Statistica Sinica Preprint No: SS-2021-0158	
Title	Rates of Bootstrap Approximation for Eigenvalues in High-Dimensional PCA
Manuscript ID	SS-2021-0158
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0158
Complete List of Authors	Junwen Yao and Miles Lopes
Corresponding Authors	Miles Lopes
E-mails	melopes@ucdavis.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

Statistica Sinica

Rates of Bootstrap Approximation for Eigenvalues in High-Dimensional PCA

Junwen Yao and Miles E. Lopes

University of California, Davis

Abstract: In the context of principal components analysis (PCA), the bootstrap is commonly applied to solve a variety of inference problems, such as constructing confidence intervals for the eigenvalues of the population covariance matrix Σ. However, when the data are high-dimensional, there are relatively few theoretical guarantees that quantify the performance of the bootstrap. Our aim in this paper is to analyze how well the bootstrap can approximate the joint distribution of the leading eigenvalues of the sample covariance matrix $\widehat{\Sigma}$, and we establish non-asymptotic rates of approximation with respect to the multivariate Kolmogorov metric. Under certain assumptions, we show that the bootstrap can achieve a dimension-free rate of $\mathtt{r}(\Sigma)/\sqrt{n}$ up to logarithmic factors, where $\mathtt{r}(\Sigma)$ is the effective rank of Σ, and n is the sample size. From a methodological standpoint, we show that applying a transformation to the eigenvalues of $\widehat{\Sigma}$ before bootstrapping is an important consideration in high-dimensional settings.

Key words and phrases: Bootstrap, covariance matrices, high-dimensional statistics, principal components analysis.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

1. Introduction

The applications of the bootstrap in principal components analysis (PCA) go back almost as far as the advent of the bootstrap itself (Diaconis and Efron, 1983), and over the years such applications have become part of standard practice in multivariate analysis (Davison and Hinkley, 1997; Jolliffe, 2002; Olive, 2017). With regard to theory, there is also a well-established set of asymptotic results showing that the bootstrap generally works in the context of PCA with low-dimensional data (Beran and Srivastava, 1985; Eaton and Tyler, 1991). Furthermore, in aberrant situations where the bootstrap is known to encounter difficulty in low dimensions, such as in the case of tied population eigenvalues, various remedies have been proposed and analyzed (Beran and Srivastava, 1985; Dümbgen, 1993; Hall et al., 2009).

However, in the context of PCA with high-dimensional data, the relationship between theory and practice is quite different. On one hand, bootstrap methods are popular among practitioners for solving inference problems related to high-dimensional PCA (e.g., Wagner, 2015; Fisher et al., 2016; Webb-Vargas et al., 2017; Terry et al., 2018; Li and Ralph, 2019; Nguyen and Holmes, 2019; Stewart et al., 2019). On the other hand, the theory for describing these methods is relatively incomplete.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

To develop a more precise understanding of the bootstrap in this context, we focus on the fundamental problem of approximating the joint distribution of the leading eigenvalues $\lambda_1(\widehat{\Sigma}), \ldots, \lambda_k(\widehat{\Sigma})$ of a sample covariance matrix $\widehat{\Sigma} \in \mathbb{R}^{p \times p}$, where $k < p$. (Precise defnitions will be given later.) Because the fluctuations of these eigenvalues are relevant to many inference tasks, this problem plays a central role in multivariate analysis, and is also of broad interest in other areas, such as signal processing (Couillet and Debbah, 2011) and finance (Ruppert and Matteson, 2015). Below, we summarize some examples of inference tasks involving sample eigenvalues. These tasks are illustrated using real-data examples based on stock market returns in Section S9 of the Supplementary Material.

- *Selecting principal components.* A key step in any implementation of PCA is to choose the number of principal components, and many established techniques for making this choice are informed by the distributions of eigenvalue-based statistics. Examples of these statistics include eigengaps $\lambda_j(\widehat{\Sigma}) - \lambda_{j+1}(\widehat{\Sigma})$, the proportions of explained variance $(\lambda_1(\widehat{\Sigma}) + \cdots + \lambda_k(\widehat{\Sigma}))/\operatorname{tr}(\widehat{\Sigma})$, as well as the componentwise proportions $\lambda_j(\widehat{\Sigma})/\operatorname{tr}(\widehat{\Sigma})$ for $j = 1, \ldots, k$. Other selection rules are based on confidence intervals for the eigenvalues $\lambda_1(\Sigma), \ldots, \lambda_k(\Sigma)$ of the population covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$. The construction of such

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

intervals is linked directly to the distribution of the eigenvalues of $\widehat{\Sigma}$. For a general overview of selection rules, we refer to Jolliffe (2002).

- *Quantifying uncertainty.* The eigenvalues of a population covariance matrix arise as unknown parameters of interest in many situations beyond the selection of principal components. For instance, these parameters govern the performance of statistical methods for covariance estimation, regression, and classification (Ledoit and Wolf, 2012; Hsu et al., 2014; Dobriban and Wager, 2018). These parameters also have domain-specific meanings in applications ranging from portfolio selection to ecology (Fabozzi et al., 2007; Chen et al., 2019). Consequently, it becomes necessary to quantify the uncertainty associated with the population eigenvalues, such as in constructing confidence intervals for them—and again, this leads to the use of distributional approximation results for the sample eigenvalues.

Although there is an extensive literature on distributional approximations for sample eigenvalues, this body of work primarily focuses on asymptotic results involving analytical formulas. Roughly speaking, the bulk of the literature can be divided into two parts, dealing either with classical asymptotics, where p is held fixed as $n \to \infty$ (Anderson, 2003), or with high-dimensional asymptotics, where p/n converges to a positive constant

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

as p and n diverge simultaneously (Bai and Silverstein, 2010). In either case, an essential limitation is that asymptotic results do not usually quantify how close the limiting distribution is to the finite-sample distribution. In more practical terms, this means it is often difficult to know if tests statistics and confidence intervals are well calibrated (i.e., if their actual levels and coverage probabilities are close to the nominal values). A second limitation is that approximations based on analytical formulas are often tied to specific model assumptions, which can make it difficult to adapt such formulas outside of a given model.

With regard to the second limitation, bootstrap methods have an advantage insofar as they do not rely on formulas, and hence can be applied in a more flexible manner. Nevertheless, the existing work on bootstrap methods for PCA still tends to suffer from the first limitation above, since the results are generally asymptotic (Beran and Srivastava, 1985; Eaton and Tyler, 1991; El Karoui and Purdom, 2019). Thus, a key motivation for our work is to provide results that explicitly quantify the accuracy of bootstrap approximation in terms of the sample size n and the effective rank of Σ. (For example, our results can be used to quantify how close the coverage probabilities of bootstrap confidence intervals are to the nominal values.) Another motivation for our study is that, until quite recently, most of the lit-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

erature on bootstrap methods for PCA has been limited to low-dimensional settings. Consequently, it is of general interest to establish a more complete theoretical description of bootstrap methods for high-dimensional PCA—a point that was highlighted in a recent survey on this topic (Johnstone and Paul, 2018, §X.C).

1.1 Contributions

Let $X_1, \dots, X_n \in \mathbb{R}^p$ be centered i.i.d. observations with population covariance matrix $\Sigma = \mathbb{E}[X_1 X_1^\top]$. Also, let $\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^n X_i X_i^\top$ denote the associated sample covariance matrix, and let $\widehat{\Sigma}^\star = \frac{1}{n}\sum_{i=1}^n X_i^\star (X_i^\star)^\top$ be its bootstrap version, formed from random vectors $X_1^\star, \dots, X_n^\star$ that are sampled with replacement from the observations. In addition, let the eigenvalues of a symmetric matrix $A \in \mathbb{R}^{p\times p}$ be denoted as $\lambda_1(A) \geq \cdots \geq \lambda_p(A)$, and let $\boldsymbol{\lambda}_k(A) = (\lambda_1(A), \dots, \lambda_k(A))$ for a fixed integer $k < p$.

In this notation, our goal is to establish non-asymptotic bounds on the multivariate Kolmogorov distance

$$\Delta_n = \sup_{t\in\mathbb{R}^k} \left| \mathbb{P}\Big(\sqrt{n}\big(\boldsymbol{\lambda}_k(\widehat{\Sigma}) - \boldsymbol{\lambda}_k(\Sigma)\big) \preceq t\Big) - \mathbb{P}\Big(\sqrt{n}\big(\boldsymbol{\lambda}_k(\widehat{\Sigma}^\star) - \boldsymbol{\lambda}_k(\widehat{\Sigma})\big) \preceq t \,\Big|\, X\Big) \right|,$$

where the relation $v \preceq w$ between two vectors $v, w \in \mathbb{R}^k$ means $v_j \leq w_j$ for all $j = 1, \dots, k$, and $\mathbb{P}(\cdot\,|X)$ refers to probability that is conditional on $X_1, \dots, X_n$. Under certain conditions, our central result (Theorem 1) shows

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

that the dimension-free bound

$$\Delta_n \;\leq\; \frac{C_n\,\mathtt{r}(\Sigma)}{\sqrt{n}} \tag{1.1}$$

holds with high probability, where $C_n > 0$ is a polylogarithmic function of n, and the quantity $\mathtt{r}(\Sigma)$ is the effective rank of Σ, defined by $\mathtt{r}(\Sigma) = \mathrm{tr}(\Sigma)/\lambda_1(\Sigma)$.

Several aspects of the bound (1.1) and the parameter $\mathtt{r}(\Sigma)$ are worth noting. First, the effective rank satisfies $1 \leq \mathtt{r}(\Sigma) \leq p$ whenever Σ is nonzero, and can be interpreted as a proxy for the number of "dominant" principal components of Σ. Hence, even in very high-dimensional settings, where $n \ll p$, the bound (1.1) shows that the bootstrap can perform well if the number of dominant components is not too large, which is precisely the situation where high-dimensional PCA is of greatest interest. Meanwhile, even in situations where $\mathtt{r}(\Sigma)$ is moderately large, e.g., $\mathtt{r}(\Sigma) \to \infty$ with $\mathtt{r}(\Sigma) = o(\sqrt{n})$, the bound (1.1) is still able to quantify the accuracy of the bootstrap. Indeed, both of these points are borne out by our numerical experiments in Section 3, which confirm that the performance of the bootstrap is governed more by $\mathtt{r}(\Sigma)$ than it is by p, and that the bootstrap can still be accurate when $\mathtt{r}(\Sigma)$ is moderately large. More generally, it should also be mentioned that effective rank has attracted attention in

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

many other aspects of high-dimensional PCA (e.g., Lounici, 2014; Bunea and Xiao, 2015; Koltchinskii and Lounici, 2017; Jung et al., 2018; Naumov et al., 2019; Koltchinskii et al., 2020).

As an alternative to approximating the distribution of $\sqrt{n}\big(\boldsymbol{\lambda}_k(\widehat{\Sigma}) - \boldsymbol{\lambda}_k(\Sigma)\big)$ by bootstrapping in a direct manner, it can be advantageous to use a *transformation* prior to bootstrapping, which is a fundamental topic in the bootstrap literature (e.g., DiCiccio, 1984; Tibshirani, 1988; Konishi, 1991; DiCiccio and Efron, 1996; Davison and Hinkley, 1997; Chernick, 2011). To be more specific, let h be a univariate scalar function, referred to as a transformation, and for any symmetric matrix $A \in \mathbb{R}^{p\times p}$, let $\boldsymbol{h}(\boldsymbol{\lambda}_k(A)) = (h(\lambda_1(A)), \ldots, h(\lambda_k(A)))$. Then, the conditional distribution of $\sqrt{n}(\boldsymbol{h}(\boldsymbol{\lambda}_k(\widehat{\Sigma}^\star)) - \boldsymbol{h}(\boldsymbol{\lambda}_k(\widehat{\Sigma})))$ given the observations can be used to approximate the distribution of $\sqrt{n}(\boldsymbol{h}(\boldsymbol{\lambda}_k(\widehat{\Sigma})) - \boldsymbol{h}(\boldsymbol{\lambda}_k(\Sigma)))$. (Additional discussion is provided in Sections 2 and 3.) For instance, a classical choice of transformation is $h(x) = \log(x)$, because it is known to be variance-stabilizing under certain conditions when $n \to \infty$, with p held fixed (Beran and Srivastava, 1985). With this in mind, a second contribution our analysis is an extended version of the bound (1.1) that can accommodate certain transformations (see Theorem 2).

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

From a more methodological standpoint, our numerical experiments also shed new light on the role of transformations in bootstrap methods for high-dimensional PCA. Although we confirm that the classical logarithm transformation can be beneficial in low dimensions, we show that it is less effective when $\mathtt{r}(\Sigma)$ is moderately large. Consequently, we explore some alternative transformations, and provide numerical results demonstrating that there are opportunities to improve upon $h(x) = \log(x)$ in high dimensions. To put such empirical findings into perspective, we are not aware any prior work investigating how transformations can be used to enhance bootstrap methods in this context.

1.2 Related work

Quite recently, there has been an acceleration in the pace of research on bootstrap methods for high-dimensional sample covariance matrices, as evidenced in the papers (Han et al., 2018; Johnstone and Paul, 2018; El Karoui and Purdom, 2019; Lopes et al., 2019, 2022; Naumov et al., 2019). Among these, the most relevant to our work is (El Karoui and Purdom, 2019), which examines both the successes and failures of the bootstrap in doing inference with the leading eigenvalues of $\widehat{\Sigma}$. In the negative direction, that paper focuses on a specialized model with $\lambda_1(\Sigma) > 1$ and $\lambda_2(\Sigma) = \cdots = \lambda_p(\Sigma) = 1$.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

This model also corresponds to a very large effective rank $\mathtt{r}(\Sigma) \asymp p$ that makes dimension reduction via PCA inherently difficult. In the positive direction, that paper deals with a different situation where Σ is assumed to have a near low-rank structure of the form

$$\Sigma = \begin{pmatrix} A & B \\ B^\top & C(\eta) \end{pmatrix}, \tag{1.2}$$

where A is of size $k \times k$ with $k \asymp 1$, and the diagonal blocks satisfy $\lambda_1(A) \asymp 1$, and $\lambda_1(C(\eta)) \lesssim n^{-\eta}$ for a fixed parameter $\eta > 1/2$. Working under an elliptical model, (El Karoui and Purdom, 2019) show that the bootstrap consistently approximates the distribution of $\sqrt{n}\big(\boldsymbol{\lambda}_k(\widehat{\Sigma}) - \boldsymbol{\lambda}_k(\Sigma)\big)$ in an asymptotic framework where $p/n \lesssim 1$. In relation to our work, the most crucial distinction is that our results quantify the accuracy of the bootstrap with non-asymptotic *rates of approximation*. To illustrate the significance of this, note that our bound (1.1) provides an explicit link between the size of $\mathtt{r}(\Sigma)$ and the accuracy of the bootstrap, whereas in an asymptotic setup, the effect of $\mathtt{r}(\Sigma)$ is hidden—because it "washes out in the limit." Our numerical experiments will also confirm that different sizes of $\mathtt{r}(\Sigma)$ can have an appreciable effect on the finite-sample accuracy of the bootstrap. In this way, our work indicates that the quantity $\mathtt{r}(\Sigma)/\sqrt{n}$ serves as a type of conceptual diagnostic for assessing the reliability of the bootstrap in

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

high-dimensional PCA.

Beyond these points of contrast with (El Karoui and Purdom, 2019), there are several distinctions with regard to model assumptions. First, we work in a dimension-free setting where there are no restrictions on the size of p with respect to n. Second, the model based on (1.2) implicitly requires that $\lambda_j(\Sigma) \lesssim n^{-\eta}$ for all $j \geq k+1$, whereas this constraint on Σ is not used here. Third, it is straightforward to check that in the model based on (1.2) with $p/n \lesssim 1$, the condition $\eta > 1/2$ *implies* $\mathtt{r}(\Sigma) = o(n^{1/2-\epsilon})$ for some fixed $\epsilon > 0$. Hence, under these conditions, our bound (1.1) also implies bootstrap consistency. In addition, the bound (1.1) can ensure bootstrap consistency in models that are outside the scope of (1.2). For example, this occurs if $p \asymp e^{m(n)}$ for some sequence of integers satisfying $m(n) = o(\sqrt{n}/C_n)$, and if $\lambda_j(\Sigma) \asymp j^{-1}$.

Other works on bootstrap methods related to high-dimensional sample covariance matrices deal with models or statistics that are qualitatively different from those considered here. The papers (Han et al., 2018; Lopes et al., 2022) look at bootstrapping the operator norm error $\sqrt{n}\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}}$, as well as variants of this statistic, such as $\sup_{u \in \mathcal{U}} \sqrt{n}|u^\top(\widehat{\Sigma} - \Sigma)u|/u^\top \Sigma u$, where $\mathcal{U}$ is a set of sparse vectors in the unit sphere of $\mathbb{R}^p$. In a different

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

direction, (Lopes et al., 2019) focus on "linear spectral statistics" of the form $\frac{1}{p}\sum_{j=1}^{p} f(\lambda_j(\widehat{\Sigma}))$, where $f : [0, \infty) \to \mathbb{R}$ is a smooth function. They show that a type of parametric bootstrap procedure consistently approximates the distributions of such statistics when p/n converges to a positive limit. Lastly, (Naumov et al., 2019) study statistics related to the eigenvectors of $\widehat{\Sigma}$.

Notation. For a random variable X and an integer $q \in \{1, 2\}$, define the ψ_q-Orlicz norm as $\|X\|_{\psi_q} = \inf\{t > 0 \mid \mathbb{E}[\exp(|X|^q/t^q)] \leq 2\}$. The random variable X is said to be sub-exponential if $\|X\|_{\psi_1}$ is finite, and sub-Gaussian if $\|X\|_{\psi_2}$ is finite. In addition, for any $q \geq 1$, the L_q-norm of X is defined as $\|X\|_q = (\mathbb{E}[|X|^q])^{1/q}$. For any vectors $u, v \in \mathbb{R}^p$, their inner product is $\langle u, v\rangle = \sum_{j=1}^{p} u_j v_j$. For any real numbers a and b, the expression $a \ll b$ is used in an informal sense to mean that b is much larger than a. Also, we use the notation $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. If $\{a_n\}$ and $\{b_n\}$ are two sequences on non-negative numbers, then the relation $a_n \lesssim b_n$ means that there is a positive constant c not depending on n such that $a_n \leq c\, b_n$ holds for all large n. When both of the conditions $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold, we write $a_n \asymp b_n$.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

2. Main results

We consider a sequence of models indexed by n, in which all parameters may depend on n, except when stated otherwise. In particular, the dimension $p = p(n)$ is allowed to have an arbitrary dependence on n. Likewise, if a parameter does not depend on n, then it is understood not to depend on p either. One of the few parameters that will be treated as fixed with respect to n is the positive integer $k < p$.

Assumption 1 (Data-generating model)**.**

(a). There is a non-zero positive-semidefinite matrix $\Sigma \in \mathbb{R}^{p\times p}$*, such that the ith observation is generated as* $X_i = \Sigma^{1/2} Z_i$ *for all* $i = 1, \ldots, n$*, where* $Z_1, \ldots, Z_n \in \mathbb{R}^p$ *are i.i.d. random vectors with* $\mathbb{E}[Z_1] = 0$*, and* $\mathbb{E}[Z_1 Z_1^\top] = I_p$*.*

(b). The eigenvalues of Σ *satisfy* $\min_{1\le j\le k} \big(\lambda_j(\Sigma) - \lambda_{j+1}(\Sigma)\big) \gtrsim \lambda_1(\Sigma)$*.*

(c). Let $u_j \in \mathbb{R}^p$ *denote the* j*th eigenvector of* Σ*, and let* $\Gamma \in \mathbb{R}^{k\times k}$ *have entries given by* $\Gamma_{jj'} = \mathbb{E}[(\langle u_j, Z_1\rangle^2 - 1)(\langle u_{j'}, Z_1\rangle^2 - 1)]$ *for all* $1 \le j, j' \le k$*. Then, the matrix* Γ *satisfies* $\lambda_k(\Gamma) \gtrsim 1$*.*

In connection with the model described by Assumption 1, our results will make reference to a moment parameter defined as $\beta_q = \max_{1\le j\le p} \|\langle u_j, Z_1\rangle^2\|_q$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

for any $q \geq 1$.

Remarks. Regarding Assumption 1.(b), it ensures that there is some degree of separation between the leading eigenvalues of Σ. In less compact notation, the assumption states that there is a fixed constant $c > 0$ such that the inequality $\lambda_j(\Sigma) - \lambda_{j+1}(\Sigma) \geq c\lambda_1(\Sigma)$ holds for all $j = 1, \ldots, k$ and all large n. (There is no restriction on the size of c.) In general, a separation condition on the leading eigenvalues is unavoidable, because it is known both theoretically and empirically that the bootstrap can fail to approximate the distribution of $\sqrt{n}\big(\boldsymbol{\lambda}_k(\widehat{\Sigma}) - \boldsymbol{\lambda}_k(\Sigma)\big)$ if the leading population eigenvalues are not distinct (Beran and Srivastava, 1987; Hall et al., 2009). In more technical terms, the source of this issue can be explained briefly as follows: If $\mathcal{S}^{p\times p}$ denotes the space of real symmetric $p \times p$ matrices, and if $\lambda_j(\cdot)$ is viewed as a functional from $\mathcal{S}^{p\times p}$ to $\mathbb{R}$, then $\lambda_j(\cdot)$ can be non-differentiable at Σ when $\lambda_j(\Sigma)$ is a repeated eigenvalue (i.e., with multiplicity greater than one). In turn, this lack of smoothness makes it difficult for the bootstrap to approximate the distribution of $\sqrt{n}(\lambda_j(\widehat{\Sigma}) - \lambda_j(\Sigma))$.

To interpret Assumption 1.(c), the matrix Γ serves a technical role as a surrogate for the correlation matrix of $\sqrt{n}\big(\boldsymbol{\lambda}_k(\widehat{\Sigma}) - \boldsymbol{\lambda}_k(\Sigma)\big)$. Hence, the lower bound $\lambda_k(\Gamma) \gtrsim 1$ can be viewed as a type of non-degeneracy condition

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

for the distribution of interest. The proposition below gives examples of well-established models in which Assumption 1.(c) holds. Namely, parts (i) and (ii) below correspond to *Marčenko-Pastur models* and *elliptical models*, respectively. The latter case also illustrates that the entries of the vector Z_1 are not required to be independent.

Proposition 1. *(i) (Marčenko-Pastur case). Suppose that Assumption 1.(a) holds. In addition, suppose that the entries of Z_1 are independent, and there is a constant $\kappa > 1$ not depending on n such that $\min_{1\leq j\leq p}\mathbb{E}[Z_{1j}^4] \geq \kappa$. Then, Assumption 1.(c) holds.*

(ii) (Elliptical case). Let V be a random vector that is uniformly distributed on the unit sphere of $\mathbb{R}^p$, and let ξ be a non-negative scalar random variable independent of V that satisfies $\mathbb{E}[\xi^2] = p$ and $\mathbb{E}[\xi^4] < \infty$. Under these conditions, if Z_1 has the same distribution as ξV, then Assumption 1.(c) holds.

The proof of Proposition 1 is given in Section S1 of the Supplementary Material.

Bootstrap approximation. The following theorem is the central result of the paper, and quantifies the accuracy of the bootstrap when it is used to approximate the distribution of $\sqrt{n}(\boldsymbol{\lambda}_k(\widehat{\Sigma}) - \boldsymbol{\lambda}_k(\Sigma))$.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

Theorem 1. *Suppose that Assumption 1 holds and let* $q = 5\log(kn)$. *Then, there is a constant* $c > 0$ *not depending on* n *such that the event*

$$\sup_{t\in\mathbb{R}^k}\left|\mathbb{P}\Big(\sqrt{n}(\boldsymbol{\lambda}_k(\widehat{\Sigma})-\boldsymbol{\lambda}_k(\Sigma)) \preceq t\Big)-\mathbb{P}\Big(\sqrt{n}(\boldsymbol{\lambda}_k(\widehat{\Sigma}^\star)-\boldsymbol{\lambda}_k(\widehat{\Sigma})) \preceq t \,\Big|\, X\Big)\right| \leq \frac{c\,\log(n)\,\beta_{3q}^3\,\mathtt{r}(\Sigma)}{\sqrt{n}} \tag{2.3}$$

holds with probability at least $1 - c/n$.

Remarks. The proof of Theorem 1 is given in Section S4 of the Supplementary Material. It is possible to provide a more concrete understanding of the bound (2.3) by looking at how the factors $\mathtt{r}(\Sigma)$ and β_{3q} behave in some well-known situations. For instance, consider the class of matrices Σ whose eigenvalues have a polynomial decay profile of the form $\lambda_j(\Sigma) \asymp j^{-\gamma}$, for some fixed constant $\gamma > 0$. This class offers a convenient point of reference, because it interpolates between models that have low-dimensional structure and those that do not. Specifically, the effective rank can be related to γ as

$$\mathtt{r}(\Sigma) \asymp \begin{cases} 1 & \text{if } \gamma > 1 \\ \log(p) & \text{if } \gamma = 1 \\ p^{1-\gamma} & \text{if } \gamma < 1. \end{cases}$$

With regard to the parameter β_{3q}, its dependence on q is simple to describe in some commonly considered cases. If the entries of Z_1 are i.i.d. and sub-Gaussian, then β_{3q} grows at most linearly in q, with $\beta_{3q} \lesssim q\|Z_{11}\|_{\psi_2}^2$. Alternatively, if the entries of Z_1 are i.i.d. and sub-exponential, then β_{3q}

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

grows at most quadratically in q, with $\beta_{3q} \lesssim q^2 \|Z_{11}\|_{\psi_1}^2$. (See Chapter 2 of Vershynin (2018) for further details.) Hence, a direct consequence of Theorem 1 in such cases is that bootstrap consistency holds when $\gamma > 1/2$, $p \asymp n$ and $\|Z_{11}\|_{\psi_1} \lesssim 1$. Likewise, when $\gamma > 1$, the bound in Theorem 1 nearly achieves the *parametric rate* $n^{-1/2}$ and is not influenced by the size of p at all. This conclusion also conforms with the numerical results presented in Section 3.

From a more practical standpoint, it is possible to gauge the size of $\mathtt{r}(\Sigma)$ in an empirical way, by either estimating $\mathtt{r}(\Sigma)$ directly, or estimating upper bounds on it. Some examples of upper bounds on $\mathtt{r}(\Sigma)$ for which straightforward estimation methods are known to be effective in high dimensions include $\mathrm{tr}(\Sigma)/\max_{1\le j\le p}\Sigma_{jj}$ and $\mathrm{tr}(\Sigma)^2/\|\Sigma\|_F^2$. (Although guarantees can be established for direct estimates of $\mathtt{r}(\Sigma)$ in high-dimensions, such results can involve a more complex set of considerations than the upper bounds just mentioned.)

Transformations. To briefly review the idea of transformations, they are often used to solve inference problems involving a parameter θ and an estimator $\widehat{\theta}$ for which the distribution of $(\widehat{\theta} - \theta)$ is difficult to approximate. In certain situations, this difficulty can be alleviated if there is a monotone

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

function h for which the distribution of $(h(\widehat{\theta}) - h(\theta))$ is easier to approximate. In turn, this allows for more accurate inference on the "transformed parameter" $h(\theta)$, and then the results can be inverted to do inference on θ. In light of this, our next result shows that the rates of bootstrap approximation established in Theorem 1 remain essentially unchanged when using the class of fractional power transformations from $[0, \infty)$ to $[0, \infty)$. This class will be denoted by $\mathcal{H}$, so that if $h \in \mathcal{H}$, then $h(x) = x^a$, for some $a \in (0, 1]$.

Beyond the class of transformations just mentioned, the bootstrap can be combined with another type of transformation, known as *partial standardization* (Lopes et al., 2020). Letting $h \in \mathcal{H}$ be a given function, and letting $\varsigma_j^2 = \text{var}\left(h(\lambda_j(\widehat{\Sigma}))\right)$ for each $j = 1, \ldots, p$, this technique is well suited to bootstrapping "max statistics" of the form

$$M = \max_{1 \leq j \leq k} \frac{h(\lambda_j(\widehat{\Sigma})) - h(\lambda_j(\Sigma))}{\varsigma_j^\tau}, \tag{2.4}$$

where $\tau \in [0, 1]$ is a parameter that can be viewed as a degree of standardization. The ability to approximate the distribution of M is relevant to the construction of simultaneous confidence intervals for $\lambda_1(\Sigma), \ldots, \lambda_k(\Sigma)$. It also turns out that the choice of τ encodes a trade-off between the coverage accuracy and the width of such intervals, and that choosing an intermediate

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

value $\tau \in (0,1)$ can offer benefits in relation to $\tau = 0$ and $\tau = 1$. This is discussed in greater detail later in Section 3.

In order to state our extension of Theorem 1 in a way that handles both partial standardization and transformations $h \in \mathcal{H}$ in a unified way, we need to introduce a bit more notation. First, when considering the bootstrap counterpart of a partially standardized statistic such as (2.4), the vector $\boldsymbol{\varsigma}_k^\tau = (\varsigma_1^\tau, \ldots, \varsigma_k^\tau)$ is replaced with the estimate $\widehat{\boldsymbol{\varsigma}}_k^\tau = (\widehat{\varsigma}_1^\tau, \ldots, \widehat{\varsigma}_k^\tau)$, whose entries are defined by $\widehat{\varsigma}_j^2 = \operatorname{var}\big(h(\lambda_j(\widehat{\Sigma}^\star))\big|X\big)$ for all $j = 1, \ldots, p$. Second, the expression v/u involving vectors v and u denotes the vector obtained by entrywise division, $(v/u)_j = v_j/u_j$. (To handle the possibility zero denominators, events of the form $\{V/\widehat{\boldsymbol{\varsigma}}_k^\tau \preceq t\}$ are understood as $\{V \preceq t \odot \widehat{\boldsymbol{\varsigma}}_k^\tau\}$, where $V \in \mathbb{R}^k$ is random, $t \in \mathbb{R}^k$ is fixed, and $\odot$ is entrywise multiplication. Lemma S5.5 in the Supplementary Material also shows that such cases occur with negligible probability.) Lastly, recall that we write $\boldsymbol{h}(v) = (h(v_1), \ldots, h(v_k))$ for a k-dimensional vector v and transformation h.

Theorem 2. *Suppose that Assumption 1 holds. Fix a transformation $h \in \mathcal{H}$ and a constant $\tau \in [0,1]$ with respect to n, and let $q = 5\log(kn)$. Then, there is a constant $c > 0$ not depending on n, such that the event*

$$\sup_{t \in \mathbb{R}^k} \left| \mathbb{P}\left(\frac{\boldsymbol{h}(\boldsymbol{\lambda}_k(\widehat{\Sigma})) - \boldsymbol{h}(\boldsymbol{\lambda}_k(\Sigma))}{\boldsymbol{\varsigma}_k^\tau} \preceq t \right) - \mathbb{P}\left(\frac{\boldsymbol{h}(\boldsymbol{\lambda}_k(\widehat{\Sigma}^\star)) - \boldsymbol{h}(\boldsymbol{\lambda}_k(\widehat{\Sigma}))}{\widehat{\boldsymbol{\varsigma}}_k^\tau} \preceq t \,\middle|\, X \right) \right| \;\leq\; \frac{c\,\log(n)\,\beta_{3q}^5\,\mathfrak{r}(\Sigma)}{\sqrt{n}}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

holds with probability at least $1 - c/n$.

Remarks. The proof of Theorem 2 is given in Section S5 of the Supplementary Material. To comment on the technical relationship between Theorems 1 and 2, it is important to call attention to the differences between asymptotic and non-asymptotic analysis. When using asymptotics, the process of showing that bootstrap consistency for $\sqrt{n}\big(\boldsymbol{\lambda}_k(\widehat{\Sigma}) - \boldsymbol{\lambda}_k(\Sigma)\big)$ implies the same for $(\boldsymbol{h}(\boldsymbol{\lambda}_k(\widehat{\Sigma})) - \boldsymbol{h}(\boldsymbol{\lambda}_k(\Sigma)))/\boldsymbol{\varsigma}_k^{\tau}$ can typically be handled with a brief argument, based on the delta method and the consistency of the estimate $\widehat{\boldsymbol{\varsigma}}_k^{\tau}$. However, when taking a non-asymptotic approach, this process is much more involved.

3. Numerical Experiments

In this section, we focus on the application of constructing simultaneous confidence intervals for $\lambda_1(\Sigma), \ldots, \lambda_k(\Sigma)$. We do so in a variety of settings, corresponding to different values of n and p, as well as different values of effective rank, and different choices of transformations. In a nutshell, there are two overarching conclusions to take away from the experiments: **(1)** In situations where $n \ll p$ and $\mathtt{r}(\Sigma) \asymp 1$, the bootstrap generally produces intervals with accurate coverage, which provides a confirmation

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

of our theoretical results. **(2)** The classical log transformation mostly works well in low dimensions, but it can lead to coverage that is substantially below the nominal level when $\mathtt{r}(\Sigma)$ is moderately large. Nevertheless, we show that it is possible to find transformations that offer more reliable coverage in this challenging case. More generally, this indicates that alternative transformations are worth exploring in high-dimensional settings.

3.1 Simulation settings

The eigenvalues of the population covariance matrix Σ were chosen to have two different decay profiles:

(a) A polynomial decay profile $\lambda_j(\Sigma) = j^{-\gamma}$ for all $j = 1, \ldots, p$, with $\gamma \in \{0.7, 1.0, 1.3\}$.

(b) An exponential decay profile $\lambda_j(\Sigma) = \delta^j$ for all $j = 1, \ldots, p$, with $\delta \in \{0.7, 0.8, 0.9\}$.

As a clarification, it is important to note that the effective rank of Σ increases for larger values of δ, but decreases for larger values of γ. For the purposes of simulations, the choices (a) and (b) have the valuable property that the eigenvalues are parameterized in the same way for every choice of p, which facilitates the comparison of results across different dimensions.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

The matrix of eigenvectors for Σ was drawn uniformly from the set of $p \times p$ orthogonal matrices. The dimension p was taken from $\{10, 50, 100, 200\}$, and the sample size n ranged from 50 to 500. For each triple (n, p, γ) or (n, p, δ), the data $X_1, \ldots, X_n$ were generated in an i.i.d. manner with the following choices for the distribution of X_1:

(i) The vector $X_1 = \Sigma^{1/2}\xi V$ was generated with V being uniformly distributed on the unit sphere of $\mathbb{R}^p$, and ξ^2 being an exponential random variable independent of V with $\mathbb{E}[\xi^2] = p$.

(ii) The vector X_1 was generated from the Gaussian distribution $N(0, \Sigma)$.

For each parameter setting, we generated 1000 realizations of the dataset $X_1, \ldots, X_n$, and for each such realization, we generated $B := 1000$ sets of bootstrap samples of size n. When constructing simultaneous confidence intervals for $\lambda_1(\Sigma), \ldots, \lambda_k(\Sigma)$, the value of k was set to 5.

3.2 Bootstrap confidence intervals

For any $\alpha \in (0, 1)$, we aim to construct approximate versions of ideal random intervals $\mathcal{I}_1, \ldots, \mathcal{I}_k$ that satisfy

$$\mathbb{P}\left(\bigcap_{j=1}^{k} \{\lambda_j(\Sigma) \in \mathcal{I}_j\} \right) \geq 1 - \alpha. \tag{3.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

To this end, consider the following max and min statistics, based on any choice of partial standardization parameter $\tau \in [0, 1]$ and transformation h,

$$M = \max_{1 \le j \le k} \frac{h(\lambda_j(\widehat{\Sigma})) - h(\lambda_j(\Sigma))}{\varsigma_j^\tau}$$
$$L = \min_{1 \le j \le k} \frac{h(\lambda_j(\widehat{\Sigma})) - h(\lambda_j(\Sigma))}{\varsigma_j^\tau}.$$

Letting $q_M(\alpha)$ and $q_L(\alpha)$ denote the respective α-quantiles of M and L for any $\alpha \in (0, 1)$, it follows that the desired condition (3.5) holds if each interval $\mathcal{I}_j$ is defined as

$$\mathcal{I}_j = h^{-1}\left(\left[h(\lambda_j(\widehat{\Sigma})) - \varsigma_j^\tau q_M(1 - \tfrac{\alpha}{2}) \ , \ h(\lambda_j(\widehat{\Sigma})) - \varsigma_j^\tau q_L(\tfrac{\alpha}{2})\right]\right), \qquad (3.6)$$

with $h^{-1}([a, b])$ being understood as the preimage of $[a, b]$ under h.

To construct bootstrap intervals $\widehat{\mathcal{I}}_1, \ldots, \widehat{\mathcal{I}}_k$ based on (3.6), we need only replace $q_M(1 - \frac{\alpha}{2})$, $q_L(\frac{\alpha}{2})$, and $\varsigma_1, \ldots, \varsigma_k$ with estimates. In detail, we first estimate ς_j using the sample standard deviation of B bootstrap replicates of the form $h(\lambda_j(\widehat{\Sigma}^\star))$, denoted $\tilde{\varsigma}_j$. Next, we use the empirical $1 - \frac{\alpha}{2}$ quantile of B bootstrap replicates of the form $M^\star = \max_{1 \le j \le k}[h(\lambda_j(\widehat{\Sigma}^\star)) - h(\lambda_j(\widehat{\Sigma}))]/\tilde{\varsigma}_j^\tau$ as an estimate of $q_M(1 - \frac{\alpha}{2})$, and similarly for $q_L(\frac{\alpha}{2})$.

Regarding the use of transformations, the following three options were included in the experiments:

- *log transformation*: $h(x) = \log(x)$, with $\tau = 0$.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

- *standardization*: $h(x) = x$, with $\tau = 1$.
- *square-root transformation*: $h(x) = x^{1/2}$, with $\tau \in [0, 1]$ chosen data-adaptively.

In the case of the log transformation, the choice of $\tau = 0$ corresponds to the way that this transformation has been used in the classical literature (Beran and Srivastava, 1985), while in the case of standardization, the choice of $\tau = 1$ is definitional. For the square-root transformation, the use of a data-adaptive selection rule for $\tau \in [0, 1]$ is more nuanced, and can be informally explained in terms of the following ideas developed previously in (Lopes et al., 2020) and (Lin et al., 2021).

In essence, this choice can be understood in terms of a trade-off between two competing effects that occur in the extreme cases of $\tau = 1$ and $\tau = 0$. When using $\tau = 1$, the random variables $[\lambda_j(\widehat{\Sigma})^{1/2} - \lambda_j(\Sigma)^{1/2}]/\varsigma_j$ with $\varsigma_j^2 = \mathrm{var}(\lambda_j(\widehat{\Sigma})^{1/2})$ and $j = 1, \ldots, k$ are on approximately "equal footing", which makes the behavior of the statistic M sensitive to their joint distribution (and likewise for L). By contrast, when $\tau = 0$ is used, the variables $[\lambda_j(\widehat{\Sigma})^{1/2} - \lambda_j(\Sigma)^{1/2}]$ will tend to be on different scales, and the variable on the largest scale, say j', will be the maximizer for M relatively often. In this situation, the statistic M is governed more strongly by the marginal

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

distribution of $[\lambda_{j'}(\widehat{\Sigma})^{1/2} - \lambda_{j'}(\Sigma)^{1/2}]$. So, from this heuristic point of view, the choice of $\tau = 0$ can simplify the behavior of M relative to the case of $\tau = 1$, making the distribution of M easier to approximate. However, the choice of $\tau = 0$ also has the drawback that it can lead to simultaneous confidence intervals that are excessively wide, because the widths are no longer adapted to the different values $\varsigma_1, \ldots, \varsigma_k$ (since $\varsigma_1^0 = \cdots = \varsigma_k^0 = 1$).

To strike a balance between these competing effects, we used the following simple rule to select τ in the case of the square-root transformation. For a candidate value of τ, let $\widehat{\mathcal{I}}_1(\tau), \ldots, \widehat{\mathcal{I}}_k(\tau)$ denote the associated bootstrap intervals defined beneath equation (3.6) (so that the dependence on τ is explicit), and let $|\widehat{\mathcal{I}}_1(\tau)|, \ldots, |\widehat{\mathcal{I}}_k(\tau)|$ denote their widths. Also define $\widehat{\mu}(\tau) = \frac{1}{k}\sum_{j=1}^{k} |\widehat{\mathcal{I}}_j(\tau)|$ and $\widehat{\sigma}(\tau)^2 = \frac{1}{k}\sum_{i=1}^{k}(|\widehat{\mathcal{I}}_j(\tau)| - \widehat{\mu}(\tau))^2$. In this notation, we selected the value of τ that minimized $\widehat{\mu}(\tau) + \widehat{\sigma}(\tau)$ over the set of candidates $\{0.0, 0.1, \ldots, 0.9, 1.0\}$. Different variants of this type of criterion minimization rule have also been observed to be effective in other contexts (Lopes et al., 2020; Lin et al., 2021).

3.3 Discussion of coverage

Figure 1 contains nine panels displaying the results for the simultaneous coverage probability $\mathbb{P}(\cap_{j=1}^{5}\{\lambda_j(\Sigma) \in \widehat{\mathcal{I}}_j\})$, based on a nominal value of 95%

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

(i.e., $\alpha = 0.05$), for simulation model (i) with a polynomial decay profile for the population eigenvalues. The figure summarizes a large amount of information, because it shows how the coverage depends on n, p, the eigenvalue decay parameter γ, and the three transformations described above. For each panel, the x-axis measures n, and the y-axis measures $\mathbb{P}(\cap_{j=1}^{5}\{\lambda_j(\Sigma) \in \widehat{\mathcal{I}}_j\})$. The results corresponding to the dimensions $p = 10, 50, 100, 200$ are plotted with curves that are labeled in the legend. The three rows of panels from top to bottom correspond to the log transformation, ordinary standardization, and the square-root transformation. The three columns of panels from left to right correspond to the eigenvalue decay parameters $\gamma = 0.7, 1.0, 1.3$. In addition, Figure 2 displays analogous results for exponentially decaying population eigenvalues in model (i). Lastly, results for model (ii), as well as for a nominal value of 90% (instead of 95%), are provided in Section S8 of the Supplementary Material.

There are several notable patterns in Figures 1 to discuss. The first is that faster rates of decay tend to lead to better coverage accuracy—as anticipated by our theoretical results. In particular, when the eigenvalue decay parameter is set to $\gamma = 1.3$, the coverage is rather accurate, even when $n \ll p$. Furthermore, the accuracy is essentially unaffected by the dimen-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

sion p in this situation, as indicated by the overlap of the four curves. On the other hand, as the decay parameter becomes smaller, the three transformations perform in different ways. For instance, when $\gamma = 0.7$, $p = 200$, and $n < 200$, the log transformation yields coverage that clearly falls short of the nominal level. By contrast, the standardization and square-root transformations tend to err more safely in the conservative direction when $\gamma = 0.7$. To give some indication of the difficulty of $\gamma = 0.7$, note that if γ were decreased slightly to 0.5 with $p \gtrsim n$, this would imply $\mathtt{r}(\Sigma)/\sqrt{n} \asymp \sqrt{p/n} \gtrsim 1$, in which case bootstrap consistency would not be guaranteed. When considering all three cases $\gamma = 0.7, 1.0, 1.3$ collectively, the square-root transformation seems to yield the best overall coverage results if conservative errors are viewed as preferable to anti-conservative ones.

Turning to the coverage results for exponential spectrum decay, the log and square-root transformations continue to follow the pattern that faster decay improves coverage accuracy. (Recall that smaller δ corresponds to faster decay.) Also, the log transformation maintains its tendency to err in the anti-conservative direction, while the square-root transformation still tends to err in the conservative direction. Lastly, ordinary standardization yields larger errors in the anti-conservative direction than it did in the

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

previous context.

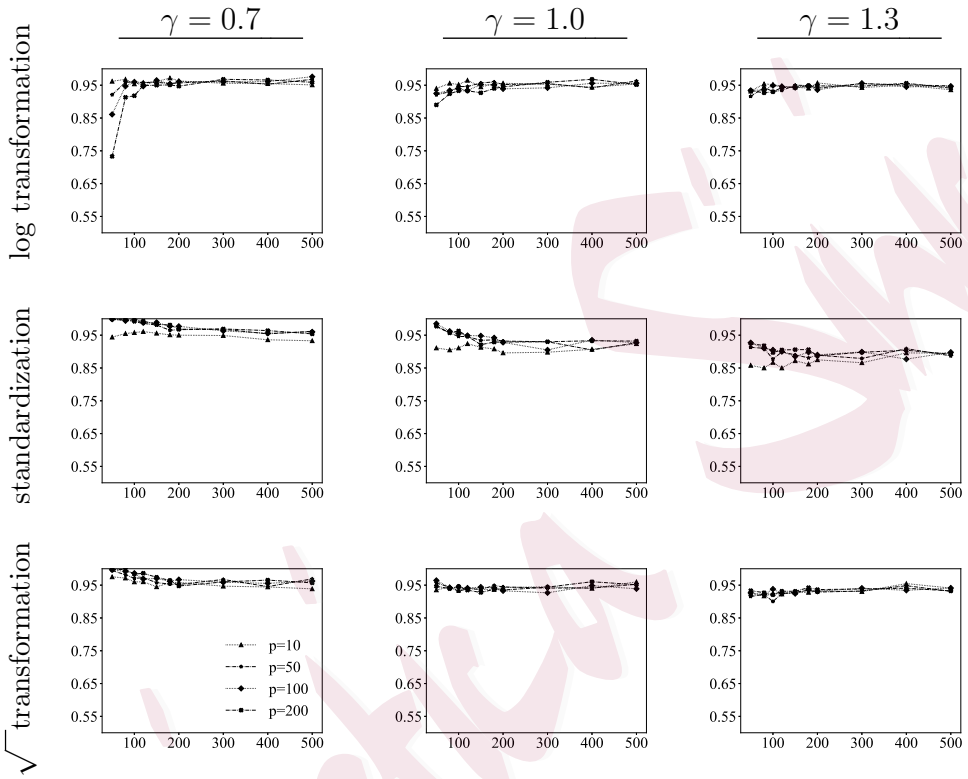

Figure 1: (Simultaneous coverage probability versus n in simulation model (i) with a polynomial decay profile). In each panel, the y-axis measures $\mathbb{P}(\cap_{j=1}^{5}\{\lambda_j(\Sigma) \in \widehat{\mathcal{I}}_j\})$ based on a nominal value of 95%, and the x-axis measures n. The curves correspond to different values of p.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

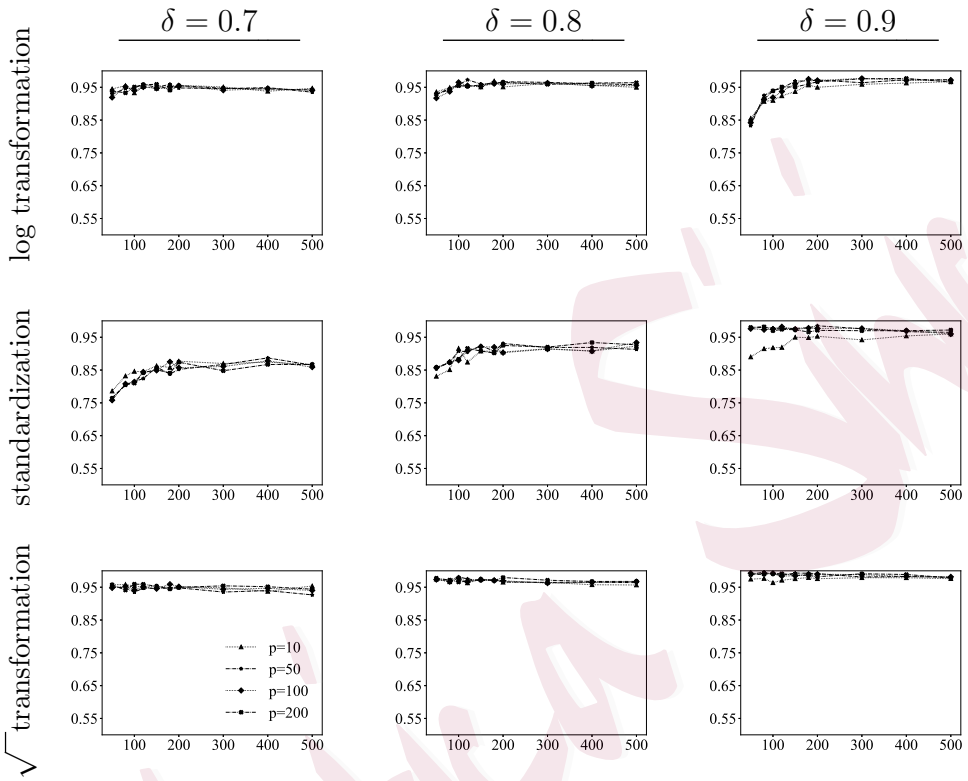

Figure 2: (Simultaneous coverage probability versus n in simulation model (i) with an exponential decay profile). The plotting scheme is the same as that described in the caption of Figure 1, except that the three columns correspond to values of the eigenvalue decay parameter δ.

3.4 Discussion of width

Beyond coverage probability, interval width is another important factor to consider when appraising confidence intervals. In Figures 3-4, the average

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

width $\mathbb{E}[|\widehat{\mathcal{I}}_1| + \cdots + |\widehat{\mathcal{I}}_k|]/k$ is plotted on the y-axis as a function of the sample size n on the x-axis, with the underlying parameter settings being organized in the same manner as in Figures 1-2. (Corresponding results for settings based on model (ii) and a nominal value of 90% are presented in Section S8 of the Supplementary Material.) With regard to the three transformations, they produce intervals that have roughly similar widths across most parameter settings. However, at a more fine-grained level, the results in the case of polynomial spectrum decay show that the log transformation tends to yield slightly shorter widths than the square-root transformation, which in turn, tends to yield slightly shorter widths than ordinary standardization. In the case of exponential spectrum decay with $\delta = 0.9$, the same pattern is also apparent, while for smaller values of δ, there is not much difference among the transformations.

Aside from the transformations, there are two other general trends to notice. Within each of the 18 panels of Figures 3-4, there is a monotone relationship between width and the dimension p, with the width increasing as the dimension increases. Similarly, the width generally also increases as the effective rank $\mathtt{r}(\Sigma)$ increases.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

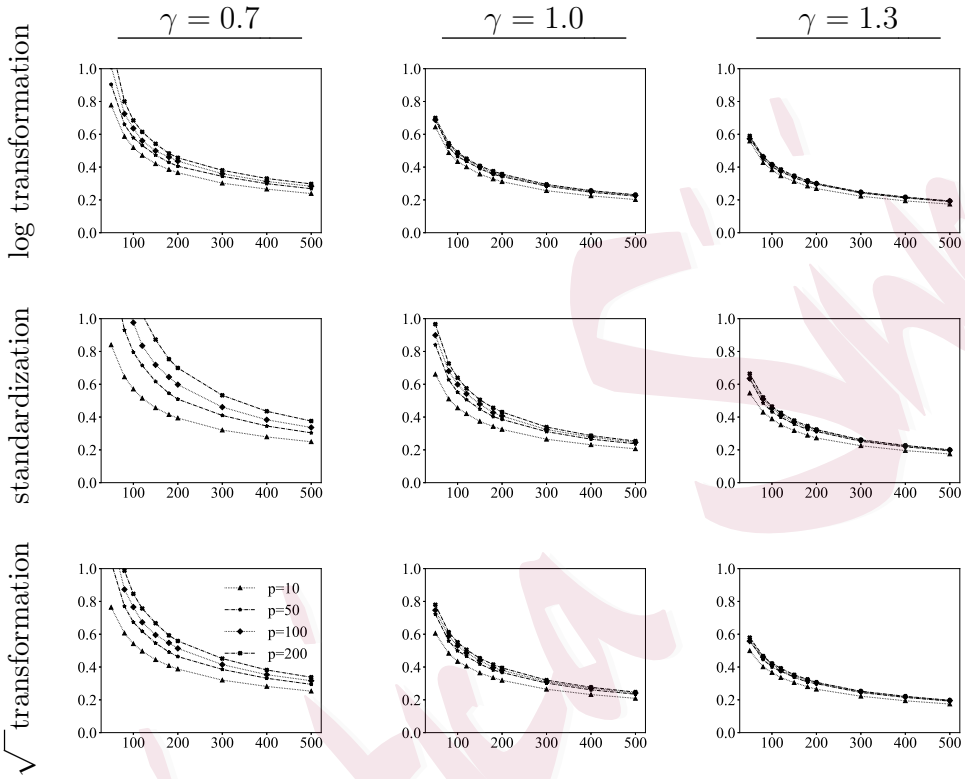

Figure 3: (Average width versus n in simulation model (i) with a polynomial decay profile). In each of the nine panels, the y-axis measures the average width $\mathbb{E}[|\widehat{\mathcal{I}}_1| + \cdots + |\widehat{\mathcal{I}}_5|]/5$, and the x-axis measures n. The curves correspond to $p = 10$, 50, 100, 200. The three rows and three columns correspond to the choices of transformations and the values of the eigenvalue decay parameter γ, respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

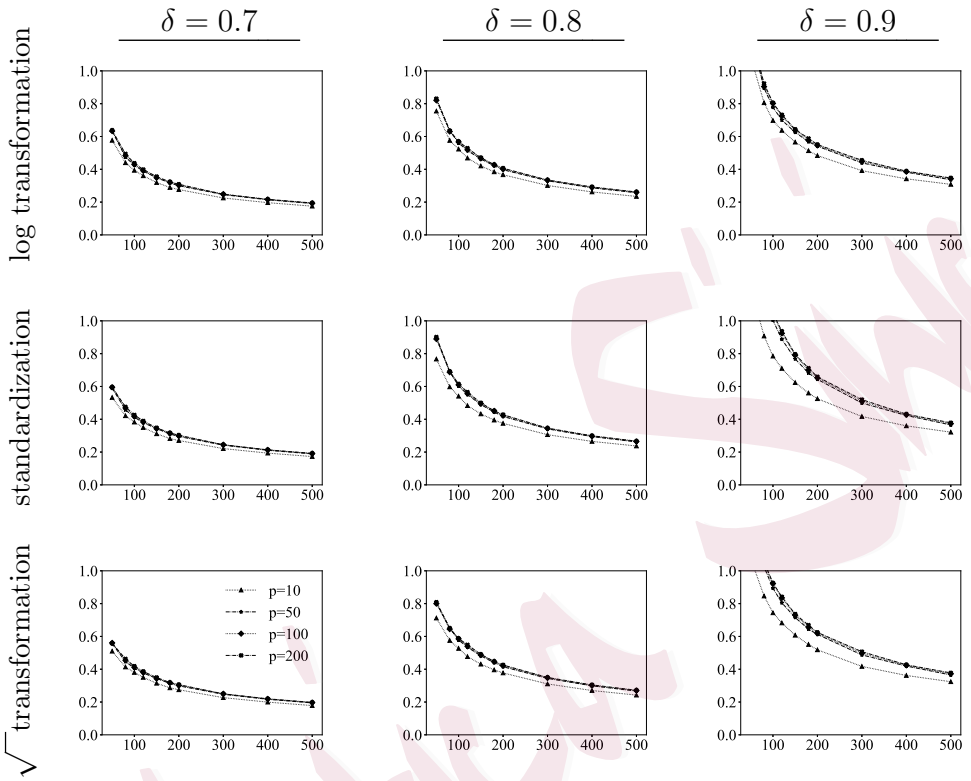

Figure 4: (Average width versus n in simulation model (i) with an exponential decay profile). The plotting scheme is the same as that described in the caption of Figure 3, except that the three columns correspond to values of the eigenvalue decay parameter δ.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

Supplementary Material

The Supplementary Material contains the proofs of all theoretical results, additional simulation results, and real-data examples based on stock market returns.

Acknowledgements

The authors gratefully acknowledge partial support from NSF grant DMS-1915786.

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley.

Bai, Z. and Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Springer.

Beran, R. and Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *Annals of Statistics*, 13(1):95–115.

Beran, R. and Srivastava, M. S. (1987). Correction: Bootstrap tests and confidence regions for functions of a covariance matrix. *Annals of Statistics*, 15(1):470–471.

Bunea, F. and Xiao, L. (2015). On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA. *Bernoulli*, 21(2):1200–1230.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

REFERENCES

Chen, S., ODea, E. B., Drake, J. M., and Epureanu, B. I. (2019). Eigenvalues of the covariance matrix as early warning signals for critical transitions in ecological systems. *Scientific Reports*, 9(1):1–14.

Chernick, M. R. (2011). *Bootstrap Methods: A Guide for Practitioners and Researchers*. Wiley.

Couillet, R. and Debbah, M. (2011). *Random Matrix Methods for Wireless Communications*. Cambridge.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge.

Diaconis, P. and Efron, B. (1983). Computer-intensive methods in statistics. *Scientific American*, 248(5):116–131.

DiCiccio, T. J. (1984). On parameter transformations and interval estimation. *Biometrika*, 71(3):477–485.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3):189–228.

Dobriban, E. and Wager, S. (2018). High-dimensional asymptotics of prediction: Ridge regression and classification. *Annals of Statistics*, 46(1):247–279.

Dümbgen, L. (1993). On nondifferentiable functions and the bootstrap. *Probability Theory and Related Fields*, 95(1):125–140.

Eaton, M. L. and Tyler, D. E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *Annals of Statistics*,

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

pages 260–271.

El Karoui, N. and Purdom, E. (2019). The non-parametric bootstrap and spectral analysis in moderate and high-dimension. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2115–2124.

Fabozzi, F. J., Kolm, P. N., Pachamanova, D. A., and Focardi, S. M. (2007). *Robust Portfolio Optimization and Management*. Wiley.

Fisher, A., Caffo, B., Schwartz, B., and Zipunnikov, V. (2016). Fast, exact bootstrap principal component analysis for $p > 1$ million. *Journal of the American Statistical Association*, 111(514):846–860.

Hall, P., Lee, Y. K., Park, B. U., and Paul, D. (2009). Tie-respecting bootstrap methods for estimating distributions of sets and functions of eigenvalues. *Bernoulli*, 15(2):380–401.

Han, F., Xu, S., and Zhou, W.-X. (2018). On Gaussian comparison inequality and its application to spectral analysis of large random matrices. *Bernoulli*, 24(3):1787–1833.

Hsu, D., Kakade, S. M., and Zhang, T. (2014). Random design analysis of ridge regression. *Foundations of Computational Mathematics*, 14(3):569–600.

Johnstone, I. M. and Paul, D. (2018). PCA in high dimensions: An orientation. *Proceedings of the IEEE*, 106(8):1277–1292.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer.

Jung, S., Lee, M. H., and Ahn, J. (2018). On the number of principal components in high

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

dimensions. *Biometrika*, 105(2):389–402.

Koltchinskii, V., Löffler, M., and Nickl, R. (2020). Efficient estimation of linear functionals of principal components. *Annals of Statistics*, 48(1):464–490.

Koltchinskii, V. and Lounici, K. (2017). Normal approximation and concentration of spectral projectors of sample covariance. *Annals of Statistics*, 45(1):121–157.

Konishi, S. (1991). Normalizing transformations and bootstrap confidence intervals. *Annals of Statistics*, 19(4):2209 – 2225.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40(2):1024–1060.

Li, H. and Ralph, P. (2019). Local PCA shows how the effect of population structure differs along the genome. *Genetics*, 211(1):289–304.

Lin, Z., Lopes, M. E., and Müller, H.-G. (2021). High-dimensional MANOVA via bootstrapping and its application to functional and sparse count data. *Journal of the American Statistical Association*, 0(0):1–15.

Lopes, M. E., Blandino, A., and Aue, A. (2019). Bootstrapping spectral statistics in high dimensions. *Biometrika*, 106(4):781–801.

Lopes, M. E., Erichson, N. B., and Mahoney, M. W. (2022+). Bootstrapping the operator norm in high dimensions: Error estimation for covariance matrices and sketching. *to appear: Bernoulli (arXiv:1909.06120)*.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

Lopes, M. E., Lin, Z., and Müller, H.-G. (2020). Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data. *Annals of Statistics*, 48(2):1214–1229.

Lounici, K. (2014). High-dimensional covariance matrix estimation with missing observations. *Bernoulli*, 20(3):1029–1058.

Naumov, A., Spokoiny, V., and Ulyanov, V. (2019). Bootstrap confidence sets for spectral projectors of sample covariance. *Probability Theory and Related Fields*, 174(3-4):1091–1132.

Nguyen, L. H. and Holmes, S. (2019). Ten quick tips for effective dimensionality reduction. *PLOS Computational Biology*, 15(6):e1006907.

Olive, D. J. (2017). *Robust Multivariate Analysis*. Springer.

Ruppert, D. and Matteson, D. S. (2015). *Statistics and Data Analysis for Financial Engineering*. Springer.

Stewart, T. A., Liang, C., Cotney, J. L., Noonan, J. P., Sanger, T. J., and Wagner, G. P. (2019). Evidence against tetrapod-wide digit identities and for a limited frame shift in bird wings. *Nature communications*, 10(1):1–13.

Terry, E. E., Zhang, X., Hoffmann, C., Hughes, L. D., Lewis, S. A., Li, J., Wallace, M. J., Riley, L. A., Douglas, C. M., and Gutierrez-Monreal, M. A. (2018). Transcriptional profiling reveals extraordinary diversity among skeletal muscle tissues. *eLife*, 7:e34613.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0158

REFERENCES

Tibshirani, R. (1988). Variance stabilization and the bootstrap. *Biometrika*, 75(3):433–444.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge.

Wagner, F. (2015). GO-PCA: An unsupervised method to explore gene expression data using prior knowledge. *PLOS One*, 10(11):e0143196.

Webb-Vargas, Y., Chen, S., Fisher, A., Mejia, A., Xu, Y., Crainiceanu, C., Caffo, B., and Lindquist, M. A. (2017). Big data and neuroimaging. *Statistics in Biosciences*, 9(2):543–558.

University of California, Davis

E-mail: jwyao@ucdavis.edu

University of California, Davis

E-mail: melopes@ucdavis.edu