

SUPERCONSISTENCY OF TESTS IN HIGH DIMENSIONS

ANDERS BREDAHL KOCK
University of Oxford

DAVID PREINERSTORFER
University of St. Gallen

To assess whether there is some signal in a big database, aggregate tests for the global null hypothesis of no effect are routinely applied in practice before more specialized analysis is carried out. Although a plethora of aggregate tests is available, each test has its strengths but also its blind spots. In a Gaussian sequence model, we study whether it is possible to obtain a test with substantially better consistency properties than the likelihood ratio (LR; i.e., Euclidean norm-based) test. We establish an impossibility result, showing that in the high-dimensional framework we consider, the set of alternatives for which a test may improve upon the LR test (i.e., its superconsistency points) is always asymptotically negligible in a relative volume sense.

1. INTRODUCTION

A major challenge in the current “big-data era” is to extract signals from huge databases. Often, an applied researcher proceeds in a two-step fashion: First, in order to decide whether there is any signal in the data at all, one performs an aggregate test of the global null hypothesis of no signal. This global null hypothesis is typically formulated as the high-dimensional target parameter being the zero vector. Second, if the global null hypothesis is rejected by the test, further analysis is undertaken to uncover the precise nature of the signal. Much research has been directed to studying properties of such a sequential rejection principle (cf. Romano and Wolf, 2005, Meinshausen, 2008, Romano, Shaikh, and Wolf, 2008, Rosenbaum, 2008, Yekutieli, 2008, Goeman and Solari, 2010, Heller et al., 2018, Bogomolov et al., 2020, and the references therein).

Using a powerful test for the global null hypothesis in the first step of such a hierarchical multistep procedure is of course crucial, and the development of tests for this hypothesis has therefore attracted much research in its own right. A typical choice, employed in, e.g., Heller, Meir, and Chatterjee (2019), is to use a test based on the Euclidean norm of the estimator. This also leads to the likelihood

We are grateful for the comments of the Editor, a Co-Editor, four referees, and the participants of the “High Voltage Econometrics” workshop, which helped to improve the previous version of the manuscript. Address correspondence to Anders Bredahl Kock, University of Oxford, 10 Manor Road, Oxford OX1 3UQ, UK; e-mail: anders.kock@economics.ox.ac.uk.

ratio (LR) test in the Gaussian sequence model they considered, which is also the framework in the present article. Although the LR test is a natural choice, one may ask: *Do tests for the global null exist that are consistent against substantially more alternatives than the LR test?* This question is practically relevant, because one can choose from a large menu of well-established tests, yet precisely which one to use is not obvious: For example, one could use tests based on other norms than the Euclidean one, a natural class of tests being based on p -norms (cf. the classic monograph of Ingster and Suslina (2003)). One could also use a test based on combining different p -norms as suggested by the power enhancement principle of Fan, Liao, and Yao (2015) and in Kock and Preinerstorfer (2021). The possibility of increasing power by combining tests has recently been applied in many types of high-dimensional testing problems (cf. Xu et al., 2016, Yang and Pan, 2017, Yu, Li, and Xue, 2020, He et al., 2021, Yu et al., 2021 [testing high-dimensional means and covariance matrices]; Zhang, Wang, and Shao, 2021 [change point detection]; Jammalamadaka, Meintanis, and Verdebout, 2020 [tests for uniformity on the sphere]; Feng et al., 2022 [tests for cross-sectional independence in high-dimensional panel data models]). Another test that has gained popularity in recent years is the Higher Criticism. This test dates back to Tukey (1976), and its strong power properties against deviations from the global null were first exhibited by Donoho and Jin (2004) and have led to much subsequent research (cf. Donoho and Jin, 2009; Hall and Jin, 2010; Arias-Castro, Candès, and Plan, 2011; Cai, Jeng, and Jin, 2011; Barnett and Lin, 2014; Li and Siegmund, 2015; Arias-Castro and Ying, 2019; Porter and Stewart, 2020). Alternatively, one could use tests based on combining p -values for coordinatewise zero restrictions. Important early works include Tippett (1931), Pearson (1933), Fisher (1934), Stouffer et al. (1949), and Simes (1986). For a review of the classic literature, see Cousins (2007), more recent contributions are Owen (2009), Duan et al. (2020), and Vovk and Wang (2020, 2021). It is crucial to highlight here that many of the above mentioned tests are consistent against strictly more alternatives than the LR test, i.e., they dominate the LR test in terms of their consistency properties; indeed, this is the main motivation of the power enhancement principle. Hence, the question of interest in the present article is not whether one can do better than the LR test at all, but whether one can do *substantially* better.

We consider the question raised in the previous paragraph from a high-dimensional perspective. In the Gaussian sequence model, we investigate whether aggregate tests can be obtained that are consistent against substantially more alternatives than the LR test. We show that relative to a uniform prior on the parameter space this is impossible: Essentially, we prove that for any given test the set of alternatives against which it is consistent, but the LR test is not, has vanishing relative Lebesgue measure. Hence, no test for the global null hypothesis can substantially improve on the LR test. From a technical perspective, our proofs are based on results by Schechtman and Schmuckenschläger (1991) concerning the asymptotic volume of intersections of p -norm balls and on the concentration of measure phenomenon.

Our finding is perhaps reminiscent of Le Cam (1953), who showed (in finite-dimensional settings and sufficiently regular models) that the set of possible superefficiency points of an estimator relative to the maximum likelihood estimator cannot be larger than a Lebesgue null set (cf. also van der Vaart (1997) for more discussion, Leeb and Pötscher (2006, 2008) for further elucidation, and, e.g., Han, Phillips, and Sul (2011) for superefficiency-type results in a nonregular model, namely a nonstationary autoregression). Note that our result does not imply that one should always use the LR test and not think carefully about the choice of test in high-dimensional testing problems. If, for example, one is interested in particular types of deviations from the null, e.g., sparse ones, there may be good reasons to use a test based on the supremum norm or the Higher Criticism. Furthermore, albeit very natural, the magnitude of the consistency set is merely one of many properties that can be used to compare tests. For example, tests are also frequently compared in terms of, e.g., their minimax detection properties or their local power against deviations from the null of a specific type. Nevertheless, in analogy to Le Cam (1953), regardless of how cleverly an alternative test is designed, the amount of alternatives against which one achieves an improvement as compared to the LR test cannot be substantial in terms of relative volume. This also supports basing a combination procedure, such as the power enhancement principle by Fan et al. (2015), on the Euclidean norm.

2. FRAMEWORK AND TERMINOLOGY

We consider the Gaussian sequence model

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

where $y_{1,d}, \dots, y_{d,d}$ are the observations, the parameters $\theta_{i,d} \in \mathbb{R}$ are unknown, $d \in \mathbb{N}$, and where the unobserved terms ε_i are independent and standard normal. Writing $\mathbf{y}_d = (y_{1,d}, \dots, y_{d,d})'$, $\boldsymbol{\varepsilon}_d = (\varepsilon_1, \dots, \varepsilon_d)'$, and $\boldsymbol{\theta}_d = (\theta_{1,d}, \dots, \theta_{d,d})' \in \mathbb{R}^d$, one can equivalently state the model in (1) as $\mathbf{y}_d = \boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d$.

In the model (1), we are interested in 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 $\mathbf{0}_d$ denotes the origin in \mathbb{R}^d . The null hypothesis $H_{0,d}$ is typically referred to as the “global null” of no effect.

The asymptotic analysis in the Gaussian sequence model (1) relies on $d \rightarrow \infty$. This is a high-dimensional regime in the sense that the number of parameters, d , tends to infinity. For each $d \in \mathbb{N}$, we observe a single realization of a d -dimensional Gaussian vector \mathbf{y}_d with mean $\boldsymbol{\theta}_d$ and identity covariance matrix. In this sense, the “sample size” is one for each d .

Remark 2.1. Although the Gaussian sequence model is an idealization, many fundamental issues of high dimensionality show up already here and insights obtained within this model carry over, at least on a conceptual level, to many

other settings. It is therefore widely recognized as an important prototypical framework in high-dimensional statistics (see, for example, Ingster and Suslina (2003), Carpentier and Verzelen (2019), Johnstone (2019), or Castillo and Roquain (2020)). To make this more precise, consider a situation where an estimator $\hat{\beta}_d$ for a target parameter $\beta_d \in \mathbb{R}^d$ is available the distribution of which is approximately normal, that is,

$$\hat{\beta}_d \approx \mathbb{N}(\beta_d, \Omega_d), \tag{3}$$

for Ω_d invertible. Suppose further that an invertible estimator $\hat{\Omega}_d \approx \Omega_d$ is at one's disposal, such that

$$\hat{\Omega}_d^{-1/2} \hat{\beta}_d \approx \mathbb{N}(\Omega_d^{-1/2} \beta_d, I_d).$$

Then, testing $\beta_d = \mathbf{0}_d$ on the basis of $\hat{\beta}_d$ and $\hat{\Omega}_d$ is approximated by testing $\theta_d := \Omega_d^{-1/2} \beta_d = \mathbf{0}_d$ in a Gaussian sequence model. Precise sets of conditions under which the above approximation statements hold depend on: (i) the formal meaning of the symbol “ \approx ”; (ii) the interplay of the dimension of the target parameter and the sample size used for computing $\hat{\beta}_d$; and (iii) particularities of the specific setup under consideration. This has been a topic of intense research interest, and sets of sufficient conditions for normal approximations in high-dimensional models can be found in, e.g., Portnoy (1985), Portnoy (1988), He and Shao (2000), Bentkus (2003), or in more recent work such as Chernozhukov, Chetverikov, and Kato (2017) and Giessing and Fan (2020). Working directly with a Gaussian sequence model allows us to bypass normal approximation results, and to focus on conceptual issues that already arise in a somewhat idealistic setup. We leave extensions to other models (e.g., via the above approximation heuristic) to future research.

Remark 2.2. To illustrate Remark 2.1 with a simple relevant model, consider a linear regression model with nonstochastic design matrix X , say, of full column rank d , and i.i.d. Gaussian errors with known variance, which we normalize to one for simplicity. Then the OLS estimator is Gaussian with expectation β_d , the regression coefficient vector, and covariance matrix $\Omega_d := (X'X)^{-1}$, which is known and can therefore be used as $\hat{\Omega}_d$. Note that in this situation the approximation statements in Remark 2.1 actually hold with equality.

For a given $d \in \mathbb{N}$, a (possibly randomized) test φ_d , say, for (2) is a (measurable) function from the sample space \mathbb{R}^d to the closed unit interval. In the asymptotic framework we consider, we are interested in properties of sequences of tests $\{\varphi_d\}$, where φ_d is a test for (2) for every $d \in \mathbb{N}$. To lighten the notation, we shall write φ_d instead of $\{\varphi_d\}$ whenever there is no risk of confusion. We are particularly interested in the consistency properties of sequences of tests. As usual, we say that a sequence of tests φ_d is consistent against the array of parameters $\vartheta = \{\theta_d : d \in \mathbb{N}\}$, where $\theta_d \in \mathbb{R}^d$ for every $d \in \mathbb{N}$, if and only if (as $d \rightarrow \infty$)

$$\mathbb{E}(\varphi_d(\theta_d + \varepsilon_d)) \rightarrow 1.$$

To every sequence of tests φ_d , we associate its *consistency set* $\mathcal{C}(\varphi_d)$, say. The consistency set $\mathcal{C}(\varphi_d)$ is the set of all arrays of parameters $\boldsymbol{\vartheta}$ the sequence of tests φ_d is consistent against. By definition,

$$\mathcal{C}(\varphi_d) \subseteq \prod_{d=1}^{\infty} \mathbb{R}^d =: \Theta,$$

the latter denoting the set of all possible arrays of parameters.

Recall that a sequence of tests φ_d is said to have *asymptotic size* $\alpha \in [0, 1]$ if $\mathbb{E}(\varphi_d(\boldsymbol{\varepsilon}_d)) \rightarrow \alpha$.

In this article, following the Neyman–Pearson paradigm, we focus on the case where $\alpha \in (0, 1)$, which we shall implicitly assume in the discussions throughout unless mentioned otherwise.

It is well known that the LR test for (2) rejects if the Euclidean norm $\|\cdot\|_2$ of the observation vector \mathbf{y}_d exceeds a critical value $\kappa_{d,2}$ chosen to satisfy the given size constraint. That is, the LR test is given by $\mathbb{1}\{\|\cdot\|_2 \geq \kappa_{d,2}\}$. For notational simplicity, we abbreviate the sequence of tests $\{\mathbb{1}\{\|\cdot\|_2 \geq \kappa_{d,2}\}\}$ by $\{2, \kappa_{d,2}\}$ and thus write $\mathcal{C}(\{2, \kappa_{d,2}\})$ for its consistency set. The following result is contained in Ingster and Suslina (2003) (cf. also Theorem 3.1 in Kock and Preinerstorfer (2021) for extensions).

THEOREM 2.1. *Let $\kappa_{d,2}$ be a sequence of critical values such that the asymptotic size of $\{2, \kappa_{d,2}\}$ is $\alpha \in (0, 1)$. Then*

$$\boldsymbol{\vartheta} \in \mathcal{C}(\{2, \kappa_{d,2}\}) \iff d^{-1/2} \|\boldsymbol{\theta}_d\|_2^2 \rightarrow \infty. \tag{4}$$

Theorem 2.1 shows that the consistency set of the LR test is precisely characterized by the asymptotic behavior of the Euclidean norms of the array of alternatives under consideration. That the consistency set of the LR test can be completely characterized in terms of the norm its test statistic is based on seems natural, but is quite specific to the LR test (see Theorem 3.1 and the ensuing discussion in Kock and Preinerstorfer (2021)).

3. SUPERCONSISTENCY POINTS

3.1. Improving on the LR Test

Although the LR test is a canonical choice of a test for the testing problem (2), there are many other reasonable tests available. For example, classic results by Birnbaum (1955) and Stein (1956) show that any test with convex acceptance region (i.e., the complement of its rejection region) is admissible. Anderson’s (1955) theorem implies that if the acceptance region is furthermore symmetric around the origin, then the test is also unbiased. Thus, any convex symmetric (around the origin) set delivers an admissible unbiased test, which is, hence, reasonable from a nonasymptotic point of view.

One class of tests that is intimately related to the LR test consists of tests based on other p -norms than the Euclidean one. For $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$ and $p \in (0, \infty]$, define the p -norm as usual via¹

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

In analogy to the LR test, p -norm-based tests reject if the p -norm of the observation vector exceeds a critical value $\kappa_{d,p}$. Special cases, which have an established tradition in high-dimensional inference, are the 1-norm and the supremum norm. We shall denote the sequence of tests $\{\mathbb{1}\{\|\cdot\|_p \geq \kappa_{d,p}\}\}$ by $\{p, \kappa_{d,p}\}$. Clearly, p -norm-based tests are unbiased and admissible for $p \in [1, \infty]$ as a consequence of the discussion in the first paragraph of this section.

Concerning the consistency sets $\mathcal{C}(\{p, \kappa_{d,p}\})$ of general p -norm-based tests, it is a somewhat surprising fact that:

- (i) $\mathcal{C}(\{p, \kappa_{d,p}\}) \subsetneq \mathcal{C}(\{q, \kappa_{d,q}\})$, for $0 < p < q < \infty$, i.e., strictly larger exponents p result in *strictly* larger consistency sets;
- (ii) this ranking does not extend to $q = \infty$, in the sense that there are alternatives the supremum norm-based test is not consistent against but against which any p -norm-based test with $p \in (0, \infty)$ is consistent and vice versa

(see Kock and Preinerstorfer (2021) for formal statements).² From (i), it follows that any p -norm-based test with $p \in (2, \infty)$ has a *strictly* larger consistency set than the LR test. We stress that this asymptotic strict domination of the LR test in terms of consistency sets is not in contradiction to its admissibility for each $d \in \mathbb{N}$.

Other tests that strictly dominate the LR test can be obtained, e.g., through combination procedures that enhance the LR test with a sequence of tests η_d that is sensitive against alternatives of a different “type” than the LR test in the sense that

$$\mathcal{C}(\eta_d) \setminus \mathcal{C}(\{2, \kappa_{d,2}\}) \neq \emptyset.$$

To see how this can be achieved, note that the consistency set of the sequence of tests ψ_d , say, where ψ_d rejects if the LR test *or* η_d rejects, contains $\mathcal{C}(\{2, \kappa_{d,2}\}) \cup \mathcal{C}(\eta_d)$, and hence dominates the LR test in terms of consistency. Essentially, this is the power enhancement principle of Fan et al. (2015) (see Kock and Preinerstorfer (2019) for related results and cf. Preinerstorfer (2021) for a nonasymptotic version of the power enhancement principle). Note that if η_d has asymptotic size 0, which is an assumption imposed on η_d in the context of the power enhancement principle,

¹Strictly speaking, $\|\cdot\|_p$ defines a norm on \mathbb{R}^d only for $p \in [1, \infty]$ and a quasi-norm for $p \in (0, 1)$.

²Recall that, throughout the present article, we implicitly impose the condition that all tests have asymptotic size in $(0, 1)$ if not otherwise mentioned.

nothing is lost in terms of asymptotic size when using ψ_d instead of the LR test, because both sequences of tests then have the same asymptotic size.³

To clarify how much can possibly be gained in terms of consistency by using a sequence of tests φ_d other than the LR test, we shall consider the corresponding set

$$\mathcal{C}(\varphi_d) \setminus \mathcal{C}(\{2, \kappa_{d,2}\}),$$

which we refer to as the *superconsistency points* of the sequence of tests φ_d (relative to the LR test). Note that the set of superconsistency points is defined for any sequence of tests, regardless of whether it dominates the LR test or not (in the sense that its consistency set includes that of the LR test).⁴ On a conceptual level, superconsistency points are related to superefficiency points of estimators relative to the maximum likelihood estimator in classic parametric theory.

3.2. The Relative Volume of the Set of Superconsistency Points

The central question we consider in this article is how “large” the set of superconsistency points $\mathcal{C}(\varphi_d) \setminus \mathcal{C}(\{2, \kappa_{d,2}\})$ can possibly be for a sequence of tests φ_d with asymptotic size in $(0, 1)$. Note that the larger $\mathcal{C}(\varphi_d) \setminus \mathcal{C}(\{2, \kappa_{d,2}\})$ is, the larger is the set of alternatives the sequence of tests φ_d is consistent against, but the LR test is not consistent against. Although we already know from the examples discussed in Section 3.1 that $\mathcal{C}(\varphi_d) \setminus \mathcal{C}(\{2, \kappa_{d,2}\})$ is nonempty for many φ_d , we here investigate whether one can *substantially* enlarge the consistency set by using another test than the LR test.

To make the above question amenable to a formal treatment, note that Theorem 2.1 implies that for any sequence of LR tests $\{2, \kappa_{d,2}\}$ with asymptotic size $\alpha \in (0, 1)$, the complement of $\mathcal{C}(\{2, \kappa_{d,2}\})$ satisfies

$$\Theta \setminus \mathcal{C}(\{2, \kappa_{d,2}\}) \supseteq \bigtimes_{d=1}^{\infty} \mathbb{B}_2^d(r_d),$$

if the sequence $r_d > 0$ is such that $r_d/d^{1/4}$ is bounded and where, for every $p > 0$, we denote by $\mathbb{B}_p^d(r)$ the closed p -norm ball with radius r centered at the origin. That is, the LR test is inconsistent against any element of $\bigtimes_{d=1}^{\infty} \mathbb{B}_2^d(r_d)$. We now investigate how many inconsistency points of the LR test can be removed from any such benchmark $\bigtimes_{d=1}^{\infty} \mathbb{B}_2^d(r_d)$ by erasing all superconsistency points of a sequence of tests φ_d .

³If η_d has a positive asymptotic size that is smaller than the asymptotic size targeted in the final combination test, one can work with an LR test with small enough asymptotic size in the combination procedure to obtain a test that dominates the LR test in terms of consistency (recall from Theorem 2.1 that the consistency set of the LR test does not depend on the specific value of the asymptotic size).

⁴To provide an example, for any $p \in (2, \infty)$, the set of superconsistency points of the p -norm-based test is fully characterized by Corollary 3.2 in Kock and Preinerstorfer (2021) (cf. also their Theorem 3.4, which essentially shows that these superconsistency points are approximately sparse and have at least one large entry).

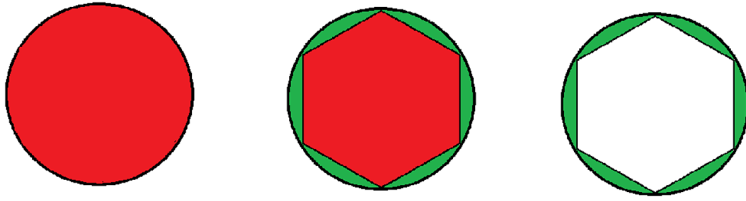


FIGURE 1. Illustration of $\mathbb{B}_2^d(r_d)$ (red), $\mathbb{D}_d \subseteq \mathbb{B}_2^d(r_d)$, and \mathbb{D}_d (green) for $d = 2$. Readers are referred to the online version for colored figures.

Formally, this is to be understood in the following sense: Let φ_d be a sequence of tests with consistency set $\mathcal{C}(\varphi_d)$, and let r_d be such that $r_d/d^{1/4}$ is bounded. Let $\mathbb{D}_d \subseteq \mathbb{B}_2^d(r_d)$ be such that

$$\bigtimes_{d=1}^{\infty} \mathbb{D}_d \subseteq \mathcal{C}(\varphi_d).$$

Note that all elements of $\bigtimes_{d=1}^{\infty} \mathbb{D}_d$ are superconsistency points of φ_d which are also contained in the benchmark $\bigtimes_{d=1}^{\infty} \mathbb{B}_2^d(r_d)$ (cf. the illustration in Figure 1). Denoting by vol_d the d -dimensional Lebesgue measure, we investigate the asymptotic behavior of the relative volume measure

$$\frac{\text{vol}_d(\mathbb{D}_d)}{\text{vol}_d(\mathbb{B}_2^d(r_d))}. \tag{5}$$

Obviously, the ratio in (5) is a number in $[0, 1]$. On the one hand, if this ratio is asymptotically close to 1, this means that, in terms of relative volume, many elements of the benchmark $\bigtimes_{d=1}^{\infty} \mathbb{B}_2^d(r_d)$ are superconsistency points of the sequence of tests φ_d . That is, one can substantially improve upon the LR test by using φ_d (or by combining the LR test with φ_d through the power enhancement principle). On the other hand, if this ratio is asymptotically close to 0, this means that in terms of relative volume only few elements of the benchmark are superconsistency points of φ_d .

Remark 3.1. One could also study the asymptotic behavior of the sequences $\text{vol}_d(\mathbb{B}_2^d(r_d)) - \text{vol}_d(\mathbb{D}_d)$ or $\text{vol}_d(\mathbb{D}_d)$ in order to determine whether one can substantially improve upon the LR test. However, these sequences both converge to 0. To see this, just note that

$$\text{vol}_d(\mathbb{B}_2^d(r_d)) = \frac{\pi^{d/2}}{\Gamma(d/2 + 1)} r_d^d \rightarrow 0,$$

in case $r_d/d^{1/4}$ is bounded as a consequence of Stirling’s approximation to the gamma function as well as $\mathbb{D}_d \subseteq \mathbb{B}_2^d(r_d)$. Thus, such “absolute” volume measures are uninformative, since even the absolute volume of $\mathbb{B}_2^d(r_d)$ tends to zero.

Remark 3.2. One may argue that, rather than (5), one should study

$$\frac{\text{vol}_d(\mathbb{D}_d)}{\text{vol}_d(\text{proj}_d[\Theta \setminus \mathcal{C}(\{2, \kappa_{d,2}\})])},$$

$\text{proj}_d(\cdot)$ denoting the projection onto the d th coordinate of its argument. However, since

$$\prod_{d=1}^K \mathbb{R}^d \times \prod_{d=K+1}^{\infty} \{0_d\} \subseteq \Theta \setminus \mathcal{C}(\{2, \kappa_{d,2}\}),$$

for all $K \in \mathbb{N}$, it follows that $\text{vol}_d(\text{proj}_d[\Theta \setminus \mathcal{C}(\{2, \kappa_{d,2}\})]) = \infty$, for all $d \in \mathbb{N}$.

We emphasize that using the (normalized) Lebesgue measure to assess the asymptotic magnitude of the set of superconsistency points is one among many possible choices. Other measures would be possible too, but the uniform prior over $\mathbb{B}_2^d(r_d)$ is a natural choice as in many situations there is no clear guidance concerning the type of alternative one wishes to favor.⁵

Note that the ratio in (5) depends on two ingredients:

1. the benchmark $\prod_{d=1}^{\infty} \mathbb{B}_2^d(r_d)$;
2. the sequence of superconsistency points $\prod_{d=1}^{\infty} \mathbb{D}_d$ which depends on the sequence of tests φ_d .

Therefore, one could suspect that the asymptotic behavior of (5) depends in a complicated way on the interplay between these two components. Nevertheless, it turns out that the asymptotic behavior of (5) has a simple description that does not depend on any of the two ingredients just described. In particular, we shall prove in Section 5 that the limit of the sequence is 0 for *all* sequences of tests φ_d . Hence, it is impossible to improve on the LR test in terms of the magnitude of its consistency set apart from a set of superconsistency points that is negligible in a relative volume sense.

In Section 4, we shall first establish this result for φ_d a sequence of p -norm-based tests with $p \in (2, \infty)$. Note that all these tests have a strictly larger consistency set than the LR test as discussed in Section 3.1. A general result, which also provides a rate of convergence, will be presented in Section 5.

4. p -NORM-BASED TESTS

We now consider the asymptotic behavior of the sequence (5) for the special case where φ_d is a sequence of p -norm-based tests with $p \in (2, \infty)$ being fixed. For this class of tests, we can exploit the characterization of their consistency sets provided in Theorem 3.1 and Corollary 3.2 of Kock and Preinerstorfer (2021), together with results from asymptotic geometry developed in Schechtman and

⁵Our results remain valid if, instead of measuring the magnitude of \mathbb{D}_d w.r.t. the uniform probability measure on $\mathbb{B}_2^d(r_d)$, one measures its magnitude w.r.t. the uniform probability measure on the Euclidean sphere of radius r_d . We will comment on this in Remark 5.1, but will focus on the uniform distribution on $\mathbb{B}_2^d(r_d)$ throughout the article.

Schmuckenschläger (1991) based on earlier results in Schechtman and Zinn (1990). These ingredients lead to a direct proof of the limit of the sequence in (5) being 0.

THEOREM 4.1. *Let $p \in (2, \infty)$, and let the sequence of critical values $\kappa_{d,p}$ be such that $\{p, \kappa_{d,p}\}$ has asymptotic size $\alpha \in (0, 1)$. Then, for any sequence $r_d > 0$ such that $r_d/d^{1/4}$ is bounded, and any sequence of nonempty Borel sets $\mathbb{D}_d \subseteq \mathbb{B}_2^d(r_d)$ such that*

$$\bigtimes_{d=1}^{\infty} \mathbb{D}_d \subseteq \mathcal{C}(\{p, \kappa_{d,p}\}), \tag{6}$$

we have

$$\lim_{d \rightarrow \infty} \frac{\text{vol}_d(\mathbb{D}_d)}{\text{vol}_d(\mathbb{B}_2^d(r_d))} = 0.$$

Proof. Let $\{p, \kappa_{d,p}\}$, r_d , and \mathbb{D}_d be as in the statement of the theorem. Corollary 3.2 in Kock and Preinerstorfer (2021) shows that $\vartheta \in \mathcal{C}(\{p, \kappa_{d,p}\})$ if and only if $d^{-1/2}(\|\theta_d\|_2^2 \vee \|\theta_d\|_p^p) \rightarrow \infty$. Together with $r_d/d^{1/4}$ being bounded, $\mathbb{D}_d \subseteq \mathbb{B}_2^d(r_d)$, and (6), this guarantees that $\tilde{s}_d/d^{1/(2p)} \rightarrow \infty$ for $\tilde{s}_d := \inf\{\|\theta_d\|_p : \theta_d \in \mathbb{D}_d\}$. The definition of \tilde{s}_d eventually implies

$$\mathbb{G}_d := \mathbb{B}_2^d(r_d) \setminus \mathbb{D}_d \supseteq \mathbb{B}_2^d(r_d) \cap \mathbb{B}_p^d(\tilde{s}_d/2).$$

Define the sequence $s_d := d^{1/(2p)-1/4}r_d > 0$, so that $s_d/d^{1/(2p)} = r_d/d^{1/4}$ is bounded. Hence, eventually $\tilde{s}_d \geq 2s_d$ and thus $\text{vol}_d(\mathbb{G}_d) \geq \text{vol}_d(\mathbb{B}_2^d(r_d) \cap \mathbb{B}_p^d(s_d))$ holds, so that the quotient

$$1 - \frac{\text{vol}_d(\mathbb{D}_d)}{\text{vol}_d(\mathbb{B}_2^d(r_d))} = \frac{\text{vol}_d(\mathbb{G}_d)}{\text{vol}_d(\mathbb{B}_2^d(r_d))}$$

is eventually not smaller than

$$\begin{aligned} \frac{\text{vol}_d(\mathbb{B}_2^d(r_d) \cap \mathbb{B}_p^d(s_d))}{\text{vol}_d(\mathbb{B}_2^d(r_d))} &= \frac{\text{vol}_d(\mathbb{B}_2^d(e_{d,2}) \cap \mathbb{B}_p^d(e_{d,2}s_d/r_d))}{\text{vol}_d(\mathbb{B}_2^d(e_{d,2}))} \\ &= \text{vol}_d(\mathbb{B}_2^d(e_{d,2}) \cap u_d \mathbb{B}_p^d(e_{d,p})), \end{aligned}$$

where $u_d := \frac{e_{d,2}}{e_{d,p}} \frac{d^{1/(2p)}}{d^{1/4}}$, $e_{d,p} := \frac{1}{2} \frac{\Gamma(1+d/p)}{\Gamma(1+1/p)}$, and consequently $\text{vol}_d(\mathbb{B}_2^d(e_{d,2})) = 1$. The main result in Schechtman and Schmuckenschläger (1991) shows that for every t large enough $\text{vol}_d(\mathbb{B}_2^d(e_{d,2}) \cap t\mathbb{B}_p^d(e_{d,p})) \rightarrow 1$, as $d \rightarrow \infty$. Therefore, we are done upon verifying that $u_d \rightarrow \infty$. This follows from the lower bound

$$\frac{e_{d,2}}{e_{d,p}} = \left[\frac{\Gamma(1+d/2)}{\Gamma(1+d/p)} \right]^{1/d} \frac{\Gamma(\frac{1}{p}+1)}{\Gamma(\frac{1}{2}+1)} \geq [d/p]^{1/2-1/p} \frac{\Gamma(\frac{1}{p}+1)}{\Gamma(\frac{1}{2}+1)},$$

where we used the inequality for ratios involving the gamma function in equation (12) of Jameson (2013) with “ $x = 1 + d/p$ ” (which is not smaller than 1) and “ $y = d(1/2 - 1/p)$ ” (which is not smaller than 0). □

Hence, even though $\mathcal{C}(\{p, \kappa_{d,p}\})$ contains the consistency set of the LR test as a *strict* subset for every $p \in (2, \infty)$ as discussed in Section 3.1, the subset of those alternatives in each benchmark $\times_{d=1}^{\infty} \mathbb{B}_2^d(r_d)$ for which the test $\{p, \kappa_{d,p}\}$ provides an improvement over the LR test is “negligible” in (relative) volume. That this result is not specific to p -norm-based tests but extends to *all* tests will be shown next.

5. UNRESTRICTED SEQUENCES OF TESTS

The proof of Theorem 4.1 builds heavily on the particular structure of the consistency set of p -norm-based tests. We shall now establish that *no* test can improve substantially on the LR test. In the absence of any structure on the tests, one can no longer exploit specific properties of the consistency set stemming from the test being based on a p -norm. Instead we rely on concentration results for Lipschitz continuous functions on the sphere as expounded in Ledoux (2001) (cf. also Ledoux (1992)).

THEOREM 5.1. *For every sequence of tests ψ_d with asymptotic size $\alpha \in (0, 1)$ and every sequence $r_d > 0$ such that $r_d/d^{1/4}$ is bounded, there exists an $\epsilon > 0$, such that for every sequence of nonempty Borel sets $\mathbb{D}_d \subseteq \mathbb{B}_2^d(r_d)$ satisfying*

$$\times_{d=1}^{\infty} \mathbb{D}_d \subseteq \mathcal{C}(\psi_d), \tag{7}$$

we have

$$\frac{\text{vol}_d(\mathbb{D}_d)}{\text{vol}_d(\mathbb{B}_2^d(r_d))} \leq \exp\left(-2\epsilon^2(d-2)/r_d^2\right) \quad \text{for all } d \text{ large enough}; \tag{8}$$

in particular, $\text{vol}_d(\mathbb{D}_d) / \text{vol}_d(\mathbb{B}_2^d(r_d))$ converges to 0 as $d \rightarrow \infty$.

The proof of Theorem 5.1 can be found in Appendix A. Note that Theorem 5.1 not only shows that the magnitude of superconsistency points is asymptotically negligible for any test, but it also shows that the measure of these points converges to zero quickly in the dimension d .

Remark 5.1 (Spherical measure instead of relative volume). One could ask what happens in the context of Theorem 5.1 if, instead of considering $\text{vol}_d(\mathbb{D}_d) / \text{vol}_d(\mathbb{B}_2^d(r_d))$ in (8), one considers $\rho_{d,r_d}(\mathbb{D}_d)$, where ρ_{d,r_d} denotes the uniform probability measure on the sphere

$$\mathbb{S}^{d-1}(r_d) := \{\xi \in \mathbb{R}^d : \|\xi\|_2 = r_d\}.$$

Inspection of the proof of Theorem 5.1 (cf. equation (A.2)) shows that the statement equally holds with $\text{vol}_d(\mathbb{D}_d)/\text{vol}_d(\mathbb{B}_2^d(r_d))$ replaced by $\rho_{d,r_d}(\mathbb{D}_d)$. That is, also with this alternative measure, one reaches the same conclusion concerning the magnitude of the set of superconsistency points of a sequence of tests relative to the LR test.

So far, all our results concerned consistency properties of tests. We were interested in the possible magnitude of the superconsistency points of a sequence of tests relative to the LR test and have seen that the magnitude of such points cannot be substantial. Although we now know that one cannot substantially improve on the LR test in terms of consistency (in the sense of Theorem 5.1), there could in principle exist sequences of tests that have larger power than the LR test on substantial portions of the parameter space (without the power there necessarily being close to 1). A nonasymptotic question one can therefore ask is: How large can such portions of the parameter space be? To answer this question, we introduce some more notation: Let $\alpha \in [0, 1]$ and denote for every $r > 0$ by $\beta_{d,\alpha}(r)$ the power of the LR test of size α against alternatives $\theta \in \mathbb{R}^d$ such that $\|\theta\|_2 = r$ (noting that the power of the LR test coincides for all such parameters as it is rotationally invariant).⁶ Denote the set of all tests $\psi : \mathbb{R}^d \rightarrow [0, 1]$ by Ψ_d , and define for every $\alpha \in [0, 1]$, $\epsilon > 0$, and $\psi \in \Psi_d$ the set $\mathbb{F}_d(\epsilon, \psi)$ as the subset of parameters against which the power of ψ exceeds the power of the LR test of the same size as ψ by more than ϵ , i.e.,

$$\mathbb{F}_d(\epsilon, \psi) := \left\{ \theta \in \mathbb{R}^d : \mathbb{E}(\psi(\theta + \mathbf{e}_d)) - \beta_{d,\alpha}(\|\theta\|_2) > \epsilon \text{ for } \alpha = \mathbb{E}(\psi(\mathbf{e}_d)) \right\}. \tag{9}$$

The question is: How large can this set be made by cleverly choosing ψ ? The following proposition (on which the proof of Theorem 5.1 rests) provides a nonasymptotic upper bound on its measure w.r.t. the uniform distribution $\rho_{d,r}$ on $\mathbb{S}^{d-1}(r)$. The upper bound decreases exponentially in d .

PROPOSITION 5.2. *For every $\epsilon > 0$, $r > 0$, $d \in \mathbb{N}$, and $\psi \in \Psi_d$, it holds that*

$$\rho_{d,r}(\mathbb{F}_d(\epsilon, \psi)) \leq \exp\left(-2\epsilon^2(d-2)/r^2\right). \tag{10}$$

Remark 5.2. Given a test $\psi \in \Psi_d$, note that $\mathbb{F}_d(\epsilon, \psi) \cap \mathbb{S}^{d-1}(r)$ is empty if $\beta_{d,\alpha}(r) + \epsilon \geq 1$, where $\alpha = \mathbb{E}(\psi(\mathbf{e}_d))$. For such values of r and ϵ , it obviously holds that $\rho_{d,r}(\mathbb{F}_d(\epsilon, \psi)) = 0$.

The proof of Proposition 5.2 is given next. It is mainly based on the observation that power functions in the model under consideration are Lipschitz continuous when restricted to spheres, and that such functions concentrate around their average.

⁶With this notation, it is worth noting that the proof of Theorem 5.1 shows that ϵ in that theorem can be chosen as $(1 - \limsup_{d \rightarrow \infty} \beta_{d,\alpha_d}(r_d))/2$, where α_d denotes the size of ψ_d .

Proof. Fix $\epsilon > 0$, $r > 0$, and $d \in \mathbb{N}$. The statement in (10) is trivially true if $d \leq 2$. Hence, we only need to consider the case where $d \geq 3$ holds. Let $\psi \in \Psi_d$ and denote its size by α . If $\alpha = 0$ or $\alpha = 1$, the inequality in (10) trivially holds as $\mathbb{F}_d(\epsilon, \psi) = \emptyset$ then follows. Hence, we only need to verify the claim for $\alpha \in (0, 1)$. It appears to be well known (a complete proof of this statement is given in Appendix B for the convenience of the reader) that the LR test (for the testing problem (2) in the model (1)) with size α maximizes the “weighted average power” (WAP)

$$\psi^* \mapsto \int_{\mathbb{S}^{d-1}(r)} \mathbb{E}(\psi^*(\boldsymbol{y} + \boldsymbol{\epsilon}_d)) d\rho_{d,r}(\boldsymbol{y}) \tag{11}$$

among all tests $\psi^* \in \Psi_d$ of size α . Therefore, $\int_{\mathbb{S}^{d-1}(r)} \mathbb{E}(\psi(\boldsymbol{y} + \boldsymbol{\epsilon}_d)) d\rho_{d,r}(\boldsymbol{y}) \leq \beta_{d,\alpha}(r)$, and we can conclude that $\rho_{d,r}(\mathbb{F}_d(\epsilon, \psi)) = \rho_{d,r}(\mathbb{F}_d(\epsilon, \psi) \cap \mathbb{S}^{d-1}(r))$ is bounded from above by

$$\rho_{d,r} \left(\left\{ \boldsymbol{\theta} \in \mathbb{S}^{d-1}(r) : \mathbb{E}(\psi(\boldsymbol{\theta} + \boldsymbol{\epsilon}_d)) \geq \int_{\mathbb{S}^{d-1}(r)} \mathbb{E}(\psi(\boldsymbol{y} + \boldsymbol{\epsilon}_d)) d\rho_{d,r}(\boldsymbol{y}) + \epsilon \right\} \right). \tag{12}$$

It is well known (and easy to verify using Pinsker’s inequality as in Lemma 2.5 in Tsybakov (2009)) that the total variation distance between two Gaussian distributions with covariance matrices equal to the identity and mean vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively, is bounded from above by $\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2/2$. This implies (by, e.g., Lemma 2.3 in Strasser (1985)) that the function $\boldsymbol{\theta} \mapsto 2\mathbb{E}(\psi(\boldsymbol{\theta} + \boldsymbol{\epsilon}_d))$ is Lipschitz continuous with constant 1. Since the geodesic distance between two points in $\mathbb{S}^{d-1}(r)$ is not smaller than the Euclidean distance between the two points, the function $\boldsymbol{\theta} \mapsto 2\mathbb{E}(\psi(\boldsymbol{\theta} + \boldsymbol{\epsilon}_d))$ is Lipschitz continuous with constant 1 on $\mathbb{S}^{d-1}(r)$ when the latter is equipped with the geodesic distance.

Multiplying both sides of the inequality in (12) by 2, and using the concentration inequality for Lipschitz continuous functions on spheres as given in the third display on page 222 of Ledoux (1992) (cf. also the discussion in Section 2.3 of Ledoux (2001)), we obtain

$$\rho_{d,r}(\mathbb{F}_d(\epsilon, \psi)) \leq \exp\left(-2\epsilon^2(d-2)/r^2\right). \tag{13}$$

6. CONCLUSION

In high-dimensional testing problems, the choice of a test implicitly or explicitly determines the type of alternative it prioritizes. In the Gaussian sequence model, the LR test is based on the Euclidean norm. Many tests exist that are consistent against alternatives the LR test is not consistent against (or are even consistent against strictly more alternatives than the LR test), i.e., they possess what we refer to as superconsistency points. We have shown that for any test, the corresponding set of superconsistency points is negligible in an asymptotic sense. This may be interpreted as a high-dimensional testing analog of Le Cam’s famous result that,

in sufficiently regular models, the set of superefficiency points relative to the maximum likelihood estimator is at most a Lebesgue null set (cf. Le Cam, 1953). In analogy to that classic finding, our result does not suggest that one should always use the LR test. However, it shows that there exists no test for which one can expect substantial improvements in terms of the magnitude of its set of superconsistency points.

APPENDIX A. Proof of Theorem 5.1

Let the sequence of tests ψ_d and r_d be as in the theorem’s statement. Denote the size of ψ_d by α_d . Since $r_d/d^{1/4}$ is bounded and $\alpha_d \rightarrow \alpha$, it follows from Theorem 2.1 that⁷

$$\beta := \limsup_{d \rightarrow \infty} \beta_{d, \alpha_d}(r_d) < 1.$$

Define $\epsilon = (1 - \beta)/2$. From (7), we obtain

$$c_d := \inf_{\theta \in \mathbb{D}_d} \mathbb{E}(\psi_d(\theta + \epsilon_d)) \rightarrow 1. \tag{A.1}$$

For all d large enough, we thus obtain $c_d > \beta_{d, \alpha_d}(r_d) + \epsilon$. Together with $\mathbb{D}_d \subseteq \mathbb{B}_2^d(r_d)$ and the function $r \mapsto \beta_{d, \alpha_d}(r)$ being nondecreasing, it therefore follows that $\mathbb{D}_d \subseteq \mathbb{F}_d(\epsilon, \psi_d)$ for all d large enough. Proposition 5.2, hence, allows us to conclude that, for all d large enough, we have

$$\rho_{d,r}(\mathbb{D}_d) \leq \rho_{d,r}(\mathbb{F}_d(\epsilon, \psi_d)) \leq \exp(-2\epsilon^2(d-2)/r^2) \quad \text{for every } r > 0. \tag{A.2}$$

For every $r > 0$, the push-forward measure of $\rho_d := \rho_{d,1}$ under the transformation $\mathbf{y} \mapsto r\mathbf{y}$, $\boldsymbol{\gamma} \in \mathbb{S}^{d-1} := \mathbb{S}^{d-1}(1)$ is $\rho_{d,r}$. Using polar coordinates (as in, e.g., Stroock (1998, Sect. 5.2)) and $\mathbb{D}_d \subseteq \mathbb{B}_2^d(r_d)$, we may express

$$\frac{\text{vol}_d(\mathbb{D}_d)}{\text{vol}_d(\mathbb{B}_2^d(r_d))} = \frac{d}{r_d^d} \int_{(0, r_d)} r^{d-1} \int_{\mathbb{S}^{d-1}} \mathbb{1}_{\mathbb{D}_d}\{r\boldsymbol{\gamma}\} d\rho_d(\boldsymbol{\gamma}) dr = \frac{d}{r_d^d} \int_{(0, r_d)} r^{d-1} \rho_{d,r}(\mathbb{D}_d) dr,$$

which, for all d large enough, we can upper bound by

$$\frac{1}{r_d^d} \int_{(0, r_d)} dr^{d-1} \exp(-2\epsilon^2(d-2)/r^2) dr \leq \exp(-2\epsilon^2(d-2)/r_d^2).$$

APPENDIX B. WAP Optimality of the LR Test

In this appendix, we provide an argument showing that (for $r > 0$) the LR test (for the testing problem (2) in the model (1)) with size α in $(0, 1)$ maximizes the weighted average power (WAP)

$$\psi^* \mapsto \int_{\mathbb{S}^{d-1}(r)} \mathbb{E}(\psi^*(\boldsymbol{\gamma} + \epsilon_d)) d\rho_{d,r}(\boldsymbol{\gamma})$$

among all tests $\psi^* \in \Psi_d$ of size α . To see this, denote the d -variate normal density with mean $\boldsymbol{\gamma}$ and identity covariance matrix by $\phi_{\boldsymbol{\gamma}}$ and note that (by the Neyman–Pearson lemma)

⁷Throughout this proof, we use the notation that was introduced in the context of Proposition 5.2.

the test which maximizes WAP (i.e., which is WAP optimal) is the LR test for the simple hypothesis where (i) the density under the null equals ϕ_{0_d} and (ii) the density under the alternative equals $\int \phi_{\mathcal{Y}} d\rho_{d,r}(\mathcal{Y})$. This test rejects for the observation y if and only if

$$\int_{\mathbb{S}^{d-1}(r)} \phi_{\mathcal{Y}}(y) / \phi_{0_d}(y) d\rho_{d,r}(\mathcal{Y}) = \exp(-r^2/2) \int_{\mathbb{S}^{d-1}(r)} \exp(\mathcal{Y}'y) d\rho_{d,r}(\mathcal{Y}) \quad (\text{B.1})$$

exceeds a critical value $C_{\alpha,r}$, say, which is chosen such that the test has size α . Note that the measure $\rho_{d,r}$ coincides with its push-forward measure under any orthonormal linear transformation $U: \mathbb{R}^d \rightarrow \mathbb{R}^d$, i.e., $\rho_{d,r}$ is “rotationally invariant.” Choosing U orthonormal and such that Uy coincides with $\|y\|_2$ times the first element of the canonical basis of \mathbb{R}^d , it follows that the integral to the right of (B.1) coincides with

$$\int_{\mathbb{S}^{d-1}(r)} \exp(\|y\|_2 \gamma_1) d\rho_{d,r}(\mathcal{Y}) = \frac{1}{2} \left(\int_{\mathbb{S}^{d-1}(r)} [\exp(\|y\|_2 \gamma_1) + \exp(-\|y\|_2 \gamma_1)] d\rho_{d,r}(\mathcal{Y}) \right).$$

Since the function $a \mapsto \exp(a\gamma_1) + \exp(-a\gamma_1)$ is nondecreasing on $[0, \infty)$ for every γ_1 , it follows that the WAP optimal test rejects if and only if $\|y\|_2$ exceeds a critical value (chosen so that the test has the right size). In other words, the LR test (for the testing problem (2) in the model (1)) is WAP optimal.

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, 170–176.
- Arias-Castro, E., E.J. Candès, & Y. Plan (2011) Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Annals of Statistics* 39, 2533–2556.
- Arias-Castro, E. & A. Ying (2019) Detection of sparse mixtures: Higher criticism and scan statistic. *Electronic Journal of Statistics* 13, 208–230.
- Barnett, I. & X. Lin (2014) Analytical P -value calculation for the higher criticism test in finite d problems. *Biometrika* 101, 964–970.
- Bentkus, V. (2003) On the dependence of the Berry–Esseen bound on dimension. *Journal of Statistical Planning and Inference* 113, 385–402.
- Birnbaum, A. (1955) Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests. *Annals of Mathematical Statistics* 26, 21–36.
- Bogomolov, M., C.B. Peterson, Y. Benjamini, & C. Sabatti (2020) Hypotheses on a tree: New error rates and testing strategies. *Biometrika* 108, 575–590.
- Cai, T., J. Jeng, & J. Jin (2011) Optimal detection of heterogeneous and heteroscedastic mixtures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73, 629–662.
- Carpentier, A. & N. Verzelen (2019) Adaptive estimation of the sparsity in the Gaussian vector model. *Annals of Statistics* 47, 93–126.
- Castillo, I. & E. Roquain (2020) On spike and slab empirical Bayes multiple testing. *Annals of Statistics* 48, 2548–2574.
- Chernozhukov, V., D. Chetverikov, & K. Kato (2017) Central limit theorems and bootstrap in high dimensions. *Annals of Probability* 45, 2309–2352.
- Cousins, R. (2007) Annotated bibliography of some papers on combining significances or p -values. *Preprint*, arXiv:0705.2209.
- Donoho, D. & J. Jin (2004) Higher criticism for detecting sparse heterogeneous mixtures. *Annals of Statistics* 32, 962–994.

- Donoho, D. & J. Jin (2009) Feature selection by higher criticism thresholding achieves the optimal phase diagram. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367, 4449–4470.
- Duan, B., A. Ramdas, S. Balakrishnan, & L. Wasserman (2020) Interactive martingale tests for the global null. *Electronic Journal of Statistics* 14, 4489–4551.
- Fan, J., Y. Liao, & J. Yao (2015) Power enhancement in high-dimensional cross-sectional tests. *Econometrica* 83, 1497–1541.
- Feng, L., T. Jiang, B. Liu, & W. Xiong (2022) Max-sum tests for cross-sectional independence of high-dimensional panel data. *Annals of Statistics* 50, 1124–1143.
- Fisher, R.A. (1934) *Statistical Methods for Research Workers*, 5th Edition. Oliver and Boyd.
- Giessing, A. & J. Fan (2020) Bootstrapping ℓ_p -statistics in high dimensions. *Preprint*, [arXiv:2006.13099](https://arxiv.org/abs/2006.13099).
- Goeman, J.J. & A. Solari (2010) The sequential rejection principle of familywise error control. *Annals of Statistics* 38, 3782–3810.
- Hall, P. & J. Jin (2010) Innovated higher criticism for detecting sparse signals in correlated noise. *Annals of Statistics* 38, 1686–1732.
- Han, C., P.C.B. Phillips, & D. Sul (2011) Uniform asymptotic normality in stationary and unit root autoregression. *Econometric Theory* 27, 1117–1151.
- He, X. & Q.-M. Shao (2000) On parameters of increasing dimensions. *Journal of Multivariate Analysis* 73, 120–135.
- He, Y., G. Xu, C. Wu, & W. Pan (2021) Asymptotically independent U-statistics in high-dimensional testing. *Annals of Statistics* 49, 154–181.
- Heller, R., N. Chatterjee, A. Krieger, & J. Shi (2018) Post-selection inference following aggregate level hypothesis testing in large-scale genomic data. *Journal of the American Statistical Association* 113, 1770–1783.
- Heller, R., A. Meir, & N. Chatterjee (2019) Post-selection estimation and testing following aggregate association tests. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 81, 547–573.
- Ingster, Y. & I.A. Suslina (2003) *Nonparametric Goodness-of-Fit Testing under Gaussian Models*. Springer.
- Jameson, G.J.O. (2013) Inequalities for gamma function ratios. *The American Mathematical Monthly* 120, 936–940.
- Jammalamadaka, S.R., S. Meintanis, & T. Verdebout (2020) On Sobolev tests of uniformity on the circle with an extension to the sphere. *Bernoulli* 26, 2226–2252.
- Johnstone, I.M. (2019). Gaussian estimation: Sequence and wavelet models. Unpublished manuscript.
- Kock, A.B. & D. Preinerstorfer (2019) Power in high-dimensional testing problems. *Econometrica* 87, 1055–1069.
- Kock, A.B. & Preinerstorfer, D. (2021). Consistency of p -norm based tests in high dimensions: Characterization, monotonicity, domination. *Preprint*, [arXiv:2103.11201](https://arxiv.org/abs/2103.11201).
- Le Cam, L. (1953) On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publications in Statistics* 1, 277–330.
- Ledoux, M. (1992) A heat semigroup approach to concentration on the sphere and on a compact Riemannian manifold. *Geometric and Functional Analysis* 2, 221–224.
- Ledoux, M. (2001) *The Concentration of Measure Phenomenon*. Mathematical Surveys and Monographs, vol. 89. American Mathematical Society.
- Leeb, H. & B.M. Pötscher (2006) Performance limits for estimators of the risk or distribution of shrinkage-type estimators, and some general lower risk-bound results. *Econometric Theory* 22, 69–97.
- Leeb, H. & B.M. Pötscher (2008) Sparse estimators and the oracle property, or the return of Hodges' estimator. *Journal of Econometrics* 142, 201–211.
- Li, J. & D. Siegmund (2015) Higher criticism: p -values and criticism. *Annals of Statistics* 43, 1323–1350.

- Meinshausen, N. (2008) Hierarchical testing of variable importance. *Biometrika* 95, 265–278.
- Owen, A. (2009) Karl Pearson's meta-analysis revisited. *Annals of Statistics* 37, 3867–3892.
- Pearson, K. (1933) On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika* 25, 379–410.
- Porter, T. & M. Stewart (2020) Beyond HC: More sensitive tests for rare/weak alternatives. *Annals of Statistics* 48, 2230–2252.
- Portnoy, S. (1985) Asymptotic behavior of M estimators of p regression parameters when p^2/n is large; II. Normal approximation. *Annals of Statistics* 13, 1403–1417.
- Portnoy, S. (1988) Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Annals of Statistics* 16, 356–366.
- Preinerstorfer, D. (2021) How to avoid the zero-power trap in testing for correlation. *Econometric Theory*, to appear.
- Romano, J. & M. Wolf (2005) Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association* 100, 94–108.
- Romano, J.P., A.M. Shaikh, & M. Wolf (2008) Formalized data snooping based on generalized error rates. *Econometric Theory* 24, 404–447.
- Rosenbaum, P. (2008) Testing hypotheses in order. *Biometrika* 95, 248–252.
- Schechtman, G. & M. Schmuckenschläger (1991) Another remark on the volume of the intersection of two L_p^n balls. In J. Lindenstrauss and V. D. Milman (eds.), *Geometric Aspects of Functional Analysis*, pp. 174–178. Springer.
- Schechtman, G. & J. Zinn (1990) On the volume of the intersection of two L_p^n balls. *Proceedings of the American Mathematical Society* 110, 217–224.
- Simes, J. (1986) An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73, 751–754.
- Stein, C. (1956) The admissibility of Hotelling's t^2 -test. *Annals of Mathematical Statistics* 27, 616–623.
- Stouffer, S.A., Suchman, E.A., DeVinney, L.C., Star, S.A., & Williams Jr, R.M. (1949) *The American Soldier: Adjustment During Army Life (Studies in Social Psychology in World War II)*, vol. 1. Princeton Univ. Press.
- Strasser, H. (1985) *Mathematical Theory of Statistics*. Walter de Gruyter.
- Stroock, D.W. (1998) *A Concise Introduction to the Theory of Integration*. Springer.
- Tippett, L.H.C. (1931) *The Methods of Statistics*. Williams & Norgate Ltd.
- Tsybakov, A.B. (2009) *Introduction to Nonparametric Estimation*. Springer.
- Tukey, J. (1976) *T13 N: The Higher Criticism. Course Notes*. Statistics 411. Princeton University.
- van der Vaart, A.W. (1997) Superefficiency. In D. Pollard, E. Torgersen, & G. L. Yang (eds.), *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, pp. 397–410. Springer.
- Vovk, V. & R. Wang (2020) Combining p -values via averaging. *Biometrika* 107, 791–808.
- Vovk, V. & R. Wang (2021) E -values: Calibration, combination and applications. *Annals of Statistics* 49, 1736–1754.
- Xu, G., L. Lin, P. Wei, & W. Pan (2016) An adaptive two-sample test for high-dimensional means. *Biometrika* 103, 609–624.
- Yang, Q. & G. Pan (2017) Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices. *Journal of the American Statistical Association* 112, 188–200.
- Yekutieli, D. (2008) Hierarchical false discovery rate-controlling methodology. *Journal of the American Statistical Association* 103, 309–316.
- Yu, X., D. Li, & L. Xue (2020) Fisher's combined probability test for high-dimensional covariance matrices. *Journal of the American Statistical Association*, to appear.
- Yu, X., D. Li, L. Xue, & R. Li (2021) Power-enhanced simultaneous test of high-dimensional mean vectors and covariance matrices with application to gene-set testing. *Journal of the American Statistical Association*, to appear.
- Zhang, Y., R. Wang, & X. Shao (2021) Adaptive inference for change points in high-dimensional data. *Journal of the American Statistical Association*, to appear.