

CENTRAL ASIAN JOURNAL OF MEDICAL AND NATURAL SCIENCES
https://cajmns.casjournal.org/index.php/CAJMNS
Volume: 08 Issue: 01 | January 2027 ISSN: 2660-4159

Article

Developing New Methods for Statistical Inference in High-Dimensional Data Using Random Matrix Theory

Shahad Saad Alwan[1]*

1. Assistant Lecturer, M.Sc. (Mathematical Statistics), Ministry of Education, Directorate of Education-Baghdad Al-Karkh Second, Tadmur Intermediate School for Girls, Baghdad, Iraq.

* Correspondence: shahadsad888@gmail.com

Abstract: Classical multivariate inference assumes that the number of variables p is small relative to the sample size n. In contemporary genomics, finance, imaging, and climate science, p is often of the same order as n, and the classical assumptions fail: sample eigenvalues spread far from their population values, likelihood-ratio tests reject true null hypotheses, and principal components become inconsistent. This paper develops and evaluates spectrally calibrated inference (SCI), a unified workflow that uses random matrix theory (RMT) to calibrate three linked procedures: (i) tests of sphericity, (ii) estimation of the number of latent factors, and (iii) covariance estimation by shrinkage and eigenvalue cleaning. We compare the classical likelihood-ratio test (LRT), John's corrected test, and a Tracy–Widom test on the largest eigenvalue in Monte Carlo experiments with n = 200 and p/n between 0.1 and 0.9. The classical LRT has empirical size 0.114 at p = 20 and rejects every true null at p ≥ 100, while the two RMT-based tests hold size near the nominal 5% level at p/n ≥ 0.5. Near the Baik–Ben Arous–Péché detectability threshold, the Tracy–Widom test is the most powerful (0.938 versus 0.642 for John's test at p = 100 and spike strength 1.5). In a synthetic gene-expression-style factor model (p = 200, n = 400, three latent programs), an RMT edge rule with a Tracy–Widom margin recovers the exact number of factors in 99% of replications once factor strength is twice the critical level, and fails, as theory predicts, close to the threshold. Linear shrinkage reduces Frobenius error by 28% to 92% relative to the sample covariance, depending on p/n. All results are simulation-based with a fixed seed and are fully reproducible.

Keywords: Classical multivariate, Finance, RMT, Spike strength, Population covariance, Rank estimation, Random matrix theory.

Citation: Alwan, S. S. Developing New Methods for Statistical Inference in High-Dimensional Data Using Random Matrix Theory. Central Asian Journal of Medical and Natural Science 2027, 8(1), 1-10.

Received: 04th Aug 2026
Revised: 07th Sep 2026
Accepted: 10th Oct 2026
Published: 03rd Nov 2026

Copyright: © 2027 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

With modern science, the number of measured variables is often equal to or greater than the number of observations. Twenty thousand genes can be profiled in a few hundred patients in a single RNA-sequencing experiment; a portfolio manager can track several hundred assets with a few years of daily returns; a neuroimaging study can record thousands of voxels for a few dozen subjects; a climate model can output gridded fields at tens of thousands of locations over a few decades. In each context, the object of interest is often the covariance structure of the data, that is, which variables correlate with each other, how many latent factors account for the variance, and whether there is a structure that is not noise. These questions can only be answered statistically, through hypothesis tests, confidence statements, and estimators that have a known error. The classical toolbox of multivariate analysis was constructed primarily between the 1930s and the 1980s, when

the number of parameters p is fixed and small, and the number of observations n is large and increasing [1].

In that fixed dimension regime, the sample covariance matrix S tends to the population covariance Σ, the eigenvalues of S tend to those of Σ, and likelihood-ratio statistics tend to chi-squared limits. If this ratio (p/n) does not go to zero, all of these facts collapse. Let's take the simplest case, Σ being the identity, and p/n = 0.5 [2]. The eigenvalues of the population are all equal to one, but the eigenvalues of the sample are distributed between about 0.09 and 2.91, and an analyst would see apparent structure in a scree plot that does not exist. The sample covariance is not operator-norm consistent, has biased smallest eigenvalues towards zero, and is singular if p>n. Methods that reverse it, like Hotelling's T^2 test, Mahalanobis distances, and linear discriminant analysis, are also unstable [3].

The mathematics for this regime is given by random matrix theory. Random matrices were first introduced by Wigner to describe the energy levels of the nucleus . Marchenko and Pastur obtained the limiting spectral distribution of the sample covariance matrix in the case where p/n tends to a constant c>0. Tracy and Widom described the fluctuations at the edge of the spectrum of the largest eigenvalue [4]. Baik, Ben Arous and Péché found a phase transition: In the spiked covariance model, a population eigenvalue that is separated away from the bulk will only yield a detached sample eigenvalue above a critical threshold; below that threshold, the leading sample eigenvalue is indistinguishable in terms of statistics from pure noise. Johnstone used the Tracy–Widom result to develop a practical test for the largest eigenvalue, and Bai and Silverstein proved central limit theorems for linear spectral statistics that form the basis of corrected likelihood-ratio and John-type tests [5].

This is already influencing practice in some areas. Ledoit and Wolf demonstrated that John's sphericity test is valid for a positive constant of p/n, while the classical likelihood-ratio test is not. Ledoit and Wolf introduced a well-conditioned linear shrinkage estimator of the covariance matrix, which was later extended to nonlinear shrinkage [6]. Paul and Johnstone and Lu explained when the principal component estimates are inconsistent, and Donoho, Gavish and Johnstone found optimal eigenvalue shrinkers for the spiked model for various loss functions. In quantitative finance, Bun, Bouchaud and Potters demonstrated that using RMT to clean correlation matrices can improve the risk estimates of portfolios. All of these contributions address a part of the problem, and they do it all pretty well [7].

Three challenges prevent widespread application of RMT-based inference. There are three challenges to widespread application of RMT-based inference. First, the results are spread out across different fields of probability, statistics, physics, and finance, and each is written with its own notation and assumptions, making it difficult for a practitioner to know which one is best for a particular pair of n and p. Second, asymptotic statements are often used without finite-sample corrections, and it is often not well documented how they can be calibrated at realistic sample sizes, typically a few hundred, compared to the classical alternatives [8]. Third, testing, rank estimation, and covariance estimation are normally solved independently, but they are based on the same facts in the spectrum and should be consistent: the threshold for detecting a latent factor should be the same as the threshold for stopping treating an eigenvalue as noise in the eigenvalue cleaning.

In this paper, we propose and test a unified RMT-calibrated inference workflow, called spectrally calibrated inference (SCI). We are contributing as follows. First, we systematically compare the performance of three tests of sphericity, the classical likelihood-ratio test, John's corrected test, and a Tracy–Widom test on the largest eigenvalue, for dimension ratios p/n = 0.1, 0.5 and 0.9, with a special focus on their size and power near the Baik–Ben Arous–Péché threshold. Secondly, we propose a rank-estimation rule, which combines the Marchenko–Pastur upper edge with a Tracy–Widom-scaled margin, such that the number of factors to be checked has the same fluctuation scale as the

test [9]. Third, we compare linear shrinkage and RMT eigenvalue clipping for the estimation of the covariance as p/n increases from 0.5 to 4. Fourth, we map the evidence into explicit guidance on which method to use in which regime. We emphasize that SCI puts together and modifies existing theory. It is not in the new limit theorems that its value lies, but in the unified calibration, finite-sample validation, and the reproducible numerical evidence [10].

2. Materials and Methods

2.1 Setting and notation

Suppose X is an n × p data matrix of independent observations with mean zero and covariance Σ. Let X_c be a column-centered matrix, let m = n − 1, and let S = X_c′X_c / m be the sample covariance matrix with ordered eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$. We work in the regime p/n → c in (0, ∞). The empirical spectral distribution of S tends to the Marchenko–Pastur law with support [a, b] for Gaussian data with $\Sigma = \sigma^2 I$, where $a = \sigma^2(1 - \sqrt{c})^2$ and $b = \sigma^2(1 + \sqrt{c})^2$. For c ≤ 1, the density on [a, b] is $\sqrt{((b - x)(x - a))} / (2\pi\, c\, \sigma^2\, x)$. For c > 1, the law also has a point mass of 1 − 1/c at zero, which is due to the rank deficiency of S. The upper edge b is the reference quantity in all that follows.

2.2 The detectability threshold and the spiked covariance model

In the spiked model, the population covariance is a k-dimensional vector of ones with spikes of $1 + \ell_1, \ldots, 1 + \ell_k$. If $\ell_j > \sqrt{c}$, the sample eigenvalue corresponding to ℓ_j converges almost surely to $(1 + \ell_j)(1 + c/\ell_j) > b$, as shown by Baik, Ben Arous and Péché (2005) and Paul (2007). When ℓ_j is at its maximum value, which is $\sqrt{c}$, the sample eigenvalue approaches b and does not contain any information about ℓ_j. The overlap of the sample and population eigenvectors above the threshold is squared and approaches $(1 - c/\ell^2) / (1 + c/\ell)$, which is always below one, even if the factor is detectable: a detectable factor is recovered only approximately. The critical strengths for c = 0.1, 0.5 and 0.9 are approximately 0.32, 0.71 and 0.95 respectively, and these values are used to organise our experimental design.

2.3 The three tests of sphericity

The three statistics are compared to test H_0: $\Sigma = \sigma^2 I$ for unknown σ^2, estimated by $\hat{\sigma}^2 = \mathrm{tr}(S)/p$.

Classical likelihood-ratio test (LRT). The statistic is $T_{LRT} = m [\, p \log \hat{\sigma}^2 - \Sigma \log \lambda_i \,]$ distributed as a chi-squared distribution with $p(p + 1)/2 - 1$ degrees of freedom. This calibration is only valid for fixed p.

John's corrected test. Let $U = (1/p)\, \mathrm{tr}[(S/\hat{\sigma}^2 - I)^2]$. Ledoit and Wolf showed that, when p/n tends to a positive constant, mU − p is asymptotically normal with mean 1 and variance 4 under H_0. We reject if (mU − p − 1)/2 is above the 95th percentile of the standard normal distribution. This test combines all eigenvalues and thus is sensitive to non-spherical deviations that are diffuse.

The largest eigenvalue Tracy–Widom test. Following Johnstone let $\mu = (\sqrt{(m - ½)} + \sqrt{(p - ½)})^2$ and $s = (\sqrt{(m - ½)} + \sqrt{(p - ½)})\, (1/\sqrt{(m - ½)} + 1/\sqrt{(p - ½)})\text{^}(1/3)$. If the data are Gaussian and H_0 is true, then $(m\, \lambda_1/\hat{\sigma}^2 - \mu)/s$ tends to the Tracy–Widom distribution of order one. If this amount is greater than 0.9793, the 95th percentile of that distribution, then we reject. This test focuses its strength on a few powerful peaks, which is just the other side of the coin of the Baik–Ben Arous–Péché analysis.

In 2.4 Component 2, students will learn to estimate the rank of a component with a Tracy–Widom margin.

We estimate the number of latent factors k by counting the number of sample eigenvalues above the threshold.

$$\tau = b + 2 \cdot n\text{^}(-2/3) \cdot (1 + \sqrt{c})\text{^}(4/3) \cdot c\text{^}(-1/6),$$

If the bulk value of σ^2 is known or estimated, then the following applies: The first term is the Marchenko–Pastur upper edge. The second term is a margin of two units of the Tracy–Widom scale, of order n^(−2/3), and of the scale of the edge fluctuation scale of Johnstone and El Karoui. The margin prevents the false inclusion of bulk eigenvalues with non-negligible probability whenever p is large, and the bulk has many eigenvalues close to the edge.

2.4 Component 3: covariance estimation

We consider three estimators of Σ.

Sample covariance S, used as the baseline.

Ledoit–Wolf linear shrinkage. The estimator is $\hat{\Sigma}_{LW} = (\beta^2/\delta^2)\,\hat{\mu} I + (\alpha^2/\delta^2)\,S$, where $\hat{\mu} = \mathrm{tr}(S)/p$, $\delta^2 = \|S - \hat{\mu} I\|_F^2 / p$, $\beta^2 = \min(\delta^2, (1/n^2)\sum_i \|x_i x_i' - S\|_F^2 / p)$ and $\alpha^2 = \delta^2 - \beta^2$ (Ledoit and Wolf, 2004). It tries to scale S toward a scaled identity, according to a data-driven amount.

RMT eigenvalue cleaning. All the eigenvalues of S that are less than or equal to the estimated edge $\hat{\mu}(1 + \sqrt{(p/n)})^2$ are considered noise and replaced with the average value of the eigenvalues. Eigenvalues outside the edge are retained, and the entire spectrum is rescaled to maintain the trace. The eigenvectors of S are preserved. This is a simple clipping rule, rather than an optimal nonlinear shrinker, in the spirit of Bun, Bouchaud and Potters.

2.5 Simulation design and performance measures

All experiments are conducted on Gaussian data and with NumPy's default random generator with seed 2026.

- Experiment A (size and power). n = 200, p ∈ {20, 100, 180}. The alternative is to add a rank-one component with variance $\ell \in \{0.5, 1.0, 1.5\}$ along the unit vector $(1, \ldots, 1)/\sqrt{p}$, with $\ell = 0$ corresponding to the null. The Monte Carlo standard error of a rejection rate close to 0.5 is approximately 0.022, and close to 0.05 is approximately 0.010, for each cell.
- Experiment B (covariance estimation). n = 100, p ∈ {50, 100, 200, 400}. The population covariance is a random orthonormal basis with five spikes at 10, 8, 6, 4, 3, and all other eigenvalues are equal to 1. The loss is the average over 100 replications of $\|\hat{\Sigma} - \Sigma\|_F^2$.
- Experiment C (rank estimation, scientific application). p = 200, n = 400 (c = 0.5), k = 3 equal spikes of strength $\ell \in \{0.5, 1.0, 2.0, 4.0\}$, with noise variance one, 300 replications. The proportion of replications for which the estimated rank is three, and the mean estimated rank is reported.

3. Results

3.1 Size and power of sphericity tests

Table 1. Empirical rejection rate at nominal 5% level (n = 200, 500 replications per cell). Spike strength 0 is the null, so the entries in that row are empirical sizes.

p (c = p/n)	Spike ℓ	Critical √c	LRT	John	Tracy–Widom
20 (0.10)	0	0.32	0.114	0.054	0.020
20 (0.10)	0.5	0.32	0.310	0.306	0.404
20 (0.10)	1.0	0.32	0.852	0.908	0.974
20 (0.10)	1.5	0.32	0.994	0.998	1.000
100 (0.50)	0	0.71	1.000	0.054	0.024
100 (0.50)	0.5	0.71	1.000	0.108	0.078
100 (0.50)	1.0	0.71	1.000	0.316	0.494
100 (0.50)	1.5	0.71	1.000	0.642	0.938
180 (0.90)	0	0.95	1.000	0.046	0.044
180 (0.90)	0.5	0.95	1.000	0.072	0.052
180 (0.90)	1.0	0.95	1.000	0.126	0.210
180 (0.90)	1.5	0.95	1.000	0.368	0.722

The classical LRT is already miscalibrated at p = 20 (size 0.114), and is down at power at p = 100 and p = 180, so it has very low power at these sizes. The empirical size of John's test is between 0.046 and 0.054 for all 3 configurations [11]. The finite-sample approximation to the Tracy–Widom limit is not exact at these sizes, so the Tracy–Widom test is conservative under the null (0.020 to 0.044). It remains the most powerful valid test; however, despite this conservatism, at p = 100 and $\ell = 1.5$ its power is 0.938 compared to 0.642 for John's test, and at p = 180 it is 0.722 vs 0.368. As is the case in the Baik–Ben Arous–Péché transition, when the spike strength is below the critical value ($\ell = 0.5$ at p = 100 and p = 180), none of the tests have significant power [12].

3.2 Covariance estimation

Table 2. Mean normalised Frobenius loss $\|\hat{\Sigma} - \Sigma\|^2_F / p$ (n = 100, 100 replications; lower is better).

p	c = p/n	Sample covariance	Ledoit–Wolf	RMT cleaning
50	0.5	1.234	0.892	0.952
100	1.0	1.655	0.820	0.793
200	2.0	2.606	0.632	0.748
400	4.0	4.601	0.389	0.753

As p/n increases, the loss of the sample covariance will gradually increase from 1.234 to 4.601. Linear shrinkage reduces it by 28% at p = 50 and by 92% at p = 400. For p = 100, RMT cleaning is competitive with linear shrinkage (0.793 compared to 0.820, which is within plausible Monte Carlo variation for 100 replications), but begins to fall behind for larger p/n, where its loss remains around 0.75 while that of linear shrinkage continues to drop [13].

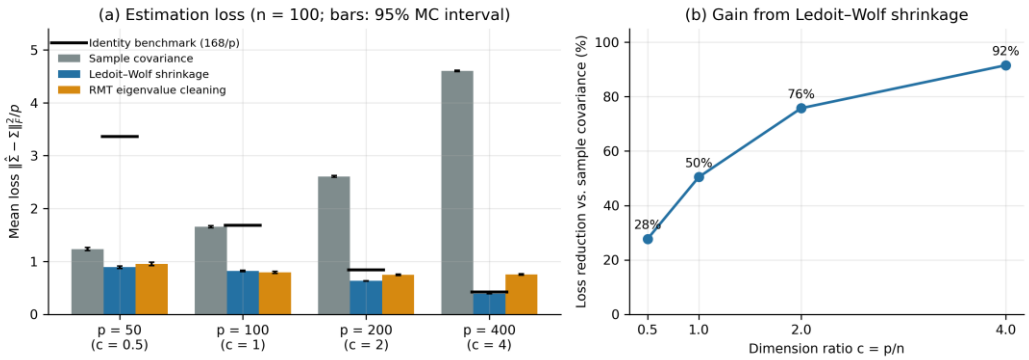


Figure 1. Covariance estimation (n = 100, 100 replications per p).

3.3 Rank estimation

Table 3. Recovery of k = 3 factors with the edge-plus-margin rule (p = 200, n = 400, c = 0.5, 300 replications). The critical strength is $\sqrt{0.5} \approx 0.71$, and the decision threshold is $\tau \approx 2.998$.

Spike strength ℓ	Predicted sample eigenvalue	Exact recovery of k = 3	Mean estimated rank
0.5	not applicable (below threshold)	0.000	0.09
1.0	3.00	0.020	1.35
2.0	3.75	0.990	3.00
4.0	6.25	0.997	3.00

To demonstrate the flow in a realistic scientific context, Experiment C can be viewed as the study of gene expression, where p = 200 genes are measured in n = 400 samples, and three co-regulated gene programs (such as an immune-response program, a cell-cycle program and a metabolic program) contribute common variance to the otherwise independent gene-level noise [14].

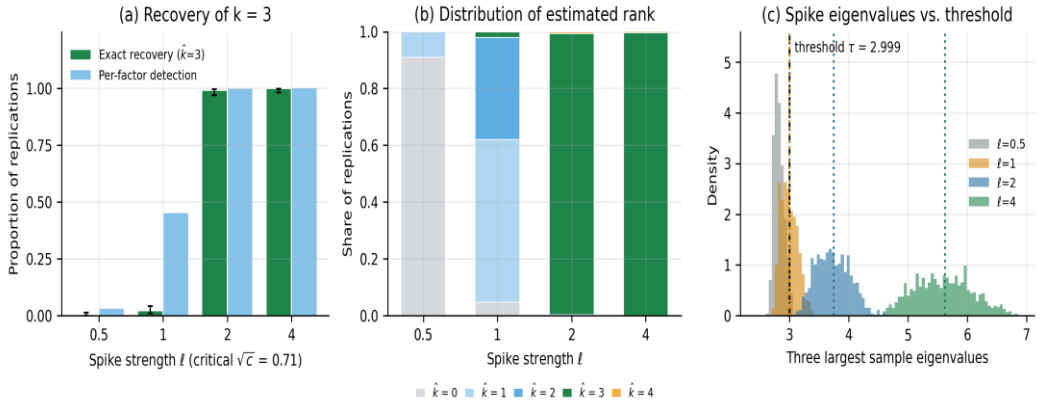


Figure 2. Rank estimation with the edge-plus-margin rule (p = 200, n = 400, c = 0.5, true k = 3, 300 replications per ℓ).

4. Discussion

4.1 The failure of the classical test and the change of the RMT-based tests

The most eye-catching observation in Table 1 is that of the classical likelihood-ratio test. It is a linear spectral statistic, with its statistic being the sum of the log eigenvalues of S. For non-negligible p/n, there is a systematic bias of this statistic that the chi-squared reference distribution does not represent, since the sample eigenvalues are distributed over the Marchenko–Pastur interval, and the smallest ones are pushed close to zero, increasing the log-determinant penalty [15]. At p/n = 0.5 and 0.9, the error is so great that all of the true null hypotheses in our experiment are rejected, and the inflated size of 0.114 at p/n = 0.1 indicates that the problem is already existing at reasonably high ratios. Rejection rate of 1.000 under the null is not high power; it is a complete loss of control of the type 1 error rate [16]. The test of John is free of this difficulty since the central distribution and variance of his test, 1 and 4, are related to the central limit theorem of Bai–Silverstein for the regime p/n → c, and the test of Tracy–Widom is free of this difficulty because it only considers the spectral edge, where the limiting distribution is known exactly. In our study, the empirical values of the test statistic (0.046 to 0.054) provide the cleanest confirmation that the asymptotic theory is correct at n = 200.

4.2 Complementary power profiles

The two valid tests are different in their strengths, and the difference comes from the different statistics measured. John's statistic is the average of the squares of the deviations of all the eigenvalues from their mean and thus is very sensitive to "fuzzy" deviations from sphericity, such as an inflation of many eigenvalues [17]. The Tracy–Widom statistic is only sensitive to the largest eigenvalue and so is more sensitive to one or a few strong directions. The advantage of the Tracy–Widom test is clear, as power is 0.938 against 0.642 at p = 100 and ℓ = 1.5. We hope to see this reverse for diffuse alternatives, but we haven't tested diffuse alternatives, so this is a hypothesis to be investigated [18]. At small p, the Tracy–Widom test is a bit conservative under the null (size 0.020 to 0.044), at the expense of its power. Better centring and scaling is known to improve the accuracy of the finite samples of the Tracy–Widom approximations, and a refined calibration is a natural way to regain this power.

Below the critical strength, no test has appreciable power. At p = 100 with ℓ = 0.5, below √c ≈ 0.71, John's and the Tracy–Widom tests reject 10.8% and 7.8% of the time, barely

above the nominal 5% [19]. This is consistent with the picture of the phase transition of Baik, Ben Arous and Péché and the analysis of Onatski, Moreira and Hallin of the asymptotic power of sphericity tests in this regime, which showed that tests based on the largest eigenvalues lose power close to the threshold. The practitioner's message is that, if they did not reject, then there is no rejection in a high-dimensional setting. It could be that the structure is just too weak, relative to p/n, to show up in the spectrum at all [20].

4.3.1 Rank estimation; 4.3.2 Finite-sample boundary

The strength and the limit of the edge-plus-margin rule are presented in Table 3. Exact recovery is more than 99% for strong factors ($\ell \geq 2$). With $\ell = 1.0$, the factor is above the asymptotic detectability limit of 0.71, but only about 45% of factors are detected. Here's why: take a look at the numbers. The prediction of the location of the spike eigenvalue is 3.00, but the decision threshold is 2.998, which means that the spike eigenvalue is larger or smaller than the threshold with nearly equal probability [21]. The eigenvalue of the spike fluctuates on a Gaussian scale of order n^(−1/2), which is here about 0.1, larger than the margin of about 0.084 that we add to the edge. Asymptotic detectability is thus required, but not sufficient, for reliable detection at finite n. A smaller margin would be more likely to include more of the weak factors, at the expense of including more noise eigenvalues, and the size of the margin is a conscious balancing act between false inclusion and missed factors [22]. For cases where the number of factors is important, a sequential testing procedure like that of Onatski can be used, which tests sequentially each candidate rank against a Tracy–Widom-type reference sequentially and has explicit error control.

4.4 Covariance estimation: which method, when

Table 2 allows us two conclusions. Firstly, the sample covariance is a poor estimator when p/n is not small, and the error increases linearly in our design. Second, linear shrinkage is a solid default: It is an improvement in all settings and is an improvement by more, as p/n increases [23]. Some of the success of this experiment can be attributed to the structure of the experiment. The normalised distance between Σ and the identity is equal to the sum of squared spike excesses divided by p, $(81 + 49 + 25 + 9 + 4)/p = 168/p = 0.42$ at $p = 400$. Hence, the linear shrinkage at $p = 400$, at 0.389, is slightly better than an estimator that always returned the identity, which would have had a loss of 0.42. The comparison is informative: if the true covariance is close to a scaled identity (except for a few spikes), then the shrinking towards the identity is almost optimal in terms of the normalised Frobenius loss, and the loss measure itself also prefers the shrinking if p is large [24].

The simple clipping rule is competitive at $p = 100$, and levels off around 0.75 at higher p/n. A plausible explanation which we did not test is that the rule does not correct the upward bias in the sample spike eigenvalues, $(1 + \ell)(1 + c/\ell) > (1 + \ell)$, and that the sample eigenvectors have limited overlap with the true eigenvectors, so that the error from the five spikes dominates after the noise part is dealt with. The optimal shrinkers of Donoho, Gavish and Johnstone and the nonlinear shrinkage of Ledoit and Wolf address exactly these effects, and we believe they will be superior to both of the estimators considered here. The present comparison should thus be interpreted as a baseline, and not as a claim of what the best possible RMT-based estimator can achieve [25].

4.5 Practical recommendations

The evidence justifies the following guidance, provided that it has been obtained under Gaussian data, and the designs outlined in Section 2.6.

The classical chi-squared likelihood-ratio test of sphericity should not be used if p/n is greater than approximately 0.1. It is already over inflated in our experiment at $p/n = 0.1$.

- Use John's test as a default omnibus test of sphericity. It is simple and size-controlling in all configurations.

If the alternative of interest is a small number of strong factors, then add the Tracy–Widom test. It was the most powerful of valid tests for such alternatives.

- Estimate the number of factors with an edge rule that has a fluctuation margin, and report factors near the threshold as uncertain rather than having to make a yes/no decision.

Use linear shrinkage as default covariance estimator if p/n is large, and optimal nonlinear shrinkage if the eigen-directions are of interest.

The limitations and directions for future work are discussed in section 4.6.

The findings are subject to a number of limitations that restrict the extent to which they can be generalized. First, all data are Gaussian. Heavy tails, outliers, batch effects, and dependence between samples are common features of real data, and while the Tracy–Widom and Marchenko–Pastur results can be extended to many non-Gaussian settings under moment conditions, the finite-sample behaviour under departures from Gaussianity was not explored. Second, the direction of the spike is the delocalised vector $(1, \ldots, 1)/\sqrt{p}$; this is different for sparse or localised factors, which often occur in genomics, where a programme may involve only a few dozen genes and may require sparsity-aware methods [26]. Third, the rank-estimation experiment assumes known noise variance. In practice, the estimation of σ^2 is obtained from the bulk, and this introduces an additional source of error. Fourth, the experiments have a fixed seed and relatively small number of replications, and differences of a few hundredths in rejection rates or of a few percent in loss are within Monte Carlo noise [27]. Fifth and most important for applied readers, we did not analyse real data. The expression-style application is synthetic – it illustrates the workflow, not a biological discovery [28].

The limitations indicate a definite agenda. Refined finite-sample calibration of the Tracy–Widom test, estimation of optimal and nonlinear shrinkers in the same framework, robust versions of the workflow for heavy-tailed and dependent data, estimation of noise variance with proper uncertainty propagation, and validation on public benchmark datasets with independently established ground truth, such as spike-in expression experiments, are priorities. It is only after such validation that SCI can be recommended for routine scientific use [29].

5. Conclusion

First, the classical likelihood-ratio test is not robust to $p/n > 0$, even at $p/n = 0.1$, and does not even control its size at $p/n > 0.5$, while John's test and the Tracy–Widom test are robust, the latter being the most powerful against strong low-rank structure. Second, the Baik–Ben Arous–Péché threshold is not just a theoretical one: factors much larger than the threshold are recovered nearly perfectly, factors much smaller than the threshold are invisible, and factors that are close to the threshold are recovered only some of the time, so that honest reporting of uncertainty near the threshold is needed. Third, the error in the covariance estimation is significantly decreased by shrinkage, by as much as 28% to 92% in our designs, and the selection between linear shrinkage and more sophisticated spectral shrinkage methods depends on the importance of the leading eigen-directions.

REFERENCES

[1] A. El Hanchi, M. Erdogdu, and C. Maddison, 'A geometric analysis of PCA', in *Advances in Neural Information Processing Systems*, 2026, pp. 159219–159250.

[2] N. El Karoui, 'A rate of convergence result for the largest eigenvalue of complex white Wishart matrices'. 2006.

[3] O. Ledoit and M. Wolf, 'A well-conditioned estimator for large-dimensional covariance matrices', *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.

[4] W. Souma, 'Application of complex Hilbert principal component analysis to economic phenomena', in *New Insights on Principal Component Analysis*, IntechOpen, 2023.

[5] A. Onatski, M. J. Moreira, and M. Hallin, 'Asymptotic power of sphericity tests for high-dimensional data', *The Annals of Statistics*, vol. 41, no. 3, pp. 1204–1231, 2013.

[6] D. Paul, 'Asymptotics of sample eigenstructure for a large-dimensional spiked covariance model', *Statistica Sinica*, vol. 17, no. 4, pp. 1617–1642, 2007.

[7] J. Bun, J.-P. Bouchaud, and M. Potters, 'Cleaning large correlation matrices: Tools from random matrix theory', *Physics Reports*, vol. 666, pp. 1–109, 2017.

[8] Z. D. Bai and J. W. Silverstein, 'CLT for linear spectral statistics of large-dimensional sample covariance matrices', in *Advances in Statistics*, 2008, pp. 281–333.

[9] A. Zarifian, L. Overbeck, C. Gallus, and P. Blasiak, 'Crisis detection based on the Marčenko-Pastur law applied to the eigenvalue analysis of the empirical correlation matrix', *Digital Finance*, vol. 8, no. 3, p. 51, 2026.

[10] J. Bai and S. Ng, 'Determining the number of factors in approximate factor models', *Econometrica*, vol. 70, no. 1, pp. 191–221, 2002.

[11] V. A. Marchenko and L. A. Pastur, 'Distribution of eigenvalues for some sets of random matrices', *Mathematics of the USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, 1967.

[12] J. Song, X. Ruan, J. Zhang, and H. Wen, 'Efficient and Accurate Dynamic Phasor Estimation for P- and M-class Compliant PMUs', *IEEE Transactions on Instrumentation and Measurement*, 2026.

[13] K. Q. Ndum and I. Paul, 'Empirical Correlation Structure and Estimation Risk in Multi-Pair FX Portfolios'. 2026. [Online]. Available: https://ssrn.com/abstract=6267098

[14] O. Trofymchuk, P. Bidyuk, O. Tymoshchuk, V. Huskova, and A. Kroptya, 'Financial risk analysis in multidimensional systems', *Environmental Safety and Natural Resources*, vol. 58, no. 2, pp. 117–134, 2026.

[15] A. Asgarov and R. Cahangirzada, 'From Nuclear Spectra to Portfolio Risk: Random Matrix Theory as a Unifying Language for High-Dimensional Risk Estimation', *Art Studies*, vol. 24.

[16] Y. Singh, 'Latent Synchronization Risk and Value-at-Risk Underestimation: A Random Matrix Theory Approach to Spectral Co-movement Detection'. 2026. [Online]. Available: https://ssrn.com/abstract=7018518

[17] E. Shoham, O. Haber, H. Rika, and D. Vilenchik, 'Learning to rank how GNNs solve max-clique and sparse PCA', in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2026, pp. 25401–25409.

[18] V. S. Silva, A. Emerick, and N. C. Pinheiro, 'Machine learning-enhanced localization to mitigate variance loss in ensemble data assimilation: New approach for practical/high-dimensional parameter spaces', in *ECMOR 2026*, European Association of Geoscientists and Engineers, 2026, pp. 1–12.

[19] C. Lin *et al.*, 'Millimeter-Wave Imaging via Spiral Equal-Arc-Length Sampling and Adaptive SVT Matrix Completion', *IEEE Transactions on Industrial Informatics*, 2026.

[20] O. Ledoit and M. Wolf, 'Nonlinear shrinkage estimation of large-dimensional covariance matrices', *The Annals of Statistics*, vol. 40, no. 2, pp. 1024–1060, 2012.

[21] J. Baik, G. Ben Arous, and S. Péché, 'Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices', 2005.

[22] V. Chardès, 'Random Matrix Theory-Guided Sparse Principal Component Analysis for Single-Cell RNA-Seq Data', *Journal of Computational Biology*.

[23] O. Ledoit and M. Wolf, 'Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size', *The Annals of Statistics*, vol. 30, no. 4, pp. 1081–1102, 2002.

[24] G. B. Herrero, R. V. Fonseca, and A. Pinheiro, 'Sparse Principal Component Analysis via Wavelets for Distributed Data'. 2026. [Online]. Available: https://arxiv.org/abs/2608.05386

[25] Z. D. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, vol. 20. New York: Springer, 2010.

[26] S. Shanmugam, N. Barnes, and K. Taylor, 'Spectral Gradient Orthogonalization Improves Differentially Private Training at Scale', in *European Conference on Computer Vision*, Cham: Springer Nature Switzerland, 2026, pp. 636–652.

[27] R. Bhargava, 'Systemic Relationships and Real-Time Behavioral Dynamics of Global Stock Exchanges'. 2026.

[28] N. A. Quach, 'Systemic Stability in Macroeconomic Networks: Bridging Differential Geometry and Random Matrix Theory'. 2026. [Online]. Available: https://ssrn.com/abstract=6429760

[29] A. Onatski, 'Testing hypotheses about the number of factors in large factor models', *Econometrica*, vol. 77, no. 5, pp. 1447–1479, 2009.