Bias-Reduced Moment Estimators of Population Spectral Distribution and Their Applications

Yingli Qin and Weiming Li

Abstract In this paper, we propose a series of bias-reduced moment estimators for the Population Spectral Distribution (PSD) of large covariance matrices, which are fundamentally important for modern high-dimensional statistics. In addition, we derive the limiting distributions of these moment estimators, which are then adopted to test the order of PSDs. The simulation study demonstrates the desirable performance of the order test in conjunction with the proposed moment estimators for the PSD of large covariance matrices.

Keywords Asymptotically normal • Consistency • Covariance matrix • High-dimension • Hypothesis testing • Moment estimator • Population spectral distribution

1 Introduction

Statistical inference concerning large covariance matrices is developing rapidly, due to the wide availability of high-dimensional data from a variety of scientific, economic, and social studies. Some specific structural assumptions about covariance matrices are often considered, e.g., sparsity in terms of population eigenvalues and eigenvectors or sparsity in terms of the entries of covariance matrices. Johnstone [11] proposes that there only exist a fixed number r of population eigenvalues separated from the bulk. In an even more extreme case, Berthet and Rigollet [4] assume $r = 1$ and the covariance matrix can be modeled as $I + \theta \nu \nu^T$, where ν is a unit length sparse vector and $\theta \in \mathbb{R}^+$. Birnbaum et al. [5] propose adaptive estimation of $r \geq 1$ individual leading eigenvectors when the ordered entries of each eigenvector decay rapidly.

Y. Qin
Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1
e-mail: yingli.qin@uwaterloo.ca

W. Li (✉)
Shanghai University of Finance and Economics, Shanghai, China
e-mail: liwm601@gmail.com

S.E. Ahmed (ed.), *Big and Complex Data Analysis*, Contributions to Statistics,
DOI 10.1007/978-3-319-41573-4_6

In high-dimensional framework, where the dimension p and the sample size n are both large, estimating Population Spectral Distribution (PSD) H_p of covariance matrix Σ_p has attracted much attention recently, see [3, 9, 12, 14, 15, 19]. In [15], the estimation is designed for discrete PSDs with finite support. In [9], the proposed method is evaluated by three simple models considered in their simulation study: $\Sigma_p = I_p, H_p = 0.5\delta_1 + 0.5\delta_2$, and a Toeplitz covariance matrix. For the first model, all population eigenvalues are equal to 1, which is a special case of order 1 discrete PSDs, i.e., $H_p = \delta_1$, while the second model is of order 2 (with mass points 1 and 2) and the third is of order p (i.e., continuous PSD as $p \to \infty$).

In this paper, our main contribution is to propose bias-reduced moment estimators for the PSD of large covariance matrices. These moment estimators can be proved to enjoy some desirable theoretical properties. We then adopt the test in [18] in conjunction with the proposed moment estimators to test the order of PSDs.

Specifically, we assume that under the null hypothesis, there are k distinct population eigenvalues $a_1, \ldots, a_k$, and their multiplicities are $p_1, \ldots, p_k$, respectively. Then the PSD H_p can be expressed as

$$H_p = w_1\delta_{a_1} + \cdots + w_k\delta_{a_k}, \tag{1}$$

where $w_i = p_i/p$ and thus $\sum_{i=1}^{k} w_i = 1$. This model has been considered in [3, 12, 14, 15, 19], where the estimation of H_p is developed by assuming the order $k = k_0$ is known. This assumption does not cause any serious problem if the true order k is smaller than k_0, since the model with higher order contains the (smaller) true model. But if $k > k_0$, then any estimation based on $k = k_0$ can surely lead to erroneous result. Another closely related work is [7], in which the authors develop a cross-validation type procedure to estimate the order k. However, their estimators cannot be used to test the order of PSDs because of the lack of asymptotic distributions. Qin and Li [18] consider the following hypotheses to find statistical evidence to support that there are no more than k_0 distinct mass points in H_p.

$$H_0 : k \le k_0 \quad \text{v.s.} \quad H_1 : k > k_0, \quad k_0 \in \mathbb{N}. \tag{2}$$

The rest of the paper is organized as follows. In the next section, we discuss the bias-reduced estimation of moments of PSDs. In Sect. 3, we reformulate the test in [18] with our proposed moment estimators. Section 4 reports simulation results. Concluding remarks are presented in Sect. 5 and proofs of the main theorems are postponed to the last section.

2 Moments of a PSD and Their Bias-Reduced Estimators

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{x}_i \in \mathbb{R}^p$, be a sequence of independent and identically distributed zero mean random vectors with a common population covariance matrix Σ_p. The sample covariance matrix is

$$S_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'.$$

Note that the population mean is assumed to be zero for simplicity, if not, one may replace S_n with its centralized version.

Let H_p be the PSD of Σ_p and F_n be the empirical spectral distribution (ESD) of S_n. Integer moments of H_p and F_n are, respectively, defined as

$$\gamma_k := \int t^k dH_p(t) \quad \text{and} \quad \hat{\beta}_k := \int x^k dF_n(x),$$

$k = 0, 1, 2, \ldots$. Unbiased estimators of γ_k's based on $\hat{\beta}_k$'s under normality are provided in [10, 21]. However, their results are limited to $k \leq 4$. In [3, 12, 13], more general moment estimators are introduced. However, their estimators are biased. Moreover, their asymptotic means and variances have no explicit forms, and are expressed through contour integrals only. In this paper, we present an explicit bias-reduced version of the estimators in [3].

Our main assumptions are listed as follows. These three assumptions are conventional conditions for the central limit theorem of linear spectral statistics, see [1, 2].

Assumption (a) The sample size n and the dimension p both tend to infinity such that $c_n := p/n \to c \in (0, \infty)$.

Assumption (b) There is a doubly infinite array of i.i.d. random variables (w_{ij}), $i, j \geq 1$, satisfying

$$E(w_{11}) = 0, \;\; E(w_{11}^2) = 1, \;\; E(w_{11}^4) < \infty,$$

such that for every given p, n pair, $W_n = (w_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$. Hence, the observed data vectors can be represented as $\mathbf{x}_j = \Sigma_p^{1/2} w_{.j}$ where $w_{.j} = (w_{ij})_{1 \leq i \leq p}$ denotes the jth column of W_n.

Assumption (c) The PSD H_p of Σ_p weakly converges to a probability distribution H, as $p \to \infty$, and the sequence of spectral norms $(||\Sigma_p||)$ is bounded.

Under the assumptions (a)–(c), the ESD F_n converges in distribution to a determinate distribution $F^{c,H}$ [20], called the limiting spectral distribution (LSD), and the moments γ_k and $\hat{\beta}_k$ also converge,

$$\gamma_k \to \tilde{\gamma}_k := \int t^k dH(t) \quad \text{and} \quad \hat{\beta}_k \to \tilde{\beta}_k := \int x^k dF^{c,H}(x).$$

Moreover, these limiting moments $\tilde{\gamma}_k$'s and $\tilde{\beta}_k$'s are linked through a series of recursive formulas [16],

$$\begin{aligned}
\tilde{\gamma}_1 &= \tilde{\beta}_1, \\
\tilde{\gamma}_2 &= \tilde{\beta}_2 - c\tilde{\gamma}_1^2, \\
\tilde{\gamma}_k &= \tilde{\beta}_k - \frac{1}{c}\sum (c\tilde{\gamma}_1)^{i_1}(c\tilde{\gamma}_2)^{i_2}\cdots(c\tilde{\gamma}_{k-1})^{i_{k-1}}\phi(i_1,\ldots,i_{k-1}), \quad k \geq 2,
\end{aligned}$$

where the sum runs over the following partitions of k:

$$(i_1,\ldots,i_{k-1}) : k = i_1 + 2i_2 + \cdots + (k-1)i_{k-1}, \quad i_l \in \mathbb{N},$$

and the coefficient $\phi(i_1,\ldots,i_{k-1}) = k!/[i_1!\cdots i_{k-1}!(k+1-i_1-\cdots i_{k-1})!]$.

Bai et al. [3] just plug $\hat{\beta}_k$'s into these recursive formulas to get the estimators of γ_k's (also estimators of $\tilde{\gamma}_k$'s).

It's obvious that the mapping from $\tilde{\beta}_k$'s to $\tilde{\gamma}_k$'s,

$$g : (\tilde{\beta}_1,\ldots,\tilde{\beta}_k)' \to (\tilde{\gamma}_1,\ldots,\tilde{\gamma}_k)', \tag{3}$$

is one-to-one and its Jacobian matrix $\partial g(\beta)/\partial\beta$ is a lower-triangular matrix with unit determinant. Therefore, the properties of the plug-in estimators are fully determined by those of $\hat{\beta}_k$'s which actually, as estimators of $\tilde{\beta}_k$'s when $H_p = H$ and $c_n = c$, are biased by the order of $O(1/p)$ [1]. In this paper, our main contribution is to correct the bias and propose bias-reduced moment estimators.

Let $q_{s,t}$ be the coefficient of z^t in the Taylor expansion of $(1+z)^{-s}$ at $z = 0$ and define three power series $P(z)$, $Q(z)$, and $R(z)$ as

$$P(z) = -1 - c\sum_{l=1}^{\infty}\tilde{\gamma}_l(-z)^l, \tag{4}$$

$$Q(z) = c\sum_{l=0}^{\infty} q_{3,l}\tilde{\gamma}_{l+2}z^l, \quad R(z) = 1 - c\sum_{l=0}^{\infty} q_{2,l}\tilde{\gamma}_{l+2}z^{l+2}. \tag{5}$$

Let μ_k ($k \geq 1$) be the coefficient of z^{k-2} in the Taylor expansion of function $P^k(z)Q(z)/R(z)$ at $z = 0$. Apparently $\mu_1 = 0$. When calculating μ_k for $k \geq 2$, it's enough to keep the terms of z^l for $l \leq k-2$ in the series P, Q, and R since higher order terms, after taking derivatives of order $k-2$, are all zero at $z = 0$. Therefore, μ_k is a function of $c, \tilde{\gamma}_1,\ldots,\tilde{\gamma}_k$, and thus a function of $c, \tilde{\beta}_1,\ldots,\tilde{\beta}_k$.

It will be shown that μ_k/p is approximately the leading term of the bias contained in $\hat{\beta}_k$, and hence we modify this estimator to

$$\hat{\beta}_k^* = \hat{\beta}_k - \frac{1}{p}\hat{\mu}_k,$$

where $\hat{\mu}_k = \mu_k(c_n, \hat{\beta}_1, \ldots, \hat{\beta}_k)$, $k = 1, 2, \ldots$. The correction can be conducted iteratively by updating $\hat{\mu}_k$ from $\hat{\beta}_k^*$'s to reduce the bias to the order of $o(1/p)$. As a consequence, we obtain bias-reduced estimators of the moments γ_k's, referred to as $\hat{\gamma}_k$'s,

$$(\hat{\gamma}_1, \ldots, \hat{\gamma}_k)' = g(\hat{\beta}_1^*, \ldots, \hat{\beta}_k^*), \tag{6}$$

$k = 1, 2, \ldots$.

Theorem 1 *Suppose that the assumptions* (a)–(c) *hold, then*

(i) *the estimator* $\hat{\gamma}_k$ $(k \geq 1)$ *is strongly consistent, i.e.,*

$$\hat{\gamma}_k - \gamma_k \xrightarrow{a.s.} 0.$$

(ii) *If in addition* $E(w_{11}^4) = 3$*, then*

$$p\,(\hat{\gamma}_1 - \gamma_1, \ldots, \hat{\gamma}_k - \gamma_k)' \xrightarrow{D} N_k(0, \Psi(k)), \tag{7}$$

where $\Psi(k) = ABA'$*,* A *is the Jacobian matrix* $\partial g(\beta)/\partial \beta$ *at* $\beta = (\tilde{\beta}_k)$*, and* $B = (b_{ij})_{1 \leq i,j \leq k}$ *with its entries*

$$b_{ij} = 2\sum_{l=0}^{i-1}(i-l)\alpha_{i,l}\alpha_{j,i+j-l},$$

where $\alpha_{s,t}$ *is the coefficient of* z^t *in the Taylor expansion of* $P^s(z)$*, the sth power of* $P(z)$ *defined in* (4) .

Theorem 1 establishes the consistency and asymptotic normality of the proposed bias-reduced moment estimators $\hat{\gamma}_k$'s. Compared with the estimators in [3], our proposed moment estimators have two main advantages: One is that the limiting mean vector in (7) is zero, which implies that our estimators reduce biases to the order of $o(1/p)$; The other is that the limiting covariance matrix in (7) is explicitly formulated.

3 Test Procedure

Define a $(k+1)\times(k+1)$ Hankel matrix $\Gamma(G,k)$ related to a distribution G,

$$\Gamma(G,k)=\begin{pmatrix} g_0 & g_1 & \cdots & g_k \\ g_1 & g_2 & \cdots & g_{k+1} \\ \vdots & \vdots & & \vdots \\ g_k & g_{k+1} & \cdots & g_{2k} \end{pmatrix},$$

where g_j is the jth moment of G, $j=0,\ldots,2k$. Write $D(k)=\det(\Gamma(H_p,k))$ then, from Proposition 1 in [12], $D(k_0)=0$ if the null hypothesis in (2) holds, otherwise $D(k_0)>0$. On the other hand, from Theorem 1, a plug-in estimator of this determinant, denoted by $\widehat{D}(k_0)$, can be obtained by replacing γ_k in $D(k_0)$ with $\hat{\gamma}_k$, defined in (6), for $k=1,\ldots,2k_0$, i.e.,

$$\widehat{D}(k_0)=\begin{vmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{k_0} \\ \hat{\gamma}_1 & \hat{\gamma}_2 & \cdots & \hat{\gamma}_{k_0+1} \\ \vdots & \vdots & & \vdots \\ \hat{\gamma}_{k_0} & \hat{\gamma}_{k_0+1} & \cdots & \hat{\gamma}_{2k_0} \end{vmatrix}.$$

We may thus reject the null hypothesis if $\widehat{D}(k_0)$ is significantly greater than zero. Applying Theorem 1 and the main theorem in [18], we may immediately derive the asymptotic distribution of $\widehat{D}(k_0)$.

Theorem 2 *Suppose that the assumptions* (a)–(c) *hold, then the statistic* $\widehat{D}(k_0)$ *is asymptotically normal, i.e.,*

$$p\left(\widehat{D}(k_0)-D(k_0)\right)\xrightarrow{D}N(0,\sigma_{k_0}^2),$$

where $\sigma_{k_0}^2=\alpha' V\Omega V'\alpha$ *with* $\alpha=vec(adj(\Gamma(H,k_0)))$*, the vectorization of the adjugate matrix of* $\Gamma(H,k_0)$*. The* $(2k_0+1)\times(2k_0+1)$ *matrix* Ω *consists of the first row and column zero and the remaining submatrix* $\Psi(2k_0)$ *defined in* (7)*, and the* $(k_0+1)^2\times(2k_0+1)$ *matrix* $V=(v_{ij})$ *is a 0-1 matrix with only* $v_{i,a_i}=1$*,* $a_i=i-\lfloor(i-1)/(k_0+1)\rfloor k_0$*,* $i=1,\ldots,(k_0+1)^2$*, where* $\lfloor x\rfloor$ *denotes the greatest integer not exceeding x.*

To present the limiting null distribution and guarantee the consistency of the order test, we need the following assumption:

Assumption (d) The order of H_p is consistent with the order of H, that is, they simultaneously satisfy the null hypothesis or the alternative in (2).

This assumption is a generalized version of the condition that the order of H_p is equal to that of its limit H, which requires the weight parameters $w_i = p_i/p$ of H_p in (1) all converge to some positive constants, which, for example, excludes the spike model $H_p = (1 - 1/p)\delta_1 + (1/p)\delta_a$, for some $a \neq 1$, see [11]. Notice that the order of H_p for their spike model is always 2 but that of H is 1.

From Theorem 1, the unknown parameters involved in the limiting variance $\sigma_{k_0}^2$ are $c, \tilde{\gamma}_1, \ldots, \tilde{\gamma}_{4k_0}$. Under the null hypothesis and Assumption (d), $\tilde{\gamma}_k$ for $k \geq 2k_0$ is a function of $\tilde{\gamma}_1, \tilde{\gamma}_2, \ldots, \tilde{\gamma}_{2k_0-1}$. A numerical algorithm for obtaining $\tilde{\gamma}_k$ from the lower moments is introduced in [12]. Therefore, under the null hypothesis, a strongly consistent estimator of $\sigma_{k_0}^2$ is $\sigma_{k_0}^2(c_n, \hat{\gamma}_1, \ldots, \hat{\gamma}_{2k_0-1})$, denoted by $\hat{\sigma}_{H_0}^2$.

Theorem 3 *Suppose that the assumptions* (a)–(d) *hold then, under the null hypothesis,*

$$\frac{p\widehat{D}(k_0)}{\hat{\sigma}_{H_0}} \xrightarrow{D} N(0, 1),$$

where $\hat{\sigma}_{H_0}$ *is the square root of* $\hat{\sigma}_{H_0}^2$.

Theorem 4 *Suppose that the assumptions* (a)–(d) *hold, then the asymptotic power of the order test tends to 1, as* $(n, p) \to \infty$.

4 Simulation

4.1 Case of Testing for Order Two PSDs

We report on simulations carried out to evaluate the performance of the order test. Samples are drawn from zero mean multivariate normal population $N_p(0, \Sigma)$. The sample size is $n = 100, 200, 300, 400, 500$ and the dimension to sample size ratio is $c = 1, 3, 5, 7$. The number of independent replications is 10,000.

We first examine empirical sizes of the test. The model under the null hypothesis is

$$H_p = w_1\delta_{a_1} + w_2\delta_{a_2},$$

where the distinct mass points are fixed at $(a_1, a_2) = (1, 4)$ and their weights are $(w_1, w_2) = (0.95, 0.05)$, $(0.9, 0.1)$, $(0.8, 0.2)$, and $(0.5, 0.5)$. Results collected in Table 1 show that, when $n = 100$, the empirical sizes are a bit smaller than the targeted nominal level 0.05; as the sample size increases, all empirical sizes approach 0.05.

Table 1 Empirical sizes in percentages of the test for PSDs of order two

	$H_p = 0.95\delta_1 + 0.05\delta_4$				$H_p = 0.9\delta_1 + 0.1\delta_4$			
n	$c = 1$	$c = 3$	$c = 5$	$c = 7$	$c = 1$	$c = 3$	$c = 5$	$c = 7$
100	2.47	3.68	3.82	4.19	3.30	3.81	4.31	4.32
200	4.00	4.70	4.94	4.11	4.85	4.74	5.10	4.14
300	4.35	4.87	4.92	4.55	4.86	4.72	4.89	4.89
400	4.89	4.76	4.73	4.99	5.03	4.99	4.90	5.05
	$H_p = 0.8\delta_1 + 0.2\delta_4$				$H_p = 0.5\delta_1 + 0.5\delta_4$			
n	$c = 1$	$c = 3$	$c = 5$	$c = 7$	$c = 1$	$c = 3$	$c = 5$	$c = 7$
100	3.52	4.30	4.44	4.56	4.57	4.36	4.34	4.25
200	4.92	4.59	4.55	5.02	5.25	4.90	4.68	4.86
300	5.27	5.07	4.51	5.15	4.95	5.33	4.97	5.06
400	5.12	5.50	4.91	4.46	5.12	4.92	4.71	5.18

The dimension to sample size ratio $c = 1, 3, 5, 7$. The nominal significant level is $\alpha = 0.05$ and the number of independent replications is 10,000

We also observe that, for small p and n, the performance of the order test in conjunction with the bias-reduced moment estimators varies slightly when the mixture proportions of H_p change. This is due to the fact that our test statistic is dependent upon the moment estimators of H_p, which are affected by the changing mixture proportions.

Next, we examine the power of the order test. Four models under the alternative hypothesis are employed:

Model 1: $H_p = 0.8\delta_1 + 0.1\delta_4 + 0.1\delta_7$,
Model 2: $H_p = 0.8\delta_1 + 0.1\delta_3 + 0.05\delta_7 + 0.05\delta_{10}$,
Model 3: $H_p = 0.8\delta_1 + 0.2 \cdot U(4, 10)$,
Model 4: $H_p = U(1, 25)$,

where $U(a, b)$ stands for a uniform distribution on the interval $(a, b) \subset \mathbb{R}^+$. The fist two models are discrete PSDs and their orders are, respectively, 3 and 4. Model 3 can be seen as a mixture of a discrete distribution and a continuous one, where 80 % of the population eigenvalues are 1 and the remaining 20 % are drawn from $U(4, 10)$. The last model is completely continuous.

Notice that the test statistic is invariant to orthonormal transformation. Hence, without loss of generality, we set Σ_p to be diagonal. For discrete PSDs, we set the diagonal entries of Σ_p according to the mixture proportions and corresponding distinct mass points, then use this (same) Σ_p for all 10,000 replications; while for continuous PSDs or PSDs with a continuous mixture component, for each of 10,000 replications, we generate a (different) set of diagonal entries for Σ_p accordingly.

Figure 1 exhibits the empirical power for Models 1–4. The results exhibit a trend that the power tends to 1 as the sample size increases, while the power deteriorates as the ratio c increases. This demonstrates that the increased dimension makes the order detection harder to achieve. The power for Model 2 is better than that for Model 1, which can be attributed to the fact that, compared with Model 1, Model

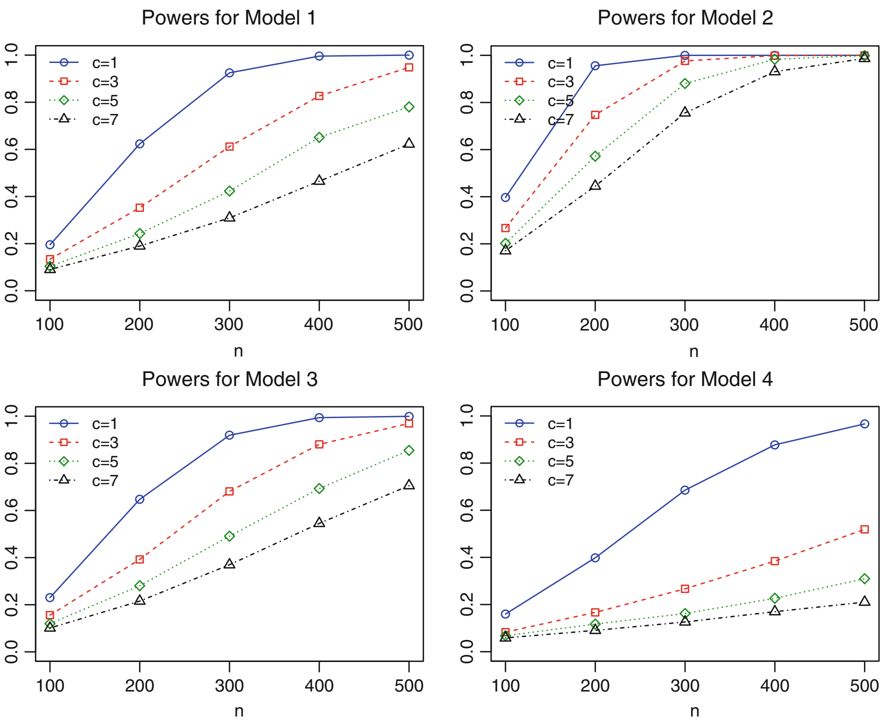

Fig. 1 Empirical powers of the test for Models 1–4 with the dimensional ratio $c = 1, 3, 5, 7$. The nominal significant level is $\alpha = 0.05$ and the number of independent replications is 10,000

2 is further away from the null hypothesis due to the existence of the largest mass point 10. Another phenomena is that the power for the pure continuous model grows slowly compared with the others, although its true order is infinity in the limit, which seems far away from the null hypothesis. A possible reason is that the moment estimators of this continuous PSD have large fluctuations comparing to those of the other discrete PSDs.

4.2 Case of Testing for Order Three PSDs

Qin and Li [18] do not provide simulation results on order three PSDs due to the unavailability of higher order moment estimators. Given the proposed bias-reduced moment estimators in this paper, we will be able to test for any order of PSDs. In this section, we examine the performance of the test for order three hypothesis. Samples are still drawn from zero mean multivariate normal population. The sample size is taken as $n = 300, 400, 500, 600$ and the dimension to sample size ratio is set to be $c = 0.3, 0.6, 0.9, 1.2$. The number of independent replications is 10,000.

Table 2 Empirical sizes in percentages of the test for PSDs of order three

	$H_p = 0.4\delta_1 + 0.4\delta_4 + 0.2\delta_7$				$H_p = 0.4\delta_1 + 0.4\delta_5 + 0.2\delta_{10}$			
n	$c = 0.3$	$c = 0.6$	$c = 0.9$	$c = 1.2$	$c = 0.3$	$c = 0.6$	$c = 0.9$	$c = 1.2$
300	2.78	4.00	4.77	4.09	3.18	4.36	4.94	4.65
400	4.13	4.82	5.00	5.32	4.24	4.91	5.18	5.92
500	4.96	5.52	5.53	5.38	4.59	5.11	5.77	5.79
600	4.92	5.39	5.51	5.82	4.83	5.65	5.51	6.00
	$H_p = 0.5\delta_1 + 0.3\delta_4 + 0.2\delta_7$				$H_p = 0.5\delta_1 + 0.3\delta_5 + 0.2\delta_{10}$			
n	$c = 0.3$	$c = 0.6$	$c = 0.9$	$c = 1.2$	$c = 0.3$	$c = 0.6$	$c = 0.9$	$c = 1.2$
300	3.05	4.56	4.44	4.70	3.37	4.71	5.13	5.45
400	4.39	5.45	5.68	5.73	4.62	5.37	6.07	5.97
500	4.54	5.84	5.90	6.03	4.86	5.72	5.66	6.19
600	5.04	5.68	5.95	5.89	5.46	5.89	6.00	6.15

The dimension to sample size ratio $c = 0.3, 0.6, 0.9, 1.2$. The nominal significant level is $\alpha = 0.05$ and the number of independent replications is 10,000

The model under the null hypothesis is

$$H_p = w_1\delta_{a_1} + w_2\delta_{a_2} + w_3\delta_{a_3},$$

where the distinct mass points are $(a_1, a_2, a_3) = (1, 4, 7), (1, 5, 10)$ and their weights are $(w_1, w_2, w_3) = (0.4, 0.4, 0.2), (0.5, 0.3, 0.2)$. Results in Table 2 show that the empirical sizes are all close to the nominal level, though their fluctuation is a bit larger than that in the test of order two.

Next, we examine the power of the order test using four models under the alternative.

Model 5: $H_p = 0.4\delta_1 + 0.3\delta_5 + 0.2\delta_{15} + 0.1\delta_{25}$,
Model 6: $H_p = 0.4\delta_1 + 0.3\delta_5 + 0.2\delta_{15} + (1/15)\delta_{25} + (1/30)\delta_{30}$,
Model 7: $H_p = 0.4\delta_1 + 0.4\delta_5 + 0.2U(10, 20)$,
Model 8:$H_p = 0.4\delta_1 + 0.3\delta_5 + 0.2\delta_{15} + 0.1U(20, 30)$.

The fist two models are discrete PSDs of orders 4 and 5, respectively, and the last two models are mixture distributions of discrete and continuous. Figure 2 illustrates the power curves for Models 5–8. It shows that this test is more difficult to gain power than the order two test since we need to estimate higher order moments of PSDs. However, we still can see that the power increases along with the increasing (n, p), which again demonstrates the consistency of the test.

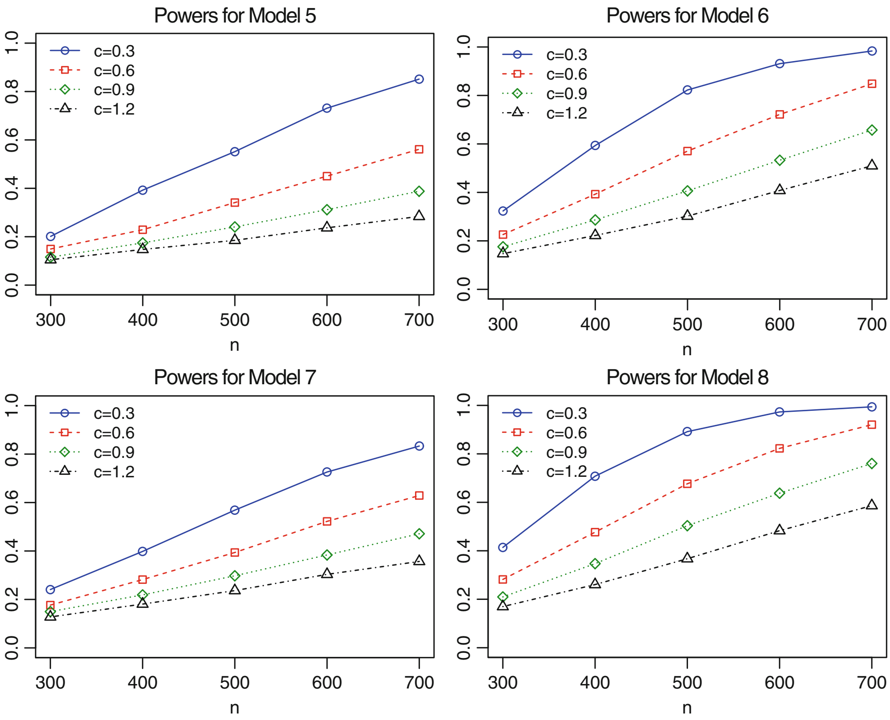

Fig. 2 Empirical powers of the test for Models 5–8 with the dimensional ratio $c = 0.3, 0.6, 0.9, 1.2$. The nominal significant level is $\alpha = 0.05$ and the number of independent replications is 10,000

5 Conclusions and Remarks

In this paper we propose bias-reduced moment estimators of PSDs, which are originally introduced in [3]. The proposed estimators successfully remove all $O(1/p)$ terms in the biases such that the asymptotic normal distributions regain zero mean. We adopt these bias-reduced estimators to a test procedure for the order of PSDs, proposed by Qin and Li [18]. Asymptotic distributions of the test statistic are presented under both the null and the alternative hypotheses as $(n, p) \to \infty$ with their ratio $p/n \to c \in (0, \infty)$. We have observed in the simulation study that the order test maintains desired nominal level and its power tends to 1 as (n, p) tend to infinity.

Recall that unbiased estimators of the first fourth moments of the PSD are given in [10, 21], referred to as $\hat{\gamma}_k^{(u)}, k = 1, 2, 3, 4$. Corresponding estimators in [3] are referred as to $\hat{\gamma}_k^{(b)}$. Some elementary calculations reveal that

$$\hat{\gamma}_1 = \hat{\gamma}_1^{(b)} = \hat{\gamma}_1^{(u)}, \quad \hat{\gamma}_2 = \hat{\gamma}_2^{(b)}\left(1 - \frac{1}{n}\right) = \hat{\gamma}_2^{(u)}\left(1 - \frac{3}{n^2} + \frac{2}{n^3}\right),$$
$$\hat{\gamma}_3 = \hat{\gamma}_3^{(b)}\left(1 - \frac{3}{n}\right) = \hat{\gamma}_3^{(u)}\left(1 - \frac{17}{n^2} + \frac{12}{n^3} + \frac{52}{n^4} - \frac{48}{n^5}\right),$$
$$\hat{\gamma}_4 = \hat{\gamma}_4^{(b)} + O_p\left(\frac{1}{n}\right) = \hat{\gamma}_4^{(u)} + O_p\left(\frac{1}{n^2}\right),$$

from which we can clearly see that these estimators are all asymptotically equivalent, while $\hat{\gamma}_k^{(b)}$ has a bias of order $O(1/p)$ and $\hat{\gamma}_k$ keeps a bias of order $O(1/p^2)$, $k = 2, 3, 4$.

It is worth noticing that the central limiting theorems of all these estimators heavily rely on the moment conditions, say $E(w_{11}^4) = 3$, of the underlying distribution. If the fourth moment is not equal to 3, then there are two additional terms appearing in the limiting mean and covariance matrix, see [17]. Moreover, these two terms are functions of both eigenvalues and eigenvectors of Σ_p (unless Σ_p is diagonal), which are currently hard to be estimated.

6 Proofs

6.1 Proof of Theorem 1

Suppose that the assumptions (a)–(c) hold, from [20], the ESD F_n converges weakly to the LSD $F^{c,H}$, and moreover the Stieltjes transform $s_n(z)$ of the ESD F_n converges almost surely to $s(z)$, the Stieltjes transform of $F^{c,H}$. Let $(\beta_1, \ldots, \beta_k)' = g^{-1}(\gamma_1, \ldots, \gamma_k)$ then,

$$\beta_j = \int t^j dF^{c_n,H_p}(t) \to \tilde{\beta}_j := \int t^j dF^{c,H}(t), \quad j \geq 1,$$

where F^{c_n,H_p} is an LSD derived from $F^{c,H}$ by replacing c and H with c_n and H_p, respectively.

When the support of H is bounded, the support of $F^{c,H}$ is also bounded. Thus, for any $z \in \mathbb{C}$ with $|z|$ large, the Stieltjes transform $s_n(z)$ and $s(z)$ can be expanded as Laurent series, and we have

$$s_n(z) = \int \frac{1}{x - z} dF_n(x) = \sum_{l=0}^{\infty} \frac{-1}{z^{l+1}} \hat{\beta}_l \xrightarrow{a.s.} s(z) = \sum_{l=0}^{\infty} \frac{-1}{z^{l+1}} \tilde{\beta}_l.$$

From this we get $\hat{\beta}_j^* - \beta_j = \hat{\beta}_j - \hat{\mu}_j/p - \beta_j \xrightarrow{a.s.} 0$, and hence

$$\hat{\gamma}_j - \gamma_j \xrightarrow{a.s.} 0, \quad j = 1, 2, \ldots,$$

as $(n, p) \to \infty$, which is the first conclusion.

For the second conclusion, applying Theorem 1.1 in [1] with $f_j(z) = z^j$, $j = 1, \ldots, k$, for real case, we obtain

$$p\left(\hat{\beta}_1 - \beta_1, \ldots, \hat{\beta}_k - \beta_k\right) \xrightarrow{D} N_k(\mu, B),$$

where the mean vector $\mu = (\mu_j)$ with

$$\mu_j = -\frac{1}{2\pi \mathrm{i}} \oint_{C_1} \frac{cz^j \underline{s}^3(z) \int t^2(1 + t\underline{s}(z))^{-3} dH(t)}{(1 - c \int \underline{s}^2(z) t^2 (1 + t\underline{s}(z))^{-2} dH(t))^2} dz, \tag{8}$$

and the covariance $B = (b_{ij})$ with its entries

$$b_{ij} = -\frac{1}{2\pi^2} \oint_{C_2} \oint_{C_1} \frac{z_1^i z_2^j}{(\underline{s}(z_1) - \underline{s}(z_2))^2} \underline{s}'(z_1) \underline{s}'(z_2) dz_1 dz_2, \tag{9}$$

where

$$\underline{s}(z) = -\frac{1-c}{z} + cs(z)$$

is companion Stieltjes transform of $F^{c,H}$ satisfying

$$z = -\frac{1}{\underline{s}(z)} + c \int \frac{t}{1 + t\underline{s}(z)} dH(t). \tag{10}$$

The contours C_1 and C_2 in (8) and (9) are simple, closed, non-overlapping, taken in the positive direction in the complex plane, and each enclosing the support of $F^{c,H}$. Then the second conclusion of this theorem follows from a standard application of the Delta method, and the remaining works are to calculate the contour integrals in (8) and (9).

Without loss of generality, let the contour C_2 enclose C_1 and both of them be away from the support S_F of $F^{c,H}$ such that

$$\max_{t \in S_H, z \in C_i} |t\underline{s}(z)| < 1,$$

where S_H is the support of H. In such a situation, for any $z \in C_1 \cup C_2$,

$$P(\underline{s}(z)) = -1 - c\sum_{l=1}^{\infty}(-\underline{s}(z))^l\tilde{\gamma}_l = -1 + c\int \frac{t\underline{s}(z)}{1+t\underline{s}(z)}dH(t),$$

$$Q(\underline{s}(z)) = c\sum_{l=0}^{\infty} q_{3,l}\tilde{\gamma}_{l+2}\underline{s}^l(z) = c\int \frac{t^2}{(1+t\underline{s}(z))^3}dH(t),$$

$$R(\underline{s}(z)) = 1 - c\sum_{l=0}^{\infty} q_{2,l}\tilde{\gamma}_{l+2}\underline{s}^{l+2}(z) = 1 - c\int \frac{(zt)^2}{(1+tz)^2}dH(t),$$

and from (10) we also get $P(\underline{s}(z)) = z\underline{s}(z)$, where the functions P, Q, and R are defined in (4)–(5). On the other hand, denote the image of C_i under $\underline{s}(z)$ be

$$\underline{s}(C_i) = \{\underline{s}(z) : z \in C_i\}, i = 1, 2.$$

Notice that $\underline{s}(z)$ is a univalent analytic function on $\mathbb{C} \setminus (S_F \cup \{0\})$, and thus C_i and $\underline{s}(C_i)$ are homeomorphic, which implies $\underline{s}(C_1)$ and $\underline{s}(C_2)$ are also simple, closed, and non-overlapping. In addition, from the open mapping theorem and the fact $\underline{s}(z) \to 0$ as $|z| \to \infty$, we may conclude that $\underline{s}(C_2)$ encloses $\underline{s}(C_1)$, and both of them have negative direction and enclose zero.

Based on these knowledge and by the equality

$$\frac{\underline{s}^2(z)}{\underline{s}'(z)} = 1 - c\int \frac{t^2\underline{s}^2(z)}{(1+t\underline{s}(z))^2}dH(t), \tag{11}$$

the integral in (8) becomes

$$\begin{aligned}\mu_j &= -\frac{1}{2\pi \mathrm{i}}\oint_{C_1} \frac{cz^j\underline{s}(z)\underline{s}'(z)\int t^2(1+t\underline{s}(z))^{-3}dH(t)}{1 - c\int \underline{s}^2(z)t^2(1+t\underline{s}(z))^{-2}dH(t)}dz\\ &= \frac{1}{2\pi \mathrm{i}}\oint_{\underline{s}(C_1)} \frac{P^j(\underline{s})Q(\underline{s})}{\underline{s}^{j-1}R(\underline{s})}d\underline{s}\\ &= \begin{cases} 0, & j = 1,\\ \frac{1}{(j-2)!}\left[P^j(z)Q(z)/R(z)\right]^{(j-2)}\Big|_{z=0}, & 2 \le j \le k,\end{cases}\end{aligned} \tag{12}$$

where the equality in (11) is obtained by taking the derivative of z on both sides of the Eq. (10), and the results in (12) are from the Cauchy integral theorem.

Finally, the integral in (9) can be simplified as

$$b_{ij} = -\frac{1}{2\pi^2}\oint_{C_2}\oint_{C_1} \frac{z_1^i z_2^j}{(\underline{s}(z_1) - \underline{s}(z_2))^2}d\underline{s}(z_1)d\underline{s}(z_2)$$

$$= -\frac{1}{2\pi^2} \oint_{\underline{s}(C_2)} \oint_{\underline{s}(C_1)} \frac{P_i(\underline{s}_1)P_j(\underline{s}_2)}{\underline{s}_1^i \underline{s}_2^j (\underline{s}_1 - \underline{s}_2)^2} d\underline{s}_1 d\underline{s}_2$$
$$= -\frac{1}{2\pi^2} \oint_{\underline{s}(C_2)} \frac{P_j(\underline{s}_2)}{\underline{s}_2^j} \left(\oint_{\underline{s}(C_1)} \frac{P_i(\underline{s}_1)}{\underline{s}_1^i (\underline{s}_1 - \underline{s}_2)^2} d\underline{s}_1 \right) d\underline{s}_2.$$

By the Cauchy integral theorem,

$$\oint_{\underline{s}(C_1)} \frac{P_i(\underline{s}_1)}{\underline{s}_1^i (\underline{s}_1 - \underline{s}_2)^2} d\underline{s}_1 = \sum_{l=0}^{i-1} \oint_{\underline{s}(C_1)} \frac{\alpha_{i,l}}{\underline{s}_1^{i-l} (\underline{s}_1 - \underline{s}_2)^2} d\underline{s}_1$$
$$= -2\pi \mathrm{i} \sum_{l=0}^{i-1} \frac{\alpha_{i,l}(i-l)}{\underline{s}_2^{i-l+1}}.$$

From similar arguments, we get

$$b_{ij} = -\frac{1}{\pi \mathrm{i}} \sum_{l=0}^{i-1} (i-l)\alpha_{i,l} \oint_{\underline{s}(C_2)} \frac{P_j(\underline{s}_2)}{\underline{s}_2^{i+j-l+1}} d\underline{s}_2$$
$$= 2 \sum_{l=0}^{i-1} (i-l)\alpha_{i,l}\alpha_{j,i+j-l},$$

$i, j = 1, \ldots, k$.

6.2 *Proof of Theorem 4*

Under the alternative hypothesis and the assumption of this theorem, we have $D(k_0) = \det(\Gamma(H_p, k_0)) \to \det(\Gamma(H, k_0)) > 0$ and

$$\hat{\sigma}_{H_0}^2 \xrightarrow{a.s.} \sigma_{H_0}^2 := \sigma_{k_0}^2(c, \tilde{\gamma}_1, \ldots, \tilde{\gamma}_{2k_0-1}, \gamma_{2k_0}^*, \ldots, \gamma_{4k_0}^*) > 0,$$

as $(n, p) \to \infty$, where γ_k^*, $2k_0 \le k \le 4k_0$, is the kth moment of a discrete random variable with only k_0 different masses, determined by its first $2k_0 - 1$ moments $\tilde{\gamma}_1, \ldots, \tilde{\gamma}_{2k_0-1}$. Therefore, for large p and n, $\hat{\sigma}_{H_0}$ exists and is positive, and

$$P\left(\frac{\widehat{D}(k_0)}{\hat{\sigma}_{H_0}/p} > z_\alpha \right) = P\left(\frac{\widehat{D}(k_0) - D(k_0)}{\sigma_{k_0}/p} > z_\alpha \frac{\hat{\sigma}_{H_0}}{\sigma_{k_0}} - \frac{D(k_0)}{\sigma_{k_0}/p} \right)$$
$$= 1 - \Phi\left(z_\alpha \frac{\sigma_{H_0}}{\sigma_{k_0}} - \frac{\det(\Gamma(H, k_0))}{\sigma_{k_0}/p} \right) + o_p(1)$$
$$\to 1,$$

as $(n, p) \to \infty$, where σ_{k_0} is the square root of $\sigma_{k_0}^2$ defined in Theorem 2 and z_α is the $1-\alpha$ quantile of standard normal population.

Acknowledgements We would like to thank Dr. S. Ejaz Ahmed for organizing this Springer refereed volume. We appreciate his tremendous efforts. Comments by the anonymous referees led to substantial improvement of the manuscript. Yingli Qin's research is partly supported by Research Incentive Fund grant No. 115953 and Natural Sciences and Engineering Research Council of Canada (NSERC) grant No. RGPIN-2016-03890. Weiming Li's research is supported by National Natural Science Foundation of China, No. 11401037 and Program for IRTSHUFE.

References

1. Bai, Z.D., Silverstein, J.W.: CLT for linear spectral statistics of large-dimensional sample covariance matrices. Ann. Probab. **32**, 553–605 (2004)
2. Bai, Z. D., Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices* (2nd ed.). New York: Springer.
3. Bai, Z.D., Chen, J.Q., Yao, J.F.: On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. Aust. N. Z. J. Stat. **52**, 423–437 (2010)
4. Berthet, Q., Rigollet, P.: Optimal detection of sparse principal components in high dimension. Ann. Statist. **41**, 1780–1815 (2013)
5. Birnbaum, A., Johnstone, I.M., Nadler, B., Paul, D.: Minimax bounds for sparse PCA with noisy high-dimensional data. Ann. Statist. **41**, 1055–1084 (2013)
6. Cai, T.T., Ma, Z.M., Wu, Y.H.: Sparse PCA: Optimal rates and adaptive estimation. Ann. Statist. **41**, 3074–3110 (2013)
7. Chen, J.Q., Delyon, B., Yao, J.F.: On a model selection problem from high-dimensional sample covariance matrices. J. Multivar. Anal. **102**, 1388–1398 (2011)
8. Dudoit, S., Fridlyand, J., Speed, T.P.: Comparison of discrimination methods for the classification of tumors using gene expression data. J. Am. Stat. Assoc. **97**, 77–87 (2002)
9. El Karoui, N.: Spectrum estimation for large dimensional covariance matrices using random matrix theory. Ann. Statist. **36**, 2757–2790 (2008)
10. Fisher, T.J., Sun, X., Gallagher, C.M.: A new test for sphericity of the covariance matrix for high dimensional data. J. Multivar. Anal. **101**, 2554–2570 (2010)
11. Johnstone, I.M.: On the distribution of the largest eigenvalue in principal components analysis. Ann. Statist. **29**, 295–327 (2001)
12. Li, W.M., Yao, J.F.: A local moment estimator of the spectrum of a large dimensional covariance matrix. Stat. Sin. **24**, 919–936 (2014)
13. Li, W.M., Yao, J.F.: On generalized expectation-based estimation of a population spectral distribution from high-dimensional data. Ann. Inst. Stat. Math. **67**, 359–373 (2015)
14. Li, W.M., Chen, J.Q., Qin, Y.L., Yao, J.F., Bai, Z.D.: Estimation of the population spectral distribution from a large dimensional sample covariance matrix. J. Stat. Plan. Inference **143**, 1887–1897 (2013)
15. Mestre, X.: Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. IEEE Trans. Inf. Theory **54**, 5113–5129 (2008)
16. Nica, A., Speicher, R.: Lectures on the Combinatorics of Free Probability. Cambridge University Press, New York (2006)
17. Pan, G.M., Zhou, W.: Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. Ann. Appl. Probab. **18**, 1232–1270 (2008)
18. Qin, Y.L., Li, W.M.: Testing the order of a population spectral distribution for high-dimensional data. Comput. Stat. Data Anal. **95**, 75–82 (2016)

19. Rao, N.R., Mingo, J.A., Speicher, R., Edelman, A.: Statistical eigen-inference from large Wishart matrices. Ann. Statist. **36**, 2850–2885 (2008)
20. Silverstein, J.W.: Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. J. Multivar. Anal. **55**, 331–339 (1995)
21. Srivastava, M.S.: Some tests concerning the covariance matrix in high dimensional data. J. Japan Stat. Soc. **35**, 251–272 (2005)