\alpha$ .

In particular it holds that  $\mathcal{C}(\psi_d) \subsetneq \mathcal{C}(\bar{\psi}_d)$ .

The proof is a special case of the example discussed in Section 1.1 of [Kock and Preinerstorfer \(2019\)](#). The simplification is due to no sufficiency argument being necessary in the present framework. We provide a complete argument for the convenience of the reader.

*Proof.* The first part of the proof is based on an argument in Section 3.4.2 of [Ingster and Suslina \(2003\)](#): For every  $d \in \mathbb{N}$ , set  $\mathbb{Q}_{d,0} := \mathbb{N}(\mathbf{0}_d, \mathbf{I}_d)$  (the distribution of  $\mathbf{y}_d$  under the null) and define the mixture  $\mathbb{Q}_d = d^{-1} \sum_{i=1}^d \mathbb{Q}_{d,i}$ , where  $\mathbb{Q}_{d,i} := \mathbb{N}(a_d \mathbf{e}_i(d), \mathbf{I}_d)$  (the distribution of  $\mathbf{y}_d$  under the 1-sparse alternative  $\boldsymbol{\theta}_d = a_d \mathbf{e}_i(d)$ ) for  $a_d := \sqrt{\log(d)/2}$ . The likelihood-ratio statistic of  $\mathbb{Q}_d$  w.r.t.  $\mathbb{Q}_{d,0}$  is given by  $L_d(z_1, \dots, z_d) = d^{-1} \sum_{i=1}^d e^{a_d z_i - a_d^2/2}$ . Denote the expectation operators w.r.t.  $\mathbb{Q}_d$  and  $\mathbb{Q}_{d,i}$  by  $\mathbb{E}_d^{\mathbb{Q}}$  and  $\mathbb{E}_{d,i}^{\mathbb{Q}}$ , respectively. It holds that

$$\left| \mathbb{E}_{d,0}^{\mathbb{Q}}(\psi_d) - d^{-1} \sum_{i=1}^d \mathbb{E}_{d,i}^{\mathbb{Q}}(\psi_d) \right|^2 = \left| \mathbb{E}_{d,0}^{\mathbb{Q}}(\psi_d(1 - L_d)) \right|^2 \leq \mathbb{E}_{d,0}^{\mathbb{Q}}((1 - L_d)^2) = \mathbb{E}_{d,0}^{\mathbb{Q}}(L_d^2) - 1, \quad (47)$$

where we used  $d^{-1} \sum_{i=1}^d \mathbb{E}_{d,i}^{\mathbb{Q}}(\psi_d) = \mathbb{E}_d^{\mathbb{Q}}(\psi_d) = \mathbb{E}_{d,0}^{\mathbb{Q}}(\psi_d L_d)$ , Jensen's inequality and  $\mathbb{E}_{d,0}^{\mathbb{Q}}(L_d) = 1$ . From the moment-generating-function of a normal distribution we obtain

$$\mathbb{E}_{d,0}^{\mathbb{Q}}(L_d^2) - 1 = d^{-2} \sum_{i=1}^d \sum_{j=1}^d \mathbb{E}_{d,0}^{\mathbb{Q}}\left(e^{a_d(z_i + z_j) - a_d^2}\right) - 1 \leq d^{-1/2} \rightarrow 0. \quad (48)$$

Given this argument, we now see that the sequence to the left in (47) converges to 0, and thus the existence of an array  $\boldsymbol{\vartheta} = \{a_d \mathbf{e}_{i(d)}(d) : d \in \mathbb{N}\} \in \Sigma_1$  follows, against which  $\psi_d$  has asymptotic power at most  $\alpha$ . Next, define (for every  $d \in \mathbb{N}$ ) the test  $\nu_d = \mathbb{1}\{y \in \mathbb{R}^d : |e'_{i(d)} y| \geq \sqrt{a_d}\}$ . Obviously, its asymptotic size equals 0, and it is consistent against  $\boldsymbol{\vartheta}$ . Hence  $\bar{\psi}_d := \min(\psi_d + \nu_d, 1)$  has the properties required in the theorem.  $\square$

*Remark A.1.* Theorem [A.1](#) is an “existence result” in the sense that it shows that there exists a sequence of tests  $\bar{\psi}_d$  with certain properties. Inspection of the proof just given, however, shows how such a sequence of tests  $\bar{\psi}_d$  can actually be obtained by “enhancing” the test  $\psi_d$  with a test  $\nu_d$  as defined in course of the proof. Note that the sequence  $i(d)$  needed in such a construction can in principle be found numerically by choosing  $i(d)$  for every  $d \in \mathbb{N}$  as an index  $i \in \{1, \dots, d\}$  such that the test  $\psi_d$  has minimal power against  $\mathbb{N}(a_d e_i(d), \mathbf{I}_d)$  (note that the  $d$  values of the power function needed can be obtained by Monte-Carlo simulations). The same remark applies a fortiori to Theorem [2.1](#).

## A.2 $p$ -norm based tests with $p \in (0, \infty)$

### A.2.1 Proof of Theorem [3.1](#)

By Remark [B.1](#), Assumption [B.1](#) holds for  $F = \Phi$ , so that the theorem is an immediate consequence of Theorem [B.4](#) (since  $\mathbb{E}(|\varepsilon|^q) < \infty$  then holds for every  $q \in (0, \infty)$ ).

### A.2.2 Proof of Corollary [3.2](#)

To show [\(11\)](#), recalling the definition of  $g_p$  from [\(9\)](#), we just combine Theorem [3.1](#) with  $\frac{1}{2}(|\cdot|^p + |\cdot|^2) \leq g_p(\cdot) \leq |\cdot|^p + |\cdot|^2$  for  $p \in [2, \infty)$ . Similarly, to show [\(12\)](#), we use Theorem [3.1](#) together with  $g_p(\cdot) \leq |\cdot|^p$  and  $g_p(\cdot) \leq |\cdot|^2$  for  $p \in (0, 2)$ . To verify the remaining statement, let  $\boldsymbol{\vartheta} \in \Theta$  be defined via

$$\theta_{i,d} = \begin{cases} d^{-1/4} & \text{for } i = 1, \dots, d-1, \\ d^{1/(2p)} & \text{for } i = d. \end{cases}$$

Then  $\|\boldsymbol{\theta}_d\|_2^2/\sqrt{d} \geq d^{\frac{1}{p}-\frac{1}{2}} \rightarrow \infty$  and  $\|\boldsymbol{\theta}_d\|_p^p/\sqrt{d} \geq \frac{d-1}{\sqrt{d}} d^{-p/4} \rightarrow \infty$ . However,

$$d^{-1/2} \sum_{i=1}^d g_p(\theta_{i,d}) = \frac{d-1}{d} + 1 \leq 2,$$

and we conclude  $\boldsymbol{\vartheta} \notin \mathcal{C}(\{p, \kappa_d\})$  with Theorem [3.1](#).

### A.2.3 Proof of Theorem 3.3

Just observe that  $0 < p < q < \infty$  implies  $g_p \leq g_q$ , which by Theorem 3.1 delivers  $\mathcal{C}(\{p, \kappa_{d,p}\}) \subseteq \mathcal{C}(\{q, \kappa_{d,q}\})$ . That the inclusion is strict follows from

$$\boldsymbol{\vartheta} := \{(d^{1/(2p)}, 0, \dots, 0) : d \in \mathbb{N}\} \in \mathcal{C}(\{q, \kappa_{d,q}\}) \setminus \mathcal{C}(\{p, \kappa_{d,p}\}),$$

which is easily seen using the equivalence from Theorem 3.1.

### A.2.4 Proof of Theorem 3.4

Fix  $p < q$ ,  $\{p, \kappa_{d,p}\}$  and  $\{q, \kappa_{d,q}\}$  as in the statement of this theorem. Let  $\boldsymbol{\vartheta} \in \mathcal{C}(\{q, \kappa_{d,q}\}) \setminus \mathcal{C}(\{p, \kappa_{d,p}\})$ , which is possible by Theorem 3.3. From Theorem 3.1 and Remark 3.1 we conclude that for every  $\delta \in (0, \infty)$  it holds that

$$G_{d,p}(\delta) := \frac{\sum_{i=1}^d \theta_{i,d}^2 \mathbb{1}\{|\theta_{i,d}| \leq \delta\} + \sum_{i=1}^d |\theta_{i,d}|^p \mathbb{1}\{|\theta_{i,d}| > \delta\}}{\sqrt{d}} = \frac{\sum_{i=1}^d g_p^{(\delta)}(\theta_{i,d})}{\sqrt{d}} \not\rightarrow \infty. \quad (49)$$

By (49), for every  $\delta \in (0, \infty)$ , we can choose a subsequence  $d'_\delta$  of  $d$  (we highlight the dependence of the subsequence on  $\delta$  in the proof to avoid confusion), such that  $\sup_{d'_\delta} G_{d'_\delta,p}(\delta) \leq D(\delta) < \infty$  for some  $D(\delta)$ . Although  $d'_\delta$  and  $D(\delta)$  depend on  $\delta$ , in general, if  $p \geq 2$  these dependencies can be avoided, because  $\sup_{d'_1} G_{d'_1,p}(\delta) \leq \sup_{d'_1} G_{d'_1,p}(1) =: D(1) < \infty$  then holds for all  $\delta \in (0, \infty)$ . To see the latter, just note that in case  $p \geq 2$  the function  $\delta \mapsto G_{d,p}(\delta)$ , for  $\delta \in (0, \infty)$ , has a maximum at  $\delta = 1$  for every  $d \in \mathbb{N}$ . In what follows we set  $d'_\delta = d'_1$  and  $D(\delta) = D(1)$  in case  $p \geq 2$ .

Irrespective of the value of  $p \in (0, \infty)$ , we obtain for every  $\delta \in (0, \infty)$  that

$$\delta^p \times \sup_{d'_\delta} \frac{\sum_{i=1}^{d'_\delta} \mathbb{1}\{|\theta_{i,d'_\delta}| > \delta\}}{\sqrt{d'_\delta}} \leq \sup_{d'_\delta} \frac{\sum_{i=1}^{d'_\delta} |\theta_{i,d'_\delta}|^p \mathbb{1}\{|\theta_{i,d'_\delta}| > \delta\}}{\sqrt{d'_\delta}} \leq \sup_{d'_\delta} G_{d'_\delta,p}(\delta) \leq D(\delta). \quad (50)$$

In case  $p \geq 2$ , we can furthermore take the supremum over  $\delta \in (0, \infty)$  (since the

subsequence chosen does not depend on  $\delta$ ) to get

$$\sup_{\delta \in (0, \infty)} \delta^p \times \sup_{d'_1} \frac{\sum_{i=1}^{d'_1} \mathbf{1} \{|\theta_{i,d'_1}| > \delta\}}{\sqrt{d'_1}} \leq D(1). \quad (51)$$

Thus, the first statements in Equation (16) and (17), respectively, hold for the subsequences  $d'_\delta$  and  $d'_1$ . It remains to verify that for every  $\delta \in (0, \infty)$  we have that  $\max_{i=1, \dots, d'_\delta} |\theta_{i,d'_\delta}| \rightarrow \infty$  (note that  $d'_\delta = d'_1$  in case  $p \geq 2$  by construction). Fix  $\delta \in (0, \infty)$ . Theorem 3.1 and Remark 3.1 show that  $\boldsymbol{\vartheta} \in \mathcal{C}(\{q, \kappa_{d,q}\})$  is equivalent to

$$\frac{\sum_{i=1}^d \theta_{i,d}^2 \mathbf{1} \{|\theta_{i,d}| \leq \delta\} + \sum_{i=1}^d |\theta_{i,d}|^q \mathbf{1} \{|\theta_{i,d}| > \delta\}}{\sqrt{d}} = \frac{\sum_{i=1}^d g_q^{(\delta)}(\theta_{i,d})}{\sqrt{d}} \rightarrow \infty \quad (52)$$

The last inequality in (50) delivers  $\sum_{i=1}^{d'_\delta} \theta_{i,d'_\delta}^2 \mathbf{1} \{|\theta_{i,d'_\delta}| \leq \delta\} / \sqrt{d'_\delta} \leq D(\delta)$  for every  $d'_\delta$ . Thus, (52) implies

$$\frac{\sum_{i=1}^{d'_\delta} |\theta_{i,d'_\delta}|^q \mathbf{1} \{|\theta_{i,d'_\delta}| > \delta\}}{\sqrt{d'_\delta}} \rightarrow \infty.$$

But for every  $d'_\delta$

$$\frac{\sum_{i=1}^{d'_\delta} |\theta_{i,d'_\delta}|^q \mathbf{1} \{|\theta_{i,d'_\delta}| > \delta\}}{\sqrt{d'_\delta}} \leq \max_{i=1, \dots, d'_\delta} |\theta_{i,d'_\delta}|^q \frac{\sum_{i=1}^{d'_\delta} \mathbf{1} \{|\theta_{i,d'_\delta}| > \delta\}}{\sqrt{d'_\delta}},$$

which, by (50), is upper bounded by  $\max_{i=1, \dots, d'_\delta} |\theta_{i,d'_\delta}|^q D(\delta) \delta^{-p}$ , which hence diverges to  $\infty$ .

### A.3 Supremum-norm based tests

In some proofs in this section we will use the following classic inequality (e.g., Feller (1968), Section 7.1) for the standard normal cdf  $\Phi$ :

$$(x^{-1} - x^{-3}) \frac{e^{-x^2/2}}{\sqrt{2\pi}} \leq 1 - \Phi(x) \leq x^{-1} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad \text{for every } x > 0. \quad (53)$$



We also need the following observation concerning  $\kappa_d$  sequences leading to  $\{\infty, \kappa_d\} \in \mathbb{T}_\alpha$  for some  $\alpha \in (0, 1)$  (which provides a more common way of writing critical values for the supremum-based test in the normal case as compared to the general case treated in Lemma B.8 further below). To obtain the result, we recall the classical limit theorem (e.g., Theorem 1.5.3 in Leadbetter et al. (1983))

$$\mathbb{P}\left(a_d \left[\max_{1 \leq i \leq d} \varepsilon_i - b_d\right] \leq x\right) \rightarrow \Lambda(x) \quad \text{for every } x \in \mathbb{R}, \quad (54)$$

where  $\Lambda(x) = \exp(-\exp(-x))$ ,  $x \in \mathbb{R}$ , denotes Gumbel's double exponential cdf, and where  $a_d = \sqrt{2 \log(d)}$  and  $b_d = \sqrt{2 \log(d)} - \frac{\log \log(d) + \log(4\pi)}{2\sqrt{2 \log(d)}}$ . For  $x \in \mathbb{R}$  set  $u_d = u_d(x) := x/a_d + b_d$ , so that

$$\mathbb{P}\left(\max_{1 \leq i \leq d} |\varepsilon_i| \leq u_d\right) = \mathbb{P}\left(\max_{1 \leq i \leq d} \varepsilon_i \leq u_d, \min_{1 \leq i \leq d} \varepsilon_i \geq -u_d\right) \rightarrow \exp(-2 \exp(-x)),$$

where we used (54) and the asymptotic independence of the minimum and maximum (e.g., Theorem 1.8.2 in Leadbetter et al. (1983)). That is,

$$\mathbb{P}\left(a_d \left[\max_{1 \leq i \leq d} |\varepsilon_i| - b_d\right] \leq x\right) \rightarrow \exp(-2 \exp(-x)) \quad \text{for every } x \in \mathbb{R}, \quad (55)$$

The following now immediately follows from (55).

**Lemma A.2.** *It holds that  $\{\infty, \kappa_d\} \in \mathbb{T}_\alpha$ ,  $\alpha \in (0, 1)$ , if and only if*

$$\kappa_d = \sqrt{2 \log(d)} - \frac{\log \log(d) + \log(4\pi)}{2\sqrt{2 \log(d)}} - \frac{\log(-\log(1-\alpha)/2)}{\sqrt{2 \log(d)}} + o\left(\frac{1}{\sqrt{\log(d)}}\right). \quad (56)$$

We can now proceed to proving the characterization of the consistency set of supremum-norm based tests. Throughout the proof we interpret sums over empty index sets as 0.

### A.3.1 Proof of Theorem 3.5

Let  $\{\infty, \kappa_d\}$  be as in the theorem statement. Proposition B.6 delivers that  $\boldsymbol{\vartheta} \in \mathcal{C}(\{\infty, \kappa_d\})$  is equivalent to the condition that every subsequence  $d'$  of  $d$  has a subsequence  $d''$  along which

$$\bar{\Phi}(\kappa_d - \|\boldsymbol{\theta}_d\|_\infty) \rightarrow 1 \quad \text{or} \quad \sum_{i=1}^d \bar{\Phi}(\kappa_d - |\theta_{i,d}|) \rightarrow \infty.$$

As a first step, we now show that this condition, and hence also  $\boldsymbol{\vartheta} \in \mathcal{C}(\{\infty, \kappa_d\})$ , is equivalent to the same condition but with  $\mathbf{c}_d = \sqrt{2 \log(d)} - \frac{\log \log(d)}{2\sqrt{2 \log(d)}}$  replacing  $\kappa_d$ .

Let  $d'$  be a subsequence of  $d$ . Lemma A.2 shows that  $\kappa_d - \mathbf{c}_d = u_d / \sqrt{2 \log(d)}$  for a bounded sequence  $u_d$ . Therefore,

$$\limsup_{d' \rightarrow \infty} \bar{\Phi}(\kappa_{d'} - \|\boldsymbol{\theta}_{d'}\|_\infty) = 1 \quad \Leftrightarrow \quad \limsup_{d' \rightarrow \infty} \bar{\Phi}(\mathbf{c}_{d'} - \|\boldsymbol{\theta}_{d'}\|_\infty) = 1, \quad (57)$$

and setting  $\mathcal{I}_d := \{i \in \{1, \dots, d\} : \kappa_d - |\theta_{i,d}| \leq 3\}$ , it is obvious that

$$\limsup_{d' \rightarrow \infty} \sum_{i \in \mathcal{I}_{d'}} \bar{\Phi}(\kappa_{d'} - |\theta_{i,d'}|) = \infty \quad \Leftrightarrow \quad \limsup_{d' \rightarrow \infty} \sum_{i \in \mathcal{I}_{d'}} \bar{\Phi}(\mathbf{c}_{d'} - |\theta_{i,d'}|) = \infty, \quad (58)$$

(as both divergences are equivalent to  $|\mathcal{I}_d| \rightarrow \infty$  along a subsequence of  $d'$ ). For  $\mathcal{I}_d^c := \{1, \dots, d\} \setminus \mathcal{I}_d$  we now show that

$$\limsup_{d' \rightarrow \infty} \sum_{i \in \mathcal{I}_{d'}^c} \bar{\Phi}(\kappa_{d'} - |\theta_{i,d'}|) = \infty \quad \Leftrightarrow \quad \limsup_{d' \rightarrow \infty} \sum_{i \in \mathcal{I}_{d'}^c} \bar{\Phi}(\mathbf{c}_{d'} - |\theta_{i,d'}|) = \infty. \quad (59)$$

To this end, let  $d \geq 2$  be large enough so that

$$|u_d| \leq \sqrt{2 \log(d)}. \quad (60)$$

Furthermore, let  $x$  and  $y$  be in  $(2, \infty)$  and such that  $x - y = u_d / \sqrt{2 \log(d)}$ . We make the following two observations:

O1: If  $\max(x, y) \geq \sqrt{2 \log(d)}$ , then, denoting  $z := \sqrt{2 \log(d)} - |u_d| / \sqrt{2 \log(d)}$ , we

have

$$\overline{\Phi}(x) \vee \overline{\Phi}(y) \leq \overline{\Phi}(z) \leq e^{-z^2/2} \leq d^{-1}e^{|u_d|},$$

where we used that  $\overline{\Phi}(w) \leq e^{-w^2/2}$  for  $w \geq 0$ .

O2: If  $\max(x, y) \leq \sqrt{2 \log(d)}$ , then  $|x^2 - y^2| \leq 2|u_d|$ , and

$$\frac{x^{-1}}{y^{-1} - y^{-3}} = \left[ (y + u_d/\sqrt{2 \log(d)})(y^{-1} - y^{-3}) \right]^{-1} \leq [(1 - y^{-1})(1 - y^{-2})]^{-1} \leq \frac{8}{3},$$

where we used (60) in the first (displayed) inequality. Now, (53) and boundedness of  $u_d$  shows that

$$0 < \frac{3}{8}e^{-\sup_{d \in \mathbb{N}} |u_d|} \leq \frac{\overline{\Phi}(x)}{\overline{\Phi}(y)} \leq \frac{8}{3}e^{\sup_{d \in \mathbb{N}} |u_d|} < \infty;$$

(by symmetry one only needs to check the upper bound).

Let  $\mathcal{I}_{d,1}^c$  denote those indices  $i$  in  $\mathcal{I}_d^c$  for which  $\max(\kappa_d - |\theta_{i,d}|, \mathbf{c}_d - |\theta_{i,d}|) \geq \sqrt{2 \log(d)}$ , and denote the set of remaining indices in  $\mathcal{I}_d^c$  by  $\mathcal{I}_{d,2}^c$ . It follows from O1 together with  $|\mathcal{I}_d^c| \leq d$ , that (to show the equivalence in (59)) it suffices to verify the equivalence with  $\mathcal{I}_{d'}^c$  replaced by  $\mathcal{I}_{d',2}^c$ . But the latter equivalence follows from O2, which establishes (59). Combining (57), (58) and (59), we have shown that  $\vartheta \in \mathcal{C}(\{\infty, \kappa_d\})$  is equivalent to every subsequence  $d'$  of  $d$  having a subsequence  $d''$  along which

$$\overline{\Phi}(\mathbf{c}_d - \|\boldsymbol{\theta}_d\|_\infty) \rightarrow 1 \quad \text{or} \quad \sum_{i=1}^d \overline{\Phi}(\mathbf{c}_d - |\theta_{i,d}|) \rightarrow \infty. \quad (61)$$

By Lemma A.2, the sequence of critical values  $\mathbf{c}_d$  satisfies  $\{\infty, \mathbf{c}_d\} \in \mathbb{T}_{\alpha^*}$  for some  $\alpha^* \in (0, 1)$ . Hence, the just-derived equivalence (61) together with (the equivalence (3  $\Leftrightarrow$  4)) in Proposition B.6 applied to  $\{\infty, \mathbf{c}_d\}$  shows that

$$\vartheta \in \mathcal{C}(\{\infty, \kappa_d\}) \Leftrightarrow \sum_{i=1}^d \frac{\overline{\Phi}(\mathbf{c}_d - |\theta_{i,d}|)}{\Phi(\mathbf{c}_d - |\theta_{i,d}|)} \rightarrow \infty. \quad (62)$$

Finally, note that by construction  $\overline{\Phi}/g_\infty > 0$  is uniformly continuous on compact

subsets of  $\mathbb{R} \setminus \{1\}$  with positive left- and right-sided limits at 1, and  $\bar{\Phi}(x)/g_\infty(x) \rightarrow 1/\sqrt{2\pi}$  as  $x \rightarrow \infty$ . Therefore, for every  $z \in \mathbb{R}$  there exists a  $C(z) \in (0, \infty)$  such that

$$\frac{\bar{\Phi}(x)}{g_\infty(x)} \in [C(z)^{-1}, C(z)] \quad \text{for every } x \geq z. \quad (63)$$

Furthermore, a sequence  $x_d$  satisfies  $g_\infty(x_d) \rightarrow \infty$  if and only if  $x_d \rightarrow -\infty$ . It follows that  $\sum_{i=1}^d g_\infty(\mathbf{c}_d - |\theta_{i,d}|) \rightarrow \infty$  if and only if every subsequence  $d'$  of  $d$  permits a subsequence  $d''$  along which  $\mathbf{c}_d - \|\boldsymbol{\theta}\|_\infty \rightarrow -\infty$  (equivalently  $\bar{\Phi}(\mathbf{c}_d - \|\boldsymbol{\theta}_d\|_\infty) \rightarrow 1$ ) or  $\sum_{i=1}^d \bar{\Phi}(\mathbf{c}_d - |\theta_{i,d}|) \rightarrow \infty$ . In other words, we have shown that the condition in (61) is equivalent to  $\sum_{i=1}^d g_\infty(\mathbf{c}_d - |\theta_{i,d}|) \rightarrow \infty$ , which concludes the proof.

### A.3.2 Proof of Theorem 3.7

Proposition B.7 (applicable due to Remark B.3) and Remark B.4 deliver the result.

### A.3.3 Proof of Theorem 4.1

It suffices to verify  $\boldsymbol{\vartheta}^\dagger \in \mathcal{C}(\{p, \kappa_{d,p}\})$  and (27). Recall that  $\boldsymbol{\theta}_d^\dagger := (\tau_d, \dots, \tau_d, 0, \dots, 0)$ ,  $\tau_d = \sqrt{2 \log(d)}/\log \log(d)$ , and that the number of non-zero entries of  $\boldsymbol{\theta}_d$  is  $\lceil \sqrt{d}/\log(d) \rceil$  (at least for  $d$  large enough).

1.: That  $\boldsymbol{\vartheta}^\dagger \in \mathcal{C}(\{p, \kappa_{d,p}\})$  follows from Corollary 3.2, noting that eventually

$$\|\boldsymbol{\theta}_d\|_p^p/\sqrt{d} \geq 2^{p/2} \log^{p/2-1}(d)/(\log \log(d))^p \rightarrow \infty,$$

the divergence following from  $p > 2$ .

2.: That  $\mathbb{P}(\|\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d\|_2 \geq \kappa_{d,2}) \rightarrow \alpha_2$  follows from Theorem B.4 (cf. also Remark B.1) and

$$\sum_{i=1}^d g_2(\theta_{i,d})/\sqrt{d} = \|\boldsymbol{\theta}_d^\dagger\|_2^2/\sqrt{d} \rightarrow 0.$$

3.: We now show that  $\mathbb{P}(\|\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d\|_\infty \leq \kappa_{d,\infty}) \rightarrow 1 - \alpha_\infty$ . Setting  $k_d :=$

$\lceil \sqrt{d}/\log(d) \rceil$ , write  $\mathbb{P}(\|\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d\|_\infty \leq \kappa_{d,\infty})$  as

$$\mathbb{P}(|\varepsilon_1 + \tau_d| \leq \kappa_{d,\infty})^{k_d} [1 - 2\Phi(-\kappa_{d,\infty})]^{d-k_d}. \quad (64)$$

Concerning the second factor in (64),  $B_d$ , say, observe that

$$B_d = [1 - 2\Phi(-\kappa_{d,\infty})]^{d-k_d} = \frac{[1 - 2\Phi(-\kappa_{d,\infty})]^d}{[1 - 2\Phi(-\kappa_{d,\infty})]^{d \times \frac{k_d}{d}}} \rightarrow 1 - \alpha_\infty,$$

where we used that  $\{\infty, \kappa_{d,\infty}\} \in \mathbb{T}_{\alpha_\infty}$  is equivalent to  $[1 - 2\Phi(-\kappa_{d,\infty})]^d \rightarrow 1 - \alpha_\infty$ , and that, as  $k_d/d \rightarrow 0$ , one also has  $[1 - 2\Phi(-\kappa_{d,\infty})]^{d \times \frac{k_d}{d}} \rightarrow 1$ . It remains to show that the first factor in (64),  $A_d$ , say, converges to 1. Note that  $\mathbb{P}(|\varepsilon_1 + \tau_d| \leq \kappa_{d,\infty})$  can be written as

$$\Phi(\kappa_{d,\infty} - \tau_d) - \Phi(-\kappa_{d,\infty} - \tau_d) \geq 1 - 2\Phi(-\kappa_{d,\infty} + \tau_d) = 1 - 2\bar{\Phi}(\kappa_{d,\infty} - \tau_d). \quad (65)$$

By (56) (and since  $\{\infty, \kappa_{d,\infty}\} \in \mathbb{T}_{\alpha_\infty}$  with  $\alpha_\infty \in (0, 1)$ ) we eventually have  $\kappa_{d,\infty} - \tau_d \geq \sqrt{\log(d)}$ . Thus, we can eventually apply Bernoulli's inequality (that is,  $(1+x)^m \geq 1+mx$  for  $m \in \mathbb{N} \cup \{0\}$  and  $x \geq -1$ ) and use (53) to conclude that

$$A_d \geq 1 - 2k_d \bar{\Phi}(\sqrt{\log(d)}) \geq 1 - 2k_d/\sqrt{d} \rightarrow 1.$$

### A.3.4 Proof of Corollary 4.2

First consider the case where  $\alpha_2$  and  $\alpha_\infty$  are both greater than 0. Then, we just note that (28) implies

$$\mathbb{E}\varphi_d(\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d) \leq \mathbb{P}(\|\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d\|_2 \geq \kappa_{d,2}) + \mathbb{P}(\|\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d\|_\infty \geq \kappa_{d,\infty}) \rightarrow \alpha_2 + \alpha_\infty < 1,$$

the convergence following from the last statement in Theorem 4.1. If any of the asymptotic sizes is 0, just pass to smaller critical values so that the asymptotic sizes are in  $(0, 1)$  and sum up to a number smaller than 1. Then, use the monotonicity of the rejection probabilities in the critical values and apply the already established

statement.

### A.3.5 Proof of Theorem 5.1

The theorem follows immediately from Theorem B.9 (cf. Remark B.1) together with Remarks B.3, B.4 and Theorem 3.5.

### A.3.6 Proof of Theorem 5.2

This is an immediate consequence of Theorem B.10 (cf. Remark B.1), noting that by Anderson's theorem

$$\mathbb{E}(\psi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d)) - \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d) \leq 1 - \mathbb{P}(\|\boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d) \leq 1 - \alpha + o(1),$$

which establishes (37) from (125).

## A.4 Joint minimax rate consistency

The following notation will be employed freely throughout the remainder of this section. For every  $p \in (0, \infty)$  define  $\lambda_p : \mathbb{R} \rightarrow [0, \infty)$  via  $s \mapsto \mathbb{E}|\varepsilon_1 + s|^p$  and abbreviate  $\sigma_p^2 := \text{Var}|\varepsilon_1|^p$ . We denote the positive square root of  $\sigma_p^2$  by  $\sigma_p$ .

We also note the following bounds concerning absolute moments of a standard normally distributed random variable, which are due to the following version of Stirling's approximation (cf. Theorem 1.5 in Batir (2008), applied with “ $x = (r - 1)/2$ ”)

$$\sqrt{2e} (r/(2e))^{r/2} \leq \Gamma((r+1)/2) < \sqrt{2\pi} (r/(2e))^{r/2}, \quad \text{for every } r > 1, \quad (66)$$

and will be used in the proof of Theorem 6.2 below.

**Lemma A.3.** *Let  $Z$  be a standard normal random variable. Then, for all  $r > 1$ ,*

$$\sqrt{2e/\pi} r^{r/2} e^{-r/2} \leq \mathbb{E}|Z|^r < \sqrt{2} r^{r/2} e^{-r/2}.$$

*Proof.* Write  $\mathbb{E}|Z|^r = \frac{1}{\sqrt{\pi}}\Gamma((r+1)/2)2^{r/2}$  and apply (66).  $\square$

#### A.4.1 Proof of Theorem 6.1

Let  $p \in (2, \infty)$  and the set of radial sequences be as in the statement of the theorem. We start with some preliminary observations. Note that for all  $q \in (0, p]$  and  $d \in \mathbb{N}$  one has  $r_{q,d} \geq r_d r_{q,d}^*$ . Furthermore, for every  $\boldsymbol{\theta}_d \in \mathbb{R}^d$ , one has that  $\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_p \geq \kappa_d$  is equivalent to

$$t_{p,d} = \frac{\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_p^p - \sum_{i=1}^d \lambda_p(\theta_{i,d})}{\sqrt{d\sigma_p^2}} + \frac{\sum_{i=1}^d (\lambda_p(\theta_{i,d}) - \lambda_p(0))}{\sqrt{d\sigma_p^2}} \geq r_d.$$

Note that  $t_{p,d}$  depends on  $\boldsymbol{\theta}_d$ , but we don't highlight this in our notation. By Lemma B.1 (cf. Remark B.1) there exists a  $c_p > 0$  such that

$$\mathbb{E}(t_{p,d}) = \frac{\sum_{i=1}^d (\lambda_p(\theta_{i,d}) - \lambda_p(0))}{\sqrt{d\sigma_p^2}} \geq c_p^{-1} \frac{\sum_{i=1}^d g_p(\theta_{i,d})}{\sqrt{d\sigma_p^2}} = c_p^{-1} \frac{\sum_{i=1}^d (\theta_{i,d}^2 \vee |\theta_{i,d}|^p)}{\sqrt{d\sigma_p^2}}, \quad (67)$$

the last equality following from  $g_p(x) = x^2 \vee |x|^p$  for  $p \in [2, \infty)$ , cf. (9). We can also bound

$$\sum_{i=1}^d (\theta_{i,d}^2 \vee |\theta_{i,d}|^p) \geq \|\boldsymbol{\theta}_d\|_s^s, \text{ for every } s \in [2, p]. \quad (68)$$

We next claim that for  $d$  large enough

$$\inf_{\boldsymbol{\theta}_d \in \mathbb{V}_{p,d}} \mathbb{E}(t_{p,d}) \geq 2r_d, \quad (69)$$

and first establish two inequalities taking care of different subsets of  $\mathbb{V}_{p,d}$ :

1. Let  $\boldsymbol{\theta}_d \in \mathbb{V}_d^q(r_{q,d})$  for  $q \in [2, p]$ : Then,  $r_{q,d}^* = d^{1/2q}$  by (40) and hence  $r_{q,d} \geq r_d d^{1/2q}$ . By (67) and (68) (applied with  $s = q$ ) one has that

$$\mathbb{E}(t_{p,d}) \geq c_p^{-1} \frac{\sum_{i=1}^d g_p(\theta_{i,d})}{\sqrt{d\sigma_p^2}} \geq c_p^{-1} \frac{\|\boldsymbol{\theta}_d\|_q^q}{\sqrt{d\sigma_p^2}} \geq c_p^{-1} \frac{r_d^q}{\sigma_p}.$$

**2.** Let  $\boldsymbol{\theta}_d \in \mathbb{V}_d^q(r_{q,d})$  for  $q \in (0, 2)$ : Then,  $r_{q,d}^* = d^{\frac{4-q}{4q}}$  by (40) and hence  $r_{q,d} \geq r_d d^{\frac{4-q}{4q}}$ . By Jensen's inequality  $\|\boldsymbol{\theta}_d\|_2 \geq d^{1/2-1/q} \|\boldsymbol{\theta}_d\|_q$ . By (67) and (68) (applied with  $s = 2$ ),

$$\mathbb{E}(t_{p,d}) \geq c_p^{-1} \frac{\sum_{i=1}^d g_p(\theta_{i,d})}{\sqrt{d\sigma_p^2}} \geq c_p^{-1} \frac{\|\boldsymbol{\theta}_d\|_2^2}{\sqrt{d\sigma_p^2}} \geq c_p^{-1} \frac{d^{1-\frac{2}{q}} \|\boldsymbol{\theta}_d\|_q^2}{\sqrt{d\sigma_p^2}} \geq c_p^{-1} \frac{d^{1-\frac{2}{q}} r_d^2 d^{\frac{4-q}{2q}}}{\sqrt{d\sigma_p^2}} = c_p^{-1} \frac{r_d^2}{\sigma_p}.$$

Since the lower bounds in the two inequality chains obtained are both greater than  $2r_d$  for  $d$  large enough (uniformly over the respective set of  $\boldsymbol{\theta}_d$  considered in each case) the claim in (69) follows.

Next, by Part **2.** of Lemma B.2 (cf. (89) and note that  $p \leq 2p - 2$ ), there exists a  $C_p \in (0, \infty)$  such that for all  $\boldsymbol{\theta}_d \in \mathbb{R}^d$

$$\text{Var}(t_{p,d}) = \sum_{i=1}^d \text{Var} \left( |\theta_{i,d} + \varepsilon_i|^p / \sqrt{d\sigma_p^2} \right) \leq 1 + \frac{C_p \sum_{i=1}^d g_{2p-2}(\theta_{i,d})}{d\sigma_p^2}.$$

Using (90) in the proof of Lemma B.3 and  $\sum_{i=1}^d g_p(\theta_{i,d}) \geq 1$  for all  $\boldsymbol{\theta}_d \in \mathbb{V}_{p,d}$  for  $d$  sufficiently large (the latter following from the displayed inequalities in **1.** and **2.** above) implies that for  $d$  sufficiently large

$$\frac{\sum_{i=1}^d g_{2p-2}(\theta_{i,d})}{d\sigma_p^2} \leq \frac{2}{(d\sigma_p^2)^{1/p}} \left( \sum_{i=1}^d g_p(\theta_{i,d}) / \sqrt{d\sigma_p^2} \right)^{\frac{2p-2}{p}} \leq \frac{1}{(d\sigma_p^2)^{1/p}} \left( c_p \mathbb{E}(t_{p,d}) \right)^{\frac{2p-2}{p}},$$

the last inequality following from (67). Thus, for  $\tau := (2p - 2)/p \in (0, 2)$  and for  $d$  sufficiently large,

$$\text{Var}(t_{p,d}) \leq 1 + \left( \mathbb{E}(t_{p,d}) \right)^\tau \text{ for every } \boldsymbol{\theta}_d \in \mathbb{V}_{p,d}. \quad (70)$$

Corollary 3.1 in Ingster and Suslina (2003) together with (69),  $r_d \rightarrow \infty$ , and (70) now proves the result.



#### A.4.2 Proof of Theorem 6.2

For every  $d \in \mathbb{N}$  and  $j = 1, \dots, p_d$  we can write  $\|\boldsymbol{\varepsilon}_d\|_j \geq \kappa_{j,d}$  as

$$\|\boldsymbol{\varepsilon}_d\|_j^j - d\lambda_j(0) \geq r_d \sqrt{d\sigma_j^2}.$$

Combined with  $\psi_d^*(\cdot) \leq \sum_{j=1}^{p_d} \mathbf{1}\{\|\cdot\|_j \geq \kappa_{j,d}\}$ , which follows from the definition of  $\psi_d^*$ , we obtain

$$\mathbb{E}\psi_d^*(\boldsymbol{\varepsilon}_d) \leq \sum_{j=1}^{p_d} \left[ \Phi(-r_d) + \sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\|\boldsymbol{\varepsilon}_d\|_j^j - d\lambda_j(0) \leq x\sqrt{d\sigma_j^2}\right) - \Phi(x) \right| \right].$$

Using the Berry-Esseen theorem and Lemma A.3, together with a straightforward computation (note that  $\sigma_p^2 = \mathbb{E}(|\varepsilon_1|^{2p}) - (\mathbb{E}(|\varepsilon_1|^p))^2$ ), it now follows that for constants  $C, D \in (0, \infty)$  (both independent of  $d$ )

$$\mathbb{E}\psi_d^*(\boldsymbol{\varepsilon}_d) \leq p_d \Phi(-r_d) + C \frac{p_d}{\sqrt{d}} \sup_{p \in [1, p_d]} \frac{\mathbb{E}|\varepsilon_1|^{3p}}{\sigma_p^3} \leq p_d \Phi(-r_d) + CD \frac{p_d}{\sqrt{d}} (3/2)^{3p_d/2} \rightarrow 0,$$

the convergence following from (44). To establish (46), it hence remains to be shown that for any  $p \in (0, \infty)$ , one has that

$$\inf_{\boldsymbol{\theta}_d \in \mathbb{V}_{p,d}} \mathbb{E}\psi_d^*(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d) \rightarrow 1, \quad \text{as } d \rightarrow \infty.$$

To this end, fix  $p \in (0, \infty)$  and let  $d$  be sufficiently large to ensure that  $J := \max(3, \lceil p \rceil) \leq p_d$  (recall that  $p_d \uparrow \infty$  by assumption). Then, it follows from the definition of  $\psi_d^*$  that

$$\psi_d^*(\cdot) \geq \mathbf{1}\{\|\cdot\|_J \geq \kappa_{J,d}\},$$

from which we can conclude that

$$\inf_{\boldsymbol{\theta}_d \in \mathbb{V}_{p,d}} \mathbb{E}\psi_d^*(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d) \geq \inf_{\boldsymbol{\theta}_d \in \mathbb{V}_{d,p}} \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_J \geq \kappa_{J,d}) \geq \inf_{\boldsymbol{\theta}_d \in \mathbb{V}_{J,d}} \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_J \geq \kappa_{J,d}).$$

From  $\bar{r}_d := \inf_{q \in (0, J]} \frac{r_{q,d}}{r_{q,d}^*} \geq r_d$  we now obtain

$$\kappa_{J,d} = \left( r_d \sqrt{d\sigma_J^2} + d\lambda_J(0) \right)^{1/J} \leq \left( \bar{r}_d \sqrt{d\sigma_J^2} + d\lambda_J(0) \right)^{1/J} =: \kappa_d,$$

so that

$$\inf_{\boldsymbol{\theta}_d \in \mathbb{V}_{J,d}} \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_J \geq \kappa_{J,d}) \geq \inf_{\boldsymbol{\theta}_d \in \mathbb{V}_{J,d}} \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_J \geq \kappa_d) \rightarrow 1,$$

where the convergence follows from Theorem 6.1 (applied with “ $p = J$ ”).

## A.5 Verification of a claim in Section 3.2

Let  $\{\infty, \kappa_d\} \in \mathbb{T}_\alpha$ ,  $\alpha \in (0, 1)$ . In Section 3.2 it was claimed that for a real sequence  $\tau_d$  and  $\boldsymbol{\nu}_d$  the vector of ones of length  $d$  it holds that  $\boldsymbol{\vartheta} := \{\tau_d \boldsymbol{\nu}_d : d \in \mathbb{N}\} \in \mathcal{C}(\{\infty, \kappa_d\})$  if and only if  $\sqrt{\log(d)}|\tau_d| \rightarrow \infty$ . To show this via Theorem 3.5, it suffices to verify that for any sequence  $\tau_d \geq 0$  it holds that

$$dg_\infty(\boldsymbol{c}_d - \tau_d) \rightarrow \infty \quad \Leftrightarrow \quad \sqrt{\log(d)}\tau_d \rightarrow \infty. \quad (71)$$

We choose  $w$  such that  $g_\infty$  is continuous and strictly decreasing (e.g., as in the example given after Equation (18)). We first verify (71) for all bounded sequences  $\tau_d \geq 0$ . In this case, for  $d$  large enough,  $dg_\infty(\boldsymbol{c}_d - \tau_d) = de^{-(\boldsymbol{c}_d - \tau_d)^2/2}/(\boldsymbol{c}_d - \tau_d)$ , the logarithm of which writes, using  $\boldsymbol{c}_d := \sqrt{2\log(d)} - \log \log(d)/(2\sqrt{2\log(d)})$  and expanding the square,

$$\begin{aligned} \log(d) - (\boldsymbol{c}_d - \tau_d)^2/2 - \log(\boldsymbol{c}_d - \tau_d) &= \log\left(\sqrt{\log(d)}/(\boldsymbol{c}_d - \tau_d)\right) + \sqrt{2\log(d)}\tau_d + O(1) \\ &= \sqrt{2\log(d)}\tau_d + O(1). \end{aligned}$$

Finally, if there existed an unbounded sequence  $\tau_d \geq 0$  with the property that one of the sequences in (71) diverges and the other doesn't, there would exist a subsequence  $d'$  along which this property is preserved and along which  $\tau_d$  diverges to  $\infty$

(otherwise this would contradict what has already been shown). Then,  $\sqrt{\log(d')} \tau_{d'} \rightarrow \infty$  and thus  $\sqrt{\log(d')} \min(\tau_{d'}, 1)$  diverges, implying  $d' g_\infty(\mathbf{c}_{d'} - \min(\tau_{d'}, 1)) \rightarrow \infty$  (by what we have already established for bounded sequences, along subsequences). Since  $g_\infty$  is strictly decreasing, we obtain  $d' g_\infty(\mathbf{c}_{d'} - \tau_{d'}) \rightarrow \infty$ , contradicting the property upon  $\tau_d$  was chosen to satisfy in the first place, as we have just shown that both sequences in (71) diverge along  $d'$ .

## B Supplementary general results *not* imposing Gaussianity

Throughout Appendix B we do no longer assume that  $\varepsilon_i$  is standard normal (although we still assume that the  $\varepsilon_i$  are i.i.d.), but we shall work under weaker conditions, which are imposed whenever needed. In particular, different (but overlapping) sets of conditions will be used depending on whether  $p \in (0, \infty)$  or  $p = \infty$ . This allows us to establish more general results.

We shall throughout denote the cdf of  $\varepsilon_i$  by  $F$ , and we write  $\bar{F} = 1 - F$ . Since the  $\varepsilon_i$  are i.i.d. throughout, an assumption imposed on the distribution of  $\varepsilon_1$  carries over to the distribution of every  $\varepsilon_i$  for  $i \in \mathbb{N}$ . In formulating our assumptions we denote  $\varepsilon_1$  by  $\varepsilon$  for simplicity.

Note that in this more general framework, for a sequence of tests  $\varphi_d$  the statement whether or not  $\varphi_d \in \mathbb{T}_\alpha$  for a given  $\alpha \in (0, 1)$  can depend on  $F$ . That is,  $\mathbb{T}_\alpha$  depends on  $F$ , which we shall highlight by writing  $\mathbb{T}_{\alpha, F}$  in this section. Similarly, the consistency set  $\mathcal{C}(\varphi_d)$  of  $\varphi_d$  may depend on  $F$ , which highlight by writing  $\mathcal{C}_F(\varphi_d)$ .

### B.1 Results on the consistency of $p$ -norm based tests with finite $p > 0$

We start with some auxiliary results.

### B.1.1 Assumptions and auxiliary results

For every  $p \in (0, \infty)$  define the function

$$\lambda_p : \mathbb{R} \rightarrow [0, \infty] \quad \text{via} \quad s \mapsto \mathbb{E}|\varepsilon + s|^p, \quad (72)$$

(this notation was already used in Appendix A.4 in the Gaussian case). Note that  $\lambda_p$  also depends on  $F$ , but we do not highlight this notationally. Our main assumption on  $\varepsilon$  in Appendix B.1 is the following.

*Assumption B.1.* The following conditions hold:

1.  $F(x) = 1 - F(-x)$  for every  $x \in \mathbb{R}$ ;
2.  $F$  is twice continuously differentiable (with first and second derivative  $f$  and  $f'$ , respectively);
3.  $f'$  restricted to  $(0, \infty)$  is non-positive, and is non-decreasing on  $(M, \infty)$  for some  $M > 0$ .

*Remark B.1.* Assumption B.1 is obviously satisfied for normal distributions with mean zero (and positive variance). But it clearly holds much more generally.

*Remark B.2.* Note that under Assumption B.1 the derivative  $f' : \mathbb{R} \rightarrow \mathbb{R}$  is bounded.

The following auxiliary results suitably extend statements contained in Lemma 3.2 of Ingster and Suslina (2003), Lemma 4.2 of Pinelis (2010) and Sections 5.7 and 5.9 in the latter reference from the Gaussian case to distributions satisfying Assumption B.1. Recall the definition of  $g_p$  from Equation (9).

**Lemma B.1.** *Under Assumption B.1 and if  $q \in (0, \infty)$  is such that  $\mathbb{E}(|\varepsilon|^q) < \infty$ , then*

$$c_p^{-1} g_p(s) \leq \lambda_p(s) - \lambda_p(0) \leq c_p g_p(s) \quad \text{for every } s \in \mathbb{R}, \quad (73)$$

for every  $p \in (0, q]$  and a suitable constant  $c_p \in (0, \infty)$  (that may depend on  $F$ ). Furthermore, for every such  $p$  the function  $\lambda_p$  is real-valued, continuous and even.

*Proof.* Let  $p \in (0, q]$ . Note that  $\lambda_p$  is continuous and real valued because of  $\mathbb{E}(|\varepsilon|^p)^{1/p} \leq \mathbb{E}(|\varepsilon|^q)^{1/q} < \infty$  together with the Dominated Convergence Theorem (note that  $(a + b)^p \leq 2^p(a^p + b^p)$  for every  $a, b \geq 0$ ). Note furthermore that  $\lambda_p$  is even since  $f$  is even. This proves the statements in the last sentence of the lemma.

We now claim that (73) follows if we can show that

$$(a) \quad \lim_{s \rightarrow 0; s \neq 0} s^{-2} (\lambda_p(s) - \lambda_p(0)) \in (0, \infty), \quad \text{and} \quad (b) \quad \lambda_p(s) > \lambda_p(0) \text{ for every } s \neq 0. \quad (74)$$

To see that this claim is correct, first note that by the Dominated Convergence Theorem

$$\lim_{s \rightarrow \pm\infty} |s|^{-p} (\lambda_p(s) - \lambda_p(0)) = \lim_{s \rightarrow \pm\infty} \mathbb{E} |s|^{-1} \varepsilon + \text{sign}(s)|^p = 1. \quad (75)$$

Next, by (74), the function  $r(s) := (\lambda_p(s) - \lambda_p(0))/g_p(s)$ ,  $s \neq 0$ , extends to a positive and continuous function on  $\mathbb{R}$ , which we denote by  $r$  as well, and for which (75) delivers

$$0 < a_0 := \inf_{s \in \mathbb{R}} r(s) \leq \sup_{s \in \mathbb{R}} r(s) =: a_1 < \infty.$$

Hence (73) follows. To verify (74) set  $I(x, s) := F(-x^{1/p} + s) + F(-x^{1/p} - s)$  for  $x \geq 0$  and  $s \in \mathbb{R}$ , and use  $F(x) = 1 - F(-x)$  to write

$$\lambda_p(s) = \int_0^\infty \mathbb{P}(|\varepsilon + s|^p > x) dx = \int_0^\infty I(x, s) dx. \quad (76)$$

For every  $x \in \mathbb{R}$ , using that  $f$  is even and the mean-value theorem,

$$\begin{aligned} \partial_s I(x, s) &= f(x^{1/p} - s) - f(x^{1/p} + s) = -2s f'(x^{1/p} + \tilde{s}), \quad \tilde{s} \in [-s, s], \\ \partial_s^2 I(x, s) &= -f'(x^{1/p} - s) - f'(x^{1/p} + s). \end{aligned} \quad (77)$$

For every  $\epsilon > 0$ , it thus follows that

$$\sup_{|s| \leq \epsilon, i=1,2} |\partial_s^i I(x, s)| \leq 2(\epsilon + 1) \sup_{|s| \leq \epsilon} |f'(x^{1/p} + s)|. \quad (78)$$

Therefore, if, for every  $\epsilon > 0$ , the function  $x \mapsto \sup_{|s| \leq \epsilon} |f'(x^{1/p} + s)|$  has a majorant  $m_\epsilon : (0, \infty) \rightarrow [0, \infty)$ , say, that is Lebesgue integrable over  $(0, \infty)$ , then, by the Dominated Convergence Theorem, we can differentiate twice under the integral in (76) at every  $s_0 \in \mathbb{R}$ . By Remark B.2, we have  $|f'| \leq L$ , say, and for  $\epsilon > 0$  set

$$m_\epsilon(x) := \begin{cases} -f'(x^{1/p} - \epsilon) & \text{if } x > (M + \epsilon)^p, \\ L & \text{if } x \in (0, (M + \epsilon)^p]. \end{cases} \quad (79)$$

By assumption  $-f'$  is non-negative on  $(0, \infty)$  and non-increasing on  $(M, \infty)$ . Thus,

$$0 \leq -f'(x^{1/p} + s) \leq -f'(x^{1/p} - \epsilon) = m_\epsilon(x) \text{ for every } x > (M + \epsilon)^p \text{ and every } |s| \leq \epsilon.$$

Hence, it remains to show that  $m_\epsilon$  is Lebesgue integrable. Since  $\mathbb{E}(|\varepsilon|^p)$  is finite, there exists a sequence  $b_m \in (1, \infty)$  such that  $b_m \uparrow \infty$  and  $b_m^p f(b_m) \rightarrow 0$ . Because  $f'$  is bounded, and by monotone convergence, it suffices to show that the sequence  $\int_{(1+\epsilon)^p}^{(b_m+\epsilon)^p} -f'(x^{1/p} - \epsilon) dx = p \int_1^{b_m} -(x + \epsilon)^{p-1} f'(x) dx$  is bounded. But integration by parts shows that

$$\int_1^{b_m} -x^p f'(x) dx \leq f(1) - b_m^p f(b_m) + p \mathbb{E}(|\varepsilon|^p) \rightarrow f(1) + p \mathbb{E}(|\varepsilon|^p), \quad (80)$$

where we used that  $\int_1^{b_m} x^{p-1} f(x) dx \leq \mathbb{E}(|\varepsilon|^p)$ , and from which the boundedness condition follows. Differentiating twice under the integral in (76) now gives

$$\lambda_p'(s) = \int_0^\infty f(x^{1/p} - s) - f(x^{1/p} + s) dx; \quad \lambda_p''(s) = - \int_0^\infty f'(x^{1/p} - s) + f'(x^{1/p} + s) dx.$$

By the Dominated Convergence Theorem both derivatives are continuous in  $s$  as  $f$  and  $f'$  are continuous (recall the majorant established above). Thus (74)(a) follows by L'Hospital' rule from  $\lambda_p'(0) = 0$  and  $\lambda_p''(0)$  positive (and finite) because  $-f'$  cannot vanish identically and is nonnegative everywhere on  $(0, \infty)$ . From Parts 1 and 3 of Assumption B.1 together with the previous display it readily follows that  $\lambda_p'(s) \geq 0$  for every  $s \geq 0$ . But then (74)(b) follows as well, because (74)(a) already shows

that  $\lambda_p(s) > \lambda_p(0)$  for every  $s$  in an open neighborhood of 0, and  $\lambda_p$  is even.  $\square$

**Lemma B.2.** *Under Assumption B.1 and if  $q \in [2, \infty)$  is such that  $\mathbb{E}(|\varepsilon|^q) < \infty$ , then the following holds:*

1. *For every  $p \in (0, \infty)$  such that  $2\lceil p/2 \rceil \leq q$  the function  $\Delta_p : \mathbb{R} \setminus (-1, 1) \rightarrow \mathbb{R}$  defined via*

$$\Delta_p(s) := s^2 (|s|^{-p} \lambda_p(s) - 1) \quad (81)$$

*is bounded and continuous*

2. *For every  $p \in (0, \infty)$  such that  $2\lceil p \rceil \leq q$  there exists a constant  $C'_p \in (0, \infty)$ , such that for every  $s \in \mathbb{R}$  we have*

$$|\text{Var}|\varepsilon + s|^p - \text{Var}|\varepsilon|^p| \leq C'_p [s^2 \mathbf{1}\{|s| \leq 1\} + (1 + |s|^{2p-2}) \mathbf{1}\{|s| > 1\}]. \quad (82)$$

3. *For every  $p \in (0, \infty)$  such that  $2\lceil 2p \rceil \leq q$ , there exists a constant  $C''_p \in (0, \infty)$ , such that for every  $s \in \mathbb{R}$  we have*

$$\mathbb{E} [|\varepsilon + s|^p - \mathbb{E}|\varepsilon + s|^p]^4 \leq C''_p [1 + |s|^{4p-2} \mathbf{1}\{|s| > 1\}]. \quad (83)$$

*Proof. 1.:* Fix  $p \in (0, \infty)$  such that  $\tilde{p} := \lceil p/2 \rceil \leq q/2$ . By Lemma B.1 the function  $\lambda_p$  is real-valued, continuous and even. These properties carry over to  $\Delta_p$ . To show that  $\Delta_p$  is bounded, it remains to verify that  $\limsup_{s \rightarrow \infty} \Delta_p(s) < \infty$  and that  $\liminf_{s \rightarrow \infty} \Delta_p(s) > -\infty$ . Concerning the former, let  $s > 0$  be large enough such that  $|s^{-2\tilde{p}} \lambda_{2\tilde{p}}(s) - 1| < 1/2$  holds (recall that  $s^{-2\tilde{p}} \lambda_{2\tilde{p}}(s) \rightarrow 1$  as  $s \rightarrow \infty$  was shown in (75)). By Jensen's inequality and the mean-value theorem (both applied to  $x \mapsto x^{p/(2\tilde{p})}$ ) there exists a constant  $D > 0$  (independent of  $s$  in the range we consider) such that

$$s^{-p} \lambda_p(s) - 1 \leq \left( \mathbb{E} \left( (\varepsilon/s + 1)^{2\tilde{p}} \right) \right)^{p/(2\tilde{p})} - 1 \leq D \left| \mathbb{E} \left( (\varepsilon/s + 1)^{2\tilde{p}} \right) - 1 \right|.$$

Now, we write

$$\mathbb{E} \left( (\varepsilon/s + 1)^{2\tilde{p}} \right) - 1 = \sum_{i=1}^{2\tilde{p}} \binom{2\tilde{p}}{i} \frac{\mathbb{E}(\varepsilon^i)}{s^i} = \sum_{i=1}^{\tilde{p}} \binom{2\tilde{p}}{2i} \frac{\mathbb{E}(\varepsilon^{2i})}{s^{2i}},$$

the last equality following from symmetry of  $F$ . Multiplying by  $s^2$  and letting  $s \rightarrow \infty$  we obtain  $\binom{2\tilde{p}}{2} \mathbb{E}(\varepsilon^2) < \infty$ , which takes care of the limit superior. Next, we observe that

$$s^{-p} \lambda_p(s) - 1 \geq \mathbb{E} \left( (\varepsilon/s + 1)^p \mathbf{1}\{|\varepsilon|/s < 1/2\} \right) - 1.$$

Expanding to the second order shows that for every  $x \in (-1/2, 1/2)$  the difference  $(x+1)^p - (1+px)$  equals  $\frac{p(p-1)}{2}(1+\tilde{x})^{p-2}x^2$  for an  $\tilde{x} \in (-1/2, 1/2)$ . In particular the difference is not smaller than  $-D'x^2$  for a constant  $D' > 0$  (treating separately the cases  $p \in (0, 1)$  and  $p \in [1, \infty)$ ). We can thus lower-bound  $s^{-p} \lambda_p(s) - 1$  further by

$$\begin{aligned} & \mathbb{E} \left( (p\varepsilon/s + 1) \mathbf{1}\{|\varepsilon|/s < 1/2\} \right) - 1 - D' \mathbb{E}((\varepsilon/s)^2) \\ &= -\mathbb{P}(|\varepsilon|/s \geq 1/2) - D' \mathbb{E}((\varepsilon/s)^2) \geq -(4 + D') \mathbb{E}(\varepsilon^2) s^{-2}; \end{aligned}$$

where we used  $\mathbb{E}(\varepsilon \mathbf{1}\{|\varepsilon|/s < 1/2\}) = 0$  (by symmetry of  $F$ ) to obtain the equality, and the last inequality follows from Markov's inequality. This proves the statement concerning the limit inferior.

**2.:** Fix  $p \in (0, \infty)$  such that  $2\lceil p \rceil \leq q$ . By symmetry of  $F$ , it suffices to prove (82) for  $s \geq 0$ . We start with the case  $s \geq 1$ . Writing  $\lambda_{\tilde{p}}(s) = s^{\tilde{p}}(s^{-2}\Delta_{\tilde{p}}(s) + 1)$  for  $\tilde{p} \in \{p, 2p\}$  shows that

$$\begin{aligned} |\mathbb{V}ar|\varepsilon + s|^p - \mathbb{V}ar|\varepsilon|^p| &= |\lambda_{2p}(s) - \lambda_p^2(s) - \lambda_{2p}(0) + \lambda_p^2(0)| \\ &\leq s^{2p-2} |\Delta_{2p}(s) - 2\Delta_p(s) - s^{-2}\Delta_p^2(s)| + |\lambda_{2p}(0) - \lambda_p^2(0)|. \end{aligned}$$

By **1.**, the upper bound is dominated by  $1 + s^{2p-2}$  times a positive constant. Hence we are done in this case. In case  $s < 1$ , applying Lemma **B.1** to  $\tilde{p} \in \{p, 2p\}$ , we have  $\lambda_{\tilde{p}}(s) - \lambda_{\tilde{p}}(0) \leq c_{\tilde{p}}s^2$ ; furthermore, the monotonicity property discussed in the



last paragraph of the proof of Lemma B.1 (and the mean-value theorem) shows that

$$0 \leq \lambda_p^2(s) - \lambda_p^2(0) \leq 2 \max_{s \in [0,1]} \lambda_p(s) [\lambda_p(s) - \lambda_p(0)] \leq 2\lambda_p(1)c_p s^2,$$

Hence, for  $s < 1$  we obtain (cf. the penultimate display)

$$|\mathbb{V}ar|\varepsilon + s|^p - \mathbb{V}ar|\varepsilon|^p| \leq |\lambda_{2p}(s) - \lambda_{2p}(0)| + |\lambda_p^2(s) - \lambda_p^2(0)| \leq (c_{2p} + 2\lambda_p(1)c_p)s^2,$$

and the statement in (82) thus follows.

**3.:** Let  $p \in (0, \infty)$  be such that  $2\lceil 2p \rceil \leq q$ . The Dominated Convergence Theorem shows that the function  $s \mapsto \mathbb{E}||\varepsilon + s|^p - \mathbb{E}|\varepsilon + s|^p|^4$  is continuous; furthermore, by symmetry of  $F$ , this function is even. To show (83), it hence suffices to verify that the function just defined divided by  $s^{4p-2}$  remains bounded as  $s \rightarrow \infty$ . To this end, let  $s \geq 1$  and write

$$\mathbb{E} [|\varepsilon + s|^p - \mathbb{E}|\varepsilon + s|^p]^4 = \sum_{i=0}^4 \binom{4}{i} \lambda_{ip}(s) (-\lambda_p(s))^{4-i}, \quad (84)$$

where we set  $\lambda_0 \equiv 1$ . Writing  $\lambda_{\tilde{p}}(s) = s^{\tilde{p}}(s^{-2}\Delta_{\tilde{p}}(s) + 1)$  for  $\tilde{p} \in \{ip : i = 0, 1, \dots, 4\}$  and  $\Delta_0 \equiv 0$ , shows that upon dividing (84) by  $s^{4p-2}$  we obtain

$$\begin{aligned} & \sum_{i=0}^4 \binom{4}{i} (\Delta_{ip}(s) + s^2) (-s^{-2}\Delta_p(s) + 1)^{4-i} \\ &= \sum_{i=0}^4 \binom{4}{i} \Delta_{ip}(s) (-s^{-2}\Delta_p(s) + 1)^{4-i} + s^2[s^{-2}\Delta_p(s)]^4, \end{aligned}$$

where we used the Binomial formula in the last equality. We now conclude with **1.**

□

From now on convergence in distribution as  $d \rightarrow \infty$  will be denoted by “ $\rightsquigarrow$ .”

**Lemma B.3.** *Under Assumption B.1 and if  $q \in [2, \infty)$  is such that  $\mathbb{E}(|\varepsilon|^q) < \infty$ , then the following holds for every  $p \in (0, \infty)$  such that  $2\lceil 2p \rceil \leq q$ : for every  $\vartheta \in \Theta$*

and every subsequence  $d'$  of  $d$  along which the sequence  $d^{-1/2} \sum_{i=1}^d g_p(\theta_{i,d})$  is bounded, we have

$$\frac{\|\boldsymbol{\theta}_{d'} + \boldsymbol{\varepsilon}_{d'}\|_p^p - \sum_{i=1}^{d'} \lambda_p(\theta_{i,d'})}{\sqrt{d' \text{Var}|\varepsilon|^p}} \rightsquigarrow \mathbb{N}(0, 1); \quad (85)$$

in particular, it holds that

$$\frac{\|\boldsymbol{\varepsilon}_d\|_p^p - d\mathbb{E}|\varepsilon|^p}{\sqrt{d\text{Var}|\varepsilon|^p}} \rightsquigarrow \mathbb{N}(0, 1), \quad (86)$$

so that for  $\alpha \in (0, 1)$  a sequence of critical values  $\kappa_d$  satisfies  $\{p, \kappa_d\} \in \mathbb{T}_{\alpha, F}$  if and only if

$$\kappa_d = \left[ \left( \Phi^{-1}(1 - \alpha) + o(1) \right) \sqrt{d\text{Var}|\varepsilon|^p} + d\mathbb{E}|\varepsilon|^p \right]^{1/p}. \quad (87)$$

*Proof.* Fix  $p \in (0, \infty)$  such that  $2[2p] \leq q$ . We give the proof under the assumption that  $d' \equiv d$ , which only simplifies the notation. To this end, we verify Lyapunov's condition (with fourth moments) for

$$\xi_{d,i} := |\theta_{i,d} + \varepsilon_i|^p - \lambda_p(\theta_{i,d}), \quad d \in \mathbb{N}, \quad i = 1, \dots, d.$$

From **3.** of Lemma **B.2** we obtain

$$\sum_{i=1}^d \mathbb{E} \left| |\theta_{i,d} + \varepsilon_i|^p - \mathbb{E}|\theta_{i,d} + \varepsilon_i|^p \right|^4 \leq C_p'' \left[ d + \sum_{i=1}^d |\theta_{i,d}|^{4p-2} \mathbf{1}\{|\theta_{i,d}| > 1\} \right].$$

The inequality  $\|\cdot\|_r^r \leq \|\cdot\|_1^r$  for  $r \geq 1$ , applied with  $r^* := 1 \vee ((4p - 2)/p)$  delivers

$$\sum_{i=1}^d |\theta_{i,d}|^{4p-2} \mathbf{1}\{|\theta_{i,d}| > 1\} \leq \sum_{i=1}^d |\theta_{i,d}|^{pr^*} \mathbf{1}\{|\theta_{i,d}| > 1\} \leq d^{\frac{r^*}{2}} \left( d^{-1/2} \sum_{i=1}^d g_p(\theta_{i,d}) \right)^{r^*}.$$

Since  $d^{-1/2} \sum_{i=1}^d g_p(\theta_{i,d})$  is bounded, we obtain

$$\sum_{i=1}^d \mathbb{E} \left| |\theta_{i,d} + \varepsilon_i|^p - \mathbb{E}|\theta_{i,d} + \varepsilon_i|^p \right|^4 \leq C_p'' \left[ d + O\left(d^{r^*/2}\right) \right] = o(d^2).$$

Hence, the Lyapunov condition and (85) follow, upon showing that

$$\frac{\sum_{i=1}^d \mathbb{V}ar|\theta_{i,d} + \varepsilon_i|^p}{d\mathbb{V}ar|\varepsilon|^p} = 1 + \frac{\sum_{i=1}^d \mathbb{V}ar|\theta_{i,d} + \varepsilon_i|^p - d\mathbb{V}ar|\varepsilon|^p}{d\mathbb{V}ar|\varepsilon|^p} \rightarrow 1. \quad (88)$$

From 2. of Lemma B.2 we obtain

$$\left| \sum_{i=1}^d \mathbb{V}ar|\theta_{i,d} + \varepsilon_i|^p - d\mathbb{V}ar|\varepsilon|^p \right| \leq 2C'_p \sum_{i=1}^d g_{p \vee (2p-2)}(\theta_{i,d}). \quad (89)$$

On the one hand, if  $p \in (0, 2]$ , then  $2p - 2 \leq p$ , and since  $d^{-1/2} \sum_{i=1}^d g_p(\theta_{i,d})$  is bounded we obtain (88) from (89). On the other hand, if  $p \in (2, \infty)$ , then  $2p - 2 > p$  and the inequality  $\|\cdot\|_r^r \leq \|\cdot\|_1^r$  (for  $r \geq 1$ ) applied with  $r^{**} := (2p - 2)/p$  delivers

$$\begin{aligned} \sum_{i=1}^d g_{2p-2}(\theta_{i,d}) &= \sum_{i:|\theta_{i,d}| \leq 1} \theta_{i,d}^2 + \sum_{i:|\theta_{i,d}| > 1} |\theta_{i,d}|^{2p-2} \\ &\leq \sum_{i=1}^d g_p(\theta_{i,d}) + \left( \sum_{i=1}^d g_p(\theta_{i,d}) \right)^{(2p-2)/p}. \end{aligned} \quad (90)$$

Boundedness of  $d^{-1/2} \sum_{i=1}^d g_p(\theta_{i,d})$  again delivers (88). The statement in (86) follows from what has been established by setting  $\boldsymbol{\vartheta}$  the zero array. The final assertion in the lemma follows immediately from that statement.  $\square$

### B.1.2 Characterization of the consistency set of $p$ -norm based tests with finite $p > 0$

We are now ready to characterize the consistency set of a  $p$ -norm based test in the general (not necessarily normal) case. This is done in the following result. Note that the range of  $p > 0$  for which a statement is made in the following theorem depends on the “highest” moment that exists for  $\varepsilon$ . In particular, if  $\mathbb{E}|\varepsilon|^q < \infty$  for all  $q \in (0, \infty)$ , as in the normal case, then one obtains a statement for all  $p \in (0, \infty)$  which coincides with the statement made in Theorem 3.1. If, on the other hand, not

all moments exist (but at least the second moment exists), then one still obtains a corresponding statement, but for a limited range of  $p$ .

**Theorem B.4.** *Under Assumption B.1 and if  $q \in [2, \infty)$  is such that  $\mathbb{E}(|\varepsilon|^q) < \infty$ , then the following holds for every  $p \in (0, \infty)$  such that  $2\lceil 2p \rceil \leq q$ : for  $\kappa_d$  such that  $\{p, \kappa_d\} \in \mathbb{T}_{\alpha, F}$ ,  $\alpha \in (0, 1)$ ,*

$$\boldsymbol{\vartheta} \in \mathcal{C}_F(\{p, \kappa_d\}) \Leftrightarrow \frac{\sum_{i=1}^d g_p(\theta_{i,d})}{\sqrt{d}} \rightarrow \infty; \quad (91)$$

furthermore  $\{p, \kappa_d\}$  has asymptotic power  $\alpha$  against  $\boldsymbol{\vartheta} \in \Theta$  if and only if the sum on the right in (91) converges to 0.

*Proof.* Let  $\alpha \in (0, 1)$ ,  $p$  be such that  $2\lceil 2p \rceil \leq q$ , and  $\boldsymbol{\vartheta} \in \Theta$ . We start with two observations:

First, note that  $\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_p \geq \kappa_d$  is equivalent to

$$\frac{\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_p^p - \sum_{i=1}^d \lambda_p(\theta_{i,d})}{\sqrt{d\text{Var}|\varepsilon|^p}} + \frac{\sum_{i=1}^d (\lambda_p(\theta_{i,d}) - \lambda_p(0))}{\sqrt{d\text{Var}|\varepsilon|^p}} \geq \frac{\kappa_d^p - d\lambda_p(0)}{\sqrt{d\text{Var}|\varepsilon|^p}} =: \bar{\kappa}_d, \quad (92)$$

where  $\bar{\kappa}_d \rightarrow \Phi^{-1}(1 - \alpha) =: z_{1-\alpha}$  follows from  $\{p, \kappa_d\} \in \mathbb{T}_{\alpha, F}$  and (87).

Second, by Lemma B.1, there exists a positive real number  $c_p$  such that

$$0 \leq c_p^{-1} \frac{\sum_{i=1}^d g_p(\theta_{i,d})}{\sqrt{d\text{Var}|\varepsilon|^p}} \leq b_d := \frac{\sum_{i=1}^d (\lambda_p(\theta_{i,d}) - \lambda_p(0))}{\sqrt{d\text{Var}|\varepsilon|^p}} \leq c_p \frac{\sum_{i=1}^d g_p(\theta_{i,d})}{\sqrt{d\text{Var}|\varepsilon|^p}} < \infty. \quad (93)$$

We now prove the equivalence in (91):

Let  $\boldsymbol{\vartheta} \in \Theta$  be such that the sequence  $\sum_{i=1}^d g_p(\theta_{i,d})/\sqrt{d}$  does not diverge to  $\infty$ . We show that  $\boldsymbol{\vartheta} \notin \mathcal{C}_F(\{p, \kappa_d\})$ . From (93) it follows that there exists a subsequence  $d'$ , say, along which  $b_d \rightarrow b \in [0, \infty)$ , say. From (85) in Lemma B.3 (applied along the subsequence  $d'$ ) it follows that the sequence of random variables to the left in (92),  $X_d$ , say, converges in distribution to  $\mathbb{N}(b, 1)$  along  $d'$ . The Portmanteau Theorem hence

implies

$$\liminf_{d \rightarrow \infty} \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_p \geq \kappa_d) \leq \lim_{d' \rightarrow \infty} \mathbb{P}(X_{d'} - \bar{\kappa}_{d'} \geq 0) = \bar{\Phi}(z_{1-\alpha} - b) < 1,$$

that is  $\boldsymbol{\nu} \notin \mathcal{C}_F(\{p, \kappa_d\})$ .

Next, let  $\boldsymbol{\nu} \in \boldsymbol{\Theta}$  be such that the sequence  $\sum_{i=1}^d g_p(\theta_{i,d})/\sqrt{d} \rightarrow \infty$ . We show that  $\boldsymbol{\nu} \in \mathcal{C}_F(\{p, \kappa_d\})$ . We first claim that the sequence of random variables

$$\frac{\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_p^p - \sum_{i=1}^d \lambda_p(\theta_{i,d})}{\sum_{i=1}^d g_p(\theta_{i,d})} = \frac{\sum_{i=1}^d [|\theta_{i,d} + \varepsilon_i|^p - \mathbb{E}(|\theta_{i,d} + \varepsilon_i|^p)]}{\sum_{i=1}^d g_p(\theta_{i,d})} \quad (94)$$

converges to 0 in probability (the quotients are well defined for  $d$  large enough). To prove this claim, since the expectation of the random variables under consideration are all 0, it is enough to verify that the sequence of their variances converges to 0. By **2.** of Lemma **B.2**, we can bound these variances via

$$0 \leq \frac{\sum_{i=1}^d \text{Var}(|\theta_{i,d} + \varepsilon_i|^p)}{\left(\sum_{i=1}^d g_p(\theta_{i,d})\right)^2} \leq \frac{d \text{Var}|\varepsilon|^p}{\left(\sum_{i=1}^d g_p(\theta_{i,d})\right)^2} + \frac{2C'_p \sum_{i=1}^d g_{p \vee (2p-2)}(\theta_{i,d})}{\left(\sum_{i=1}^d g_p(\theta_{i,d})\right)^2}. \quad (95)$$

The first ratio on the far right-hand side in (95) converges to 0. To see that also the second ratio converges to 0 we argue as around (89): if  $p \in (0, 2]$  then  $2p - 2 \leq p$  and we conclude with  $\sum_{i=1}^d g_p(\theta_{i,d}) \rightarrow \infty$ . If  $p \in (2, \infty)$  we can use the bound in (90) and conclude in the same way.

With this in mind, we now show that  $\mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_p \geq \kappa_d) \rightarrow 1$  if  $\sum_{i=1}^d g_p(\theta_{i,d})/\sqrt{d} \rightarrow \infty$ . Let  $d'$  be a subsequence of  $d$ . By (93), there exists a subsequence  $d''$  of  $d'$  along which  $b_d^* := \sum_{i=1}^d (\lambda_p(\theta_{i,d}) - \lambda_p(0))/\sum_{i=1}^d g_p(\theta_{i,d})$  converges to  $b^* \in (0, \infty)$ , say. Now, we re-write the rejection event in (92) as the event that the random variable

$$\frac{\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_p^p - \sum_{i=1}^d \lambda_p(\theta_{i,d})}{\sum_{i=1}^d g_p(\theta_{i,d})} + b_d^* - \bar{\kappa}_d \frac{\sqrt{d \text{Var}|\varepsilon|^p}}{\sum_{i=1}^d g_p(\theta_{i,d})}$$

is non-negative. By the claim established above, and since  $\bar{\kappa}_d$  converges, this sequence of random variables converges along  $d''$  in probability to  $b^* > 0$ . Hence, the

consistency follows.

The second statement in the theorem follows by slightly modifying the argument after Equation (93).  $\square$

Given Theorem B.4 we could now obtain versions of Corollary 3.2 and Theorems 3.3 and 3.4 also under the more general Assumption B.1, but for a range of powers that depends on the “highest” moment of  $\varepsilon$ , which is required to be greater than 2. In particular, if all moments exist as in the discussion before Theorem B.4, then the statements carry over identically. Instead of spelling out all the details, we illustrate one such generalization for the important monotonicity statement in Theorem 3.3, but do not provide details for other results. The statement is as follows:

**Theorem B.5.** *Under Assumption B.1 the following holds: For  $0 < p < q < \infty$  such that  $\mathbb{E}|\varepsilon|^{2\lceil 2q \rceil} < \infty$ , and sequences of tests  $\{p, \kappa_{d,p}\}$  and  $\{q, \kappa_{d,q}\}$  with asymptotic sizes in  $(0, 1)$ , we have*

$$\mathcal{C}_F(\{p, \kappa_{d,p}\}) \subsetneq \mathcal{C}_F(\{q, \kappa_{d,q}\}).$$

*Proof.* It follows from Theorem B.4 that for  $\tilde{p} \in \{p, q\}$  we have

$$\mathcal{C}_F(\{\tilde{p}, \kappa_{d,\tilde{p}}\}) = \mathcal{C}_\Phi(\{\tilde{p}, \tilde{\kappa}_{d,\tilde{p}}\}) = \mathcal{C}(\{\tilde{p}, \tilde{\kappa}_{d,\tilde{p}}\}),$$

where  $\tilde{\kappa}_{d,\tilde{p}}$  is a sequence of critical values such that  $\{\tilde{p}, \tilde{\kappa}_{d,\tilde{p}}\}$  has asymptotic size in  $(0, 1)$  under standard normal errors. The statement hence follows from Theorem 3.3.  $\square$

## B.2 Consistency of supremum-norm based tests

We first summarize some assumptions and observations needed in Appendix B.2. The assumptions differ from Assumption B.1 upon which the results in Appendix B.1 are based, in that we do not need differentiability of  $F$ . However, in two of the results we need to further restrict its tail behavior. This is because the analysis

of the supremum-norm based test relies on results from extreme-value theory. The standard normal distribution satisfies all of the conditions imposed.

The symmetry condition on  $F$  in Assumption B.1 will also be used in the context of supremum-norm based tests.

*Assumption B.2.* We have  $F(x) = 1 - F(-x)$  for every  $x \in \mathbb{R}$ .

*Assumption B.3.* The cdf  $F$  is continuous.

Under Assumptions B.2 and B.3 we have

$$2\bar{F}(x) = \mathbb{P}(\varepsilon \geq x) + \mathbb{P}(\varepsilon \leq -x) = \mathbb{P}(|\varepsilon| \geq x) \quad \text{for all } x \geq 0. \quad (96)$$

The following condition will be used whenever we rely on results in Bogachev (2006), where a discussion of this assumption can be found in Section 5; cf. also Bingham et al. (1987) for a detailed account of regularly varying functions. Note that in the following assumption we implicitly impose the condition that the support of  $\varepsilon$  is unbounded.

*Assumption B.4.* The (log-tail distribution) function

$$h(x) := -\log(\mathbb{P}(|\varepsilon| \geq x)), \quad x \in \mathbb{R}, \quad (97)$$

is normalized regularly varying at infinity with index  $\rho \in (0, \infty)$ ; that is, for every  $\epsilon > 0$  the functions  $h(x)/x^{\rho-\epsilon}$  and  $h(x)/x^{\rho+\epsilon}$  are ultimately (i.e., for  $x$  large enough) increasing and decreasing, respectively.

*Remark B.3.* That Assumption B.4 holds with  $\rho = 2$  in case  $\varepsilon$  is standard normally distributed can easily be checked making use of (96),  $\Phi'(x)/(x\bar{\Phi}(x)) \rightarrow 1$  as  $x \rightarrow \infty$ , and the characterization for normalized regular variation given in Lemma 5.2 of Bogachev (2006) (cf. also Bingham et al. (1987, p. 15)).

A generalized inverse of the non-decreasing function  $h$  defined in (97) is defined via

$$h^{\leftarrow}(x) := \inf\{z \in \mathbb{R} : h(z) > x\}, \quad (98)$$

noting that the set over which the infimum is taken is non-empty for every  $x \in \mathbb{R}$  as  $h(x) \rightarrow \infty$  as  $x \rightarrow \infty$ .

As the first main result in this section, we now provide some statements equivalent to  $\boldsymbol{\vartheta} \in \mathcal{C}_F(\{\infty, \kappa_d\})$  under very weak assumptions on  $F$ . In the fourth statement in the following proposition we interpret  $1/0 = \infty$ ,  $a + \infty = \infty$  for every  $a \in (-\infty, \infty]$ .

**Proposition B.6.** *Suppose Assumptions B.2 and B.3 hold and let  $\kappa_d$  be a sequence of real numbers such that  $\{\infty, \kappa_d\} \in \mathbb{T}_{\alpha, F}$ , for an  $\alpha \in (0, 1)$ . Then, the following statements are equivalent:*

1.  $\boldsymbol{\vartheta} \in \mathcal{C}_F(\{\infty, \kappa_d\})$ .
2.  $\mathbb{P}(\max_{i=1, \dots, d} (\varepsilon_i + |\theta_{i,d}|) \leq \kappa_d) = \prod_{i=1}^d F(\kappa_d - |\theta_{i,d}|) \rightarrow 0$ .
3. Every subsequence  $d'$  of  $d$  has a subsequence  $d''$  along which  $\overline{F}(\kappa_d - \|\boldsymbol{\theta}_d\|_\infty) \rightarrow 1$  or  $\sum_{i=1}^d \overline{F}(\kappa_d - |\theta_{i,d}|) \rightarrow \infty$ .
4.  $\sum_{i=1}^d \overline{F}(\kappa_d - |\theta_{i,d}|) / F(\kappa_d - |\theta_{i,d}|) \rightarrow \infty$ .

*Proof.* Before we establish the equivalences in the lemma, we note some trivial equivalences to  $\boldsymbol{\vartheta} \in \mathcal{C}_F(\{\infty, \kappa_d\})$  under Assumptions B.2 and B.3: by definition  $\boldsymbol{\vartheta} \in \mathcal{C}_F(\{\infty, \kappa_d\})$  is equivalent to  $\mathbb{P}(\max_{i=1, \dots, d} |\varepsilon_i + \theta_{i,d}| \geq \kappa_d) \rightarrow 1$ , which is equivalent to  $\mathbb{P}(\max_{i=1, \dots, d} |\varepsilon_i + |\theta_{i,d}|| \geq \kappa_d) \rightarrow 1$ , because  $F$  is symmetric. Furthermore,  $\boldsymbol{\vartheta} \in \mathcal{C}_F(\{\infty, \kappa_d\})$  is equivalent to  $\mathbb{P}(\max_{i=1, \dots, d} |\varepsilon_i + \theta_{i,d}| \leq \kappa_d) \rightarrow 0$  because  $F$  is continuous, which is, by symmetry of  $F$ , equivalent to  $\mathbb{P}(\max_{i=1, \dots, d} |\varepsilon_i + |\theta_{i,d}|| \leq \kappa_d) \rightarrow 0$ .

(1  $\Rightarrow$  2): We need to show that

$$a_d := \mathbb{P}\left(\max_{1 \leq i \leq d} |\varepsilon_i + |\theta_{i,d}|| \leq \kappa_d\right) \rightarrow 0 \quad \text{implies} \quad b_d := \prod_{i=1}^d F(\kappa_d - |\theta_{i,d}|) \rightarrow 0. \quad (99)$$

To do so, we argue that if  $a_d \rightarrow 0$ , then any subsequence  $d'$  of  $d$  possesses a subsequence along which  $b_d \rightarrow 0$ . Fix a subsequence  $d'$  of  $d$ . The inequality  $b_d \leq F(\kappa_d - \|\boldsymbol{\theta}_d\|_\infty)$  shows that in case  $\liminf_{d' \rightarrow \infty} F(\kappa_{d'} - \|\boldsymbol{\theta}_{d'}\|_\infty) = 0$  we are done. Hence, we may assume without loss of generality (otherwise pass to a subsequence)



that for some  $\zeta \in (0, \infty)$  we have

$$F(\kappa_{d'} - |\theta_{i,d'}|) \geq F(\kappa_{d'} - \|\boldsymbol{\theta}_{d'}\|_\infty) \geq \zeta/2 > 0 \text{ for every } d', \text{ and every } i = 1, \dots, d'. \quad (100)$$

Since  $\{\infty, \kappa_d\} \in \mathbb{T}_{\alpha, F}$ , we have  $\alpha_d := \mathbb{P}(\|\boldsymbol{\varepsilon}_d\|_\infty \geq \kappa_d) \rightarrow \alpha \in (0, 1)$ , and we can hence assume without loss of generality that for some  $\epsilon > 0$  it holds that  $\alpha_{d'} \in (\epsilon, 1 - \epsilon)$  for every  $d'$  (otherwise pass to a subsequence). Then, for every  $d'$

$$F(-\kappa_{d'} - |\theta_{i,d'}|) \leq F(-\kappa_{d'}) = \bar{F}(\kappa_{d'}) = \frac{1 - (1 - \alpha_{d'})^{1/d'}}{2} \leq \frac{1 - \epsilon^{1/d'}}{2} \leq -\frac{\log(\epsilon)}{2d'}, \quad (101)$$

where we used symmetry of  $F$  to get the first equality, (96) and  $\mathbb{P}(|\varepsilon| \leq \kappa_d) = (1 - \alpha_d)^{1/d}$  to get the second equality, and  $1 - x^{-1} \leq \log(x)$  for every  $x > 0$  to get the last inequality. Combining (100) and (101), we can write

$$\mathbb{P}(|\varepsilon_i + |\theta_{i,d'}|| \leq \kappa_{d'}) = F(\kappa_{d'} - |\theta_{i,d'}|) - F(-\kappa_{d'} - |\theta_{i,d'}|)$$

as

$$F(\kappa_{d'} - |\theta_{i,d'}|) \left[ 1 - \frac{F(-\kappa_{d'} - |\theta_{i,d'}|)}{F(\kappa_{d'} - |\theta_{i,d'}|)} \right] \geq F(\kappa_{d'} - |\theta_{i,d'}|) \left[ 1 + \frac{\log(\epsilon)}{d'\zeta} \right].$$

By (100), there exists a  $\underline{d} \in \mathbb{N}$  such that the lower bound just derived is greater than 0 for every  $d' \geq \underline{d}$ . For all  $d' \geq \underline{d}$ , it so follows that

$$b_{d'} \left[ 1 + \frac{\log(\epsilon)}{d'\zeta} \right]^{d'} \leq \prod_{i=1}^{d'} \mathbb{P}(|\varepsilon_i + |\theta_{i,d'}|| \leq \kappa_{d'}) = a_{d'} \rightarrow 0,$$

and  $b_{d'} \rightarrow 0$  thus follows from  $[1 + \frac{\log(\epsilon)}{d'\zeta}]^{d'} \rightarrow \epsilon^{1/\zeta}$  as  $d' \rightarrow \infty$ .

**(2  $\Rightarrow$  3):** Upon taking the logarithm (interpreting  $-\log(0) = \infty$  and setting  $a + \infty = \infty$  for every  $a \in (-\infty, \infty]$ ) we obtain from  $\prod_{i=1}^d F(\kappa_d - |\theta_{i,d}|) \rightarrow 0$  that

$$\sum_{i=1}^d -\log \left( 1 - \bar{F}(\kappa_d - |\theta_{i,d}|) \right) \rightarrow \infty. \quad (102)$$

Let  $d'$  be a subsequence of  $d$ . In case  $\limsup_{d' \rightarrow \infty} \bar{F}(\kappa_{d'} - \|\boldsymbol{\theta}_{d'}\|_\infty) = 1$  we are obviously done. We may thus assume (otherwise pass to a subsequence) that for some  $\epsilon \in (0, 1)$  we have  $\bar{F}(\kappa_{d'} - \|\boldsymbol{\theta}_{d'}\|_\infty) \leq 1 - \epsilon$  for every  $d'$ . Then, for every  $d'$  and every  $i = 1, \dots, d'$ ,

$$\epsilon \leq F(\kappa_{d'} - |\theta_{i,d'}|).$$

Together with  $\frac{-x}{1+x} \geq -\log(1+x)$  for  $x > -1$  we thus obtain

$$\sum_{i=1}^{d'} -\log\left(1 - \bar{F}(\kappa_{d'} - |\theta_{i,d'}|)\right) \leq \frac{1}{\epsilon} \sum_{i=1}^{d'} \bar{F}(\kappa_{d'} - |\theta_{i,d'}|),$$

and (102) thus implies  $\sum_{i=1}^{d'} \bar{F}(\kappa_{d'} - |\theta_{i,d'}|) \rightarrow \infty$ .

**(3  $\Rightarrow$  1):** Let  $d'$  be a subsequence of  $d$ . We start with two preliminary observations: On the one hand, if there exists a subsequence  $d''$  of  $d'$  along which  $\bar{F}(\kappa_{d'} - \|\boldsymbol{\theta}_{d'}\|_\infty) \rightarrow 1$ , it follows that

$$\mathbb{P}\left(\max_{i=1, \dots, d''} (\varepsilon_i + |\theta_{i,d''}|) \leq \kappa_{d''}\right) = \prod_{i=1}^{d''} F(\kappa_{d''} - |\theta_{i,d''}|) \leq F(\kappa_{d''} - \|\boldsymbol{\theta}_{d''}\|_\infty) \rightarrow 0.$$

On the other hand, i.e., if  $\limsup_{d' \rightarrow \infty} \bar{F}(\kappa_{d'} - \|\boldsymbol{\theta}_{d'}\|_\infty) < 1$ , there exists a subsequence  $d''$  of  $d'$  along which  $\sum_{i=1}^{d''} \bar{F}(\kappa_{d''} - |\theta_{i,d''}|) \rightarrow \infty$ , and the inequality  $-\log(1+x) \geq -x$  for every  $x \geq -1$  thus implies

$$\sum_{i=1}^{d''} -\log\left(1 - \bar{F}(\kappa_{d''} - |\theta_{i,d''}|)\right) \geq \sum_{i=1}^{d''} \bar{F}(\kappa_{d''} - |\theta_{i,d''}|),$$

from which (102) and thus  $\prod_{i=1}^{d''} F(\kappa_{d''} - |\theta_{i,d''}|) \rightarrow 0$  follows, from which we again conclude that  $\mathbb{P}(\max_{i=1, \dots, d''} (\varepsilon_i + |\theta_{i,d''}|) \leq \kappa_{d''}) \rightarrow 0$ .

In any case, any subsequence  $d'$  of  $d$  admits a subsequence  $d''$  along which

$$0 = \lim_{d'' \rightarrow \infty} \mathbb{P}\left(\max_{i=1, \dots, d''} (\varepsilon_i + |\theta_{i,d''}|) \leq \kappa_{d''}\right) \geq \lim_{d'' \rightarrow \infty} \mathbb{P}\left(\max_{i=1, \dots, d''} |\varepsilon_i + |\theta_{i,d''}|| \leq \kappa_{d''}\right) \geq 0.$$

Thus,  $\mathbb{P}(\max_{i=1,\dots,d} |\varepsilon_i + |\theta_{i,d}|| \leq \kappa_d) \rightarrow 0$ , which (cf. the discussion in the first paragraph of this proof) is equivalent to  $\boldsymbol{\vartheta} \in \mathcal{C}_F(\{\infty, \kappa_d\})$ .

(3  $\Leftrightarrow$  4): We use the general fact (making use of the convention spelled out before the statement of the proposition) that a triangular array  $\{x_{i,d} \in [0, 1] : d \in \mathbb{N}, i = 1, \dots, d\}$  satisfies  $\sum_{i=1}^d x_{i,d}/(1 - x_{i,d}) \rightarrow \infty$  if and only if for every subsequence  $d'$  of  $d$  there exists a subsequence  $d''$  of  $d'$ , along which

$$\max_{i=1}^d x_{i,d} \rightarrow 1 \quad \text{or} \quad \sum_{i=1}^d x_{i,d} \rightarrow \infty.$$

□

The following result gives a sufficient condition for a  $p_d$ -norm based test with  $p_d$  diverging to  $\infty$  suitably quickly to be consistent against  $\boldsymbol{\vartheta}$ . The condition is formulated in terms of the consistency behavior of supremum-norm based tests. In this sense, the result links the consistency set of  $p_d$ -norm based tests with diverging  $p_d$  to that of the supremum-norm based test.

**Proposition B.7.** *Suppose Assumption B.4 holds, and let the sequence  $p_d \in (0, \infty)$  and the sequence of critical values  $\kappa_d$  be such that the sequence of tests  $\{p_d, \kappa_d\} := \mathbf{1}\{\|\cdot\|_{p_d} \geq \kappa_d\}$  is in  $\mathbb{T}_{\alpha, F}$ ,  $\alpha \in (0, 1)$ . Under the condition that*

$$\liminf_{d \rightarrow \infty} \frac{p_d}{\rho \log(d)} > 1, \tag{103}$$

*the sequence of tests  $\{p_d, \kappa_d\}$  is consistent against  $\boldsymbol{\vartheta} \in \Theta$  if every supremum-norm based test with asymptotic size in  $(0, \alpha]$  is consistent against  $\boldsymbol{\vartheta}$ . That is*

$$\bigcap \{\mathcal{C}_F(\{\infty, \kappa_{d,\infty}\}) : \{\infty, \kappa_{d,\infty}\} \in \mathbb{T}_{\tilde{\alpha}, F}, \tilde{\alpha} \in (0, \alpha]\} \subseteq \mathcal{C}_F(\{p_d, \kappa_d\}); \tag{104}$$

*in fact, there exists an  $\alpha^* \in (0, \alpha]$  such that*

$$\bigcap \{\mathcal{C}_F(\{\infty, \kappa_{d,\infty}\}) : \{\infty, \kappa_{d,\infty}\} \in \mathbb{T}_{\tilde{\alpha}, F}, \tilde{\alpha} \in [\alpha^*, \alpha]\} \subseteq \mathcal{C}_F(\{p_d, \kappa_d\}), \tag{105}$$

*which can be chosen as  $\alpha^* = \alpha$  in case the limit inferior in (103) is  $\infty$ .*

*Remark B.4.* If  $F$  is such that the consistency set of any supremum-norm based test with asymptotic size in  $(0, 1)$  neither depends on the sequence of critical values  $\kappa_{d,\infty}$  nor on the actual value of the asymptotic size (which turns out to be the case under normality of the errors, cf. Theorem 3.5), then the intersection to the left in (104) (and thus also in (105)) coincides with any member of this set. We then see that in case  $p_d$  satisfies (103), every  $p_d$ -norm based test with asymptotic size  $\alpha$  dominates (in terms of consistency) any supremum-norm based test with asymptotic size in  $(0, 1)$ .

*Remark B.5.* Note that the intersection in (105) actually equals

$$\bigcap \{ \mathcal{C}_F(\{\infty, \kappa_{d,\infty}\}) : \{\infty, \kappa_{d,\infty}\} \in \mathbb{T}_{\alpha^*, F} \} \subseteq \mathcal{C}_F(\{p_d, \kappa_d\}),$$

which follows since for  $\alpha^* \leq \tilde{\alpha}_1 < \tilde{\alpha}_2 \leq \alpha$  the sequence of critical values corresponding to  $\tilde{\alpha}_1$  are eventually larger than those of  $\tilde{\alpha}_2$ .

*Proof.* Note that for  $d$  large enough  $h^\leftarrow(\log(d)) > 0$ , recall (98) for a definition of the generalized inverse  $h^\leftarrow$ , which we can hence without loss of generality take for granted throughout. We start with two preliminary observations:

First, we note that Proposition 10.1 in Bogachev (2006) establishes that

$$\frac{\rho \log(d)}{h^\leftarrow(\log(d))} (\|\varepsilon_d\|_\infty - h^\leftarrow(\log(d))) \rightsquigarrow \Lambda, \quad (106)$$

where  $\Lambda(x) = \exp(-\exp(-x))$ ,  $x \in \mathbb{R}$ , denotes Gumbel's double exponential cdf.

Secondly, we fix an  $a \in (0, 1)$  such that

$$\liminf_{d \rightarrow \infty} p_d / (\rho \log(d)/a) > 1; \quad (107)$$

such an  $a$  exists due to (103). Abbreviating  $r_d := \rho \log(d)/a$ , Part (c) of Theorem 2.7 in Bogachev (2006) (applied with “ $t = r_d$ ”, “ $\alpha = a$ ”, and “ $N(t) = d$ ,” and noting that  $h$  is eventually continuous and strictly increasing, cf. the discussion after Condition 5.1 in Bogachev (2006) and also Theorem 1.5.5 in Bingham et al. (1987)) then establishes

$$\frac{\rho \log(d)}{h^\leftarrow(\log(d))} (\|\varepsilon_d\|_{r_d} - h^\leftarrow(\log(d))) \rightsquigarrow \Lambda_a, \quad (108)$$

where  $\Lambda_a$  is the distribution of  $a \log(Z_a)$ , for  $Z_a$  a random variable with a stable law with characteristic exponent  $a$  and skewness parameter 1 (the characteristic function of  $Z_a$  can be found in Equation 2.9 of [Bogachev \(2006\)](#)); in particular the cdf corresponding to  $\Lambda_a$  is continuous.

Given the two preliminary observations made, we now argue as follows: By [\(107\)](#), and since we are only concerned with asymptotic statements, we may for simplicity of notation and without loss of generality assume that  $p_d \geq r_d \geq 1$  holds for every  $d$ . With this in mind, we have, for every  $d \in \mathbb{N}$ ,

$$\|\cdot\|_\infty \leq \|\cdot\|_{p_d} \leq \|\cdot\|_{r_d}. \quad (109)$$

From  $\{p_d, \kappa_d\} \in \mathbb{T}_{\alpha, F}$  (cf. also [\(3\)](#)) and [\(109\)](#) we can hence conclude that

$$\limsup_{d \rightarrow \infty} \mathbb{P}(\|\boldsymbol{\varepsilon}_d\|_\infty \geq \kappa_d) \leq \alpha \leq \liminf_{d \rightarrow \infty} \mathbb{P}(\|\boldsymbol{\varepsilon}_d\|_{r_d} \geq \kappa_d). \quad (110)$$

Together with [\(106\)](#) and [\(108\)](#), continuity of  $\Lambda$  and  $\Lambda_a$ , Polya's theorem,  $\alpha \in (0, 1)$ , and denoting

$$u_d := \frac{\rho \log(d)}{h^{\leftarrow}(\log(d))} (\kappa_d - h^{\leftarrow}(\log(d))),$$

this implies

$$-\infty < \Lambda^{-1}(1 - \alpha) \leq \liminf_{d \rightarrow \infty} u_d \leq \limsup_{d \rightarrow \infty} u_d \leq \Lambda_a^{-1}(1 - \alpha) < \infty. \quad (111)$$

Set  $u'_d := \sup_{m \geq d} u_m \rightarrow \limsup_{d \rightarrow \infty} u_d =: \bar{u}$ . By [\(106\)](#), continuity of  $\Lambda$ , Polya's theorem and [\(111\)](#) the supremum-norm based test with sequence of critical values

$$\kappa'_d := h^{\leftarrow}(\log(d)) + u'_d \frac{h^{\leftarrow}(\log(d))}{\rho \log(d)} \geq h^{\leftarrow}(\log(d)) + u_d \frac{h^{\leftarrow}(\log(d))}{\rho \log(d)} = \kappa_d \quad (112)$$

has asymptotic size  $1 - \Lambda(\bar{u}) \in [\alpha^*, \alpha]$  where  $\alpha^* := 1 - \Lambda(\Lambda_a^{-1}(1 - \alpha)) \in (0, \alpha]$ , the inclusion following from [\(111\)](#). That is  $\{\infty, \kappa'_d\} \in \mathbb{T}_{\tilde{\alpha}, F}$  for some  $\tilde{\alpha} \in [\alpha^*, \alpha]$ . Now, let  $\boldsymbol{\vartheta}$  be an element of the intersection in [\(105\)](#). In particular, it follows that  $\boldsymbol{\vartheta} \in \mathcal{C}(\{\infty, \kappa'_d\})$ , i.e.  $\mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_\infty \geq \kappa'_d) \rightarrow 1$ . From [\(109\)](#) and [\(112\)](#) it follows

that  $\mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_{p_d} \geq \kappa_d) \geq \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_\infty \geq \kappa'_d)$ , which establishes the inclusion in (105) (and thus also the weaker statement in (104)). To prove the last statement, it suffices to note that  $\alpha^* \rightarrow \alpha$  for  $a \rightarrow 0$  because  $\Lambda_a \rightarrow \Lambda$  uniformly as  $a \rightarrow 0$ , as established in Theorem 10.2 of Bogachev (2006) (together with  $\Lambda$  being strictly increasing and continuous).  $\square$

The following result is an immediate consequence of Proposition 10.1 in Bogachev (2006); cf. Equation (106) in the proof of Proposition B.7.

**Lemma B.8.** *Suppose Assumption B.4 holds. Then, a sequence of critical values  $\kappa_d$  satisfies  $\{\infty, \kappa_d\} \in \mathbb{T}_{\alpha, F}$  if and only if*

$$\kappa_d = h^{\leftarrow}(\log(d)) \left( 1 + \frac{\Lambda^{-1}(1 - \alpha) + o(1)}{\rho \log(d)} \right), \quad (113)$$

where  $\Lambda(x) = \exp(-\exp(-x))$ ,  $x \in \mathbb{R}$ , denotes Gumbel's double exponential cdf.

### B.3 Question 1 and related results in the non-Gaussian case

We now show that Question 1 can be answered in the affirmative also in the non-Gaussian case under the condition that  $\varepsilon$  has moments of all orders (one can of course also formulate related results concerning tests that dominate all  $p$ -norm based tests up to a certain order, in case  $\varepsilon$  only permits absolute moments up to some order, but we do not spell out the details).

The following theorem contains a general version of the first part of Theorem 5.1. The theorem also contains a generalized version of the second part of Theorem 5.1, which is based on Proposition B.7, and which, by Remark B.4, simplifies to the second statement in Theorem 5.1 under additional conditions.

**Theorem B.9.** *Suppose Assumption B.1 holds and that  $\mathbb{E}(|\varepsilon|^p) < \infty$  for every  $p \in (0, \infty)$ . Let  $p_d, m_d, \alpha, \mathcal{A}$  and  $\mathbb{M}$  be as in the statement of Theorem 5.1. For every  $d \in \mathbb{N}$  and every  $j = 1, \dots, m_d$ , choose  $\kappa_{j,d} > 0$  and  $c_d \in (0, 1]$  such that*

$$\mathbb{P}(\|\boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_{j,d}) = \alpha_{j,d} \quad \text{and} \quad \mathbb{E}(\psi_d(\boldsymbol{\varepsilon}_d)) = \alpha, \quad (114)$$

where

$$\psi_d(\cdot) := \mathbb{1} \left\{ \max_{j=1, \dots, m_d} \kappa_{j,d}^{-1} \|\cdot\|_{p_j} \geq c_d \right\}. \quad (115)$$

Then, the following holds.

1. The sequence of tests  $\psi_d$  has the property

$$\mathcal{C}_F(\{p, \kappa_d\}) \subseteq \mathcal{C}_F(\psi_d), \text{ for every } p \in (0, \infty) \text{ and every } \{p, \kappa_d\} \in \mathbb{T}_{\alpha, F}. \quad (116)$$

2. If Assumption B.4 is satisfied and

$$\liminf_{d \rightarrow \infty} \frac{p_{m_d}}{\rho \log(d)} > 1, \quad (117)$$

then there exists a  $\delta^* \leq \delta := \lim_{d \rightarrow \infty} \alpha_{m_d, d}$  such that  $\psi_d$  is consistent against any  $\vartheta$  that every supremum-norm based test with asymptotic size in  $[\delta^*, \delta]$  is consistent against, i.e.,

$$\bigcap \left\{ \mathcal{C}_F(\{\infty, \kappa_{d, \infty}\}) : \{\infty, \kappa_{d, \infty}\} \in \mathbb{T}_{\tilde{\alpha}, F}, \tilde{\alpha} \in [\delta^*, \delta] \right\} \subseteq \mathcal{C}_F(\psi_d); \quad (118)$$

furthermore, in case the limit inferior in (117) is  $\infty$  one can choose  $\delta^* = \delta$ .

*Proof.* We prove the two statements in the theorem separately:

**Part 1:** That  $\kappa_{j,d} > 0$  as requested in (114) exists follows from the cdf of  $\|\boldsymbol{\varepsilon}_d\|_p$  being continuous, non-decreasing on  $(0, \infty)$  and zero everywhere else, together with  $\alpha_{j,d} \in (0, 1)$ . To show that  $c_d \in (0, 1]$  as in (114) exists, fix  $d \in \mathbb{N}$  and note that by (114) and a union bound

$$\mathbb{P} \left( \max_{j=1, \dots, m_d} \kappa_{j,d}^{-1} \|\boldsymbol{\varepsilon}_d\|_{p_j} \geq 1 \right) \leq \sum_{j=1}^{m_d} \mathbb{P} (\|\boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_{j,d}) = \sum_{j=1}^d \alpha_{j,d} = \alpha;$$

Furthermore, observe that the function

$$c \mapsto \mathbb{P} \left( \max_{j=1, \dots, m_d} \kappa_{j,d}^{-1} \|\boldsymbol{\varepsilon}_d\|_{p_j} \geq c \right) \quad (119)$$

is continuous and non-increasing on  $(0, \infty)$ . Since the limit of this function as  $c \rightarrow 0$  ( $c \rightarrow \infty$ ) is  $1 > \alpha$  (is  $0 < \alpha$ ), the existence of  $c_d \leq 1$  as required in (114) follows.

Next, let  $p \in (0, \infty)$  and  $\{p, \kappa_d\} \in \mathbb{T}_{\alpha, F}$ . Since  $\mathbb{M} \subseteq \mathbb{N}$  is unbounded, and because  $p_j$  is strictly increasing and unbounded, there exists a  $J \in \mathbb{M}$  (not depending on  $d$ ) such that  $p_J \geq p$ . By assumption,  $m_j \rightarrow \infty$  as  $j \rightarrow \infty$ , from which it follows that eventually  $m_d \geq J$ . For all such  $d$ ,  $c_d \leq 1$  shows that

$$\psi_d \geq \mathbf{1}\{\|\cdot\|_{p_J} \geq c_d \kappa_{J,d}\} \geq \mathbf{1}\{\|\cdot\|_{p_J} \geq \kappa_{J,d}\}. \quad (120)$$

The sequence  $\kappa_{J,d}$  was chosen such that the test at the far-right in (120) has null-rejection probability  $\alpha_{J,d}$ . Furthermore, we obtain  $\lim_{d \rightarrow \infty} \alpha_{J,d} =: \underline{\alpha}_J \in (0, 1)$  from  $J \in \mathbb{M}$ , (29), and  $\underline{\alpha}_J \leq \alpha < 1$ . Hence,  $\{p_J, \kappa_{J,d}\} \in \mathbb{T}_{\underline{\alpha}_J, F}$ . Theorem B.5 now shows that

$$\mathcal{C}_F(\{p, \kappa_d\}) \subseteq \mathcal{C}_F(\{p_J, \kappa_{J,d}\}) \subseteq \mathcal{C}_F(\psi_d), \quad (121)$$

the second inclusion following from (120).

**Part 2:** Note that

$$\psi_d \geq \mathbf{1}\{\|\cdot\|_{p_{m_d}} \geq c_d \kappa_{m_d,d}\} \geq \mathbf{1}\{\|\cdot\|_{p_{m_d}} \geq \kappa_{m_d,d}\}. \quad (122)$$

From (114) and (29) it follows that  $\{p_{m_d}, \kappa_{m_d,d}\}$ , say, the sequence of tests to the far-right in (122) has asymptotic size  $\delta \in (0, 1)$ . The condition in (117) together with Proposition B.7 allows us to conclude that for some  $\delta^* \in (0, \delta]$  we have

$$\bigcap \{\mathcal{C}_F(\{\infty, \kappa_{d,\infty}\}) : \{\infty, \kappa_{d,\infty}\} \in \mathbb{T}_{\tilde{\alpha}, F}, \tilde{\alpha} \in [\delta^*, \delta]\} \subseteq \mathcal{C}_F(\{p_{m_d}, \kappa_{m_d,d}\}) \subseteq \mathcal{C}_F(\psi_d),$$

where the last inclusion followed from (122). The final statement in Part 2 follows from the last statement in Proposition B.7.  $\square$

The following result quantifies the closeness of the power function of  $\psi_d$  to any  $p_j$ -norm based test used in its construction. It is a general version of Theorem 5.2 in the non-Gaussian case.



**Theorem B.10.** *In the context of Theorem B.9, fix  $j \in \mathbb{M}$  and set*

$$\lim_{d \rightarrow \infty} \alpha_{j,d} =: \underline{\alpha}_j > 0. \quad (123)$$

*For any sequence of critical values  $\kappa_d$  such that  $\{p_j, \kappa_d\} \in \mathbb{T}_{\alpha, F}$ , the sequence of tests  $\psi_d$  as defined in (115) satisfies*

$$\limsup_{d \rightarrow \infty} \sup_{\boldsymbol{\theta}_d \in \mathbb{R}^d} \left[ \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d) - \mathbb{E}(\psi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d)) \right] \leq \frac{\Phi^{-1}(1 - \underline{\alpha}_j) - \Phi^{-1}(1 - \alpha)}{\sqrt{2\pi}}, \quad (124)$$

*for  $\Phi$  the cdf of the standard normal distribution, and*

$$\liminf_{d \rightarrow \infty} \sup_{\boldsymbol{\theta}_d \in \mathbb{R}^d} \left[ \mathbb{E}(\psi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d)) - \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d) \right] \geq 1 - \alpha. \quad (125)$$

*Proof.* We make some initial observations. Unboundedness of  $\mathbb{M}$  together with (29) implies  $\underline{\alpha}_j \in (0, \alpha)$ , and from (114) it follows that  $\{p_j, \kappa_{j,d}\} \in \mathbb{T}_{\underline{\alpha}_j, F}$ . Furthermore, for every  $d \in \mathbb{N}$  and  $\boldsymbol{\theta}_d \in \mathbb{R}^d$ , the definition of  $\psi_d$  in Equation (115) and  $c_d \in (0, 1]$  readily shows that

$$\mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d) - \mathbb{E}(\psi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d)) \leq \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d) - \mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_{j,d}). \quad (126)$$

Next, let  $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$  and a subsequence  $d'$  of  $d$  be such that the limit superior to the left in (125) coincides with

$$\lim_{d' \rightarrow \infty} \left[ \mathbb{P}(\|\boldsymbol{\theta}_{d'} + \boldsymbol{\varepsilon}_{d'}\|_{p_j} \geq \kappa_d) - \mathbb{E}(\psi_{d'}(\boldsymbol{\theta}_{d'} + \boldsymbol{\varepsilon}_{d'})) \right] =: \mathfrak{s}. \quad (127)$$

It remains to verify that  $\mathfrak{s} \leq \phi(0)(\Phi^{-1}(1 - \underline{\alpha}_j) - \Phi^{-1}(1 - \alpha))$ . Without loss of generality (pass to a further subsequence if necessary) we can assume that  $\sum_{i=1}^{d'} g_{p_j}(\theta_{i,d'})/\sqrt{d'} \rightarrow a \in [0, \infty]$ . We consider two cases:

**Case 1:** If  $a = \infty$ , we may define  $\boldsymbol{\vartheta}^* \in \boldsymbol{\Theta}$  such that  $\boldsymbol{\theta}_{d'} = \boldsymbol{\theta}_{d'}^*$  holds along  $d'$  and such that  $\sum_{i=1}^{d'} g_{p_j}(\theta_{i,d'}^*)/\sqrt{d'} \rightarrow \infty$ . By Theorem B.4,  $\boldsymbol{\vartheta}^* \in \mathcal{C}_F(\{p_j, \kappa_{j,d}\})$  and  $\boldsymbol{\vartheta}^* \in$

$\mathcal{C}_F(\{p_j, \kappa_d\})$ , from which it follows that

$$\mathbb{P}(\|\boldsymbol{\theta}_d^* + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d) - \mathbb{P}(\|\boldsymbol{\theta}_d^* + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_{j,d}) \rightarrow 0. \quad (128)$$

In particular, it follows that (as  $d' \rightarrow \infty$ )

$$\mathbb{P}(\|\boldsymbol{\theta}_{d'} + \boldsymbol{\varepsilon}_{d'}\|_{p_j} \geq \kappa_{d'}) - \mathbb{P}(\|\boldsymbol{\theta}_{d'} + \boldsymbol{\varepsilon}_{d'}\|_{p_j} \geq \kappa_{j,d'}) \rightarrow 0. \quad (129)$$

Hence, in this case  $\mathfrak{s} \leq 0$  by (126), and we are done as  $\underline{\alpha}_j < \alpha$ .

**Case 2:** Suppose now that  $a \in [0, \infty)$ . Lemma B.3 shows that

$$\frac{\|\boldsymbol{\theta}_{d'} + \boldsymbol{\varepsilon}_{d'}\|_p^p - \sum_{i=1}^{d'} \lambda_p(\theta_{i,d'})}{\sqrt{d' \text{Var}(|\varepsilon|^p)}} \rightsquigarrow \mathbb{N}(0, 1) \quad (130)$$

and that

$$\frac{\kappa_{d'}^{p_j} - d' \lambda_{p_j}(0)}{\sqrt{d' \text{Var}(|\varepsilon|^{p_j})}} \rightarrow \Phi^{-1}(1 - \alpha) \quad \text{and} \quad \frac{\kappa_{j,d'}^{p_j} - d' \lambda_{p_j}(0)}{\sqrt{d' \text{Var}(|\varepsilon|^{p_j})}} \rightarrow \Phi^{-1}(1 - \underline{\alpha}_j). \quad (131)$$

Furthermore, since  $a$  is finite and by Lemma B.1, along a subsequence  $d''$  of  $d'$  we have

$$\frac{\sum_{i=1}^{d''} \lambda_{p_j}(\theta_{i,d''}) - \lambda_{p_j}(0)}{\sqrt{d'' \text{Var}(|\varepsilon|^{p_j})}} \rightarrow b \in [0, \infty). \quad (132)$$

Combining the observations in the previous three displays using a decomposition similar to (92) (and using Polya's theorem), we obtain

$$\begin{aligned} \mathbb{P}(\|\boldsymbol{\theta}_{d''} + \boldsymbol{\varepsilon}_{d''}\|_{p_j} \geq \kappa_{d''}) &\rightarrow \bar{\Phi}(\Phi^{-1}(1 - \alpha) - b) \\ \mathbb{P}(\|\boldsymbol{\theta}_{d''} + \boldsymbol{\varepsilon}_{d''}\|_{p_j} \geq \kappa_{j,d''}) &\rightarrow \bar{\Phi}(\Phi^{-1}(1 - \underline{\alpha}_j) - b), \end{aligned}$$

so that the upper bound in (126) converges along  $d''$  to

$$\Phi(\Phi^{-1}(1 - \underline{\alpha}_j) - b) - \Phi(\Phi^{-1}(1 - \alpha) - b) \leq \phi(0) \left( \Phi^{-1}(1 - \underline{\alpha}_j) - \Phi^{-1}(1 - \alpha) \right),$$

the inequality following from the mean-value theorem.

To prove the statement in (125), it suffices to exhibit an array  $\boldsymbol{\vartheta} \in \mathcal{C}_F(\psi_d)$  such that

$$\mathbb{P}(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d) \rightarrow \alpha. \quad (133)$$

To this end, define  $\boldsymbol{\vartheta}$  via  $\boldsymbol{\theta}_d := (d^{1/(4p_j)}, 0, \dots, 0)$ ,  $d \in \mathbb{N}$ . Theorem B.4 verifies (133), and shows that  $\boldsymbol{\vartheta} \in \mathcal{C}_F(\{4p_j, \kappa_d^*\})$  for  $\{4p_j, \kappa_d^*\} \in \mathbb{T}_{\alpha, F}$ ,  $\alpha \in (0, 1)$ , from which it follows from Theorem B.9 that  $\boldsymbol{\vartheta} \in \mathcal{C}(\psi_d)$ .  $\square$