APPENDIX

Supplementary material for "The non-parametric bootstrap and spectral analysis in moderate and high-dimension"

by Noureddine El Karoui and Elizabeth Purdom

Contents

S1 Review of existing results in random matrix theory

In what follows, we give a review of the existing technical literature regarding the distribution of eigenvalues in high dimensions. We divide this review into two parts: results concerning the bulk of the spectral distribution and those concerning the edge. Bulk results are concerned with characterizing the spectral distribution of $\widehat{\Sigma}$, i.e with the (random) probability measure with distribution

$$dF_p(x) = \frac{1}{p}\sum_{i=1}^{n} \delta_{\lambda_i(\widehat{\Sigma})}(x) \; .$$

Edge results are concerned with the fluctuation behavior of the eigenvalues that are at the edge of the spectrum of the matrices of interest, such as the distribution of the largest eigenvalue.

Notations We call $\lambda_1(M)$ the largest eigenvalue of a symmetric matrix M. We call $\lambda_1(M) \geq \lambda_2(M) \geq \lambda_3(M) \geq \lambda_p(M)$ the ordered eigenvalues of the $p \times p$ matrix M. If $Z \sim \mathcal{N}_{\mathbb{C}}(0, \Sigma)$, Z has a complex normal distribution, i.e $Z = \frac{1}{\sqrt{2}}(Z_1 + iZ_2)$ where Z_1 and Z_2 are independent with $Z_i \sim \mathcal{N}(0, \Sigma)$. We call $\mathbb{C}^+$ the set of complex numbers with positive imaginary part.

S1.1 Bulk results

Bulk results are concerned with the spectral distribution of $\widehat{\Sigma}$, $F_p(x)$, as defined above.

An efficient way to characterize the limiting behavior of F_p is through its Stieltjes transform:

$$\text{for } z = u + iv \text{ with } v > 0, \;\; m_p(z) = \frac{1}{p}\text{trace}\left((\widehat{\Sigma} - z\text{Id}_p)^{-1}\right) \; .$$

Note that $m_p(z) : \mathbb{C}^+ \mapsto \mathbb{C}^+$. We have of course

$$m_p(z) = \int \frac{dF_p(\lambda)}{\lambda - z} = \frac{1}{p}\sum_{i=1}^{p} \frac{1}{\lambda_i(\widehat{\Sigma}) - z} \; .$$

An important result in this area is the so-called Marchenko-Pastur equation [20, 27], which states the following

Theorem S1.1. *Suppose $X_i \overset{iid}{\backsim} \Sigma^{1/2} Z_i$, where Z_i has i.i.d entries, with mean 0, variance 1 and 4 moments. Suppose that the spectral distribution of Σ has a limit H in the sense of weak convergence of probability measures and $p/n \to r \in (0,1)$. Then*

$$F_p \Longrightarrow F a.s \; ,$$

where F is a deterministic probability distribution.

Call $v_p(z) = (1 - p/n)\frac{-1}{z} + \frac{p}{n} m_p(z)$. Then $v_p(z) \to v_F(z)$ a.s. The Stieltjes transform of F can be characterized through the equation

$$-\frac{1}{v_F(z)} = z - r\int \frac{\lambda dH_\infty(\lambda)}{1 + \lambda v_F(z)} \; , \forall z \in \mathbb{C}^+$$

At an intuitive level, this result means that the histogram of eigenvalues of $\widehat{\Sigma}$ is asymptotically non-random, and its limiting shape, which depends on the ratio p/n, is characterized by the Marchenko-Pastur distribution.

A generalization of this result to the case of elliptical predictors was obtained in [12]. For the purpose of the current paper, the main result of [12] states the following:

Theorem S1.2. *Suppose $X_i \overset{iid}{\backsim} \Sigma^{1/2} Z_i$, where Z_i has i.i.d entries, with mean 0, variance 1 and 4 moments. Suppose that the spectral distribution of Σ has a limit H, that H has one moment and $p/n \to r \in (0, \infty)$. Consider the matrix*

$$B_n = \frac{1}{n} \sum_{i=1}^{n} w_i X_i X_i' \ .$$

Assume that the weights $\{w_i\}_{i=1}^n$ are independent of X_i's. Call ν_n the empirical distribution of the weights w_i's and suppose that $\nu_n \Longrightarrow \nu$.

Then $B_n \Longrightarrow B$ a.s, where B is a deterministic probability distribution; furthermore the Stieltjes transform of B, m, satisfies the system

$$m(z) = \int \frac{dH(\tau)}{\tau \int \frac{w^2}{1+rw^2\gamma(z)} d\nu(w) - z} \quad \text{and}$$
$$\gamma(z) = \int \frac{\tau dH(\tau)}{\tau \int \frac{w^2}{1+rw^2\gamma(z)} d\nu(w) - z} \ .$$

where $\gamma(z)$ is the only solution of this equation mapping $\mathbb{C}^+$ into $\mathbb{C}^+$.

Theorem S1.2 is interesting statistically because it shows that the limiting spectral distribution of weighted covariance matrices is completely different from that of unweighted covariance matrices, even when $\mathbf{E}(w_i) = 1$. This is in very sharp contrast with the low-dimensional case. (Note that as shown in [11], Theorem S1.2 holds for many other distributions for X_i's that the one mentioned in our statement.)

In the context of the current paper, this result is especially useful since bootstrapping a covariance matrix amounts to moving from an unweighted to a weighted covariance matrix.

S1.1.1 Consequences of the previous results for the bootstrap

As recalled above, in the case of a Gaussian design the spectral distribution of $\widehat{\Sigma}$, $L_n(\widehat{\Sigma})$, has a non-random limit, $\mathcal{L}(\Sigma)$ satisfying the Marchenko-Pastur equation (see [20],[28], [27],[2]). The results above also say that when sampling both w_i's and X_i's, with $\{w_i\}_{i=1}^n$ independent of $\{X_i\}_{i=1}^n$, that $\widehat{\Sigma}_w$ has a non-random limiting distribution, $\mathcal{L}(\Sigma, w)$. $\mathcal{L}(\Sigma, w)$ can be implicitly characterized by a pair of equations for a variant of the Stieltjes transform of $\widehat{\Sigma}_w$ (see [8], [12]). As can be seen from above, this limit distribution is in general hard to characterize analytically, however it is not the same as that of $\widehat{\Sigma}$, i.e. $\mathcal{L}(\Sigma, w) \neq \mathcal{L}(\Sigma)$.

Furthermore, this discrepancy between the sample distribution and the bootstrap distribution also happens almost surely for the bootstrap distribution of eigenvalues generated from a specific realization of the design matrix. Specifically, let us call $L_n(\widehat{\Sigma}_w)|\{X_i\}_{i=1}^n$ the bootstrapped spectral distribution of $\widehat{\Sigma}$ for a specific design matrix X. Then it is clear by a simple conditioning argument, that if $L_n(\widehat{\Sigma}_w)$ is the spectral distribution of $\widehat{\Sigma}_w$,

$$L_n(\widehat{\Sigma}_w)|\{X_i\}_{i=1}^n \Longrightarrow \mathcal{L}(\Sigma, w) \ \ a.s$$

where the $a.s$ statement refers to the design matrix. The limit , $\mathcal{L}(\Sigma, w)$ is different from the limiting spectral distribution of $\widehat{\Sigma}$, $\mathcal{L}(\Sigma)$ when p, n tend to infinity and $p/n \to r \in (0, \infty)$. *Hence the bootstrapped distribution of the eigenvalues of $\widehat{\Sigma}$ is in general biased.*

One could argue that this is not disqualifying since the empirical spectral distribution is itself biased for the population spectral distribution. However, as the system above shows, the relationship between the spectral distributions of the bootstrapped sample covariance and sample covariance is different from that between the spectral distribution of the sample covariance and the population covariance. This helps explain why the bootstrap fails in our context.

S1.2 Edge results

The first result ([17]) concerns the distribution of the largest eigenvalue of $\widehat{\Sigma}$ in the case that might be considered the "null" case for PCA – the predictors are all independent with covariance matrix equal to Id_p.

Theorem S1.3. *Suppose $X_i \overset{iid}{\backsim} \mathcal{N}(0, \text{Id}_p)$. Assume that $p/n \to r \in (0,1)$. Then*

$$n^{2/3} \frac{\lambda_1(\widehat{\Sigma}) - \mu_{n,p}}{\sigma_{n,p}} \Longrightarrow TW_1 \; .$$

We have for instance $\mu_{n,p} = (1 + \sqrt{p/n})^2$ and $\sigma_{n,p} = (1 + \sqrt{p/n})(1 + \sqrt{n/p})^{1/3}$. The result for the case $r \in (1, \infty)$ follows immediately by changing the role of p and n; see [17] for details.

TW_1 refers to the Tracy-Widom distribution appearing in the study of the Gaussian Orthogonal ensemble (GOE); details about its density can be found in [17] for instance. Further details about $\mu_{n,p}$ and $\sigma_{n,p}$ are in the appendix; they both converge to finite, non-zero limit. This result implies, among other things, that the standard estimate $\lambda_1(\widehat{\Sigma})$ is a biased estimator of the true $\lambda_1(\Sigma)$ when p/n is not close to zero, overestimating the true size of $\lambda_1(\Sigma)$.

From the point of view of PCA, $\lambda_1(\Sigma) > 1$ corresponds to the scenario of the "alternative" hypothesis, and an important question is how well can we differentiate when the data came from the alternative distribution rather than the null. [4] shows that the distribution of $\lambda_1(\widehat{\Sigma})$ given in Theorem S1.3 also describes the distribution of $\lambda_1(\widehat{\Sigma})$ when Σ is a finite rank perturbations of the Id_p, provided none of the eigenvalues of Σ are too separated from each other. In practical terms, this signifies that $\lambda_1(\widehat{\Sigma})$ has - asymptotically - the exact same distribution under the null as under the alternative and therefore no ability to differentiate the null and the alternative, provided the alternative is not far away from the null.

Following a question posed by Johnstone, Baik, Ben-Arous and Péché obtained the following result [4], which gives the point at which the alternative hypothesis is sufficiently removed from the null so that the distribution of $\lambda_1(\widehat{\Sigma})$ is stochastically different from that of $\lambda_1(\widehat{\Sigma})$ under the null.

Theorem S1.4. *Suppose $X_i \overset{iid}{\backsim} \mathcal{N}_{\mathbb{C}}(0, \Sigma)$. Suppose that $\lambda_1(\Sigma) = 1 + \eta\sqrt{p/n}$ and $\lambda_i(\Sigma) = 1$ for $i > 1$.*

1. *If $0 < \eta < 1$, then*

$$n^{2/3} \frac{\lambda_1(\widehat{\Sigma}) - \mu_{n,p}}{\sigma_{n,p}} \Longrightarrow TW_2 \; .$$

2. *On the other hand, if $\eta > 1$, then*

$$\sqrt{n} \frac{\lambda_1(\widehat{\Sigma}) - \mu_{\eta,n,p}}{\sigma_{\eta,n,p}} \Longrightarrow \mathcal{N}(0,1) \; .$$

Here,

$$\mu_{\eta,n,p} = \lambda_1 (1 + \frac{\sqrt{p/n}}{\eta}) \text{ and } \sigma_{\eta,n,p} = \lambda_1 \sqrt{1 - \eta^{-2}} \; .$$

We note that we can rewrite the previous quantities solely as functions of λ_1, specifically

$$\mu_{\eta,n,p} = \lambda_1 \left(1 + \frac{p/n}{\lambda_1 - 1}\right) \text{ and } \sigma_{\eta,n,p} = \lambda_1 \sqrt{1 - \frac{n}{p}(\lambda_1 - 1)^{-2}} .$$

This representation shows that $\mu_{\eta,n,p}$ is an increasing function of λ_1 on $(1+\sqrt{p/n}, \infty)$ and therefore it would be easy to estimate $\lambda_1(\Sigma)$ from $\lambda_1(\widehat{\Sigma})$. In particular, it is very simple to build confidence intervals in this context.

S1.2.1 Gaussian Phase Transition

These results show that if the largest eigenvalue is changed from 1 to $\lambda_1(\Sigma) > 1 + \sqrt{p/n}$ and the other eigenvalues remain the same, then $\lambda_1(\widehat{\Sigma})$ has Gaussian fluctuations and they are of order $n^{-1/2}$. See also [23]. In other words, there is a phase-transition: if λ_1 is sufficiently large, i.e larger than $1 + \sqrt{p/n}$ the largest eigenvalue of $\widehat{\Sigma}$ has Gaussian fluctuations. If it is not large enough, i.e smaller than $1 + \sqrt{p/n}$, the fluctuations are Tracy-Widom, and in fact are the same if $\lambda_1(\Sigma) = 1$. The value $1 + \sqrt{p/n}$ is therefore called the *Gaussian phase transition*. Statistically, it is hard to build confidence intervals for λ_1 in the latter case - but it is very easy to do so in the first case where λ_1 is sufficiently large. Part of our simulation study investigates whether the bootstrap is capable of capturing this statistically interesting phase transition.

Similar results were obtained in [11] for general Σ in the complex Gaussian case and extended to the real case in [19]. [10] showed that Theorem S1.3 holds when $p/n \to 0$ and $p \to \infty$ at any rate. See also the interesting [23] and [16]. We finally note that the main result in [4] is slightly more general than Theorem S1.4 but we just need that version for the current paper.

S2 Description of Simulations and other Numerics

For each of 1,000 simulations, we generate a $n \times p$ data matrix X. For each X, we calculate either the top eigenvalue (or the gap statistic) from the sample covariance matrix. Specifically, we perform the SVD of X using the ARPACK numeric routines (implemented in the package `rARPACK` in R) to find the top five singular values of X and get estimates $\hat{\lambda}_i$ by multiplying the singular values of X by $1/n$.

For each simulation, we perform bootstrap resampling of the n rows of X to get a bootstrap resample X^{*b} and λ_i^{*b}; we repeat the bootstrap resampling $B = 999$ times for each simulation, resulting in 999 values of λ_i^{*b} for each simulation.

In all simulations, we only consider the case where only λ_1 is allowed to differ from the rest of the eigenvalues. Therefore for all eigen values except the first, $\lambda_i = 1$. For λ_1, we consider $\lambda_1 = 1 + c\sqrt{\frac{p}{n}}$ for the following c: 0, 0.9, 1.1, 1.5, 2.0, 3.0, 6.0, 11.0, 50.0, 100.0, and 1000.0 (not all of these values are shown in figures or tables accompanying this manuscript). The results shown in this manuscript set $n = 1,000$, though $n = 500$ was also simulated.

Generating X We generate X as $X = Z\Sigma$ where $\Sigma = V\Lambda V'$, and Λ is a diagonal matrix of eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$. We assume that there is no structure in the true eigenvectors, and generate V as the right eigenvectors of the SVD of a nxp matrix with entries i.i.d $N(0,1)$.

$Z = DZ_0$ is a nxp matrix, with Z_0 having entries i.i.d. $N(0,1)$ and D is a diagonal matrix D. If D is the identity matrix, Z will be i.i.d. normally distributed; otherwise Z will be i.i.d with an elliptical distribution. We simulated under the following distributions for the diagonal entries of D to create elliptical distribution for Z,

- $D_{ii} \sim N(0,1)$
- $D_{ii} \sim Unif(1/2, \frac{\sqrt{3}\sqrt{4-1/4}}{2} - \frac{1}{4})$
- $D_{ii} \sim Exp(\sqrt{2})$

In this manuscript, we concentrated only on $D_{ii} \sim Exp(\sqrt{2})$, the "Elliptical Exponential" distribution. This was because its behavior resulted in an elliptical distribution for Z with properties the most different from when Z is normal. The remaining choices for the distribution of D result in elliptical distributions between that of the Elliptical Exponential and the Normal. The results from when $D_{ii} \sim Unif$ were generally fairly similar to when Z is normal and results from when $D_{ii} \sim N(0,1)$ were more different, though not as extreme as the exponential weights.

S3 Proofs

S3.1 Proof of Lemma 3.1

Proof of Lemma 3.1. We recall the following result from a simple application of the Sherman-Morrison-Woodbury formula (see [15] and [2]): if M is a symmetric matrix, q is a real vector $v > 0$ and and $z = u + iv \in \mathbb{C}^+$,

$$|\text{trace}\left((M + qq' - z\text{Id}_p)^{-1}\right) - \text{trace}\left((M - z\text{Id}_p)^{-1}\right)| \leq \frac{1}{v} .$$

We use bounded martingale difference arguments as in [14], [22], [12].

• **Case 1: independent weights w_i**

Consider the filtration $\{\mathcal{F}_i\}_{i=0}^n$, with $\mathcal{F}_i = \sigma(w_1, \ldots, w_i)$ - the σ-field generated by $w_1, \ldots, w_i$ - and $\mathcal{F}_0 = \emptyset$.

Call $S_w^{(i)} = S_w - \frac{1}{n} w_i X_i X_i' - z\text{Id}_p$. In light of the result we just mentioned,

$$\frac{1}{p}|\text{trace}\left([S_w^{(i)}]^{-1}\right) - \text{trace}\left([S_w]^{-1}\right)| \leq \frac{1}{pv} .$$

In particular, this implies, since $\mathbf{E}\left(\text{trace}\left([S_w^{(i)}]^{-1}\right)|\mathcal{F}_i\right) = \mathbf{E}\left(\text{trace}\left([S_w^{(i)}]^{-1}\right)|\mathcal{F}_{i-1}\right)$ that

$$\frac{1}{p}|\mathbf{E}\left(\text{trace}\left([S_w]^{-1}\right)|\mathcal{F}_i\right) - \mathbf{E}\left(\text{trace}\left([S_w]^{-1}\right)|\mathcal{F}_{i-1}\right)| \leq \frac{2}{pv} .$$

Hence, $d_i = \frac{1}{p}\mathbf{E}\left(\text{trace}\left([S_w]^{-1}\right)|\mathcal{F}_i\right) - \frac{1}{p}\mathbf{E}\left(\text{trace}\left([S_w]^{-1}\right)|\mathcal{F}_{i-1}\right)$ is a bounded martingale-difference sequence. We can therefore apply Azuma's inequality ([18], p. 68), to get

$$P(|m_p(z) - \mathbf{E}\left(m_p(z)\right)| > t) \leq C \exp(-c\frac{p^2 v^2 t^2}{n}) .$$

In [12] it is shown that we can take $C = 4$ and $c = 1/16$.

• **Case 2: multinomial weights**

In this case, the previous result cannot be applied directly because the weights are not independent, since they must sum to n. However, to draw according to a Multinomial($n, 1/n$), we can simply pick an index from $\{1, \ldots, n\}$ uniformly and repeat the operation n times independently. Let $I(k)$ be the value of the index picked on the k-th draw from our sampling scheme. Clearly, the bootstrapped covariance matrix can be written as

$$S_w = \frac{1}{n}\sum_{k=1}^{n} X_{I(k)}X'_{I(k)} .$$

Consider the filtration $\{\mathcal{F}_i\}_{i=0}^n$, with $\mathcal{F}_i = \sigma(I(1),\ldots,I(i))$ - the σ-field generated by $I(1),\ldots,I(i)$ - and $\mathcal{F}_0 = \emptyset$. Clearly S_w is a sum of rank-1, independent matrices. So, if $S_w(k) = S_w - X_{I(k)}X'_{I(k)}/n$,

$$\mathbf{E}\left(\text{trace}\left([S_w^{(i)}]^{-1}\right)|\mathcal{F}_i\right) = \mathbf{E}\left(\text{trace}\left([S_w^{(i)}]^{-1}\right)|\mathcal{F}_{i-1}\right) .$$

The same argument as above therefore applies and the theorem is shown. □

S3.2 Proofs of Theorem 3.1 and 3.2

Proof of Theorem 3.1. Recall that Wielandt's Theorem (see p.261 in [9]) gives

$$\sup_{1\leq i\leq q} 0 \leq \lambda_i(S_n) - \lambda_i(T_n) \leq \frac{\lambda_{max}(U_nU_n')}{\lambda_q(T_n) - n^{-\alpha}\lambda_{max}(V_n)} ,$$

provided $\lambda_q(T_n) > n^{-\alpha}\lambda_{max}(V_n)$.

Recall that the Schur complement formula gives

$$n^{-\alpha}V_n \succeq U_n'T_n^{-1}U_n \succeq U_n'U_n/\lambda_{max}(T_n) ,$$

where the second inequality is a standard application of Lemma V.1.5 in [7]. Since $\lambda_{max}(U_nU_n') = \lambda_{max}(U_n'U_n)$ by simply writing the singular value decomposition of U_n, we conclude that

$$\lambda_{max}(T_n)n^{-\alpha}|||V_n|||_2 \geq \lambda_{max}(U_nU_n') .$$

So we conclude that provided $\lambda_q(T_n) > n^{-\alpha}\lambda_{max}(V_n)$,

$$\sup_{1\leq i\leq q} 0 \leq \lambda_i(S_n) - \lambda_i(T_n) \leq n^{-\alpha}\frac{\lambda_{max}(V_n)}{\lambda_q(T_n) - n^{-\alpha}\lambda_{max}(V_n)} . \tag{1}$$

• **Proof of Equation** (1) Note that under assumption **A2**, standard results in random matrix theory [13, 25, 26] guarantee that $|||V_n|||_2 = \mathrm{O}_P(1)$. Furthermore, standard results in classic multivariate analysis [1] show that $\lambda_q(T_n) \to \lambda_q(\Sigma_{11})$ in probability. Hence, we have

$$\frac{\lambda_{max}(V_n)}{\lambda_q(T_n) - n^{-\alpha}\lambda_{max}(V_n)} = \mathrm{O}_P(1) .$$

We therefore have

$$\sup_{1\leq i\leq q} \sqrt{n}(\lambda_i(S_n) - \lambda_i(T_n)) = \mathrm{O}_P(n^{1/2-\alpha}) .$$

• **Proof of Equation** (2) We note that if D_w is the diagonal matrix with the bootstrap weights on the diagonal, we have

$$S_n^* = \frac{1}{n}X'D_wX .$$

Therefore, we see that $|||T_n^*|||_2 = \mathrm{O}_{P,w}(1)$, $\lambda_q(T_n^*) \to_{P,w} \lambda_q(\Sigma_{11})$ by the law of large numbers (provided $\mathbf{E}(w_i) = 1$; the case of Multinomial$(n, 1/n)$ weights is also easy to deal with by the technique described in the previous subsection for instance) and $|||V_n^*|||_2 = \mathrm{O}_{P,w}(\text{polyLog}(n))$ provided $\|w\|_\infty = \text{polyLog}(n)$.

We can then conclude that

$$\sup_{1 \leq i \leq q} \sqrt{n}(\lambda_i(S_n^*) - \lambda_i(T_n^*)) = \mathrm{o}_{P,w}(1) \ .$$

□

Preliminary remarks concerning Theorem 3.2 Before we prove these theorems, we recall the definitions of bootstrap consistency.

Definition 1. Suppose $\widehat{\theta}_n(X_1, \ldots, X_n)$ is a statistic, $\widehat{\theta}_n^*$ is its bootstrapped version. Suppose that $\widehat{\theta}_n \Longrightarrow T$. We say that *the bootstrap is consistent in probability* if

$$\widehat{\theta}_n^* \Longrightarrow_w T \quad \text{in } P_{X_1,\ldots,X_n} - \text{probability}.$$

Here $\Longrightarrow_w$ refers to weak convergence of $\widehat{\theta}_n^*$ under the bootstrap weight distribution; the convergence in probability is with respect to the joint distribution of $X_1, \ldots, X_n$, which we denote $P_{X_1,\ldots,X_n}$. For simplicity, we often abbreviate $P_{X_1,\ldots,X_n}$ by P.

We say that *the bootstrap is strongly consistent* if

$$\widehat{\theta}_n^* \Longrightarrow_w T \quad \text{a.s } P_{X_1,\ldots,X_n}.$$

Key results of [5, 6] and [9], p.269 show that in the classical low-dimensional case, where p is fixed and $n \to \infty$, if all eigenvalues of Σ are simple, the bootstrap distribution of the eigenvalues of S_n is strongly consistent. On the other hand, it is known from these papers that the bootstrap distribution of the eigenvalues of S_n is inconsistent when the eigenvalues of Σ have multiplicities higher than 1.

Proof of Theorem 3.2. The results from Theorem 3.1 imply that

$$\sup_{1 \leq i \leq q} \left| \left[\sqrt{n}(\lambda_i(S_n^*) - \lambda_i(S_n))\right] - \left[\sqrt{n}(\lambda_i(T_n^*) - \lambda_i(T_n))\right] \right| = \mathrm{o}_{P,w}(1) \ . \tag{2}$$

The arguments used in the proof of Theorem 3.1 also apply to Σ and show that

$$\sup_{1 \leq i \leq q} |\lambda_i(\Sigma_n) - \lambda_i(\Sigma_{11})| \leq n^{-\alpha} \frac{\lambda_{max}(\Sigma_{22})\lambda_{max}(\Sigma_{11})}{\lambda_q(\Sigma_{11}) - n^{-\alpha}\lambda_{max}(\Sigma_{22})}$$

Hence, when $\alpha > 1/2 + \epsilon$, we have

$$\sqrt{n} \sup_{1 \leq i \leq q} |\lambda_i(\Sigma_n) - \lambda_i(\Sigma_{11})| = \mathrm{o}(1) \ .$$

Therefore,

$$\sup_{1 \leq i \leq q} \sqrt{n} \, |\lambda_i(S_n) - \lambda_i(\Sigma_n) - [\lambda_i(T_n) - \lambda_i(\Sigma_{11})]| = \mathrm{o}_P(1) \ .$$

Hence, the q largest eigenvalues of S_n have the same limiting fluctuation behavior as the q largest eigenvalues of T_n (classical results [1] show that $\sqrt{n}$ is the correct order of fluctuations). The same is true for the bootstrapped version of their distributions, according to Equation (2).

Since the aforementioned results of [5, 6, 9] show consistency of the bootstrap distribution of the eigenvalues of T_n, this result carries over to the bootstrap distribution of S_n. □

S3.2.1 Discussion of the assumptions of Theorems 3.1 and 3.2

Recall our assumptions: **A1** We assume that $|||\Sigma_{22}|||_2 = \mathrm{O}(1)$ and that $\lambda_{min}(\Sigma_{11}) > \eta > 0$. We assume that Σ_{11} is $q \times q$ with q fixed. **A2** X_i's are i.i.d with $X_i = r_i Z_i$, where $Z_i \sim \mathcal{N}(0, \Sigma_n)$, and $0 < \delta_0 < r_i < \gamma_0$ is a bounded random variable independent of Z_i, with $\mathbf{E}\left(r_i^2\right) = 1$.**A3** The bootstrap weights w_i have infinitely many moments, $\|w\|_\infty = \mathrm{O}(\text{polyLog}(n))$ and $\mathbf{E}\left(w_i\right) = 1$. These weights can either be independent or Multinomial$(n, 1/n)$. **A4** p/n remains bounded as n and p tend to infinity.

Distributional assumptions on X_i's The assumption that $Z_i \sim \mathcal{N}(0, \Sigma_n)$ is not critical: most of our arguments could be adapted to handle the case where $Z_i = \Sigma_n^{1/2} Y_i$, where Y_i has independent, mean 0, variance 1 entries, with sufficiently many moments. This simply requires appealing to slightly different random matrix results that exist in the literature. Also, the first q coordinates of X_i could have a much more general distribution than the elliptical distributions we consider here, as our proof simply requires control of $|||V_n|||_2$, which is where we appeal to random matrix theory. Doing this entails minor technical modifications to the proof, but since it might reduce clarity, we leave them to the interested reader.

Assumptions on Σ The block representation assumptions are made for analytic convenience and can be easily dispensed of: eigenvalues are of course unaffected by rotations, so we simply chose to write Σ in a basis that gave us this nice block format. As long as the ratio between the q-th eigenvalue of Σ and $q+1$st is of order n^α, our results hold. Furthermore, our results also handle a situation similar to ours, where for instance the top q largest eigenvalues of Σ grow like n^α and the $q+1$st is of order 1, by simple rescaling, using for instance the fact that trace (T_n) /trace $(\Sigma_{11}) \to 1$ in probability.

Strong consistency of the bootstrap We have chosen to present our results using convergence in probability statements, as we think they better reflect the questions encountered in practice. However, a quick scan through the proofs show that all the approximation results could be extended to a.s convergence: the random matrix results we rely on hold a.s, and the low-dimensional bootstrap results we use also hold a.s in low-dimension.

S3.3 On bootstrap bias for the top eigenvalue in the non well-separated case

Recall that if M is a symmetric matrix, the application $M \mapsto \lambda_1(M)$ is convex. We assume that the bootstrap weights w_i's have (bootstrap) mean 1, i.e. $\mathbf{E}^*\left(w_i\right) = 1$.

For the sake of simplicity we consider the case where X_i's are i.i.d $\mathcal{N}(0, \Sigma_n)$, in which case the maximum likelihood estimate of covariance is $\widehat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^n X_i X_i'$. Applying Jensen's inequality to $\widehat{\Sigma}_n$, we see that

$$\mathbf{E}\left(\lambda_1(\widehat{\Sigma}_n)\right) \geq \lambda_1(\Sigma) \ .$$

Now the bootstrapped version of $\widehat{\Sigma}_n$ is $\widehat{\Sigma}_n^* = \frac{1}{n}\sum_{i=1}^n w_i X_i X_i'$. Hence we have for the bootstrap expectation, when $\mathbf{E}^*\left(w_i\right) = 1$

$$\mathbf{E}^*\left(\lambda_1(\widehat{\Sigma}_n^*)\right) \geq \lambda_1(\mathbf{E}^*\left(\widehat{\Sigma}_n^*\right)) = \lambda_1(\widehat{\Sigma}_n) \ .$$

In other words, Jensen's inequality, applied in two slightly different manners, shows that the sample largest eigenvalue is potentially biased and so is the sample largest bootstrap eigenvalue. As we have recalled in this supplementary material, $\lambda_1(\widehat{\Sigma}_n)$ is severely biased when p/n is not close to 0. A naive application of the bootstrap to estimate the sampling distribution of such a biased estimator is clearly a bad idea, a fact which is not perhaps sufficiently appreciated in the context of the bootstrap for eigenvalues. But this does not rule out in general that the bootstrap could work for estimating simpler quantities than the sampling distribution, such as the bias of the largest eigenvalue.

We have the following simple observation.

Fact S3.1. *Suppose X_i's are i.i.d $\mathcal{N}(0, \mathrm{Id}_p)$. Suppose the bootstrap weights have distribution $\mathcal{W}$. Consider the random matrix $S_w = \frac{1}{n}\sum_{i=1}^n w_i X_i X_i'$ where X_i's and w_i's are drawn independently has the property that its largest eigenvalue converges in probability (with respect to $P_{\{X_i\}_{i=1^n}} \otimes P_{\mathcal{W}}$ to the edge of the limiting empirical spectral distribution of $S_{\mathcal{W}}$, denoted by $e_{\mathcal{W}}$.*

Then

$$\lambda_1(S_w^*) - e_{\mathcal{W}} = \mathrm{o}_{P_{\mathcal{W}}}(1) \text{ with probability 1 wrt} P_{X_1,\ldots,X_n} \; .$$

This very simple fact follows immediately from the fact that if $\lambda_1(S_w)$ converges in probability to a constant jointly with respect to $\mathcal{W}$ and $P_{X_1,\ldots,X_n}$, it converges to the same constant conditionally on X_i's, with probability 1 with respect to X_i's.

An idealized and simplified example For technical reasons, let us now assume that w's are drawn i.i.d according to a $Poisson(1)$ distribution conditioned to be less than K, where K is a fixed number, say $K = 20$ when n is a few 100's like in our paper. We use this distribution as an approximation of bootstrap weights truncated to not be too large. (Recall that the marginal distribution of a $Mult(n, 1/n)$ converges to a $Poisson(1)$ distribution.) For this distribution of weights, results such as [11, 21, 3] apply, after we recall that if $M = DX$, its eigenvalues are the same as those of $M' = X'D'$ and understanding the limiting spectral distribution of S_w is the same as understanding the spectrum of a sample covariance matrix computed from $\tilde{X}_i$ where $\tilde{X}_i \sim \mathcal{N}(0, \Sigma_w)$ data with $\Sigma_w = \mathrm{diag}(w)$. Naturally in this formulation the role of n and p has been switched. In particular, rewriting slightly the results of these papers, the largest eigenvalue of $\frac{n}{p} S_w$ converges to $\mu_{n,p}(\mathcal{W})$ with

$$\mu_{n,p}(\mathcal{W}) = \frac{1}{c_{n,p}}\left(1 + \frac{n}{p}\int \frac{w c_{n,p}}{1 - w c_{n,p}} dH_n(w)\right) \text{ , where } n\int\left(\frac{w c_{n,p}}{1 - w c_{n,p}}\right)^2 dH_n(w) = p$$

where H_n is the empirical distribution of w_i's, i.e. $dH_n(w) = \frac{1}{n}\sum_{i=1}^n \delta_{w_i}$ and $c_{n,p}$ is the unique solution of the equation on the right hand side in $(0, 1/\max_{1 \leq i \leq n} w_i)$.

In general, we will have

$$\mu_{n,p}(\mathcal{W}) - (1 + \sqrt{p/n})^2 \neq (1 + \sqrt{(p/n)})^2 - 1 \; .$$

In other words, the bootstrap estimate of bias (or its technical idealization here) is not going to be a consistent estimate of the bias of the largest eigenvalue.

More generally and related to these arguments, it is clear from [24] or [11] and their main results recalled above that the bootstrap weight distribution affects the *support* of the limiting spectral distributions, and thus the range of values the top eigenvalues can take. As we just did, looking carefully at the discrepancy in the support of the sample spectral distribution and the bootstrap weight distribution gives strong theoretical intuition as to why the (bootstrap) bias in the bootstrapped $\lambda_1(\widehat{\Sigma}^*)$ is unlikely to be the same as the bias of the sample estimate $\lambda_1(\widehat{\Sigma})$.

The situation is however complicated by the fact that the largest eigenvalue of the sample covariance matrix does not always converge to the end point of its limiting spectral distribution [26] and that as we recalled rather subtle phase-transitions can occur for these statistics [4]. This is why the mathematically motivated intuition given above applies only to situations where the extreme eigenvalues of Σ are not well-separated from the bulk. In this case it has been shown in a variety of situations [11, 19] that the sample extreme eigenvalues stay close to the edge of the limiting spectral distribution, so that using support of the limiting spectral distribution as a proxy for the location of the extreme eigenvalues provides a plausible explanation for the problems we observed with the bootstrap estimate of bias.

We note however that the supremum of the support of the limiting spectral distribution provides a lower bound on the top eigenvalue; hence using this proxy for the top eigenvalue would help when showing that the bootstrap estimate of bias is positively biased.

References

[1] T. W. Anderson. *An introduction to multivariate statistical analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition, 2003.

[2] Z. D. Bai. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica*, 9(3):611–677, 1999. With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author.

[3] Z. D. Bai and Jack W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1):316–345, 1998.

[4] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for non-null complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.

[5] Rudolf Beran and Muni S. Srivastava. Bootstrap tests and confidence regions for functions of a covariance matrix. *Ann. Statist.*, 13(1):95–115, 1985.

[6] Rudolf Beran and Muni S. Srivastava. Correction: "Bootstrap tests and confidence regions for functions of a covariance matrix" [Ann. Statist. **13** (1985), no. 1, 95–115; MR0773155 (86g:62054)]. *Ann. Statist.*, 15(1):470–471, 1987.

[7] Rajendra Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[8] A. Boutet de Monvel, A. Khorunzhy, and V. Vasilchuk. Limiting eigenvalue distribution of random matrices with correlated entries. *Markov Process. Related Fields*, 2(4):607–636, 1996.

[9] Morris L. Eaton and David E. Tyler. On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *Ann. Statist.*, 19(1):260–271, 1991.

[10] Noureddine El Karoui. On the largest eigenvalue of Wishart matrices with identity covariance when n, p and $p/n \to \infty$. *arXiv:math.ST/0309355*, September 2003.

[11] Noureddine El Karoui. Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability*, 35(2):663–714, March 2007.

[12] Noureddine El Karoui. Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *The Annals of Applied Probability*, 19(6):2362–2405, December 2009.

[13] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.

[14] F. Götze and A. N. Tikhomirov. Limit theorems for spectra of random matrices with martingale structure. In *Stein's method and applications*, volume 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 181–193. Singapore Univ. Press, Singapore, 2005.

[15] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[16] Iain M. Johnstone. High dimensional statistical inference and random matrices. In *International Congress of Mathematicians. Vol. I*, pages 307–333. Eur. Math. Soc., Zürich, 2007.

[17] I.M. Johnstone. On the distribution of the largest eigenvalue in principal component analysis. *Ann. Statist.*, 29(2):295–327, 2001.

[18] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[19] Ji oon Lee and Kevin Schnelli. Tracy-widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *arxiv:1409.4979*, 2014.

[20] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967.

[21] Alexei Onatski. The Tracy-Widom limit for the largest eigenvalues of singular complex Wishart matrices. *Ann. Appl. Probab.*, 18(2):470–490, 2008.

[22] A. Pajor and L. Pastur. On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution. *Studia Math.*, 195(1):11–29, 2009.

[23] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4):1617–1642, October 2007.

[24] Debashis Paul and Jack W. Silverstein. No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix. *J. Multivariate Anal.*, 100(1):37–57, 2009.

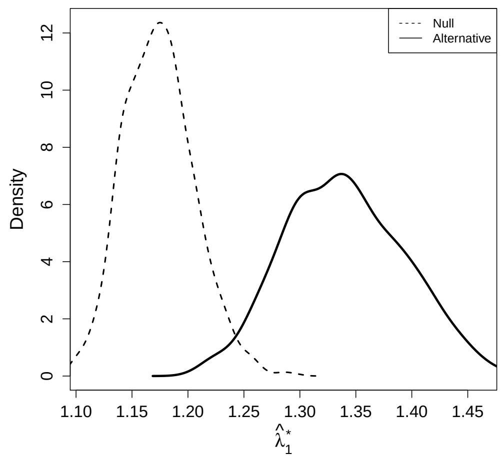
[25] Jack W. Silverstein. The smallest eigenvalue of a large-dimensional Wishart matrix. *Ann. Probab.*, 13(4):1364–1368, 1985.

[26] Jack W. Silverstein. On the weak limit of the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivariate Anal.*, 30(2):307–311, 1989.

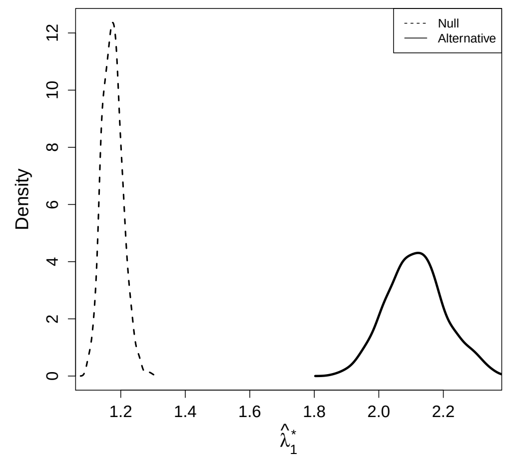
[27] Jack W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, 55(2):331–339, 1995.

[28] Kenneth W. Wachter. The strong limits of random matrix spectra for sample matrices of independent elements. *Annals of Probability*, 6(1):1–18, 1978.

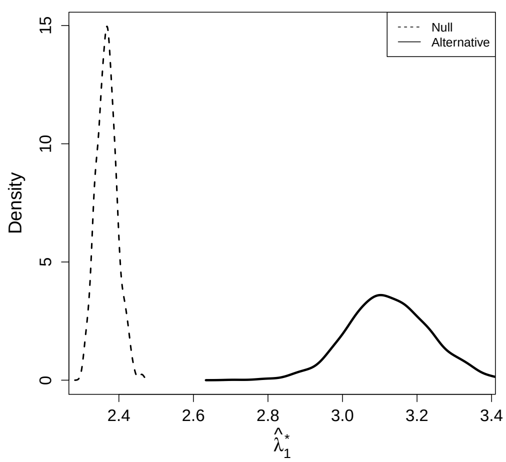
S4 Supplementary Figures



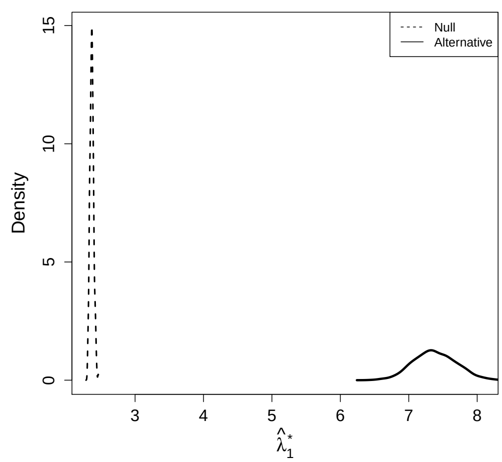
(a) $\lambda_1^{Alt} = 1 + 3\sqrt{r}$, r=0.01



(b) $\lambda_1^{Alt} = 1 + 11\sqrt{r}$, r=0.01

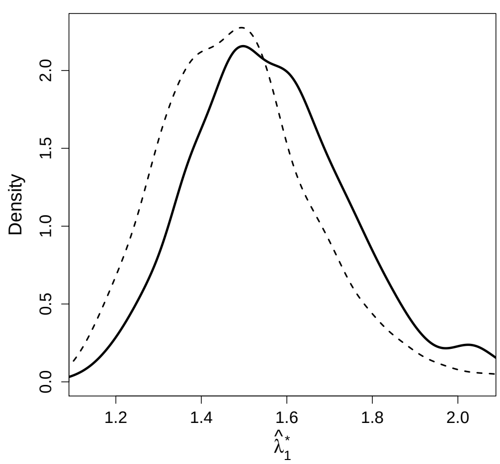


(c) $\lambda_1^{Alt} = 1 + 3\sqrt{r}$, r=0.03

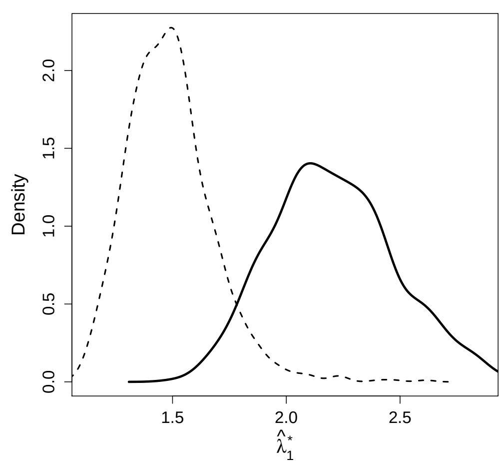


(d) $\lambda_1^{Alt} = 1 + 11\sqrt{r}$, r=0.3

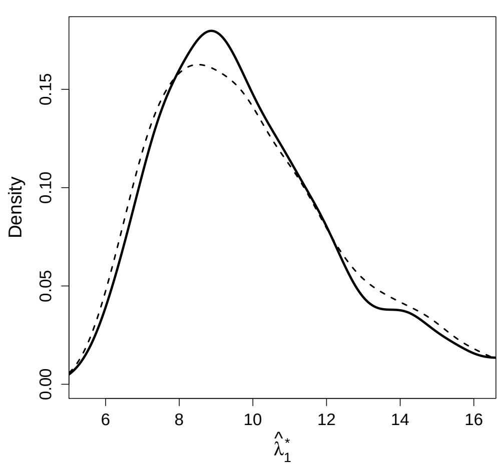
Figure S1: **Top Eigenvalue: Distribution of Largest Eigenvalue, Null versus Alternative,** $X_i \sim$ **Normal**



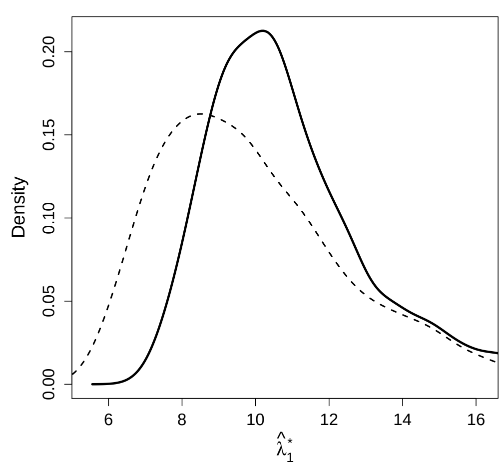
(a) $\lambda_1^{Alt} = 1 + 3\sqrt{r}$, r=0.01



(b) $\lambda_1^{Alt} = 1 + 11\sqrt{r}$, r=0.01

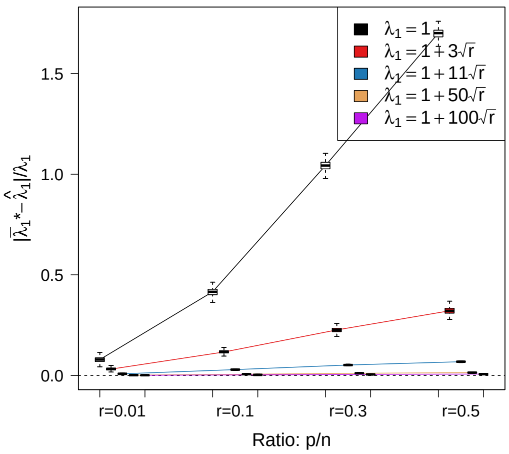


(c) $\lambda_1^{Alt} = 1 + 3\sqrt{r}$, r=0.03

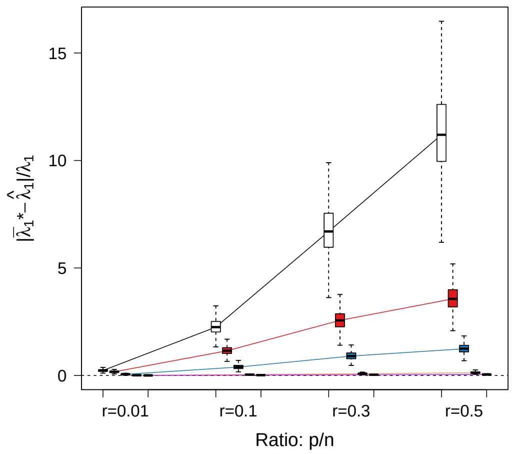


(d) $\lambda_1^{Alt} = 1 + 11\sqrt{r}$, r=0.3

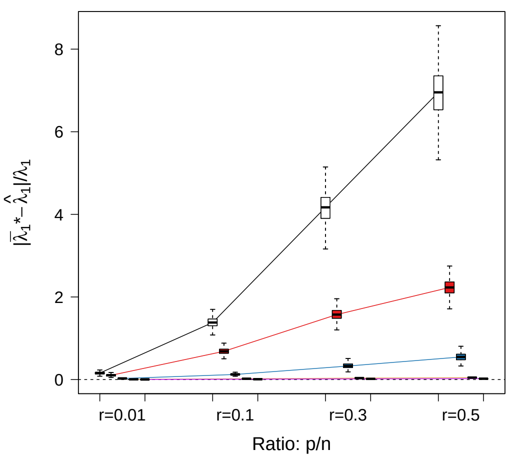
Figure S2: **Top Eigenvalue: Distribution of Largest Eigenvalue, Null versus Alternative,** $X_i \sim$ **Ellip. Exp**



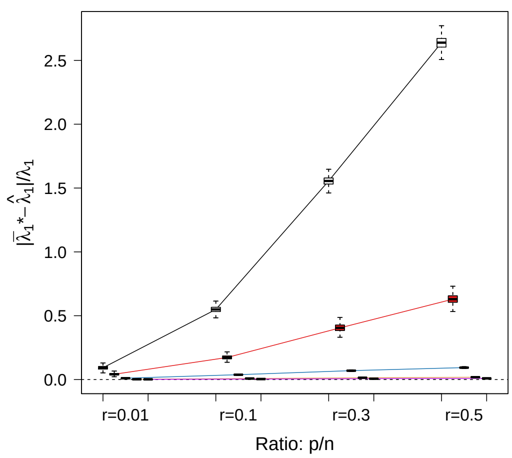
(a) $X_i \sim$ Normal



(b) $X_i \sim$ Ellip. Exp

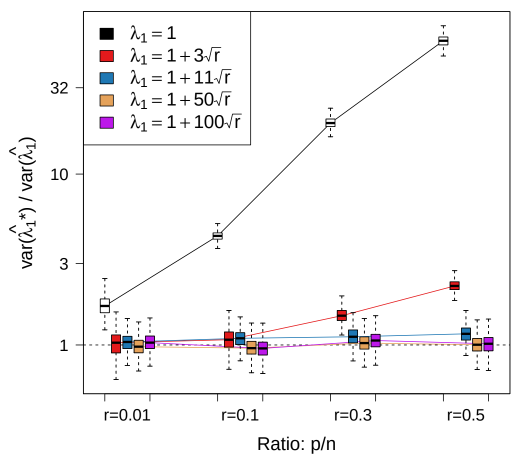


(c) $X_i \sim$ Ellip. Normal

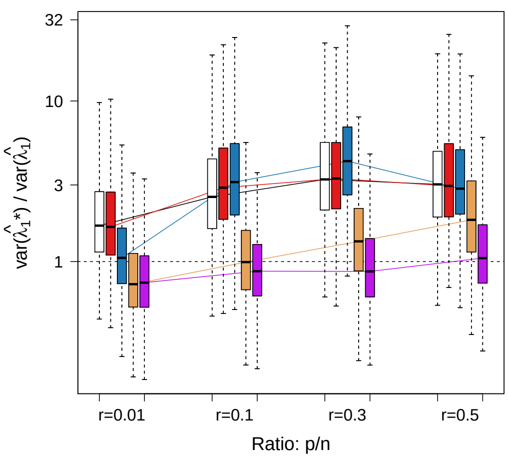


(d) $X_i \sim$ Ellip. Uniform

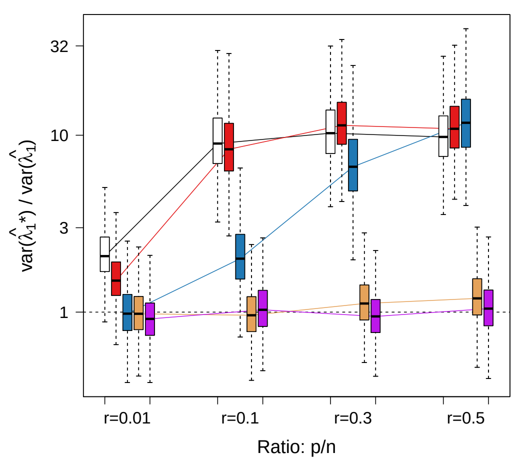
Figure S3: **Top Eigenvalue: Bias of Largest Bootstrap Eigenvalue, n=1,000:** Plotted are boxplots of the bootstrap estimate of bias covering larger values of λ_1 than shown in the main text. Unlike the main text, here we scale the bias by the true λ_1 so as to make the comparisons more comparable ($|\bar{\lambda}_1^* - \hat{\lambda}_1|/\lambda_1$). See the legend of Figure 1 in the main text for more details.



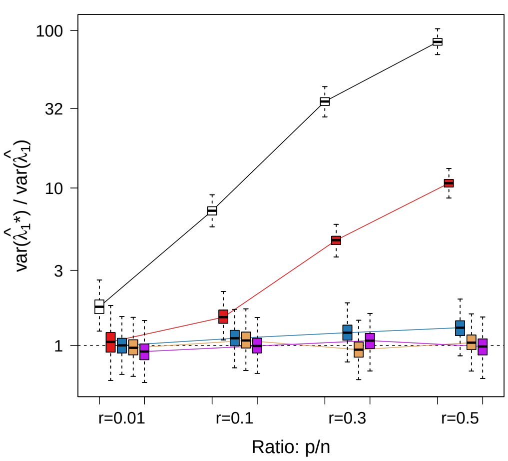
(a) $X_i \sim$ Normal



(b) $X_i \sim$ Ellip. Exp

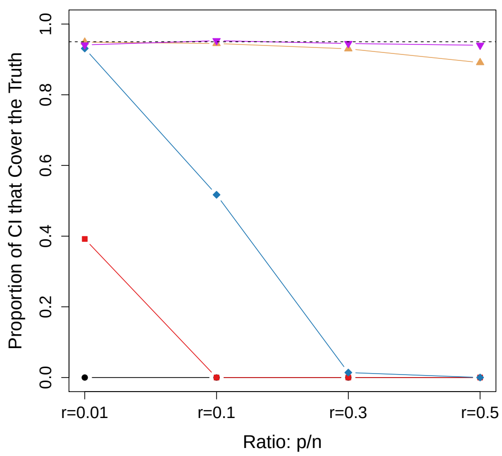


(c) $X_i \sim$ Ellip Normal

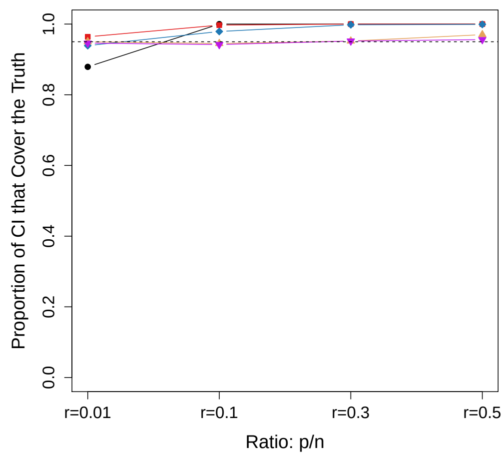


(d) $X_i \sim$ Ellip. Uniform

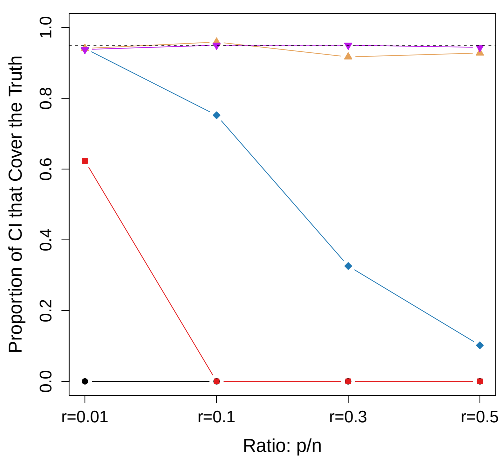
Figure S4: **Top Eigenvalue: Ratio of Bootstrap Estimate of Variance to True Variance for Largest Eigenvalue, n=1,000:** Plotted are boxplots of the bootstrap estimate of variance, showing larger values of λ_1 than shown in the main text. See the legend of Figure 1 in the main text for more details.



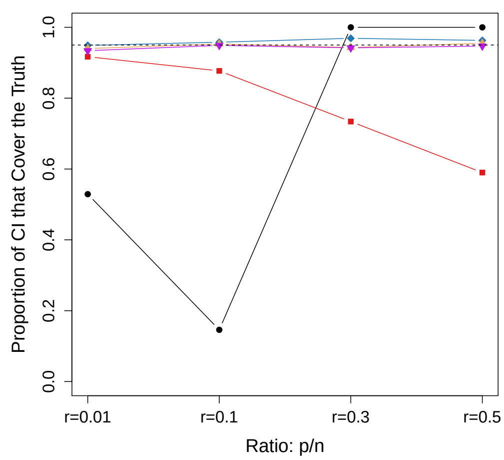
(a) $X_i \sim$ Ellip Normal
Percentile Intervals



(b) $X_i \sim$ Ellip Normal
Normal-based Intervals

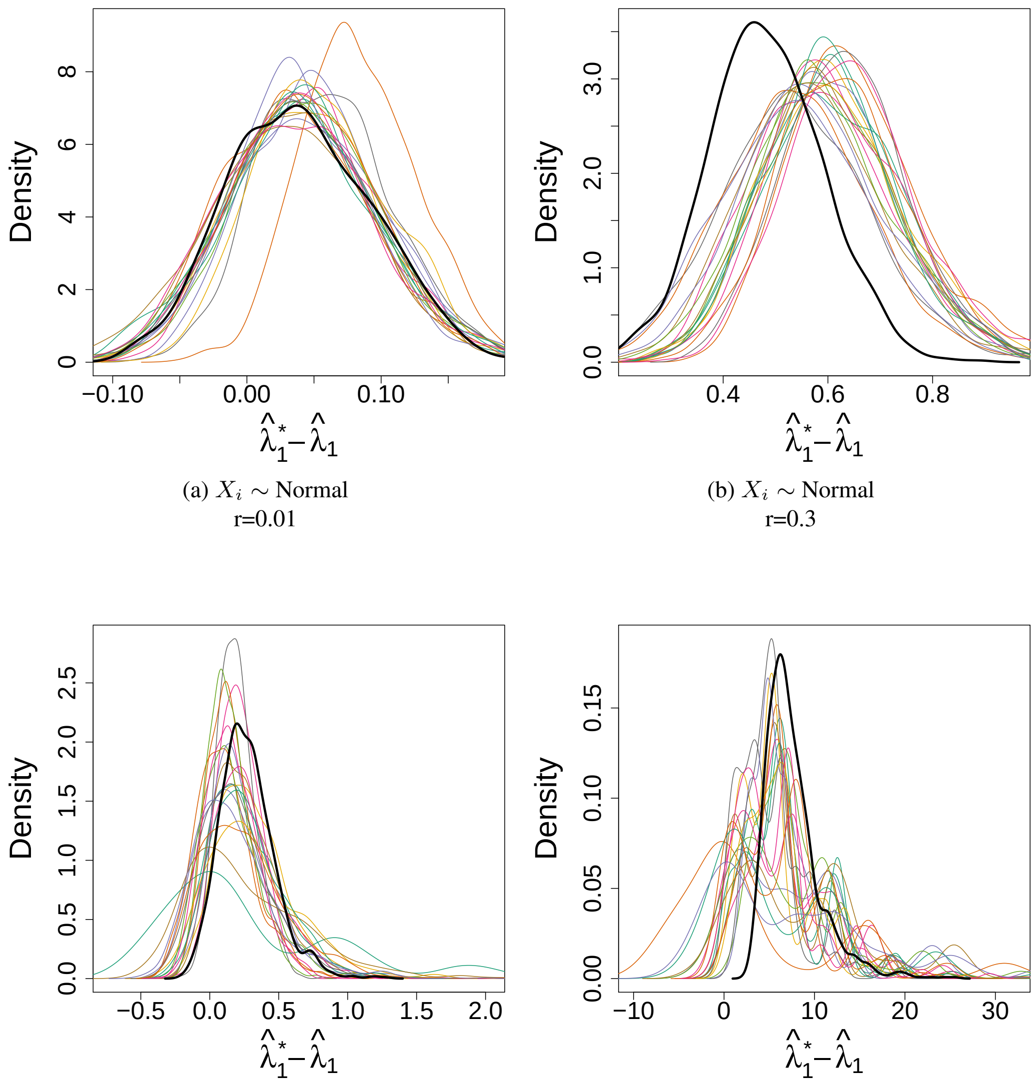


(c) $X_i \sim$ Ellip Uniform
Percentile Intervals



(d) $X_i \sim$ Ellip Uniform
Normal-based Intervals

Figure S5: **Top Eigenvalue: 95% CI Coverage,** $n = 1,000$ **for additional distributions:** Plotted are the corresponding CI Coverage plots for when X_i follows an elliptical distribution with Normal and Uniform weights. See Figure 3 for more details.



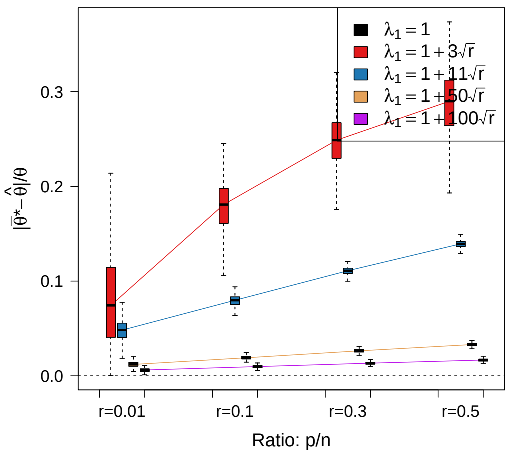
(a) $X_i \sim$ Normal
r=0.01

(b) $X_i \sim$ Normal
r=0.3

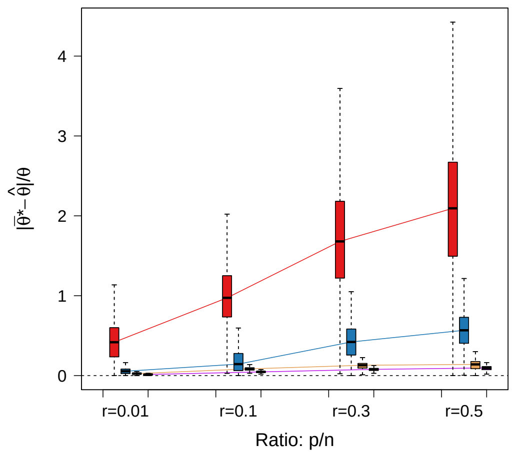
(c) $X_i \sim$ Ellip. Exp
r=0.01

(d) $X_i \sim$ Ellip. Exp
r=0.3

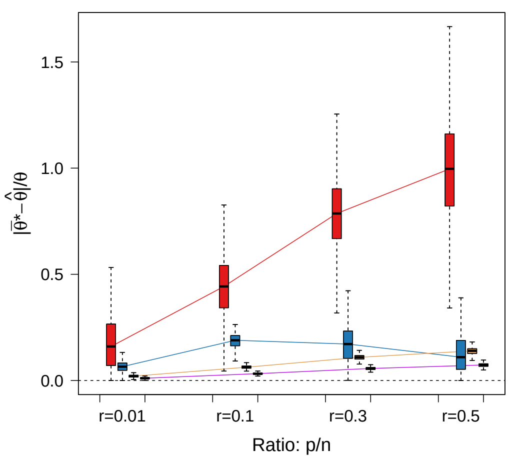
Figure S6: **Top Eigenvalue: Bootstrap distribution of $\hat{\lambda}_1^*$ when $\lambda_1 = 1 + 3\sqrt{r}$, n=1,000:** Plotted are the estimated density of twenty simulations of the bootstrap distribution of $\hat{\lambda}_1^{*b} - \hat{\lambda}_1$, with $b = 1, \ldots, 999$. The solid black line line represents the distribution of $\hat{\lambda}_1 - \lambda_1$ over 1,000 simulations.



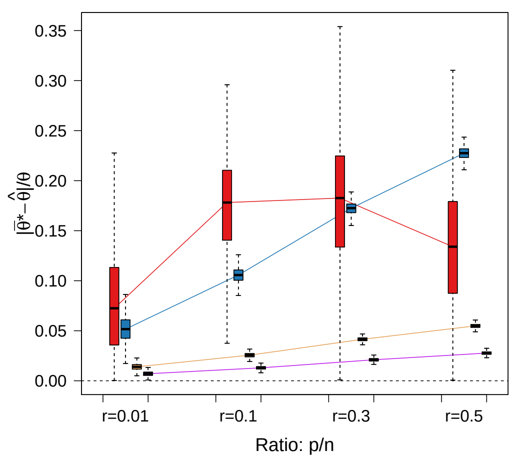
(a) $X_i \sim$ Normal



(b) $X_i \sim$ Ellip. Exp

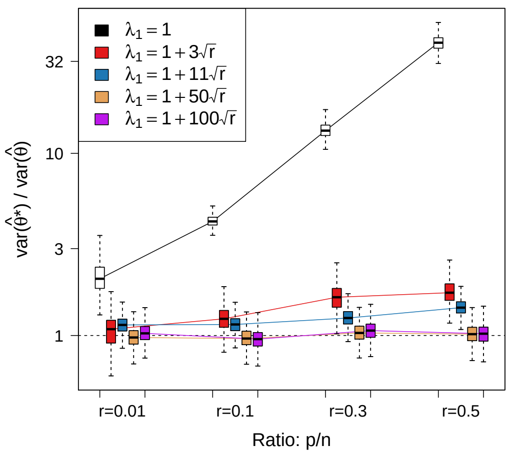


(c) $X_i \sim$ Ellip. Norm

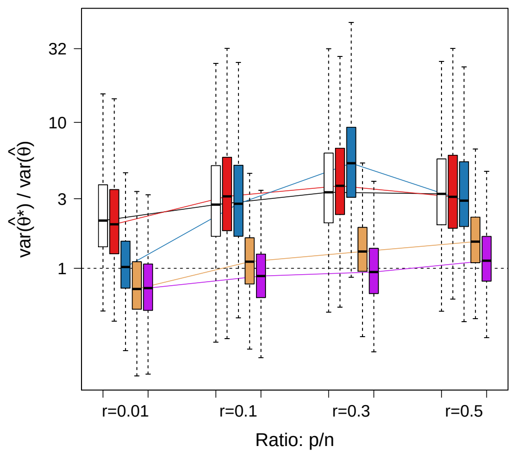


(d) $X_i \sim$ Ellip. Unif

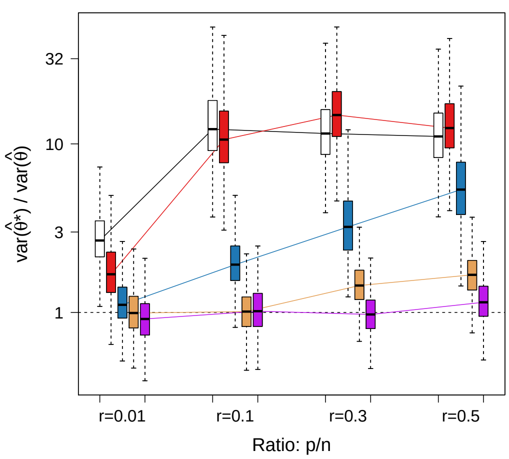
Figure S7: **Gap Statistic: Ratio of Bias of Bootstrap to true Gap Statistic.** Note that the true Gap Statistic when $\lambda_1 = 1$ is zero, so that the ratio is not well defined and hence not plotted . See also Supplementary Figures S8-S11 for the median bias values and the legend of Figure 1 in the main text for more information about this plot. Note that the y-axis is different for each of the distributions, and differs from that of Supplementary Figure S3.



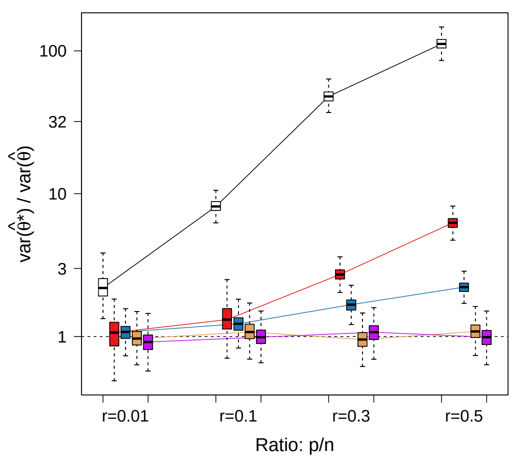
(a) $X_i \sim$ Normal



(b) $X_i \sim$ Ellip. Exp

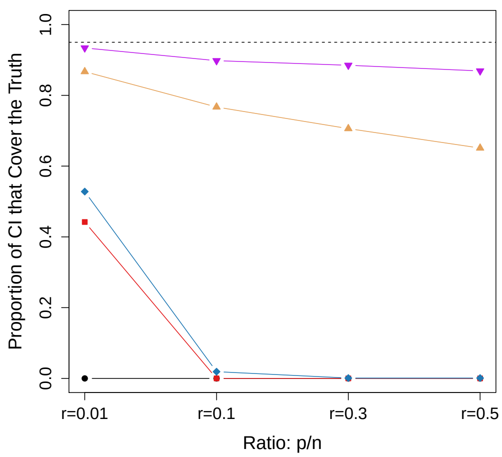


(c) $X_i \sim$ Ellip. Normal

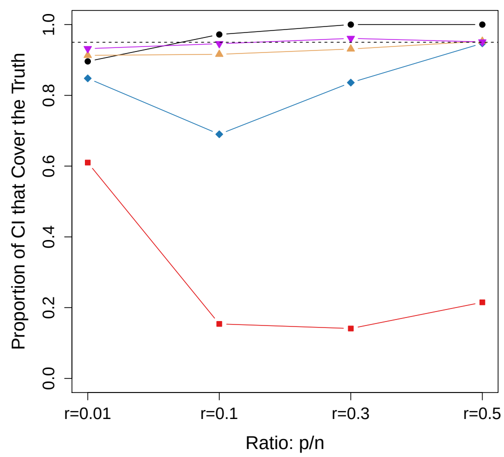


(d) $X_i \sim$ Ellip. Uniform

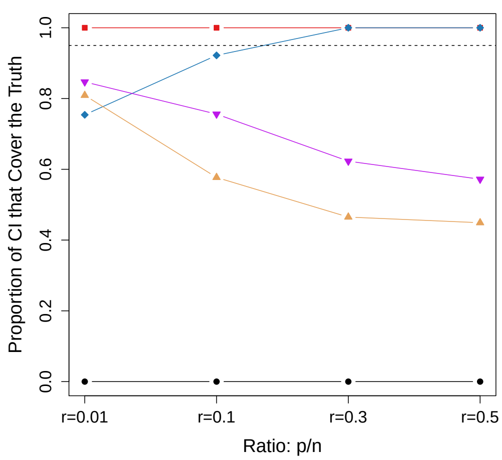
Figure S8: **Gap Statistic: Ratio of Bootstrap Estimate of Variance to True Variance.**



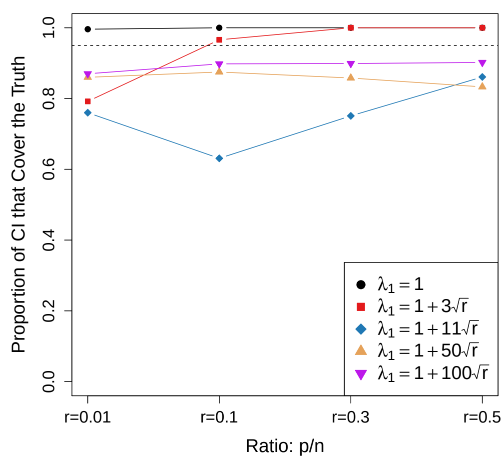
(a) $X_i \sim$ Normal
Percentile Intervals



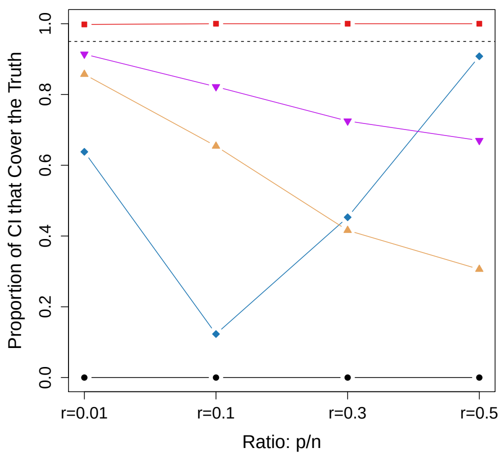
(b) $X_i \sim$ Normal
Normal-based Intervals



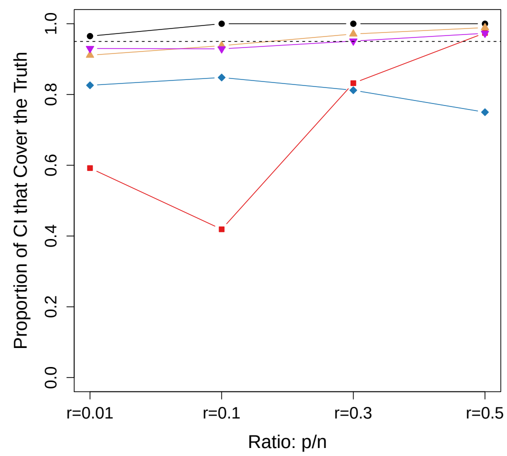
(c) $X_i \sim$ Ellip Exp
Percentile Intervals



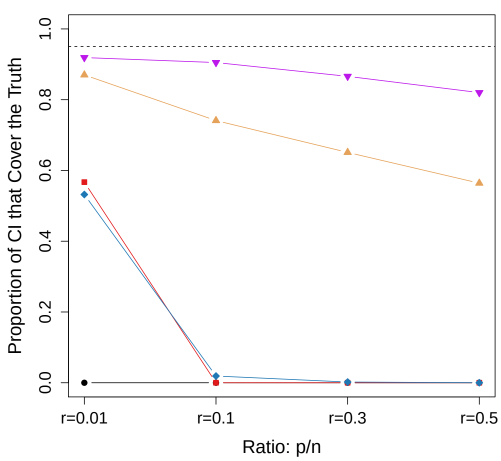
(d) $X_i \sim$ Ellip Exp
Normal-based Intervals



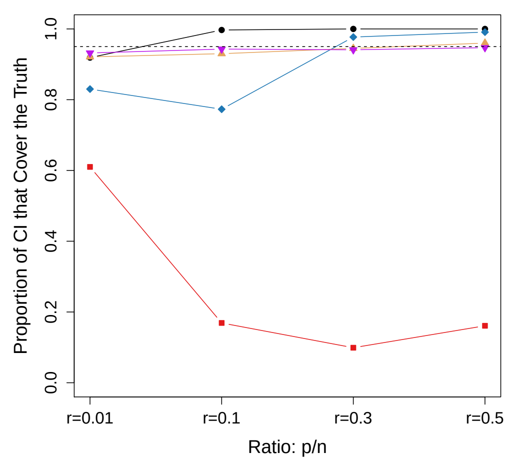
(e) $X_i \sim$ Ellip Normal
Percentile Intervals



(f) $X_i \sim$ Ellip Normal
Normal-based Intervals

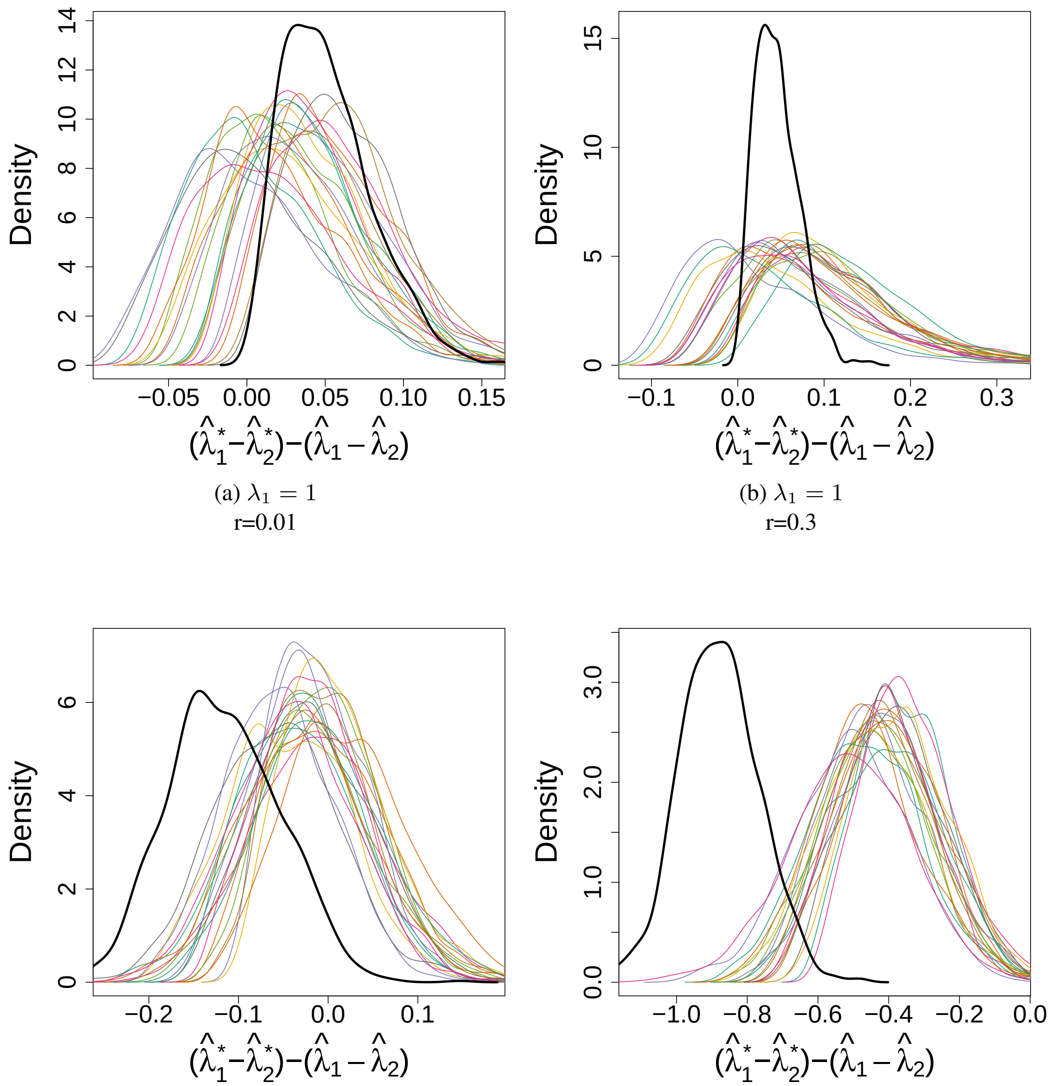


(g) $X_i \sim$ Ellip Uniform
Percentile Intervals



(h) $X_i \sim$ Ellip Uniform
Normal-based Intervals

Figure S8: **Gap Statistic: 95% CI Coverage,** $n = 1,000$**:**



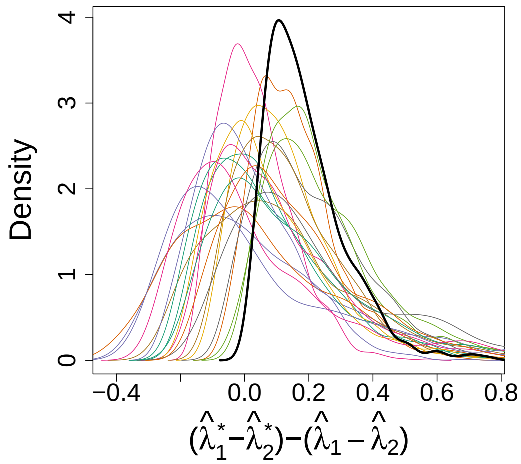
(a) $\lambda_1 = 1$
r=0.01

(b) $\lambda_1 = 1$
r=0.3

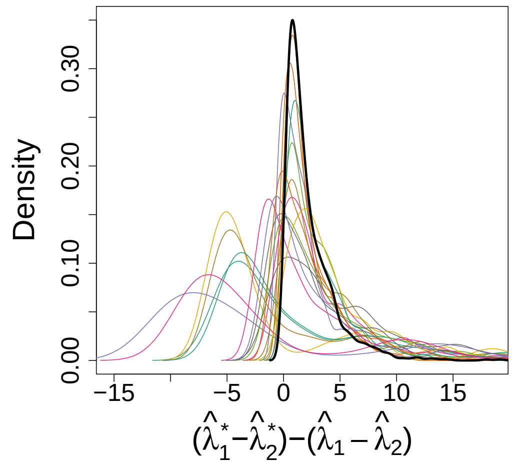
(c) $\lambda_1 = 1 + 3\sqrt{r}$
r=0.01

(d) $\lambda_1 = 1 + 3\sqrt{r}$
r=0.3

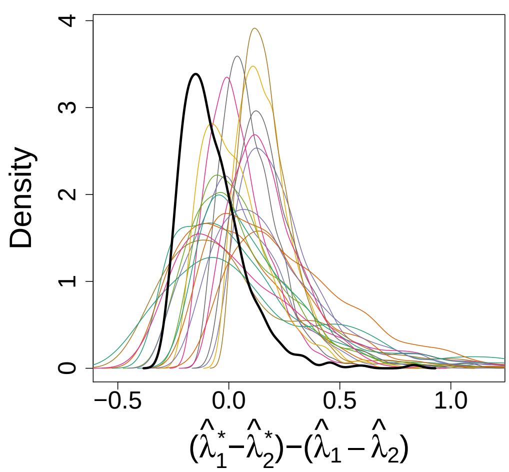
Figure S9: **Gap Statistic: Bootstrap distribution,** X_i $\sim$ **Normal, n=1,000:** Plotted are the estimated density of twenty simulations of the bootstrap distribution of $(\hat{\lambda}_1^{*b} - \hat{\lambda}_2^{*b}) - (\hat{\lambda}_1 - \hat{\lambda}_2)$, with $b = 1, \ldots, 999$. The solid black line line represents the distribution of $(\hat{\lambda}_1 - \hat{\lambda}_2) - (\lambda_1 - \lambda_2)$ over 1,000 simulations.



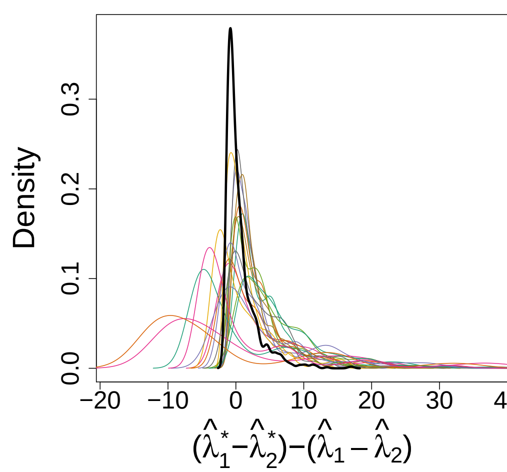
(a) $\lambda_1 = 1$
r=0.01



(b) $\lambda_1 = 1$
r=0.3

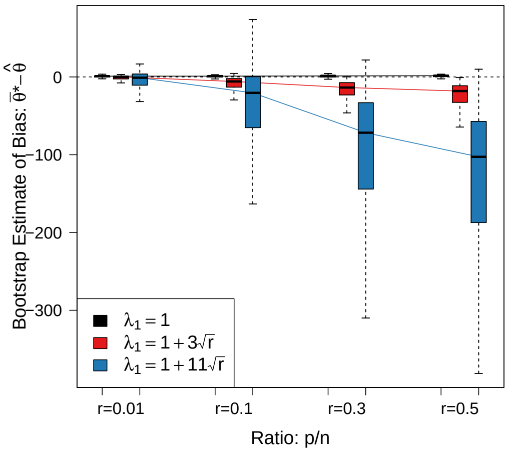


(c) $\lambda_1 = 1 + 3\sqrt{r}$
r=0.01

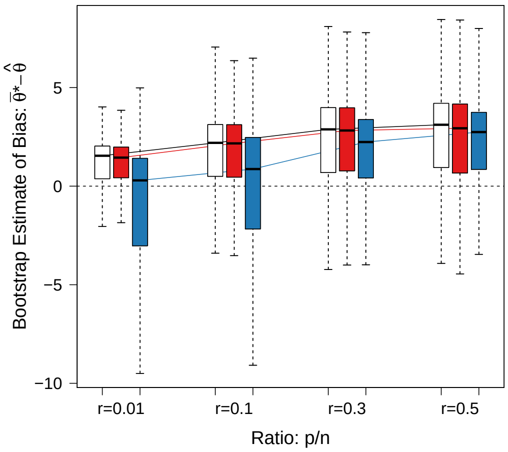


(d) $\lambda_1 = 1 + 3\sqrt{r}$
r=0.3

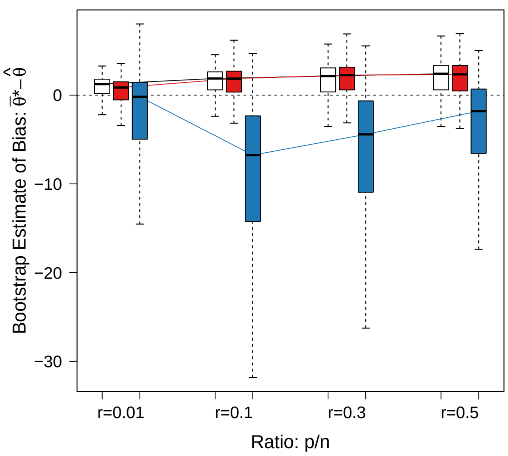
Figure S10: **Gap Statistic: Bootstrap distribution,** $X_i \sim$ **Ellip Exp, n=1,000:** Plotted are the estimated density of twenty simulations of the bootstrap distribution of $(\hat{\lambda}_1^{*b} - \hat{\lambda}_2^{*b}) - (\hat{\lambda}_1 - \hat{\lambda}_2)$, with $b = 1, \ldots, 999$. The solid black line line represents the distribution of $(\hat{\lambda}_1 - \hat{\lambda}_2) - (\lambda_1 - \lambda_2)$ over 1,000 simulations.



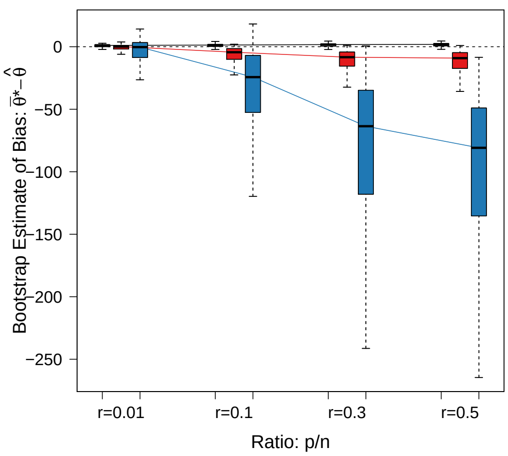
(a) $X_i \sim$ Normal



(b) $X_i \sim$ Ellip. Exp

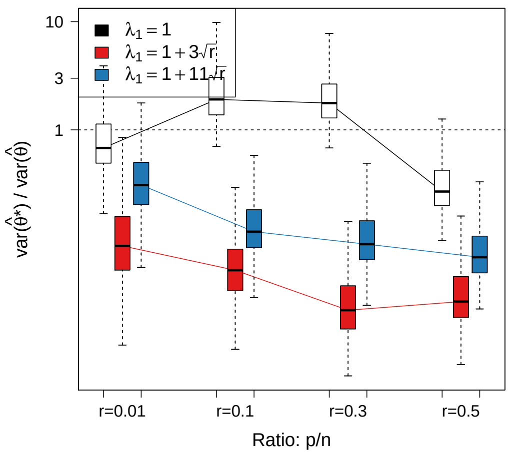


(c) $X_i \sim$ Ellip. Norm

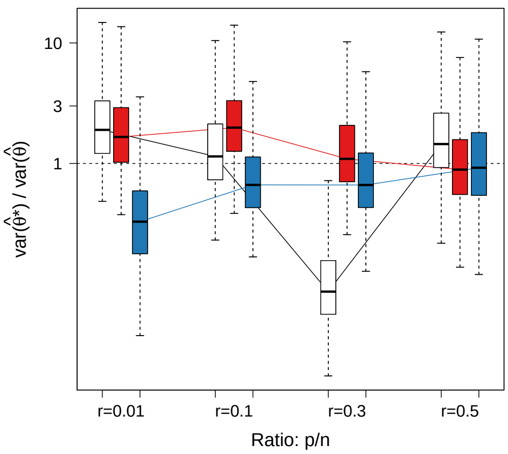


(d) $X_i \sim$ Ellip. Unif

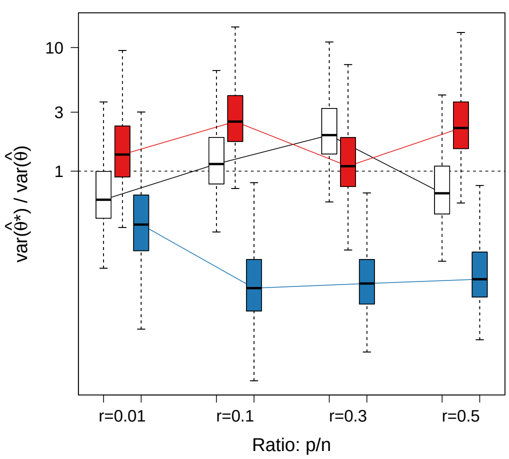
Figure S11: **Gap Ratio Statistic: Bias of Bootstrap.** Note that the Gap Ratio is not well defined in the population for this simulation (since $\lambda_2 = \lambda_3$) so we can not scale the Gap Ratio by the true value of the Gap Ratio as was done for other plots in the Supplementary Figures. Instead we plot the actual bias, as in Figure 1. For this reason, we only show the smaller values of λ_1 (otherwise, without scaling, the plot is dominated by the bias of large values of λ_1, even though the relative value of the bias is small). Similarly, the true bias of the estimate $\hat{\lambda}_1 - \hat{\lambda}_2$ is not a well-defined quantity and hence is not plotted.



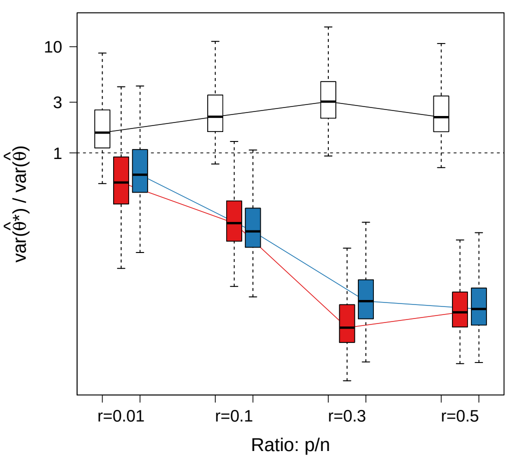
(a) $X_i \sim$ Normal



(b) $X_i \sim$ Ellip. Exp

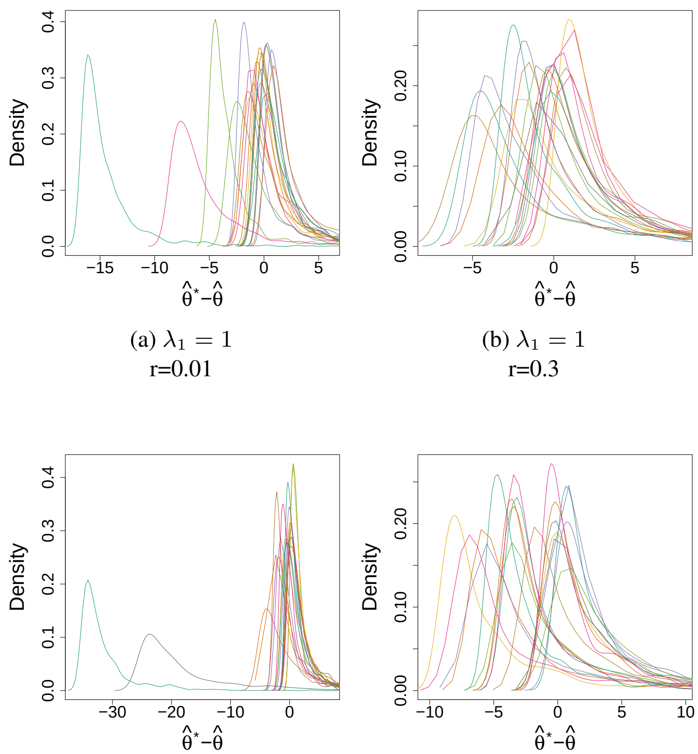


(c) $X_i \sim$ Ellip. Normal



(d) $X_i \sim$ Ellip. Uniform

Figure S12: **Gap Ratio Statistic: Ratio of Bootstrap Estimate of Variance to True Variance.**



(a) $\lambda_1 = 1$
r=0.01

(b) $\lambda_1 = 1$
r=0.3

(c) $\lambda_1 = 1 + 3\sqrt{r}$
r=0.01

(d) $\lambda_1 = 1 + 3\sqrt{r}$
r=0.3

Figure S13: **Gap Ratio Statistic: Bootstrap distribution,** $X_i \sim$ **Ellip Exp, n=1,000:** Plotted are the estimated density of twenty simulations of the bootstrap distribution of $(\hat{\lambda}_1^{*b} - \hat{\lambda}_2^{*b}) - (\hat{\lambda}_1 - \hat{\lambda}_2)$, with $b = 1, \ldots, 999$. The solid black line line represents the distribution of $(\hat{\lambda}_1 - \hat{\lambda}_2) - (\lambda_1 - \lambda_2)$ over 1,000 simulations.

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.08	0.04	0.02	0.01	0.01
$r = 0.1$	0.42	0.23	0.13	0.11	0.10
$r = 0.3$	1.04	0.60	0.37	0.31	0.31
$r = 0.5$	1.70	1.00	0.60	0.52	0.51

(a) Bootstrap Median Estimate of Bias

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.17	0.04	0.02	0.02	0.02
$r = 0.1$	0.70	0.20	0.12	0.12	0.12
$r = 0.3$	1.37	0.48	0.36	0.28	0.38
$r = 0.5$	1.88	0.73	0.55	0.48	0.45

(b) True Bias

Table S1: **Top Eigenvalue: Median value of Bootstrap and True values of Bias, $Z \sim$ Normal** This tables give the median values of the boxplots plotted in Figure 1, as well as the true bias values (*) in the plots. See figure caption for more details.

S5 Supplementary Tables

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.23	0.22	0.12	0.06	0.05
$r = 0.1$	2.25	2.24	1.75	0.67	0.58
$r = 0.3$	6.70	6.78	6.35	2.37	1.88
$r = 0.5$	11.19	11.12	10.91	4.70	3.37

(a) Bootstrap Median Estimate of Bias

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.49	0.27	0.12	0.06	0.01
$r = 0.1$	3.34	2.39	1.10	0.54	0.75
$r = 0.3$	9.17	7.57	4.10	1.92	1.69
$r = 0.5$	14.93	12.76	8.08	3.69	3.45

(b) True Bias

Table S2: **Top Eigenvalue: Median value of Bootstrap and True values of Bias, $Z \sim$ Ellip Exp** This tables give the median values of the boxplots plotted in Figure 1, as well as the true bias values (*) in the plots. See figure caption for more details.

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.16	0.13	0.05	0.03	0.03
$r = 0.1$	1.38	1.33	0.55	0.32	0.30
$r = 0.3$	4.17	4.16	2.30	1.00	0.93
$r = 0.5$	6.95	6.96	4.80	1.72	1.57

(a) Bootstrap Median Estimate of Bias

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.32	0.13	0.05	0.04	0.07
$r = 0.1$	1.65	0.83	0.41	0.35	0.13
$r = 0.3$	4.14	2.52	1.20	0.78	0.92
$r = 0.5$	6.57	4.45	2.01	1.64	1.76

(b) True Bias

Table S3: **Top Eigenvalue: Median value of Bootstrap and True values of Bias, $Z \sim$ Ellip Norm**

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.09	0.05	0.02	0.01	0.01
$r = 0.1$	0.55	0.34	0.17	0.14	0.13
$r = 0.3$	1.56	1.07	0.49	0.40	0.39
$r = 0.5$	2.64	1.96	0.82	0.67	0.65

(a) Bootstrap Median Estimate of Bias

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.20	0.05	0.02	0.00	0.02
$r = 0.1$	0.83	0.27	0.16	0.12	0.22
$r = 0.3$	1.66	0.62	0.45	0.45	0.28
$r = 0.5$	2.33	0.98	0.72	0.63	0.63

(b) True Bias

Table S4: **Top Eigenvalue: Median value of Bootstrap and True values of Bias, $Z \sim$ Ellip Uniform**

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	1.69	1.03	1.04	0.98	1.03
$r = 0.1$	4.35	1.07	1.09	0.96	0.95
$r = 0.3$	19.88	1.48	1.12	1.03	1.06
$r = 0.5$	60.27	2.21	1.16	1.00	1.02

(a) $Z \sim$ Normal

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	1.67	1.64	1.05	0.72	0.74
$r = 0.1$	2.53	2.88	3.13	0.99	0.87
$r = 0.3$	3.25	3.29	4.22	1.34	0.87
$r = 0.5$	3.02	2.96	2.84	1.82	1.05

(b) $Z \sim$ Ellip Exp

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	2.07	1.51	0.98	0.98	0.92
$r = 0.1$	8.98	8.33	2.01	0.96	1.03
$r = 0.3$	10.27	11.37	6.63	1.12	0.95
$r = 0.5$	9.78	10.87	11.76	1.20	1.05

(c) $Z \sim$ Ellip Norm

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	1.76	1.05	1.00	0.97	0.92
$r = 0.1$	7.17	1.51	1.11	1.08	0.99
$r = 0.3$	35.38	4.66	1.21	0.94	1.07
$r = 0.5$	84.48	10.72	1.30	1.04	0.99

(d) $Z \sim$ Ellip Uniform

Table S5: **Top Eigenvalue: Median value of ratio of bootstrap estimate of variance to true variance for** $n = 1000$ This tables give the median values of the boxplots plotted in Figure 1.

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	0.71	0.94	0.94	0.95	0.36	0.92	0.95	0.94	0.95
$r = 0.1$	0.00	0.00	0.78	0.94	0.95	0.00	0.89	0.96	0.94	0.95
$r = 0.3$	0.00	0.00	0.39	0.94	0.96	0.04	0.84	0.96	0.95	0.96
$r = 0.5$	0.00	0.00	0.16	0.90	0.94	1.00	0.73	0.96	0.96	0.95

(a) $Z \sim$ Normal

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	0.30	0.94	0.92	0.92	0.99	0.99	0.95	0.91	0.90
$r = 0.1$	0.00	0.00	0.20	0.95	0.94	1.00	1.00	1.00	0.94	0.93
$r = 0.3$	0.00	0.00	0.00	0.90	0.94	1.00	1.00	1.00	0.95	0.94
$r = 0.5$	0.00	0.00	0.00	0.82	0.95	1.00	1.00	1.00	0.98	0.96

(b) $Z \sim$ Ellip Exp

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	0.39	0.93	0.95	0.94	0.88	0.96	0.94	0.95	0.95
$r = 0.1$	0.00	0.00	0.52	0.94	0.95	1.00	1.00	0.98	0.94	0.94
$r = 0.3$	0.00	0.00	0.01	0.93	0.94	1.00	1.00	1.00	0.95	0.95
$r = 0.5$	0.00	0.00	0.00	0.89	0.94	1.00	1.00	1.00	0.97	0.96

(c) $Z \sim$ Ellip Norm

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	0.62	0.94	0.94	0.94	0.53	0.92	0.95	0.94	0.93
$r = 0.1$	0.00	0.00	0.75	0.96	0.95	0.15	0.88	0.96	0.95	0.95
$r = 0.3$	0.00	0.00	0.33	0.92	0.95	1.00	0.73	0.97	0.94	0.94
$r = 0.5$	0.00	0.00	0.10	0.93	0.94	1.00	0.59	0.96	0.95	0.95

(d) $Z \sim$ Ellip Uniform

Table S6: **Top Eigenvalue: Median value of 95% CI Coverage of true λ_1 for $n = 1000$.** This tables give the percentage of CI intervals (out of 1,000 simulations) that cover the true λ_1 as plotted in Figure 3.

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	0.00	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.00
$r = 0.1$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$r = 0.3$	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
$r = 0.5$	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00

(a) $Z \sim$ Normal

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	0.00	0.00	0.00	0.00	0.99	0.92	0.04	0.00	0.00
$r = 0.1$	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.73	0.00	0.00
$r = 0.3$	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.01	0.00
$r = 0.5$	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.04	0.00

(b) $Z \sim$ Ellip Exp

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	0.00	0.00	0.00	0.00	0.88	0.47	0.00	0.00	0.00
$r = 0.1$	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00
$r = 0.3$	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.21	0.00	0.00
$r = 0.5$	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.77	0.00	0.00

(c) $Z \sim$ Ellip Norm

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	0.00	0.00	0.00	0.00	0.53	0.01	0.00	0.00	0.00
$r = 0.1$	0.00	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00
$r = 0.3$	0.00	0.00	0.00	0.00	0.00	1.00	0.01	0.00	0.00	0.00
$r = 0.5$	0.00	0.00	0.00	0.00	0.00	1.00	0.37	0.00	0.00	0.00

(d) $Z \sim$ Ellip Uniform

Table S7: **Top Eigenvalue: Median value of 95% CI Coverage of null value 1 for $n = 1000$.** This tables give the percentage of CI intervals (out of 1,000 simulations) that cover the value $\lambda_1 = 1$ for different values of the true λ_1.

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.03	-0.02	-0.05	-0.06	-0.06
$r = 0.1$	0.05	-0.17	-0.28	-0.30	-0.31
$r = 0.3$	0.10	-0.41	-0.67	-0.72	-0.73
$r = 0.5$	0.18	-0.61	-1.08	-1.17	-1.18

(a) Bootstrap Median Estimate of Bias

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.05	-0.12	-0.14	-0.14	-0.15
$r = 0.1$	0.04	-0.49	-0.58	-0.58	-0.58
$r = 0.3$	0.05	-0.88	-1.01	-1.08	-0.98
$r = 0.5$	0.05	-1.15	-1.33	-1.40	-1.43

(b) True Bias

Table S8: **Gap Statistic: Median value of Bootstrap and True values of Bias, $Z \sim$ Normal**

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.14	0.13	-0.04	-0.12	-0.14
$r = 0.1$	0.93	0.92	0.40	-1.33	-1.47
$r = 0.3$	2.74	2.76	2.54	-3.50	-4.27
$r = 0.5$	4.56	4.44	4.41	-4.82	-6.76

(a) Bootstrap Median Estimate of Bias

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.19	-0.07	-0.30	-0.37	-0.41
$r = 0.1$	0.85	-0.11	-1.88	-2.63	-2.41
$r = 0.3$	2.32	0.64	-3.51	-6.70	-7.02
$r = 0.5$	3.76	1.71	-3.74	-10.42	-11.33

(b) True Bias

Table S9: **Gap Statistic: Median value of Bootstrap and True values of Bias, $Z \sim$ Ellip Exp** This tables give the median values of the boxplots plotted in Figure 1, as well as the true bias values (*) in the plots. See figure caption for more details.

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.08	0.05	-0.07	-0.10	-0.11
$r = 0.1$	0.44	0.42	-0.66	-0.99	-1.02
$r = 0.3$	1.28	1.29	-1.00	-2.99	-3.10
$r = 0.5$	2.10	2.11	-0.34	-4.94	-5.17

(a) Bootstrap Median Estimate of Bias

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.11	-0.13	-0.24	-0.25	-0.23
$r = 0.1$	0.21	-0.70	-1.22	-1.28	-1.51
$r = 0.3$	0.56	-1.11	-2.83	-3.25	-3.13
$r = 0.5$	0.95	-1.21	-4.35	-4.80	-4.70

(b) True Bias

Table S10: **Gap Statistic: Median value of Bootstrap and True values of Bias, $Z \sim$ Ellip Norm**

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.03	-0.02	-0.06	-0.07	-0.07
$r = 0.1$	0.08	-0.17	-0.37	-0.41	-0.41
$r = 0.3$	0.24	-0.30	-1.04	-1.14	-1.15
$r = 0.5$	0.43	-0.28	-1.77	-1.94	-1.96

(a) Bootstrap Median Estimate of Bias

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	0.06	-0.12	-0.16	-0.17	-0.17
$r = 0.1$	0.05	-0.55	-0.66	-0.70	-0.60
$r = 0.3$	0.06	-1.02	-1.20	-1.20	-1.37
$r = 0.5$	0.07	-1.34	-1.60	-1.70	-1.69

(b) True Bias

Table S11: **Gap Statistic: Median value of Bootstrap and True values of Bias, $Z \sim$ Ellip Uniform**

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	2.05	1.08	1.14	0.98	1.03
$r = 0.1$	4.23	1.23	1.15	0.96	0.96
$r = 0.3$	13.33	1.62	1.25	1.03	1.06
$r = 0.5$	40.50	1.72	1.42	1.02	1.02

(a) $Z \sim$ Normal

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	2.12	2.00	1.02	0.72	0.73
$r = 0.1$	2.73	3.11	2.77	1.11	0.89
$r = 0.3$	3.32	3.68	5.25	1.30	0.94
$r = 0.5$	3.24	3.09	2.91	1.52	1.13

(b) $Z \sim$ Ellip Exp

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	2.67	1.68	1.11	0.99	0.91
$r = 0.1$	12.22	10.60	1.92	1.01	1.02
$r = 0.3$	11.52	14.83	3.22	1.44	0.97
$r = 0.5$	11.07	12.41	5.36	1.67	1.15

(c) $Z \sim$ Ellip Norm

	$\lambda_1 = 1$	$\lambda_1 = 1 + 3\sqrt{r}$	$\lambda_1 = 1 + 11\sqrt{r}$	$\lambda_1 = 1 + 50\sqrt{r}$	$\lambda_1 = 1 + 100\sqrt{r}$
$r = 0.01$	2.19	1.07	1.08	0.97	0.92
$r = 0.1$	8.14	1.31	1.23	1.08	0.99
$r = 0.3$	47.97	2.72	1.67	0.95	1.07
$r = 0.5$	111.75	6.25	2.22	1.09	0.99

(d) $Z \sim$ Ellip Uniform

Table S12: **Gap Statistic: Median value of ratio of bootstrap estimate of variance to true variance for** $n = 1000$ This tables give the median values of the boxplots plotted in SupplementaryFigure S8.

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	0.44	0.53	0.87	0.93	0.90	0.61	0.85	0.91	0.93
$r = 0.1$	0.00	0.00	0.02	0.77	0.90	0.97	0.15	0.69	0.92	0.95
$r = 0.3$	0.00	0.00	0.00	0.71	0.89	1.00	0.14	0.84	0.93	0.96
$r = 0.5$	0.00	0.00	0.00	0.65	0.87	1.00	0.21	0.95	0.95	0.95

(a) $Z \sim$ Normal

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	1.00	0.75	0.81	0.85	1.00	0.79	0.76	0.86	0.87
$r = 0.1$	0.00	1.00	0.92	0.58	0.76	1.00	0.97	0.63	0.88	0.90
$r = 0.3$	0.00	1.00	1.00	0.47	0.62	1.00	1.00	0.75	0.86	0.90
$r = 0.5$	0.00	1.00	1.00	0.45	0.57	1.00	1.00	0.86	0.83	0.90

(b) $Z \sim$ Ellip Exp

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	1.00	0.64	0.86	0.91	0.96	0.59	0.83	0.91	0.93
$r = 0.1$	0.00	1.00	0.12	0.65	0.82	1.00	0.42	0.85	0.94	0.93
$r = 0.3$	0.00	1.00	0.45	0.42	0.72	1.00	0.83	0.81	0.97	0.95
$r = 0.5$	0.00	1.00	0.91	0.31	0.67	1.00	0.97	0.75	0.99	0.97

(c) $Z \sim$ Ellip Norm

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	0.57	0.53	0.87	0.92	0.92	0.61	0.83	0.92	0.93
$r = 0.1$	0.00	0.00	0.02	0.74	0.91	1.00	0.17	0.77	0.93	0.94
$r = 0.3$	0.00	0.00	0.00	0.65	0.87	1.00	0.10	0.98	0.94	0.94
$r = 0.5$	0.00	0.00	0.00	0.56	0.82	1.00	0.16	0.99	0.96	0.95

(d) $Z \sim$ Ellip Uniform

Table S13: **Gap Statistic: Median value of 95% CI Coverage of true Gap for $n = 1000$.** This tables give the percentage of CI intervals (out of 1,000 simulations) that cover the true $\lambda_1 - \lambda_2$ as plotted in Supplementary Figure S8.

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	0.00	0.85	0.00	0.00	0.00	0.00	0.95	0.00	0.00
$r = 0.1$	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00
$r = 0.3$	0.00	0.00	0.00	0.00	0.00	0.00	0.78	0.00	0.00	0.00
$r = 0.5$	0.00	0.01	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00

(a) $Z \sim$ Normal

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.28	0.33	0.87	0.00	0.00	0.07	0.08	0.84	0.00	0.00
$r = 0.1$	1.00	1.00	1.00	0.00	0.00	0.95	0.95	0.96	0.01	0.00
$r = 0.3$	1.00	1.00	0.99	0.00	0.00	1.00	1.00	1.00	0.04	0.00
$r = 0.5$	0.97	0.97	0.92	0.00	0.00	1.00	1.00	1.00	0.10	0.01

(b) $Z \sim$ Ellip Exp

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	0.00	0.79	0.00	0.00	0.00	0.00	0.92	0.00	0.00
$r = 0.1$	1.00	1.00	0.96	0.00	0.00	0.27	0.35	0.07	0.00	0.00
$r = 0.3$	1.00	1.00	0.99	0.00	0.00	1.00	1.00	0.39	0.00	0.00
$r = 0.5$	1.00	1.00	1.00	0.00	0.00	1.00	1.00	0.78	0.00	0.00

(c) $Z \sim$ Ellip Norm

	Percentile					Normal				
	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$	$\lambda_1 = 1$	$\lambda_1 = 3\sqrt{r}$	$\lambda_1 = 11\sqrt{r}$	$\lambda_1 = 50\sqrt{r}$	$\lambda_1 = 100\sqrt{r}$
$r = 0.01$	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.94	0.00	0.00
$r = 0.1$	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00
$r = 0.3$	0.07	0.03	0.00	0.00	0.00	0.00	0.90	0.00	0.00	0.00
$r = 0.5$	1.00	1.00	0.00	0.00	0.00	0.00	0.98	0.00	0.00	0.00

(d) $Z \sim$ Ellip Uniform

Table S14: **Gap Statistic: Median value of 95% CI Coverage of null value 0 for $n = 1000$.** This tables give the percentage of CI intervals (out of 1,000 simulations) that cover the value $\lambda_1 - \lambda_2 = 0$ for different values of the true λ_1.