UNIVERSITY OF CAMBRIDGE

Faculty of Economics

Cambridge Working Paper Economics

Cambridge Working Paper Economics: 1807

LOCAL ASYMPTOTIC NORMALITY OF THE SPECTRUM OF HIGH-DIMENSIONAL SPIKED F-RATIOS

Prathapasinghe Dharmawansa	Iain M. Johnstone	Alexei Onatski
(University of Moratuwa)	(Stanford University)	(University of Cambridge)

25 January 2018

We consider two types of spiked multivariate F distributions: a scaled distribution with the scale matrix equal to a rank-k perturbation of the identity, and a distribution with trivial scale, but rank-*k* non-centrality. The eigenvalues of the rank-*k* matrix (*spikes*) parameterize the joint distribution of the eigenvalues of the corresponding F matrix. We show that, for the *spikes* located above a *phase transition* threshold, the asymptotic behavior of the log ratio of the joint density of the eigenvalues of the F matrix to their joint density under a local deviation from these values depends only on the *k* of the largest eigenvalues {code} Furthermore, we show that {code} are asymptotically jointly normal, and the *statistical experiment* of observing all the eigenvalues of the F matrix converges in the Le Cam sense to a Gaussian shift experiment that depends on the asymptotic means and variances of {code} . In particular, the best statistical inference about sufficiently large *spikes* in the local asymptotic regime is based on the *k* of the largest eigenvalues only.

Submitted to the Annals of Statistics

LOCAL ASYMPTOTIC NORMALITY OF THE SPECTRUM OF HIGH-DIMENSIONAL SPIKED F-RATIOS

BY PRATHAPASINGHE DHARMAWANSA, IAIN M. JOHNSTONE AND ALEXEI ONATSKI

We consider two types of *spiked* multivariate F distributions: a scaled distribution with the scale matrix equal to a rank-k perturbation of the identity, and a distribution with trivial scale, but rank-k non-centrality. The eigenvalues of the rank-k matrix (*spikes*) parameterize the joint distribution of the eigenvalues of the corresponding F matrix. We show that, for the *spikes* located above a *phase transition* threshold, the asymptotic behavior of the log ratio of the joint density of the eigenvalues of the F matrix to their joint density under a local deviation from these values depends only on the k of the largest eigenvalues $\lambda_1, ..., \lambda_k$. Furthermore, we show that $\lambda_1, ..., \lambda_k$ are asymptotically jointly normal, and the *statistical experiment* of observing all the eigenvalues of the F matrix converges in the Le Cam sense to a *Gaussian shift experiment* that depends on the asymptotic means and variances of $\lambda_1, ..., \lambda_k$. In particular, the best statistical inference about sufficiently large *spikes* in the local asymptotic regime is based on the k of the largest eigenvalues only.

KEY WORDS: Spiked F-ratio, Local Asymptotic Normality, multivariate F distribution, phase transition, super-critical regime, asymptotic normality of eigenvalues, limits of statistical experiments.

1. Introduction.

The roots of the equation

$$\det(\mathbf{H} - \lambda \mathbf{E}) = 0, \tag{1}$$

or equivalently the eigenvalues of the $\mathbf{F}$-ratio $\mathbf{E}^{-1}\mathbf{H}$, where matrices $\mathbf{H}$ and $\mathbf{E}$ are the 'hypothesis' and 'error' sums of squares, are fundamental for the multivariate statistics. They form the basis for many invariant tests, including the classical tests of the equality of two covariance matrices and of the linear hypotheses in the multivariate regression. In this paper, we study the behavior of these roots when the $\mathbf{F}$-ratio matrix is high dimensional as is often the case in the contemporaneous statistical applications.

We assume that under the null, both $\mathbf{H}$ and $\mathbf{E}$ are central Wisharts, whereas under the alternative, the 'hypothesis' sum of squares matrix $\mathbf{H}$ contains a low-rank structure. This structure is revealed either in a low-rank difference between the covariance parameters of $\mathbf{H}$ and $\mathbf{E}$, or in a low-rank non-centrality in $\mathbf{H}$. The former corresponds to testing the equality of two covariance matrices, whereas the latter corresponds to testing linear hypotheses in multivariate regression.

We call the eigenvalues of the low-rank difference between the parameters of $\mathbf{H}$ and $\mathbf{E}$ *the spikes*. Spiked models have attracted much recent research attention. They were introduced in Johnstone (2001) as a useful abstraction capturing the fact that the high-dimensional sample variation often concentrates along a small number of distinct directions.

In the case of testing the equality of two covariance matrices, these directions may correspond to a few signals that are present only in one of the two samples. We will refer to this as *the signal detection* case (SigD). In the regression context (REG), an example of a low-rank alternative would be a one-way MANOVA with unequal group means that belong to the same low dimensional hyperplane. Another regression example is the structural break in the number of factors in mean, with a small number of additional factors potentially born by a break event which splits the sample.

The focus of this paper is on the $\mathbf{F}$-ratios of high dimensionality p. We consider the asymptotic regime where p goes to infinity proportionally to the 'sample sizes' represented by the 'hypothesis' (matrix $\mathbf{H}$) and 'error' (matrix $\mathbf{E}$) degrees of freedom (d.f.). Our main results can be summarized as follows.

First, we establish a *phase transition* threshold such that if the spikes are below it, or *sub-critical*, then any finite number of the largest eigenvalues of the $\mathbf{F}$-ratio almost surely (a.s.) converge to the upper boundary of the support of the limiting spectral distribution of $\mathbf{E}^{-1}\mathbf{H}$, derived by Wachter (1980). In contrast, when m of the spikes are *super-critical*, the m of the largest eigenvalues of the $\mathbf{F}$-ratio a.s. converge to locations strictly above the upper boundary of the Wachter distribution. The threshold turns out to be the same for SigD and REG cases.

Second, we prove the *joint asymptotic normality* of the m of the largest eigenvalues of the $\mathbf{F}$-ratio that correspond to the *super-critical* spikes. We derive explicit formulas for the asymptotic means and variances. In both SigD and REG cases, the asymptotic variance is highly sensitive to the ratio of the dimensionality to the 'error d.f.'. Even small non-zero values of this ratio lead to very substantial variance increases.

Third, and most important, we establish quadratic asymptotic approximations to the *likelihood ratios* corresponding to *local alternatives* for k super-critical spikes. We find that the approximations depend only on the k largest eigenvalues of the $\mathbf{F}$-ratio, and that the statistical experiment of observing all the eigenvalues is *Locally Asymptotically Normal* (LAN). The limiting experiment is a simple k-dimensional *Gaussian shift*. This result implies that the *asymptotically optimal inference* on the k *super-critical* spikes can be based exclusively on the k largest eigenvalues of the $\mathbf{F}$-ratio.

We conduct a small-scale Monte Carlo experiment to assess the quality of the LAN confidence sets for super-critical spikes. The experiment shows that the coverage rate of the sets is very close to the nominal one. Moreover, surprisingly, the coverage rate remains good even for low dimensional data. **Iain: I thought it would be nice to add the Monte Carlo results from your second to the last Lumini slide. I have added some experiments and removed reference on bootstrap intervals because I thought that the consensus is that they do not work.**

The previous literature on the eigenvalues of $\mathbf{F}$-ratios is vast and old. The finite sample null distribution of the eigenvalues was independently derived by Fisher, Girshick, Hsu, Roy, and Mood in 1939 (see Wilks (1962) for citations). The non-null distributions were classified in James (1964). There have been many subsequent finite sample research papers. These papers are typically motivated by the fact that the power of various tests in MANOVA context depend on the population non-centrality of an $\mathbf{F}$-ratio. To choose between the available tests, one can use the eigenvalues of the $\mathbf{F}$-ratio constructed from a preliminary sample to estimate the non-centrality (see Leung and Muirhead (1987)). As another motivation, Sheena et al (2004) cite the need for estimating the non-centrality of an $\mathbf{F}$-ratio in constructing modified model selection criteria.

In the context of high dimensional data, much recent research focuses on the eigenvalues of sample covariance matrices, which can be viewed as degenerate $\mathbf{F}$-ratios with $\mathbf{E} = I_p$. Baik et al (2005) derive the asymptotic distributions of a few of the largest eigenvalues of *complex Wisharts*. Paul (2007) establishes the asymptotic normality of the fluctuations of a few of the largest eigenvalues of *real Wisharts* in the *super-critical* case. Féral and Péché (2009), Benaych-Georges et al (2011) and Bao et al (2014) show that the fluctuations in the *sub-critical* real case have the *Tracy-Widom distribution*, while Mo (2012) and Bloemendal and Viràg (2011, 2013) establish the asymptotic distribution of a different type in the critical regime.

In a setting of two independent and not necessarily normal samples with different covariances (SigD case), the *phase transition* phenomenon has been studied in Nadakuditi and Silverstein (2010). They obtain a formula for the threshold and establish the a.s. limits of the largest eigenvalues corresponding to the super-critical spikes. The asymptotic distribution of the eigenvalues is described in their paper as an open problem. Our paper solves this problem for the case of two normal samples, including the REG case, which was not covered by Nadakuditi and Silverstein (2010).

We expect that our asymptotic normality results can be extended to the **F**-ratios constructed from non-normal samples. In the one-sample case, an extension of Paul's (2007) asymptotic normality results has been done in Bai and Yao (2008). For non-degenerate **F**-ratios, our asymptotic normality result for SigD case has been recently extended by Wang and Yao (2015). **Iain: Bai and Yao's theorem is incorrect, Wang and Yao's proofs are incomplete. How should we handle this?**

The focus of this paper on normal data is dictated by our main goal: establishing the LAN property for the eigenvalues of the **F**-ratio. To reach this goal, we derive an asymptotic approximation to a log likelihood process by representing it in the form of a multiple contour integral, and applying the *Laplace approximation* method. The explicit form of the joint distribution of the eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$ is known only in the normal case, and we need such an explicit form for our analysis.

A decision-theoretic approach to the finite sample estimation of the eigenvalues of the "ratio" of the population covariances of **H** and **E**, or the eigenvalues of the non-centrality parameter of **H** was taken in many previous studies (see Sheena et al (2004), Bilodeau and Srivastava (1992), and references therein). In one of the first such studies, Muirhead and Verathaworn (1985) explain that the ideal decision-theoretic approach that directly analyzes expected loss with respect to the joint distribution of the eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$ "does not seem feasible due primarily to the complexity of the distribution of the ordered latent roots..." Instead, they focus on deriving an optimal estimator from a particular class.

The proportional asymptotics used in this paper preserves a salient feature of the finite sample, by making the dimensionality of the data non-negligible relative to the sample size. From this perspective, our LAN result can be viewed as an asymptotic implementation of the ideal decision-theoretic approach to the finite sample estimation. We overcome the complexity of the joint distribution of the eigenvalues by using a tractable multiple contour integral representation of the log likelihood process, which follows from the multiple contour integral representation of hypergeometric functions of two matrix arguments, established in Onatski (2013), Dharmawansa and Johnstone (2014), and Passemier et al (2014).

The LAN result of this paper stays in sharp contrast to the asymptotic behavior of the likelihood ratio in the *sub-critical* regime. In a separate paper, we show that the statistical experiment of observing the eigenvalues of an **F**-ratio with a single sub-critical spike is not LAN. The corresponding likelihood ratio depends only on a smooth functional of the empirical distribution of *all* the eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$, so that asymptotically optimal inference about the spike may ignore information contained in the largest eigenvalue. This is totally different from what happens in the super-critical regime, as our LAN result implies that the asymptotically optimal inference about super-critical spikes can be based on the corresponding largest eigenvalues only.

The rest of the paper is structured as follows. In the next section, we describe our setting. In Section 3, we explore the phase transition and derive the a.s. limits of the super-critical eigenvalues. In Section 4, we establish the asymptotic normality of the super-critical eigenvalues. In Section 5, we derive an asymptotic approximation to the joint distribution of all the eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$ for the case of k *super-critical* spikes. In Section 6, we show that the likelihood ratio in the local parameter space is asymptotically equivalent to a linear combination of k of the largest eigenvalues, and establish the LAN property. Section 7 concludes.

2. Setup.

Suppose that

$$(n_1+k)\mathbf{H} \sim W_p\left(n_1+k, \Sigma_1, \Omega_1\right) \quad \text{ and } \quad n_2\mathbf{E} \sim W_p\left(n_2, \Sigma_2\right)$$

are independent non-central and central Wishart matrices respectively. For the non-centrality parameter Ω_1, we use a symmetric version of the definition in Muirhead (1982, p. 442). That is, if Z is an $n \times p$ matrix distributed as $N\left(M, I_n \otimes \Sigma\right)$, then $Z'Z \sim W_p\left(n, \Sigma, \Omega\right)$ with the non-centrality parameter $\Omega = \Sigma^{-1/2}M'M\Sigma^{-1/2}$. We are interested in the eigenvalues $\lambda_{p1} \geq ... \geq \lambda_{pp}$ of $\mathbf{F} \equiv \mathbf{E}^{-1}\mathbf{H}$.

In what follows, we will assume that $\Sigma_2 = I_p$. This assumption is without loss of generality because the eigenvalues of $\mathbf{F}$ do not change under the transformation $\mathbf{H} \mapsto \Sigma_2^{-1/2}\mathbf{H}\Sigma_2^{-1/2}$, $\mathbf{E} \mapsto \Sigma_2^{-1/2}\mathbf{E}\Sigma_2^{-1/2}$. We will consider two different settings for the parameters Σ_1 and Ω_1.

1. **Spiked covariance (SigD):** $\Sigma_1 = I_p + LsL'$ and $\Omega_1 = 0$, where $s = \mathrm{diag}\left\{s_1, ..., s_k\right\}$ with $s_1 > ... > s_k > 0$ is the diagonal matrix of the *covariance spikes*, and L is a $p \times k$ matrix with orthogonal columns, which consists of nuisance parameters.
2. **Spiked non-centrality (REG):** $\Sigma_1 = I_p$ and $\Omega_1 = n_H LsL'$ with $n_H = n_1 + k$, where L and s are as defined above, but the diagonal elements of s are interpreted as *non-centrality spikes*.

Is is convenient to think of $\mathbf{H}$ as the sample covariance matrix $\mathbf{X}\mathbf{X}'/n_H$ of a sample $\mathbf{X}$ having the factor structure

$$\mathbf{X} = LF' + \varepsilon \tag{2}$$

with L, F, and ε playing the roles of the normalized factor loadings, factors, and idiosyncratic terms, respectively. Matrices F and ε are mutually independent, and independent from $\mathbf{E}$. The entries of ε are i.i.d. standard normals, and the distribution of F depends on the setting. For SigD, $F \sim N\left(0, I_{n_H} \otimes s\right)$, whereas for REG, F is a deterministic matrix such that $F'F/n_H = s$. With this interpretation, SigD and REG describe, respectively, distributions of $\mathbf{H}$ which are unconditional and conditional on the factors. In both cases the spike parameters s_j measure the j-th factor's variability or 'strength'.

Let us introduce a convenient representation for the eigenvalues of $\mathbf{F}$. First, note that these eigenvalues are invariant with respect to the simultaneous transformations

$$\mathbf{X} \mapsto U\mathbf{X}V \equiv X \quad \text{and} \quad \mathbf{E} \mapsto U\mathbf{E}U' \equiv E, \tag{3}$$

where U is a random matrix uniformly distributed over the orthogonal group $\mathcal{O}(p)$, and $V \in \mathcal{O}(n_H)$ is such that the submatrix of its first k columns equals $F\left(F'F\right)^{-1/2}$.

Note that UL can be represented as

$$UL = v(v'v)^{-1/2} \equiv vW_v^{-1/2},$$

where v is a $p \times k$ matrix with i.i.d. standard normal entries, and $W_v \sim W_k\left(p, I_k\right)$. Furthermore,

$$X = [vW_v^{-1/2}s^{1/2}W_F^{1/2}, 0] + \epsilon,$$

where v, ϵ, and W_F are mutually independent, the entries of ϵ are i.i.d. standard normals, and the distribution of W_F depends on the setting. For SigD, $W_F \sim W_k\left(n_H, I_k\right)$, whereas for REG, $W_F = n_H I_k$.

Let us denote the submatrix of the first k columns of ϵ as u. Then

$$XX' = \xi\xi' + n_1 H, \tag{4}$$

where $n_1 H \sim W_p\left(n_1, I_p\right)$, H and $\xi\xi'$ are mutually independent, and independent from E, and

$$\xi = vW_v^{-1/2}s^{1/2}W_F^{1/2} + u. \tag{5}$$

Using (3) and (4), we obtain the following convenient representation for the eigenvalues $\lambda_{p1} \geq ... \geq \lambda_{pp}$ of $\mathbf{F}$. Let $\hat{\lambda}_{p1} \geq ... \geq \hat{\lambda}_{pp}$ be the roots of the equation

$$\det\left(\xi\xi'/n_1 + H - xE\right) = 0. \tag{6}$$

Then

$$\lambda_{pj} = n_1 \hat{\lambda}_{pj} / \left(n_1 + k\right). \tag{7}$$

This representation is convenient because the roots of (6) can be viewed and analyzed as perturbations of the roots of equation $\det(H - xE) = 0$ caused by adding the low-rank matrix $\xi\xi'/n_1$ to H. Here ξ, H, and E are independent and

$$n_1 H \sim W_p(n_1, I_p), \qquad n_2 E \sim W_p(n_2, I_p).$$

If $x \in \mathbb{R}$ is such that $H - xE$ is invertible, then

$$\left(\xi\xi'/n_1 + H - xE\right)^{-1} = Q - Q\xi\left(I_k + \xi' Q \xi / n_1\right)^{-1} \xi' Q / n_1,$$

where $Q \equiv (H - xE)^{-1}$. Therefore, if x is a root of the equation

$$\det\left(I_k + \xi'\left(H - xE\right)^{-1} \xi / n_1\right) = 0, \tag{8}$$

then it also solves (6), and hence, behavior of the roots of (6) can be inferred from that of the random matrix-valued function

$$M(x) = \xi'\left(H - xE\right)^{-1} \xi / n_1. \tag{9}$$

This is the main idea of the analysis in the next section.

3. Phase transition and almost sure limits. We will consider the proportional asymptotic regime where n_1, n_2, and p diverge to infinity so that

$$c_1 \equiv p/n_1 \to \gamma_1 \text{ and } c_2 \equiv p/n_2 \to \gamma_2 \text{ with } \gamma_j \in (0,1).$$

Let $n = (n_1, n_2)$ and $\gamma = (\gamma_1, \gamma_2)$. We will abbreviate the above asymptotics as $p, n \to_\gamma \infty$.

As follows from Wachter's (1980) work (see also Yin et al. (1983) and Silverstein (1985)), as $p, n \to_\gamma \infty$, the empirical distribution of the eigenvalues of $E^{-1}H$ converges in probability to the distribution with density

$$\frac{1-\gamma_2}{2\pi} \frac{\sqrt{\left(\beta_+ - \lambda\right)\left(\lambda - \beta_-\right)}}{\lambda\left(\gamma_1 + \gamma_2 \lambda\right)} \mathbf{1}\left\{\beta_- \le \lambda \le \beta_+\right\}. \tag{10}$$

The upper and the lower boundaries of the support of this density are

$$\beta_\pm = \left(\frac{1 \pm \rho}{1 - \gamma_2}\right)^2, \text{ where } \rho = \sqrt{\gamma_1 + \gamma_2 - \gamma_1\gamma_2}.$$

The results of Silverstein and Bai (1995) and Silverstein (1995) show that the empirical distribution converges not only in probability, but also a.s. Furthermore, a simple extension of Theorem 1.1 of Bai and Silverstein (1998) that covers random E (Lemma SM1 in the Supplementary Material) implies that the largest eigenvalue of $E^{-1}H$ a.s. converges to β_+.

The latter convergence, together with (7) and Weyl's inequalities for the eigenvalues of a sum of two Hermitian matrices (see Theorem 4.3.7 in Horn and Johnson (1985)), imply that the $k+1$-th largest eigenvalue of $\mathbf{F}$, $\lambda_{p,k+1}$, a.s. converges to β_+. Those of the k largest eigenvalues that remain separated from β_+ as $p, n \to_\gamma \infty$, must correspond to solutions of (8). Below, we study these solutions in detail. The Supplementary Material (SM) contains proofs of the following three auxiliary lemmas.

LEMMA 1. *For any $x > \beta_+$, as $p, n \to_\gamma \infty$,*

$$\operatorname{tr}(H - xE)^{-1}/p \overset{a.s.}{\to} m(0;x) \text{ and} \tag{11}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\operatorname{tr}(H - xE)^{-1}/p \overset{a.s.}{\to} \frac{\mathrm{d}}{\mathrm{d}x}m(0;x), \tag{12}$$

where $m(0;x) = \lim_{z\to 0} m(z;x)$, and $m \equiv m(z;x) \in \mathbb{C}^+$ is an analytic function of $z \in \mathbb{C}^+$ that satisfies equation

$$z - \frac{1}{1+\gamma_1 m} = -\frac{1}{m} - \frac{x}{1-\gamma_2 x m}. \tag{13}$$

LEMMA 2. *For any $x > \beta_+$, as $p, n \to_\gamma \infty$,*

$$\left\| M(x) - (s + \gamma_1 I_k)\operatorname{tr}(H - xE)^{-1}/p \right\| \overset{a.s.}{\to} 0 \text{ and}$$
$$\left\| \frac{\mathrm{d}}{\mathrm{d}x} M(x) - (s + \gamma_1 I_k)\frac{\mathrm{d}}{\mathrm{d}x}\operatorname{tr}(H - xE)^{-1}/p \right\| \overset{a.s.}{\to} 0,$$

where $\|\cdot\|$ denotes the spectral norm.

In the next lemma, and throughout this paper, the statement "for sufficiently large p" abbreviates "for sufficiently large p and n along the sequence $p, n \to_\gamma \infty$".

LEMMA 3. *(i) For any $\varepsilon > 0$, the eigenvalues of $M(x)$ are strictly increasing functions of $x \in (\beta_+ + \varepsilon, \infty)$ for sufficiently large p, a.s.;*
(ii) $m(0;x)$ is a strictly increasing, continuous function of $x \in (\beta_+, \infty)$;
(iii) $\lim_{x\to\infty} m(0;x) = 0$, and $\lim_{x\downarrow\beta_+} m(0;x)(s_i + \gamma_1) < -1$ if and only if $s_i > \bar{s}$, where

$$\bar{s} = (\gamma_2 + \rho)/(1 - \gamma_2).$$

Let $\hat{\lambda}_{p1} \geq ... \geq \hat{\lambda}_{pk}$ be the k largest solutions of equation (8). By Lemmas 1–3, if

$$s_1 > ... > s_q > \bar{s} > s_{q+1} > ... > s_k, \tag{14}$$

then $\hat{\lambda}_{pi} \overset{a.s.}{\to} x_i$, where x_i, $i = 1, ..., q$, are such that

$$1 + (s_i + \gamma_1)\, m(0; x_i) = 0 \tag{15}$$

and $m(0; x_i)$ satisfies (13) with x replaced by x_i. In particular,

$$\frac{1}{1 + \gamma_1 m(0; x_i)} - \frac{1}{m(0; x_i)} - \frac{x_i}{1 - \gamma_2 x_i m(0; x_i)} = 0. \tag{16}$$

Combining (15) and (16), we obtain

$$\frac{1}{s_i} + 1 - \frac{x_i}{s_i + \gamma_1 + \gamma_2 x_i} = 0,$$

which implies that

$$x_i = \frac{(s_i + \gamma_1)\,(s_i + 1)}{s_i - \gamma_2\,(s_i + 1)}. \tag{17}$$

By (7), $n_1 \hat{\lambda}_{pi} / (n_1 + k)$, $i = 1, ..., q$, must be the q largest eigenvalues of $\mathbf{F}$, and thus, x_i, $i = 1, ..., q$, describe their a.s. limits. Since there are only q roots of (8) that are asymptotically separated from β_+ and are located above β_+, the other $k - q$ of the largest eigenvalues of $\mathbf{F}$ must a.s. converge to β_+. To summarize, the following theorem holds.

THEOREM 4. *Suppose that $s_1 > ... > s_q > \bar{s} > s_{q+1} > ... > s_k$, and let λ_{pi} be the i-th largest eigenvalue of $\mathbf{F}$. Then for $i \leq q$,*

$$\lambda_{pi} \overset{a.s.}{\to} \frac{(s_i + \gamma_1)\,(s_i + 1)}{s_i - \gamma_2\,(s_i + 1)}$$

as $p, n \to_\gamma \infty$. For $q < i \leq k$, $\lambda_{pi} \overset{a.s.}{\to} \beta_+$.

As follows from Theorem 4, $\bar{s} = (\gamma_2 + \rho)\,/\,(1 - \gamma_2)$ is the phase transition threshold for the eigenvalues of the spiked $\mathbf{F}$-ratio. The value of this threshold diverges to infinity when $\gamma_2 \to 1$. Note that when γ_2 is close to one, the smallest eigenvalue of E is close to zero, which makes E^{-1} a particularly bad estimator of the inverse of the population covariance Σ_2^{-1}. When $\gamma_2 \to 0$, the threshold converges to $\sqrt{\gamma_1}$, which is the phase transition threshold for the eigenvalues of one spiked Wishart matrix. In such a case, x_i converges to $(s_i+\gamma_1)(s_i+1)/s_i$, which is the a.s. limit of the i-th largest eigenvalue of the spiked Wishart when the i-th spike s_i is above $\sqrt{\gamma_1}$.

When both γ_1 and γ_2 converge to zero, x_i converges to s_i+1, which is the population analogue of λ_{pi}. For positive γ_1 and γ_2, λ_{pi} is an upward biased estimator of s_i+1. The relative bias $\lambda_{pi}/(s_i+1)$ converges to $1/(1-\gamma_2)$ when the spike s_i diverges to infinity. The sizes of the relative and absolute biases are very senistive to the value of γ_2. They quickly increase when γ_2 rises above zero. Such a behavior is illustrated in Figure 1.

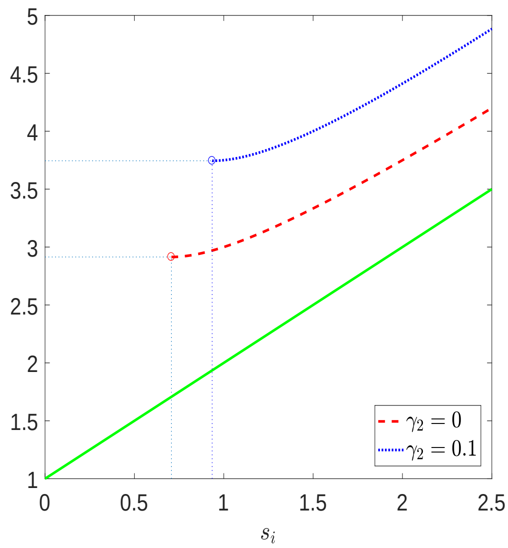

FIG 1. *The a.s. limits of λ_{pi} as functions of s_i for $s_i > \bar{s}$. The top line corresponds to $\gamma_1 = 1/2, \gamma_2 = 1/10$. The dashed line corresponds to the same γ_1, but $\gamma_2 = 0$. The straight line is the population analogue of λ_{pi}, $s_i + 1$.*

4. Asymptotic normality. In this section, we will assume that (14) holds, so that only q eigenvalues of $\mathbf{F}$ separate from the bulk asymptotically. We would like to study their fluctuations around the corresponding a.s. limits. Theorem 4 shows that the limits x_i depend on γ_1 and γ_2. Because of this dependence, the rate of the convergence has to depend on the rates of the convergences $c_1 \to \gamma_1$ and $c_2 \to \gamma_2$. However, as will be shown below, the latter rates do not affect the fluctuations of λ_i around

$$l_i = \frac{(s_i+c_1)(s_i+1)}{s_i - c_2(s_i+1)},$$

which are obtained from x_i by replacing γ_1 and γ_2 by $c_1 \equiv p/n_1$ and $c_2 \equiv p/n_2$ in equation (17).

Similar to x_i, which are linked to the Stieltjes transform of the limiting spectral distribution of $H - xE$ via (15), l_i also can be linked to the limiting Stieltjes transform, albeit under a slightly different asymptotic regime. Precisely, let $m_p(z; x)$ be the Stieltjes transform of the limiting spectral distribution of $H - xE$ as n_1, n_2, and p diverge to infinity so that p/n_1 and p/n_2 *remain fixed*. Then, similarly to (15), we have

$$1 + (s_i + c_1)\, m_p(0; l_i) = 0. \tag{18}$$

This equation will be useful in our analysis below, where we maintain the assumption that p/n_1 and p/n_2 are not necessarily fixed, but converge to γ_1 and γ_2, respectively.

Recall that, by (7), $\lambda_{pi} = n_1 \hat{\lambda}_{pi} / (n_1 + k)$, where $\hat{\lambda}_{pi}$, $i = 1, ..., q$, satisfy (8). Clearly, the asymptotic distributions of $\sqrt{p}\,(\lambda_{pi} - l_i)$ and $\sqrt{p}\left(\hat{\lambda}_{pi} - l_i\right)$, $i = 1, ..., q$, coincide. Therefore, below we will study the asymptotic behavior of the latter. By the standard Taylor expansion argument,

$$\sqrt{p}\left(\hat{\lambda}_{pi} - l_i\right) = -\frac{\sqrt{p}\det \mathcal{M}(l_i)}{\frac{\mathrm{d}}{\mathrm{d}x}\det \mathcal{M}(l_i) + \frac{1}{2}\left(\hat{\lambda}_{pi} - l_i\right)\frac{\mathrm{d}^2}{\mathrm{d}x^2}\det \mathcal{M}\left(\tilde{l}_i\right)}, \tag{19}$$

$i = 1, ..., q$, where $\mathcal{M}(x) = I_k + M(x)$, and $\tilde{l}_i \in \left[l_i, \hat{\lambda}_{pi}\right]$.

We have (see, for example, Magnus and Neudecker (1999) pp. 149–150)

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}x}\det \mathcal{M}(l_i) = \det \mathcal{M}(l_i)\operatorname{tr} J(l_i), \text{ and} \\
&\frac{\mathrm{d}^2}{\mathrm{d}x^2}\det \mathcal{M}(l_i) = \det \mathcal{M}(l_i)\left\{\operatorname{tr} R(l_i) + (\operatorname{tr} J(l_i))^2 - \operatorname{tr} J^2(l_i)\right\},
\end{aligned}$$

where

$$J(x) = \mathcal{M}(x)^{-1}\frac{\mathrm{d}}{\mathrm{d}x}M(x), \text{ and } R(x) = \mathcal{M}(x)^{-1}\frac{\mathrm{d}^2}{\mathrm{d}x^2}M(x).$$

Since the event

$$\det \mathcal{M}(l_i) = 0 \text{ or } 1 + M_{ii}(l_i) = 0 \text{ for some } i = 1, ..., q$$

happens with probability zero, we can simultaneously multiply the numerator and denominator of (19) by $(1 + M_{ii}(l_i))\,/\det \mathcal{M}(l_i)$ to obtain

$$\sqrt{p}\left(\hat{\lambda}_{pi} - l_i\right) = -\frac{\sqrt{p}\,(1 + M_{ii}(l_i))}{\sigma(l_i) + \frac{1}{2}\left(\hat{\lambda}_{pi} - l_i\right)\delta(l_i)}, \tag{20}$$

where

$$\begin{aligned}\sigma(l_i) &= (1+M_{ii}(l_i))\operatorname{tr} J(l_i)\,, \text{ and}\\ \delta(l_i) &= (1+M_{ii}(l_i))\left\{\operatorname{tr} R(l_i) + (\operatorname{tr} J(l_i))^2 - \operatorname{tr} J^2(l_i)\right\}.\end{aligned}$$

A proof of the following lemma is given in the SM.

LEMMA 5. *For any $i = 1, ..., q$, we have: (i) $\sigma(l_i) \xrightarrow{\mathrm{P}} (s_i+\gamma_1)\frac{\mathrm{d}}{\mathrm{d}x}m(0;x_i)$; (ii) $\delta(l_i) = O(1)$ a.s.*

Equation (20), Lemma 5, and the Slutsky theorem imply that, for the purpose of establishing convergence in distribution of $\sqrt{p}\left(\hat\lambda_{pi} - l_i\right)$, $i = 1, ..., q$, we may focus on the numerator of (20)

$$Z_{ii}(l_i) \equiv \sqrt{p}\,(1+M_{ii}(l_i)) = \sqrt{p}\,[M_{ii}(l_i) - (s_i + c_1)\, m_p(0; l_i)]\,,$$

where the last equality follows from (18).

The random variable Z_{ii} is the entry of the matrix

$$Z(l_i) = \sqrt{p}\,[M(l_i) - (s + c_1 I_k)\, m_p(0; l_i)]$$

that belongs to the i-th row and the i-th column. Let us now introduce new notations. Let

$$\begin{aligned}D &= (W_F/n_1)^{1/2}\, s^{1/2}\,(W_v/p)^{-1/2}\,,\\ G &= (H - l_i E)^{-1}/p,\\ \Delta_F &= \sqrt{n_1}\left((W_F/n_1)^{1/2} - I_k\right), \text{ and}\\ \Delta_v &= \sqrt{p}\,(W_v/p - I_k)\,.\end{aligned}$$

Then, using equations (9) and (5), we obtain the following decomposition.

$$Z(l_i) = \sum\nolimits_{v=1}^{7} Z^{(v)},$$

where

$$\begin{aligned}Z^{(1)} &= D\sqrt{p}\,(v'Gv - I_k \operatorname{tr} G)\, D',\\ Z^{(2)} &= (\operatorname{tr} G)\, D\,(W_v/p)^{-1/2}\, s^{1/2}\sqrt{c_1}\Delta_F,\\ Z^{(3)} &= \operatorname{tr} G\sqrt{c_1}\Delta_F s^{1/2}\,(W_v/p)^{-1}\, s^{1/2},\\ Z^{(4)} &= -(\operatorname{tr} G)\, s^{1/2}\Delta_v\,(W_v/p)^{-1}\, s^{1/2},\\ Z^{(5)} &= \sqrt{c_1}\sqrt{p}\,(Dv'Gu + u'GvD')\,,\\ Z^{(6)} &= c_1\sqrt{p}\,(u'Gu - I_k \operatorname{tr} G)\,,\end{aligned}$$

and

$$Z^{(7)} = (s + c_1 I_k)\sqrt{p}\,(\operatorname{tr} G - m_p(0; l_i)).$$

For the last term, $Z^{(7)}$, we have the following lemma.

LEMMA 6. $Z^{(7)} \xrightarrow{a.s.} 0$.

A proof of this lemma is given in the SM. Had l_i been negative, $H - l_i E$ would have been having the form YTY' with $Y \sim N(0, I_p \otimes I_{n_1+n_2})$ and a positive definite diagonal T with converging spectral distribution. Then Lemma 6 would have been following from the results of Bai and Silverstein (2004). Our proof extends Bai and Silverstein's (2004) arguments to the case of negative l_i.

Further, the asymptotic behavior of the terms $Z^{(2)}$ and $Z^{(3)}$ differ depending on the setting. Recall that for SigD, $W_F \sim W_k(n_H, I_k)$. Then, since

$$\Delta_F = \sqrt{n_1}\,(W_F/n_1 - I_k)/2 + o_{\mathrm{P}}(1),$$

a standard CLT together with Lemma 1 imply that

$$\operatorname{diag}\left(Z^{(2)} + Z^{(3)}\right) \xrightarrow{d} N\left(0, 2\gamma_1 m^2(0; x_i) s^2\right). \tag{21}$$

The latter limit is independent from the limits of $Z^{(j)}$, $j \neq 2, 3$, because W_F is independent from u and v.

In contrast, for REG, we have $W_F = n_H I_k$, and $\Delta_F = o(1)$. Therefore,

$$\operatorname{diag}\left(Z^{(2)} + Z^{(3)}\right) \xrightarrow{\mathrm{P}} 0. \tag{22}$$

Let us now establish the convergence of the remaining components $Z^{(j)}$. Let b_i and B_i be such that $[b_i, B_i]$ includes the support of the limiting spectral distribution, $F(\lambda; x_i)$, of $H - l_i E$. Moreover, let $[b_i, B_i]$ be such that none of the eigenvalues $\mu_{1,i} \geq ... \geq \mu_{p,i}$ of $H - l_i E$ lies outside $[b_i, B_i]$ for sufficiently large p, a.s. Further, let g_j with $j = 1, ..., J$, where J is an arbitrary positive integer, be functions which are continuous on $[b_i, B_i]$ and let ζ denote a $p \times q$ matrix with i.i.d. $N(0, 1)$ entries, independent from H and E. Finally, let

$$\Theta = \{(j, s, t) : j = 1, ..., J; 1 \leq s \leq t \leq q\}.$$

The following lemma can be viewed as a special case of Theorem 5.2 in Capitaine et al (2009) or Theorem 7.2 of Bai and Yao (2008), modified to fit the needs of this paper. For readers' convenience, its proof is given in the SM.

LEMMA 7. *The joint distribution of random variables*

$$\left\{ \frac{1}{\sqrt{p}} \sum\nolimits_{r=1}^{p} g_j \left(\mu_{r,i} \right) \left(\zeta_{rs} \zeta_{rt} - \delta_{st} \right), (j, s, t) \in \Theta \right\}$$

weakly converges to a multivariate normal. The covariance between components (j, s, t) *and* (j_1, s_1, t_1) *of the limiting distribution is equal to* 0 *when* $(s, t) \neq (s_1, t_1)$, *and to* $(1 + \delta_{st}) \int g_j(\lambda) g_{j_1}(\lambda) \, \mathrm{d}F(\lambda; x_i)$ *when* $(s, t) = (s_1, t_1)$.

Note that all entries of $Z^{(j)}$, $j = 1, 4, 5, 6$ are linear combinations of the terms having the form considered in Lemma 7, with weights converging in probability to finite constants. Take for example $Z^{(1)}$. Its entries are linear combinations of the entries of

$$\frac{1}{\sqrt{p}} v' \left(H - l_i E \right)^{-1} v - I_k \frac{1}{\sqrt{p}} \operatorname{tr} \left(H - l_i E \right)^{-1},$$

which, in turn, can be represented in the form $\frac{1}{\sqrt{p}} \sum\nolimits_{r=1}^{p} \left(\mu_{r,i} \right)^{-1} \left(\zeta_{rs} \zeta_{rt} - \delta_{st} \right)$. The matrix ζ is obtained by multiplying $[u, v]$ from the left by the eigenvector matrix of $H - l_i E$.

Lemma 7 implies that vector $\left(Z_{ii}^{(1)}, Z_{ii}^{(4)}, Z_{ii}^{(5)}, Z_{ii}^{(6)} \right)$ converges in distribution to a four-dimensional normal vector with zero mean and the following covariance matrix

$$\begin{pmatrix} 2s_i^2 m'(0; x_i) & -2s_i^2 m^2(0; x_i) & 0 & 0 \\ -2s_i^2 m^2(0; x_i) & 2s_i^2 m^2(0; x_i) & 0 & 0 \\ 0 & 0 & 4\gamma_1 s_i m'(0; x_i) & 0 \\ 0 & 0 & 0 & 2\gamma_1^2 m'(0; x_i) \end{pmatrix}.$$

Combining this result with Lemma 6, and convergencies (21) and (22), we obtain, for SigD,

$$Z_{ii}(l_i) \xrightarrow{d} N \left(0, 2 \left(s_i + \gamma_1 \right)^2 m'(0; x_i) - 2s_i^2 \left(1 - \gamma_1 \right) m^2(0; x_i) \right), \tag{23}$$

and, for REG,

$$Z_{ii}(l_i) \xrightarrow{d} N \left(0, 2 \left(s_i + \gamma_1 \right)^2 m'(0; x_i) - 2s_i^2 m^2(0; x_i) \right). \tag{24}$$

To establish the joint convergence of $Z_{ii}(l_i)$, $i = 1, ..., q$, we need another lemma. For each $i = 1, ..., q$, let $g_{j,i}$, with $j = 1, ..., J$, be functions continuous on $[b_i, B_i]$.

LEMMA 8. *For any set of pairs* $\{(s_i, t_i) : i = 1, ..., q\}$ *such that* $(s_{i_1}, t_{i_1}) \neq (s_{i_2}, t_{i_2})$ *for any* $i_1 \neq i_2$, *the joint distribution of random variables*

$$\left\{ \frac{1}{\sqrt{p}} \sum\nolimits_{r=1}^{p} g_{j,i}\left(\mu_{r,i}\right)\left(\zeta_{rs_i}\zeta_{rt_i} - \delta_{s_i t_i}\right), i = 1, ..., q \right\}$$

weakly converges to a multivariate normal. The covariance between components i_1 *and* i_2 *of the limiting distribution is equal to* 0 *when* $i_1 \neq i_2$.

A proof of this lemma is very similar to that of Lemma 7, and we do not report it. Lemma 8 implies that $Z_{ii}(l_i)$, $i = 1, ..., q$ jointly converge to a q-dimensional normal vector with a diagonal covariance matrix. This result, together with equations (7, 20), Lemma 5, and convergences (23, 24) establish the following lemma.

LEMMA 9. *The joint asymptotic distribution of* $\sqrt{p}\left(\lambda_{pi} - l_i\right), i = 1, ..., q$ *is normal, with diagonal covariance matrix. For SigD, the* i*-th diagonal element of the covariance matrix equals*

$$\frac{2\left(s_i + \gamma_1\right)^2 m'\left(0; x_i\right) - 2s_i^2\left(1 - \gamma_1\right) m^2(0; x_i)}{\left(s_i + \gamma_1\right)^2 \left(\frac{\mathrm{d}}{\mathrm{d}x} m(0; x_i)\right)^2}. \tag{25}$$

For REG, it equals

$$\frac{2\left(s_i + \gamma_1\right)^2 m'\left(0; x_i\right) - 2s_i^2 m^2(0; x_i)}{\left(s_i + \gamma_1\right)^2 \left(\frac{\mathrm{d}}{\mathrm{d}x} m(0; x_i)\right)^2}. \tag{26}$$

In the SM, we establish the following explicit expressions for $m^2\left(0; x_i\right)$, $m'\left(0; x_i\right)$, and $\frac{\mathrm{d}}{\mathrm{d}x} m(0; x_i)$:

$$m^2\left(0; x_i\right) = \left(s_i + \gamma_1\right)^{-2}, \tag{27}$$

$$m'\left(0; x_i\right) = -\frac{s_i^2}{\left(s_i + \gamma_1\right)^2 \left(\gamma_1 + \gamma_2\left(1 + s_i\right)^2 - s_i^2\right)}, \tag{28}$$

$$\mathrm{d}m\left(0; x_i\right)/\mathrm{d}x = \frac{-\left(\gamma_2\left(1 + s_i\right) - s_i\right)^2}{\left(s_i + \gamma_1\right)^2 \left(\gamma_1 + \gamma_2\left(1 + s_i\right)^2 - s_i^2\right)}. \tag{29}$$

Using (27), (28), and (29) in (25) and (26), we obtain

THEOREM 10. *Let* $\bar{s} = (\gamma_2 + \rho)/(1-\gamma_2)$, $\underline{s} = -(\gamma_1+\gamma_2)/(\gamma_2+\rho)$, *and* $d(s_i) = (1-\gamma_2)s_i - \gamma_2$. *Then, for any* $s_1 > ... > s_q > \bar{s}$, *the joint asymptotic distribution of* $\sqrt{p}\,(\lambda_{pi} - l_i)$, $i = 1, ..., q$ *is normal with diagonal covariance matrix. In particular,*

$$\sqrt{p}\,(\lambda_{pi} - l_i) \xrightarrow{d} N\left(0, \tau^2(s_i)\right), \tag{30}$$

where $\tau^2(s_i) = w(s_i) \times t(s_i)$ *with*

$$t(s_i) = (1-\gamma_2)(s_i - \underline{s})(s_i - \bar{s})/d^2(s_i)$$

and

$$w(s_i) = \begin{cases} 2\,(\rho s_i(1+s_i)/d(s_i))^2 & \textit{for SigD,} \\ 2\,(\rho s_i(1+s_i)/d(s_i))^2 - 2\gamma_1 s_i^2 t(s_i) & \textit{for REG.} \end{cases} \tag{31}$$

REMARK 11. *The* $w(s_i)$ *component of the asymptotic variance will play the role of the scaling factor in our LAN result below. The fact that the asymptotic variance is smaller for REG than for SigD accords with intuition. Indeed, as discussed above, REG corresponds to the analysis conditional on factors* F, *whereas SigD corresponds to the unconditional analysis. The factors' variance adds to the asymptotic variance of* λ_{pi}.

Similarly to the bias discussed in the previous section, the asymptotic variance of λ_{pi} is sensitive to the size of γ_2. Figure 2 shows that even a small increase in γ_2 may lead to a large increase in the variance.

As the value of the spike s_i approaches the phase transition threshold $\bar{s}$ from above, the asymptotic variance converges to zero. As $s_i \to \infty$, the standard deviation increases linearly in s_i so that the coefficient of variation does not approach zero for large spikes. The limit of the squared coefficient of variation equals

$$\lim_{s_i\to\infty} CV^2(\lambda_{pi}) = \begin{cases} 2\rho^2/(1-\gamma_2) & \text{for SigD} \\ 2\gamma_2/(1-\gamma_2) & \text{for REG.} \end{cases}$$

Again, it is sensitive to the value of γ_2, approaching to infinity as $\gamma_2 \to 1$.

For SigD, when $\gamma_2 \to 0$, the asymptotic variance of λ_{pi} converges to the correct asymptotic variance

$$2\gamma_1\,(s_i+1)^2\left(s_i^2 - \gamma_1\right)/s_i^2$$

of the i-th largest eigenvalue in the spiked Wishart model as derived in Paul (2007). For REG, it converges to the asymptotic variance of the i-th largest eigenvalue in the Wishart model with non-centrality spikes, derived in Onatski (2007).

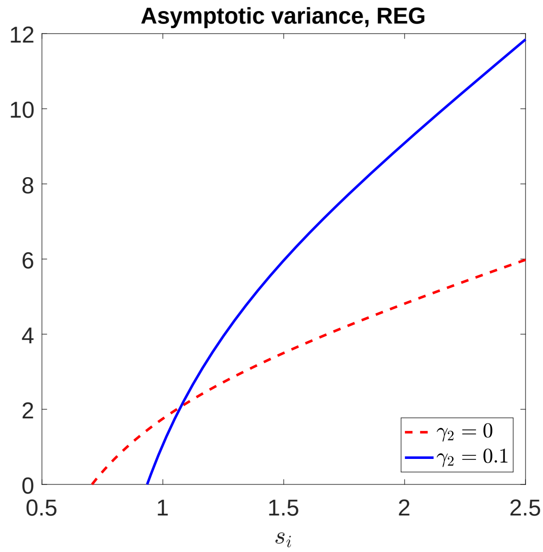

FIG 2. *The asymptotic variance of λ_{pi}, REG case. The solid and dashed lines correspond to $\gamma_2 = 0.1$ and $\gamma_2 = 0$, respectively. For both lines, $\gamma_1 = 1/2$.*

5. Analysis of the joint density of eigenvalues. In the rest of the paper we study the statistical experiment of observing the eigenvalues of $\mathbf{F}$ when the k spikes are local to some fixed points $s_{01} > ... > s_{0k}$ above the phase transition threshold $\bar{s}$. The asymptotics of such an experiment can be characterized by that of the likelihood ratio corresponding to the null and alternative hypotheses

$$H_0 : s^{\text{true}} = s_0 \text{ and } H_1 : s^{\text{true}} = s \equiv s_0 + \delta/\sqrt{p},$$

where $s_0 = \text{diag}\{s_{01}, ..., s_{0k}\}$, and $\delta = \text{diag}\{\delta_1, ..., \delta_k\}$ is the diagonal matrix of local parameters $\delta_j \in \mathbb{R}$. Here we instruduce notation s^{true} for the true values of the spikes to contrast them with the spike *parameters*, s.

When $s^{\text{true}} = s$, the joint density of the p eigenvalues of the multivariate Beta matrix $(n_H\mathbf{H} + n_2\mathbf{E})^{-1} n_H\mathbf{H}$ has the following form (see James (1964), Khatri (1967), and Muirhead (1982), pp. 312–314):

$$f^{\text{SigD}}(x; s) = \frac{Z_{p,n}^{\text{SigD}}(x)}{\det(I_k + s)^{n_H/2}} \, {}_1F_0 \left(N/2; Ls(I_k + s)^{-1}L', x\right), \tag{32}$$

whereas for REG, we have

$$f^{\text{REG}}(x; s) = \frac{Z_{p,n}^{\text{REG}}(x)}{\text{etr}\{n_H s/2\}} \, {}_1F_1 \left(N/2, n_H/2; n_H L s L'/2, x\right). \tag{33}$$

Here the argument of the density, x, is a $p\times p$ real diagonal matrix; ${}_1F_0$ and ${}_1F_1$ are the hypergeometric functions of two matrix arguments; $N=n_H+n_2$; L is the $p\times k$ matrix of nuisance parameters, or factor loadings, as in (2); and $Z_{p,n}^{\mathrm{Case}}(x)$ with Case $=\mathrm{SigD},\mathrm{REG}$ depend on n_H,n_2,p and x, but not on s.

Let $\tilde\lambda_{pj}$ be the eigenvalues of $(n_H\mathbf{H}+n_2\mathbf{E})^{-1}n_H\mathbf{H}$ for some arbitrary value of s^{true}, not necessarily equal to s, and let $\tilde\Lambda=\operatorname{diag}\left\{\tilde\lambda_{p1},\cdots,\tilde\lambda_{pp}\right\}$. We would like to study the asymptotic behavior, under the null hypothesis, of the likelihood ratios

$$f^{\mathrm{Case}}(\tilde\Lambda;s)/f^{\mathrm{Case}}(\tilde\Lambda;s_0)$$

with Case $=\mathrm{SigD},\mathrm{REG}$ as $p,n\to_\gamma\infty$.

The eigenvalues $\tilde\lambda_{pj}$ are related to the eigenvalues λ_{pj} of the $\mathbf{F}$-ratio as follows

$$\tilde\lambda_{pj}=\alpha_n\lambda_{pj}/\left(1+\alpha_n\lambda_{pj}\right),\text{ where }\alpha_n=n_H/n_2;$$

For the purpose of the analysis of the likelihood ratios, we find it more convenient to work with $\tilde\lambda_{pj}$ rather than with λ_{pj}.

First, we use Lemma 1 of Passemier et al (2014) to rewrite $f^{\mathrm{SigD}}(\tilde\Lambda;s)$ and $f^{\mathrm{REG}}(\tilde\Lambda;s)$ in the form of repeated contour integrals that involve hypergeometric functions of two matrix arguments of fixed dimension $k\times k$. Let Z be a $k\times k$ diagonal matrix with complex variables z_j along the diagonal, and let

$$\omega=\prod_{j\neq i}^{k}\left(1-z_iz_j^{-1}\right)^{1/2}\prod_{j=1}^{k}\left[z_j^{-(p-k+1)/2}\prod_{s=1}^{p}\left(1-\tilde\lambda_{ps}z_j^{-1}\right)^{-1/2}\right],\tag{34}$$

where the principal branches of all the fractional powers are taken. Let $C_{p,n}^{\mathrm{Case}}(\tilde\Lambda)$ be some real quantity that depend on 'Case', n_H,n_2,p and $\tilde\Lambda$, but not on s; and let

$$k_{p,n}^{\mathrm{SigD}}(s)=\left[\det(I_k+s)\right]^{\frac{p-k-n_H-1}{2}}\left[\det s\right]^{-\frac{p-k-1}{2}},\tag{35}$$

and

$$k_{p,n}^{\mathrm{REG}}(s)=\operatorname{etr}\left\{-n_Hs/2\right\}\left[\det s\right]^{-\frac{p-k-1}{2}}.\tag{36}$$

LEMMA 12. *Let $\tilde{\mathcal{K}}$ be a counter-clockwise oriented contour in the complex plane that encircles zero and $\tilde\lambda_{pj}$, $j=1,...,p$, and intersects each of the rays $\{z:\arg z=\varphi\}$, $\varphi\in(-\pi,\pi]$ only once. Then, for even $p-k+1$, we have*

$$f^{\mathrm{Case}}(\tilde\Lambda;s)=\frac{C_{p,n}^{\mathrm{Case}}(\tilde\Lambda)k_{p,n}^{\mathrm{Case}}(s)}{(2\mathrm{i})^k}\int_{\tilde{\mathcal{K}}}\cdots\int_{\tilde{\mathcal{K}}}\omega\mathcal{F}_{p,n}^{\mathrm{Case}}\prod_{i=1}^{k}\mathrm{d}z_i,\tag{37}$$

where i *is the imaginary unit,* Case = SigD, REG*,*

$$\mathcal{F}_{p,n}^{\text{SigD}} = {}_1F_0\left(\frac{N-p+k+1}{2}; s(I_k+s)^{-1}, Z\right), \tag{38}$$

and

$$\mathcal{F}_{p,n}^{\text{REG}} = {}_1F_1\left(\frac{N-p+k+1}{2}, \frac{n_H-p+k+1}{2}; \frac{n_H}{2}s, Z\right). \tag{39}$$

The lemma is a direct corollary of Lemma 1 of Passemier et al (2014). The requirement that $\tilde{\mathcal{K}}$ intersects each of the rays emanating from $z=0$ only once ensures that the branches of the fractional powers in ω are principal. Indeed, Onatski's (2013) Lemma 1, which Lemma 1 of Passemier et al (2014) is based on, is proven first under the assumption that $\tilde{\mathcal{K}}$ is the unit circle and the principal branches of the fractional powers in ω are used. Then the contour is deformed without changing the value of the integrals. When $\tilde{\mathcal{K}}$ is deformed so that the rays $\{z : \arg z = \varphi\}, \varphi \in (-\pi, \pi]$ are intersected by $\tilde{\mathcal{K}}$ only once, the arguments of the fractional power functions in ω never hit the negative semi-axis (note that $z_j^{-(p-k+1)/2}$ is not a fractional power when $p-k+1$ is even), and therefore, the principal branches of the fractional powers should still be used after the deformation of $\tilde{\mathcal{K}}$.

In future work, it would be interesting to relax the technical requirement that $p-k+1$ is even. In a previous version of this paper, we provide such a relaxation for the case of a single spike, $k=1$. An extension to $k>1$ requires a separate non-trivial effort.

5.1. *Contour deformation.* Let us deform the contour of integration $\tilde{\mathcal{K}}$ into contour $\mathcal{K}$ as shown on Figure 3. Parts $\mathcal{K}_j^+$ and $\mathcal{K}_j^-$, $j=1,...,k$, of $\mathcal{K}$ are shown non-overlaping with the real axis to enhance visibility. In fact, these parts coincide with the axis. The position $\tilde{x}_0$ of a kink in $\mathcal{K}$ is fixed so that

$$\alpha\beta_+/\left(1+\alpha\beta_+\right) < \tilde{x}_0 < \alpha x_k/\left(1+\alpha x_k\right)$$

with $\alpha = \lim \alpha_n = \gamma_2/\gamma_1$, and

$$x_j = \lim l_j = \frac{\left(s_{0j}+\gamma_1\right)\left(s_{0j}+1\right)}{s_{0j}-\left(s_{0j}+1\right)\gamma_2}, \qquad j=1,...,k. \tag{40}$$

As follows from our results in the previous sections, under the null,

$$\tilde{\lambda}_{pk} \overset{a.s.}{\to} \alpha x_k/\left(1+\alpha x_k\right), \tag{41}$$

and

$$\tilde{\lambda}_{p,k+1} \overset{a.s.}{\to} \alpha\beta_+/\left(1+\alpha\beta_+\right), \tag{42}$$

so $\tilde{x}_0 \in \left(\tilde{\lambda}_{p,k+1}, \tilde{\lambda}_{pk}\right)$ for sufficiently large p, a.s.

The radius of the circles around $\tilde{\lambda}_{pj}$ with $j = 1, ..., k$ can be chosen arbitrarily small. Since, as can be seen from (34), the singularities of the integrand at $\tilde{\lambda}_{pj}$ are of the inverse square-root-type, the contribution of the circles to the integral disappear in the limit when the radius tends to zero. Below, we will consider this limiting version of $\mathcal{K}$, that is, the contour with the horizontal part given by the two differently oriented copies of $[\tilde{x}_0, \tilde{\lambda}_{p1}]$, where the points $\tilde{\lambda}_{p1}, ..., \tilde{\lambda}_{pk}$ are excluded.

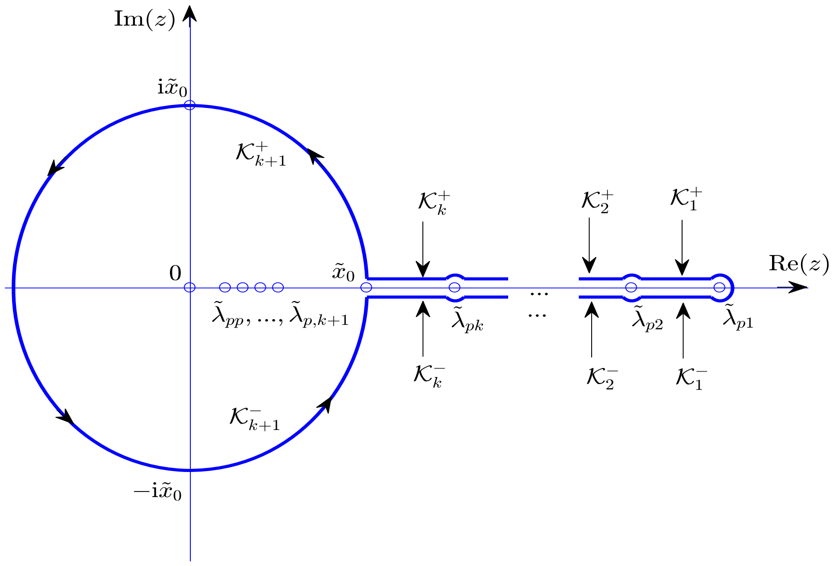

FIG 3. *Deformed contour $\mathcal{K}$.*

Since contour $\mathcal{K}$ has common intervals with the ray $\{z : \arg z = 0\}$, some of the arguments of the fractional power functions involved in ω are real and negative. Therefore, care should be taken to identify the branches used.

Suppose that

$$z_{j_1}, ..., z_{j_r} \in \mathcal{K} \cap [\tilde{x}_0, \tilde{\lambda}_{p1}],$$

where $r \leq k$ and $z_{j_1} < ... < z_{j_r}$, and let all z_j with $j \notin \{j_1, ..., j_r\}$ belong to $\mathcal{K} \backslash [\tilde{x}_0, \tilde{\lambda}_1]$. To simplify notation, we may assume that $j_t = t$. Since ω is symmetric in $z_1, ..., z_k$, this assumption is without loss of generality. Then the parts of ω that need the branch identification are

$$\prod_{j>i}^{r}\left(1-z_j z_i^{-1}\right)^{1/2} \text{ and } \left(1-\tilde{\lambda}_{ps} z_i^{-1}\right)^{-1/2} \text{ for } \tilde{\lambda}_{ps} > z_i.$$

In the SM, we prove the following lemma.

LEMMA 13. *Suppose that $z_1, ..., z_r \in \mathcal{K} \cap [\tilde{x}_0, \tilde{\lambda}_{p1}]$ are such that $z_1 < ... < z_r$, and let $sgn_{z_i} = +1$ if z_i belongs to the "upper" portion of $\mathcal{K} \cap [\tilde{x}_0, \tilde{\lambda}_{p1}]$, that is, the portion oriented from $\tilde{\lambda}_{p1}$ to $\tilde{x}_0$, and $sgn_{z_i} = -1$ if z_i belongs to the "lower" portion of $\mathcal{K} \cap [\tilde{x}_0, \tilde{\lambda}_{p1}]$, that is, the portion oriented from $\tilde{x}_0$ to $\tilde{\lambda}_{p1}$. Then for $j > i$, we have*

$$\left(1-z_j z_i^{-1}\right)^{1/2} = \mathrm{i} \times sgn_{z_i} \left|1-z_j z_i^{-1}\right|^{1/2},$$

while for $\tilde{\lambda}_{ps} > z_i$, we have

$$\left(1-\tilde{\lambda}_{ps} z_i^{-1}\right)^{-1/2} = -\mathrm{i} \times sgn_{z_i} \left|1-\tilde{\lambda}_{ps} z_i^{-1}\right|^{-1/2}.$$

5.2. *Decomposition of the contour integral.* Let us split $\mathcal{K}$ into $2\times(k+1)$ parts

$$\mathcal{K} = \bigcup_{i=1}^{k+1}\left\{\mathcal{K}_i^+ \cup \mathcal{K}_i^-\right\}$$

as shown on Figure 3, and let $\mathcal{K}_i = \mathcal{K}_i^+ \cup \mathcal{K}_i^-$. For any $\sigma = (\sigma_1, ..., \sigma_k)$ with $\sigma_i \in \{1, ..., k+1\}$, let

$$\mathcal{I}_\sigma = \frac{1}{(2\mathrm{i})^k}\int_{\mathcal{K}_{\sigma_k}} ... \int_{\mathcal{K}_{\sigma_1}} \omega \mathcal{F}_{p,n} \prod_{i=1}^{k} \mathrm{d}z_i.$$

Since $\omega\mathcal{F}_{p,n}$ is symmetric in the variables z_i, we may permute them so that $z_1 < ... < z_r$ are in $\mathcal{K} \cap \left[\tilde{x}_0, \tilde{\lambda}_{p1}\right]$ and $z_{r+1}, ..., z_k$ lie in $\mathcal{K}_{k+1}$. Let $\mathcal{S}^r$ denote the simplex defined by $z_1 < ... < z_r$. Consider a sequence σ with $\sigma_r \le ... \le \sigma_1 \le k$ and $\sigma_{r+1} = ... = \sigma_k = k+1$. Consider the iterated integral

$$\mathcal{I}'_\sigma = \frac{1}{(2\mathrm{i})^k}\int_{\mathcal{K}_\sigma \cap \mathcal{S}^r} \omega \mathcal{F}_{p,n} \prod_{i=1}^{k} \mathrm{d}z_i. \tag{43}$$

In the SM, we show that $\mathcal{I}'_\sigma$ vanishes if there are any repeats in σ on the real axis.

If $\mathcal{I}'_\sigma$ vanishes, $\mathcal{I}_\sigma$ vanishes too. Therefore, the components of the integral

$$\mathcal{I} \equiv \frac{1}{(2\mathrm{i})^k}\int_{\mathcal{K}}\cdots\int_{\mathcal{K}}\omega\mathcal{F}_{p,n}\prod_{j=1}^k \mathrm{d}z_j$$

represented by $\mathcal{I}_\sigma$ with repeated $\sigma_j \leq k$ equal zero. This implies the following lemma. Let τ be any subset of $\{1, 2, ..., k\}$, and let $\sigma_\tau = (\sigma_{1\tau}, ..., \sigma_{k\tau})$, where

$$\sigma_{j\tau} = \begin{cases} k+1 & \text{if } j \in \tau \\ j & \text{if } j \notin \tau \end{cases}.$$

LEMMA 14. *Let T be the set of all the subsets of $\{1, 2, ..., k\}$. Then,*

$$\mathcal{I} = \sum_{\tau\in T}\frac{k!}{|\tau|!}\mathcal{I}_{\sigma_\tau}. \tag{44}$$

REMARK 15. *The multiplier $k!/|\tau|!$ in the latter expression counts the number of integrals $\mathcal{I}_\sigma$, which are different from $\mathcal{I}_{\sigma_\tau}$ only by permutation of the variables of integration, $z_1, ..., z_k$.*

Below, we will show that, asymptotically as $p, n \to_\gamma \infty$, all integrals $\mathcal{I}_{\sigma_\tau}$ are dominated by

$$\mathcal{I}_{\sigma_\varnothing} = \frac{1}{(2\mathrm{i})^k}\int_{\mathcal{K}_k}\cdots\int_{\mathcal{K}_1}\omega\mathcal{F}_{p,n}\prod_{j=1}^k \mathrm{d}z_j$$

so that $\mathcal{I}$ is asymptotically equivalent to $k!\mathcal{I}_{\sigma_\varnothing}$. Using Lemma 13, it is straightforward to verify that

$$\mathcal{I}_{\sigma_\varnothing} = \int_{\tilde{x}_0}^{\tilde{\lambda}_{pk}}\int_{\tilde{\lambda}_{pk}}^{\tilde{\lambda}_{p,k-1}}\cdots\int_{\tilde{\lambda}_{p2}}^{\tilde{\lambda}_{p1}}|\omega|\,\mathcal{F}_{p,n}\prod_{j=1}^k \mathrm{d}z_j \tag{45}$$

Note that the constant $(2\mathrm{i})^k$ in the denominator has canceled out.

To study the asymptotics of $\mathcal{I}_{\sigma_\varnothing}$, we will use Laplace approximation for the integrals involved in the above expression. However, first, we need to replace $\mathcal{F}_{p,n}$, that involves the hypergeometric function ${}_1F_0$ for SigD and ${}_1F_1$ for REG, by tractable approximations. This requires a separate Laplace approximation step.

5.3. *Laplace approximations for $\mathcal{F}_{p,n}$.* As follows from equation (38),

$$\begin{aligned}\mathcal{F}_{p,n}^{\mathrm{SigD}} &= {}_1F_0\left(m; s(I_k+s)^{-1}, Z\right)\\ &= [\det(I-Z)]^{-m}\,{}_1F_0\left(m; -(I_k+s)^{-1}, Z(I_k-Z)^{-1}\right),\end{aligned}$$

where $m=(N-p+k+1)/2$. Chang (1970) studies the asymptotic behavior of ${}_1F_0(m,-A,B)$ for fixed diagonal matrices A and B as $m\to\infty$. The following lemma uses a minor modification of Chang's Theorem 1 to derive a Laplace approximation for $\mathcal{F}_{p,n}^{\mathrm{SigD}}$ that is uniform over a set of diagonal matrices s and Z (see the SM for a proof).

LEMMA 16. *For $Z=\mathrm{diag}\{z_1,...,z_k\}$ such that $1>z_1>...>z_k>0$, and for $s=\mathrm{diag}\{s_1,...,s_k\}$ such that $s_1>...>s_k>0$, as $m\to\infty$, we have*

$$\mathcal{F}_{p,n}^{\mathrm{SigD}} = \Gamma_k(k/2)\,\pi^{-k(k+1)/4}\left(D^{\mathrm{SigD}}\right)^{-m}\prod_{i<j}^{k}(mc_{ij})^{-1/2}(1+o(1)),\tag{46}$$

where $\Gamma_k(x)=\pi^{k(k-1)/4}\prod_{i=1}^{k}\Gamma(x-(i-1)/2)$ is the multivariate Gamma function,

$$D^{\mathrm{SigD}}=\det\left(I_k-s(I_k+s)^{-1}Z\right),\ c_{ij}=\frac{(s_i-s_j)(z_i-z_j)}{(1+s_i-s_iz_i)(1+s_j-s_jz_j)},$$

and $o(1)\to 0$ uniformly on any compact subsets of the simplexes $1>z_1>...>z_k>0$ and $s_1>...>s_k>0$.

For REG case, we have from (39)

$$\mathcal{F}_{p,n}^{\mathrm{REG}} = {}_1F_1\left(Ta+(k+1)/2, Tb+(k+1)/2, Ts/2, Z\right),$$

where $T=n_H$, $a=(N-p)/(2n_H)$, and $b=(n_H-p)/(2n_H)$. Note that for sufficiently large p, as $p,n\to_\gamma\infty$, we must have $a\in(1/2,\infty)$ and $b\in(0,1/2)$.

The asymptotics of ${}_1F_1(\alpha,\beta;A,B)$ where α and A diverge to ∞ at the same rate was studied in Glynn (1980). We however need the asymptotics of this function when not only $\alpha\equiv Ta+(k+1)/2$ and $A\equiv Ts/2$, but also $\beta\equiv Tb+(k+1)/2$ diverge to infinity. Following Glynn's (1980) strategy of proof, we derive the following result. Its proof is reported in the SM.

LEMMA 17. *Suppose that a and b belong to compact subsets of $(1/2,\infty)$ and $(0,1/2)$, respectively, while the diagonal entries of $s=\mathrm{diag}(s_1,...,s_k)$*

and $Z = \operatorname{diag}(z_1, ..., z_k)$ *belong to compact subsets of the simplexes* $s_1 > ... > s_k > 0$ *and* $z_1 > ... > z_k > 0$*. Then, as* $T \to \infty$*, we have*

$$\mathcal{F}_{p,n}^{\mathrm{REG}} = (T/2)^{-k(k-1)/4} \frac{\Gamma_k(k/2)}{\pi^{k(k+1)/4}} \frac{b^{kTb+k(k+1)/4}}{a^{kTa+k(k+1)/4}} \times D^{\mathrm{REG}} \prod_{i<j} (z_i - z_j)^{-1/2} (s_i - s_j)^{-1/2} \times (1 + o(1)), \tag{47}$$

where

$$D^{\mathrm{REG}} = \prod_{j=1}^{k} e^{Tz_{j+}} \frac{(z_{j+} + a)^{aT}}{(z_{j+} + b)^{bT}} \left(\frac{z_{j+}(z_{j+} + a)}{z_{j+}^2 + a\zeta_j} \right)^{\frac{1}{2}},$$

$$z_{j+} = \frac{1}{2} \left\{ \zeta_j - b + \sqrt{(b - \zeta_j)^2 + 4a\zeta_j} \right\} \text{ with } \zeta_j = z_j s_j / 2,$$

and $o(1) \to 0$ *uniformly over* a, b, s *and* Z *that satisfy the above requirements.*

5.4. *Laplace approximations for* $\mathcal{I}_{\sigma_\varnothing}$. Note that the asymptotic approximations (46) and (47) do not hold for z_i, $i = 1, ..., k$, that may approach one another. Therefore, we shall, first, analyze a multiple integral with trimmed integration domains

$$\overline{\mathcal{I}}_{\sigma_\varnothing} = \int_{\tilde{x}_0}^{\tilde{\lambda}_{pk}} \int_{\tilde{\lambda}_{pk}+\varepsilon}^{\tilde{\lambda}_{p,k-1}} ... \int_{\tilde{\lambda}_{p2}+\varepsilon}^{\tilde{\lambda}_{p1}} |\omega| \, \mathcal{F}_{p,n} \prod_{j=1}^{k} \mathrm{d}z_j, \tag{48}$$

where ε is a fixed small positive number. Then, we will show that $\overline{\mathcal{I}}_{\sigma_\varnothing}$ is asymptotically equivalent to $\mathcal{I}_{\sigma_\varnothing}$.

Although the strategy of such an analysis is the same for SigD and REG, the details are different. We start from the SigD case and then turn to the REG case.

SigD case. First, we use Lemma 16 to obtain

$$\overline{\mathcal{I}}_{\sigma_\varnothing}^{\mathrm{SigD}} = b^{\mathrm{SigD}}(s) \int_{\tilde{x}_0}^{\tilde{\lambda}_{pk}} \int_{\tilde{\lambda}_{pk}+\varepsilon}^{\tilde{\lambda}_{p,k-1}} ... \int_{\tilde{\lambda}_{p2}+\varepsilon}^{\tilde{\lambda}_{p1}} G^{\mathrm{SigD}} \times (1 + o(1)) \prod_{j=1}^{k} \mathrm{d}z_j, \tag{49}$$

where

$$G^{\mathrm{SigD}} = |\omega| \left(D^{\mathrm{SigD}} \right)^{-\frac{N-p+2}{2}} \prod_{i<j}^{k} (z_i - z_j)^{-1/2},$$

$$b^{\mathrm{SigD}}(s) = \frac{\Gamma_k(k/2)}{\pi^{k(k+1)/4} m^{k(k-1)/4}} \prod_{i<j}^{k} \left(\frac{(1 + s_i)(1 + s_j)}{s_i - s_j} \right)^{1/2}, \tag{50}$$

and $o(1)$ converges to zero as $p, n \to_\gamma \infty$, uniformly over $s = s_0 + \delta/\sqrt{p}$ such that $(\delta_1, ..., \delta_k)$ belongs to a compact subset of $\mathbb{R}^k$, and over Z such that $(z_1, ..., z_r)$ belongs to the trimmed domain of integration.

Consider the inner-most integral in (49),

$$\overline{\mathcal{I}}_{\sigma_\varnothing, in}^{\mathrm{SigD}} = \int_{\tilde{\lambda}_{p2+\varepsilon}}^{\tilde{\lambda}_{p1}} G^{\mathrm{SigD}} \times (1 + o(1))\, \mathrm{d}z_1.$$

Using the definition of ω, we rewrite this integral in the following form

$$\overline{\mathcal{I}}_{\sigma_\varnothing, in}^{\mathrm{SigD}} = G_{-1}^{\mathrm{SigD}} \int_{\tilde{\lambda}_{p2+\varepsilon}}^{\tilde{\lambda}_{p1}} e^{-pf_1(z_1)} g_1(z_1)(1 + o(1))\, \mathrm{d}z_1, \tag{51}$$

where f_1, g_1, and G_{-1}^{SigD} are defined in Table 1.

TABLE 1

Definition of f_1, g_1, and $G_{\mathrm{SigD},-1}$ used in equation (51), and of H_j used in Lemma 18.

Quantity		Definition
$f_1(z_1)$	$=$	$\frac{N-p+2}{2p} \ln\left(1 - \frac{s_1 z_1}{1+s_1}\right) + \frac{1}{2p}\sum_{j=k+1}^{p} \ln\left(z_1 - \tilde{\lambda}_{pj}\right)$
$g_1(z_1)$	$=$	$\left(\tilde{\lambda}_{p1} - z_1\right)^{-1/2} \prod_{j=2}^{k} \left((z_1 - z_j) / \left(z_1 - \tilde{\lambda}_{pj}\right)\right)^{1/2}$
G_{-1}^{SigD}	$=$	$\lvert\omega_{-1}\rvert \left(D_{-1}^{\mathrm{SigD}}\right)^{-\frac{N-p+2}{2}} \prod_{i,j=2,\dots,k; i<j} (z_i - z_j)^{-1/2}$
D_{-1}^{SigD}	$=$	$\det\left(I_{k-1} - s_{-1}(I_{k-1} + s_{-1})^{-1} Z_{-1}\right)$
ω_{-1}	$=$	$\prod_{i,j=2,\dots,k; j\neq i} \left(1 - z_i z_j^{-1}\right)^{1/2} \prod_{j=2}^{k} \left[z_j^{-(p-k+2)/2} \prod_{s=1}^{p} \left(1 - \tilde{\lambda}_{ps} z_j^{-1}\right)^{-1/2}\right]$
Z_{-1}	$=$	$\mathrm{diag}\{z_2, ..., z_k\}$
s_{-1}	$=$	$\mathrm{diag}\{s_2, ..., s_k\}$
H_j	$=$	$H_{j,\mathrm{num}}/H_{j,\mathrm{den}}$
$H_{j,\mathrm{num}}$	$=$	$(1-\gamma_2)s_{0j}(\gamma_1 + \gamma_2 + \gamma_2 s_{0j})\left(1 + s_{0j} - \beta_+^{1/2}\right)\left(1 + s_{0j} - \beta_-^{1/2}\right)$
$H_{j,\mathrm{den}}$	$=$	$2\gamma_1\gamma_2(s_{0j} - \gamma_2 s_{0j} - \gamma_2)(1 + s_{0j})(\gamma_1 + s_{0j})$

Using the Laplace method to approximate the integral in (51) (see the SM for details), and then repeating the procedure for the second, third, etc. to the inner-most integral in (49) and combining the results, we obtain the following lemma.

LEMMA 18. *Under the null hypothesis $H_0 : s^{\mathrm{true}} = s_0$, as $p, n \to_\gamma \infty$,*

$$\overline{\mathcal{I}}_{\sigma_\varnothing}^{\mathrm{SigD}} = b^{\mathrm{SigD}}(s) \prod_{j=1}^{k} \left[\Omega_j^{\mathrm{SigD}} (pH_j/\pi)^{-1/2} \prod_{s=k+1}^{p} \left(\tilde{\lambda}_{pj} - \tilde{\lambda}_{ps}\right)^{-1/2}\right](1 + o(1)),$$

where $\Omega_j^{\mathrm{SigD}} = \left(1 - \frac{s_j}{1+s_j}\tilde{\lambda}_{pj}\right)^{-\frac{N-p+2}{2}}$, H_j *are as defined in Table 1, and* $o(1) \to 0$ *uniformly over* $s = s_0 + \delta/\sqrt{p}$ *with* $(\delta_1, ..., \delta_k)$ *from a compact subset of* $\mathbb{R}^k$*, a.s.*

REG case. First, we use Lemma 17, to obtain

$$\overline{\mathcal{I}}_{\sigma_\varnothing}^{\mathrm{REG}} = b^{REG}(s) \int_{\tilde{x}_0}^{\tilde{\lambda}_{pk}} \int_{\tilde{\lambda}_{pk}+\varepsilon}^{\tilde{\lambda}_{p,k-1}} ... \int_{\tilde{\lambda}_{p2}+\varepsilon}^{\tilde{\lambda}_{p1}} G^{\mathrm{REG}} \times (1 + o(1)) \prod_{j=1}^{k} \mathrm{d}z_j, \tag{52}$$

where

$$G^{\mathrm{REG}} = |\omega| \, D^{\mathrm{REG}} \prod_{i<j} (z_i - z_j)^{-1/2}$$

with

$$D^{\mathrm{REG}} = \prod_{j=1}^{k} e^{n_H z_{j+}} \frac{(z_{j+} + a)^{an_H}}{(z_{j+} + b)^{bn_H}} \left(\frac{z_{j+} (z_{j+} + a)}{z_{j+}^2 + a\zeta_j} \right)^{\frac{1}{2}},$$

$$a = (N-p)/(2n_H), \; b = (n_H - p)/(2n_H), \text{ and}$$

$$b^{\mathrm{REG}}(s) = \left(\frac{n_H}{2}\right)^{-k(k-1)/4} \frac{\Gamma_k(k/2)}{\pi^{k(k+1)/4}} \frac{b^{kn_H b + k(k+1)/4}}{a^{kn_H a + k(k+1)/4}} \prod_{i<j} (s_i - s_j)^{-1/2}. \tag{53}$$

The same uniformity properties of $o(1)$ as in the case of (49) apply.

Consider the inner-most integrals in (52),

$$\overline{\mathcal{I}}_{\sigma_\varnothing, in}^{\mathrm{REG}} = \int_{\tilde{\lambda}_{p2}+\varepsilon}^{\tilde{\lambda}_{p1}} G^{\mathrm{REG}} \times (1 + o(1)) \, \mathrm{d}z_1.$$

Using the definition of ω, we rewrite this integral in the following form

$$\overline{\mathcal{I}}_{\sigma_\varnothing, in}^{\mathrm{REG}} = G_{-1}^{\mathrm{REG}} \int_{\tilde{\lambda}_{p2}+\varepsilon}^{\tilde{\lambda}_{p1}} e^{-n_H f_2(z_1)} g_2(z_1) (1 + o(1)) \, \mathrm{d}z_1, \tag{54}$$

where f_2, g_2, and G_{-1}^{REG} are defined in Table 2.

Similarly to the integral in (51), the one in (54) can be analyzed using the Laplace approximation steps. Repeating the procedure for the second, third, etc. to the inner-most integral in (52) and combining the results, we obtain the following lemma. See the SM for a detailed proof.

LEMMA 19. *Under the null hypothesis $H_0 : s^{\text{true}} = s_0$, as $p, n \to_\gamma \infty$,*

$$\overline{\mathcal{I}}^{\text{REG}}_{\sigma_\varnothing} = b^{\text{REG}}(s) \prod_{j=1}^{k} \left[\Omega_j^{\text{REG}} (pH_j/\pi)^{-1/2} \prod_{s=k+1}^{p} \left(\tilde{\lambda}_{pj} - \tilde{\lambda}_{ps} \right)^{-1/2} \right] (1 + o(1)),$$

where

$$\Omega_j^{\text{REG}} = e^{n_H \bar{z}_{j+}} \frac{(\bar{z}_{j+} + a)^{an_H}}{(\bar{z}_{j+} + b)^{bn_H}} \frac{\gamma_1 + \gamma_2 + \gamma_2 s_{0j}}{((\gamma_1 + \gamma_2 + \gamma_2 s_{0j})^2 - \rho^2 \gamma_1)^{1/2}}$$

and $\bar{z}_{j+}$ is the value of z_{j+} that corresponds to $\zeta_j = \tilde{\lambda}_{pj} s_j/2$. The H_j are as defined in Table 1, and $o(1) \to 0$ uniformly over $s = s_0 + \delta/\sqrt{p}$ with $(\delta_1, ..., \delta_k)$ from a compact subset of $\mathbb{R}^k$, a.s.

Now let us show that $\overline{\mathcal{I}}_{\sigma_\varnothing}$ is asymptotically equivalent to $\mathcal{I}_{\sigma_\varnothing}$. By definition,

$$\mathcal{I}_{\sigma_\varnothing} - \overline{\mathcal{I}}_{\sigma_\varnothing} = \sum_{\mathcal{D}} \int_{\tilde{x}_0}^{\tilde{\lambda}_{pk}} \int_{\mathcal{D}_{k-1}} ... \int_{\mathcal{D}_1} |\omega| \, \mathcal{F}_{p,n} \prod_{j=1}^{k} \mathrm{d}z_j,$$

where the sum runs over all $\mathcal{D}_j$ that are represented by either $[\tilde{\lambda}_{j+1} + \varepsilon, \tilde{\lambda}_j]$ or $[\tilde{\lambda}_{j+1}, \tilde{\lambda}_{j+1} + \varepsilon]$, and at least one $\mathcal{D}_j$, $j = 1, ..., k-1$, is represented by $[\tilde{\lambda}_{j+1}, \tilde{\lambda}_{j+1} + \varepsilon]$. All terms in this sum cam be analyzed similarly. Let us explain the main idea of the analysis using the term

$$J \equiv \int_{\tilde{x}_0}^{\tilde{\lambda}_{pk}} \int_{\tilde{\lambda}_{pk}+\varepsilon}^{\tilde{\lambda}_{p,k-1}} ... \int_{\tilde{\lambda}_{p3}+\varepsilon}^{\tilde{\lambda}_{p2}} \int_{\tilde{\lambda}_{p2}}^{\tilde{\lambda}_{p2}+\varepsilon} |\omega| \, \mathcal{F}_{p,n} \prod_{j=1}^{k} \mathrm{d}z_j.$$

Since the lower integration limit of the inner-most integral coincides with the upper integration limit of the second inner-most integral, we cannot

TABLE 2
Definition of f_2, g_2, and $G_{\text{REG},-1}$ used in equation (54). Quantities ω_{-1}, Z_{-1}, and s_{-1} are as defined in Table 1.

Quantity		Definition
$f_2(z_1)$	$=$	$-z_{1+} - a\ln(z_{1+} + a) + b\ln(z_{1+} + b) + \frac{1}{2n_H} \sum_{j=k+1}^{p} \ln\left(z_1 - \tilde{\lambda}_{pj}\right)$
$g_2(z_1)$	$=$	$\left(\frac{z_{1+}(z_{1+}+a)}{z_{1+}^2 + a\zeta_1} \right)^{\frac{1}{2}} \left(\tilde{\lambda}_{p1} - z_1 \right)^{-1/2} \prod_{j=2}^{k} \left(\frac{z_1 - z_j}{z_1 - \tilde{\lambda}_{pj}} \right)^{1/2}$
G_{-1}^{REG}	$=$	$\lvert\omega_{-1}\rvert D_{-1}^{\text{REG}} \prod_{i,j=2,\ldots,k;i<j} (z_i - z_j)^{-1/2}$
D_{-1}^{REG}	$=$	$\prod_{j=2}^{k} e^{n_H z_{j+}} \frac{(z_{j+}+a)^{an_H}}{(z_{j+}+b)^{bn_H}} \left(\frac{z_{j+}(z_{j+}+a)}{z_{j+}^2 + a\zeta_j} \right)^{\frac{1}{2}}$

use Lemmas 16 and 17 to approximate $\mathcal{F}_{p,n}$ uniformly over the integration domain of J. However, we can obtain an upper bound on $|J|$ that can be analyzed using these lemmas.

The key is to observe that $\mathcal{F}_{p,n}$ viewed as a function of $Z \equiv \mathrm{diag}\{z_1, ..., z_k\}$ is positive and monotonically increasing in each of $0 < z_j < 1$, $j = 1, ..., k$, for all $s = s_0 + \delta/\sqrt{p}$ with $(\delta_1, ..., \delta_k)$ from a compact subset of $\mathbb{R}^k$ and all sufficiently large p. This follows from the representation of ${}_1F_0$ and ${}_1F_1$ in the series of zonal polynomials and from the monotonicity of the zonal polynomials of Z in each of $0 < z_j < 1$, $j = 1, ..., k$. Such a monotonicity follows from the fact that zonal polynomials are linear combinations of monomial symmetric functions of z_j with positive coefficients (see Chattopahyay and Pillai (1970), Lemma 2).

Let us make the dependence of $\mathcal{F}_{p,n}$ on Z explicit by writing $\mathcal{F}_{p,n}(Z)$. The positivity and monotonicity of $\mathcal{F}_{p,n}(Z)$ yield the following bound

$$|J| \leq \int_{\tilde{x}_0}^{\tilde{\lambda}_{pk}} \int_{\tilde{\lambda}_{pk}+\varepsilon}^{\tilde{\lambda}_{p,k-1}} \cdots \int_{\tilde{\lambda}_{p3}+\varepsilon}^{\tilde{\lambda}_{p2}} \int_{\tilde{\lambda}_{p2}}^{\tilde{\lambda}_{p2}+\varepsilon} |\omega| \, \mathcal{F}_{p,n}\left(\tilde{Z}_1\right) \prod_{j=1}^{k} \mathrm{d}z_j, \tag{55}$$

where $\tilde{Z}_1 = \mathrm{diag}\left\{\tilde{\lambda}_{p2} + \varepsilon, z_2, ..., z_k\right\}$.

In constrast to $\mathcal{F}_{p,n}(Z)$, function $\mathcal{F}_{p,n}\left(\tilde{Z}_1\right)$ can be approximated using Lemmas 16 and 17. Exploiting such an approximation to show that the right hand side of (55) is asymptotically dominated by $\overline{\mathcal{I}}_{\sigma_\varnothing}$ yields the following lemma. Its proof is given in the SM.

LEMMA 20. *Under the null hypothesis $H_0 : s^{\mathrm{true}} = s_0$, as $p, n \to_\gamma \infty$,*

$$\mathcal{I}_{\sigma_\varnothing} = \overline{\mathcal{I}}_{\sigma_\varnothing} \left(1 + o(1)\right),$$

where $o(1) \to 0$ uniformly over $s = s_0 + \delta/\sqrt{p}$ with $(\delta_1, ..., \delta_k)$ from a compact subset of $\mathbb{R}^k$, a.s.

5.5. *Asymptotic negligibility of $\mathcal{I}_{\sigma_\tau}$ with $\tau \neq \varnothing$, and a summary.* To finish our asymptotic analysis of the joint densities of the eigenvalues, we need to show that integrals $\mathcal{I}_{\sigma_\tau}$ with $\tau \neq \varnothing$ are asymptotically dominated by $\mathcal{I}_{\sigma_\varnothing}$. This can be established similarly to Lemma 20.

Consider, for example, $\tau = \{1, 2, ..., k\}$. By definition,

$$\mathcal{I}_{\sigma_\tau} = \frac{1}{(2\mathrm{i})^k} \int_{\mathcal{K}_{k+1}} \cdots \int_{\mathcal{K}_{k+1}} \omega \mathcal{F}_{p,n}\left(Z\right) \prod_{j=1}^{k} \mathrm{d}z_j.$$

Since $\mathcal{K}_{k+1}$ is not a subset of $\mathbb{R}$, we cannot use Lemmas 16 and 17 to approximate $\mathcal{F}_{p,n}(Z)$. However, it is easy to obtain an upper bound on $|\mathcal{F}_{p,n}(Z)|$ that can be approximated using those lemmas.

Indeed, the representation of ${}_1F_0$ and ${}_1F_1$ in the series of zonal polynomials and the fact that these polynomials are linear combinations of monomial symmetric functions with positive coefficients (see Chattopahyay and Pillai (1970), Lemma 2) yield the following inequality

$$|\mathcal{F}_{p,n}(Z)| \le \mathcal{F}_{p,n}(|Z|),$$

where $|Z| = \operatorname{diag}\{|z_1|, ..., |z_k|\}$. Therefore, we have

$$|\mathcal{I}_{\sigma_\tau}| \le \frac{1}{2^k} \int_{\mathcal{K}_{k+1}} ... \int_{\mathcal{K}_{k+1}} |\omega| \, \mathcal{F}_{p,n}(\tilde{x}_0 I_k) \prod_{j=1}^{k} |\mathrm{d}z_j| \, .$$

Further, by the monotonicity of zonal polynomials, $\mathcal{F}_{p,n}(\tilde{x}_0 I_k) \le \mathcal{F}_{p,n}(Z_\eta)$, where $Z_\eta = \operatorname{diag}\{\tilde{x}_0 + k\eta, ..., \tilde{x}_0 + 2\eta, \tilde{x}_0 + \eta\}$ and η is a fixed small positive number. Therefore,

$$|\mathcal{I}_{\sigma_\tau}| \le \frac{1}{2^k} \int_{\mathcal{K}_{k+1}} ... \int_{\mathcal{K}_{k+1}} |\omega| \, \mathcal{F}_{p,n}(Z_\eta) \prod_{j=1}^{k} |\mathrm{d}z_j| \, .$$

Function $\mathcal{F}_{p,n}(Z_\eta)$ can now be approximated using Lemmas 16 and 17, which yields the following lemma (see the SM for a proof).

LEMMA 21. *Under the null hypothesis $H_0 : s^{\mathrm{true}} = s_0$, as $p, n \to_\gamma \infty$, for any $\tau \neq \varnothing$,*

$$\mathcal{I}_{\sigma_\tau} = o(1)\mathcal{I}_{\sigma_\varnothing},$$

where $o(1) \to 0$ uniformly over $s = s_0 + \delta/\sqrt{p}$ with $(\delta_1, ..., \delta_k)$ from a compact subset of $\mathbb{R}^k$, a.s.

In conclusion of this section, we formulate a theorem that describes the asymptotic behavior of the joint density of the eigenvalues of the multivariate Beta matrix $(n_H\mathbf{H} + n_2\mathbf{E})^{-1} n_H\mathbf{H}$ by combining results of Lemma 12 with those of Lemmas 18–21. For the reader's convenience, we reproduce the definitions of the quantities used in the statement of the theorem in Table 3.

THEOREM 22. *Under the null hypothesis $H_0 : s^{\mathrm{true}} = s_0$, as $p, n \to_\gamma \infty$ while $p - k + 1$ remains even,*

$$\begin{aligned} f^{\mathrm{Case}}(\tilde{\Lambda}; s) &= k! C_{p,n}^{\mathrm{Case}}(\tilde{\Lambda}) k_{p,n}^{\mathrm{Case}}(s) b^{\mathrm{Case}}(s) \\ &\quad \times \prod_{j=1}^{k} \left[\Omega_j^{\mathrm{Case}} (pH_j/\pi)^{-1/2} \prod_{s=k+1}^{p} (\tilde{\lambda}_{pj} - \tilde{\lambda}_{ps})^{-1/2} \right] (1 + o(1)), \end{aligned}$$

where $o(1) \to 0$ *uniformly over* $s = s_0 + \delta/\sqrt{p}$ *with* $(\delta_1, ..., \delta_k)$ *from a compact subset of* $\mathbb{R}^k$*, a.s.*

TABLE 3
Definitions of $C_{p,n}^{\text{Case}}(\tilde{\Lambda})$, $k_{p,n}^{\text{Case}}(s)$, $b^{\text{Case}}(s)$, Ω_j^{Case}, *and* H_j.

Quantity		Definition
$C_{p,n}^{\text{Case}}(\tilde{\Lambda})$	$=$	A quantity that depends on 'Case', n_H, n_2, p and $\tilde{\Lambda}$, but not on s
$k_{p,n}^{\text{SigD}}(s)$	$=$	$[\det(I_k + s)]^{\frac{p-k-n_H-1}{2}} [\det s]^{-\frac{p-k-1}{2}}$,
$k_{p,n}^{\text{REG}}(s)$	$=$	$\text{etr}\{-n_H s/2\} [\det s]^{-\frac{p-k-1}{2}}$,
$b^{\text{SigD}}(s)$	$=$	$\frac{\Gamma_k(k/2)}{\pi^{k(k+1)/4} m^{k(k-1)/4}} \prod_{i<j}^k \left(\frac{(1+s_i)(1+s_j)}{s_i - s_j}\right)^{1/2}$,
$b^{\text{REG}}(s)$	$=$	$\left(\frac{n_H}{2}\right)^{-k(k-1)/4} \frac{\Gamma_k(k/2)}{\pi^{k(k+1)/4}} \frac{b^{kn_H b + k(k+1)/4}}{a^{kn_H a + k(k+1)/4}} \prod_{i<j} (s_i - s_j)^{-1/2}$,
Ω_j^{SigD}	$=$	$\left(1 - \frac{s_j}{1+s_j} \tilde{\lambda}_{pj}\right)^{-\frac{N-p+2}{2}}$,
Ω_j^{REG}	$=$	$e^{n_H \bar{z}_{j+}} \frac{(\bar{z}_{j+} + a)^{an_H}}{(\bar{z}_{j+} + b)^{bn_H}} \frac{\gamma_1 + \gamma_2 + \gamma_2 s_{0j}}{((\gamma_1 + \gamma_2 + \gamma_2 s_{0j})^2 - \rho^2 \gamma_1)^{1/2}}$,
H_j	$=$	$H_{j,\text{num}}/H_{j,\text{den}}$
$H_{j,\text{num}}$	$=$	$(1-\gamma_2) s_{0j} (\gamma_1 + \gamma_2 + \gamma_2 s_{0j}) \left(1 + s_{0j} - \beta_+^{1/2}\right) \left(1 + s_{0j} - \beta_-^{1/2}\right)$
$H_{j,\text{den}}$	$=$	$2\gamma_1\gamma_2 (s_{0j} - \gamma_2 s_{0j} - \gamma_2)(1 + s_{0j})(\gamma_1 + s_{0j})$

6. Local Asymptotic Normality. Our goal is to understand the asymptotic behavior of the likelihood ratios for the eigenvalues of the **F**-ratio (or, equivalently, of the multivatiate Beta) at local alternatives to the null of k supercritical spikes:

$$H_0 : s^{\text{true}} = s_0, \; H_1 : s^{\text{true}} = s_0 + \delta/\sqrt{p}.$$

Let us, first, reparametrize the alternative by considering new local parameters

$$\theta_j = \delta_j / w(s_{0j}) \text{ for } j = 1, ..., k,$$

where $w(s_{0j})$ is the component of the asymptotic variance of λ_{pj} defined in Theorem 10. That is,

$$w(s_{0j}) = \begin{cases} 2\left(\rho s_{0j}(1+s_{0j})/d(s_{0j})\right)^2 & \text{for SigD,} \\ 2\left(\rho s_{0j}(1+s_{0j})/d(s_{0j})\right)^2 - 2\gamma_1 s_{0j}^2 t(s_{0j}) & \text{for REG} \end{cases}$$

with

$$\begin{aligned} t(s_{0j}) &= (1-\gamma_2)(s_{0j} - \underline{s})(s_{0j} - \bar{s})/d^2(s_{0j}) \text{ and} \\ d(s_{0j}) &= (1-\gamma_2)s_{0j} - \gamma_2. \end{aligned}$$

Let $\Lambda = \mathrm{diag}\{\lambda_{p1}, ..., \lambda_{pp}\}$. Denote the likelihood ratio as

$$L_p(\theta, \Lambda) \equiv f(\tilde{\Lambda}; s)/f(\tilde{\Lambda}; s_0).$$

We write $L_p(\theta, \Lambda)$ instead of $L_p(\theta, \tilde{\Lambda})$ to emphasize the fact that the likelihood ratio remains the same whether we define it as the ratio of the joint densities of the eigenvalues of the multivariate Beta $(n_H\mathbf{H} + n_2\mathbf{E})^{-1} n_H\mathbf{H}$ (the diagonal elements of $\tilde{\Lambda}$) or of the eigenvalues of the $\mathbf{F}$-ratio (the diagonal elements of Λ).

Using Theorem 22 to express $\ln L_p(\theta, \Lambda)$ in terms of elementary functions of θ, expanding the result in the powers of $\theta p^{-1/2}$ up to and including terms with $\theta^2 p^{-1}$, and invoking Theorem 4 yields the following theorem. Its proof can be found in the SM.

THEOREM 23. *Under the null hypothesis $H_0 : s^{\mathrm{true}} = s_0$, as $p, n \to_\gamma \infty$ while $p - k + 1$ remains even,*

$$\ln L_p(\theta, \Lambda) = \sum_{j=1}^{k} \left\{ \theta_j \sqrt{p}(\lambda_{pj} - l_j) - \frac{1}{2}\theta_j^2 \tau^2(s_{0j}) \right\} + o_{\mathrm{P}}(1),$$

where $l_j = (s_{0j} + c_1)(s_{0j} + 1)/((1 - c_2)s_{0j} - c_2)$, $\tau^2(s_{0j}) = w(s_{0j})t(s_{0j})$, and $o_{\mathrm{P}}(1) \to 0$ in probability, uniformly in $(\theta_1, ..., \theta_k)$ from any compact subset of $\mathbb{R}^k$.

Theorem 23 together with the joint asymptotic normality of $\sqrt{p}\,(\lambda_{pj} - l_j)$, $j = 1, ..., k$, established in Theorem 10 imply, via Le Cam's First Lemma (see van der Vaart (1998), p.88), that the sequences of the probability measures $\{\mathbb{P}_{s_0,p}\}$ and $\{\mathbb{P}_{s,p}\}$ describing the joint distribution of the eigenvalues of $\mathbf{F}$ under the null $H_0 : s^{\mathrm{true}} = s_0$ and under the local alternative $H_1 : s^{\mathrm{true}} = s_0 + \theta w/\sqrt{p}$, where

$$\theta = \mathrm{diag}\{\theta_1, ..., \theta_k\} \text{ and } w = \mathrm{diag}\{w(s_{01}), ..., w(s_{0k})\},$$

are mutually contiguous. Moreover, the experiments

$$\mathcal{E}_{\theta,s_0,p} \equiv \left((\lambda_{p1}, ..., \lambda_{pp}) \sim \mathbb{P}_{s_0+\theta w/\sqrt{p},p} : \theta \in \mathbb{R}^k \right)$$

converge to the Gaussian shift experiment

$$\mathcal{E}_{\theta,s_0} \equiv \left(Y \sim N\left(\mu\left(\theta\right), \mathcal{T}\right) : \theta \in \mathbb{R}^k \right),$$

where

$$\begin{aligned} Y &= \sqrt{p}(\lambda_{p1} - l_1, ..., \lambda_{pk} - l_k)', \\ \mu(\theta) &= (\theta_1 \tau^2(s_{01}), ..., \theta_k \tau^2(s_{0k})),' \text{ and} \\ \mathcal{T} &= \operatorname{diag}\{\tau^2(s_{01}), ..., \tau^2(s_{0k})\}. \end{aligned}$$

In particular, these experiments are LAN.

As discussed in the introduction, the LAN property of the experiments $\mathcal{E}_{\theta,s_0,p}$ imply that the asymptotically efficient tests of hypotheses about super-critical spikes are based on $\lambda_{p1}, ..., \lambda_{pk}$. Such tests may ignore information contained in the other eigenvalues of the $\mathbf{F}$-ratio.

Here, we will illustrate the LAN property by constructing LAN confidence sets for s. The likelihood ratio confidence set, CS_{LR}, is the set of all s that are not rejected by the likelihood ratio test. Asymptotically, this will coincide with the set of all s_0 such that the hypothesis $H_0 : \theta = 0$ is not rejected in the limiting experiement $\mathcal{E}_{\theta,s_0}$. Therefore, we find the asymptotic $100(1-\alpha)\%$ CS_{LR} for supercritical spikes, by collecting all s_0 that satisfy the inequality

$$\sum_{j=1}^{k} \frac{p(\lambda_{pj} - l_j(s_{0j}))^2}{\hat{\tau}^2(s_{0j})} \leq \chi^2_\alpha(k),$$

where $\chi^2_\alpha(k)$ is the critical value of the chi-squard distribution with k degrees of freedom and the other quantities are as defined in Table 4.

TABLE 4
Definitions of $l(x)$, $\hat{\tau}(x)$ and related quantities.

Quantity		Definition
$l(x)$	$=$	$(x + c_1)(x + 1)/((1 - c_2)x - c_2)$,
$\hat{\tau}^2(x)$	$=$	$\hat{w}(x)\hat{t}(x)$,
$\hat{w}(x)$	$=$	$2(rx(1 + x)/\hat{d}(x))^2$ for SigD,
$\hat{w}(x)$	$=$	$2(rx(1 + x)/\hat{d}(x))^2 - 2c_1x^2\hat{t}(x)$ for REG,
$\hat{t}(x)$	$=$	$(1 - c_2)(x - \underline{\hat{s}})(x - \hat{\bar{s}})/\hat{d}^2(x)$,,
$\hat{d}(x)$	$=$	$(1 - c_2)x - c_2$,
$\underline{\hat{s}}$	$=$	$-(c_1 + c_2)/(c_2 + r)$,
$\hat{\bar{s}}$	$=$	$(c_2 + r)/(1 - c_2)$,
r^2	$=$	$c_1 + c_2 - c_1c_2$.

Figure 4 shows the 95% asymptotic confidence sets for (s_1, s_2) when $c_1 = c_2 = 1/2$, $p = 100$, and $\lambda_{p1} = 45, \lambda_{p2} = 25$. The outer and inner ovals represent the confidence sets for SigD and REG cases, respectively. It is worth noting that the asymptotic confidence sets do not necessarily preserve the ranking $s_1 > s_2$. Indeed, in the figure, the confidence set for the SigD case intersects the 45-degree (dashed) line. Of course, this undesirable phenomenon will not be observed for sufficiently large p.

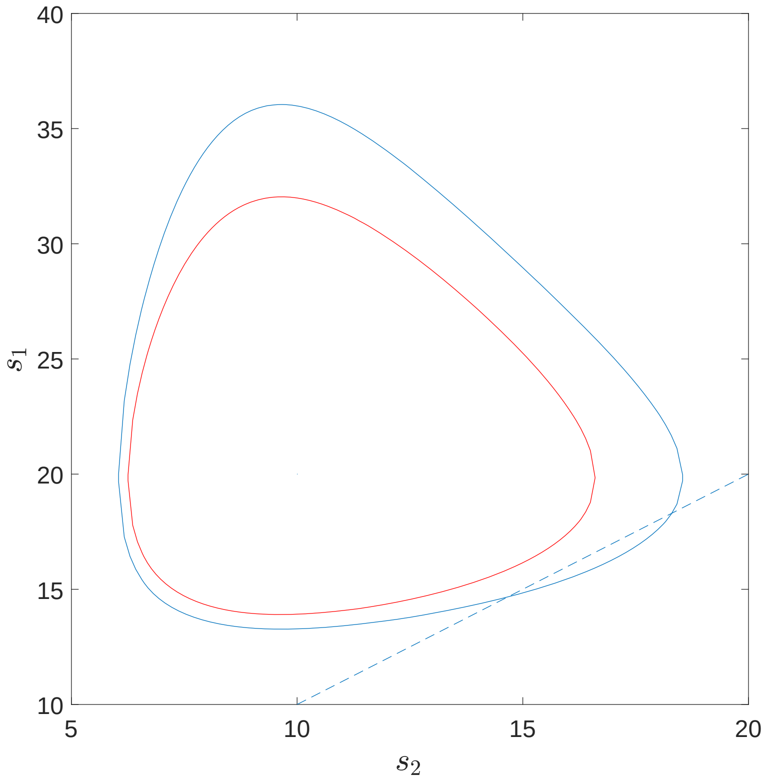

FIG 4. *Confidence sets.*

To assess the quality of the *LAN* confidence sets, we conduct a small-scale Monte Carlo experiment. Specifically, we generate 10,000 replications of $\mathbf{H}$ and $\mathbf{E}$, distributed as

$$(n_1 + k)\mathbf{H} \sim W_p\left(n_1 + k, \Sigma_1, \Omega_1\right) \quad \text{and} \quad n_2\mathbf{E} \sim W_p\left(n_2, I\right)$$

and compute the eigenvalues of the corresponding $\mathbf{F}$-ratio, $\mathbf{E}^{-1}\mathbf{H}$. We consider various values of n_1, n_2, and p, and several different values of the spikes.

The Monte Carlo coverage rates of the nominal 95% confidence sets are reported in Table 5 for $k = 1$ and Table 6 for $k = 2$.

TABLE 5
Coverage probabilities, nominal 95% confidence sets. Single spike.

$k=1$		SigD			REG		
		$s=3$	$s=5$	$s=10$	$s=3$	$s=5$	$s=10$
$c_2=0, n_1=p=100$	$\hat{\bar{s}}=1.0$	95.8	95.1	95.5	95.1	94.9	95.3
$n_1=n_2=100, p=50$	$\hat{\bar{s}}=2.7$	79.1	94.2	94.3	79.6	94.4	94.3
$n_1=n_2=100, p=5$	$\hat{\bar{s}}=0.4$	94.9	94.9	95.1	94.2	95.0	94.6
$n_1=n_2=100, p=2$	$\hat{\bar{s}}=0.2$	95.0	95.3	95.3	94.9	95.1	94.5

TABLE 6
Coverage probabilities, nominal 95% confidence sets. Two spikes.

		SigD				
$k=2$		$s_1=4$	$s_1=5$	$s_1=10$	$s_1=6$	$s_1=20$
		$s_2=3$	$s_2=3$	$s_2=3$	$s_2=5$	$s_2=10$
$c_2=0, n_1=p=100$	$\hat{\bar{s}}=1.0$	95.9	95.9	95.6	95.4	95.6
$n_1=n_2=100, p=50$	$\hat{\bar{s}}=2.7$	68.1	75.4	80.8	92.3	94.9
$n_1=n_2=100, p=5$	$\hat{\bar{s}}=0.4$	94.8	95.3	94.3	94.1	95.2
$n_1=n_2=100, p=2$	$\hat{\bar{s}}=0.2$	94.7	95.4	95.1	94.0	94.9
		REG				
$k=2$		$s_1=4$	$s_1=5$	$s_1=10$	$s_1=6$	$s_1=20$
		$s_2=3$	$s_2=3$	$s_2=3$	$s_2=5$	$s_2=10$
$c_2=0, n_1=p=100$	$\hat{\bar{s}}=1.0$	95.5	95.5	95.3	95.5	95.1
$n_1=n_2=100, p=50$	$\hat{\bar{s}}=2.7$	70.4	76.1	80.5	92.2	94.2
$n_1=n_2=100, p=5$	$\hat{\bar{s}}=0.4$	94.8	95.5	95.1	93.9	94.5
$n_1=n_2=100, p=2$	$\hat{\bar{s}}=0.2$	95.0	95.4	94.4	94.0	94.4

Surprisingly, for $k = 1$, the coverage rate of the nominal 95% confidence intervals remains good even for an extremely small dimensionality – $p = 2$. The worst reported coverage corresponds to $n_1 = n_2 = 100, p = 50$ and $s = 3$. This is an example of a situation where the true spike is close to the phase transition threshold, given here by $\hat{\bar{s}} \equiv (c_2 + r)/(1 - c_2) \approx 2.7$. A further analysis suggests that the reason for the poor coverage in this situation is that the finite sample variance of λ_{p1} is substantially larger than its asymptotic counterpart.

For $k = 2$, the results are similar. Again, surprisingly, the coverage remains good even when $p = 2$. The worst results correspond to situations where a spike is close to the phase transition. A particularly unfavourable situation arises when both spikes are close to the threshold – $s_1 = 4, s_2 = 3$ with $\hat{\bar{s}} = 2.7$. A more detailed analysis shows that in such a case the asymptotic variances are smaller than the finite sample ones. In addition, the smallest of the spikes tends to lie below the corresponding a.s. limit, whereas the largest one tends to lie above the corresponding a.s. limit. These deviations become smaller when at least one of the spikes is moved away from the phase transition threshold.

7. Conclusion. In this paper, we establish the Local Asymptotic Normality of the experiments of observing the eigenvalues of the F-ratio $\mathbf{F} \equiv \mathbf{E}^{-1}\mathbf{H}$ of two large-dimensional Wishart matrices. The experiments are parameterized by the values of a finite number k of spikes that describe the "ratio" of the covariance parameters of $\mathbf{H}$ and $\mathbf{E}$, or, in the case of equal covariance parameters, the non-centrality parameter of $\mathbf{H}$.

We find that the asymptotic behavior of the log ratio of the joint density of the eigenvalues of $\mathbf{F}$, which corresponds to super-critical spikes, to their joint density under a local deviation from these values depends only on the k of the largest eigenvalues $\lambda_{p1}, ..., \lambda_{pk}$. This implies, in particular, that the best statistical inference about k super-critical spikes in the local asymptotic regime is based on the k largest eigenvalue only. A small-scale Monte Carlo analysis shows that LAN confidence sets for super-critical spikes have good coverage properties even for extremely small values of the dimensionality p.

As a by-product of our analysis, we establish the joint asymptotic normality of a few of the largest eigenvalues of $\mathbf{F}$ that correspond to the super-critical spikes. We derive explicit formulas for the phase transition threshold, for the almost sure limits of the super-critical eigenvalues, and for the asymptotic variances of their fluctuations around these limits.

References.

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

[2] Bai, Z. D. and Silverstein, J. W. (2004). CLT for Linear Spectral Statistics of Large-Dimensional Sample Covariance Matrices. *Annals of Probability* **32** 553–605.

[3] Bai, Z.D. and Yao, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques* **44** 447–474.

[4] Baik, J., Ben Arous, G. and Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability* **33** 1643–1697.

[5] Bao, Z., Hu, J., Pan, G. and Zhou, W. (2014). Canonical correlation coefficients of high-dimensional normal vectors: finite rank case. *arXiv 1407.7194*

[6] Benaych-Georges, F., Guionnet, A. and Maida, M. (2011). Fluctuations of the Extreme Eigenvalues of Finite Rank Deformations of Random Matrices. *Electronic Journal of Probability* **16** 1621–1662.

[7] Bilodeau, M. and Srivastava, M. S. (1992). Estimation of the Eigenvalues of $\Sigma_1\Sigma_2^{-1}$. *Journal of Multivariate Analysis* **41** 1–13.

[8] Bloemendal, A. and Virág, B. (2011). Limits of spiked random matrices II. *arXiv: 1109.3704v1.*

[9] Bloemendal, A. and Virág, B. (2013). Limits of spiked random matrices I. *Probability Theory and Related Fields* **156** 795–825.

[10] Capitaine, M., |textscDonati-Martin, C. and Féral, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. *The Annals of Probability* **37** 1–47.

[11] Chang, T.C. (1970). On an Asymptotic Representation of the Distribution of the Characteristic Roots of $S_1S_2^{-1}$. *The Annals of Mathematical Statistics* **41** (2), 440–445.

[12] Chattopahyay, A.K. and Pillai, K.C.S. (1970). On the maximization of an integral of a matrix function over the group of orthogonal matrices. Mimeo Series No. 248, Department of Statistics, Purdue University.

[13] Dharmawansa, P. and Johnstone, I. M. (2014). Joint density of eigenvalues in spiked multivariate models. *Stat* **3** 240–249.

[14] Dharmawansa, P., Johnstone, I.M. and Onatski, A. (2017). Supplementary material for "Local Asymptotic Normality of the spectrum of high-dimensional spiked F-ratios". *manuscript*

[15] Féral, D. and Péché, S. (2009). The largest eigenvalues of sample covariance matrices for a spiked population: diagonal case. *Journal of Mathematical Physics* **50** 073302.

[16] Glynn, W.J. (1980). Asymptotic representations of the densities of canonical correlations and latent roots in MANOVA when the population parameters have arbitrary multiplicity. *The Annals of Statistics* **8** 958–976.

[17] Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis.* Cambridge University Press.

[18] James, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Annals of Mathematical Statistics* **35** 475–501.

[19] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics* **29**(2) 295–327.

[20] Khatri, C. G. (1967). Some distributional problems associated with the characteristics roots of $S_1S_2^{-1}$. *Ann. Math. Stat.* **38**, no. 3, 944–948.

[21] Leung, P. L., and Muirhead, R. (1987). Estimation of Prameter Matrices and Eigenvalues in MANOVA and Canonical Correlation Analysis. *Annals of Statistics* **15** 1651–1666.

[22] Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley.

[23] Mo, M.Y. (2012). The rank 1 real Wishart spiked model. *Communications on Pure and Applied Mathematics* **65** 1528–1638.

[24] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* John Wiley &

Son, Hoboken, New Jersey.

[25] Muirhead, R.J. and Verathaworn, T. (1985). On estimating the latent roots of $\Sigma_1\Sigma_2^{-1}$. in Krishnaiah, P.R. (ed.) *Multivariate Analysis - VI*, Elsevier Science Publishers B.V., 431–447.

[26] Nadakuditi, R. R. and Silverstein, J. W. (2010). Fundamental Limit of Sample Generalized Eigenvalue Based Detection of Signals in Noise Using Relatively Few Signal-Bearing and Noise-Only Samples. *IEEE Journal of Selected Topics in Signal Processing* **4** 468–480.

[27] Onatski, A. (2007). Asymptotics of the principal components estimator of large factor models with weak factors and i.i.d. Gaussian noise. *manuscript*, University of Cambridge.

[28] Onatski, A. (2013). Detection of weak signals in high-dimensional complex-valued data. *Random Matrices: Theory and Applications* **3** (1).

[29] Paul, D. (2007). Asymptotics of the leading sample eigenvalues for a spiked covariance model. *Statistica Sinica* **17** 1617–1642.

[30] Passemier, D., McKay, M. R. and Chen, Y. (2014). Asymptotic Linear Spectral Statistics for Spiked Hermitian Random Matrix Models. *arXiv:1402.6419v1*

[31] Sheena, Y., Gupta, A. K. and Fujikoshi, Y. (2004). Estimation of the eigenvalues of noncentrality parameter in matrix variate noncentral beta distribution. *Annals of the Institute of Statistical Mathematics* **56** 101–125.

[32] Silverstein, J. W. (1985). The Limiting Eigenvalue Distribution of a Multivariate F Matrix. *SIAM Journal of Mathematical Analysis* **16** 641–646.
bibitemr33 Silverstein, J. W. and Bai, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis* **54** 175–192.

[33] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

[34] Wachter, K. (1980). The limiting empirical measure of multiple discriminant ratios. *The Annals of Statistics* **8** 937–957.

[35] Wang, Q. and Yao, J. (2015). Extreme eigenvalues of large-dimensional spiked Fisher matrices with application. *ArXiv: 1504.05087*.

[36] Wilks, S. S. (1962). *Mathematical Statistics*. Wiley, New York.

[37] Yin, Y. Q., Bai, Z. D. and Krishnaiah, P. R. (1983). Limiting Behavior of the Eigenvalues of a Multivariate F Matrix. *Journal of Multivariate Analysis* **13** 508–516.

E-mail: prathapa@uom.lk E-mail: imj@stanford.edu

E-mail: ao319@cam.ac.uk

SUPPLEMATARY MATERIAL FOR "LOCAL ASYMPTOTIC NORMALITY OF THE SPECTRUM OF HIGH-DIMENSIONAL SPIKED F-RATIOS."

By Prathapasinghe Dharmawansa Iain M. Johnstone and Alexei Onatski

This note contains supplementary material for Dharmawansa et al (2016) (DJO in what follows). It is lined up with sections in the main text to make it relatively easy to see how and where the proof details fit in.

CONTENTS

1. Introduction.

There is no supplementary material for the Introduction section of DJO.

2. Setup.

There is no supplementary material for the Setup section of DJO.

3. Phase transition and almost sure limits.

3.1. *An extension of Bai and Silverstein's (1998) Theorem 1.1.* In this subsection, we will use notations from Bai and Silverstein (1998). Their Theorem 1.1 covers only nonrandom T_n. Remark 6.5 on page 125 of Bai and Silverstein (2010) points out that the theorem is easily extendable to the cases of random T_n, as long as it is independent from X_n, its limiting spectral distribution is nonrandom, and condition (f) of the theorem that "interval $[a,b]$ with $a>0$ lies outside the suport of $F^{c,H}$ and F^{c_n,H_n} for all large n" holds a.s. (here Bai and Silverstein's c and c_n correspond to our γ_1 and c_1, respectively.)

This still assumes that $\|T_n\|$ is bounded (assumption (d) of the theorem). Unfortunately, if $T_n = E^{-2}$, where E is a sample covariance matrix, as in DJO, $\|T_n\|$ can be larger than any positive number with a small probability. Note however, that as long as $\gamma_2 \in (0,1)$, $\|E^{-2}\|$ a.s. converges to a finite number. This helps because, as the following lemma shows, Theorem 1.1 remains valid if assumption (d) is replaced by

(d*) There exists $C > 0$, such that $\limsup_{p,n\to_\gamma\infty} \|T_n\| < C$, a.s.

LEMMA 1. *Assumption (d) of Theorem 1.1 can be replaced by (d*) without changing the validity of the theorem.*

Proof: Define event

$$\Omega = \left\{ \left\| E^{-1} \right\| < C \text{ for all } p > p_0 \right\}.$$

By (d*), $\Pr(\Omega) \to 1$ as $p_0 \to \infty$. On the other hand, Theorem 1.1 holds conditionally on Ω. That is, the conditional probability $\Pr(\Omega_1|\Omega) = 1$, where Ω_1 is the event that there exists p_1 s.t. for any $p > p_1$, all the eigenvalues of $E^{-1}H$ do not belong to $[a,b]$. Therefore, $\Pr(\Omega_1 \cap \Omega) = \Pr(\Omega) \to 1$ as $p_0 \to \infty$. But $\Pr(\Omega_1) \geq \Pr(\Omega_1 \cap \Omega)$. Hence, we must have $\Pr(\Omega_1) = 1$. $\square$

Lemma 1 implies that the largest eigenvalue of $E^{-1}H$ a.s. converges to β_+.

3.2. *Proof of Lemma DJO1 about the convergence of* $\operatorname{tr}(H - xE)^{-1}/p$*.* Let $x \in \mathbb{R}$ be such that $x > \beta_+$, and let $\hat{F}(\lambda; x)$ be the empirical distribution function of the eigenvalues of $H - xE$. For any $z \in \mathbb{C}^+$, let

$$\hat{m}(z;x) = \int (\lambda - z)^{-1} \, \mathrm{d}\hat{F}(\lambda; x)$$

be the Stieltjes transform of $\hat{F}(\lambda; x)$. Note that matrix $H - xE$ can be represented in the form YTY'/p, where Y is a $p \times (n_1 + n_2)$ matrix with i.i.d. standard normal entries and T is a diagonal matrix with the first n_1 and the last n_2 diagonal elements equal to c_1 and $-xc_2$, respectively. Therefore, by Theorem 1.1 of Silverstein and Bai (1995), for any $z \in \mathbb{C}^+$, $\hat{m}(z;x)$ a.s. converges to $m(z;x) \in \mathbb{C}^+$, which is an analytic function in the domain $z \in \mathbb{C}^+$ that solves the functional equation (DJO13).

By Lemma 1, the largest eigenvalue of $E^{-1}H$ a.s. converges to β_+. Therefore, for any $x > \beta_+$, the largest eigenvalue of $H - xE$ is a.s. asymptotically bounded away from the positive semi-axis. Hence, $\hat{m}(z;x)$ is analytic and bounded in a small disc D around $z = 0$ for all sufficiently large p and n, a.s. By Vitali's theorem (see Titchmarsh (1939), p. 168), $\hat{m}(z;x)$ is a.s. converging to an analytic function in D. Since, in $D \cap \mathbb{C}^+$, the limiting function is $m(z;x)$, we have

$$\operatorname{tr}(H - xE)^{-1}/p = \hat{m}(0;x) \stackrel{a.s.}{\to} m(0;x),$$

where $m(0;x) = \lim_{z\to 0} m(z;x)$. Further, $\operatorname{tr}(H - \zeta E)^{-1}/p$ is an analytic bounded function of ζ in a small disk D_x around x, for all sufficiently large p and n, a.s. Therefore, by Vitali's theorem its a.s. limit $f(\zeta)$ is analytic in D_x, and

$$\frac{\mathrm{d}}{\mathrm{d}\zeta} \operatorname{tr}(H - \zeta E)^{-1}/p \stackrel{a.s.}{\to} \frac{\mathrm{d}}{\mathrm{d}\zeta} f(\zeta)$$

in D_x. On the other hand, we know that $f(\zeta) = m(0;\zeta)$ for $\zeta \in \mathbb{R} \cap D_x$. Therefore, we have (DJO12).

3.3. *Proof of Lemma DJO2 about the asymptotic proportionality of $M(x)$ and $s+\gamma_1 I_k$.* The convergences stated in Lemma DJO2 follow from (DJO5), (DJO9), and Lemma 11 stated below. □

LEMMA 2. *Let C be a random $p\times p$ matrix, independent from u and v, which are as defined in Section DJO2, and such that $p\,\|C\|$ is bounded for all sufficiently large p, a.s. Then, as $p\to\infty$,*

$$\left\|v'Cv-(\operatorname{tr}C)\,I_k\right\|\overset{a.s.}{\to}0 \text{ and } \left\|v'Cu\right\|\overset{a.s.}{\to}0.$$

Proof: This lemma follows from the Borel-Cantelli lemma, and the upper bounds on the fourth moments of the entries $v'Cv-(\operatorname{tr}C)\,I_k$ and $v'Cu$ established by Lemma 2.7 of Bai and Silverstein (1998). □

3.4. *Proof of Lemma DJO3 about properties of functions $M(x)$ and $m(0;x)$.* Let $\mu_1\in(0,\infty)$ be the largest eigenvalue of $E^{-1}H$. For any $x_1>x_2>\mu_1$, matrix $(H-x_1E)^{-1}-(H-x_2E)^{-1}$ is positive definite, a.s. Part (i) follows from this, from the definition (DJO9) of $M(x)$, and from the fact that $\mu_1\overset{a.s.}{\to}\beta_+$. Part (i) together with Lemmas DJO1 and DJO2 imply that $m(0;x)$ is increasing on (β_+,∞). It is strictly increasing because, otherwise, equation (DJO13) would not be satisfied for some $z\in\mathbb{C}^+$ that are sufficiently close to zero. The continuity follows from the analyticity of $m(0;x)$ established in the proof of Lemma DJO1. Finally, $\lim_{x\to\infty}m(0;x)=0$ is implied by (ii) and (DJO11). Equation (DJO13) implies that

$$\lim_{x\downarrow\beta_+}m(0;x)=(\gamma_2-1)/[(\rho+1)\rho],$$

which, in its turn, implies the second statement of (iii).

4. Asymptotic normality.

4.1. *Proof of Lemma DJO5 about $\sigma(l_i)$ and $\delta(l_i)$.* By Lemmas DJO1 and DJO2,

$$\frac{\mathrm{d}}{\mathrm{d}x}M(l_i)\overset{a.s.}{\to}(s+\gamma_1 I_k)\,\frac{\mathrm{d}}{\mathrm{d}x}m(0;x_i). \tag{1}$$

Further,

$$(1+M_{ii}(l_i))\,(I_k+M(l_i))^{-1}\overset{a.s.}{\to}\operatorname{diag}\{0,...,0,1,0,...,0\} \tag{2}$$

with 1 at the i-th place on the diagonal. The latter convergence follows from the fact that $I_k+M(l_i)$ can be viewed as a small perturbation of a diagonal matrix

$$I_k+(s+\gamma_1 I_k)\,m(0;x_i),$$

which has non-zero diagonal elements, except at the i-th position. The eigenvalue perturbation formulae (see, for example, (2.33) on p.79 of Kato (1980)) will then lead to (2). Combining (10) and (2), and using the definition of $\sigma(l_i)$, we obtain (i).

To establish (ii), we note that $(1+M_{ii}(l_i))\operatorname{tr}R\,(l_i)=O_{\mathrm{P}}(1)$ by an argument similar to that used to establish (i). Further, $(\operatorname{tr}J\,(l_i))^2-\operatorname{tr}J^2\,(l_i)$ is a linear function of the only eigenvalue of $J\,(l_i)$ that diverges to infinity. By the eigenvalue perturbation formulae, such an eigenvalue equals $(1+M_{ii}(l_i))^{-1}\,O(1)$ a.s. Therefore,

$$(1+M_{ii}(l_i))\left((\operatorname{tr}J\,(l_i))^2-\operatorname{tr}J^2\,(l_i)\right)=O(1),$$

which concludes the proof of (ii).

4.2. *Proof of Lemma DJO6 that $Z^{(7)}$ a.s. converges to zero.* Recall that

$$Z^{(7)} = (s + c_1 I_k)\sqrt{p}\,(\operatorname{tr} G - m_p(0; l_i))\,, \tag{3}$$

where

$$G = (H - l_i E)^{-1}/p\ ,\ l_i = \frac{(s_i + c_1)(s_i + 1)}{s_i - c_2(s_i + 1)}$$

and $m_p(z; l_i)$ is the Stieltjes transform of the limiting spectral distribution of $H - l_i E$ as n_1, n_2, and p diverge to infinity so that c_1 and c_2 (and thus l_i too) *remain fixed.*

Let $\hat{F}(\lambda; x)$ be the empirical distribution of the eigenvalues of $H - xE$ and $\hat{m}(z; x)$ be its Stieltjes transform. That is,

$$\hat{m}(z; x) \equiv \int (\lambda - z)^{-1} \mathrm{d}\hat{F}(\lambda; x) \equiv \operatorname{tr}(H - xE - zI_p)^{-1}/p.$$

Then, to establish Lemma DJO6, it is sufficient to prove that

$$\sqrt{p}(\hat{m}(0; l_i) - m_p(0; l_i)) \overset{a.s.}{\to} 0. \tag{4}$$

REMARK 3. *By definition, $l_i \to x_i$, where x_i is the a.s. limit of the i-th largest super-critical eigenvalue of $\mathbf{F}$. Therefore, $l_i > \beta_+$ for sufficiently large p, where β_+ is the a.s. limit of the largest eigenvalue of $E^{-1}H$ as $p, n \to_\gamma \infty$. This fact implies that, $H - l_i E$ is a.s. negative definite for sufficiently large p, and $\hat{m}\,(z; l_i)$ amd $m_p\,(z; l_i)$ are well defined for $z = 0$ by analytic continuation form the upper half of the complex plane.*

Equations for m_p. By the definition of H and E,

$$H - l_i E = YTY'/p,$$

where $Y \sim N(0, I_p \otimes I_{n_1+n_2})$ and $T = \operatorname{diag}(t_1, ..., t_{n_1+n_2})$ with

$$t_j = \begin{cases} c_1 \text{ for } j \le n_1, \\ -l_i c_2 \text{ for } n_1 < j \le n_1 + n_2. \end{cases}.$$

Denote the empirical distribution of $\{t_1, ..., t_{n_1+n_2}\}$ as $\hat{\mathcal{T}}(t)$.

Results of Silverstein and Bai (1995) imply that, as $p, n \to_\gamma \infty$, $\hat{F}(\lambda; l_i)$ a.s. weakly converges to a distribution $F^{\gamma,\mathcal{T}}(\lambda)$, whose Stieltjes transform $m(z)$ satisfies equation (compare to equation DJO13)

$$m(z) = -\left(z - \frac{\gamma_1 + \gamma_2}{\gamma_1\gamma_2} \int \frac{t\mathrm{d}\mathcal{T}(t)}{1 + tm(z)}\right)^{-1}, \tag{5}$$

where $\mathcal{T}(t)$ is the limit of $\hat{\mathcal{T}}(t)$. More explicitly, $m(z)$ satisfies

$$m(z) = -\left(z - \frac{1}{1 + \gamma_1 m(z)} + \frac{x_i}{1 - x_i\gamma_2 m(z)}\right)^{-1}. \tag{6}$$

Function $m_p(z) \equiv m_p(z; l_i)$ is the analogue of $m(z)$ uner the fixed $p/n_1 \equiv c_1$ and $p/n_2 \equiv c_2$ asymptotics. That is, it satisfies equation

$$m_p(z) = -\left(z - \frac{c_1 + c_2}{c_1 c_2} \int \frac{t\mathrm{d}\hat{\mathcal{T}}(t)}{1 + tm_p(z)}\right)^{-1}. \tag{7}$$

For future reference, notice that

$$zm_p(z) - \frac{c_1+c_2}{c_1c_2}\int \frac{tm_p(z)\mathrm{d}\hat{\mathcal{T}}(t)}{1+tm_p(z)} + 1 = 0. \tag{8}$$

REMARK 4. *Had matrix $T = \operatorname{diag}(t_1, ..., t_n)$ been positive semi-definite, our 'target' equation (4) would have been following from results of Bai and Silverstein (2004). Our strategy of the proof of (4) will be to extend some of Bai and Silverstein's (2004) analysis to cover T that are not positive semi-definite.*

An upper bound on $|\sqrt{p}\,(\hat{m}\,(0;l_i) - m_p\,(0;l_i))|$. Suppose that $\mathcal{K}$ is a contour in the complex plane that does not encircle zero, but encircles all the eigenvalues of $H - l_iE$ and the support of the limiting (under fixed c_1, c_2 asymptotics) spectral distribution $F^{c,\hat{\mathcal{T}}}$ of $H - l_iE$. Then, we have

$$\oint_{\mathcal{K}} \frac{\hat{m}(z;l_i) - m_p(z;l_i)}{z}\mathrm{d}z = \oint_{\mathcal{K}}\int \frac{\mathrm{d}\hat{F}(\lambda;l_i) - \mathrm{d}F^{c,\hat{\mathcal{T}}}(\lambda)}{z(\lambda - z)}\mathrm{d}z$$

Intechanging the order of the integrals, multiplying by $-1/2\pi\mathrm{i}$, and using Cauchy's residue theorem, we obtain

$$-\frac{1}{2\pi\mathrm{i}}\oint_{\mathcal{K}} \frac{\hat{m}(z;l_i) - m_p(z;l_i)}{z}\mathrm{d}z = \hat{m}(0;l_i) - m_p(0;l_i). \tag{9}$$

To prove the a.s. convergence of $\sqrt{p}(\hat{m}(0;l_i) - m_p(0;l_i))$ to zero, we will analyze the behavior of $\sqrt{p}(\hat{m}(z;l_i) - m_p(z;l_i))$ along $\mathcal{K}$. But first, let us explicitly construct such a contour.

Since l_i converges to x_i, which lies above β_+, we will assume without loss of generality that $l_i \in [x_- x_+]$, where x_+ and x_- are fixed real numbers satisfying the inequality $x_+ > x_- > \beta_+$. Let a_- be the lower bound of the support of the limiting spectral distribution (LSD) of $H - x_+E$ and let a_+ be the upper bound of the support of the LSD of $H - x_-E$. Note that, almost surely, $H - x_+E < H - x_-E < 0$ for sufficiently large p, and thus, $a_- \le a_+ \le 0$. In fact, it easy to see that $a_+ < 0$.

LEMMA 5. $a_+ < 0$.

Proof: Let $A(x) = H - xE$. Consider the following decomposition

$$A(x) = \left(H - \frac{x+\beta_+}{2}E\right) - \frac{x-\beta_+}{2}E \equiv A_1(x) + A_2(x).$$

For the largest eigenvalue of $A_2(x_-)$, we have

$$\lambda_{\max}\left(A_2(x_-)\right) \overset{a.s.}{\to} -\frac{x_- - \beta_+}{2}\left(1-\sqrt{\gamma_2}\right)^2 \equiv \bar{a}_+ < 0. \tag{10}$$

On the other hand, since β_+ is the a.s. limit of the largest eigenvalue of $E^{-1}H$, $A_1(x_-)$ is a.s. negative semi-definite for sufficiently large p. Hence, $\lambda_{\max}\left(A(x_-)\right) < \bar{a}_+$ for sufficiently large p, a.s., which implies that $a_+ < \bar{a}_+ < 0$.□

We say that a sequence of events Q_p occurs with overwhelming probability (w.o.p.) if $\Pr\left(Q_p^c\right) = o\left(p^{-t}\right)$ for each fixed $t > 0$. Often, we will simply say that Q_p occurs w.o.p. omitting the words "the sequence of events".

LEMMA 6. *The sequence of events*

$$Q_p = \{\lambda_{\max}(A(l_i)) < \bar{a}_+,\ \lambda_{\min}(A(l_i)) > -4x_+\}$$

occurs w.o.p.

Proof: Consider the sequene of events

$$\hat{Q}_p = \{\lambda_{\max}(A(x_-)) < \bar{a}_+,\ \lambda_{\min}(A(x_+)) > -4x_+\}.$$

Since $l_i \in [x_-, x_+]$, $\hat{Q}_p \subseteq Q_p$ and it is sufficient to prove that $\hat{Q}_p$ occurs w.o.p.

The decomposition

$$A(x) = \left(H - \frac{x + 3\beta_+}{4}E\right) - \frac{3\,(x - \beta_+)}{4}E$$

and the definition $A(x) = H - xE$ show that event $\hat{Q}_p^c$ implies $\cup_{i=1}^4 R_{pi}$, where

$$\begin{aligned}
R_{p1} &= \left\{\lambda_{\max}\left(-\frac{3\,(x_- - \beta_+)}{4}E\right) \geq \bar{a}_+\right\} = \left\{\lambda_{\min}(E) \leq \frac{2}{3}(1 - \sqrt{\gamma_2})^2\right\},\\
R_{p2} &= \left\{H - \frac{x_- + 3\beta_+}{4}E < 0\right\}^c = \left\{\lambda_{\max}\left(E^{-1}H\right) \geq \frac{x_- + 3\beta_+}{4}\right\},\\
R_{p3} &= \{\lambda_{\min}(-x_+E) \leq 4x_+\} = \{\lambda_{\max}(E) \geq 4\},\\
R_{p4} &= \{H > 0\}^c = \{\lambda_{\min}(H) \leq 0\}.
\end{aligned}$$

The Gaussian concentration inequalities for the largest and smallest singular values of Wishart matrices imply that R_{p1}^c, R_{p3}^c, and R_{p4}^c occur w.o.p. Further, as follows, for example, from the proof of Theorem 11.3.2 in Muirhead (1982), the largest root of the equation

$$\det\{n_1H - x\,(n_1H + n_2E)\} = 0 \tag{11}$$

is distributed as the first squared sample canonical correlation coefficient r_1^2 between columns of Z_1 and Z_2, where Z_1 and Z_2 are independent $p\times(n_1+n_2)$ and $n_1\times(n_1+n_2)$ matrices with independent $N(0,1)$ entries. In the next subsection of this note, we show that such a squared sample canonical correlation coefficient satisfies the following concentration inequality

$$\Pr\left\{r_1^2 > \mathbb{E}r_1^2 + t\right\} \leq 2\exp\left\{-\frac{(n_1+n_2)t^2}{2\times 16^2}\right\},\ t > 0. \tag{12}$$

This probability bound is not the best possible, but sufficient for our purposes. Indeed, note that the largest root of (11) equals

$$\lambda_{\max}\left(E^{-1}H\right)/\left(n_2/n_1 + \lambda_{\max}\left(E^{-1}H\right)\right).$$

This equality, the fact that $\lambda_{\max}\left(E^{-1}H\right)$ a.s. converges to β_+, and the concentration inequality (12) imply that R_{p2}^c occurs w.o.p. Since $\hat{Q}_p$ is implied by $\cap_{i=1}^4 R_{pi}^c$, $\hat{Q}_p$ also occurs w.o.p.□

REMARK 7. *The bounds $-4x_+$ and $\bar{a}_+$ on the smallest and the largest eigenvalues of $A(l_i)$ are rough, but they are sufficient for our purposes.*

Now we are ready to construct contour $\mathcal{K}$. It is the rectangle shown in Figure 1. The contour intersects the real axis at $-5x_+$ and $\bar{a}_+/2$, so that the bounds $-4x_+$ and $\bar{a}_+$ remain inside the contour, but zero lies outside the contour. It is symmetric around the real axis.

Separate, but related, arguments for bounding $\sqrt{p}(\hat{m}(z;l_i) - m_p(z;l_i))$ are needed for the horizontal and vertical segments of the contour, $\mathcal{K}_H$ and $\mathcal{K}_V$ respectively. Small vertical intervals $\mathcal{K}_{0p} = \{z \in \mathcal{K} : |\mathrm{Im}\, z| \le p^{-2}\}$ about the real axis will be excluded from many bounds and handled separately. Accordingly, we write $\mathcal{K}_p$ for $\mathcal{K}\backslash\mathcal{K}_{0p}$ and $\mathcal{K}_{V,p}$ for $\mathcal{K}_V\backslash\mathcal{K}_{0p}$. Without loss of generality, we set $|\mathrm{Im}\, z| = \min\{-\bar{a}_+/2, x_+\} \equiv \eta$ for $z \in \mathcal{K}_H$. The purpose of such a setting is to have a distance between $[-4x_+, \bar{a}_+]$ and γ be bounded from below by η.

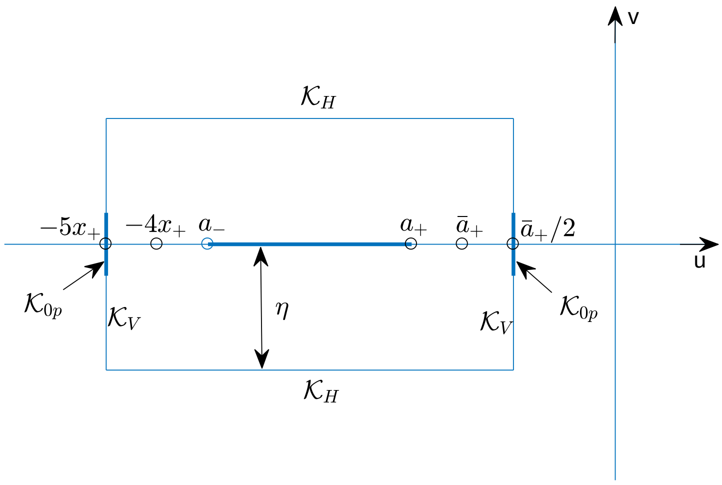

FIG 1. *Contour $\mathcal{K}$ in $u + iv$ plane*

Define 'deterministic' and 'stochastic' terms by

$$I_p^{(D)} = \oint_{\mathcal{K}_p} \sqrt{p}\,|\mathbb{E}\hat{m}(z;l_i) - m_p(z;l_i)|\,\mathrm{d}z \text{ and}$$
$$I_p^{(S)} = \oint_{\mathcal{K}_p} \sqrt{p}\,|\hat{m}(z;l_i) - \mathbb{E}\hat{m}(z;l_i)|\,\mathrm{d}z,$$

and an exceptional term near the real axis by

$$I_p^{(E)} = \oint_{\mathcal{K}_{0p}} \sqrt{p}\,|\hat{m}(z;l_i) - m_p(z;l_i)|\,\mathrm{d}z.$$

Write $\|z^{-1}\|_{\infty,\mathcal{K}} = \sup\{z^{-1} : z \in \mathcal{K}\}$. From (9), we have on event Q_p,

$$|\sqrt{p}\,(\hat{m}\,(0;l_i) - m_p\,(0;l_i))| \le (2\pi)^{-1} \left\|z^{-1}\right\|_{\infty,\mathcal{K}} \left\{I_p^{(D)} + I_p^{(S)} + I_p^{(E)}\right\}. \tag{13}$$

First reduction. Let us show that the proof of the convergence (4) can be reduced to verifying the stochastic bounds

$$\sup_{z\in\mathcal{K}_p} \mathbb{E}\,|\hat{m}\,(z;l_i) - \mathbb{E}\hat{m}\,(z;l_i)|^{2l} \le C_l p^{-2l}, \;\; l = 1, 2. \tag{14}$$

and the deterministic convergence

$$\sup_{z\in\mathcal{K}_p}\sqrt{p}\left|\mathbb{E}\hat{m}\left(z;l_i\right)-m_p\left(z;l_i\right)\right|\to 0. \tag{15}$$

Note that (15) implies that $I_p^{(D)}\to 0$.

For the stochastic bounds, write $|\mathcal{K}_p|$ for the length of $\mathcal{K}_p$, and make use of Hölder's inequality for $\left(I_p^{(S)}\right)^4$ and then (14) to bound

$$\begin{aligned}\Pr\left(I_p^{(S)}>\varepsilon\right) &\le \varepsilon^{-4}\mathbb{E}\left(I_p^{(S)}\right)^4\\ &\le \varepsilon^{-4}\left|\mathcal{K}_p\right|^3\oint_{\mathcal{K}_p}p^2\mathbb{E}\left|\hat{m}(z;l_i)-\mathbb{E}\hat{m}\left(z;l_i\right)\right|^4|\mathrm{d}z|\\ &\le \varepsilon^{-4}\left|\mathcal{K}_p\right|^4p^2\sup_{z\in\mathcal{K}_p}\mathbb{E}\left|\hat{m}(z;l_i)-\mathbb{E}\hat{m}\left(z;l_i\right)\right|^4\\ &\le \varepsilon^{-4}\left|\mathcal{K}_p\right|^4C_2p^{-2}.\end{aligned}$$

Since this sequence is summable in p, we have $I_p^{(S)}\to 0$, a.s.

We turn to the exceptional term $I_p^{(E)}$. When $z\in\mathcal{K}_{0p}$ and event Q_p occurs, we may bound $\hat{m}\left(z;l_i\right)=\frac{1}{p}\sum_{j=1}^p\left(\mu_{pj}-z\right)^{-1}$ (where μ_{pj} with $j=1,...,p$ are the eigenvalues of $A(l_i)\equiv H-l_iE$) using

$$\max_j\left|\mu_{pj}-z\right|^{-1}\le\max\left\{-2/\bar{a}_+,1/x_+\right\}\equiv\eta^{-1},$$

so that $\left|\hat{m}\left(z;l_i\right)\right|\le\eta^{-1}$. Further, for sufficiently large p, we have $\left|m_p\left(z;l_i\right)\right|\le\eta^{-1}$ for $z\in\mathcal{K}_{0p}$. Consequently,

$$\begin{aligned}\Pr\left\{I_p^{(E)}>\varepsilon,Q_p\right\} &\le \varepsilon^{-1}\mathbb{E}\left(I_p^{(E)}\mathbf{1}_{Q_p}\right)\\ &\le \varepsilon^{-1}\left|\mathcal{K}_{0p}\right|\sup_{z\in\mathcal{K}_{0p}}\left\{\sqrt{p}\left(\left|\hat{m}\left(z;l_i\right)\right|+\left|m_p\left(z;l_i\right)\right|\right)\mathbf{1}_{Q_p}\right\}\\ &\le \varepsilon^{-1}4p^{-2}2\eta^{-1}p^{1/2}=C\varepsilon^{-1}p^{-3/2},\end{aligned}$$

where C denotes a constant. Again this is summable in p, and since Q_p occurs w.o.p., it follows that $I_p^{(E)}\overset{a.s.}{\to}0$ also. In summary, referring to (13), we see that in order to show a.s. convergence in (4), it remains to establish (14) and (15).

We begin with some preliminary results. Two tools for handling fluctuations are then introduced: first, moment bounds for deviations of quadratic forms, and then, the martingale difference structure. Then we proceed to bound the deterministic term in (13). After all this, we are ready to attack the stochastic bounds (14).

Preliminary results. The approach consists in careful analysis of the perturbations induced in the resolvent of $A(l_i)=YTY'/p$ by deletion of a single column from Y. Thus, let $q_k=Y_{\cdot k}/\sqrt{p}$ (the k-th column of Y divided by $\sqrt{p}$) and $A_k=A(l_i)-t_kq_kq_k'$. Consider events

$$Q_{pk}=\left\{\lambda_{\max}\left(A_k\right)<\bar{a}_+,\ \lambda_{\min}\left(A_k\right)>-4x_+\right\}.$$

Similar to Q_p, events Q_{pk} occur w.o.p.

Let

$$Q_{p,all}=Q_p\cap Q_{p1}\cap...\cap Q_{pn}.$$

Then, $Q_{p,all}$ occur w.o.p. This follows from the equality $\max_{j=1,...,p} \Pr\left\{Q_{pj}^c\right\} = o\left(p^{-t}\right)$ for each fixed $t > 0$. The equality is true because, first, each Q_{pj}, $j = 1, ..., p$ occurs w.o.p., and, second, $\Pr\left\{Q_{pj}^c\right\}$ takes on only two possible values, depending on whether $t_k = c_1$ or $t_k = -l_i c_2$. In particular, although there is a proliferation of the number of events involved in the construction of $Q_{p,all}$ as $p \to \infty$, the probabilities of these events approach one uniformly.

Let

$$D = (A(l_i) - zI_p)^{-1} \text{ and } D_k = (A_k - zI_p)^{-1},$$

and let

$$\beta_k = 1 + t_k q_k' (A_k - zI)^{-1} q_k.$$

An important identity to be used later is

$$q_k' D = q_k' D_k / \beta_k. \tag{16}$$

The following lemma establishes a useful bound on β_k.

LEMMA 8. *Suppose that event $Q_{p,all}$ holds. Then, there exists a constant c, that depends only on $\bar{a}_+, x_+$, and $\mathcal{K}$, such that, for any $z = u + iv \in \mathcal{K}$,*

$$|\beta_k| \geq \sqrt{\frac{v^2}{c^2 + v^2}}.$$

Proof: Let e_j be a normalized eigenvector corresponding to the j-th largest eigenvalue, μ_{jk}, of A_k. Then, we have

$$\beta_k = 1 + t_k q_k' (A_k - zI)^{-1} q_k = 1 + t_k \sum_{j=1}^{p} \frac{\left(q_k' e_j\right)^2}{\mu_{jk} - z}. \tag{17}$$

Consider the case where $v > 0$. When $Q_{p,all}$ holds, $\mu_{jk} \in [-4x_+, \bar{a}_+]$, and

$$\arg(\mu_{jk} - z) \in [\arg(-4x_+ - z), \arg(\bar{a}_+ - z)],$$

where arg belongs to $(-\pi, \pi)$, and $\arg(-4x_+ - z) < \arg(\bar{a}_+ - z) < 0$. Let us denote $\arg(-4x_+ - z)$ as $-\varphi_L$ and $\arg(\bar{a}_+ - z)$ as $-\varphi_R$. Note that

$$\arg\left\{(\mu_{jk} - z)^{-1}\right\} \in [\varphi_R, \varphi_L],$$

and thus,

$$\varphi_R \leq \arg \sum_{j=1}^{p} \left(q_k' e_j\right)^2 / (\mu_{jk} - z) \leq \varphi_L.$$

These inequalities and equation (17) imply that, when $Q_{p,all}$ holds, $|\beta_k|$ cannot be smaller than the distance from the origin to the cone $\{1 + \rho e^{i\varphi} : \rho \in \mathbb{R}, \varphi_R \leq \varphi \leq \varphi_L\}$, which equals $\min\{\sin\varphi_L, \sin\varphi_R\}$. On the other hand,

$$\sin\varphi_L = \sin\arctan\frac{v}{|-4x_+ - u|} \text{ and } \sin\varphi_R = \sin\arctan\frac{v}{|\bar{a}_+ - u|},$$

so that there exists c that depends only on $\bar{a}_+, x_+$, and $\mathcal{K}$, such that, when $Q_{p,all}$ holds,

$$|\beta_k| \geq \sin\arctan\frac{v}{c} = \sqrt{\frac{v^2}{c^2 + v^2}}.$$

The case where $v < 0$ leads to the same conclusion in a similar way. □

REMARK 9. *In the case where all t_k are non-negative (which was studied by Bai and Silverstein (2004)), $|\beta_k|$ is always bounded by $|v/z|$. In contrast, when some t_k are negative, the corresponding $|\beta_k|$ can be arbitrarily close to zero with small, but positive, probability. This is why the bound in Lemma 8 is conditioned on the event $Q_{p,all}$.*

For the analysis of the deterministic term $I_p^{(D)}$, we will need the following preliminary convergence result.

LEMMA 10. *$\mathbb{E}\hat{m}(z;l_i) - m(z) \to 0$ and $m_p(z;l_i) - m(z) \to 0$, uniformly in $z \in \mathcal{K}_p$. Here $m(z)$ is the Stieltjes transform of the LSD of $A(l_i) \equiv H - l_i E$ as $p, n \to_\gamma \infty$.*

Proof: For $m_p(z;l_i)$, such a convergence is a consequence of the fact that $F^{c,\hat{\mathcal{T}}}$, that is, the LSD of $A(l_i)$ under the fixed c_1, c_2 asymptotic regime converges to $F^{\gamma,\mathcal{T}}$ as $c_1 \to \gamma_1$ and $c_2 \to \gamma_2$. Moreover, the supports of $F^{c,\hat{\mathcal{T}}}$ and $F^{\gamma,\mathcal{T}}$ coinside asymptotically, and lie at a positive distance from contour $\mathcal{K}$.

For $\mathbb{E}\hat{m}(z;l_i)$, note that, since the spectral distribution $\hat{F}_p$ of $A(l_i)$ a.s. converges to $F^{c,\mathcal{T}}$ (we denote this as $\hat{F}_p \xrightarrow{d} F^{c,\mathcal{T}}$, a.s.), we have, by the dominated convergence theorem,

$$\mathbb{E}\hat{F}_p \xrightarrow{d} F^{c,\mathcal{T}}. \tag{18}$$

Further, since $\left\{(\lambda - z)^{-1} : z \in \mathcal{K}_H\right\}$ is a family of bounded equicontinuous functions of $\lambda \in \mathbb{R}$, (18) implies that $\mathbb{E}\hat{m} \to m$, uniformly in $z \in \mathcal{K}_H$. Next,

$$\begin{aligned}\mathbb{E}\hat{m} - m &= \int (\lambda - z)^{-1} \mathbf{1}_{[-4x_+,\bar{a}_+]}(\lambda)\, \mathrm{d}\left(\mathbb{E}\hat{F}_p - F^{c,\mathcal{T}}\right) \\ &\quad + \int (\lambda - z)^{-1} \mathbf{1}_{[-4x_+,\bar{a}_+]^c}(\lambda)\, \mathrm{d}\mathbb{E}\hat{F}_p.\end{aligned}$$

The first integral converges to zero uniformly in $z \in \mathcal{K}_{pV}$ because

$$\left\{(\lambda - z)^{-1} \mathbf{1}_{[-4x_+,\bar{a}_+]}(\lambda) : z \in \mathcal{K}_{pV}\right\}$$

is a family of bounded equicontinuous functions of $\lambda \in \mathbb{R}$. For the second integral, we have

$$\sup_{z \in \mathcal{K}_{pV}} \left| \int (\lambda - z)^{-1} \mathbf{1}_{[-4x_+,\bar{a}_+]^c}(\lambda)\, \mathrm{d}\mathbb{E}F_p \right| \le p^2 \Pr\left\{Q_p^c\right\} \to 0. \ \square$$

Deviations of quadratic forms and the martingale difference structure.

LEMMA 11. *Let C be a $p \times p$ non-random matrix, and $r \ge 1$. Suppose that ξ_j, $j = 1, ..., p$ are independent mean zero random variables with $\mathbb{E}|\xi_j|^2 = 1$ and $\mathbb{E}|\xi_j|^l = \nu_l$ for $l \le 2r$. Then*

$$\mathbb{E}\left|\xi^* C \xi - trC\right|^r \le K_r \left(\left[\nu_4 tr\left(CC^*\right)\right]^{r/2} + \nu_{2r} tr\left(CC^*\right)^{r/2}\right)$$

This is Lemma 2.7 in Bai and Silverstein (1998).

Introduce increasing σ-fields $\mathcal{F}_k = \sigma\{q_j : j \le k\}$ and let $\mathbb{E}_k$ denote $\mathbb{E}(\cdot|\mathcal{F}_k)$. For $y = \mathrm{tr}D$, write

$$y - \mathbb{E}y = \sum_{k=1}^{n_1+n_2} \mathbb{E}_k y - \mathbb{E}_{k-1} y.$$

Now introduce $y_k = \mathrm{tr} D_k$ and observe that $\mathbb{E}_k y_k = \mathbb{E}_{k-1} y_k$. Therefore,

$$p\left(\hat{m}\left(z;l_i\right) - \mathbb{E}\hat{m}\left(z;l_i\right)\right) = y - \mathbb{E}y = \sum_{k=1}^{n_1+n_2} z_k,$$

$$z_k = \left(\mathbb{E}_k - \mathbb{E}_{k-1}\right)\left(y - y_k\right). \tag{19}$$

We have $\mathbb{E}_{k-1} z_k = 0$ for $k = 1, ..., n_1 + n_2$ and so the z_k form a martingale difference sequence. The differences are orthogonal, so

$$\mathbb{E}\left(y - \mathbb{E}y\right)^2 = \sum_{k=1}^{n_1+n_2} \mathbb{E}z_k^2.$$

Let us establish some bounds on $y - y_k$. We have

$$\begin{aligned} D &= \left(A_k - zI_p + t_k q_k q_k'\right)^{-1} \\ &= D_k - t_k D_k q_k \left(1 + t_k q_k' D_k q_k\right)^{-1} q_k' D_k. \end{aligned} \tag{20}$$

Therefore,

$$y - y_k = \mathrm{tr}\left(D - D_k\right) = -\frac{t_k}{1 + t_k q_k' D_k q_k} q_k' D_k^2 q_k. \tag{21}$$

Let e_{jk} be a normalized eigenvector of A_k corresponding to its j-th largest eigenvalue μ_{jk}. We have

$$\left|1 + t_k q_k' D_k q_k\right| \geq \left|t_k\right| \left|\mathrm{Im}\, q_k' D_k q_k\right| = \left|t_k\right| \sum_{j=1}^{p} \frac{\left(q_k' e_{jk}\right)^2}{\left|\mu_{jk} - z\right|^2} \left|v\right|$$

and

$$\left|t_k q_k' D_k^2 q_k\right| \leq \left|t_k\right| \sum_{j=1}^{p} \frac{\left(q_k' e_{jk}\right)^2}{\left|\mu_{jk} - z\right|^2}$$

Therefore,

$$\left|y - y_k\right| \leq 1/\left|v\right|. \tag{22}$$

This bound can be very large when v is small. Therefore, for $z \in \mathcal{K}_V$, we will need another bound. Denote the eigenvalues of $A(l_i)$ as $\mu_1 \geq \mu_2 \geq ... \geq \mu_p$. We would like to show that if $Q_{p,all}$ holds and $z \in \mathcal{K}_V$, we have

$$\left|y - y_k\right| < C$$

for some constant C. There are two vertical sections of $\mathcal{K}_V$. Let us denote the "left" vertical section $\mathcal{K}_{VL}$. Consider $z \in \mathcal{K}_{VL}$, let $z_L = -5x_+$ denote the intercection of $\mathcal{K}_{VL}$ with real axis.

Recall that the eigenvalues of A_k are denoted as $\mu_{1k} \geq \mu_{2k} \geq ... \geq \mu_{pk}$. By interlacing inequality, if $t_k > 0$, then

$$\mu_1 \geq \mu_{1k} \geq ... \geq \mu_p \geq \mu_{pk}. \tag{23}$$

If $t_k < 0$, then

$$\mu_{1k} \geq \mu_1 \geq ... \geq \mu_{pk} \geq \mu_p \tag{24}$$

Now,

$$y - y_k = \mathrm{Re}\,[\mathrm{tr}\, D - \mathrm{tr}\, D_k] + \mathrm{i}\,\mathrm{Im}\,[\mathrm{tr}\, D - \mathrm{tr}\, D_k] \equiv u_{1k} + \mathrm{i}u_{2k}.$$

For u_{1k} and $z \in \mathcal{K}_{VL}$, we have

$$u_{1k} = \sum_{j=1}^{p} \frac{\mu_j - z_L}{(\mu_j - z_L)^2 + v^2} - \sum_{s=1}^{p} \frac{\mu_{sk} - z_L}{(\mu_{sk} - z_L)^2 + v^2}.$$

Since $|v| \leq \eta \equiv \min\{-\bar{a}_+/2, x_+\}$, the ratios in the above displayed expression are strictly decreasing functions of $\mu_j, \mu_{sk} \geq -4x_+$. Therefore, by interlacing inequalities (23, 24), if $t_k > 0$, then

$$-\frac{\mu_{pk} - z_L}{(\mu_{pk} - z_L)^2 + v^2} < u_{1k} < 0. \tag{25}$$

If $t_k < 0$, then

$$0 < u_{1k} < \frac{\mu_p - z_L}{(\mu_p - z_L)^2 + v^2}. \tag{26}$$

Similarly,

$$u_{2k} = \sum_{j=1}^{p} \frac{v}{(\mu_j - z_L)^2 + v^2} - \sum_{s=1}^{p} \frac{v}{(\mu_{sk} - z_L)^2 + v^2}$$

and the ratios in the above displayed expression are strictly decreasing (increasing) functions of $\mu_j, \mu_{sk} \geq -4x_+$ when $v > 0$ ($v < 0$). Therefore, by (23, 24), we have, if $t_k > 0$,

$$-\frac{v}{(\mu_{pk} - z_L)^2 + v^2} < u_{2k} < 0. \tag{27}$$

If $t_k < 0$, then

$$0 < u_{2k} < \frac{v}{(\mu_p - z_L)^2 + v^2}. \tag{28}$$

From (25-28), we see that, if $Q_{p,all}$ holds, then, for $z \in \mathcal{K}_{VL}$ with $v \leq \eta$,

$$\begin{aligned} |y - y_k| &\leq \frac{|v|}{\eta^2 + v^2} + \frac{\eta}{\eta^2 + v^2} \\ &\leq \frac{\eta}{2\eta^2} + \frac{\eta}{\eta^2} = \frac{3}{2}\eta^{-1}. \end{aligned} \tag{29}$$

Similarly, we can show that the same inequality holds for $z \in \mathcal{K}_{VR}$ (the "right" portion of $\mathcal{K}_V$).

Another bound on β_k. The bound on β_k obtained in Lemma 8 will be sufficient for our analysis in cases where $z \in \mathcal{K}_H$. However, for $z \in \mathcal{K}_V$, it may be too close to zero, and we need another bound.

As follows from (6), for any t from the support of $\mathcal{T}$ (in our case there are only two such t: γ_1 and $-x_i\gamma_2$), $1 + tm(z) \neq 0$ for $z \in \mathcal{K}$. Therefore,

$$\nu \equiv \min_{t \in \{\gamma_1, -x_i\gamma_2\}} \inf_{z \in \mathcal{K}} |1 + tm(z)| > 0.$$

Since $\mathbb{E}\hat{m}(z; l_i) - m(z) \to 0$ uniformly in $z \in \mathcal{K}_p$, and since $c_1, c_2, l_i \to \gamma_1, \gamma_2, x_i$, we have

$$\min_{t\in\{c_1,-l_ic_2\}} \inf_{z\in\mathcal{K}_p} |1 + t\mathbb{E}\hat{m}(z; l_i)| > 2\nu/3$$

for sufficiently large p.

Consider the event $|\beta_k| \le \nu/3$. If this event holds, then $|\beta_k - 1 - t_k\mathbb{E}\hat{m}(z; l_i)| > \nu/3$ for sufficiently large p and any $z \in \mathcal{K}_p$. Recalling the definition of β_k, we obtain

$$|t_k| \left|q_k' D_k q_k - \mathbb{E}\hat{m}(z; l_i)\right| \equiv |t_k| |\varepsilon_k| > \nu/3.$$

Let us show that the sequence of events $|t_k| |\varepsilon_k| \le \nu/3$ occurs w.o.p. Note that this would imply that the sequence of events $|\beta_k| > \nu/3$ also occurs w.o.p.

We have

$$|\varepsilon_k| \le \left|q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right| + \left|\frac{1}{p}(\operatorname{tr} D_k - \operatorname{tr} D)\right| + \left|\frac{1}{p}(\operatorname{tr} D - \mathbb{E}\operatorname{tr} D)\right|$$

Bound on $\frac{1}{p}(\operatorname{tr} D - \mathbb{E}\operatorname{tr} D)$.

By the Burkholder inequality (see Burkholder, 1973, and Theorem 2.10 of Hall and Heyde, 1980), we have, for any $1 < r < \infty$,

$$\begin{aligned} &\mathbb{E}|\operatorname{tr} D - \mathbb{E}\operatorname{tr} D|^r = \mathbb{E}|y - \mathbb{E}y|^r \\ \le\;& C_r\mathbb{E}\left|\sum\nolimits_{k=1}^{n_1+n_2} [(\mathbb{E}_k - \mathbb{E}_{k-1})(y - y_k)]^2\right|^{r/2}. \end{aligned}$$

Therefore,

$$\begin{aligned} &\Pr\left(\frac{|t_k|}{p}|\operatorname{tr} D - \mathbb{E}\operatorname{tr} D| > \frac{\nu}{9}\right) \le \left(\frac{9|t_k|}{\nu p}\right)^r \mathbb{E}|\operatorname{tr} D - \mathbb{E}\operatorname{tr} D|^r \\ \le\;& \left(\frac{9|t_k|}{\nu p}\right)^r C_r\mathbb{E}\left|\sum_{k=1}^{n_1+n_2} [(\mathbb{E}_k - \mathbb{E}_{k-1})(y - y_k)]^2\right|^{r/2}. \end{aligned}$$

For $z \in \mathcal{K}_H$, using (22), we obtain

$$\Pr\left(\frac{|t_k|}{p}|\operatorname{tr} D - \mathbb{E}\operatorname{tr} D| > \frac{\nu}{9}\right) \le \left(\frac{9|t_k|}{\nu}\right)^r p^{-r/2} C_r \left(\frac{2}{\eta}\right)^r \left(\frac{n_1+n_2}{p}\right)^{r/2}$$

Since r is an arbitrary number larger than one, the sequence of events $\frac{|t_k|}{p}|\operatorname{tr} D - \mathbb{E}\operatorname{tr} D| \le \frac{\nu}{9}$ $k = 1, 2, ...$ occurs w.o.p.

For $z \in \mathcal{K}_{pV}$, we need another estimate. We have

$$\begin{aligned} &\Pr\left(\frac{|t_k|}{p}|\operatorname{tr} D - \mathbb{E}\operatorname{tr} D| > \frac{\nu}{9}, Q_{p,all}\right) \\ \le\;& \Pr\left(\frac{|t_k|}{p}\left|\operatorname{tr} D\mathbf{1}_{Q_{p,all}} - \mathbb{E}\left[\operatorname{tr} D\left(\mathbf{1}_{Q_{p,all}} + \mathbf{1}_{Q^c_{p,all}}\right)\right]\right| > \frac{\nu}{9}\right) \end{aligned}$$

But, for $z \in \mathcal{K}_{pV}$, $|\operatorname{tr} D| \le p^3$, and therefore, $\mathbb{E}\left[\operatorname{tr} D\mathbf{1}_{Q^c_{p,all}}\right] \le p^3 \Pr\left(Q^c_{p,all}\right) \to 0$ as $p \to \infty$. In particular, for sufficiently large p, $\frac{|t_k|}{p}\left|\mathbb{E}\left[\operatorname{tr} D\mathbf{1}_{Q^c_{p,all}}\right]\right| < \frac{\nu}{9} - \frac{\nu}{10}$, and we can write, for sufficiently large p and any $r > 1$,

$$
\begin{aligned}
& \Pr\left(\frac{|t_k|}{p}\left|\operatorname{tr} D - \mathbb{E}\operatorname{tr} D\right| > \frac{\nu}{9}, Q_{p,all}\right) \\
\leq\ & \Pr\left(\frac{|t_k|}{p}\left|\operatorname{tr} D\mathbf{1}_{Q_{p,all}} - \mathbb{E}\left[\operatorname{tr} D\mathbf{1}_{Q_{p,all}}\right]\right| > \frac{\nu}{10}\right) \\
\leq\ & \left(\frac{10\,|t_k|}{\nu p}\right)^r \mathbb{E}\left|\operatorname{tr} D\mathbf{1}_{Q_{p,all}} - \mathbb{E}\left[\operatorname{tr} D\mathbf{1}_{Q_{p,all}}\right]\right|^r \\
\leq\ & \left(\frac{10\,|t_k|}{\nu p}\right)^r C_r \mathbb{E}\left|\sum_{k=1}^{n_1+n_2}\left[(\mathbb{E}_k - \mathbb{E}_{k-1})\left(y\mathbf{1}_{Q_{p,all}} - y_k\mathbf{1}_{Q_{p,all}}\right)\right]^2\right|^{r/2} \\
\leq\ & \left(\frac{10\,|t_k|}{\nu}\right)^r p^{-r/2} C_r \left(\frac{n_1+n_2}{p}\right)^{r/2}\left(3\eta^{-1}\right)^r.
\end{aligned}
$$

The last inequality follows from (29). Since $Q_{p,all}$ occur w.o.p., the obtained upper bound on $\Pr\left(\frac{|t_k|}{p}\left|\operatorname{tr} D - \mathbb{E}\operatorname{tr} D\right| > \frac{\nu}{9}, Q_{p,all}\right)$ implies that $\frac{|t_k|}{p}\left|\operatorname{tr} D - \mathbb{E}\operatorname{tr} D\right| \leq \frac{\nu}{9}$ occur w.o.p. too.

Bound on $q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k$.

By Chebyshev's inequality, for any $r \geq 1$,

$$
\Pr\left(|t_k|\left|q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right| > \frac{\nu}{9}\right) \leq \left(\frac{9}{\nu\,|t_k|}\right)^r \mathbb{E}\left|q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right|^r.
$$

Now, by Lemma 11, we have

$$
\mathbb{E}\left(\left|q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right|^r |D_k\right) \leq p^{-r} K_r\left(\left[\nu_4 tr\left(D_k D_k^*\right)\right]^{r/2} + \nu_{2r} tr\left(D_k D_k^*\right)^{r/2}\right).
$$

Since $\|D_k\| \leq 1/v$, we have, for $z \in \mathcal{K}_H$,

$$
\mathbb{E}\left(\left|q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right|^r |D_k\right) \leq p^{-r} K_r\left(\nu_4 \eta^{-r} p^{r/2} + \nu_{2r} p \eta^{-r}\right).
$$

Therefore,

$$
\Pr\left(|t_k|\left|q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right| > \frac{\nu}{9}\right) \leq \left(\frac{9}{\nu\,|t_k|}\right)^r p^{-r} K_r\left(\nu_4 \eta^{-r} p^{r/2} + \nu_{2r} p \eta^{-r}\right),
$$

and $|t_k|\left|q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right| \leq \frac{\nu}{9}$ occur w.o.p.

For $z \in \mathcal{K}_{pV}$, we have

$$
\begin{aligned}
& \Pr\left(|t_k|\left|q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right| > \frac{\nu}{9}, Q_{p,k}\right) \\
=\ & \Pr\left(|t_k|\left|q_k' D_k q_k \mathbf{1}_{Q_{p,k}} - \frac{1}{p}\operatorname{tr} D_k \mathbf{1}_{Q_{p,k}}\right| > \frac{\nu}{9}\right)
\end{aligned}
$$

and since the eigenvalues of $D_k\mathbf{1}_{Q_{p,k}}$ are bounded by η^{-1}, we get, using the above line of arguments,

$$
\Pr\left(|t_k|\left|q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right| > \frac{\nu}{9}, Q_{p,k}\right) \leq \left(\frac{9}{\nu\,|t_k|}\right)^r p^{-r} K_r\left(\nu_4 \eta^{-r} p^{r/2} + \nu_{2r} p \eta^{-r}\right).
$$

Since $Q_{p,k}$ occur w.o.p., $|t_k|\left|q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right| \leq \frac{\nu}{9}$ occur w.o.p. too.

Bound on $\frac{1}{p}(\operatorname{tr} D_k - \operatorname{tr} D)$.

Inequalities (22) and (29) imply that

$$\left|\frac{1}{p}(\operatorname{tr} D_k - \operatorname{tr} D)\right| \le \frac{1}{p\eta}$$

for $z \in \mathcal{K}_H$, and

$$\left|\frac{1}{p}\left(\operatorname{tr} D_k \mathbf{1}_{Q_{p,all}} - \operatorname{tr} D \mathbf{1}_{Q_{p,all}}\right)\right| < \frac{1}{p}\frac{3}{2}\eta^{-1}$$

for $z \in \mathcal{K}_V$. Therefore, events

$$\left|\frac{|t_k|}{p}(\operatorname{tr} D_k - \operatorname{tr} D)\right| \le \frac{\nu}{9}$$

occur w.o.p.

To summarize, we have shown that $|t_k|\,|\varepsilon_k| \le \nu/3$ occur w.o.p. This implies that

$$|\beta_k| > \nu/3$$

occur w.o.p. for any $z \in \mathcal{K}_p$.

The deterministic term.

Let

$$W_p = \cap_{k=1}^{n_1+n_2}\{|\beta_k| > \nu/3\} \cap Q_{p,all}.$$

Note that W_p occur w.o.p. This implies that $\sqrt{p}\,(\mathbb{E}\hat{m}(z;l_i) - m_p(z;l_i)) \to 0$ uniformly in $z \in \mathcal{K}_p$ if and only if $\sqrt{p}\left(\mathbb{E}\left[\hat{m}(z;l_i)\mathbf{1}_{W_p}\right] - m_p(z;l_i)\right) \to 0$ uniformly in $z \in \mathcal{K}_p$. Indeed, for $z \in \mathcal{K}_p$, $|\hat{m}(z;l_i)| \le 1/|\operatorname{Im} z| \le p^{-2}$, and

$$\sqrt{p}\mathbb{E}\left|\hat{m}(z;l_i)\mathbf{1}_{W_p^c}\right| \le p^{-2}\sqrt{p}\Pr\left\{W_p^c\right\} \to 0.$$

Let us denote $\hat{m}(z;l_i)\mathbf{1}_{W_p}$ as $\bar{m}(z;l_i) \equiv \bar{m}(z)$ for brevity. Note that Lemma 10 implies that $\mathbb{E}\bar{m}(z) - m(z) \to 0$ uniformly in $z \in \mathcal{K}_p$.

Consider

$$\delta(z) = z\mathbb{E}\bar{m}(z) - \frac{c_1+c_2}{c_1c_2}\int \frac{t\mathbb{E}\bar{m}(z)\mathrm{d}\hat{\mathcal{T}}(t)}{1+t\mathbb{E}\bar{m}(z)} + 1,$$

which is the left hand side of (8) where $m_p(z)$ is replaced by $\mathbb{E}\bar{m}(z)$. We have

$$\mathbb{E}\bar{m}(z) = -\left(z - \frac{c_1+c_2}{c_1c_2}\int \frac{t\mathrm{d}\hat{\mathcal{T}}(t)}{1+t\mathbb{E}\bar{m}(z)} - \frac{\delta(z)}{\mathbb{E}\bar{m}(z)}\right)^{-1}.$$

Subtracting (7) and rearranging, we obtain

$$\mathbb{E}\bar{m}(z) - m_p(z) = -\delta m_p\left[1 - m_p\mathbb{E}\bar{m}\frac{c_1+c_2}{c_1c_2}\int \frac{t^2\mathrm{d}\hat{\mathcal{T}}(t)}{(1+t\mathbb{E}\bar{m})(1+tm_p)}\right]^{-1},$$

where we omit the dependence of δ, m_p and $\bar{m}$ on z andl_i to make the displayed formula easier to read. We will omit this dependence in what follows to make notations more compact.

To establish that $\sqrt{p}\,|\mathbb{E}\bar{m} - m_p| \to 0$, uniformly in $z \in \mathcal{K}_p$, it is sufficient to show that

(a) for sufficiently large p, $|m_p|$ is bounded, uniformly in $z \in \mathcal{K}_p$,

(b) for sufficiently large p, the absolute value of the square bracket in (30) is bounded away from zero, uniformly in $z \in \mathcal{K}_p$,

(c) $\sqrt{p}\,|\delta| \to 0$, uniformly in $z \in \mathcal{K}_p$.

(a) follows from the fact that m_p converges to m, uniformly in $z \in \mathcal{K}_p$, which was established in Lemma 10.

To establish (b), it is sufficient to show that there exists $\xi \in (0,1)$ such that, for sufficiently large p,

$$|\mathbb{E}\bar{m}|^2 \frac{c_1+c_2}{c_1c_2} \int \frac{t^2 \mathrm{d}\hat{\mathcal{T}}(t)}{|1+t\mathbb{E}\bar{m}|^2} < \xi \tag{30}$$

and

$$|m_p|^2 \frac{c_1+c_2}{c_1c_2} \int \frac{t^2 \mathrm{d}\hat{\mathcal{T}}(t)}{|1+tm_p|^2} < \xi, \tag{31}$$

uniformly in $z \in \mathcal{K}_p$. Indeed, then (b) follows by the Cauchy–Schwarz inequality.

In fact, since m_p and $\mathbb{E}\bar{m}$ converge to m, uniformly in $z \in \mathcal{K}_p$, it is sufficient to prove that there exists $\xi \in (0,1)$

$$|m|^2 \frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2} \int \frac{t^2 \mathrm{d}\mathcal{T}(t)}{|1+tm|^2} < \xi, \tag{32}$$

uniformly in $z \in \mathcal{K}$. Using (5), we obtain

$$\begin{aligned} |m|^2 \frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2} \int \frac{t^2 \mathrm{d}\mathcal{T}(t)}{|1+tm|^2} &= \frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2} \int \frac{t^2 \mathrm{d}\mathcal{T}(t)}{|1+tm|^2} \\ &\quad \times \left| z - \frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2} \int \frac{t \mathrm{d}\mathcal{T}(t)}{1+tm} \right|^{-2}. \end{aligned} \tag{33}$$

Using (5) again, we obtain for $z = u + iv$,

$$\operatorname{Im} m = \left(v + \operatorname{Im} m \frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2} \int \frac{t^2 \mathrm{d}\mathcal{T}(t)}{|1+tm|^2} \right) \left| z - \frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2} \int \frac{t \mathrm{d}\mathcal{T}(t)}{1+tm} \right|^{-2}.$$

Combining this with (33), we obtain

$$\begin{aligned} |m|^2 \frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2} \int \frac{t^2 \mathrm{d}\mathcal{T}(t)}{|1+tm|^2} &= \operatorname{Im} m \frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2} \int \frac{t^2 \mathrm{d}\mathcal{T}(t)}{|1+tm|^2} \\ &\quad \times \left(v + \operatorname{Im} m \frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2} \int \frac{t^2 \mathrm{d}\mathcal{T}(t)}{|1+tm|^2} \right)^{-1}. \end{aligned}$$

On the other hand, $\operatorname{Im} m = \int v \frac{\mathrm{d}F^{\gamma,\mathcal{T}}}{|\lambda - z|^2}$. Therefore,

$$\begin{aligned} |m|^2 \frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2} \int \frac{t^2 \mathrm{d}\mathcal{T}(t)}{|1+tm|^2} &= \int \frac{\mathrm{d}F^{\gamma,\mathcal{T}}}{|\lambda - z|^2} \frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2} \int \frac{t^2 \mathrm{d}\mathcal{T}(t)}{|1+tm|^2} \\ &\quad \times \left(1 + \int \frac{\mathrm{d}F^{\gamma,\mathcal{T}}}{|\lambda - z|^2} \frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2} \int \frac{t^2 \mathrm{d}\mathcal{T}(t)}{|1+tm|^2} \right)^{-1}. \end{aligned}$$

The right hand side of the above equality is smaller than one for any $z \in \mathcal{K}$. Since it is continuous on $z \in \mathcal{K}$, there exists $\xi \in (0,1)$ such that (32) holds, uniformly in $z \in \mathcal{K}$.

It remains to establish (c). It is sufficient to show that $\sqrt{p}\,|\bar{\delta}| \to 0$, where

$$\bar{\delta} = \delta - \Pr\left\{W_p^c\right\}.$$

Define

$$u = \frac{c_1 + c_2}{c_1 c_2} \int \frac{t \mathrm{d}\hat{T}(t)}{1 + t\mathbb{E}\bar{m}} = \frac{1}{p} \sum_{k=1}^{n_1+n_2} \frac{t_k}{1 + t_k \mathbb{E}\bar{m}}.$$

We have

$$\begin{aligned} \bar{\delta} &= \Pr\left\{W_p\right\} - (u - z)\,\mathbb{E}\bar{m} = \frac{1}{p} \operatorname{tr} \mathbb{E}\left[(A(l_i) - uI)\,(A(l_i) - zI)^{-1}\, \mathbf{1}_{W_p}\right] \\ &= \frac{1}{p} \operatorname{tr} \mathbb{E}\left[\sum_{k=1}^{n_1+n_2} t_k q_k q_k' D \mathbf{1}_{W_p} - uD\mathbf{1}_{W_p}\right] \end{aligned}$$

Using the identity (16), we obtain

$$\bar{\delta} = \frac{1}{p} \sum_{k=1}^{n_1+n_2} \mathbb{E}\left[\frac{t_k q_k' D_k q_k}{\beta_k} \mathbf{1}_{W_p}\right] - y\mathbb{E}\bar{m}.$$

Using the definition of u, we can continue

$$\bar{\delta} = \frac{1}{p} \sum_{k=1}^{n_1+n_2} \frac{t_k}{1 + t_k \mathbb{E}\bar{m}} \mathbb{E}\left[\frac{t_k q_k' D_k q_k}{\beta_k} \mathbf{1}_{W_p} \frac{1 + t_k \mathbb{E}\bar{m}}{t_k} - \mathbb{E}\bar{m}\right].$$

The term in the square brackets equals

$$\begin{aligned} & \frac{q_k' D_k q_k}{\beta_k} \mathbf{1}_{W_p} + \left[\frac{t_k q_k' D_k q_k}{\beta_k} \mathbf{1}_{W_p} - 1\right] \mathbb{E}\bar{m} \\ =\ & \frac{q_k' D_k q_k}{\beta_k} \mathbf{1}_{W_p} - \frac{1}{\beta_k} \mathbf{1}_{W_p} \mathbb{E}\bar{m} - \mathbf{1}_{W_p^c} \mathbb{E}\bar{m} \\ =\ & \frac{\varepsilon_k}{\beta_k} \mathbf{1}_{W_p} + \frac{\mathbb{E}\hat{m} - \mathbb{E}\bar{m}}{\beta_k} \mathbf{1}_{W_p} - \mathbf{1}_{W_p^c} \mathbb{E}\bar{m}, \end{aligned}$$

where

$$\varepsilon_k = q_k' D_k q_k - \mathbb{E}\hat{m}.$$

Therefore, to establish (c), it is sufficient to show that

$$\left|\mathbb{E} \frac{\varepsilon_k}{\beta_k} \mathbf{1}_{W_p}\right| = O\left(p^{-1/2}\right), \tag{34}$$

$$\left|\mathbb{E} \frac{\mathbb{E}\hat{m} - \mathbb{E}\bar{m}}{\beta_k} \mathbf{1}_{W_p}\right| = O\left(p^{-1/2}\right), \tag{35}$$

and

$$\mathbb{E}\left|\mathbf{1}_{W_p^c} \mathbb{E}\bar{m}\right| = O\left(p^{-1/2}\right), \tag{36}$$

uniformly in $z \in \mathcal{K}_p$. It is because $\max_k \left|\frac{t_k}{1+t_k \mathbb{E}\bar{m}}\right|$ is bounded uniformly in $z \in \mathcal{K}_p$, which follows from the uniform in $z \in \mathcal{K}_p$ convergence of $\mathbb{E}\hat{m}$ (and thus, of $\mathbb{E}\bar{m}$) to m, established by Lemma 10.

Equality (36) immediately follows from the fact that W_p occur w.o.p. and from the boundedness of $\mathbb{E}\bar{m}$ (it converges to $m(z)$ which is bounded on $z \in \mathcal{K}$). For (35), we have

$$\mathbb{E}\frac{\mathbb{E}\hat{m} - \mathbb{E}\bar{m}}{\beta_k}\mathbf{1}_{W_p} = \mathbb{E}\left(\hat{m}\mathbf{1}_{W_p^c}\right)\mathbb{E}\left(\beta_k^{-1}\mathbf{1}_{W_p}\right).$$

But

$$\mathbb{E}\left(\beta_k^{-1}\mathbf{1}_{W_p}\right) \leq \frac{3}{\nu}. \tag{37}$$

Further, $|\hat{m}| \leq 1/|v| \leq p^2$ for $z \in \mathcal{K}_p$, and therefore,

$$\mathbb{E}\left(\hat{m}\mathbf{1}_{W_p^c}\right) \leq p^2 \Pr\left\{W_p^c\right\}. \tag{38}$$

Inequalities (37), (38), and the fact that W_p occur w.o.p. imply (35).

The following lemma subsumes (34) by proving a stronger statement.

Lemma 12. *There exists a constant C such that, for sufficiently large p, for any $z \in \mathcal{K}_p$,*

$$\left|\mathbb{E}\frac{\varepsilon_k}{\beta_k}\mathbf{1}_{W_p}\right| \leq Cp^{-1}.$$

Proof: We need the following decomposition of $1/\beta_k$:

$$\frac{1}{\beta_k} = \frac{1}{b_k} - \frac{1}{\beta_k}\frac{1}{b_k}\gamma_k,$$

where

$$b_k = 1 + \frac{1}{p}t_k \mathbb{E}\operatorname{tr} D_k$$

and

$$\begin{aligned}
\gamma_k &= t_k\left[q_k' D_k q_k - \frac{1}{p}\mathbb{E}\operatorname{tr} D_k\right] \\
&= t_k \varepsilon_k + t_k \frac{1}{p}\mathbb{E}\left[\operatorname{tr} D - \operatorname{tr} D_k\right].
\end{aligned}$$

Using the decomposition, we obtain

$$\left|\mathbb{E}\frac{\varepsilon_k}{\beta_k}\mathbf{1}_{W_p}\right| \leq \left|\mathbb{E}\frac{\varepsilon_k}{b_k}\mathbf{1}_{W_p}\right| + \left|\mathbb{E}\frac{t_k\varepsilon_k^2}{\beta_k b_k}\mathbf{1}_{W_p}\right| + \left|\frac{1}{p}\mathbb{E}\left[\operatorname{tr} D - \operatorname{tr} D_k\right]\right|\left|\mathbb{E}\frac{t_k\varepsilon_k}{\beta_k b_k}\mathbf{1}_{W_p}\right|.$$

Bound on $\left|\mathbb{E}\frac{\varepsilon_k}{b_k}\mathbf{1}_{W_p}\right|$.

We have

$$\mathbb{E}\left[\frac{\varepsilon_k}{b_k}\mathbf{1}_{W_p}\right] = \frac{1}{b_k}\mathbb{E}\varepsilon_k - \frac{1}{b_k}\mathbb{E}\left[\varepsilon_k \mathbf{1}_{W_p^c}\right].$$

Since

$$\mathbb{E}\left(\varepsilon_k | D_k\right) = \frac{1}{p}\left(\operatorname{tr} D_k - \mathbb{E}\operatorname{tr} D\right),$$

we have

$$\mathbb{E}\left(\varepsilon_k\right)=\frac{1}{p}\mathbb{E}\left(\operatorname{tr}D_k-\operatorname{tr}D\right)=-\frac{1}{p}\mathbb{E}\left(y-y_k\right) \tag{39}$$

By (22), we have

$$\frac{1}{p}\left|\mathbb{E}\left(y-y_k\right)\right|<\frac{1}{p\eta}\text{ for }z\in\mathcal{K}_H, \tag{40}$$

and by (29), we have

$$\begin{aligned}\frac{1}{p}\left|\mathbb{E}\left(y-y_k\right)\right| &< \frac{1}{p}\frac{3}{2}\eta+\frac{1}{p}\left|\mathbb{E}\left(\operatorname{tr}D_k\mathbf{1}_{W_p^c}-\operatorname{tr}D\mathbf{1}_{W_p^c}\right)\right| \\ &\leq \frac{1}{p}\frac{3}{2}\eta+\frac{1}{p}\left(2p^3\right)\Pr\left(W_p^c\right)\leq\frac{1}{p}2\eta^{-1}\end{aligned} \tag{41}$$

for sufficiently large p and $z\in\mathcal{K}_{pV}$. Using (40) and (41) in (39), we obtain

$$\left|\mathbb{E}\left(\varepsilon_k\right)\right|\leq\frac{1}{p}\frac{3}{\eta} \tag{42}$$

for any $z\in\mathcal{K}_p$.

Further, since the eigenvalues of D_k are no larger than $1/\left|v\right|$ by absolute value, and since $1/\left|v\right|\leq p^2$ for $z\in\mathcal{K}_p$, we have

$$\begin{aligned}\left|\mathbb{E}\left[\varepsilon_k\mathbf{1}_{W_p^c}\right]\right| &\leq p^2\left|\mathbb{E}\left[q_k'q_k\mathbf{1}_{W_p^c}\right]\right|+\mathbb{E}\hat{m}\Pr\left(W_p^c\right) \\ &\leq p^2\left(\mathbb{E}\left[\left(q_k'q_k\right)^2\right]\right)^{1/2}\Pr\left(W_p^c\right)^{1/2}+\mathbb{E}\hat{m}\Pr\left(W_p^c\right).\end{aligned}$$

Therefore,

$$\left|\mathbb{E}\left[\varepsilon_k\mathbf{1}_{W_p^c}\right]\right|\leq p^{-1}C \tag{43}$$

for sufficiently large p and some C.

Next, by Lemma 10, $b_k=1+\frac{1}{p}t_k\mathbb{E}\operatorname{tr}D_k\to1+t_km(z)$ and hence,

$$\left|b_k\right|>\nu/2 \tag{44}$$

for sufficiently large p. Combining (42), (43), and (44), we obtain

$$\left|\mathbb{E}\frac{\varepsilon_k}{b_k}\mathbf{1}_{W_p}\right|\leq p^{-1}C \tag{45}$$

for sufficiently large p and some C.

Bound on $\left|\mathbb{E}\frac{t_k\varepsilon_k^2}{\beta_kb_k}\mathbf{1}_{W_p}\right|$.

By (44), for sufficiently large p, we have

$$\left|\mathbb{E}\frac{t_k\varepsilon_k^2}{\beta_kb_k}\mathbf{1}_{W_p}\right|\leq\frac{6t_k}{\nu^2}\mathbb{E}\left(\varepsilon_k^2\mathbf{1}_{W_p}\right). \tag{46}$$

Consider the decomposition

$$\varepsilon_k = \left(q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right) + \frac{1}{p}\left(\operatorname{tr} D_k - \mathbb{E}\operatorname{tr} D\right) \tag{47}$$

By Lemma 11,

$$\mathbb{E}\left(\left(q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right)^2 |D_k\right) \leq \frac{1}{p^2} K_2\left(\left[\nu_4 \operatorname{tr}\left(D_k D_k^*\right)\right]^{1/2} + \nu_4 \operatorname{tr}\left(D_k D_k^*\right)\right)$$

Since the eigenvalues of D_k are bounded by absolute value by

$$c_z \equiv \min\left\{\frac{1}{|v|}, \eta^{-1}\mathbf{1}_{Q_{p,all}}\mathbf{1}_{z\in\mathcal{K}_V}\right\},$$

we have

$$\mathbb{E}\left(\left(q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right)^2 |D_k\right) \leq \frac{1}{p^2} K_2\left(\left[\nu_4 p c_z^2\right]^{1/2} + \nu_4 p c_z^2\right)$$

and

$$\begin{aligned}\mathbb{E}\left(q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right)^2 \quad &\leq \quad \frac{1}{p^2} K_2\left(\nu_4 p\right)^{1/2} p^2 \Pr\left\{Q_{p,all}^c\right\} + \frac{1}{p^2} K_2\left(\nu_4 p\right)^{1/2} \eta^{-1} \\ &\quad + \frac{1}{p^2} K_2 \nu_4 p p^4 \Pr\left\{Q_{p,all}^c\right\} + \frac{1}{p^2} K_2 \nu_4 p \eta^{-2}\end{aligned}$$

Therefore, for sufficiently large p,

$$\mathbb{E}\left(q_k' D_k q_k - \frac{1}{p}\operatorname{tr} D_k\right)^2 \leq C p^{-1} \tag{48}$$

for some C.

For the second part of the decomposition (47), we have

$$\mathbb{E}\frac{1}{p^2}\left(\operatorname{tr} D_k - \mathbb{E}\operatorname{tr} D\right)^2 \leq \mathbb{E}\frac{2}{p^2}\left[\left(\operatorname{tr} D_k - \mathbb{E}\operatorname{tr} D_k\right)^2 + \left(\mathbb{E}\operatorname{tr} D_k - \mathbb{E}\operatorname{tr} D\right)^2\right] \tag{49}$$

Note that $\left(\mathbb{E}\left[\operatorname{tr} D_k - \operatorname{tr} D\right]\right)^2$ is bounded by (40) and (41). Let us now prove that

$$\mathbb{E}\left(\operatorname{tr} D_k - \mathbb{E}\operatorname{tr} D_k\right)^2 \leq Cp$$

for some C. We will prove this inequality for D_k replaced by D to ease notation. The proof for D_k is very similar.

Recall that

$$\operatorname{tr} D - \mathbb{E}\operatorname{tr} D = \sum_{k=1}^{n_1+n_2}\left(\mathbb{E}_k - \mathbb{E}_{k-1}\right)\left(y - y_k\right)$$

and

$$\mathbb{E}\left(\operatorname{tr} D - \mathbb{E}\operatorname{tr} D\right)^2 = \sum_{k=1}^{n_1+n_2}\mathbb{E} z_k^2$$

where $z_k = \left(\mathbb{E}_k - \mathbb{E}_{k-1}\right)\left(y - y_k\right)$.

Using (22) and (29), we obtain

$$\begin{aligned}\mathbb{E}z_k^2 &= \mathbb{E}\left[\mathbb{E}_k\left(y-y_k\right)-\mathbb{E}_{k-1}\left(y-y_k\right)\right]^2 \leq \\ &\quad 2\mathbb{E}\left[\mathbb{E}_k\left(y-y_k\right)\right]^2+2\mathbb{E}\left[\mathbb{E}_{k-1}\left(y-y_k\right)\right]^2 \\ &\leq 2\mathbb{E}\mathbb{E}_k\left(y-y_k\right)^2+2\mathbb{E}\mathbb{E}_{k-1}\left(y-y_k\right)^2 \\ &= 4\mathbb{E}\left(y-y_k\right)^2 \\ &\leq 4p^4\Pr\left\{Q_{p,all}^c\right\}+4\left\{\frac{1}{\eta^2}+\left(\frac{3}{2\eta}\right)^2\right\}.\end{aligned}$$

Hence, $\mathbb{E}\left(\operatorname{tr}D-\mathbb{E}\operatorname{tr}D\right)^2 \leq Cp$ (and $\mathbb{E}\left(\operatorname{tr}D_k-\mathbb{E}\operatorname{tr}D_k\right)^2 \leq Cp$) for some C, and, from (49), we have

$$\mathbb{E}\frac{1}{p^2}\left(\operatorname{tr}D_k-\mathbb{E}\operatorname{tr}D\right)^2 \leq Cp^{-1} \tag{50}$$

for some C. Now, (47), (48), and (50) imply that

$$\mathbb{E}\left(\varepsilon_k^2\right) \leq Cp^{-1} \tag{51}$$

for some C. Therefore, by (46),

$$\left|\mathbb{E}\frac{t_k\varepsilon_k^2}{\beta_k b_k}\mathbf{1}_{W_p}\right| \leq Cp^{-1} \tag{52}$$

for some C.

Bound on $\left|\frac{1}{p}\mathbb{E}\left[\operatorname{tr}D-\operatorname{tr}D_k\right]\right|\left|\mathbb{E}\frac{t_k\varepsilon_k}{\beta_k b_k}\mathbf{1}_{W_p}\right|$.

By (40) and (41),

$$\left|\frac{1}{p}\mathbb{E}\left[\operatorname{tr}D-\operatorname{tr}D_k\right]\right| \leq p^{-1}\frac{3}{\eta}. \tag{53}$$

Further, by the Cauchy-Schwarz inequality,

$$\left|\mathbb{E}\frac{t_k\varepsilon_k}{\beta_k b_k}\mathbf{1}_{W_p}\right|^2 \leq \mathbb{E}\varepsilon_k^2\mathbb{E}\left(\frac{t_k^2}{\beta_k^2 b_k^2}\mathbf{1}_{W_p}\right).$$

Inequality (51) and the boundedness of $\beta_k^2 b_k^2$ away from zero on W_p imply that the right hand side of the above inequality is bounded. From this and (53) we see that

$$\left|\frac{1}{p}\mathbb{E}\left[\operatorname{tr}D-\operatorname{tr}D_k\right]\right|\left|\mathbb{E}\frac{t_k\varepsilon_k}{\beta_k b_k}\mathbf{1}_{W_p}\right| \leq Cp^{-1} \tag{54}$$

for some C.

The Lemma follows from (45), (52), and (54). □

The validity of (36) is implied by the validity of the Lemma, and thus, $\sqrt{p}\left|\bar{\delta}\right| \to 0$. This concludes our proof of the deterministic term's convergence (15).

The stochastic term.

To get the correct order of magnitude for the fluctuations of

$$z_k = \left(\mathbb{E}_k-\mathbb{E}_{k-1}\right)\left(y-y_k\right),$$

we need a finer decomposition. First, define some conditional means and residuals:

$$\begin{aligned}\beta_k^0 &= \mathbb{E}\left(\beta_k|D_k\right) = 1 + \frac{1}{p} t_k \operatorname{tr} D_k, \\ \gamma_k^0 &= \mathbb{E}\left(\gamma_k|D_k\right) = -\frac{1}{p} t_k \operatorname{tr} D_k^2,\end{aligned}$$

$$\begin{aligned}\varepsilon_{1k} &= \beta_k - \beta_k^0 = t_k \left(q_k' D_k q_k - \frac{1}{p} \operatorname{tr} D_k \right), \\ \varepsilon_{2k} &= \gamma_k - \gamma_k^0 = -t_k \left(q_k' D_k^2 q_k - \frac{1}{p} \operatorname{tr} D_k^2 \right).\end{aligned}$$

Rewrite $y - y_k$ in terms of these means and residulas

$$\begin{aligned}y - y_k &= \frac{\gamma_k}{\beta_k} = \frac{\gamma_k^0}{\beta_k^0} - \frac{\varepsilon_{1k}\gamma_k}{\beta_k^0\beta_k} + \frac{\varepsilon_{2k}}{\beta_k^0} \\ &= r_{0k} + r_{1k} + r_{2k}\end{aligned}$$

The integrable terms[1] r_{0k} do not contain any variables from $\mathcal{F}_k \backslash \mathcal{F}_{k-1}$. Therefore,

$$\left(\mathbb{E}_k - \mathbb{E}_{k-1}\right) r_{0k} = 0.$$

Thus, they disappear from the martingale differences and

$$z_k = \left(\mathbb{E}_k - \mathbb{E}_{k-1}\right)\left(y - y_k\right) = \left(\mathbb{E}_k - \mathbb{E}_{k-1}\right)\left(r_{1k} + r_{2k}\right)$$

Bounds on r_{mk}.
Crude bounds. By (22),

$$\left|\gamma_k / \beta_k\right| \leq 1 / \left|v\right|.$$

Further

$$\begin{aligned}\left|\frac{\gamma_k^0}{\beta_k^0}\right| &\leq \frac{\frac{1}{p}\left|t_k\right| \sum_{j=1}^p \left|\mu_{jk} - z\right|^{-2}}{\left|\operatorname{Im} \beta_k^0\right|} \\ &= \frac{\frac{1}{p}\left|t_k\right| \sum_{j=1}^p \left|\mu_{jk} - z\right|^{-2}}{\frac{1}{p}\left|t_k\right| \left|v\right| \sum_{j=1}^p \left|\mu_{jk} - z\right|^{-2}} = \frac{1}{\left|v\right|}\end{aligned}$$

Next,

$$\left|\beta_k^0\right| > \frac{1}{p}\left|t_k\right| \left|\operatorname{Im} D_k\right| = \frac{1}{p}\left|t_k\right| \sum_{j=1}^p \frac{v}{\left|\mu_{jk} - z\right|^2},$$

whereas

$$\left|\varepsilon_{2k}\right| \leq \left|t_k\right| \sum_{j=1}^p \frac{\left(q_k' e_j\right)^2 + 1/p}{\left|\mu_{jk} - z\right|^2} \leq \left|t_k\right| \left(q_k' q_k + 1/p\right) \sum_{j=1}^p \frac{1}{\left|\mu_{jk} - z\right|^2}$$

[1] $\left|\gamma_k^0/\beta_k^0\right| \leq \frac{1}{p} t_k \sum_{j=1}^p \left|\mu_{jk} - z\right|^{-2} / \left(\left|\operatorname{Im} \beta_k^0\right|\right)$. but

$$\operatorname{Im} \beta_k^0 = \frac{1}{p} t_k v \sum_{j=1}^p \left|\mu_{jk} - z\right|^{-2}$$

therefore, $\left|\gamma_k^0/\beta_k^0\right| \leq 1/v$, and thus, γ_k^0/β_k^0 is integrable.

Therefore,

$$|r_{2k}| \leq \frac{|t_k| \left(q_k' q_k + 1/p\right) \sum_{j=1}^{p} \frac{1}{|\mu_{jk}-z|^2}}{\frac{1}{p} |t_k| \sum_{j=1}^{p} \frac{v}{|\mu_{jk}-z|^2}} = \frac{1}{v} \left(p q_k' q_k + 1\right)$$

Finally,

$$|r_{1k}| \leq \left|\frac{\gamma_k}{\beta_k}\right| + |r_{0k}| + |r_{2k}| \leq \frac{1}{v} \left(p q_k' q_k + 3\right).$$

Fine bounds.

Consider the event

$$B_k = \left\{|\beta_k| > \nu/3, \left|\beta_k^0\right| > \nu/3, p q_k' q_k \leq 2p\right\} \cap Q_{p,all}$$

On this event, which occurs w.o.p. (a proof of the fact that $|\beta_k^0| > \nu/3$ occur w.o.p. is actually contained in the above proof of the fact that $|\beta_k| > \nu/3$ occur w.o.p.), we have

$$|\gamma_k| \leq |t_k| \, 2\eta$$

for $z \in \mathcal{K}_p$ (note that the distance from any μ_{jk} to $\mathcal{K}$ is bounded by η on $Q_{p,all}$). Therefore, on B_k, we have

$$|r_{1k}| \leq |\varepsilon_{1k}| \frac{9 \, |t_k| \, 2}{\nu^2 \eta}$$

and

$$|r_{2k}| \leq |\varepsilon_{1k}| \, 3/\nu.$$

Bounds on ε_{1k} and ε_{2k}.

We have

$$\mathbb{E}_{k-1}\left[|\varepsilon_{1k}|^{2l} \mathbf{1}_{Q_{pk}}\right] = p^{-2l} \mathbb{E}_{k-1}\left[\mathbf{1}_{Q_{pk}} \mathbb{E}\left(\left|p q_k' D_k q_k - \operatorname{tr} D_k\right|^{2l} |D_k\right)\right]$$

Using Lemma 11, we continue, for $l \geq 1$,

$$\begin{aligned} \mathbb{E}_{k-1}\left[|\varepsilon_{1k}|^{2l} \mathbf{1}_{Q_{pk}}\right] &\leq p^{-2l} \mathbb{E}_{k-1}\left[\mathbf{1}_{Q_{pk}} K_{2l} \left(\nu_4^l p^l \eta^{-2l} + \nu_{4l} p \eta^{-2l}\right)\right] \\ &\leq \tilde{K}_l p^{-l} \eta^{-2l} \end{aligned}$$

Similarly,

$$\mathbb{E}_{k-1}\left[|\varepsilon_{2k}|^{2l} \mathbf{1}_{Q_{pk}}\right] \leq \tilde{K}_l p^{-l} \eta^{-4l}$$

Let us prove the following lemma.

Lemma 13. *For any $l \geq 1$, $\mathbb{E}\,|z_k|^{2l} \leq C p^{-l}$ uniformly in $z \in \mathcal{K}_p$.*

Proof: Set $W = r_{1k} + r_{2k}$ and observe that

$$\begin{aligned} \mathbb{E}_{k-1} |z_k|^{2l} &= \mathbb{E}_{k-1} \left|\mathbb{E}_k W - \mathbb{E}_{k-1} W\right|^{2l} \leq c_{2l} \left(\mathbb{E}_{k-1} \left|\mathbb{E}_k W\right|^{2l} + \mathbb{E}_{k-1} \left|\mathbb{E}_{k-1} W\right|^{2l}\right) \\ &\leq c_{2l} \left(\mathbb{E}_{k-1} \mathbb{E}_k |W|^{2l} + \mathbb{E}_{k-1} \mathbb{E}_{k-1} |W|^{2l}\right) = 2 c_{2l} \mathbb{E}_{k-1} |W|^{2l} \\ &\leq c_l' \left(\mathbb{E}_{k-1} |r_{1k}|^{2l} + \mathbb{E}_{k-1} |r_{2k}|^{2l}\right) \end{aligned}$$

using $(a+b)^p \leq c_p (a^p + b^p)$ for $c_p = 2^{p-1}$ and $p \geq 2$.

Further

$$\begin{aligned}
\mathbb{E}_{k-1} |r_{1k}|^{2l} &= \mathbb{E}_{k-1} \left(|r_{1k}|^{2l} \mathbf{1}_{B_k} \right) + \mathbb{E}_{k-1} \left(|r_{1k}|^{2l} \mathbf{1}_{B_k^c} \right) \\
&\leq \left(\frac{9 |t_k| 2}{\nu^2 \eta} \right)^{2l} \mathbb{E}_{k-1} \left(|\varepsilon_{1k}|^{2l} \mathbf{1}_{B_k} \right) + \mathbb{E}_{k-1} \left(|r_{1k}|^{2l} \mathbf{1}_{B_k^c} \right) \\
&\leq \left(\frac{9 |t_k| 2}{\nu^2 \eta} \right)^{2l} \mathbb{E}_{k-1} \left(|\varepsilon_{1k}|^{2l} \mathbf{1}_{Q_{pk}} \right) + \mathbb{E}_{k-1} \left(|r_{1k}|^{2l} \mathbf{1}_{B_k^c} \right)
\end{aligned}$$

Taking unconditional expectations, we get

$$\begin{aligned}
\mathbb{E} |r_{1k}|^{2l} &\leq \left(\frac{9 |t_k| 2}{\nu^2 \eta} \right)^{2l} \mathbb{E} \left(|\varepsilon_{1k}|^{2l} \mathbf{1}_{Q_{pk}} \right) + \mathbb{E} \left(|r_{1k}|^{2l} \mathbf{1}_{B_k^c} \right) \\
&\leq \left(\frac{9 |t_k| 2}{\nu^2 \eta} \right)^{2l} \tilde{K}_l p^{-l} \eta^{-2l} + \mathbb{E} \left(\frac{1}{\upsilon^{2l}} |p q_k' q_k + 3|^{2l} \mathbf{1}_{B_k^c} \right) \\
&\leq \left(\frac{9 |t_k| 2}{\nu^2 \eta} \right)^{2l} \tilde{K}_l p^{-l} \eta^{-2l} + \frac{1}{\upsilon^{2l}} \mathbb{E} \left(|p q_k' q_k + 3|^{4l} \right)^{1/2} [\Pr(B_k^c)]^{1/2} \\
&\leq C p^{-l}
\end{aligned}$$

A similar argument shows that

$$\mathbb{E} |r_{2k}|^{2l} \leq C p^{-l}.$$

□

To establish the stochastic bounds (14), we need to show that $\mathbb{E} |y - \mathbb{E} y|^{2l} \leq C_l$ uniformly in $z \in \mathcal{K}_p$. By Rosenthal's inequality (see Theorem 2.12 of Hall and Heyde), we have for any $l \geq 1$,

$$\mathbb{E} |y - \mathbb{E} y|^{2l} \leq C_l \mathbb{E} \left[\left(\sum_{k=1}^{n_1+n_2} \mathbb{E}_{k-1} z_k^2 \right)^l \right] + C_l \sum_{k=1}^{n_1+n_2} \mathbb{E} |z_k|^{2l}$$

That the second sum on the right is uniformly bounded follows immediately from Lemma 13.

Turn to the first sum. First, obtain the bound

$$\left(\sum_{k=1}^{n_1+n_2} \mathbb{E}_{k-1} z_k^2 \right)^l \leq c_l \left(R_{1p}^l + R_{2p}^l \right),$$

where

$$R_{mp} = \sum_{k=1}^{n_1+n_2} \mathbb{E}_{k-1} |r_{mk}|^2$$

Recall that

$$\begin{aligned}
\mathbb{E}_{k-1} |r_{1k}|^2 &\leq \left(\frac{9 |t_k| 2}{\nu^2 \eta} \right)^2 \mathbb{E}_{k-1} \left(|\varepsilon_{1k}|^2 \mathbf{1}_{Q_{pk}} \right) + \mathbb{E}_{k-1} \left(|r_{1k}|^2 \mathbf{1}_{B_k^c} \right) \\
&\leq \left(\frac{9 |t_k| 2}{\nu^2 \eta} \right)^2 \tilde{K}_1 p^{-1} \eta^{-2} + \mathbb{E}_{k-1} \left(|r_{1k}|^2 \mathbf{1}_{B_k^c} \right)
\end{aligned}$$

Therefore,

$$R_{1p} \leq C + \sum_{k=1}^{n_1+n_2} \mathbb{E}_{k-1} \left(|r_{1k}|^2 \mathbf{1}_{B_k^c} \right)$$

and

$$R_{1p}^l \leq c_l \left[C^l + \left\{ \sum_{k=1}^{n_1+n_2} \mathbb{E}_{k-1} \left(|r_{1k}|^2 \mathbf{1}_{B_k^c} \right) \right\}^l \right]$$

On the other hand, for $l \geq 2$

$$\begin{aligned} \left\{ \sum_{k=1}^{n_1+n_2} \mathbb{E}_{k-1} \left(|r_{1k}|^2 \mathbf{1}_{B_k^c} \right) \right\}^l &\leq (n_1+n_2)^{l-1} \sum_{k=1}^{n_1+n_2} \left[\mathbb{E}_{k-1} \left(|r_{1k}|^2 \mathbf{1}_{B_k^c} \right) \right]^l \\ &\leq (n_1+n_2)^{l-1} \sum_{k=1}^{n_1+n_2} \mathbb{E}_{k-1} \left(|r_{1k}|^{2l} \right) \mathbb{E}_{k-1} \mathbf{1}_{B_k^c} \\ &\leq (n_1+n_2)^{l-1} \sum_{k=1}^{n_1+n_2} \mathbb{E}_{k-1} \left(\left| \frac{1}{v} \left(p q_k' q_k + 3 \right) \right|^{2l} \right) \mathbb{E}_{k-1} \mathbf{1}_{B_k^c} \\ &\leq p^K \sum_{k=1}^{n_1+n_2} \mathbb{E}_{k-1} \mathbf{1}_{B_k^c} \end{aligned}$$

for some K. Taking unconditional expectations, we obtain the boundedness of R_{1p}^l. The boundedness of R_{2p}^l is established similarly. This completes the proof of (14) and hence, of Lemma DJO6.

4.3. *Proof of Lemma DJO7 (a CLT for quadratic forms).* We will need the following two lemmas.

LEMMA 14. *(McLeish 1974) Let $\{X_{pr}, \mathcal{G}_{pr}, r = 1, ..., p\}$ be a martingale difference array on the probability triple $(\Omega, \mathcal{G}, P)$. If the following conditions are satisfied: a) Lindeberg's condition: for all $\varepsilon > 0$, $\sum_r \int_{|X_{pr}|>\varepsilon} X_{pr}^2 \mathrm{d}P \to 0$ as $p \to \infty$; b) $\sum_r X_{pr}^2 \xrightarrow{\mathrm{P}} 1$, then $\sum_r X_{pr} \xrightarrow{d} N(0,1)$.*

Proof: This is a consequence of Theorem (2.3) of McLeish (1974). Two conditions of the theorem: i) $\max_{r \leq p} |X_{pr}|$ is uniformly bounded in L_2 norm, and ii) $\max_{r \leq p} |X_{pr}| \xrightarrow{\mathrm{P}} 0$, are replaced here by the Lindeberg condition. □

LEMMA 15. *(Hall and Heyde) Let $\{X_{pr}, \mathcal{G}_{pr}, r = 1, ..., p\}$ be a martingale difference array, and define $V_{pj}^2 = \sum_{r=1}^j E\left(X_{pr}^2 | \mathcal{G}_{p,r-1} \right)$ and $U_{pj}^2 = \sum_{r=1}^j X_{pr}^2$ for $r = 1, ..., p$. Suppose that the conditional variances V_{pp}^2 are tight, that is $\sup_p P\left(V_{pp}^2 > \varepsilon \right) \to 0$ as $\varepsilon \to \infty$, and that the conditional Lindeberg condition holds, that is, for all $\varepsilon > 0$, $\sum_r E\left[X_{pr}^2 \mathbf{1}\{|X_{pr}| > \varepsilon\} | \mathcal{G}_{p,r-1} \right] \xrightarrow{\mathrm{P}} 0$. Then $\max_j \left| U_{pj}^2 - V_{pj}^2 \right| \xrightarrow{\mathrm{P}} 0$.*

Proof: This is a shortened version of Theorem 2.23 in Hall and Heyde (1980). □

Let $f_j(\lambda)$, $j = 1, ..., J$, be such that $f_j(\lambda) = g_j(\lambda)$ for $\lambda \in [b_i, B_i]$ and $f_j(\lambda) = 0$ otherwise. Consider random variables

$$X_{pr} = \frac{1}{\sqrt{p}} \sum\nolimits_{(j,s,t) \in \Theta} \gamma_{jst} f_j(\mu_{r,i}) (\zeta_{rs}\zeta_{rt} - \delta_{st}),$$

where γ_{jst} are some constants. Let $\mathcal{G}_{pR}$ be the σ-algebra generated by $\mu_{1,i}, ..., \mu_{p,i}$ and ζ_{rs} with $r = 1, ..., R$; $s = 1, ..., q$. Clearly, $\{X_{pr}, \mathcal{G}_{pr}, r = 1, ..., p\}$ form a martingale difference array. Let

K be the number of different triples $(j,s,t)\in\Theta$. Consider an arbitrary order in Θ. In Hölder's inequality

$$\sum\nolimits_{a=1}^{K} y_a z_a \le \left(\sum\nolimits_{a=1}^{K} (y_a)^b\right)^{1/b}\left(\sum\nolimits_{a=1}^{K} (z_a)^c\right)^{1/c},$$

which holds for $y_a>0$, $z_a>0$, $b>1$, $c>1$, and $1/b+1/c=1$, take

$$y_a=\left|\frac{1}{\sqrt{p}}\gamma_{jst} f_j\left(\mu_{r,i}\right)\left(\zeta_{rs}\zeta_{rt}-\delta_{st}\right)\right|,$$

where (j,s,t) is the a-th triple in Θ, $z_a=1$, and $b=2+\delta$ for some $\delta>0$. Then, the inequality implies that

$$|X_{pr}|^{2+\delta}\le K^{1+\delta}\rho_i^{2+\delta}\sum\nolimits_{(j,s,t)\in\Theta}\left|\frac{1}{\sqrt{p}}\gamma_{jst}\left(\zeta_{rs}\zeta_{rt}-\delta_{st}\right)\right|^{2+\delta}, \tag{55}$$

where

$$\rho_i=\max_{j=1,\dots,J}\sup_{\lambda\in[b_i,B_i]}\left|g_j(\lambda)\right|.$$

Since ζ_{rs} are i.i.d. $N(0,1)$, (55) implies that $\sum\nolimits_{r=1}^{p} E\left|X_{pr}\right|^{2+\delta}\to 0$ as $p\to\infty$, which means that the Lyapunov condition holds for X_{pr}. As is well known, Lyapunov's condition implies Lindeberg's condition. Hence, condition a) of Lemma 14 is satisfied for X_{pr}.

Let us consider $\sum\nolimits_{r=1}^{p} X_{pr}^2$. Since the convergence in mean implies the convergence in probability, the conditional Lindeberg condition is satisfied for X_{pr} because the unconditional Lindeberg condition is satisfied as checked above. Further, in notations of Lemma 15, it is easy to see that

$$V_{pp}^2=\sum\nolimits_{j,j_1}\left[\left(\sum\nolimits_{1\le s\le t\le q}\gamma_{jst}\gamma_{j_1st}\left(1+\delta_{st}\right)\right)\frac{1}{p}\sum\nolimits_{r=1}^{p} f_j\left(\mu_{r,i}\right) f_{j_1}\left(\mu_{r,i}\right)\right].$$

The convergence of the empirical distribution of $\mu_{1,i},\dots,\mu_{p,i}$ to $F(\lambda;x_i)$ and the equality of g_j and f_j on the support of $F(\lambda;x_i)$ implies that

$$V_{pp}^2\xrightarrow{\mathrm{P}}\Sigma\equiv\sum\nolimits_{j,j_1}\left[\left(\sum\nolimits_{1\le s\le t\le q}\gamma_{jst}\gamma_{j_1st}\left(1+\delta_{st}\right)\right)\int g_j(\lambda)\, g_{j_1}(\lambda)\,\mathrm{d}F(\lambda;x_i)\right].$$

In particular, V_{pp}^2 is tight and Lemma 15 applies. Therefore, $\sum\nolimits_{r=1}^{p} X_{pr}^2$ converges to the same limit as V_{pp}^2. Thus, by Lemma 14, we get $\sum\nolimits_{r=1}^{p} X_{pr}\xrightarrow{d} N(0,\Sigma)$.

Finally, let

$$Y_{pr}=\frac{1}{\sqrt{p}}\sum\nolimits_{(j,s,t)\in\Theta}\gamma_{jst} g_j\left(\mu_{r,i}\right)\left(\zeta_{rs}\zeta_{rt}-\delta_{st}\right).$$

Since

$$\Pr\left(\sum\nolimits_{r=1}^{p} X_{pr}\ne\sum\nolimits_{r=1}^{p} Y_{pr}\right)\to 0$$

as $p\to\infty$, we have $\sum\nolimits_{r=1}^{p} Y_{pr}\xrightarrow{d} N(0,\Sigma)$. Lemma DJO7 follows from this convergence via the Cramer-Wold device.

4.4. *Derivation of equations (DJO27-29).* Expression (DJO27) immediately follows from (DJO15). For (DJO28), differentiating identity (DJO13) with respect to z, we obtain

$$1+\frac{\gamma_1 m'(z;x)}{(1+\gamma_1 m(z;x))^2}=\frac{m'(z;x)}{m^2(z;x)}+\frac{-x^2\gamma_2 m'(z;x)}{(1-\gamma_2 x m(z;x))^2}.$$

Setting $z=0$ and $x=x_i$, and using the fact that

$$m(0;x_i)=-(s_i+\gamma_1)^{-1}, \tag{56}$$

which follows from (DJO15), we obtain

$$1+\frac{\gamma_1 m'(0;x_i)}{\left(1-\gamma_1(s_i+\gamma_1)^{-1}\right)^2}=\frac{m'(0;x_i)}{(s_i+\gamma_1)^{-2}}+\frac{-x_i^2\gamma_2 m'(0;x_i)}{\left(1+\gamma_2 x_i(s_i+\gamma_1)^{-1}\right)^2}.$$

Using the definition (DJO17) of x_i, we obtain

$$1+\frac{\gamma_1 m'(0;x_i)}{\left(1-\gamma_1(s_i+\gamma_1)^{-1}\right)^2}=\frac{m'(0;x_i)}{(s_i+\gamma_1)^{-2}}-\frac{(s_i+\gamma_1)^2(s_i+1)^2\gamma_2 m'(0;x_i)}{s_i^2},$$

which implies (DJO28).

Finally, differentiating identity (DJO13) with respect to x, we obtain

$$\frac{\gamma_1 \mathrm{d}m(z;x)/\mathrm{d}x}{(1+\gamma_1 m(z;x))^2}=\frac{\mathrm{d}m(z;x)/\mathrm{d}x}{(m(z;x))^2}+\frac{-1+\gamma_2 x m(z;x)-x(\gamma_2 m(z;x)+\gamma_2 x\mathrm{d}m(z;x)/\mathrm{d}x)}{(1-\gamma_2 x m(z;x))^2}.$$

Setting $z=0$ and $x=x_i$, we obtain

$$\frac{\gamma_1 \mathrm{d}m(0;x_i)/\mathrm{d}x}{(1+\gamma_1 m(0;x_i))^2}=\frac{\mathrm{d}m(0;x_i)/\mathrm{d}x}{(m(0;x_i))^2}+\frac{-1-\gamma_2 x_i^2\mathrm{d}m(0;x_i)/\mathrm{d}x}{(1-\gamma_2 x_i m(0;x_i))^2}.$$

This equality, the definition (DJO17) of x_i, and equation (56) imply (DJO29).

5. Analysis of the joint density of eigenvalues.

5.1. *Proof of Lemma DJO13 about branch determination on the horizontal part of $\mathcal{K}$.* To determine the branches, we will view the part of $\mathcal{K}$ on the real axis as the limit of a wedge-like contour

$$\mathcal{W}=(\tilde{x}_0+\mathrm{i}\varepsilon,\tilde{\lambda}_{p1})\cup(\tilde{x}_0-\mathrm{i}\varepsilon,\tilde{\lambda}_{p1})$$

as $\varepsilon\downarrow 0$, where i is the imaginary unit. Contour $\mathcal{W}$ intersects with each of the rays $\{z:\arg z=\varphi\}$, $\varphi\in(-\pi,\pi]$ no more than once, and therefore, the branches of all the fractional powers in ω must be principal as discussed in DJO. As $\varepsilon\downarrow 0$, we identify the branches by continuity as follows.

The situation will depend on which of $z_1,...,z_r$ belong to the "upper" and which of them belong to the "lower" parts of $\mathcal{K}\cap[\tilde{x}_0,\tilde{\lambda}_{p1}]$, that is the parts that are oriented from $\tilde{\lambda}_{p1}$ to $\tilde{x}_0$, and from $\tilde{x}_0$ to $\tilde{\lambda}_{p1}$, respectively.

There are 2^r possible scenarios: $(sgn_{z_1}=\pm1,...,sgn_{z_r}=\pm1)$, where $sgn_{z_j}=+1$ means that z_j belongs to the "upper" part, and $sgn_j=-1$ means that z_j belongs to the "lower" part of $\mathcal{K}\cap[\tilde{x}_0,\tilde{\lambda}_{p1}]$. Consider a particular scenario $(sgn_{z_1},..,sgn_{z_r})$. Deforming $\mathcal{K}\cap[\tilde{x}_0,\tilde{\lambda}_{p1}]$ to the wedge-like contour $\mathcal{W}$, we move z_j to

$$z_{j\varepsilon}=z_j+\mathrm{i}\times sgn_{z_j}\frac{\tilde{\lambda}_{p1}-z_j}{\tilde{\lambda}_{p1}-\tilde{x}_0}\varepsilon.$$

Since on $\mathcal{W}$, the principal branches of fractional powers are taken, the sign of the imaginary part of $\left(1 - z_{j\varepsilon} z_{i\varepsilon}^{-1}\right)^{1/2}$ for $j > i$ must be equal to sgn_{z_i}. Therefore, for $j > i$,

$$\operatorname{sgn} \operatorname{Im} \left(1 - z_j z_i^{-1}\right)^{1/2} = \lim_{\varepsilon \downarrow 0} \operatorname{sgn} \operatorname{Im} \left(1 - z_{j\varepsilon} z_{i\varepsilon}^{-1}\right)^{1/2} = sgn_{z_i}.$$

Similarly, for $\tilde{\lambda}_{ps}$ and z_i such that $\tilde{\lambda}_{ps} > z_i$, we have

$$\operatorname{sgn} \operatorname{Im} \left(1 - \tilde{\lambda}_s z_i^{-1}\right)^{-1/2} = \lim_{\varepsilon \downarrow 0} \operatorname{sgn} \operatorname{Im} \left(1 - \tilde{\lambda}_{ps} z_{i\varepsilon}^{-1}\right)^{-1/2} = -sgn_{z_i}.$$

5.2. *Proof of the fact that $\mathcal{I}'_\sigma$ vanishes if there are any repeats in σ.* Suppose, specifically, that $\sigma_{a+1} = ... = \sigma_{a+b} = i_0 \leq k$ for some $b \geq 2$, so that the variables $z_{a+1}, ..., z_{a+b}$ lie within the same segment. Because of the branch effects described in Lemma DJO13, it helps the bookkeeping to first factor out from ω the terms that depend on sgn_{z_i}, $i = 1, ..., r$. To this end, let $v_i = \#\left\{s : \tilde{\lambda}_{ps} > z_i\right\}$ and $v_+ = \sum_{i=1}^{r} v_i$. Then we factorize $\omega = S\tilde{\omega}$ so that $\tilde{\omega}$ does not depend on $\{sgn_{z_i}, i = 1, ..., r\}$ while from Lemma DJO13 we have

$$S = \mathrm{i}^{r(r-1)/2} (-\mathrm{i})^{v_+} \prod_{i=1}^{r} sgn_{z_i}^{r-i+v_i}.$$

It will be enough to show that the inner integral within (DJO43) given by

$$\mathcal{I}'^{,in}_\sigma = \int_{\mathcal{K}_{i_0}^b \cap \mathcal{S}^b} S\tilde{\omega}\mathcal{F}_{p,n} \prod_{i=a+1}^{a+b} \mathrm{d}z_i = 0, \tag{57}$$

where $\mathcal{S}^b$ denotes the region $z_{a+1} < ... < z_{a+b}$. To this end, we factorize $S = S_0 S_1$, where S_0 has all the terms that do not involve $z_{a+1}, ..., z_{a+b}$ and, if we let v denote the common value of $v_{a+1} = ... = v_{a+b}$, we have

$$S_1 = \prod_{i=a+1}^{a+b} sgn_{z_i}^{r-i+v_i}.$$

Write $[\alpha, \beta]$ for the interval $[\tilde{\lambda}_{p,i_0+1}, \tilde{\lambda}_{p,i_0}]$. Recall that each of the b copies of $\mathcal{K}_{i_0}$ is the union of two contour segments, namely $\mathcal{K}_{i_0}^-$, which traverses $[\alpha, \beta]$ left-to-right, and on which $sgn_z = -1$, and $\mathcal{K}_{i_0}^+$, which traverses $[\alpha, \beta]$ right-to-left, and on which $sgn_z = 1$. Let $sgn_j = sgn_{z_{a+j}}$ and decompose the inner integral (57) over all combinations of these contour segments to get

$$\mathcal{I}'^{,in}_\sigma = S_0 \sum_{sgn_1} ... \sum_{sgn_b} \int_{[\alpha,\beta]^b \cap \mathcal{S}^b} S_1 S_2 \tilde{\omega} \mathcal{F}_{p,n} \prod_{i=a+1}^{a+b} \mathrm{d}z_i,$$

where each sum is over $sgn_j = \pm 1$ and the term $S_2 = \prod_{j=1}^{b} (-sgn_j)$ counts whether z_{a+j} traverses $\mathcal{K}_{i_0}^-$ or $\mathcal{K}_{i_0}^+$. Now we can evaluate

$$\sum_{sgn_1} ... \sum_{sgn_b} S_1 S_2 = (-1)^b \prod_{j=1}^{b} \sum_{sgn_j} sgn_j^{r-a-j+v+1} = 0$$

when $b \geq 2$. Indeed, each sum on the right is of the form $1 + (-1)^l$, which vanishes when l is odd, and this must occur for at least one term if $b \geq 2$.

5.3. *Proof of Lemma DJO16 about the Laplace approximation for* $\mathcal{F}_{p,n}^{\mathrm{SigD}}$. Let $A = \mathrm{diag}\{a_1, ..., a_k\}$ with $0 < a_1 < ... < a_k$ and $B = \mathrm{diag}\{b_1, ..., b_k\}$ with $b_1 > ... > b_k > 0$. We have the following lemma.

LEMMA 16. *Let* $c_{ij} = (a_j - a_i)(b_i - b_j) / \{(1 + a_i b_i)(1 + a_j b_j)\}$. *Then, as* $T \to \infty$,

$$ {}_1F_0\left(T/2; -A, B\right) = \Gamma_k\left(k/2\right)\pi^{-k(k+1)/4}\prod_{i=1}^{k}\left(1 + a_i b_i\right)^{-T/2}\prod_{i<j}^{k}\left(Tc_{ij}/2\right)^{-1/2}\left(1 + o(1)\right), \tag{58} $$

where $o(1) \to 0$ *uniforms on any compact subsects of the simplexes* $0 < a_1 < ... < a_k$ *and* $b_1 > ... > b_k > 0$.

This lemma is a minor extension of Chang's (1970) Theorem 1, which establishes (58) for fixed A and B (with both sides of (58) divided by the volume of the orthogonal group $\mathcal{O}(k)$). To show that the $o(1)$ is uniform on the set of A and B described in the lemma, it is sufficient to replace Hsu's (1948) Lemma 1 by Glynn's (1980) Theorem 2.1 in the proof of Chang (1970). Lemma DJO16 is a corollary to Lemma 16.

5.4. *Proof of Lemma DJO17 about the Laplace approximation for* $\mathcal{F}_{p,n}^{\mathrm{REG}}$. The proof below uses many ideas from Glynn (1980). First, let us represent $\mathcal{F}_{p,n}^{\mathrm{REG}}$ in terms of ${}_0F_1$. Using the identities (see James' (1964) equations (30-31) or Glynn's equations (5.1-5.2))

$$ \mathcal{F}_{p,n}^{\mathrm{REG}} = \int_{\mathcal{O}(k)} {}_1F_1\left(\alpha, \beta; \frac{T}{2}s^{1/2}HZH's^{1/2}\right)(\mathrm{d}H), $$

where $\alpha = Ta + \frac{k+1}{2}$, $\beta = Tb + \frac{k+1}{2}$ and $(\mathrm{d}H)$ is the normalized invariant measure on the orthogonal group $\mathcal{O}(k)$, and

$$ {}_1F_1\left(\alpha, \beta; \frac{T}{2}s^{1/2}HZH's^{1/2}\right) = \left(\Gamma_k\left(\alpha\right)\right)^{-1}\int_{\Sigma>0}\mathrm{etr}\left(-\Sigma\right)\left|\Sigma\right|^{Ta}{}_0F_1\left(\beta; \frac{T}{2}s^{1/2}HZH's^{1/2}\Sigma\right)(\mathrm{d}\Sigma), $$

where $\Sigma > 0$ is a positive deifinite $k \times k$ matrix, we obtain

$$ \mathcal{F}_{p,n}^{\mathrm{REG}} = \left(\Gamma_k\left(\alpha\right)\right)^{-1}\int_{\mathcal{O}(k)}\int_{\Sigma>0}\mathrm{etr}\left(-\Sigma\right)\left|\Sigma\right|^{Ta}{}_0F_1\left(\beta; \frac{T}{2}s^{1/2}HZH's^{1/2}\Sigma\right)(\mathrm{d}\Sigma)(\mathrm{d}H) $$

Next, let us change variables of integration $\Sigma \mapsto Q, R$, where $\Sigma = \frac{T}{2}Q'R^2Q$,

$$ R = \mathrm{diag}\left(r_1, ..., r_k\right) $$

with $r_1 \geq ... \geq r_k \geq 0$, and $Q \in \mathcal{O}\left(k\right)$. For this transformation, we have (see, for example, Herz (1955), p. 479)

$$ (\mathrm{d}\Sigma) = \mathcal{V}_k\left(\frac{T}{2}\right)^{\frac{k^2+k}{2}}2^k\left|R\right|\prod_{i<j}\left(r_i^2 - r_j^2\right)(\mathrm{d}R)(\mathrm{d}Q), $$

where

$$ \mathcal{V}_k = 2^k\pi^{k^2/2}/\Gamma_k\left(k/2\right) $$

is the volume of $\mathcal{O}\left(k\right)$, and $(\mathrm{d}Q)$ is the normalized invariant measure on $\mathcal{O}\left(k\right)$.

The transformation is one-to-2^k becase Q is only determined up to a left-multiplication by a diagonal matrix with ± 1 coefficients along the diagonal. Therefore, we have

$$\begin{aligned}\mathcal{F}_{p,n}^{\mathrm{REG}} &= C_T \int_{\mathcal{O}(k)\times D(R)\times\mathcal{O}(k)} \operatorname{etr}\left(-TR^2/2\right)|R|^{2\alpha-k}\prod_{i<j}\left(r_i^2-r_j^2\right)\\ &\quad\times {}_0F_1\left(\beta;\frac{T^2}{4}RQs^{1/2}HZH's^{1/2}Q'R\right)(\mathrm{d}Q)(\mathrm{d}R)(\mathrm{d}H),\end{aligned}\tag{59}$$

where

$$C_T = \mathcal{V}_k\left(T/2\right)^{k\alpha}/\Gamma_k\left(\alpha\right),\ \text{and}\ D(R)=\{R: r_1>...>r_k>0\}.$$

Consider Herz' integral representation for ${}_0F_1$ (see Butler and Wood (2003), equation (12))

$${}_0F_1\left(\beta,\frac{1}{4}\Theta\Theta'\right) = c\int_{\mathcal{U}} \operatorname{etr}\{\Theta Y\}\left|I_k - YY'\right|^{(2\beta-2k-1)/2}(\mathrm{d}Y),\tag{60}$$

where Θ is a $k\times k$ matrix, $\mathcal{U}=\{Y: YY'<I_k\}$ and $c=\pi^{-k^2/2}\Gamma_k(\beta)/\Gamma_k(\beta-k/2)$. This representation is valid when $\beta\geq k$, which holds for sufficiently large T. Using (60) in (59), we obtain

$$\begin{aligned}\mathcal{F}_{p,n}^{\mathrm{REG}} &= \tilde{C}_T \int_{\mathcal{O}(k)\times D(R)\times\mathcal{O}(k)\times\mathcal{U}} \operatorname{etr}\left[-TR^2/2 + TRQs^{1/2}HZ^{1/2}Y\right]\\ &\quad\times |R|^{2Ta+1}\prod_{i<j}\left(r_i^2-r_j^2\right)\left|I_k - YY'\right|^{Tb-k/2}\mathrm{d}z,\end{aligned}$$

where

$$\mathrm{d}z = (\mathrm{d}Q)(\mathrm{d}R)(\mathrm{d}H)(\mathrm{d}Y),\ \text{and}\ \tilde{C}_T = \frac{2^k\left(T/2\right)^{k\alpha}\Gamma_k(\beta)}{\Gamma_k(\alpha)\Gamma_k(k/2)\Gamma_k(\beta-k/2)}.$$

Now let us make the change of variables $Y\mapsto U,V,\sigma$, where $Y=U\sigma V$ is a singular value decomposition of Y with

$$\sigma = \operatorname{diag}(\sigma_1,...,\sigma_k),\ \text{with}\ \sigma_1\geq...\geq\sigma_k\geq 0.$$

For such a change of variables, we have

$$(\mathrm{d}Y) = \mathcal{V}_k^2\prod_{i<j}\left(\sigma_i^2-\sigma_j^2\right)(\mathrm{d}U)(\mathrm{d}V)(\mathrm{d}\sigma),$$

where $(\mathrm{d}U)$ and $(\mathrm{d}V)$ are the normalized invariant measures on $\mathcal{O}_k$.

The transformation $Y\mapsto U,V,\sigma$ is one-to-2^k. Therefore, we obtain

$$\begin{aligned}\mathcal{F}_{p,n}^{\mathrm{REG}} &= \hat{C}_T\int_{\Lambda}\operatorname{etr}\left[T\left(-R^2/2 + RQs^{1/2}HZ^{1/2}U\sigma V\right)\right]\\ &\quad\times |R|^{2Ta+1}\prod_{i<j}\left(r_i^2-r_j^2\right)\left(\sigma_i^2-\sigma_j^2\right)\left|I_k-\sigma^2\right|^{Tb-k/2}\mathrm{d}\xi,\end{aligned}\tag{61}$$

where

$$\Lambda = \mathcal{O}(k)\times D(R)\times\mathcal{O}(k)\times\bar{D}(\sigma)\times\mathcal{O}(k)\times\mathcal{O}(k)$$

with $\bar{D}(\sigma)=\{\sigma: 1\geq\sigma_1\geq...\geq\sigma_k\geq 0\}$,

$$\mathrm{d}\xi = (\mathrm{d}Q)(\mathrm{d}R)(\mathrm{d}H)(\mathrm{d}\sigma)(\mathrm{d}U)(\mathrm{d}V),\ \text{and}\ \hat{C}_T = \frac{2^{2k}\pi^{k^2}\left(T/2\right)^{k\alpha}\Gamma_k(\beta)}{\Gamma_k(\alpha)\left(\Gamma_k(k/2)\right)^3\Gamma_k(\beta-k/2)}.$$

Equation (61) can be rewritten in the form amenable to the Laplace approximation method as follows

$$\mathcal{F}_{p,n}^{\mathrm{REG}} = \hat{C}_T \int_{\Lambda} h f^T \mathrm{d}\xi, \tag{62}$$

where

$$f = \mathrm{etr}\left\{-R^2/2 + RQs^{1/2}HZ^{1/2}U\sigma V\right\} \left|I_k - \sigma^2\right|^b |R|^{2a},$$

and

$$h = \left|I_k - \sigma^2\right|^{-k/2} |R| \prod_{i<j} \left(r_i^2 - r_j^2\right)\left(\sigma_i^2 - \sigma_j^2\right).$$

By Lemma 4.2 of Glynn (1980), the maximum of $\mathrm{etr}\left\{RQs^{1/2}HZ^{1/2}U\sigma V\right\}$ over $(Q, H, U, V) \in \mathcal{O}(k)^4$ is achieved at 2^{3k} points, where $Q, H, U,$ and V are diagonal with values ± 1 along the diagonal, and such that $QHUV = I_k$. The value of $\mathrm{etr}\left\{RQs^{1/2}HZ^{1/2}U\Sigma V\right\}$ at the maximum is $\mathrm{etr}\left\{\sum_{j=1}^k r_j\sigma_j\sqrt{2\zeta_j}\right\}$, where

$$\zeta_j = z_j s_j/2.$$

We introduce this notation because it simplifies some expressions later on.

This implies that the maximum of f over Λ is achieved at 2^{3k} points with Q, H, U, V as above, and with r_j and σ_j satisfying the following first order conditions for maximisation with respect to r_j and σ_j, $j = 1, ..., k$,

$$-r_j + \sigma_j\sqrt{2\zeta_j} + 2a/r_j = 0 \tag{63}$$

and

$$r_j\sqrt{2\zeta_j} - \frac{2b\sigma_j}{1-\sigma_j^2} = 0. \tag{64}$$

From (63) we obtain

$$\sigma_j = \frac{r_j^2 - 2a}{r_j\sqrt{2\zeta_j}}.$$

Using this in (64) we obatin

$$\left(r_j^2 - 2a\right)^2 + 2\left(b - \zeta_j\right)\left(r_j^2 - 2a\right) - 4a\zeta_j = 0.$$

Let

$$z_{j+} = \frac{1}{2}\left\{\zeta_j - b + \sqrt{\left(b-\zeta_j\right)^2 + 4a\zeta_j}\right\}$$

be the positive solution of the quadratic equation $z_j^2 + \left(b - \zeta_j\right)z_j - a\zeta_j = 0$. Then, $r_j^2 - 2a = 2z_{j+}$, and thus

$$r_j^2 = 2\left(z_{j+} + a\right), \text{ and} \tag{65}$$

$$\sigma_j^2 = \frac{1}{\zeta_j}\frac{z_{j+}^2}{z_{j+} + a}. \tag{66}$$

We have introduced the notation z_{j+} here for compatibility with the previous version of DJO, that studied the one dimensional case $k = 1$.

Let us verify the second order conditions for the maximum at the above r_j^2 and σ_j^2. The matrix of the second derivatives of $\ln f$ with respect to r_j and σ_j is

$$\begin{pmatrix} -1-2a/r_j^2 & \sqrt{2\zeta_j} \\ \sqrt{2\zeta_j} & -2b\left(1+\sigma_j^2\right)/\left(1-\sigma_j^2\right)^2 \end{pmatrix}.$$

Its value at the critical point is

$$\begin{pmatrix} -\left(z_{j+}+2a\right)/\left(z_{j+}+a\right) & \sqrt{2\zeta_j} \\ \sqrt{2\zeta_j} & -2\zeta_j\left(z_{j+}+a\right)/b-2\zeta_j^2\left(z_{j+}+a\right)^2/\left(bz_{j+}^2\right) \end{pmatrix}.$$

By inspection, the diagonal elements of this matrix are negative, whereas the determinant

$$4\zeta_j\left(a+z_{j+}\right)\left(a\zeta_j+z_{j+}^2\right)/\left(bz_{j+}^2\right)>0,$$

so that the second order condition for the maximum is satisfied.

Note that the value of $\ln f$ at the maximum is

$$\max\ln f=\sum_{j=1}^k\left\{-r_j^2/2+r_j\sigma_j\sqrt{2\zeta_j}+b\ln\left(1-\sigma_j^2\right)+2a\ln r_j\right\},$$

where $r_j>0$ and $\sigma_j>0$ are given by (65) and (66). Expressing $\max\ln f$ in terms of a,b,ζ_j and z_{j+}, we obtain

$$\max\ln f=\sum_{j=1}^k\left\{z_{j+}-b\ln\left(z_{j+}+b\right)+a\ln\left(z_{j+}+a\right)+a\ln 2+b\ln b-a\right\}.$$

Since the maximum is achieved at the 2^{3k} points, the integral over Λ in (62) can be replaced, for the purpose of the asymptotic analysis, by 2^{3k} times the integral over

$$\Lambda^+=\mathcal{O}^+(k)\times D(R)\times\mathcal{O}^+(k)\times\bar{D}\left(\sigma\right)\times\mathcal{O}^+\left(k\right)\times\mathcal{O}^+\left(k\right),$$

where $\mathcal{O}^+\left(k\right)$ denotes the set of k-dimensional orthogonal matrices with positive diagonal elements. Since $Q,H,U,$ and V are proper in $\mathcal{O}^+\left(k\right)$, they can be parameterized as

$$Q=\exp\left\{G\right\},\ H=\exp\left\{F\right\},\ U=\exp\left\{A\right\},\ \text{and}\ V=\exp\left\{B\right\},$$

where $G,F,A,$ and B are $k\times k$ skew symmetric.

Anderson (1965) shows that the Jacobian of the transfomation $Q\mapsto S$ equals

$$J_1=J\left(Q\mapsto G\right)=\Gamma_k\left(k/2\right)2^{-k}\pi^{-k^2/2}\left(1+O\left(G_{ij}^2\right)\right),$$

where $O\left(G_{ij}^2\right)$ denotes the terms that are at least quadratic in the elements of G. Similar expressions hold for the Jacobians J_2,J_3 and J_4 of the transformations $H\mapsto F$, $U\mapsto A$, and $V\mapsto B$, respectively. By making this change of variables, we arrive at the following asymptotic representation

$$\mathcal{F}_{p,n}^{\mathrm{REG}}\sim c_T\int_{\Xi}gf^T\left(\mathrm{d}\tilde{\xi}\right), \tag{67}$$

where

$$c_n = 2^{3k}\hat{C}_T = \frac{2^{5k}\pi^{k^2}(T/2)^{k\alpha}\Gamma_k(\beta)}{\Gamma_k(\alpha)(\Gamma_k(k/2))^3\Gamma_k(\beta-k/2)},$$

Ξ is the image of Λ^+ under the transformation $Q, H, U, V \mapsto G, F, A, B$,

$$\begin{aligned} g &= J_1J_2J_3J_4\left|I_k-\sigma^2\right|^{-k/2}|R|\prod_{i<j}\left(r_i^2-r_j^2\right)\left(\sigma_i^2-\sigma_j^2\right), \text{ and} \\ f &= \operatorname{etr}\left\{-R^2/2+Re^Gs^{1/2}e^FZ^{1/2}e^A\sigma e^B\right\}\left|I_k-\sigma^2\right|^b|R|^{2a}. \end{aligned}$$

Expanding e^G, e^F, e^A, and e^B into powers of G, F, A, and B, we have tr $\left[Re^Gs^{1/2}e^FZ^{1/2}e^A\sigma e^B\right]$ equals

$$\begin{aligned} &\sum_{j=1}^k r_js_j^{1/2}z_j^{1/2}\sigma_j-\frac{1}{2}\sum_{j<i}^k\left(r_js_j^{1/2}z_j^{1/2}\sigma_j+r_is_i^{1/2}z_i^{1/2}\sigma_i\right)\left(G_{ji}^2+F_{ji}^2+A_{ji}^2+B_{ji}^2\right) \\ &-\sum_{j<i}^k\left(r_js_i^{1/2}z_j^{1/2}\sigma_j+r_is_j^{1/2}z_i^{1/2}\sigma_i\right)G_{ji}F_{ji}-\sum_{j<i}^k\left(r_js_i^{1/2}z_i^{1/2}\sigma_j+r_is_j^{1/2}z_j^{1/2}\sigma_i\right)G_{ji}A_{ji} \\ &-\sum_{j<i}^k\left(r_js_i^{1/2}z_i^{1/2}\sigma_i+r_is_j^{1/2}z_j^{1/2}\sigma_j\right)G_{ji}B_{ji}-\sum_{j<i}^k\left(r_js_j^{1/2}z_i^{1/2}\sigma_j+r_is_i^{1/2}z_j^{1/2}\sigma_i\right)F_{ji}A_{ji} \\ &-\sum_{j<i}^k\left(r_js_j^{1/2}z_i^{1/2}\sigma_i+r_is_i^{1/2}z_j^{1/2}\sigma_j\right)F_{ji}B_{ji}-\sum_{j<i}^k\left(r_js_j^{1/2}z_j^{1/2}\sigma_i+r_is_i^{1/2}z_i^{1/2}\sigma_j\right)A_{ji}B_{ji} \\ &+\text{h.o.t.}, \end{aligned}$$

where h.o.t. stands for higher order terms (in G_{ji}, F_{ji}, A_{ji}, and B_{ji}).

The maximum value of f is obtained at a single point ξ in the interior of Ξ. The Hessian of $-\ln f$, Δ, reduces to a product of determinants of matrices which are at most 4×4. A direct calculation (using MAPLE symbolic algebra) gives

$$\begin{aligned} \Delta &= \prod_{j=1}^k\frac{4\zeta_j(a+z_{j+})\left(a\zeta_j+z_{j+}^2\right)}{bz_{j+}^2}\prod_{j<i}^k\left\{(z_i-z_j)(s_i-s_j)\left(r_i^2-r_j^2\right)\right. \\ &\qquad\left.\left(\sigma_i^2-\sigma_j^2\right)\left(s_i^{1/2}z_i^{1/2}\sigma_ir_i-s_j^{1/2}z_j^{1/2}\sigma_jr_j\right)^2\right\}. \end{aligned}$$

Further, using Stirling's formula, we have, at $G=F=A=B=0$,

$$c_TJ_1J_2J_3J_4 \sim \frac{\Gamma_k(k/2)}{\pi^{k^2+k(k+1)/4}}\frac{T^{k(3k+1)/4}}{2^{kTa+k^2/2}}\frac{b^{k^2/2}e^{kTa}}{a^{kTa+k(k+1)/4}}.$$

The statement of Lemma DJO17 now follows by applying Glynn's (1980) Theorem 2.1 to (67). We obtain

$$\begin{aligned} \mathcal{F}_{p,n}^{\mathrm{REG}} &\sim \left(\frac{T}{2}\right)^{-k(k-1)/4}\frac{\Gamma_k(k/2)}{\pi^{k(k+1)/4}}\frac{b^{kTb+k(k+1)/4}}{a^{kTa+k(k+1)/4}} \\ &\quad\times\prod_{j=1}^k e^{Tz_{j+}}\frac{(z_{j+}+a)^{aT}}{(z_{j+}+b)^{bT}}\left(\frac{z_{j+}(z_{j+}+a)}{z_{j+}^2+a\zeta_j}\right)^{\frac{1}{2}} \\ &\quad\times\prod_{i<j}(z_i-z_j)^{-1/2}(s_i-s_j)^{-1/2}. \end{aligned}$$

By Glynn's theorem, this asymptotic approximation is uniform over a and b that belong to compact subsets of $(1/2,\infty)$ and $(0,1/2)$, respectively, and over s and Z, such that their diagonal entries belong to compact subsets of the simplexes $s_1 > ... > s_k > 0$ and $z_1 > ... > z_k > 0$.

5.5. *Proof of Lemma DJO18 about the Laplace approximation for* $\overline{\mathcal{I}}^{\text{SigD}}_{\sigma_\varnothing}$. As shown in DJO, the inner-most integral in the multiple integral representation (DJO49) of $\overline{\mathcal{I}}^{\text{SigD}}_{\sigma_\varnothing}$ equals

$$\overline{\mathcal{I}}^{\text{SigD}}_{\sigma_\varnothing,in} = G^{\text{SigD}}_{-1}\int_{\tilde\lambda_{p2}+\varepsilon}^{\tilde\lambda_{p1}} e^{-pf_1(z_1)}g_1(z_1)(1+o(1))\,\mathrm{d}z_1. \tag{68}$$

First, we will apply the Laplace method (see Olver (1997), p. 81–82) to the integral

$$I(p) \equiv \int_{\tilde\lambda_{p2}+\varepsilon}^{\tilde\lambda_{p1}} e^{-pf_1(z_1)}g_1(z_1)\,\mathrm{d}z_1.$$

To line up our analysis with that on pages 81–82 of Olver (1997) and to simplify notations, let us rewrite $I(p)$ as

$$I(p) = \int_{-\tilde\lambda_{p1}}^{-\tilde\lambda_{p2}-\varepsilon} e^{-pf(t)}g(t)\mathrm{d}t,$$

where $t = -z_1$,

$$\begin{aligned} f(t) &= \frac{N-p+2}{2p}\ln\left(1+\frac{s_1 t}{1+s_1}\right) + \frac{1}{2p}\sum_{j=k+1}^{p}\ln\left(-t-\tilde\lambda_{pj}\right), \text{ and} \\ g(t) &= \left(\tilde\lambda_{p1}+t\right)^{-1/2}\prod_{j=2}^{k}\left((-t-z_j)/\left(-t-\tilde\lambda_{pj}\right)\right)^{1/2}. \end{aligned}$$

Let us show that under the null hypothesis, $f(t)$ has positive continuous derivative for t from the interval $[-\tilde\lambda_{p1}, -\tilde\lambda_{p2}-\varepsilon]$ for sufficiently large p, a.s. All statements made in this section should be understood as holding for sufficiently large p, almost surely, and we will omit this qualification to avoid frequent repetitions. We have

$$f'(t) = \frac{N-p+2}{2p}\frac{s_1}{1+s_1+s_1 t} + \frac{1}{2p}\sum_{j=k+1}^{p}\left(t+\tilde\lambda_{pj}\right)^{-1}. \tag{69}$$

Clearly, $f'(t)$ is continuous and strictly decreasing for all $t \in [-\tilde\lambda_{p1}, -\tilde\lambda_{p2}-\varepsilon]$. Under the null, the minimum of $f'(t)$ on $t \in [-\tilde\lambda_{p1}, -\tilde\lambda_{p2}-\varepsilon]$ converges to

$$\frac{\rho^2 s_{01}}{2\gamma_1\gamma_2(1+s_{01}-s_{01}(\lim\tilde\lambda_{p2}+\varepsilon))} + \frac{1}{2}\tilde m(\lim\tilde\lambda_{p2}+\varepsilon) \tag{70}$$

uniformly over $s = s_0 + \delta/\sqrt{p}$ with $(\delta_1,...,\delta_k)$ from a compact subset of $\mathbb{R}^k$, a.s. Here $\tilde m(x)$ is the Stieltjes transform (or rather its analytic continuation to a point on the real line) of the limiting spectral distribution of the multivariate Beta matrix $(n_H\mathbf{H}+n_2\mathbf{E})^{-1}n_H\mathbf{H}$ and

$$\lim\tilde\lambda_{p2} \equiv \frac{(s_{02}+\gamma_1)(s_{02}+1)\gamma_2}{s_{02}(\gamma_1+s_{02}\gamma_2+\gamma_2)}$$

is the a.s. limit of $\tilde\lambda_{p2}$ (see equation (DJO41)).

Consider the following function of two real variables

$$\Psi\left(y, s_{01}\right) \equiv \frac{\rho^2 s_{01}}{2\gamma_1\gamma_2(1+s_{01}-s_{01}y)}+\frac{1}{2}\tilde{m}(y).$$

From the above discussion, we see that this is the value of the a.s. limit of $f'(t)$ as $t \to -y$. The function is well defined for all $y \le 1$ that lie above the upper boundary of the support of the limiting spectral distribution of the multivariate Beta matrix, and for all $s_{01} > \bar{s}$. It is also well defined for positive s_{01} and all, but one, values $y > 1$, but we focus on the supercritical spikes (hence $s_{01} > \bar{s}$) and on y that may be equal to a limit of an eigenvalue of a multivariate Beta matrix (hence $y \le 1$).

Recall that $\tilde{\lambda}_{pj} = \alpha_n\lambda_{pj}/\left(1+\alpha_n\lambda_{pj}\right)$, where $\alpha_n = n_H/n_2$ and λ_{pj} is the j-th largest eigenvalue of $\mathbf{F} = \mathbf{E}^{-1}\mathbf{H}$. This implies that the upper boundary of the support of the limiting spectral distribution of the multivariate Beta matrix $\left(n_H\mathbf{H}+n_2\mathbf{E}\right)^{-1} n_H\mathbf{H}$ equals

$$\tilde{\beta}_+ \equiv \alpha\beta_+/(1+\alpha\beta_+), \text{ where } \alpha = \gamma_2/\gamma_1,$$

and that

$$\tilde{m}(y) = \alpha x + 1 + \alpha^{-1}(\alpha x+1)^2 m(x), \tag{71}$$

where $y = \alpha x/\left(1+\alpha x\right)$ and $m(x)$ is the Stieltjes transform of the limiting spectral distribution of $\mathbf{F}$. It is well known (see, for example, p. 79 of Bai and Silverstein's (2010) book), that

$$m(x) = \frac{1}{x\gamma_1} - \frac{1}{x} - \frac{\gamma_1\left(x(1-\gamma_2)+1-\gamma_1\right)+2x\gamma_2-\gamma_1\sqrt{\left((1-\gamma_1)+x(1-\gamma_2)\right)^2-4x}}{2x\gamma_1(\gamma_1+x\gamma_2)}. \tag{72}$$

Now returning to function Ψ, note that this function is increasing in both y and s_{01} on $y \in (\tilde{\beta}_+, 1]$ and $s_{01} > \bar{s}$. Let us compute its limit as $y \to \tilde{\beta}_+$ and $s_{01} \to \bar{s}$. Then $x \equiv y/(\alpha-\alpha y) \to \beta_+ \equiv ((1+\rho)/(1-\gamma_2))^2$, the square root in the above definition of m converges to zero, and after some algebra, we obtain

$$\frac{1}{2}\tilde{m}(y) \to -\frac{\rho((1+\rho)^2\gamma_2+\gamma_1(1-\gamma_2)^2)}{2\gamma_1\gamma_2(1-\gamma_2)(1+\rho)}.$$

On the other hand, as $s_{01} \to \bar{s} \equiv (\gamma_2+\rho)/(1-\gamma_2)$ and $y \to \tilde{\beta}_+$,

$$\frac{\rho^2 s_{01}}{2\gamma_1\gamma_2(1+s_{01}-s_{01}y)} \to \frac{\rho((1+\rho)^2\gamma_2+\gamma_1(1-\gamma_2)^2)}{2\gamma_1\gamma_2(1-\gamma_2)(1+\rho)},$$

so that $\Psi(y, s_{01}) \to 0$. Since $\Psi(y, s_{01})$ is increasing, the limit (70) of $f'(t)$ must be positive. Moreover, there exists a small positive number such that $f'(t)$ is above this number for all $t \in [-\tilde{\lambda}_{p1}, -\tilde{\lambda}_{p2}-\varepsilon]$ and all $s = s_0 + \delta/\sqrt{p}$ with $(\delta_1, ..., \delta_k)$ from a compact subset of $\mathbb{R}^k$, for sufficiently large p, almost surely.

Since $f(t)$ is strictly increasing on $[-\tilde{\lambda}_{p1}, -\tilde{\lambda}_{p2}-\varepsilon]$, the main contribution to the integral $I(p)$ comes from the vicinity of $-\tilde{\lambda}_{p1}$. From (69), we see that

$$f'(-\tilde{\lambda}_{p1}) \overset{a.s.}{\to} \frac{\rho^2 s_{01}}{2\gamma_1\gamma_2(1+s_{01}-s_{01}\lim\tilde{\lambda}_{p1})}+\frac{1}{2}\tilde{m}(\lim\tilde{\lambda}_{p1}), \tag{73}$$

where

$$\lim\tilde{\lambda}_{p1} \equiv \frac{\left(s_{01}+\gamma_1\right)\left(s_{01}+1\right)\gamma_2}{s_{01}\left(\gamma_1+s_{01}\gamma_2+\gamma_2\right)}.$$

Using (71) and (72), we obtain, after some algebra

$$\tilde{m}(\lim \tilde{\lambda}_{p1}) = \frac{\rho^2}{\gamma_1\gamma_2}\frac{s_{01}(\gamma_1+\gamma_2+\gamma_2 s_{01})}{(\gamma_1+s_{01})(\gamma_2+\gamma_2 s_{01}-s_{01})}. \tag{74}$$

Further,

$$\frac{\rho^2 s_{01}}{\gamma_1\gamma_2(1+s_{01}-s_{01}\lim\tilde{\lambda}_{p1})} = \frac{s_{01}\,(\gamma_1+\gamma_2+\gamma_2 s_{01})}{\gamma_1\gamma_2\,(1+s_{01})}.$$

These equations together with (73) yield

$$f'(-\tilde{\lambda}_{p1}) \overset{a.s.}{\to} H_1 \equiv \frac{(1-\gamma_2)s_{01}(\gamma_1+\gamma_2+\gamma_2 s_{01})(1+s_{01}-\beta_+^{1/2})(1+s_{01}-\beta_-^{1/2})}{2\gamma_1\gamma_2(\gamma_1+s_{01})(1+s_{01})(s_{01}-\gamma_2 s_{01}-\gamma_2)}. \tag{75}$$

The rest of our proof of Lemma DJO18 closely follows Olver (1997), p. 81–82. Consider a new variable of integration

$$v = f(t) - f(-\tilde{\lambda}_{p1}),$$

and let c be a fixed number that belongs to the a.s. limit of $(-\tilde{\lambda}_{p1}, -\tilde{\lambda}_{p2}-\varepsilon)$ as $p, n \to_\gamma \infty$, under the null hypothesis. We have

$$e^{pf(-\tilde{\lambda}_{p1})}\int_{-\tilde{\lambda}_{p1}}^{c} e^{-pf(t)}g(t)\mathrm{d}t = \int_0^{\kappa} e^{-pv}q(v)\mathrm{d}v, \tag{76}$$

where

$$\kappa = f(c) - f(-\tilde{\lambda}_{p1}) \text{ and } q(v) = g(t)\,(\mathrm{d}t/\mathrm{d}v) = g(t)/f'(t).$$

Note that there exist fixed non-random $0 < K_1 < K_2 < \infty$, such that $\kappa \in [K_1, K_2]$.

By definition of $f(t)$ and $g(t)$, as $t \to -\tilde{\lambda}_{p1}$,

$$\left(f(t) - f(-\tilde{\lambda}_{p1})\right) / \left(t + \tilde{\lambda}_{p1}\right) \to F, \text{ and } g(t)/(t+\tilde{\lambda}_{p1})^{-1/2} \to G,$$

where

$$F \equiv f'(-\tilde{\lambda}_{p1}), \text{ and } G \equiv \prod_{j=2}^{k}\left(\left(\tilde{\lambda}_{p1}-z_j\right)/\left(\tilde{\lambda}_{p1}-\tilde{\lambda}_{pj}\right)\right)^{1/2}.$$

These convergences, together with the above definitions of v and $q(v)$ imply that

$$q(v) \sim Gv^{-1/2}/F^{1/2} \text{ as } v \to 0+. \tag{77}$$

This means that the ratio of the left hand side of (77) to its right hand side converges to 1 as $v \to 0+$. This convergece is uniform in $s = s_0 + \delta/\sqrt{p}$ with $(\delta_1, ..., \delta_k)$ from a compact subset of $\mathbb{R}^k$.

Now, closely following Olver, rearrange the integral (76) in the form

$$\int_0^{\kappa} e^{-pv}q(v)\mathrm{d}v = GF^{-1/2}\left\{\int_0^{\infty} e^{-pv}v^{-1/2}\mathrm{d}v - \varepsilon_1(p)\right\} + \varepsilon_2(p), \tag{78}$$

where

$$\varepsilon_1 = \int_{\kappa}^{\infty} e^{-pv}v^{-1/2}\mathrm{d}v \text{ and } \varepsilon_2 = \int_0^{\kappa} e^{-pv}\left\{q(v) - GF^{-1/2}v^{-1/2}\right\}\mathrm{d}v.$$

Since $\kappa \in [K_1, K_2]$ for all sufficiently large p,

$$\int_0^{\infty} e^{-pv}v^{-1/2}\mathrm{d}v - \varepsilon_1(p) \sim \Gamma(1/2)p^{-1/2} = \sqrt{\pi/p} \text{ as } p \to \infty.$$

Further, the above results on the derivative of f yield

$$F \sim H_1 \equiv \frac{(1-\gamma_2)s_{01}(\gamma_1+\gamma_2+\gamma_2 s_{01})(1+s_{01}-\beta_+^{1/2})(1+s_{01}-\beta_-^{1/2})}{2\gamma_1\gamma_2(\gamma_1+s_{01})(1+s_{01})(s_{01}-\gamma_2 s_{01}-\gamma_2)}.$$

Therefore, for the first term on the right hand side of (78), we have

$$GF^{-1/2}\left\{\int_0^\infty e^{-pv}v^{-1/2}\mathrm{d}v-\varepsilon_1(p)\right\}\sim(pH_1/\pi)^{-1/2}\prod_{j=2}^k\left(\left(\tilde\lambda_{p1}-z_j\right)/\left(\tilde\lambda_{p1}-\tilde\lambda_{pj}\right)\right)^{1/2}. \tag{79}$$

That is, the ratio of the left to the right hand sides of the above display converges to one as $p,n\to_\gamma\infty$, uniformly in $s=s_0+\delta/\sqrt{p}$ with $(\delta_1,...,\delta_k)$ from a compact subset of $\mathbb{R}^k$, almost surely.

Next, by (77), for an arbitrarily small positive τ, we can choose κ so that

$$\left|q(v)-GF^{-1/2}v^{-1/2}\right|<\tau\,|G|\,F^{-1/2}v^{-1/2}$$

for all $v\in(0,\kappa]$ and all $s=s_0+\delta/\sqrt{p}$ with $(\delta_1,...,\delta_k)$ from a compact subset of $\mathbb{R}^k$. Therefore, ε_2 is asymptotically dominated by the first term on the right hand side of (78).

Finally, let

$$\eta=\inf_{[c,-\tilde\lambda_{p2}-\varepsilon)}\left\{f(t)-f(-\tilde\lambda_{p1})\right\}.$$

Since $f(t)$ is strictly increasing on $[-\tilde\lambda_{p1},-\tilde\lambda_{p2}-\varepsilon]$, η is larger than some positive number for any $s=s_0+\delta/\sqrt{p}$ with $(\delta_1,...,\delta_k)$ from a compact subset of $\mathbb{R}^k$, for all sufficiently large p, a.s. Therefore,

$$\left|e^{pf(-\tilde\lambda_{p1})}\int_c^{-\tilde\lambda_{p2}-\varepsilon}e^{-pf(t)}g(t)\mathrm{d}t\right|\le e^{-p\theta}\int_c^{-\tilde\lambda_{p2}-\varepsilon}|g(t)|\,\mathrm{d}t,$$

which is dominated by the right hand side of (79).

Summing up, we have established the following lemma.

LEMMA 17. *As $p,n\to_\gamma\infty$,*

$$\overline{\mathcal{I}}^{\mathrm{SigD}}_{\sigma_\varnothing,in}=G^{\mathrm{SigD}}_{-1}e^{-pf_1(\tilde\lambda_{p1})}(pH_1/\pi)^{-1/2}\prod_{j=2}^k\left(\left(\tilde\lambda_{p1}-z_j\right)/\left(\tilde\lambda_{p1}-\tilde\lambda_{pj}\right)\right)^{1/2}(1+o(1)),$$

where $o(1)\to0$ uniformly over $s=s_0+\delta/\sqrt{p}$ with $(\delta_1,...,\delta_k)$ from a compact subset of $\mathbb{R}^k$ and over $z_2,...,z_k$ that belong to the (trimmed) domain of integration in (DJO49), a.s.

Repeating the above analysis for the second, third, etc. to the inner-most integral in (DJO49) and combining the results, we obtain Lemma DJO18.

5.6. *Proof of Lemma DJO19 about the Laplace approximation for $\overline{\mathcal{I}}^{\mathrm{REG}}_{\sigma_\varnothing}$.* The proof is very similar to that of Lemma DJO18. As shown in DJO, the inner-most integral in the multiple integral representation (DJO52) of $\overline{\mathcal{I}}^{\mathrm{REG}}_{\sigma_\varnothing}$ equals

$$\overline{\mathcal{I}}^{\mathrm{REG}}_{\sigma_\varnothing,in}=G^{\mathrm{REG}}_{-1}\int_{\tilde\lambda_{p2}+\varepsilon}^{\tilde\lambda_{p1}}e^{-n_Hf_2(z_1)}g_2(z_1)(1+o(1))\,\mathrm{d}z_1. \tag{80}$$

First, we will apply the Laplace method to the integral

$$I(p) \equiv \int_{\tilde{\lambda}_{p2}+\varepsilon}^{\tilde{\lambda}_{p1}} e^{-n_H f_2(z_1)} g_2(z_1)\,\mathrm{d}z_1.$$

To line up our analysis with that on pages 81–82 of Olver (1997) and to simplify notations, let us rewrite $I(p)$ as

$$I(p) = \int_{-\tilde{\lambda}_{p1}}^{-\tilde{\lambda}_{p2}-\varepsilon} e^{-n_H f(t)} g(t)\mathrm{d}t,$$

where $t = -z_1$,

$$\begin{aligned} f(t) &= -z_{1+} - a\ln(z_{1+}+a) + b\ln(z_{1+}+b) + \frac{1}{2n_H}\sum_{j=k+1}^{p} \ln\left(-t-\tilde{\lambda}_{pj}\right), \text{ and} \\ g(t) &= \left(\frac{z_{1+}(z_{1+}+a)}{z_{1+}^2 + a\zeta_1}\right)^{\frac{1}{2}} \left(\tilde{\lambda}_{p1}+t\right)^{-1/2} \prod_{j=2}^{k} \left(\frac{-t-z_j}{-t-\tilde{\lambda}_{pj}}\right)^{1/2}. \end{aligned}$$

Here

$$z_{j+} = \frac{1}{2}\left\{\zeta_j - b + \sqrt{(b-\zeta_j)^2 + 4a\zeta_j}\right\} \text{ with } \zeta_j = -ts_j/2,$$
$$a = (N-p)/(2n_H) \equiv (n_H + n_2 - p)/(2n_H), \text{ and } b = (n_H - p)/(2n_H).$$

Let us show that under the null hypothesis, $f(t)$ has positive continuous derivative for t from the interval $[-\tilde{\lambda}_{p1}, -\tilde{\lambda}_{p2}-\varepsilon]$ for sufficiently large p, a.s. Express $f'(t)$ as

$$f'(t) = \varphi'(t) + w'(t),$$

where

$$\varphi(t) \equiv -z_{1+} - a\ln(z_{1+}+a) + b\ln(z_{1+}+b) \text{ and } w(t) \equiv \frac{1}{2n_H}\sum_{j=k+1}^{p} \ln(-t-\tilde{\lambda}_{pj}).$$

Clearly, function $w(t)$ is decreasing and concave on $t \in [-\tilde{\lambda}_{p1}, -\tilde{\lambda}_{p2}-\varepsilon]$. We will now show that $\varphi(t)$ is increasing and convex.

Since

$$z_{1+}^2 + (b-\zeta_1)z_{1+} - a\zeta_1 = 0,$$

we have

$$\zeta_1 = z_{1+}(z_{1+}+b)/(z_{1+}+a). \tag{81}$$

Therefore, $z_{1+} > \zeta_1 > 0$. Further,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} z_{1+} &= \frac{\partial z_{1+}}{\partial \zeta_1}\frac{\mathrm{d}\zeta_1}{\mathrm{d}t} = \frac{1}{2}\left\{1 + \frac{\zeta_1 - b + 2a}{\sqrt{(b-\zeta_1)^2 + 4a\zeta_1}}\right\}\frac{-s_1}{2} \\ &= \frac{z_{1+}+a}{2z_{1+}+b-\zeta_1}\frac{-s_1}{2} = \frac{(a+z_{1+})^2}{ab+2az_{1+}+z_{1+}^2}\frac{-s_1}{2} < 0. \end{aligned} \tag{82}$$

On the other hand,

$$\frac{\partial}{\partial z_{1+}}\varphi = -\frac{ab+2az_{1+}+z_{1+}^2}{(b+z_{1+})(a+z_{1+})} < 0. \tag{83}$$

Therefore, $\varphi'(t)$ is positive on $t \in [-\tilde{\lambda}_{p1}, -\tilde{\lambda}_{p2} - \varepsilon]$. Further,

$$\frac{\partial^2}{\partial z_{1+}^2}\varphi = -\frac{b}{(z_{1+}+b)^2} + \frac{a}{(z_{1+}+a)^2} = \frac{(a-b)\left(z_{1+}^2 - ab\right)}{(z_{1+}+b)^2 (z_{1+}+a)^2},$$

and using (82), we also have

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} z_{+1} = -\frac{s_1^2}{2} \frac{a\,(a+z_{1+})^3 (a-b)}{\left(ab + 2az_{1+} + z_{1+}^2\right)^3}.$$

Hence,

$$\varphi''(t) = \frac{\partial^2 \varphi}{\partial z_{1+}^2}\left(\frac{\mathrm{d}z_{1+}}{\mathrm{d}t}\right)^2 + \frac{\partial \varphi}{\partial z_{1+}}\frac{\mathrm{d}^2 z_{1+}}{\mathrm{d}t^2} = \frac{s_1^2}{4}\frac{(a+z_{1+})^2 (a-b)}{(b+z_{1+})^2\left(ab+2az_{1+}+z_{1+}^2\right)} > 0.$$

The concavity of $w(t)$ and the convexity of $\varphi(t)$ imply that

$$\min_{t\in\left[-\tilde{\lambda}_{p1}, -\tilde{\lambda}_{p2}-\varepsilon\right]} f'(t) > \varphi'\left(-\tilde{\lambda}_{p1}\right) + w'\left(-\tilde{\lambda}_{p2} - \varepsilon\right). \tag{84}$$

On the other hand, using, first, (82) and (83), and then (81) and the definition of ζ_1, we obtain

$$\varphi'(t) = \frac{s_1(a+z_{1+})}{2(b+z_{1+})} = -z_{1+}/t.$$

This and the fact that

$$\tilde{\lambda}_{p1} \overset{a.s.}{\to} \frac{(s_{01}+\gamma_1)(s_{01}+1)\gamma_2}{s_{01}(\gamma_1 + s_{01}\gamma_2 + \gamma_2)}$$

yield, after some algebra,

$$\varphi'(-\tilde{\lambda}_{p1}) \overset{a.s.}{\to} \frac{s_{01}(\gamma_1 + s_{01}\gamma_2 + \gamma_2)}{2(1+s_{01})\gamma_2}. \tag{85}$$

For $w'(-\tilde{\lambda}_{p2} - \varepsilon)$, we have

$$w'(-\tilde{\lambda}_{p2} - \varepsilon) \overset{a.s.}{\to} \frac{\gamma_1}{2}\tilde{m}(\lim \tilde{\lambda}_{p2} + \varepsilon).$$

Hence, the right hand side of (84) a.s. converges to

$$\Pi(\lim \tilde{\lambda}_{p2} + \varepsilon, s_{01}) \equiv \frac{s_{01}(\gamma_1 + s_{01}\gamma_2 + \gamma_2)}{2(1+s_{01})\gamma_2} + \frac{\gamma_1}{2}\tilde{m}(\lim \tilde{\lambda}_{p2} + \varepsilon).$$

This convergence is uniform in $s = s_0 + \delta/\sqrt{p}$ with $(\delta_1, ..., \delta_k)$ from a compact subset of $\mathbb{R}^k$.

Now note that $\Pi(y, s_{01})$ is strictly increasing function of y and s_{01} on $y \in (\tilde{\beta}_+, 1]$ and $s_{01} > \bar{s}$. Using the same tools as in the above proof of Lemma DJO18 (more specifically, those used for the analysis of $\Psi(y, s_{01})$), we find that the limit of $\Pi(y, s_{01})$ as $y \to \tilde{\beta}_+$ and $s_{01} \to \bar{s}$ is zero. Therefore, $\Pi(\lim \tilde{\lambda}_{p2} + \varepsilon, s_{01})$ is positive, and thus, by (84), $f'(t)$ is positive on $t \in [-\tilde{\lambda}_{p1}, -\tilde{\lambda}_{p2} - \varepsilon]$. Moreover, there exists a small positive number that is smaller than $f'(t)$ for all $t \in [-\tilde{\lambda}_{p1}, -\tilde{\lambda}_{p2} - \varepsilon]$, all $s = s_0 + \delta/\sqrt{p}$ with $(\delta_1, ..., \delta_k)$ from a compact subset of $\mathbb{R}^k$, and all sufficiently large p, a.s.

Since $f(t)$ is strictly increasing on $[-\tilde{\lambda}_{p1}, -\tilde{\lambda}_{p2} - \varepsilon]$, the main contribution to the integral $I(p)$ comes from the vicinity of $-\tilde{\lambda}_{p1}$. Using (85) and (74), we obtain

$$f'(-\tilde{\lambda}_{p1}) \overset{a.s.}{\to} \gamma_1 H_1.$$

The rest of the proof is almost identical to that of Lemma DJO18. The only notable difference is that function $g(t)$ has an additional multiplicative term $\left(\frac{z_{1+}(z_{1+}+a)}{z_{1+}^2+a\zeta_1}\right)^{\frac{1}{2}}$. It is straightforward to verify that, at $t=-\tilde{\lambda}_{p1}$, this term a.s. converges to

$$\frac{\gamma_1+\gamma_2+\gamma_2 s_{01}}{((\gamma_1+\gamma_2+\gamma_2 s_{01})^2-\rho^2\gamma_1)^{1/2}},$$

which explains the presence of the last term in the definition of Ω_j^{REG} given in Lemma DJO19.

5.7. *Proof of Lemma DJO20 about the asymptotic equivalence of* $\mathcal{I}_{\sigma_\varnothing}$ *and* $\overline{\mathcal{I}}_{\sigma_\varnothing}$. By definition,

$$\mathcal{I}_{\sigma_\varnothing}-\overline{\mathcal{I}}_{\sigma_\varnothing}=\sum_{\mathcal{D}}\int_{\tilde{x}_0}^{\tilde{\lambda}_{pk}}\int_{\mathcal{D}_{k-1}}\cdots\int_{\mathcal{D}_1}|\omega|\,\mathcal{F}_{p,n}\prod_{j=1}^{k}\mathrm{d}z_j, \tag{86}$$

where the sum runs over all $\mathcal{D}_j$ that are represented by either $[\tilde{\lambda}_{j+1}+\varepsilon,\tilde{\lambda}_j]$ or $[\tilde{\lambda}_{j+1},\tilde{\lambda}_{j+1}+\varepsilon]$, and at least one $\mathcal{D}_j$, $j=1,...,k-1$, is represented by $[\tilde{\lambda}_{j+1},\tilde{\lambda}_{j+1}+\varepsilon]$. All terms in the above sum can be analyzed symilarly, and here we will focus on the term

$$J\equiv\int_{\tilde{x}_0}^{\tilde{\lambda}_{pk}}\int_{\tilde{\lambda}_{pk}+\varepsilon}^{\tilde{\lambda}_{p,k-1}}\cdots\int_{\tilde{\lambda}_{p3}+\varepsilon}^{\tilde{\lambda}_{p2}}\int_{\tilde{\lambda}_{p2}}^{\tilde{\lambda}_{p2}+\varepsilon}|\omega|\,\mathcal{F}_{p,n}\prod_{j=1}^{k}\mathrm{d}z_j.$$

As is explained in DJO, we have

$$|J|\le\bar{J}\equiv\int_{\tilde{x}_0}^{\tilde{\lambda}_{pk}}\int_{\tilde{\lambda}_{pk}+\varepsilon}^{\tilde{\lambda}_{p,k-1}}\cdots\int_{\tilde{\lambda}_{p3}+\varepsilon}^{\tilde{\lambda}_{p2}}\int_{\tilde{\lambda}_{p2}}^{\tilde{\lambda}_{p2}+\varepsilon}|\omega|\,\mathcal{F}_{p,n}\left(\tilde{Z}_1\right)\prod_{j=1}^{k}\mathrm{d}z_j, \tag{87}$$

where $\tilde{Z}_1=\mathrm{diag}\left\{\tilde{\lambda}_{p2}+\varepsilon,z_2,...,z_k\right\}$.

Case SigD. Using Lemma DJO16 to approximate $\mathcal{F}_{p,n}\left(\tilde{Z}_1\right)$, we obtain

$$\bar{J}=b^{\mathrm{SigD}}(1+o(1))\int_{\tilde{x}_0}^{\tilde{\lambda}_{pk}}\int_{\tilde{\lambda}_{pk}+\varepsilon}^{\tilde{\lambda}_{p,k-1}}\cdots\int_{\tilde{\lambda}_{p3}+\varepsilon}^{\tilde{\lambda}_{p2}}\int_{\tilde{\lambda}_{p2}}^{\tilde{\lambda}_{p2}+\varepsilon}e^{-pf_1(\tilde{z}_1)}\tilde{g}_1\mathrm{d}z_1\prod_{j=2}^{k}\left\{e^{-pf_j(z_j)}g_j\mathrm{d}z_j\right\}, \tag{88}$$

where $o(1)\to 0$ uniformly in $s=s_0+\delta/\sqrt{p}$ with $(\delta_1,...,\delta_k)$ from a compact subset of $\mathbb{R}^k$, and the definitions of f_j, g_j, $\tilde{z}_1$, and $\tilde{g}_1$ are summarized in Table 1.

TABLE 1
Definition of f_j, g_j, $\tilde{z}_1$, and $\tilde{g}_1$ used in equation (88).

Quantity		Definition
$f_j(z)$	$=$	$\frac{N-p+2}{2p}\ln\left(1-\frac{s_j z}{1+s_j}\right)+\frac{1}{2p}\sum_{i=k+1}^{p}\ln\left(z-\tilde{\lambda}_{pi}\right),j=1,...,k,$
g_j	$=$	$\prod_{s=1}^{j}\left(\tilde{\lambda}_{ps}-z_j\right)^{-1/2}\prod_{i=j+1}^{k}\left((z_j-z_i)/\left(z_j-\tilde{\lambda}_{pi}\right)\right)^{1/2},j=2,...,k,$
$\tilde{z}_1$	$=$	$\tilde{\lambda}_{p2}+\varepsilon,$
$\tilde{g}_1$	$=$	$\left(\tilde{\lambda}_{p1}-\tilde{z}_1\right)^{-1/2}\prod_{i=2}^{k}\left\{\left(\frac{\tilde{z}_1-z_i}{\tilde{z}_1-\tilde{\lambda}_{pi}}\right)^{1/2}\frac{z_1-z_i}{\tilde{z}_1-z_i}\right\}\prod_{s=1}^{p}\left\|\frac{z_1-\tilde{\lambda}_{ps}}{\tilde{z}_1-\tilde{\lambda}_{ps}}\right\|^{-1/2}.$

As follows from the proof of Lemma DJO18, $f_j(z), j=1,...,k$ are strictly decreasing functions for z from the integration domain of J. Therefore, we have

$$\begin{aligned}\bar{J} &\leq b^{\mathrm{SigD}}(1+o(1))\int_{\tilde{x}_0}^{\tilde{\lambda}_{pk}}\int_{\tilde{\lambda}_{pk}+\varepsilon}^{\tilde{\lambda}_{p,k-1}}\cdots\int_{\tilde{\lambda}_{p3}+\varepsilon}^{\tilde{\lambda}_{p2}}\int_{\tilde{\lambda}_{p2}}^{\tilde{\lambda}_{p2}+\varepsilon} e^{-pf_1(\tilde{z}_1)}\tilde{g}_1\mathrm{d}z_1\prod_{j=2}^{k}\left\{e^{-pf_j(\tilde{\lambda}_{pj})}g_j\mathrm{d}z_j\right\}\\ &\leq C^{\mathrm{SigD}}e^{-pf_1(\tilde{\lambda}_{p2}+\varepsilon)}\prod_{j=2}^{k}e^{-pf_j(\tilde{\lambda}_{pj})},\end{aligned}$$

where C^{SigD} is a positive quantity that depends only on $s_{01},...,s_{0k}$.

In the proof of Lemma DJO18, we have seen that not only $f_1(z)$ is strictly decreasing in z on $z\in[\tilde{\lambda}_{p2},\tilde{\lambda}_{p1}]$, but also there exists a fixed negative number such that the derivative of $f_1(z)$ is smaller than that number. Therefore, there exists $\eta>0$, such that $f_1(\tilde{\lambda}_{p2}+\varepsilon)>f_1(\tilde{\lambda}_{p1})+\eta$ and

$$\bar{J}\leq C^{\mathrm{SigD}}e^{-p\eta}\prod_{j=1}^{k}e^{-pf_j(\tilde{\lambda}_{pj})}.$$

On the other hand, by Lemma DJO18, the right hand side of the displayed inequality is asymptotically dominated by $\overline{\mathcal{I}}_{\sigma_\varnothing}$. Repeating the above arguments for the other components of $\mathcal{I}_{\sigma_\varnothing}-\overline{\mathcal{I}}_{\sigma_\varnothing}$ (that is, the components of the sum in (86) other than J), we establish Lemma DJO20 for the SigD case.

Case REG. A proof of this case is similar to that for SigD. Using Lemma DJO17 to approximate $\mathcal{F}_{p,n}\left(\tilde{Z}_1\right)$, we obtain

$$\bar{J}=b^{\mathrm{REG}}(1+o(1))\int_{\tilde{x}_0}^{\tilde{\lambda}_{pk}}\int_{\tilde{\lambda}_{pk}+\varepsilon}^{\tilde{\lambda}_{p,k-1}}\cdots\int_{\tilde{\lambda}_{p3}+\varepsilon}^{\tilde{\lambda}_{p2}}\int_{\tilde{\lambda}_{p2}}^{\tilde{\lambda}_{p2}+\varepsilon} e^{-n_Hf_1(\tilde{z}_1)}\tilde{g}_1\mathrm{d}z_1\prod_{j=2}^{k}\left\{e^{-n_Hf_j(z_j)}g_j\mathrm{d}z_j\right\},\tag{89}$$

where $o(1)\to 0$ uniformly in $s=s_0+\delta/\sqrt{p}$ with $(\delta_1,...,\delta_k)$ from a compact subset of $\mathbb{R}^k$, and the definitions of f_j, g_j, $\tilde{z}_1$, and $\tilde{g}_1$ are summarized in Table 2.

Proceeding exactly as in SigD case, we obtain inequality

$$\bar{J}\leq C^{\mathrm{REG}}e^{-n_H\eta}\prod_{j=1}^{k}e^{-n_Hf_j(\tilde{\lambda}_{pj})},$$

so Lemma DJO20 follows by a similar argument.

Table 2
Definition of f_j, g_j, $\tilde{z}_1$, and $\tilde{g}_1$ used in equation (89).

Quantity		Definition
$f_j(z_j)$	$=$	$-z_{j+}-a\ln(z_{j+}+a)+b\ln(z_{j+}+b)+\frac{1}{2n_H}\sum_{i=k+1}^{p}\ln(z_j-\tilde{\lambda}_{pi}),\ j=2,...,k,$
g_j	$=$	$\left(\frac{z_{j+}(z_{j+}+a)}{z_{j+}^2+a\zeta_j}\right)^{1/2}\prod_{s=1}^{j}\left(\tilde{\lambda}_{ps}-z_j\right)^{-1/2}\prod_{i=j+1}^{k}\left(\frac{z_j-z_i}{z_j-\tilde{\lambda}_{pi}}\right)^{1/2},\ j=2,...,k,$
$\tilde{z}_1$	$=$	$\tilde{\lambda}_{p2}+\varepsilon,$
$f_1(\tilde{z}_1)$	$=$	$-\tilde{z}_{1+}-a\ln(\tilde{z}_{1+}+a)+b\ln(\tilde{z}_{1+}+b)+\frac{1}{2n_H}\sum_{i=k+1}^{p}\ln(\tilde{z}_1-\tilde{\lambda}_{pi}),$
$\tilde{g}_1$	$=$	$\left(\frac{\tilde{z}_{1+}(\tilde{z}_{1+}+a)}{\tilde{z}_{1+}^2+a\tilde{\zeta}_1}\right)^{1/2}\left(\tilde{\lambda}_{p1}-\tilde{z}_1\right)^{-1/2}\prod_{i=2}^{k}\left\{\left(\frac{\tilde{z}_1-z_i}{\tilde{z}_1-\tilde{\lambda}_{pi}}\right)^{1/2}\frac{z_1-z_i}{\tilde{z}_1-z_i}\right\}\prod_{s=1}^{p}\left\lvert\frac{z_1-\tilde{\lambda}_{ps}}{\tilde{z}_1-\tilde{\lambda}_{ps}}\right\rvert^{-1/2}.$
$\tilde{z}_{1+}$	$=$	$\frac{1}{2}\left\{\tilde{\zeta}_1-b+\sqrt{\left(b-\tilde{\zeta}_1\right)^2+4a\tilde{\zeta}_1}\right\},$
$\tilde{\zeta}_1$	$=$	$\tilde{z}_1s_1/2.$

5.8. *Proof of Lemma DJO21 about the asymptotic negligibility of $\mathcal{I}_{\sigma_\tau}$ with $\tau \neq \varnothing$.* We consider here only the case of $\tau = 1, 2, ..., k$. The analysis for the other subsets $\tau \subseteq 1, 2, ..., k$ is very similar.

As shown in DJO,

$$|\mathcal{I}_{\sigma_\tau}| \leq \frac{1}{2^k} \int_{\mathcal{K}_{k+1}} \cdots \int_{\mathcal{K}_{k+1}} |\omega| \, \mathcal{F}_{p,n}(Z_\eta) \prod_{j=1}^{k} |\mathrm{d}z_j| \,,$$

where $Z_\eta = \operatorname{diag}\{\tilde{x}_0 + k\eta, ..., \tilde{x}_0 + 2\eta, \tilde{x}_0 + \eta\}$ and η is a fixed small positive number. Function $\mathcal{F}_{p,n}(Z_\eta)$ can now be approximated using Lemmas DJO16 and DJO17.

Case SigD. Using Lemma DJO16, we obtain

$$|\mathcal{I}_{\sigma_\tau}| \leq C \prod_{j=1}^{k} \left(1 - \frac{s_j(\tilde{x}_0 + j\eta)}{1 + s_j}\right)^{-\frac{N-p+2}{2}} \int_{\mathcal{K}_{k+1}} \cdots \int_{\mathcal{K}_{k+1}} |\omega| \prod_{j=1}^{k} |\mathrm{d}z_j| \,, \tag{90}$$

where C is some positive constant. The above inequality holds uniformly in $s = s_0 + \delta/\sqrt{p}$ with $(\delta_1, ..., \delta_k)$ from a compact subset of $\mathbb{R}^k$, for sufficiently large p, a.s.

Using the definitions of ω and $\mathcal{K}_{k+1}$, we obtain

$$\int_{\mathcal{K}_{k+1}} \cdots \int_{\mathcal{K}_{k+1}} |\omega| \prod_{j=1}^{k} |\mathrm{d}z_j| \leq C_1 \prod_{j=1}^{k} \prod_{s=k+1}^{p} (\tilde{x}_0 - \tilde{\lambda}_{ps})^{-1/2}, \tag{91}$$

where C_1 is a positive constant. Combining this inequality with (90), we obtain

$$|\mathcal{I}_{\sigma_\tau}| \leq C_2 \prod_{j=1}^{k} \left[\left(1 - \frac{s_j(\tilde{x}_0 + j\eta)}{1 + s_j}\right)^{-\frac{N-p+2}{2}} \prod_{s=k+1}^{p} (\tilde{x}_0 - \tilde{\lambda}_{ps})^{-1/2} \right].$$

Now recall (from the proof of Lemma DJO18) that functions

$$f_j(z) = \frac{N - p + 2}{2p} \ln\left(1 - \frac{s_j z}{1 + s_j}\right) + \frac{1}{2p} \sum_{j=k+1}^{p} \ln\left(z - \tilde{\lambda}_{ps}\right), \; j = 1, ..., k,$$

are decreasing on $z \in [\tilde{x}_0, 1]$ and have there derivatives that are bounded away from zero, for sufficiently large p, a.s. This implies that there exists a small positive η_1 such that

$$|\mathcal{I}_{\sigma_\tau}| \leq C_3 e^{-p\eta_1} \prod_{j=1}^{k} e^{-p f_j(\tilde{\lambda}_{pj})},$$

and therefore, by Lemma DJO18, $\mathcal{I}_{\sigma_\tau}$ is asymptotically dominated by $\mathcal{I}_{\sigma_\varnothing}$.

Case REG. Using Lemma DJO17, we obtain

$$|\mathcal{I}_{\sigma_\tau}| \leq C \prod_{j=1}^{k} e^{-n_H \varphi_j(\tilde{x}_0 + j\eta)} \int_{\mathcal{K}_{k+1}} \cdots \int_{\mathcal{K}_{k+1}} |\omega| \prod_{j=1}^{k} |\mathrm{d}z_j| \,, \tag{92}$$

where C is some positive constant and

$$\begin{aligned}
\varphi_j(z) &= -z_+ - a\ln(z_+ + a) + b\ln(z_+ + b), \\
z_+ &= \frac{1}{2}\left\{\zeta - b + \sqrt{(b-\zeta)^2 + 4a\zeta}\right\} \text{ with } \zeta = z s_j/2.
\end{aligned}$$

Using (91) in (92), we obtain

$$|\mathcal{I}_{\sigma_\tau}| \leq C_2 \prod_{j=1}^{k} \left[e^{-n_H \varphi_j(\tilde{x}_0 + j\eta)} \prod_{s=k+1}^{p} (\tilde{x}_0 - \tilde{\lambda}_{ps})^{-1/2} \right].$$

Now recall (from the proof of Lemma DJO19) that functions

$$f_j(z) = \varphi_j(z) + \frac{1}{2n_H} \sum_{j=k+1}^{p} \ln\left(z - \tilde{\lambda}_{ps}\right),\ j = 1, ..., k,$$

are decreasing on $z \in [\tilde{x}_0, 1]$ and have there derivatives that are bounded away from zero, for sufficiently large p, a.s. This implies that there exists a small positive η_1 such that

$$|\mathcal{I}_{\sigma_\tau}| \leq C_3 e^{-p\eta_1} \prod_{j=1}^{k} e^{-p f_j(\tilde{\lambda}_{pj})},$$

and therefore, by Lemma DJO19, $\mathcal{I}_{\sigma_\tau}$ is asymptotically dominated by $\mathcal{I}_{\sigma_\varnothing}$.

6. Local Asymptotic Normality.

6.1. *Proof of Theorem DJO23 about the quadratic approximation of the log likelihood ratio.* Case SigD. Theorem DJO22 yields

$$L_p^{\mathrm{SigD}}(\theta, \Lambda) = (1 + o(1)) \frac{k_{p,n}^{\mathrm{SigD}}(s)}{k_{p,n}^{\mathrm{SigD}}(s_0)} \prod_{j=1}^{k} \left(\frac{1 - \frac{s_j}{1+s_j} \tilde{\lambda}_{pj}}{1 - \frac{s_{0j}}{1+s_{0j}} \tilde{\lambda}_{pj}} \right)^{\frac{p-N-2}{2}}. \tag{93}$$

The right hand side does not depend on $b^{\mathrm{SigD}}(s)/b^{\mathrm{SigD}}(s_0)$ because the latter ratio is asymptotically equivalent to one. Taking logarithm of both sides of (93) and simplifying, we obtain

$$\ln L_p^{\mathrm{SigD}}(\theta, \Lambda) = \sum_{j=1}^{k} \left\{ \frac{n_2}{2} \ln \frac{1 + s_j}{1 + s_{0j}} - \frac{p}{2} \ln \frac{s_j}{s_{0j}} - \frac{N - p}{2} \ln \frac{1 + s_j - s_j \tilde{\lambda}_{pj}}{1 + s_{0j} - s_{0j} \tilde{\lambda}_{pj}} \right\} + o(1). \tag{94}$$

Using the identity

$$\tilde{\lambda}_{pj} = \frac{\alpha_n \lambda_{pj}}{1 + \alpha_n \lambda_{pj}} \text{ with } \alpha_n = n_H / n_2, \tag{95}$$

we rewrite (94) as

$$\ln L_p^{\mathrm{SigD}}(\theta, \Lambda) = \sum_{j=1}^{k} \left\{ \frac{n_2}{2} \ln \frac{1 + s_j}{1 + s_{0j}} - \frac{p}{2} \ln \frac{s_j}{s_{0j}} - \frac{N - p}{2} \ln \frac{1 + s_j + \alpha_n \lambda_{pj}}{1 + s_{0j} + \alpha_n \lambda_{pj}} \right\} + o(1). \tag{96}$$

Expanding the logarithms in (96) in the powers of $\theta p^{-1/2}$ up to and including terms $\theta^2 p^{-1}$, we obtain

$$\begin{aligned}
\ln L_p^{\mathrm{SigD}}(\theta, \Lambda) = \sum_{j=1}^{k} \quad & \left\{ \frac{n_2}{2} \left(\theta_j p^{-1/2} \frac{w(s_{0j})}{1 + s_{0j}} - \frac{1}{2} \theta_j^2 p^{-1} \frac{w^2(s_{0j})}{(1 + s_{0j})^2} + o(p^{-1}) \right) \right. \\
& - \frac{p}{2} \left(\theta_j p^{-1/2} \frac{w(s_{0j})}{s_{0j}} - \frac{1}{2} \theta_j^2 p^{-1} \frac{w^2(s_{0j})}{s_{0j}^2} + o(p^{-1}) \right) \\
& \left. - \frac{N - p}{2} \left(\theta_j p^{-1/2} \frac{w(s_{0j})}{1 + s_{0j} + \alpha_n \lambda_{pj}} - \frac{1}{2} \theta_j^2 p^{-1} \frac{w^2(s_{0j})}{(1 + s_{0j} + \alpha_n \lambda_{pj})^2} + o(p^{-1}) \right) \right\} \\
& + o(1).
\end{aligned}$$

Consider, first, the terms linear in θ_j. They can be rewritten as

$$T_{\theta_j} \equiv \frac{1}{2}\theta_j p^{1/2} w(s_{0j}) \left(\frac{1}{c_2} \frac{1}{1+s_{0j}} - \frac{1}{s_{0j}} - \frac{r^2}{c_1 c_2} \frac{1}{1+s_{0j}+\alpha_n \lambda_{pj}} \right).$$

Expanding the last term in the brackets around $\lambda_{pj} = l_j$ and using the fact that, by Theorem DJO10, $\sqrt{p}(\lambda_{pj} - l_j) = O_{\mathrm{P}}(1)$, we get

$$T_{\theta_j} = \frac{1}{2}\theta_j p^{1/2} w(s_{0j}) \left(\frac{1}{c_2} \frac{1}{1+s_{0j}} - \frac{1}{s_{0j}} - \frac{r^2}{c_1 c_2} \frac{1}{1+s_{0j}+\alpha_n l_j} + \frac{r^2}{c_1^2} \frac{\lambda_{pj} - l_j}{(1+s_{0j}+\alpha_n l_j)^2} \right) + o_{\mathrm{P}}(1).$$

Simplifying this using identities

$$\begin{aligned} \alpha_n &= c_2/c_1, \\ r^2 &= c_1 + c_2 - c_1 c_2, \text{ and} \\ l_j &= (s_{0j}+c_1)(s_{0j}+1)/((1-c_2)s_{0j} - c_2), \end{aligned} \tag{97}$$

we obtain

$$T_{\theta_j} = \theta_j \sqrt{p}(\lambda_{pj} - l_j) w(s_{0j}) \frac{((1-c_2)s_{0j} - c_2)^2}{2r^2(1+s_{0j})^2 s_{0j}^2} + o_{\mathrm{P}}(1).$$

On the other hand, for SigD,

$$w(s_{0j}) = 2(\rho s_{0j}(1+s_{0j})/((1-\gamma_2)s_{0j} - \gamma_2))^2, \tag{98}$$

which yields

$$T_{\theta_j} = \theta_j \sqrt{p}(\lambda_{pj} - l_j) + o_{\mathrm{P}}(1). \tag{99}$$

Now, for the quadratic terms in θ_j, we have

$$T_{\theta_j^2} \equiv \frac{1}{4}\theta_j^2 w^2(s_{0j}) \left(-\frac{1}{c_2(1+s_{0j})^2} + \frac{1}{s_{0j}^2} + \frac{r^2}{c_1 c_2 (1+s_{0j}+\alpha_n \lambda_{pj})^2} \right).$$

Using identities (97) and the fact that $\lambda_{pj} - l_j = o_{\mathrm{P}}(1)$, we obtain

$$T_{\theta_j^2} = -\frac{1}{4}\theta_j^2 w^2(s_{0j}) \frac{(1-c_2)s_{0j}^2 - 2c_2 s_{0j} - c_1 - c_2}{(1+s_{0j})^2 s_{0j}^2 r^2} + o_{\mathrm{P}}(1). \tag{100}$$

Recall that the asymptotic variance of λ_{pj} equals $\tau^2(s_{0j}) = w(s_{0j})t(s_{0j})$, (see Theorem DJO10), where

$$t(s_{0j}) \equiv (1-\gamma_2)(s_{0j} - \underline{s})(s_{0j} - \bar{s})/d^2(s_{0j}) = \frac{(1-\gamma_2)s_{0j}^2 + 2\gamma_2 s_{0j} - \gamma_1 - \gamma_2}{((1-\gamma_2)s_{0j} - \gamma_2)^2},$$

Using this with (98) in (100), we obtain

$$T_{\theta_j^2} = -\frac{1}{2}\theta_j^2 \tau^2(s_{0j}) + o_{\mathrm{P}}(1). \tag{101}$$

Combining (99) and (101) yields

$$\ln L_p^{\mathrm{SigD}}(\theta, \Lambda) = \sum_{j=1}^{k} \left\{ \theta_j \sqrt{p}(\lambda_{pj} - l_j) - \frac{1}{2}\theta_j^2 \tau^2(s_{0j}) \right\} + o_{\mathrm{P}}(1).$$

It remains to note that, by construction, the above $o_{\mathrm{P}}(1)$ term is uniform in $(\theta_1, ..., \theta_k)$ from any compact subset of $\mathbb{R}^k$.

Case REG. Theorem DJO22 yields

$$L_p^{\mathrm{REG}}(\theta, \Lambda) = (1 + o(1)) \frac{k_{p,n}^{\mathrm{REG}}(s)}{k_{p,n}^{\mathrm{REG}}(s_0)} \prod_{j=1}^{k} e^{-n_H(\varphi(\bar{z}_{j+}) - \varphi(\bar{z}_{0j+}))}, \tag{102}$$

where

$$\begin{aligned} \bar{z}_{j+} &= \frac{1}{2}\left\{\tilde{\lambda}_{pj} s_j/2 - b + \sqrt{(b - \tilde{\lambda}_{pj} s_j/2)^2 + 2a\tilde{\lambda}_{pj} s_j}\right\}, \\ \bar{z}_{0j+} &= \frac{1}{2}\left\{\tilde{\lambda}_{pj} s_{0j}/2 - b + \sqrt{(b - \tilde{\lambda}_{pj} s_{0j}/2)^2 + 2a\tilde{\lambda}_{pj} s_{0j}}\right\}, \end{aligned}$$

and

$$\varphi(z) = -z - a\ln(z + a) + b\ln(z + b).$$

Taking logarithm of both sides of (102) and simplifying, we obtain

$$\ln L_p^{\mathrm{REG}}(\theta, \Lambda) = \sum_{j=1}^{k} \left\{ -\frac{n_H}{2}(s_j - s_{0j}) - \frac{p}{2} \ln \frac{s_j}{s_{0j}} - n_H(\varphi(\bar{z}_{j+}) - \varphi(\bar{z}_{0j+})) \right\} + o(1). \tag{103}$$

Expanding the difference $\bar{z}_{j+} - \bar{z}_{0j+}$ in the powers of $\theta p^{-1/2}$ up to and including terms $\theta^2 p^{-1}$, we obtain

$$\bar{z}_{j+} - \bar{z}_{0j+} = \frac{1}{2} R_0^{-1} \tilde{\lambda}_{pj} (\bar{z}_{0j+} + a)\theta_j p^{-1/2} w(s_{0j}) + \frac{a(b-a)}{4} R_0^{-3} \tilde{\lambda}_{pj}^2 \theta_j^2 p^{-1} w^2(s_{0j}) + o_{\mathrm{P}}(p^{-1}),$$

where

$$R_0 = \sqrt{(b - \tilde{\lambda}_{pj} s_{0j}/2)^2 + 2a\tilde{\lambda}_{pj} s_{0j}}.$$

On the other hand,

$$\ln \frac{\bar{z}_{j+} + a}{\bar{z}_{0j+} + a} = \frac{\bar{z}_{j+} - \bar{z}_{0j+}}{\bar{z}_{0j+} + a} - \frac{(\bar{z}_{j+} - \bar{z}_{0j+})^2}{2(\bar{z}_{0j+} + a)^2} + o_{\mathrm{P}}(p^{-1})$$

and

$$\ln \frac{\bar{z}_{j+} + b}{\bar{z}_{0j+} + b} = \frac{\bar{z}_{j+} - \bar{z}_{0j+}}{\bar{z}_{0j+} + b} - \frac{(\bar{z}_{j+} - \bar{z}_{0j+})^2}{2(\bar{z}_{0j+} + b)^2} + o_{\mathrm{P}}(p^{-1})$$

Therefore,

$$\begin{aligned} \varphi(\bar{z}_{0j+}) - \varphi(\bar{z}_{j+}) &= \frac{1}{2} R_0^{-1} \tilde{\lambda}_{pj} \theta_j p^{-1/2} w(s_{0j}) \frac{\bar{z}_{0j+}^2 + 2a\bar{z}_{0j+} + ab}{\bar{z}_{0j+} + b} \\ &\quad + \frac{a(b-a)}{4} R_0^{-3} \tilde{\lambda}_{pj}^2 \theta_j^2 p^{-1} w^2(s_{0j}) \frac{\bar{z}_{0j+}^2 + 2a\bar{z}_{0j+} + ab}{(\bar{z}_{0j+} + a)(\bar{z}_{0j+} + b)} \\ &\quad + \frac{b-a}{8} R_0^{-2} \tilde{\lambda}_{pj}^2 \theta_j^2 p^{-1} w^2(s_{0j}) \frac{\bar{z}_{0j+}^2 - ab}{(\bar{z}_{0j+} + b)^2} + o_{\mathrm{P}}(p^{-1}) \end{aligned} \tag{104}$$

By Theorem DJO4, as $p, n \to_\gamma \infty$,

$$\tilde{\lambda}_{pj} \overset{a.s.}{\to} \frac{\gamma_2(s_{0j} + \gamma_1)(s_{0j} + 1)}{s_{0j}((\gamma_1 + \gamma_2(1 + s_{0j}))},$$

Further, by definition, $b \to \frac{1}{2}(1-\gamma_1)$ and $a \to \frac{1}{2}(1+\gamma_1/\gamma_2-\gamma_1)$. These convergences imply that

$$\begin{aligned} R_0 &\overset{a.s.}{\to} \frac{1}{2}\frac{\gamma_2 s_{0j}^2 + 2s_{0j}(\gamma_1+\gamma_2)+\gamma_1^2+\gamma_1+\gamma_2}{\gamma_1+\gamma_2(1+s_{0j})}, \text{ and} \\ \bar{z}_{0j+} &\overset{a.s.}{\to} \frac{1}{2}(s_{0j}+\gamma_1). \end{aligned}$$

Using these results, we can simplify the last two lines of (104) to obtain

$$\begin{aligned} \varphi(\bar{z}_{0j+})-\varphi(\bar{z}_{j+}) &= \frac{1}{2}R_0^{-1}\tilde{\lambda}_{pj}\theta_j p^{-1/2}w(s_{0j})\frac{\bar{z}_{0j+}^2+2a\bar{z}_{0j+}+ab}{\bar{z}_{0j+}+b} \\ &\quad -\frac{\gamma_1(s_{0j}+\gamma_1)^2\theta_j^2 p^{-1}w^2(s_{0j})}{4s_{0j}^2(\gamma_2 s_{0j}^2+2s_{0j}(\gamma_1+\gamma_2)+\gamma_1^2+\gamma_1+\gamma_2)} \\ &\quad +o_{\mathrm{P}}(p^{-1}). \end{aligned} \tag{105}$$

For the first two terms in the figure brackets in (103), we have

$$-\frac{1}{2}(s_j-s_{0j})-\frac{c_1}{2}\ln\frac{s_j}{s_{0j}} = -\frac{s_{0j}+c_1}{2s_{0j}}\theta_j p^{-1/2}w(s_{0j})+\frac{c_1}{4}\frac{\theta_j^2 p^{-1}w^2(s_{0j})}{s_{0j}^2}+o(p^{-1}). \tag{106}$$

Using (105) and (106) in (103) yields

$$\begin{aligned} \ln L_p^{\mathrm{REG}}(\theta,\Lambda) = \sum_{j=1}^k &\Big\{ -n_H\frac{s_{0j}+c_1}{2s_{0j}}\theta_j p^{-1/2}w(s_{0j}) \\ &+n_H\frac{1}{2}R_0^{-1}\tilde{\lambda}_{pj}\theta_j p^{-1/2}w(s_{0j})\frac{\bar{z}_{0j+}^2+2a\bar{z}_{0j+}+ab}{\bar{z}_{0j+}+b} \\ &-\frac{1}{2}\theta_j^2\tau^2(s_{0j})\Big\}+o_{\mathrm{P}}(1). \end{aligned} \tag{107}$$

Finally, expand the second line of (107) in the powers of $\lambda_{pj}-l_j$ up to and including linear terms. To derive such an expansion, note that

$$\begin{aligned} \tilde{\lambda}_{pj} &= \frac{c_2(s_{0j}+c_1)(s_{0j}+1)}{s_{0j}(c_1+c_2+s_{0j}c_2)}+\frac{c_1c_2(-s_{0j}+c_2+s_{0j}c_2)^2}{s_{0j}^2(c_1+c_2+s_{0j}c_2)^2}(\lambda_{pj}-l_j)+o(\lambda_{pj}-l_j) \\ \bar{z}_{0j+} &= \frac{s_{0j}+c_1}{2}+\frac{c_1(c_2+s_{0j}c_2-s_{0j})^2}{2s_{0j}(c_1+c_2+c_1^2+2s_{0j}(c_1+c_2)+s_{0j}^2c_2)}(\lambda_{pj}-l_j)+o(\lambda_{pj}-l_j), \end{aligned}$$

and

$$R_0 = 2\bar{z}_{0j+}+\frac{1}{2}(1-c_1)-\frac{s_{0j}}{2}\tilde{\lambda}_{pj}.$$

Using these equations, we obtain, after some algebra,

$$\begin{aligned} R_0^{-1}\tilde{\lambda}_{pj}\frac{\bar{z}_{0j+}^2+2a\bar{z}_{0j+}+ab}{\bar{z}_{0j+}+b} &= \frac{s_{0j}+c_1}{s_{0j}}+\frac{c_1(c_2+s_{0j}c_2-s_{0j})^2}{s_{0j}^2(c_1+c_2+c_1^2+2s_{0j}(c_1+c_2)+s_{0j}^2c_2)}(\lambda_{pj}-l_j) \\ &\quad +o(\lambda_{pj}-l_j) \\ &= \frac{s_{0j}+c_1}{s_{0j}}+\frac{2c_1}{w(s_{0j})}(\lambda_{pj}-l_j)+o_{\mathrm{P}}(p^{-1/2}). \end{aligned}$$

Using this equality in the second line of (107) and simplifying, we obtain

$$\ln L_p^{\mathrm{REG}}(\theta, \Lambda) = \sum_{j=1}^{k} \left\{ \theta_j \sqrt{p}(\lambda_{pj} - l_j) - \frac{1}{2}\theta_j^2 \tau^2(s_{0j}) \right\} + o_{\mathrm{P}}(1).$$

Similarly to the SigD case, the above $o_{\mathrm{P}}(1)$ term is uniform in $(\theta_1, ..., \theta_k)$ from any compact subset of $\mathbb{R}^k$ by construction.

References.

[1] Anderson, G. A. (1965). An asymptotic expansion for the distribution of the latent roots of the estimated covariance matrix. *Ann. Math. Statist.* **36** 1153–1173.

[2] Bai, Z. D. and Silverstein, J. W. (1998). No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large-Dimensional Sample Covariance Matrices. *Annals of Probability* **26** 316–345.

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

[4] Bai, Z. D. and Silverstein, J. W. (2004). CLT for Linear Spectral Statistics of Large-Dimensional Sample Covariance Matrices. *Annals of Probability* **32** 553–605.

[5] Bai, Z. D. and Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices.* Second edition. Springer.

[6] Burkholder, D. L. (1973). Distribution function inequalities for martingales. *Annals of Probability* **1** 19–42.

[7] Butler, R. W. and Wood, A. T. A. (2003). Laplace approximation for Bessel functions of matrix argument. *Journal of Computational and Applied Mathematics* **155** 359–382.

[8] Chang, T.C. (1970). On an Asymptotic Representation of the Distribution of the Characteristic Roots of $S_1 S_2^{-1}$. *The Annals of Mathematical Statistics* **41** (2), 440–445.

[9] Dharmawansa, P., Johnstone, I.M. and Onatski, A. (2016). Local Asymptotic Normality of the spectrum of high-dimensional spiked F-ratios. *manuscript*

[10] Glynn, W.J. (1980). Asymptotic representations of the densities of canonical correlations and latent roots in MANOVA when the population parameters have arbitrary multiplicity. *The Annals of Statistics* **8** 958–976.

[11] Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application*, New York: Academic Press.

[12] Herz, C. S. (1955). Bessel functions of matrix argument. *Annals of Mathematics* **61** 474–523.

[13] Hsu, L.C. (1948). A Theorem on the Asymptotic Behavior of a Multiple Integral. *Duke Mathematical Journal* **15** 623–632.

[14] James, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Annals of Mathematical Statistics* **35** 475–501.

[15] Kato, T. (1980). *Perturbation Theory for Linear Operators*, Springer-Verlag. Berlin, Heidelberg, New York.

[16] McLeish, D.L. (1974). Dependent Central Limit Theorems and Invariance Principles. *Annals of Probability* **2**, No. 4, 620–628.

[17] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* John Wiley & Son, Hoboken, New Jersey.

[18] Olver, F. W. J. (1997). *Asymptotics and Special Functions.* A K Peters, Natick, Massachusetts.

[19] Silverstein, J. W. and Bai, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis* **54** 175–192.

[20] Titchmarsh, E. C. (1939). *The Theory of Functions.* Second edition. Oxford University Press. London.

E-mail: prathapa@uom.lk

E-mail: imj@stanford.edu

E-mail: ao319@cam.ac.uk