The Annals of Probability
2004, Vol. 32, No. 1A, 553–605

CLT FOR LINEAR SPECTRAL STATISTICS OF LARGE-DIMENSIONAL SAMPLE COVARIANCE MATRICES

BY Z. D. BAI[1] AND JACK W. SILVERSTEIN[2]

North East Normal University and National University of Singapore and North Carolina State University

Let $B_n = (1/N)T_n^{1/2}X_nX_n^*T_n^{1/2}$ where $X_n = (X_{ij})$ is $n \times N$ with i.i.d. complex standardized entries having finite fourth moment, and $T_n^{1/2}$ is a Hermitian square root of the nonnegative definite Hermitian matrix T_n. The limiting behavior, as $n \to \infty$ with n/N approaching a positive constant, of functionals of the eigenvalues of B_n, where each is given equal weight, is studied. Due to the limiting behavior of the empirical spectral distribution of B_n, it is known that these linear spectral statistics converges a.s. to a nonrandom quantity. This paper shows their rate of convergence to be $1/n$ by proving, after proper scaling, that they form a tight sequence. Moreover, if $\mathsf{E}X_{11}^2 = 0$ and $\mathsf{E}|X_{11}|^4 = 2$, or if X_{11} and T_n are real and $\mathsf{E}X_{11}^4 = 3$, they are shown to have Gaussian limits.

1. Introduction. Due to the rapid development of modern technology, statisticians are confronted with the task of analyzing data with ever increasing dimension. For example, stock market analysis can now include a large number of companies. The study of DNA can now incorporate a sizable number of its base pairs. Computers can easily perform computations with high-dimensional data. Indeed, within several milli-seconds, a mainframe can complete the spectral decomposition of a 1000×1000 symmetric matrix, a feat unachievable only 20 years ago. In the past, so-called *dimension reduction* schemes played the main role in dealing with high-dimensional data, but a large portion of information contained in the original data would inevitably get lost. For example, in variable selection of multivariate linear regression models, one will lose all information contained in the unselected variables; in principal component analysis, all information contained in the components deemed "nonprincipal" would be gone. Now when dimension reduction is performed it is usually not due to computational restrictions.

However, even though the technology exists to compute much of what is needed, there is a fundamental problem with the analytical tools used by statisticians.

Received June 2002; revised February 2003.
[1]Supported in part by NSFC Grant 201471000 and NUS Grant R-155-000-040-112.
[2]Supported in part by NSF Grant DMS-97-03591.
AMS 2000 subject classifications. Primary 15A52, 60F05; secondary 62H99.
Key words and phrases. Linear spectral statistics, random matrix, empirical distribution function of eigenvalues, Stieltjes transform.

Their use relies on their asymptotic behavior as the number of samples increase. It is to be expected that larger dimension will require larger samples in order to maintain the same level of behavior. But the required increase is typically orders of magnitude larger than the dimension, sample sizes that are simply unattainable in most situations. With a necessary limitation on the number of samples, many frequently used statistics in multivariate analysis perform in a completely different manner than they do on data of low dimension with no restriction on sample size. Some methods behave very poorly [see Bai and Saranadasa (1996)], and some are even not applicable [see Dempster (1958)]. Consider the following example.

Let X_{ij} be i.i.d. standard normal variables. Write

$$S_N = \left(\frac{1}{N} \sum_{k=1}^{N} X_{ik} X_{jk} \right)_{i,j=1}^{n},$$

which can be considered as a sample covariance matrix, N samples of an n-dimensional mean zero random vector with population matrix I. An important statistic in multivariate analysis is

$$L_N = \ln(\det S_N) = \sum_{j=1}^{n} \ln(\lambda_{N,j}),$$

where $\lambda_{N,j}$, $j = 1, \ldots, n$, are the eigenvalues of S_N. When n is fixed, $\lambda_{N,j} \to 1$ almost surely as $N \to \infty$ and thus $L_N \xrightarrow{\text{a.s.}} 0$.

Further, by taking a Taylor expansion on $\ln(1 + x)$, one can show that

$$\sqrt{N/n}\, L_N \xrightarrow{D} N(0, 2),$$

for any fixed n. This suggests the possibility that L_N is asymptotically normal, provided that $n = O(N)$. However, this is not the case. Let us see what happens when $n/N \to c \in (0, 1)$ as $n \to \infty$. Using results on the limiting spectral distribution of $\{S_N\}$ [see Marčenko and Pastur (1967) and Bai (1999)], we have, with probability 1,

$$\begin{aligned} \frac{1}{n} L_N &\to \int_{a(c)}^{b(c)} \frac{\ln x}{2\pi x c} \sqrt{\big(b(c) - x\big)\big(x - a(c)\big)}\, dx \\ &= \frac{c-1}{c} \ln(1 - c) - 1 \equiv d(c) < 0, \end{aligned} \tag{1.1}$$

where $a(c) = (1 - \sqrt{c})^2$, $b(c) = (1 + \sqrt{c})^2$ (see Section 5 for a derivation of the integral). This shows that almost surely

$$\sqrt{N/n}\, L_N \sim d(c)\sqrt{Nn} \to -\infty.$$

Thus, any test which assumes asymptotic normality of $\sqrt{N/n}\, L_N$ will result in a serious error.

Besides demonstrating problems with relying on traditional methodology when sample size is restricted, the example introduces one of several results that can be used to handle data with large dimension n, proportional to N, the sample size. They are limit theorems, as n approaches infinity, on the eigenvalues of a class of random matrices of sample covariance type [Yin and Krishnaiah (1983), Yin (1986), Silverstein (1995) and Bai and Silverstein (1998, 1999)]. They take the form

$$B_n = \frac{1}{N} T_n^{1/2} X_n X_n^* T_n^{1/2},$$

where $X_n = (X_{ij}^n)$ is $n \times N$, $X_{ij}^n \in \mathbb{C}$ are i.i.d. with $\mathsf{E}|X_{11}^n - \mathsf{E}X_{11}^n|^2 = 1$, $T_n^{1/2}$ is $n \times n$ random Hermitian, with X_n and $T_n^{1/2}$ independent. When X_{11}^n is known to have mean zero and T_n is nonrandom, B_n can be viewed as a sample covariance matrix, which includes any Wishart matrix, formed from N samples of the random vector $T_n^{1/2} X_{\cdot 1}^n$ ($X_{\cdot 1}^n$ denoting the first column of X_n), which has population covariance matrix $T_n \equiv (T_n^{1/2})^2$. Besides sample covariance matrices, B_n, whose eigenvalues are the same as those of $(1/N) X_n X_n^* T_n$, models the spectral behavior of other matrices important to multivariate statistics, in particular multivariate F matrices, where X_{11}^n is $N(0, 1)$ and T_n is the inverse of another Wishart matrix.

The basic limit theorem on the eigenvalues of B_n concerns its empirical spectral distribution F^{B_n}, where for any matrix A with real eigenvalues, F^A denotes the empirical distribution function of the eigenvalues of A, that is, if A is $n \times n$ then

$$F^A(x) = \frac{1}{n}(\text{number of eigenvalues of } A \le x).$$

If:

1. for all n, i, j, X_{ij}^n are i.d.,
2. with probability 1, $F^{T_n} \xrightarrow{D} H$, a proper cumulative distribution function (c.d.f.) and
3. $n/N \to c > 0$ as $n \to \infty$,

then with probability 1 F^{B_n} converges in distribution to $F^{c,H}$, a nonrandom proper c.d.f.

The case when H distributes its mass at one positive number (called the Pastur–Marcĕnko law), as in the above example, is one of seven nontrivial cases where an explicit expression for $F^{c,H}$ is known (the multivariate F matrix case [Silverstein (1985)] and, as to be seen below, when H is discrete with at most three positive mass points with or without mass at zero). However, a good deal of information, including a way to compute $F^{c,H}$, can be extracted out of an equation satisfied by its *Stieltjes transform*, defined for any c.d.f. G to be

$$m_G(z) \equiv \int \frac{1}{\lambda - z} \, dG(\lambda), \qquad \Im z \ne 0.$$

We see that $m_G(\bar{z}) = \overline{m_G(z)}$. For each $z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C} : \Im z > 0\}$, the Stieltjes transform $m(z) \equiv m_{F^{c,H}}(z)$ is the unique solution to

$$m = \int \frac{1}{\lambda(1 - c - czm) - z} \, dH(\lambda)$$

in the set $\{m \in \mathbb{C} : -\frac{1-c}{z} + cm \in \mathbb{C}^+\}$. The equation takes on a simpler form when $F^{c,H}$ is replaced by

$$\underline{F}^{c,H} \equiv (1 - c)I_{[0,\infty)} + cF^{c,H}$$

(I_A denoting the indicator function on the set A), which is the limiting empirical distribution function of $\underline{B}_n \equiv (1/N)X_n^* T_n X_n$ (the spectra of which differs from that of B_n by $|n - N|$ zeros). Its Stieltjes transform

$$\underline{m}(z) \equiv m_{\underline{F}^{c,H}}(z) = -\frac{1 - c}{z} + cm(z)$$

has inverse

$$z = z(\underline{m}) = -\frac{1}{\underline{m}} + c \int \frac{t}{1 + t\underline{m}} \, dH(t). \tag{1.2}$$

Using (1.2) it is shown in Silverstein and Choi (1995) that, on $(0, \infty)$, $F^{c,H}$ has a continuous density, is analytic inside its support and is given by

$$\begin{aligned} f^{c,H}(x) &= c^{-1} \frac{d}{dx} \underline{F}^{c,H}(x) \\ &= (c\pi)^{-1} \Im \underline{m}(x) \equiv (c\pi)^{-1} \lim_{z \to x} \Im \underline{m}(z). \end{aligned} \tag{1.3}$$

Also, $F^{c,H}(0) = \max[1 - c^{-1}, H(0)]$. Moreover, considering (1.2) for $\underline{m}$ real, the range of values where it is increasing constitutes the complement of the support of $F^{c,H}$ on $(0, \infty)$ [Marčenko and Pastur (1967) and Silverstein and Choi (1995)]. From (1.2) and (1.3) $f^{c,H}(x)$ can be computed using Newton's method for each $x \in (0, \infty)$ inside its support [see Bai and Silverstein (1998) for an illustration of the density when $c = 0.1$ and H places mass 0.2, 0.4, and 0.4 at, resp., 1, 3 and 10].

Notice in (1.2) when H is discrete with at most three positive mass points the density has an explicit expression, since $\underline{m}(z)$ is the root of a polynomial of degree at most four.

Convergence in distribution of F^{B_n} of course reveals no information on the number of eigenvalues of B_n appearing on any interval $[a, b]$ outside the support of $F^{c,H}$, other than the number is almost surely $o(n)$. In Bai and Silverstein (1998) it is shown that, with probability 1, no eigenvalues of B_n appear in $[a, b]$ for all n large under the following additional assumptions:

1′. X_n is the first n rows and N columns of a doubly infinite array of i.i.d. random variables, with $\mathsf{E}X_{11} = 0$, $\mathsf{E}|X_{11}|^2 = 1$ and $\mathsf{E}|X_{11}|^4 < \infty$, and

2′. T_n is nonrandom, $\|T_n\|$, the spectral norm of T_n, is bounded in n, and
3′. $[a, b]$ lies in an open subset of $(0, \infty)$ which is outside the support of F^{c_n, H_n} for all n large, where $c_n \equiv n/N$ and $H_n \equiv F^{T_n}$.

The result extends what has been previously known on the extreme eigenvalues of $(1/N)X_n X_n^*$ $(T_n = I)$. Let $\lambda_{\max}^A$, $\lambda_{\min}^A$ denote, respectively, the largest and smallest eigenvalues of the Hermitian matrix A. Under condition 1′, Yin, Bai and Krishnaiah (1988) showed, as $n \to \infty$

$$\lambda_{\max}^{(1/N)X_n X_n^*} \xrightarrow{\text{a.s.}} (1 + \sqrt{c})^2,$$

while in Bai and Yin (1993) for $c \le 1$

$$\lambda_{\min}^{(1/N)X_n X_n^*} \xrightarrow{\text{a.s.}} (1 - \sqrt{c})^2.$$

If $[a, b]$ separates the support of $F^{c,H}$ in $(0, \infty)$ into two nonempty sets, then associated with it is another interval J which separates the eigenvalues of T_n for all n large. The mass F^{c_n, H_n} places, say, to the right of b equals the proportion of eigenvalues of T_n lying to the right of J. In Bai and Silverstein (1999) it is proved that, with probability 1, the number of eigenvalues of B_n and T_n lying on the same side of their respective intervals is the same for all n large.

The above two results are intuitively plausible when viewing B_n as an approximation of T_n, especially when c_n is small (it can be shown that $F^{c,H} \xrightarrow{D} H$ as $c \to 0$). However, regardless of the size of c, when separation in the support of $F^{c,H}$ on $(0, \infty)$ associated with a gap in the spectrum of T_n occurs, there will be exact splitting of the eigenvalues of B_n.

These results can be used in applications where location of eigenvalues of the population covariance matrix is needed, as in the detection problem in array signal processing [see Silverstein and Combettes (1992)]. Here, each entry of the sampled vector is a reading off a sensor, due to an unknown number q of sources emitting signals in a noise-filled environment $(q < n)$. The problem is to determine q. The smallest eigenvalue of the population covariance matrix is positive with multiplicity $n - q$ (the so-called "noise" eigenvalues). The traditional approach has been to sample enough times so that the sample covariance matrix is close to the population matrix, relying on fixed dimension, large sample asymptotic analysis. However, it may be impossible to sample enough times if q is sizable. The above results show that in order to determine the number of sources, simply sample enough times so that the eigenvalues of B_n split into two discernable groups. The number on the right will, with high probability, equal q.

The results also enable us to understand the true behavior of statistics such as L_N in the above example when n and N are large but on the same order of magnitude; L_N is not close to zero, rather $n^{-1}L_N$ is close to the quantity $d(c)$ in (1.1), or perhaps more appropriately $d(c_n)$.

However, in order to fully utilize $n^{-1}L_N$, typically in hypothesis testing, it is important to establish the limiting distribution of $L_N - nd(c_n)$. We come to the main purpose of this paper, to study the limiting distribution of normalized spectral functionals like $L_N - nd(c)$, and as a by-product, the rate of convergence of statistics such as $n^{-1}L_N$, functionals of the eigenvalues of B_n where each is given equal weight. We will call them *linear spectral statistics*, quantities of the form

$$\frac{1}{n}\sum_{j=1}^{n} f(\lambda_j) \quad (\lambda_1,\dots,\lambda_n \text{ denoting the eigenvalues of } B_n) = \int f(x)\,dF^{B_n}(x),$$

where f is a function on $[0,\infty)$.

We will show, under the assumption $\mathsf{E}|X_{11}|^4 < \infty$ and the analyticity of f, the rate $\int f(x)\,dF^{B_n}(x) - \int f(x)\,dF^{c_n,H_n}(x)$ approaches zero is essentially $1/n$. Define

$$G_n(x) = n\bigl[F^{B_n}(x) - F^{c_n,H_n}(x)\bigr].$$

The main result is stated in the following theorem.

THEOREM 1.1. *Assume*:

(a) *For each n $X_{ij} = X_{ij}^n$, $i \le n$, $j \le N$ are i.i.d., i.d. for all n, i, j, $\mathsf{E}X_{11} = 0$, $\mathsf{E}|X_{11}|^2 = 1$, $\mathsf{E}|X_{11}|^4 < \infty$, $n/N \to c$, and*

(b) *T_n is $n \times n$ nonrandom Hermitian nonnegative definite with spectral norm bounded in n, with $F^{T_n} \xrightarrow{D} H$, a proper c.d.f.*

Let $f_1,\dots,f_k$ be functions on $\mathbb{R}$ analytic on an open interval containing

$$\left[\liminf_n \lambda_{\min}^{T_n} I_{(0,1)}(c)(1-\sqrt{c})^2, \limsup_n \lambda_{\max}^{T_n}(1+\sqrt{c})^2\right] \tag{1.4}$$

Then:

(i) *the random vector*

$$\left(\int f_1(x)\,dG_n(x),\dots,\int f_k(x)\,dG_n(x)\right) \tag{1.5}$$

forms a tight sequence in n.

(ii) *If X_{11} and T_n are real and $\mathsf{E}(X_{11}^4) = 3$, then* (1.5) *converges weakly to a Gaussian vector $(X_{f_1},\dots,X_{f_k})$, with means*

$$\mathsf{E}X_f = -\frac{1}{2\pi i}\int f(z)\frac{c\int \underline{m}(z)^3 t^2(1+t\underline{m}(z))^{-3}\,dH(t)}{(1-c\int \underline{m}(z)^2 t^2 (1+t\underline{m}(z))^{-2}\,dH(t))^2}\,dz \tag{1.6}$$

and covariance function

$$\begin{aligned}&\operatorname{Cov}(X_f, X_g)\\ &\quad = -\frac{1}{2\pi^2}\iint \frac{f(z_1)g(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2}\frac{d}{dz_1}\underline{m}(z_1)\frac{d}{dz_2}\underline{m}(z_2)\,dz_1\,dz_2\end{aligned} \tag{1.7}$$

($f, g \in \{f_1, \dots, f_k\}$). *The contours in* (1.6) *and* (1.7) [*two in* (1.7), *which we may assume to be nonoverlapping*] *are closed and are taken in the positive direction in the complex plane*, *each enclosing the support of* $F^{c,H}$.

(iii) *If* X_{11} *is complex with* $\mathsf{E}(X_{11}^2) = 0$ *and* $\mathsf{E}(|X_{11}|^4) = 2$, *then* (ii) *also holds*, *except the means are zero and the covariance function is* $1/2$ *the function given in* (1.7).

This theorem can be viewed as an extension of results obtained in Jonsson (1982) where the entries of X_n are Gaussian and $T_n = I$ and is consistent with central limit theorem results on linear statistics of eigenvalues of other classes of random matrices [see, e.g., Johansson (1998), Sinai and Soshnikov (1998), Soshnikov (2000) and Diaconis and Evans (2001)]. As will be seen, the techniques and arguments used to prove the theorem, which rely heavily on properties of the Stieltjes transform of F^{B_n}, have nothing in common with any of the tools used in these other papers.

We begin the proof of Theorem 1.1 here with the replacement of the entries of X_n with truncated and centralized variables. For $m = 1, 2, \dots$ find n_m ($n_m > n_{m-1}$) satisfying

$$m^4 \int_{\{|X_{11}| \geq \sqrt{n}/m\}} |X_{11}|^4 < 2^{-m}$$

for all $n \geq n_m$. Define $\delta_n = 1/m$ for all $n \in [n_m, n_{m+1})$ ($= 1$ for $n < n_1$). Then, as $n \to \infty$, $\delta_n \to 0$ and

$$\delta_n^{-4} \int_{\{|X_{11}| \geq \delta_n \sqrt{n}\}} |X_{11}|^4 \to 0. \tag{1.8}$$

Let now for each n δ_n be the larger of δ_n constructed above and the δ_n created in the proof of Lemma 2.2 of Yin, Bai and Krishnaiah (1988) with $r = 1/2$ and satisfying $\delta_n n^{1/4} \to \infty$. Let $\widehat{B}_n = (1/N) T_n^{1/2} \widehat{X}_n \widehat{X}_n^* T_n^{1/2}$ with $\widehat{X}_n$ $n \times N$ having (i, j)th entry $X_{ij} I_{\{|X_{ij}| < \delta_n \sqrt{n}\}}$. We have then

$$\mathsf{P}(B_n \neq \widehat{B}_n) \leq nNP\big(|X_{11}| \geq \delta_n \sqrt{n}\big) \leq K \delta_n^{-4} \int_{\{|X_{11}| \geq \delta_n \sqrt{n}\}} |X_{11}|^4 = o(1).$$

Define $\widetilde{B}_n = (1/N) T_n^{1/2} \widetilde{X}_n \widetilde{X}_n^* T_n^{1/2}$ with $\widetilde{X}_n$ $n \times N$ having (i, j)th entry $(\widehat{X}_{ij} - \mathsf{E}\widehat{X}_{ij})/\sigma_n$, where $\sigma_n^2 = \mathsf{E}|\widehat{X}_{ij} - \mathsf{E}\widehat{X}_{ij}|^2$. From Yin, Bai and Krishnaiah (1988) we know that both $\limsup_n \lambda_{\max}^{\widehat{B}_n}$ and $\limsup_n \lambda_{\max}^{\widetilde{B}_n}$ are almost surely bounded by $\limsup_n \|T_n\| (1 + \sqrt{c})^2$. We use $\widehat{G}_n(x)$ and $\widetilde{G}_n(x)$ to denote the analogues of $G_n(x)$ with the matrix B_n replaced by $\widehat{B}_n$ and $\widetilde{B}_n$, respectively. Let λ_i^A denote the ith smallest eigenvalue of Hermitian A. Using the same approach and bounds that are used in the proof of Lemma 2.7 of Bai (1999), we have, for each $j = 1, 2, \ldots, k$,

$$\begin{aligned}
&\left| \int f_j(x)\, d\widehat{G}_n(x) - \int f_j(x)\, d\widetilde{G}_n(x) \right| \\
&\qquad \le K_j \sum_{k=1}^{n} |\lambda_k^{\widehat{B}_n} - \lambda_k^{\widetilde{B}_n}| \\
&\qquad \le 2K_j \big(N^{-1} \operatorname{tr} T_n^{1/2}(\widehat{X}_n - \widetilde{X}_n)(\widehat{X}_n - \widetilde{X}_n)^* T_n^{1/2}\big)^{1/2} \big(n\big(\lambda_{\max}^{\widehat{B}_n} + \lambda_{\max}^{\widetilde{B}_n}\big)\big)^{1/2},
\end{aligned}$$

where K_j is a bound on $|f_j'(z)|$. From (1.8) we have

$$|\sigma_n^2 - 1| \le 2 \int_{\{|X_{11}| \ge 2\delta_n \sqrt{n}\}} |X_{11}|^2 \le 2\delta_n^{-2} n^{-1} \int_{\{|X_{11}| \ge \delta_n \sqrt{n}\}} |X_{11}|^4 = o(\delta_n^2 n^{-1}).$$

Moreover,

$$|\mathsf{E}\widehat{X}_{11}| = \left| \int_{\{|X_{11}| \ge \delta_n \sqrt{n}\}} X_{11} \right| = o(\delta_n n^{-3/2}).$$

These give us

$$\begin{aligned}
&\big(N^{-1} \operatorname{tr} T_n^{1/2}(\widehat{X}_n - \widetilde{X}_n)(\widehat{X}_n - \widetilde{X}_n)^* T_n^{1/2}\big)^{1/2} \\
&\qquad \le \big(N^{-1}(1 - 1/\sigma_n)^2 \operatorname{tr} \widehat{B}_n\big)^{1/2} + \big(N^{-1} \|T_n\| \sigma^{-2} \operatorname{tr} \mathsf{E}\widehat{X}_n \mathsf{E}\widehat{X}_n^*\big)^{1/2} \\
&\qquad \le \left(\frac{(1 - \sigma_n^2)^2}{\sigma^2 (1 + \sigma^2)^2} \frac{n}{N} \lambda_{\max}^{\widehat{B}_n} \right)^{1/2} + (n\|T_n\|)^{1/2} \sigma^{-1} |\mathsf{E}\widehat{X}_{11}| \\
&\qquad = o(\delta_n n^{-1/2}) \big(\lambda_{\max}^{\widehat{B}_n}\big)^{1/2} + o(\delta_n n^{-1}).
\end{aligned}$$

From the above estimates, we obtain

$$\int f_j(x)\, dG_n(x) = \int f_j(x)\, d\widetilde{G}_n(x) + o_p(1)$$

[$o_p(1) \xrightarrow{\text{i.p.}} 0$ as $n \to \infty$.] Therefore, we only need to find the limiting distribution of $\{\int f_j(x)\, d\widetilde{G}_n(x),\ j = 1, \ldots, k\}$. Hence, in the sequel, we shall assume the underlying variables are truncated at $\delta_n \sqrt{n}$, centralized and renormalized. For simplicity, we shall suppress all sub- or superscripts on the variables and assume that $|X_{ij}| < \delta_n \sqrt{n}$, $\mathsf{E}X_{ij} = 0$, $\mathsf{E}|X_{ij}|^2 = 1$, $\mathsf{E}|X_{ij}|^4 < \infty$, and for assumption (ii) of Theorem 1.1 $\mathsf{E}|X_{11}|^4 = 3 + o(1)$, while for assumption (iii) $\mathsf{E}X_{11}^2 = o(1/n)$ and $\mathsf{E}|X_{11}|^4 = 2 + o(1)$.

Since the truncation steps are identical to those in Yin, Bai and Krishnaiah (1988) we have for any $\eta > (1 + \sqrt{c})^2$ the existence of $\{k_n\}$ for which

$$\frac{k_n}{\ln n} \to \infty \quad \text{and} \quad \mathsf{E}\|(1/N) X_n X_n^*\|^{k_n} \le \eta^{k_n}$$

for all n large. Therefore,

$$P(\|B_n\| \geq \eta) = o(n^{-\ell}), \tag{1.9a}$$

for any $\eta > \limsup \|T\|(1+\sqrt{c})^2$ and any positive ℓ. By modifying the proof in Bai and Yin (1993) on the smallest eigenvalue of $(1/N)X_n X_n^*$ it follows that when $\liminf_n \lambda_{\min}^T I_{(0,1)}(c)(1-\sqrt{c})^2 > 0$

$$P\big(\lambda_{\min}^{B_n} \leq \eta\big) = o(n^{-\ell}), \tag{1.9b}$$

whenever $0 < \eta < \liminf_n \lambda_{\min}^T I_{(0,1)}(c)(1-\sqrt{c})^2$. The modification is given in the Appendix.

After truncation and centralization, our proof of the main theorem relies on establishing limiting results on

$$M_n(z) = n\big[m_{F^{B_n}}(z) - m_{F^{c_n,H_n}}(z)\big] = N\big[m_{F^{\underline{B}_n}}(z) - m_{\underline{F}^{c_n,H_n}}(z)\big],$$

or more precisely, on $\widehat{M}_n(\cdot)$, a truncated version of $M_n(\cdot)$ when viewed as a random two-dimensional process defined on a contour $\mathcal{C}$ of the complex plane, described as follows. Let $v_0 > 0$ be arbitrary. Let x_r be any number greater than the right end point of interval (1.4). Let x_l be any negative number if the left end point of (1.4) is zero. Otherwise choose $x_l \in (0, \liminf_n \lambda_{\min}^{T_n} I_{(0,1)}(c)(1-\sqrt{c})^2)$. Let

$$\mathcal{C}_u = \{x + iv_0 : x \in [x_l, x_r]\}.$$

Then

$$\mathcal{C} \equiv \{x_l + iv : v \in [0, v_0]\} \cup \mathcal{C}_u \cup \{x_r + iv : v \in [0, v_0]\}.$$

We define now the subsets $\mathcal{C}_n$ of $\mathcal{C}$ on which $M_n(\cdot)$ agrees with $\widehat{M}_n(\cdot)$. Choose sequence $\{\varepsilon_n\}$ decreasing to zero satisfying for some $\alpha \in (0,1)$

$$\varepsilon_n \geq n^{-\alpha}. \tag{1.10}$$

Let

$$\mathcal{C}_l = \begin{cases} \{x_l + iv : v \in [n^{-1}\varepsilon_n, v_0]\}, & \text{if } x_l > 0, \\ \{x_l + iv : v \in [0, v_0]\}, & \text{if } x_l < 0, \end{cases}$$

and

$$\mathcal{C}_r = \{x_r + iv : v \in [n^{-1}\varepsilon_n, v_0]\}.$$

Then $\mathcal{C}_n = \mathcal{C}_l \cup \mathcal{C}_u \cup \mathcal{C}_r$. The process $\widehat{M}_n(\cdot)$ can now be defined. For $z = x + iv$ we have

$$\widehat{M}_n(z) = \begin{cases} M_n(z), & \text{for } z \in \mathcal{C}_n, \\ M_n(x_r + in^{-1}\varepsilon_n), & \text{for } x = x_r,\ v \in [0, n^{-1}\varepsilon_n], \\ & \text{and if } x_l > 0, \\ M_n(x_l + in^{-1}\varepsilon_n), & \text{for } x = x_l,\ v \in [0, n^{-1}\varepsilon_n]. \end{cases} \tag{1.11}$$

$\widehat{M}_n(\cdot)$ is viewed as a random element in the metric space $C(\mathcal{C}, \mathbb{R}^2)$ of continuous functions from $\mathcal{C}$ to $\mathbb{R}^2$. All of Chapter 2 of Billingsley (1968) applies to continuous functions from a set such as $\mathcal{C}$ (homeomorphic to $[0, 1]$) to finite-dimensional Euclidean space, with $|\cdot|$ interpreted as Euclidean distance.

Most of the paper will deal with proving the following lemma.

LEMMA 1.1. *Under conditions* (a) *and* (b) *of Theorem* 1.1 $\{\widehat{M}_n(\cdot)\}$ *forms a tight sequence on* $\mathcal{C}$. *Moreover, if assumptions in* (ii) *or* (iii) *of Theorem* 1.1 *on* X_{11} *hold, then* $\widehat{M}_n(\cdot)$ *converges weakly to a two-dimensional Gaussian process* $M(\cdot)$ *satisfying for* $z \in \mathcal{C}$ *under the assumptions in* (ii),

$$\mathsf{E}M(z) = \frac{c \int \underline{m}(z)^3 t^2 (1 + t\underline{m}(z))^{-3}\, dH(t)}{(1 - c\int \underline{m}(z)^2 t^2 (1 + t\underline{m}(z))^{-2}\, dH(t))^2} \tag{1.12}$$

and for $z_1, z_2 \in \mathcal{C} \cup \bar{\mathcal{C}}$, *with* $\bar{\mathcal{C}} \equiv \{\bar{z} : z \in \mathcal{C}\}$,

$$\begin{aligned} \operatorname{Cov}\big(M(z_1), M(z_2)\big) &\equiv \mathsf{E}\big[\big(M(z_1) - \mathsf{E}M(z_1)\big)\big(M(z_2) - \mathsf{E}M(z_2)\big)\big] \\ &= \frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} - \frac{1}{(z_1 - z_2)^2}, \end{aligned} \tag{1.13}$$

while under the assumptions in (iii) $\mathsf{E}M(z) = 0$, *and the "covariance" function analogous to* (1.13) *is* 1/2 *the right-hand side of* (1.13).

We show now how Theorem 1.1 follows from the above lemma. We use the identity

$$\int f(x)\, dG(x) = -\frac{1}{2\pi i} \int f(z) m_G(z)\, dz \tag{1.14}$$

valid for c.d.f. G and f analytic on an open set containing the support of G. The complex integral on the right-hand side is over any positively oriented contour enclosing the support of G and on which f is analytic. Choose v_0, x_r and x_l so that $f_1, \dots, f_k$ are all analytic on and inside the resulting $\mathcal{C} \cup \bar{\mathcal{C}}$.

Due to the a.s. convergence of the extreme eigenvalues of $(1/N)X_nX_n^*$ and the bounds

$$\lambda_{\max}^{AB} \le \lambda_{\max}^{A}\lambda_{\max}^{B} \qquad \lambda_{\min}^{AB} \ge \lambda_{\min}^{A}\lambda_{\min}^{B},$$

valid for $n \times n$ Hermitian nonnegative definite A and B, we have with probability 1

$$\liminf_{n\to\infty} \min\big(x_r - \lambda_{\max}^{B_n}, \lambda_{\min}^{B_n} - x_l\big) > 0.$$

It also follows that the support of F^{c_n, H_n} is contained in

$$\big[\lambda_{\min}^{T_n} I_{(0,1)}(c_n)\big(1 - \sqrt{c_n}\big)^2, \lambda_{\max}^{T_n}\big(1 + \sqrt{c_n}\big)^2\big].$$

Therefore for any $f \in \{f_1, \dots, f_k\}$, with probability 1

$$\int f(x)\, dG_n(x) = -\frac{1}{2\pi i} \int f(z) M_n(z)\, dz$$

for all n large, where the complex integral is over $\mathcal{C} \cup \bar{\mathcal{C}}$. Moreover, with probability 1, for all n large,

$$\begin{aligned}
&\left| \int f(z) \big(M_n(z) - \widehat{M}_n(z)\big)\, dz \right| \\
&\quad \le 4K\varepsilon_n \Big(\big| \max\big(\lambda_{\max}^{T_n}(1+\sqrt{c_n})^2, \lambda_{\max}^{B_n}\big) - x_r \big|^{-1} \\
&\qquad\qquad + \big| \min\big(\lambda_{\min}^{T_n} I_{(0,1)}(c_n)(1-\sqrt{c_n})^2, \lambda_{\min}^{B_n}\big) - x_l \big|^{-1} \Big),
\end{aligned}$$

which converges to zero as $n \to \infty$. Here K is a bound on f over $\mathcal{C}$.

Since

$$\widehat{M}_n(\cdot) \to \left(-\frac{1}{2\pi i} \int f_1(z) \widehat{M}_n(z)\, dz, \dots, -\frac{1}{2\pi i} \int f_k(z) \widehat{M}_n(z)\, dz \right)$$

is a continuous mapping of $C(\mathcal{C}, \mathbb{R}^2)$ into $\mathbb{R}^k$, it follows that the above vector and, subsequently, (1.5) form tight sequences. Letting $M(\cdot)$ denote the limit of any weakly converging subsequence of $\{\widehat{M}_n(\cdot)\}$ we have the weak limit of (1.5) equal in distribution to

$$\left(-\frac{1}{2\pi i} \int f_1(z) M(z)\, dz, \dots, -\frac{1}{2\pi i} \int f_k(z)\, M(z)\, dz \right).$$

The fact that this vector, under the assumptions in (ii) or (iii), is multivariate Gaussian follows from the fact that Riemann sums corresponding to these integrals are multivariate Gaussian and that weak limits of Gaussian vectors can only be Gaussian. The limiting expressions for the mean and covariance follow immediately.

Notice the assumptions in (ii) and (iii) require X_{11} to have the same first, second and fourth moments of either a real or complex Gaussian variable, the latter having real and imaginary parts i.i.d. $N(0, 1/2)$. We will use the terms "RG" and "CG" to refer to these conditions.

The reason why concrete results are at present only obtained for the assumptions in (ii) and (iii) is mainly due to the identity

$$\begin{aligned}
&\mathsf{E}(X_{\cdot 1}^* A X_{\cdot 1} - \operatorname{tr} A)(X_{\cdot 1}^* B X_{\cdot 1} - \operatorname{tr} B) \\
&\quad = (\mathsf{E}|X_{11}|^4 - |\mathsf{E}X_{11}^2|^2 - 2) \sum_{i=1}^n a_{ii} b_{ii} + |\mathsf{E}X_{11}^2|^2 \operatorname{tr} AB^T + \operatorname{tr} AB
\end{aligned} \tag{1.15}$$

valid for $n \times n$ $A = (a_{ij})$ and $B = (b_{ij})$, which is needed in several places in the proof of Lemma 1.1. The assumptions in (iii) leave only the last term on the right-hand side, whereas those in (ii) leave the last two, but in this case the matrix B will always be symmetric. This also accounts for the relation between the two covariance functions and the difficulty in obtaining explicit results more generally. As will be seen in the proof, whenever (1.15) is used, little is known about the limiting behavior of $\sum a_{ii} b_{ii}$.

Simple substitution reveals

$$\text{RHS of (1.7)} = -\frac{1}{2\pi^2} \iint \frac{f(z(m_1)) g(z(m_2))}{(m_1 - m_2)^2} \, d(m_1) \, d(m_2). \tag{1.16}$$

However, the contours depend on the z_1, z_2 contours and cannot be arbitrarily chosen. It is also true that

$$\begin{aligned} (1.7) &= \frac{1}{\pi^2} \iint f'(x) g'(y) \ln \left| \frac{\underline{m}(x) - \overline{\underline{m}}(y)}{\underline{m}(x) - \underline{m}(y)} \right| dx\, dy \\ &= \frac{1}{2\pi^2} \iint f'(x) g'(y) \ln \left(1 + 4 \frac{\underline{m}_i(x) \underline{m}_i(y)}{|\underline{m}(x) - \underline{m}(y)|^2} \right) dx\, dy \end{aligned} \tag{1.17}$$

and

$$\mathsf{E} X_f = \frac{1}{2\pi} \int f'(x) \arg \left(1 - c \int \frac{t^2 \underline{m}^2(x)}{(1 + t\underline{m}(x))^2} \, dH(t) \right) dx. \tag{1.18}$$

Here for $0 \neq x \in \mathbb{R}$

$$\underline{m}(x) = \lim_{z \to x} \underline{m}(z), \qquad z \in \mathbb{C}^+, \tag{1.19}$$

known to exist and to satisfy (1.2) [see Silverstein and Choi (1995)], and $\underline{m}_i(x) = \Im \, \underline{m}(x)$. The term

$$j(x) = \arg \left(1 - c \int \frac{t^2 \underline{m}^2(x)}{(1 + t\underline{m}(x))^2} \, dH(t) \right)$$

in (1.18) is well defined for almost every x and takes values in $(-\pi/2, \pi/2)$. Section 5 contains proofs of (1.17) and (1.18), along with showing

$$k(x, y) \equiv \ln \left(1 + 4 \frac{\underline{m}_i(x) \underline{m}_i(y)}{|\underline{m}(x) - \underline{m}(y)|^2} \right) \tag{1.20}$$

to be Lebesgue integrable on $\mathbb{R}^2$. It is interesting to note that the support of $k(x, y)$ matches the support of $f^{c,H}$ on $\mathbb{R} - \{0\}$: $k(x, y) = 0 \Leftrightarrow \min(f^{c,H}(x), f^{c,H}(y)) = 0$. We also have $f^{c,H}(x) = 0 \Rightarrow j(x) = 0$.

Section 5 also contains derivations of the relevant quantities associated with the example given at the beginning of this section. The linear spectral statistic $(1/n)L_N$ has a.s. limit $d(c)$ as stated in (1.1). The quantity $L_N - nd(n/N)$ converges weakly to a Gaussian random variable $X_{\ln}$ with

$$\mathsf{E} X_{\ln} = \tfrac{1}{2} \ln(1 - c) \tag{1.21}$$

and

$$\operatorname{Var} X_{\ln} = -2\ln(1-c). \tag{1.22}$$

Results on both $L_N - \mathsf{E}L_N$ and $n[\int x^r\, dF^{S_N}(x) - \mathsf{E}\int x^r\, dF^{S_N}(x)]$ for positive integer r are derived in Jonsson (1982). Included in Section 5 are derivations of the following expressions for means and covariances, in this case ($H = I_{[1,\infty)}$). We have

$$\mathsf{E}X_{x^r} = \tfrac{1}{4}\big((1-\sqrt{c})^{2r} + (1+\sqrt{c})^{2r}\big) - \tfrac{1}{2}\sum_{j=0}^{r}\binom{r}{j}^2 c^j \tag{1.23}$$

and

$$\begin{aligned}\operatorname{Cov}(X_{x^{r_1}}, X_{x^{r_2}}) = 2c^{r_1+r_2}\sum_{k_1=0}^{r_1-1}\sum_{k_2=0}^{r_2}\binom{r_1}{k_1}\binom{r_2}{k_2}\left(\frac{1-c}{c}\right)^{k_1+k_2}&\\ \times\sum_{\ell=1}^{r_1-k_1}\ell\binom{2r_1-1-(k_1+\ell)}{r_1-1}&\\ \times\binom{2r_2-1-k_2+\ell}{r_2-1}&.\end{aligned} \tag{1.24}$$

It is noteworthy to mention here a consequence of (1.17), namely that if the assumptions in (ii) or (iii) of Theorem 1.1 were to hold, then G_n, considered as a random element in $D[0,\infty)$ (the space of functions on $[0,\infty)$ that are right-continuous with left-hand side limits, together with the Skorohod metric) cannot form a tight sequence in $D[0,\infty)$. Indeed, under the assumptions of either one, if $G(x)$ were a weak limit of a subsequence, then, because of Theorem 1.1, it is straightforward to conclude that for any x_0 in the interior of the support of F and positive ε,

$$\int_{x_0}^{x_0+\varepsilon} G(x)\, dx$$

would be Gaussian, and therefore so would

$$G(x_0) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{x_0}^{x_0+\varepsilon} G(x)\, dx.$$

However, the variance would necessarily be

$$\lim_{\varepsilon\to 0}\frac{1}{2\pi^2}\frac{1}{\varepsilon^2}\int_{x_0}^{x_0+\varepsilon}\int_{x_0}^{x_0+\varepsilon} k(x,y)\, dx\, dy = \infty.$$

Still, under the assumptions in (ii) or (iii), a limit may exist for $\{G_n\}$ when G_n is viewed as a linear functional

$$f \longrightarrow \int f(x)\, dG_n(x),$$

that is, a limit expressed in terms of a measure in a space of generalized functions. The characterization of the limiting measure of course depends on the space, which in turn relies on the set of test functions, which for now is restricted to functions analytic on the support of F. Work in this area is currently being pursued.

We emphasize here the importance of studying $G_n(x)$ which essentially balances $F^{B_n}(x)$ with F^{c_n,H_n}, and not $F^{c,H}$ or $\mathsf{E}F^{B_n}(x)$. $F^{c,H}$ cannot be used simply because the convergence of $c_n \to c$ and that of $H_n \to H$ can be arbitrarily slow. It should be viewed as a mathematical convenience because the result is expressed as a limit theorem. From the point of view of statistical inference, the choice of F^{c_n,H_n} over $\mathsf{E}F^{B_n}(x)$ is made simply because much is known of the former, while little is analytically known about the latter.

The proof of Lemma 1.1 is divided into three sections. Sections 2 and 3 handle the limiting behavior of the centralized M_n, while Section 4 analyzes the nonrandom part. In each of the three sections the reader will be referred to work done in Bai and Silverstein (1998).

2. Convergence of finite-dimensional distributions. Write for $z \in \mathcal{C}_n$, $M_n(z) = M_n^1(z) + M_n^2(z)$ where

$$M_n^1(z) = n\big[m_{F^{B_n}}(z) - \mathsf{E}m_{F^{B_n}}(z)\big]$$

and

$$M_n^2(z) = n\big[m_{\mathsf{E}F^{B_n}}(z) - m_{F^{c_n,H_n}}(z)\big].$$

In this section we will show for any positive integer r, the sum

$$\sum_{i=1}^{r} \alpha_i M_n^1(z_i) \qquad (\Im z_i \neq 0)$$

whenever it is real, is tight, and, under the assumptions in (ii) or (iii) of Theorem 1.1, will converge in distribution to a Gaussian random variable. Formula (1.13) will also be derived. We begin with a list of results.

LEMMA 2.1 [Burkholder (1973)]. *Let $\{X_k\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$. Then, for $p > 1$,*

$$\mathsf{E}\Big|\sum X_k\Big|^p \le K_p \mathsf{E}\Big(\sum |X_k|^2\Big)^{p/2}.$$

(Note: The reference considers only real variables. Extending to complex variables is straightforward.)

LEMMA 2.2 [Lemma 2.7 in Bai and Silverstein (1998)]. *For $X = (X_1, \dots, X_n)^T$ i.i.d. standardized (complex) entries, C $n \times n$ matrix (complex) we have, for any $p \ge 2$,*

$$\mathsf{E}|X^*CX - \operatorname{tr} C|^p \le K_p\big((\mathsf{E}|X_1|^4 \operatorname{tr} CC^*)^{p/2} + \mathsf{E}|X_1|^{2p} \operatorname{tr}(CC^*)^{p/2}\big).$$

LEMMA 2.3. *Let $f_1, f_2, \dots$ be analytic in D, a connected open set of $\mathbb{C}$, satisfying $|f_n(z)| \le M$ for every n and z in D, and $f_n(z)$ converges, as $n \to \infty$ for each z in a subset of D having a limit point in D. Then there exists a function f, analytic in D for which $f_n(z) \to f(z)$ and $f_n'(z) \to f'(z)$ for all $z \in D$. Moreover, on any set bounded by a contour interior to D the convergence is uniform and $\{f_n'(z)\}$ is uniformly bounded by $2M/\varepsilon$, where ε is the distance between the contour and the boundary of D.*

PROOF. The conclusions on $\{f_n\}$ are from Vitali's convergence theorem [see Titchmarsh (1939), page 168]. Those on $\{f_n'\}$ follow from the dominated convergence theorem (d.c.t.) and the identity

$$f_n'(z) = \frac{1}{2\pi i}\int_C \frac{f_n(w)}{(w-z)^2}\,dw. \qquad \square$$

LEMMA 2.4 [Theorem 35.12 of Billingsley (1995)]. *Suppose for each n $Y_{n1}, Y_{n2}, \dots, Y_{nr_n}$ is a real martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_{nj}\}$ having second moments. If as $n \to \infty$,*

$$\text{(i)} \qquad \sum_{j=1}^{r_n} \mathsf{E}(Y_{nj}^2 | \mathcal{F}_{n,j-1}) \xrightarrow{i.p.} \sigma^2,$$

where σ^2 is a positive constant, and for each $\varepsilon > 0$,

$$\text{(ii)} \qquad \sum_{j=1}^{r_n} \mathsf{E}\big(Y_{nj}^2 I_{(|Y_{nj}| \ge \varepsilon)}\big) \to 0$$

then

$$\sum_{j=1}^{r_n} Y_{nj} \xrightarrow{D} N(0, \sigma^2).$$

Recalling the truncation and centralization steps, we get from Lemma 2.2

$$\begin{aligned}\mathsf{E}|X_{\cdot 1}^* C X_{\cdot 1} - \operatorname{tr} C|^p &\le K_p \|C\|^p \big[n^{p/2} + \delta_n^{(2p-4)} n^{(p-1)}\big] \\ &\le \underline{K}_p \|C\|^p \delta_n^{(2p-4)} n^{(p-1)}, \qquad p \ge 2.\end{aligned} \tag{2.1}$$

Let $v = \Im z$. For the following analysis we will assume $v > 0$. To facilitate notation, we will let $T = T_n$. Because of assumption (2′) we may assume $\|T\| \le 1$ for all n. Constants appearing in inequalities will be denoted by K and may take on different values from one expression to the next. Let $r_j = (1/\sqrt{N}) T^{1/2} X_{\cdot j}$, $D(z) = B_n - zI$, $D_j(z) = D(z) - r_j r_j^*$,

$$\varepsilon_j(z) = r_j^* D_j^{-1}(z) r_j - \frac{1}{N}\operatorname{tr} T D_j^{-1}(z),$$

$$\delta_j(z) = r_j^* D_j^{-2}(z) r_j - \frac{1}{N}\operatorname{tr} T D_j^{-2}(z) = \frac{d}{dz}\varepsilon_j(z)$$

and

$$\beta_j(z) = \frac{1}{1 + r_j^* D_j^{-1}(z) r_j},$$

$$\bar{\beta}_j(z) = \frac{1}{1 + N^{-1} \operatorname{tr} T_n D_j^{-1}(z)},$$

$$b_n(z) = \frac{1}{1 + N^{-1} \mathsf{E} \operatorname{tr} T_n D_1^{-1}(z)}.$$

All of the three latter quantities are bounded in absolute value by $|z|/v$ [see (3.4) of Bai and Silverstein (1998)]. We have

$$D^{-1}(z) - D_j^{-1}(z) = -D_j^{-1}(z) r_j r_j^* D_j^{-1}(z) \beta_j(z)$$

and from Lemma 2.10 of Bai and Silverstein (1998) for any $n \times n$ A

$$\bigl|\operatorname{tr}\bigl(D^{-1}(z) - D_j^{-1}(z)\bigr)A\bigr| \le \frac{\|A\|}{\Im z}. \tag{2.2}$$

For nonrandom $n \times n$ A_k, $k = 1, \ldots, p$ and B_l, $l = 1, \ldots, q$, we shall establish the following inequality:

$$\begin{aligned}
&\left|\mathsf{E}\left(\prod_{k=1}^{p} r_1^* A_k r_1 \prod_{l=1}^{q} (r_1^* B_l r_1 - N^{-1} \operatorname{tr} T B_l)\right)\right| \\
&\qquad \le K N^{-(1\wedge q)} \delta_n^{(2q-4)\vee 0} \prod_{k=1}^{p} \|A_k\| \prod_{l=1}^{q} \|B_l\|, \qquad p \ge 0,\ q \ge 0.
\end{aligned} \tag{2.3}$$

When $p = 0$, $q = 1$, the left-hand side is 0. When $p = 0$, $q > 1$, (2.3) is a consequence of (2.1) and Hölder's inequality. If $p \ge 1$, then by induction on p we have

$$\begin{aligned}
&\left|\mathsf{E}\left(\prod_{k=1}^{p} r_1^* A_k r_1 \prod_{l=1}^{q} (r_1^* B_l r_1 - N^{-1} \operatorname{tr} T B_l)\right)\right| \\
&\quad \le \left|\mathsf{E}\left(\prod_{k=1}^{p-1} r_1^* A_k r_1 (r_1^* A_p r_1 - N^{-1} \operatorname{tr} T A_p) \prod_{l=1}^{q} (r_1^* B_l r_1 - N^{-1} \operatorname{tr} T B_l)\right)\right| \\
&\qquad + n N^{-1} \|A_p\| \left|\mathsf{E}\left(\prod_{k=1}^{p-1} r_1^* A_k r_1 \prod_{l=1}^{q} (r_1^* B_l r_1 - N^{-1} \operatorname{tr} T B_l)\right)\right| \\
&\quad \le K N^{-1} \delta_n^{(2q-4)\vee 0} \prod_{k=1}^{p} \|A_k\| \prod_{l=1}^{q} \|B_l\|.
\end{aligned}$$

We have proved the case where $q > 0$. When $q = 0$, (2.3) is a trivial consequence of (2.1).

Let $\mathsf{E}_0(\cdot)$ denote expectation and $\mathsf{E}_j(\cdot)$ denote conditional expectation with respect to the σ-field generated by $r_1, \dots, r_j$.

We have

$$
\begin{aligned}
n\big[m_{F^{B_n}}(z) - \mathsf{E}m_{F^{B_n}}(z)\big]
&= \operatorname{tr}[D^{-1}(z) - \mathsf{E}D^{-1}(z)] \\
&= \sum_{j=1}^{N} \operatorname{tr}\mathsf{E}_j D^{-1}(z) - \operatorname{tr}\mathsf{E}_{j-1} D^{-1}(z) \\
&= \sum_{j=1}^{N} \operatorname{tr}\mathsf{E}_j[D^{-1}(z) - D_j^{-1}(z)] - \operatorname{tr}\mathsf{E}_{j-1}[D^{-1}(z) - D_j^{-1}(z)] \\
&= -\sum_{j=1}^{N} (\mathsf{E}_j - \mathsf{E}_{j-1})\beta_j(z) r_j^* D_j^{-2}(z) r_j.
\end{aligned}
$$

Write $\beta_j(z) = \bar\beta_j(z) - \beta_j(z)\bar\beta_j(z)\varepsilon_j(z) = \bar\beta_j(z) - \bar\beta_j^2(z)\varepsilon_j(z) + \bar\beta_j^2(z) \times \beta_j(z)\varepsilon_j^2(z)$. We have then

$$
\begin{aligned}
&(\mathsf{E}_j - \mathsf{E}_{j-1})\beta_j(z) r_j^* D_j^{-2}(z) r_j \\
&\quad = (\mathsf{E}_j - \mathsf{E}_{j-1})\bigg(\bar\beta_j(z)\delta_j(z) - \bar\beta_j^2(z)\varepsilon_j(z)\delta_j(z) \\
&\qquad\qquad - \bar\beta_j^2(z)\varepsilon_j(z)\frac{1}{N}\operatorname{tr} T D_j^{-2}(z) + \bar\beta_j^2(z)\beta_j(z)\varepsilon_j^2(z) r_j^* D_j^{-2}(z) r_j\bigg) \\
&\quad = \mathsf{E}_j\bigg(\bar\beta_j(z)\delta_j(z) - \bar\beta_j^2(z)\varepsilon_j(z)\frac{1}{N}\operatorname{tr} T D_j^{-2}(z)\bigg) \\
&\qquad - (\mathsf{E}_j - \mathsf{E}_{j-1})\bar\beta_j^2(z)\big(\varepsilon_j(z)\delta_j(z) - \beta_j(z) r_j D_j^{-2}(z) r_j \varepsilon_j^2(z)\big).
\end{aligned}
$$

Using (2.3), we have

$$
\begin{aligned}
\mathsf{E}\bigg|\sum_{j=1}^{N}(\mathsf{E}_j - \mathsf{E}_{j-1})\bar\beta_j^2(z)\varepsilon_j(z)\delta_j(z)\bigg|^2 &= \sum_{j=1}^{N}\mathsf{E}|(\mathsf{E}_j - \mathsf{E}_{j-1})\bar\beta_j^2(z)\varepsilon_j(z)\delta_j(z)|^2 \\
&\le 4\sum_{j=1}^{N}\mathsf{E}|\bar\beta_j^2(z)\varepsilon_j(z)\delta_j(z)|^2 = o(1).
\end{aligned}
$$

Therefore, $\sum_{j=1}^{N}(\mathsf{E}_j - \mathsf{E}_{j-1})\bar\beta_j^2(z)\varepsilon_j(z)\delta_j(z)$ converges to zero in probability.

By the same argument, we have

$$
\sum_{j=1}^{N}(\mathsf{E}_j - \mathsf{E}_{j-1})\beta_j(z) r_j D_j^{-2}(z) r_j \varepsilon_j^2(z) \xrightarrow{\text{i.p.}} 0.
$$

Therefore we need only consider the sum

$$\sum_{i=1}^{r}\alpha_i\sum_{j=1}^{N}Y_j(z_i)=\sum_{j=1}^{N}\sum_{i=1}^{r}\alpha_i Y_j(z_i),$$

where

$$Y_j(z)=-\mathsf{E}_j\left(\bar\beta_j(z)\delta_j(z)-\bar\beta_j^2(z)\varepsilon_j(z)\frac{1}{N}\operatorname{tr}TD_j^{-2}(z)\right)=-\mathsf{E}_j\frac{d}{dz}\bar\beta_j(z)\varepsilon_j(z).$$

Again, by using (2.3), we obtain

$$\mathsf{E}|Y_j(z)|^4\le K\left(\frac{|z|^4}{v^4}\mathsf{E}|\delta_j(z)|^4+\frac{|z|^8}{v^{16}}\left(\frac{n}{N}\right)^4\mathsf{E}|\varepsilon_j(z)|^4\right)=o(N^{-1}),$$

which implies for any $\varepsilon>0$

$$\sum_{j=1}^{N}\mathsf{E}\left(\left|\sum_{i=1}^{r}\alpha_i Y_j(z_i)\right|^2 I_{(|\sum_{i=1}^{r}\alpha_i Y_j(z_i)|\ge\varepsilon)}\right)\le\frac{1}{\varepsilon^2}\sum_{j=1}^{N}\mathsf{E}\left|\sum_{i=1}^{r}\alpha_i Y_j(z_i)\right|^4\to 0$$

as $n\to\infty$. Therefore condition (ii) of Lemma 2.4 is satisfied and it is enough to prove, under the assumptions in (ii) or (iii) of Theorem 1.1, for z_1, z_2 with nonzero imaginary parts

$$\sum_{j=1}^{N}\mathsf{E}_{j-1}[Y_j(z_1)Y_j(z_2)] \tag{2.4}$$

converges in probability to a constant (and to determine the constant).

We show here for future use the tightness of the sequence $\{\sum_{i=1}^{r}\alpha_i M_n^1(z_i)\}$. From (2.3) we easily get $\mathsf{E}|Y_j(z)|^2=O(N^{-1})$, so that

$$\begin{aligned}\mathsf{E}\left|\sum_{i=1}^{r}\alpha_i\sum_{j=1}^{N}Y_j(z_i)\right|^2&=\sum_{j=1}^{N}\mathsf{E}\left|\sum_{i=1}^{r}\alpha_i Y_j(z_i)\right|^2\\&\le r\sum_{j=1}^{N}\sum_{i=1}^{r}|\alpha_i|^2\mathsf{E}\big|Y_j(z_i)\big|^2\le K.\end{aligned}\tag{2.5}$$

Consider the sum

$$\sum_{j=1}^{N}\mathsf{E}_{j-1}\big[\mathsf{E}_j\big(\bar\beta_j(z_1)\varepsilon_j(z_1)\big)\mathsf{E}_j\big(\bar\beta_j(z_2)\varepsilon_j(z_2)\big)\big].\tag{2.6}$$

In the jth term (viewed as an expectation with respect to $r_{j+1},\dots,r_N$) we apply the d.c.t. to the difference quotient defined by $\bar\beta_j(z)\varepsilon_j(z)$ to get

$$\frac{\partial^2}{\partial z_2\,\partial z_1}(2.6)=(2.4).$$

Let v_0 be a lower bound on $|\Im z_i|$. For each j let $A_j^i = (1/N)T^{1/2}\mathsf{E}_j D_j^{-1}(z_i) \times T^{1/2}$, $i = 1, 2$. Then $\operatorname{tr} A_j^i A_j^{i\,*} \le n(v_0 N)^{-2}$. Using (2.1) we see, therefore, that (2.6) is bounded, with a bound depending only on $|z_i|$ and v_0.

We can then appeal to Lemma 2.3. Suppose (2.6) converges in probability to a nonrandom limit for each $z_k, z_l \in \{z_i\} \subset \underline{D} \equiv \{z : v_0 < |\Im z| < K\}$ ($K > v_0$ arbitrary), a sequence having two limit points, one on each side of the real axis. Then by a diagonalization argument, for any subsequence of the natural numbers, there is a further subsequence such that, with probability one, (2.6) converges for each pair z_k, z_l. Write (2.6) as $f_n(z_1, z_2)$. We concentrate on this subsequence and on one realization for which convergence holds. For each $z_l \in \{z_i\}$ we apply Lemma 2.3 on each of $\{z : v_0/2 < \Im z < K\}$ and $\{z : -K < \Im z < -v_0/2\}$ to get convergence of $f_n(z, z_l)$ to a function $f(z, z_l)$, analytic for $z \in \underline{D}$ satisfying $\frac{\partial}{\partial z} f_n(z, z_l) \to \frac{\partial}{\partial z} f(z, z_l)$. From Lemma 2.3 we see that $\frac{\partial}{\partial z} f_n(z, w)$ is bounded in w and n for all $w \in \underline{D}$. Applying again Lemma 2.3 on the remaining variable we see that $f_n(z, w) \to f(z, w)$, analytic for $w \in \underline{D}$ and $\frac{\partial^2}{\partial u\, \partial z} f_n(z, w) \to \frac{\partial^2}{\partial w\, \partial z} f_n(z, w)$. Since $f(z, w)$ does not depend on the realization nor on the subsequence, we have convergence in probability of (2.6) to f and (2.4) to the mixed partials of f. Therefore we need only show (2.6) converges in probability and to determine its limit.

From the derivation above (4.3) of Bai and Silverstein (1998) we get

$$\mathsf{E}|\bar{\beta}_j(z_i) - b_n(z_i)|^2 \le K \frac{|z_i|^4}{v_0^6} N^{-1}.$$

This implies

$$\begin{aligned} \mathsf{E}\Big|&\mathsf{E}_{j-1}\big[\mathsf{E}_j\big(\bar{\beta}_j(z_1)\varepsilon_j(z_1)\big)\mathsf{E}_j\big(\bar{\beta}_j(z_2)\varepsilon_j(z_2)\big)\big] \\ &- \mathsf{E}_{j-1}\big[\mathsf{E}_j\big(b_n(z_1)\varepsilon_j(z_1)\big)\mathsf{E}_j\big(b_n(z_2)\varepsilon_j(z_2)\big)\big]\Big| = O(N^{-3/2}) \end{aligned}$$

from which we get

$$\begin{aligned} &\sum_{j=1}^N \mathsf{E}_{j-1}\big[\mathsf{E}_j\big(\bar{\beta}_j(z_1)\varepsilon_j(z_1)\big)\mathsf{E}_j\big(\hat{\beta}_j(z_2)\varepsilon_j(z_2)\big)\big] \\ &\qquad - b_n(z_1)b_n(z_2)\sum_{j=1}^N \mathsf{E}_{j-1}\big[\mathsf{E}_j(\varepsilon_j(z_1))\mathsf{E}_j(\varepsilon_j(z_2))\big] \xrightarrow{\text{i.p.}} 0. \end{aligned}$$

Thus the goal is to show

$$b_n(z_1)b_n(z_2)\sum_{j=1}^N \mathsf{E}_{j-1}\big[\mathsf{E}_j(\varepsilon_j(z_1))\mathsf{E}_j(\varepsilon_j(z_2))\big] \tag{2.7}$$

converges in probability, and to determine its limit. The latter's second mixed partial derivative will yield the limit of (2.4).

We now assume the CG case, namely $\mathsf{E}X_{11}^2 = o(1/n)$ and $\mathsf{E}|X_{11}|^4 = 2 + o(1)$, so that, using (1.15), (2.7) becomes

$$b_n(z_1)b_n(z_2)\frac{1}{N^2}\sum_{j=1}^{N}\big(\operatorname{tr} T^{1/2}\mathsf{E}_j\big(D_j^{-1}(z_1)\big)T\mathsf{E}_j\big(D_j^{-1}(z_2)\big)T^{1/2} + o(1)A_n\big),$$

where

$$|A_n| \le K\Big(\operatorname{tr} T\mathsf{E}_j\big(D_j^{-1}(z_1)\big)T\mathsf{E}_j\big(\bar{D}_j^{-1}(z_1)\big) \times \operatorname{tr} T\mathsf{E}_j\big(D_j^{-1}(z_2)\big)T\mathsf{E}_j\big(\bar{D}_j^{-1}(z_2)\big)\Big)^{1/2} = O(N).$$

Thus we study

$$b_n(z_1)b_n(z_2)\frac{1}{N^2}\sum_{j=1}^{N}\operatorname{tr} T^{1/2}\mathsf{E}_j\big(D_j^{-1}(z_1)\big)T\mathsf{E}_j\big(D_j^{-1}(z_2)\big)T^{1/2}. \tag{2.8}$$

The RG case $[T_n,\ X_{11}$ real, $\mathsf{E}|X_{11}|^4 = 3 + o(1)]$ will be double that of the limit of (2.8).

Let $D_{ij}(z) = D(z) - r_ir_i^* - r_jr_j^*$,

$$\beta_{ij}(z) = \frac{1}{1 + r_i^*D_{ij}^{-1}(z)r_i} \quad \text{and} \quad b_1(z) = \frac{1}{1 + N^{-1}\mathsf{E}\operatorname{tr} TD_{12}^{-1}(z)}.$$

We write

$$D_j(z_1) + z_1I - \frac{N-1}{N}b_1(z_1)T = \sum_{i\ne j}^{N} r_ir_i^* - \frac{N-1}{N}b_1(z_1)T.$$

Multiplying by $(z_1I - \frac{N-1}{N}b_1(z_1)T)^{-1}$ on the left-hand side, $D_j^{-1}(z_1)$ on the right-hand side and using

$$r_i^*D_j^{-1}(z_1) = \beta_{ij}(z_1)r_i^*D_{ij}^{-1}(z_1)$$

we get

$$\begin{aligned} D_j^{-1}(z_1) &= -\Big(z_1I - \frac{N-1}{N}b_1(z_1)T\Big)^{-1} \\ &\quad + \sum_{i\ne j}\beta_{ij}(z_1)\Big(z_1I - \frac{N-1}{N}b_1(z_1)T\Big)^{-1} r_ir_i^*D_{ij}^{-1}(z_1) \\ &\quad - \frac{N-1}{N}b_1(z_1)\Big(z_1I - \frac{N-1}{N}b_1(z_1)T\Big)^{-1} TD_j^{-1}(z_1) \\ &= -\Big(z_1I - \frac{N-1}{N}b_1(z_1)T\Big)^{-1} + b_1(z_1)A(z_1) + B(z_1) + C(z_1), \end{aligned} \tag{2.9}$$

where

$$A(z_1) = \sum_{i\neq j}\left(z_1 I - \frac{N-1}{N} b_1(z_1) T\right)^{-1} (r_i r_i^* - N^{-1} T) D_{ij}^{-1}(z_1),$$

$$B(z_1) = \sum_{i\neq j}\bigl(\beta_{ij}(z_1) - b_1(z_1)\bigr)\left(z_1 I - \frac{N-1}{N} b_1(z_1) T\right)^{-1} r_i r_i^* D_{ij}^{-1}(z_1)$$

and

$$C(z_1) = N^{-1} b_1(z_1)\left(z_1 I - \frac{N-1}{N} b_1(z_1) T\right)^{-1} T \times \sum_{i\neq j}\bigl(D_{ij}^{-1}(z_1) - D_j^{-1}(z_1)\bigr).$$

It is easy to verify for any real t,

$$\left|1 - \frac{t}{z(1+N^{-1}\mathsf{E}\operatorname{tr} T D_{12}^{-1}(z))}\right|^{-1} \le \frac{|z(1+N^{-1}\mathsf{E}\operatorname{tr} T D_{12}^{-1}(z))|}{\Im z(1+N^{-1}\mathsf{E}\operatorname{tr} T D_{12}^{-1}(z))} \le \frac{|z|(1+n/(Nv_0))}{v_0}.$$

Thus

$$\left\|\left(z_1 I - \frac{N-1}{N} b_1(z_1) T\right)^{-1}\right\| \le \frac{1+n/(Nv_0)}{v_0}. \tag{2.10}$$

Let M be $n\times n$ and let $|||M|||$ denote a nonrandom bound on the spectral norm of M for all parameters governing M and under all realizations of M. From (4.3) of Bai and Silverstein (1998), (2.3) and (2.10) we get

$$\begin{aligned} &\mathsf{E}|\operatorname{tr} B(z_1) M| \\ &\quad\le N\mathsf{E}^{1/2}\bigl(|\beta_{12}(z_1) - b_1(z_1)|^2\bigr) \\ &\qquad\times \mathsf{E}^{1/2}\left(\left|r_i^* D_{ij}^{-1}(z_1) M\left(z_1 I - \frac{N-1}{N} b_1(z_1) T\right)^{-1} r_i\right|^2\right) \\ &\quad\le K\,|||M|||\,\frac{|z_1|^2(1+n/(Nv_0))}{v_0^5} N^{1/2}. \end{aligned} \tag{2.11}$$

From (2.2) we have

$$|\operatorname{tr} C(z_1) M| \le |||M|||\,\frac{|z_1|(1+n/(Nv_0))}{v_0^3}. \tag{2.12}$$

From (2.3) and (2.10) we get, for M nonrandom,

$$\begin{aligned}
&\mathsf{E}|\operatorname{tr} A(z_1)M| \\
&\quad \le K\mathsf{E}^{1/2} \operatorname{tr} T^{1/2} D_{ij}^{-1}(z_1) M \biggl(z_1 I - \frac{N-1}{N} b_1(z_1) T \biggr)^{-1} \\
&\qquad \times T \biggl(\bar{z}_1 I - \frac{N-1}{N} b_1(\bar{z}_1) T \biggr)^{-1} M^* D_{ij}^{-1}(\bar{z}_1) T^{1/2} \\
&\quad \le K \|M\| \frac{(1 + n/(N v_0))}{v_0^2} N^{1/2}.
\end{aligned} \tag{2.13}$$

We write [using the identity above (2.2)]

$$\operatorname{tr} \mathsf{E}_j (A(z_1)) T D_j^{-1}(z_2) T = A_1(z_1, z_2) + A_2(z_1, z_2) + A_3(z_1, z_2), \tag{2.14}$$

where

$$\begin{aligned}
A_1(z_1, z_2) &= -\operatorname{tr} \sum_{i<j} \biggl(z_1 I - \frac{N-1}{N} b_1(z_1) T \biggr)^{-1} r_i r_i^* \mathsf{E}_j \bigl(D_{ij}^{-1}(z_1) \bigr) \\
&\qquad \times T \beta_{ij}(z_2) D_{ij}^{-1}(z_2) r_i r_i^* D_{ij}^{-1}(z_2) T \\
&= -\sum_{i<j} \beta_{ij}(z_2) r_i^* \mathsf{E}_j \bigl(D_{ij}^{-1}(z_1) \bigr) T D_{ij}^{-1}(z_2) r_i r_i^* \\
&\qquad \times D_{ij}^{-1}(z_2) T \biggl(z_1 I - \frac{N-1}{N} b_1(z_1) T \biggr)^{-1} r_i, \\
A_2(z_1, z_2) &= -\operatorname{tr} \sum_{i<j} \biggl(z_1 I - \frac{N-1}{N} b_1(z_1) T \biggr)^{-1} N^{-1} T \\
&\qquad \times \mathsf{E}_j \bigl(D_{ij}^{-1}(z_1) \bigr) T \bigl(D_j^{-1}(z_2) - D_{ij}^{-1}(z_2) \bigr) T
\end{aligned}$$

and

$$\begin{aligned}
A_3(z_1, z_2) &= \operatorname{tr} \sum_{i<j} \biggl(z_1 I - \frac{N-1}{N} b_1(z_1) T \biggr)^{-1} (r_i r_i^* - N^{-1} T) \\
&\qquad \times \mathsf{E}_j \bigl(D_{ij}^{-1}(z_1) \bigr) T D_{ij}^{-1}(z_2) T.
\end{aligned}$$

We get from (2.2) and (2.10)

$$|A_2(z_1, z_2)| \le \frac{(1 + n/(N v_0))}{v_0^2} \tag{2.15}$$

and similarly to (2.11) we have

$$\mathsf{E}|A_3(z_1, z_2)| \le \frac{(1 + n/(Nv_0))}{v_0^3} N^{1/2}.$$

Using (2.1), (2.3) and (4.3) of Bai and Silverstein (1998) we have for $i < j$

$$\begin{aligned}
\mathsf{E}\bigg|&\beta_{ij}(z_2) r_i^* \mathsf{E}_j\big(D_{ij}^{-1}(z_1)\big) T D_{ij}^{-1}(z_2) r_i r_i^* \\
&\times D_{ij}^{-1}(z_2) T \left(z_1 I - \frac{N-1}{N} b_1(z_1) T\right)^{-1} r_i \\
&- b_1(z_2) N^{-2} \operatorname{tr}\big(\mathsf{E}_j\big(D_{ij}^{-1}(z_1)\big) T D_{ij}^{-1}(z_2) T\big) \\
&\quad \times \operatorname{tr}\left(D_{ij}^{-1}(z_2) T \left(z_1 I - \frac{N-1}{N} b_1(z_1) T\right)^{-1} T\right)\bigg| \\
&\quad \le K N^{-1/2}
\end{aligned}$$

(K now depending as well on the z_i and on n/N). Using (2.2) we have

$$\begin{aligned}
&\bigg|\operatorname{tr}\big(\mathsf{E}_j\big(D_{ij}^{-1}(z_1)\big) T D_{ij}^{-1}(z_2) T\big) \operatorname{tr}\left(D_{ij}^{-1}(z_2) T \left(z_1 I - \frac{N-1}{N} b_1(z_1) T\right)^{-1} T\right) \\
&\quad - \operatorname{tr}\big(\mathsf{E}_j\big(D_j^{-1}(z_1)\big) T D_j^{-1}(z_2) T\big) \operatorname{tr}\left(D_j^{-1}(z_2) T \left(z_1 I - \frac{N-1}{N} b_1(z_1) T\right)^{-1} T\right)\bigg| \\
&\qquad \le K N.
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathsf{E}\bigg|&A_1(z_1, z_2) + \frac{j-1}{N^2} b_1(z_2) \operatorname{tr}\big(\mathsf{E}_j\big(D_j^{-1}(z_1)\big) T D_j^{-1}(z_2) T\big) \\
&\qquad \times \operatorname{tr}\left(D_j^{-1}(z_2) T \left(z_1 I - \frac{N-1}{N} b_1(z_1) T\right)^{-1} T\right)\bigg| \qquad (2.16) \\
&\le K N^{1/2}.
\end{aligned}$$

Therefore, from (2.9)–(2.16) we can write

$$\begin{aligned}
&\operatorname{tr}\big(\mathsf{E}_j\big(D_j^{-1}(z_1)\big) T D_j^{-1}(z_2) T\big) \\
&\quad \times \left[1 + \frac{j-1}{N^2} b_1(z_1) b_1(z_2) \operatorname{tr}\left(D_j^{-1}(z_2) T \left(z_1 I - \frac{N-1}{N} b_1(z_1) T\right)^{-1} T\right)\right] \\
&\quad = -\operatorname{tr}\left(\left(z_1 I - \frac{N-1}{N} b_1(z_1) T\right)^{-1} T D_j^{-1}(z_2) T\right) + A_4(z_1, z_2),
\end{aligned}$$

where

$$\mathsf{E}|A_4(z_1, z_2)| \le K N^{1/2}.$$

Using the expression for $D_j^{-1}(z_2)$ in (2.9) and (2.11)–(2.13) we find that

$$\begin{aligned}
&\operatorname{tr}\big(\mathsf{E}_j(D_j^{-1}(z_1))TD_j^{-1}(z_2)T\big)\\
&\quad\times\bigg[1-\frac{j-1}{N^2}b_1(z_1)b_1(z_2)\\
&\qquad\qquad\times\operatorname{tr}\bigg(\bigg(z_2I-\frac{N-1}{N}b_1(z_2)T\bigg)^{-1}T\bigg(z_1I-\frac{N-1}{N}b_1(z_1)T\bigg)^{-1}T\bigg)\bigg]\\
&\quad=\operatorname{tr}\bigg(\bigg(z_2I-\frac{N-1}{N}b_1(z_2)T\bigg)^{-1}T\bigg(z_1I-\frac{N-1}{N}b_1(z_1)T\bigg)^{-1}T\bigg)\\
&\qquad+A_5(z_1,z_2),
\end{aligned}$$

where

$$\mathsf{E}|A_5(z_1,z_2)|\le KN^{1/2}.$$

From (2.2) we have

$$|b_1(z)-b_n(z)|\le KN^{-1}.$$

From (4.3) of Bai and Silverstein (1998) we have

$$|b_n(z)-\mathsf{E}\beta_1(z)|\le KN^{-1/2}.$$

From the formula

$$\underline{m}_n=-\frac{1}{zN}\sum_{j=1}^{N}\beta_j(z)$$

[(2.2) of Silverstein (1995)] we get $\mathsf{E}\beta_1(z)=-z\mathsf{E}\underline{m}_n(z)$. Section 5 of Bai and Silverstein (1998) proves that

$$|\mathsf{E}\underline{m}_n(z)-\underline{m}_n^0(z)|\le KN^{-1}.$$

Therefore we have

$$|b_1(z)+z\underline{m}_n^0(z)|\le KN^{-1/2}, \tag{2.17}$$

so that we can write

$$\begin{aligned}
&\operatorname{tr}\big(\mathsf{E}_j(D_j^{-1}(z_1))TD_j^{-1}(z_2)T\big)\\
&\quad\times\bigg[1-\frac{j-1}{N^2}\underline{m}_n^0(z_1)\underline{m}_n^0(z_2)\\
&\qquad\qquad\times\operatorname{tr}\big((I+\underline{m}_n^0(z_2)T)^{-1}T(I+\underline{m}_n^0(z_1)T)^{-1}T\big)\bigg]\\
&\quad=\frac{1}{z_1z_2}\operatorname{tr}\big((I+\underline{m}_n^0(z_2)T)^{-1}T(I+\underline{m}_n^0(z_1)T)^{-1}T\big)+A_6(z_1,z_2),
\end{aligned}\tag{2.18}$$

where

$$\mathsf{E}|A_6(z_1, z_2)| \le K N^{1/2}.$$

Rewrite (2.18) as

$$\begin{aligned}
&\operatorname{tr}\big(\mathsf{E}_j\big(D_j^{-1}(z_1)\big)T D_j^{-1}(z_2)T\big)\\
&\quad\times\left[1 - \frac{j-1}{N} c_n \underline{m}_n^0(z_1)\underline{m}_n^0(z_2)\int \frac{t^2\,dH_n(t)}{(1+t\underline{m}_n^0(z_1))(1+t\underline{m}_n^0(z_2))}\right]\\
&\quad= \frac{Nc_n}{z_1 z_2}\int \frac{t^2\,dH_n(t)}{(1+t\underline{m}_n^0(z_1))(1+t\underline{m}_n^0(z_2))} + A_6(z_1, z_2).
\end{aligned}$$

Using (3.9) and (3.16) in Bai and Silverstein (1998) we find

$$\begin{aligned}
(2.19)\quad &\left|c_n \underline{m}_n^0(z_1)\underline{m}_n^0(z_2)\int \frac{t^2\,dH_n(t)}{(1+t\underline{m}_n^0(z_1))(1+t\underline{m}_n^0(z_2))}\right|\\
&= \left|c_n\left[\int \frac{t^2\,dH_n(t)}{(1+t\underline{m}_n^0(z_1))(1+t\underline{m}_n^0(z_2))}\right]\right.\\
&\qquad\left.\times\left[\left(-z_1 + c_n\int \frac{t\,dH_n(t)}{1+t\underline{m}_n^0(z_1)}\right)\left(-z_2 + c_n\int \frac{t\,dH_n(t)}{1+t\underline{m}_n^0(z_2)}\right)\right]^{-1}\right|\\
&\le \left(c_n\int \frac{t^2\,dH_n(t)}{|1+t\underline{m}_n^0(z_1)|^2}\left|-z_1 + c_n\int \frac{t\,dH_n(t)}{1+t\underline{m}_n^0(z_1)}\right|^{-2}\right)^{1/2}\\
&\qquad\times\left(c_n\int \frac{t^2\,dH_n(t)}{|1+t\underline{m}_n^0(z_2)|^2}\left|-z_2 + c_n\int \frac{t\,dH_n(t)}{1+t\underline{m}_n^0(z_2)}\right|^{-2}\right)^{1/2}\\
&= \left(\left(\Im\,\underline{m}_n^0(z_1)c_n\int \frac{t^2\,dH_n(t)}{|1+t\underline{m}_n^0(z_1)|^2}\right)\right.\\
&\qquad\left.\times\left(\Im\,z_1 + \Im\,\underline{m}_n^0(z_1)c_n\int \frac{t^2\,dH_n(t)}{|1+t\underline{m}_n^0(z_1)|^2}\right)^{-1}\right)^{1/2}\\
&\qquad\times\left(\left(\Im\,\underline{m}_n^0(z_2)c_n\int \frac{t^2\,dH_n(t)}{|1+t\underline{m}_n^0(z_2)|^2}\right)\right.\\
&\qquad\quad\left.\times\left(\Im\,z_2 + \Im\,\underline{m}_n^0(z_2)c_n\int \frac{t^2\,dH_n(t)}{|1+t\underline{m}_n^0(z_2)|^2}\right)^{-1}\right)^{1/2} < 1
\end{aligned}$$

since

$$\Im\,z\left(\Im\,\underline{m}_n^0(z)c_n\int \frac{t^2\,dH_n(t)}{|1+t\underline{m}_n^0(z)|^2}\right)^{-1}$$

is bounded away from 0. Therefore using (2.17) and letting $a_n(z_1, z_2)$ denote the expression inside the absolute value sign in (2.19) we find that (2.8) can be written as

$$a_n(z_1,z_2)\frac{1}{N}\sum_{j=1}^{N}\frac{1}{1-((j-1)/N)a_n(z_1,z_2)}+A_7(z_1,z_2),$$

where

$$\mathsf{E}|A_7(z_1,z_2)|\le KN^{-1/2}.$$

We see then that

$$(2.8)\ \xrightarrow{\text{i.p.}} a(z_1,z_2)\int_0^1\frac{1}{1-ta(z_1,z_2)}\,dt=\int_0^{a(z_1,z_2)}\frac{1}{1-z}\,dz$$

where

$$\begin{aligned}a(z_1,z_2)&=c\underline{m}(z_1)\underline{m}(z_2)\int\frac{t^2\,dH(t)}{(1+t\underline{m}(z_1))(1+t\underline{m}(z_2))}\\&=\frac{\underline{m}(z_1)\underline{m}(z_2)}{\underline{m}(z_2)-\underline{m}(z_1)}\left(c\int\frac{t\,dH(t)}{1+t\underline{m}(z_1)}-c\int\frac{t\,dH(t)}{1+t\underline{m}(z_2)}\right)\\&=1+\frac{\underline{m}(z_1)\underline{m}(z_2)(z_1-z_2)}{\underline{m}(z_2)-\underline{m}(z_1)}.\end{aligned}$$

Thus the i.p. limit of (2.4) under the CG case is

$$\begin{aligned}&\frac{\partial^2}{\partial z_2\,\partial z_1}\int_0^{a(z_1,z_2)}\frac{1}{1-z}\,dz\\&\quad=\frac{\partial}{\partial z_2}\left(\frac{\partial a(z_1,z_2)/\partial z_1}{1-a(z_1,z_2)}\right)\\&\quad=\frac{\partial}{\partial z_2}\Bigg[\frac{(\underline{m}(z_2)-\underline{m}(z_1))(\underline{m}'(z_1)\underline{m}(z_2)(z_1-z_2)+\underline{m}(z_1)\underline{m}(z_2))}{(\underline{m}(z_2)-\underline{m}(z_1))^2}\\&\qquad\qquad\qquad+\frac{\underline{m}(z_1)\underline{m}(z_2)(z_1-z_2)\underline{m}'(z_1)}{(\underline{m}(z_2)-\underline{m}(z_1))^2}\Bigg]\\&\qquad\times\frac{\underline{m}(z_2)-\underline{m}(z_1)}{\underline{m}(z_1)\underline{m}(z_2)(z_2-z_1)}\\&\quad=-\frac{\partial}{\partial z_2}\left(\frac{\underline{m}'(z_1)}{\underline{m}(z_1)}+\frac{1}{z_1-z_2}+\frac{\underline{m}'(z_1)}{\underline{m}(z_2)-\underline{m}(z_1)}\right)\\&\quad=\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_2)-\underline{m}(z_1))^2}-\frac{1}{(z_1-z_2)^2}\end{aligned}$$

which is (1.13).

half of

3. Tightness of $M_n^1(z)$. We proceed to prove tightness of the sequence of random functions $\widehat{M}_n^1(z)$ for $z \in \mathcal{C}$ defined by (1.11). We will use Theorem 12.3 [Billingsley (1968), page 96]. It is easy to verify from the proof of the Arzela–Ascoli theorem [Billingsley (1968), page 221] that condition (i) of Theorem 12.3 can be replaced with the assumption of tightness at any point in [0, 1]. From (2.5) we see that this condition is satisfied. We will verify condition (ii) of Theorem 12.3 by proving the moment condition (12.51) of Billingsley (1968). We will show

$$\sup_{n; z_1, z_2 \in \mathcal{C}_n} \frac{\mathsf{E}|M_n^1(z_1) - M_n^1(z_2)|^2}{|z_1 - z_2|^2}$$

is finite.

We claim that moments of $\|D^{-1}(z)\|$, $\|D_j^{-1}(z)\|$ and $\|D_{ij}^{-1}(z)\|$ are bounded in n and $z \in \mathcal{C}_n$. This is clearly true for $z \in \mathcal{C}_u$ and for $z \in \mathcal{C}_l$ if $x_l < 0$. For $z \in \mathcal{C}_r$ or, if $x_l > 0$, $z \in \mathcal{C}_l$, we use (1.9) and (1.10) on, for example $B_{(1)} = B_n - r_1 r_1^*$, to get

$$\begin{aligned} \mathsf{E}\|D_j^{-1}(z)\|^p &\le K_1 + v^{-p} \mathsf{P}\big(\|B_{(1)}\| \ge \eta_r \text{ or } \lambda_{\min}^{B_{(1)}} \le \eta_l\big) \\ &\le K_1 + K_2 n^p \varepsilon^{-p} n^{-\ell} \le K \end{aligned}$$

for suitably large ℓ. Here, η_r is any fixed number between $\limsup_n \|T\|(1 + \sqrt{c})^2$ and x_r, and, if $x_l > 0$, η_l is any fixed number between x_l and $\liminf_n \lambda_{\min}^T (1 - \sqrt{c})^2$ (take $\eta_l < 0$ if $x_l < 0$). Therefore for any positive p,

$$\max\big(\mathsf{E}\|D^{-1}(z)\|^p, \mathsf{E}\|D_j^{-1}(z)\|^p, \mathsf{E}\|D_{ij}^{-1}(z)\|^p\big) \le K_p. \tag{3.1}$$

We can use the above argument to extend (2.3). Using (1.8) and (2.3) we get

$$\begin{aligned} &\left|\mathsf{E}\left(a(v) \prod_{l=1}^{q} \big(r_1^* B_l(v) r_1 - N^{-1} \operatorname{tr} T B_l(v)\big)\right)\right| \\ &\qquad \le K N^{-(1 \wedge q)} \delta_n^{(2q-4) \vee 0}, \qquad q \ge 0, \end{aligned} \tag{3.2}$$

where now the matrices $B_l(v)$ are independent of r_1 and

$$\max\big(|a(v)|, \|B_l(v)\|\big) \le K\big(1 + n^s I\big(\|B_n\| \ge \eta_r \text{ or } \lambda_{\min}^{\widetilde{B}} \le \eta_l\big)\big)$$

for some positive s, with $\widetilde{B}$ being B_n or B_n with one or two of the r_j's removed.

We would like to inform the reader that in applications of (3.2), $a(v)$ is a product of factors of the form $\beta_1(z)$ or $r_1^* A(z) r_1$ and A is a product of one or several $D_1^{-1}(z) D_1^{-1}(z_j)$, $j = 1, 2$ or similarly defined D^{-1} matrices. The matrices B_l also have this form. For example, we have $|r_1^* D_1^{-1}(z_1) D_1^{-1}(z_2) r_1| \le |r_1|^2 \|D_1^{-1}(z_1) D_1^{-1}(z_2)\| \le K^2 \eta_r + |z| n^{3+2\alpha} I(\|B_n\| \ge \eta_r \text{ or } \lambda_{\min}^{B_{(1)}} \le \eta_l)$, where K can be taken to be $\max((x_r - \eta_r)^{-1}, (\eta_l - x_l)^{-1}, v_0^{-1})$, and where we have used (1.10) and the fact that $|r_1|^2 \le \eta_r$ if $\|B_n\| < \eta_r$ and $|r_1|^2 \le n$ otherwise. We have $\|B_l\|$ obviously satisfying this condition. We also have $\beta_1(z)$ satisfying this condition since from (3.3) (see below) $|\beta_1(z)| = |1 - r_1^* D^{-1} r_1| \le 1 + K\eta_r + |z| n^{2+\alpha} I(\|B_n\| \ge \eta_r \text{ or } \lambda_{\min}^{B_n} \le \eta_l)$. In the sequel, we shall freely use (3.2) without verifying these conditions, even similarly defined β_j functions and A, B matrices.

We have

$$\begin{aligned} D^{-1}(z) - D_j^{-1}(z) &= -\frac{D_j^{-1}(z) r_j r_j^* D_j^{-1}(z)}{1 + r_j^* D_j^{-1}(z) r_j} \\ &= -\beta_j(z) D_j^{-1}(z) r_j r_j^* D_j^{-1}(z). \end{aligned} \tag{3.3}$$

Let

$$\gamma_j(z) = r_j^* D_j^{-1}(z) r_j - N^{-1} \mathsf{E}\big(\operatorname{tr}\big(D_j^{-1}(z) T\big)\big).$$

We first derive bounds on the moments of $\gamma_j(z)$ and $\varepsilon_j(z)$. Using (3.2) we have

$$\mathsf{E}|\varepsilon_j(z)|^p \le K_p N^{-1} \delta_n^{2p-4} \qquad p \text{ even.} \tag{3.4}$$

It should be noted that constants obtained do not depend on $z \in \mathcal{C}_n$.

Using Lemma 2.1, (3.2), and Hölder's inequality, we have, for all even p,

$$\begin{aligned} \mathsf{E}|\gamma_j(z) - \varepsilon_j(z)|^p &= \mathsf{E}|\gamma_1(z) - \varepsilon_1(z)|^p \\ &= \mathsf{E}\left| \frac{1}{N} \sum_{j=2}^N \mathsf{E}_j \operatorname{tr} T D_1(z)^{-1} - \mathsf{E}_{j-1} \operatorname{tr} T D_1^{-1}(z) \right|^p \\ &= \mathsf{E}\left| \frac{1}{N} \sum_{j=2}^N \mathsf{E}_j \operatorname{tr} T\big(D_1^{-1}(z) - D_{1j}^{-1}(z)\big) \right. \\ &\qquad \left. - \mathsf{E}_{j-1} \operatorname{tr} T\big(D_1^{-1}(z) - D_{1j}^{-1}(z)\big) \right|^p \\ &= \frac{1}{N^p} \mathsf{E}\left| \sum_{j=2}^N (\mathsf{E}_j - \mathsf{E}_{j-1}) \beta_{1j}(z) r_j^* D_{1j}^{-1}(z) T D_{1j}^{-1}(z) r_j \right|^p \\ &\le \frac{K_p}{N^p} \mathsf{E}\left(\sum_{j=2}^N \big|(\mathsf{E}_j - \mathsf{E}_{j-1}) \beta_{1j}(z) r_j^* D_{1j}^{-1}(z) T D_{1j}^{-1}(z) r_j\big|^2 \right)^{p/2} \\ &\le \frac{K_p}{N^{1+p/2}} \sum_{j=2}^N \mathsf{E}\big|(\mathsf{E}_j - \mathsf{E}_{j-1}) \beta_{1j}(z) r_j^* D_{1j}^{-1}(z) T D_{1j}^{-1}(z) r_j\big|^p \\ &\le \frac{K_p}{N^{p/2}} \mathsf{E}\big|\beta_{12}(z) r_2^* D_{12}^{-1}(z) T D_{12}^{-1}(z) r_2\big|^p \\ &\le \frac{K_p}{N^{p/2}}. \end{aligned}$$

Therefore

$$\mathsf{E}|\gamma_j|^p \le K_p N^{-1} \delta_n^{2p-4}, \qquad p \ge 2. \tag{3.5}$$

We next prove that $b_n(z)$ is bounded for all n. From (3.2) we find, for any $p \ge 1$,

$$\mathsf{E}|\beta_1(z)|^p \le K_p. \tag{3.6}$$

Since $b_n = \beta_1(z) + \beta_1(z) b_n(z) \gamma_1(z)$ we get from (3.5), (3.6)

$$|b_n(z)| = |\mathsf{E}\beta_1(z) + \mathsf{E}\beta_1(z) b_n(z)\gamma_1(z)| \le K_1 + K_2 |b_n(z)| N^{-1/2}.$$

Thus for all n large,

$$|b_n(z)| \le \frac{K_1}{1 - K_2 N^{-1/2}}$$

and subsequently $b_n(z)$ is bounded for all n.

From (3.3) we have

$$\begin{aligned}
&D^{-1}(z_1) D^{-1}(z_2) - D_j^{-1}(z_1) D_j^{-1}(z_2) \\
&\quad = \bigl(D^{-1}(z_1) - D_j^{-1}(z_1)\bigr)\bigl(D^{-1}(z_2) - D_j^{-1}(z_2)\bigr) \\
&\qquad + \bigl(D^{-1}(z_1) - D_j^{-1}(z_1)\bigr) D_j^{-1}(z_2) \\
&\qquad + D_j^{-1}(z_1)\bigl(D^{-1}(z_2) - D_j^{-1}(z_2)\bigr) \\
&\quad = \beta_j(z_1)\beta_j(z_2) D_j^{-1}(z_1) r_j r_j^* D_j^{-1}(z_1) D_j^{-1}(z_2) r_j r_j^* D_j^{-1}(z_2) \\
&\qquad - \beta_j(z_1) D_j^{-1}(z_1) r_j r_j^* D_j^{-1}(z_1) D_j^{-1}(z_2) \\
&\qquad - \beta_j(z_2) D_j^{-1}(z_1) D_j^{-1}(z_2) r_j r_j^* D_j^{-1}(z_2).
\end{aligned}$$

Therefore

$$\begin{aligned}
&\operatorname{tr}\bigl(D^{-1}(z_1) D^{-1}(z_2) - D_j^{-1}(z_1) D_j^{-1}(z_2)\bigr) \\
&\quad = \beta_j(z_1)\beta_j(z_2)\bigl(r_j^* D_j^{-1}(z_1) D_j^{-1}(z_2) r_j\bigr)^2 \\
&\qquad - \beta_j(z_1) r_j^* D_j^{-2}(z_1) D_j^{-1}(z_2) r_j - \beta_j(z_2) r_j^* D_j^{-2}(z_2) D_j^{-1}(z_1) r_j.
\end{aligned} \tag{3.7}$$

We write

$$\begin{aligned}
m_n(z_1) - m_n(z_2) &= \frac{1}{n} \operatorname{tr}\bigl(D^{-1}(z_1) - D^{-1}(z_2)\bigr) \\
&= \frac{1}{n}(z_1 - z_2) \operatorname{tr} D^{-1}(z_1) D^{-1}(z_2).
\end{aligned}$$

Therefore, from (3.7) we have

$$
\begin{aligned}
&n\frac{m_n(z_1)-m_n(z_2)-\mathsf{E}(m_n(z_1)-m_n(z_2))}{z_1-z_2}\\
&\quad=\sum_{j=1}^{N}(\mathsf{E}_j-\mathsf{E}_{j-1})\operatorname{tr}D^{-1}(z_1)D^{-1}(z_2)\\
&\quad=\sum_{j=1}^{N}(\mathsf{E}_j-\mathsf{E}_{j-1})\beta_j(z_1)\beta_j(z_2)\big(r_j^*D_j^{-1}(z_1)D_j^{-1}(z_2)r_j\big)^2\\
&\qquad-\sum_{j=1}^{N}(\mathsf{E}_j-\mathsf{E}_{j-1})\beta_j(z_1)r_j^*D_j^{-2}(z_1)D_j^{-1}(z_2)r_j\\
&\qquad-\sum_{j=1}^{N}(\mathsf{E}_j-\mathsf{E}_{j-1})\beta_j(z_2)r_j^*D_j^{-2}(z_2)D_j^{-1}(z_1)r_j.
\end{aligned}
\tag{3.8}
$$

Our goal is to show that the absolute second moment of (3.8) is bounded. We begin with the second sum in (3.8). We have

$$
\begin{aligned}
&\sum_{j=1}^{N}(\mathsf{E}_j-\mathsf{E}_{j-1})\beta_j(z_1)r_j^*D_j^{-2}(z_1)D_j^{-1}(z_2)r_j\\
&\quad=\sum_{j=1}^{N}(\mathsf{E}_j-\mathsf{E}_{j-1})\big(b_n(z_1)r_j^*D_j^{-2}(z_1)D_j^{-1}(z_2)r_j\\
&\qquad\qquad-\beta_j(z_1)b_n(z_1)r_j^*D_j^{-2}(z_1)D_j^{-1}(z_2)r_j\gamma_j(z_1)\big)\\
&\quad=b_n(z_1)\sum_{j=1}^{N}\mathsf{E}_j\big(r_j^*D_j^{-2}(z_1)D_j^{-1}(z_2)r_j\\
&\qquad\qquad-N^{-1}\operatorname{tr}T^{1/2}D_j^{-2}(z_1)D_j^{-1}(z_2)T^{1/2}\big)\\
&\qquad-b_n(z_1)\sum_{j=1}^{N}(\mathsf{E}_j-\mathsf{E}_{j-1})\beta_j(z_1)r_j^*D_j^{-2}(z_1)D_j^{-1}(z_2)r_j\gamma_j(z_1)\\
&\quad\equiv b_n(z_1)W_1-b_n(z_1)W_2.
\end{aligned}
$$

Using (3.2) we have

$$
\begin{aligned}
\mathsf{E}|W_1|^2&=\sum_{j=1}^{N}\mathsf{E}\big|\mathsf{E}_j\big(r_j^*D_j^{-2}(z_1)D_j^{-1}(z_2)r_j-N^{-1}\operatorname{tr}T^{1/2}D_j^{-2}(z_1)D_j^{-1}(z_2)T^{1/2}\big)\big|^2\\
&\le K.
\end{aligned}
$$

Using (3.5), and the bounds for $\beta_1(z_1)$ and $r_1^* D_1^{-2}(z_1) D_1^{-1}(z_2) r_1$ given in the remark to (3.2), we have

$$\begin{aligned}
\mathsf{E}|W_2|^2 &= \sum_{j=1}^{N} \mathsf{E}\big|(\mathsf{E}_j - \mathsf{E}_{j-1})\beta_j(z_1) r_j^* D_j^{-2}(z_1) D_j^{-1}(z_2) r_j \gamma_j(z_1)\big|^2 \\
&\le KN\big[\mathsf{E}|\gamma_1(z_1)|^2 + v^{-10} n^2 P\big(\|B_n\| > \eta_r \text{ or } \lambda_{\min}^{B_{(1)}} < \eta_l\big)\big] \\
&\le K.
\end{aligned}$$

This argument of course handles the third sum in (3.8).

For the first sum in (3.8) we have

$$\begin{aligned}
&\sum_{j=1}^{N} (\mathsf{E}_j - \mathsf{E}_{j-1})\beta_j(z_1)\beta_j(z_2)\big(r_j^* D_j^{-1}(z_1) D_j^{-1}(z_2) r_j\big)^2 \\
&\quad = b_n(z_1) b_n(z_2) \sum_{j=1}^{N} (\mathsf{E}_j - \mathsf{E}_{j-1})\big[\big(r_j^* D_j^{-1}(z_1) D_j^{-1}(z_2) r_j\big)^2 \\
&\qquad\qquad - \big(N^{-1} \operatorname{tr} T^{1/2} D_j^{-1}(z_1) D_j^{-1}(z_2) T^{1/2}\big)^2\big] \\
&\qquad - b_n(z_2) \sum_{j=1}^{N} (\mathsf{E}_j - \mathsf{E}_{j-1})\beta_j(z_1)\beta_j(z_2)\big(r_j^* D_j^{-1}(z_1) D_j^{-1}(z_2) r_j\big)^2 \gamma_j(z_2) \\
&\qquad - b_n(z_1) b_n(z_2) \sum_{j=1}^{N} (\mathsf{E}_j - \mathsf{E}_{j-1})\beta_j(z_1)\big(r_j^* D_j^{-1}(z_1) D_j^{-1}(z_2) r_j\big)^2 \gamma_j(z_1) \\
&\quad = b_n(z_1) b_n(z_2) Y_1 - b_n(z_2) Y_2 - b_n(z_1) b_n(z_2) Y_3.
\end{aligned}$$

Both Y_2 and Y_3 are handled the same way as W_2 above. Using (3.2) we have

$$\begin{aligned}
\mathsf{E}|Y_1|^2 &\le N\mathsf{E}\big|\big(r_1^* D_1^{-1}(z_1) D_1^{-1}(z_2) r_1\big)^2 \\
&\qquad - \big(N^{-1} \operatorname{tr} T^{1/2} D_1^{-1}(z_1) D_1^{-1}(z_2) T^{1/2}\big)^2\big|^2 \\
&\le N\Big(2\mathsf{E}\big|\big(r_1^* D_1^{-1}(z_1) D_1^{-1}(z_2) r_1 \\
&\qquad - N^{-1} \operatorname{tr} T^{1/2} D_1^{-1}(z_1) D_1^{-1}(z_2) T^{1/2}\big)\big|^4 \\
&\qquad + 4(nN^{-1})^2 \mathsf{E}\big|\big(r_1^* D_1^{-1}(z_1) D_1^{-1}(z_2) r_1 \\
&\qquad\qquad - N^{-1} \operatorname{tr} T^{1/2} D_1^{-1}(z_1) D_1^{-1}(z_2) T^{1/2}\big) \\
&\qquad\qquad\qquad \times \|D_1^{-1}(z_1) D_1^{-1}(z_2)\|\big|^2\Big) \\
&\le K.
\end{aligned}$$

Therefore, condition (ii) of Theorem 12.3 in Billingsley (1968) is satisfied, and we conclude that $\{\widehat{M}_n^1(z)\}$ is tight.

4. Convergence of $M_n^2(z)$. The proof of Lemma 1.1 is complete with the verification of $\{M_n^2(z)\}$ for $z \in \mathcal{C}_n$ to be bounded and form an equicontinuous family, and convergence to (1.12) under the assumptions in (ii) of Theorem 1.1 and to zero under those in (iii).

In order to simplify the exposition, we let $\mathcal{C}_1 = \mathcal{C}_u$ or $\mathcal{C}_u \cup \mathcal{C}_l$ if $x_l < 0$, and $\mathcal{C}_2 = \mathcal{C}_2(n) = \mathcal{C}_r$ or $\mathcal{C}_r \cup \mathcal{C}_l$ if $x_l > 0$. We begin with proving

$$\sup_{z \in \mathcal{C}_n} |\mathsf{E}\underline{m}_n(z) - \underline{m}(z)| \to 0 \qquad \text{as } n \to \infty. \tag{4.1}$$

Since $F^{\underline{B}_n} \xrightarrow{D} F^{c,H}$ almost surely, we get from d.c.t. $\mathsf{E}F^{\underline{B}_n} \xrightarrow{D} F^{c,H}$. It is easy to verify that $\mathsf{E}F^{\underline{B}_n}$ is a proper c.d.f. Since, as z ranges in $\mathcal{C}_1$, the functions $(\lambda - z)^{-1}$ in $\lambda \in [0, \infty)$ form a bounded, equicontinuous family, it follows [see, e.g., Billingsley (1968), Problem 8, page 17] that

$$\sup_{z \in \mathcal{C}_1} |\mathsf{E}\underline{m}_n(z) - \underline{m}(z)| \to 0.$$

For $z \in \mathcal{C}_2$ we write (η_l, η_r defined as in the previous section)

$$\begin{aligned}\mathsf{E}\underline{m}_n(z) - \underline{m}(z) &= \int \frac{1}{\lambda - z} I_{[\eta_l, \eta_r]}(\lambda)\, d\big(\mathsf{E}F^{\underline{B}_n}(\lambda) - F^{c,H}(\lambda)\big) \\ &\quad + \mathsf{E}\int \frac{1}{\lambda - z} I_{[\eta_l, \eta_r]^c}(\lambda)\, dF^{\underline{B}_n}(\lambda).\end{aligned}$$

As above, the first term converges uniformly to zero. For the second term we use (1.9) with $\ell \geq 2$ to get

$$\begin{aligned}&\sup_{z \in \mathcal{C}_2} \left| \mathsf{E}\int \frac{1}{\lambda - z} I_{[\eta_l, \eta_r]^c}(\lambda)\, dF^{\underline{B}_n}(\lambda) \right| \\ &\qquad \leq (\varepsilon_n/n)^{-1} \mathsf{P}\big(\|B_n\| \geq \eta_r \text{ or } \lambda_{\min}^{B_n} \leq \eta_l\big) \\ &\qquad \leq K n \varepsilon^{-1} n^{-\ell} \to 0.\end{aligned}$$

Thus (4.1) holds.

From the fact that $F^{c_n, H_n} \xrightarrow{D} F^{c,H}$ [see Bai and Silverstein (1998), below (3.10)] along with the fact that $\mathcal{C}$ lies outside the support of $F^{c,H}$, it is straightforward to verify that

$$\sup_{z \in \mathcal{C}} |\underline{m}_n^0(z) - \underline{m}(z)| \to 0 \qquad \text{as } n \to \infty. \tag{4.2}$$

We now show that

$$\sup_{n\, z \in \mathcal{C}_n} \big\| \big(\mathsf{E}\underline{m}_n(z) T + I\big)^{-1} \big\| < \infty. \tag{4.3}$$

From Lemma 2.11 of Bai and Silverstein (1998), $\|(\mathsf{E}\underline{m}_n(z)T + I)^{-1}\|$ is bounded by $\max(2, 4v_0^{-1})$ on $\mathcal{C}_u$. Let $x = x_l$ or x_r. Since x is outside the support of $\underline{F}^{c,H}$ it follows from Theorem 4.1 of Silverstein and Choi (1995) that for any t in the support of H $\underline{m}(x)t + 1 \neq 0$. Choose any t_0 in the support of H. Since $\underline{m}(z)$ is continuous on $\mathcal{C}^0 \equiv \{x + iv : v \in [0, v_0]\}$, there exist positive constants δ_1 and μ_0 such that

$$\inf_{z\in\mathcal{C}^0} |\underline{m}(z)t_0 + 1| > \delta_1 \quad \text{and} \quad \sup_{z\in\mathcal{C}^0} |\underline{m}(z)| < \mu_0.$$

Using $H_n \xrightarrow{D} H$ and (4.1), for all large n, there exists an eigenvalue λ^T of T such that $|\lambda^T - t_0| < \delta_1/4\mu_0$ and $\sup_{z\in\mathcal{C}_l\cup\mathcal{C}_r} |\mathsf{E}\underline{m}_n(z) - \underline{m}(z)| < \delta_1/4$. Therefore, we have

$$\inf_{z\in\mathcal{C}_l\cup\mathcal{C}_r} |\mathsf{E}\underline{m}_n(z)\lambda^T + 1| > \delta_1/2,$$

which completes the proof of (4.3).

Next we show the existence of $\xi \in (0, 1)$ such that for all n large

$$\sup_{z\in\mathcal{C}_n} \left| c_n \mathsf{E}\underline{m}_n(z)^2 \int \frac{t^2}{(1 + t\mathsf{E}\underline{m}_n(z))^2} \, dH_n(t) \right| < \xi. \tag{4.4}$$

From the identity (1.1) of Bai and Silverstein (1998),

$$\underline{m}(z) = \left(-z + c\int \frac{t}{1 + t\underline{m}(z)} \, dH(t)\right)^{-1}$$

valid for $z = x + iv$ outside the support of $F^{c,H}$; we find

$$\begin{aligned}\Im\,\underline{m}(z) &= \left(v + \Im\,\underline{m}(z)c\int \frac{t^2}{|1 + t\underline{m}(z)|^2} \, dH(t)\right) \\ &\quad \times \left|-z + c\int \frac{t}{1 + t\underline{m}(z)} \, dH(t)\right|^{-2}.\end{aligned}$$

Therefore

$$\begin{aligned}&\left|c\underline{m}(z)^2 \int \frac{t^2}{(1 + t\underline{m}(z))^2} \, dH(t)\right| \\ &\quad \le \left(c\int \frac{t^2}{|1 + t\underline{m}(z)|^2} \, dH(t)\right)\left|-z + c\int \frac{t}{1 + t\underline{m}(z)} \, dH(t)\right|^{-2} \\ &\quad = \left(\Im\,\underline{m}(z)c\int \frac{t^2}{|1 + t\underline{m}(z)|^2} \, dH(t)\right) \\ &\qquad \times \left[v + \Im\,\underline{m}(z)c\int \frac{t^2}{|1 + t\underline{m}(z)|^2} \, dH(t)\right]^{-1}\end{aligned} \tag{4.5}$$

$$= \left(c \int \frac{1}{|x-z|^2} \, dF^{c,H}(x) \int \frac{t^2}{|1+t\underline{m}(z)|^2} \, dH(t) \right)$$
$$\times \left[1 + c \int \frac{1}{|x-z|^2} \, dF^{c,H}(x) \int \frac{t^2}{|1+t\underline{m}(z)|^2} \, dH(t) \right]^{-1} < 1,$$

for all $z \in \mathcal{C}$. By continuity, we have the existence of $\xi_1 < 1$ such that

$$\sup_{z\in\mathcal{C}} \left| c\underline{m}(z)^2 \int \frac{t^2}{(1+t\underline{m}(z))^2} \, dH(t) \right| < \xi_1. \tag{4.6}$$

Therefore, using (4.1), (4.4) follows.

We proceed with some improved bounds on quantities appearing earlier.

Let M be nonrandom $n \times n$. Then, using (3.2) and the argument used to derive the bound on $\mathsf{E}|W_2|$, we find

$$\begin{aligned}
&\mathsf{E}|\operatorname{tr} D^{-1}M - \mathsf{E}\operatorname{tr} D^{-1}M|^2 \\
&\quad = \mathsf{E}\left| \sum_{j=1}^{N} \mathsf{E}_j \operatorname{tr} D^{-1}M - \mathsf{E}_{j-1} \operatorname{tr} D^{-1}M \right|^2 \\
&\quad = \mathsf{E}\left| \sum_{j=1}^{N} (\mathsf{E}_j - \mathsf{E}_{j-1}) \operatorname{tr}(D^{-1} - D_j^{-1})M \right|^2 \\
&\quad = \sum_{j=1}^{N} \mathsf{E}\big|(\mathsf{E}_j - \mathsf{E}_{j-1})\beta_j r_j^* D_j^{-1} M D_j^{-1} r_j\big|^2 \qquad (4.7)\\
&\quad \le 2 \sum_{j=1}^{N} \mathsf{E}\big|[\beta_j (r_j^* D_j^{-1} M D_j^{-1} r_j) - N^{-1} \operatorname{tr}(T D_j^{-1} M D_j^{-1})|^2 \\
&\qquad\qquad + \mathsf{E}|\beta_j - \bar{\beta}_j|^2 |N^{-1} \operatorname{tr}(T D_j^{-1} M D_j^{-1})|^2] \\
&\quad \le K\|M\|^2.
\end{aligned}$$

The same argument holds for D_1^{-1} so we also have

$$\mathsf{E}|\operatorname{tr} D_1^{-1}M - \mathsf{E}\operatorname{tr} D_1^{-1}M|^2 \le K\|M\|^2. \tag{4.8}$$

Our next task is to investigate the limiting behavior of

$$\begin{aligned}
&N\left(c_n \int \frac{dH_n(t)}{1+t\mathsf{E}\underline{m}_n} + z c_n \mathsf{E} m_n \right) \\
&\quad = N\mathsf{E}\beta_1 \left[r_1^* D_1^{-1} (\mathsf{E}\underline{m}_n T + I)^{-1} r_1 - \frac{1}{N} \mathsf{E}\operatorname{tr}(\mathsf{E}\underline{m}_n T + I)^{-1} T D^{-1} \right]
\end{aligned}$$

for $z \in \mathcal{C}_n$ [see (5.2) in Bai and Silverstein (1998)]. Throughout the following, all bounds, including $O(\cdot)$ and $o(\cdot)$ expressions, and convergence statements hold uniformly for $z \in \mathcal{C}_n$.

We have

$$\begin{aligned}
&\mathsf{E}\operatorname{tr}(\mathsf{E}\underline{m}_n T+I)^{-1}TD_1^{-1}-\mathsf{E}\operatorname{tr}(\mathsf{E}\underline{m}_n T+I)^{-1}TD^{-1}\\
&\qquad=\mathsf{E}\beta_1\operatorname{tr}(\mathsf{E}\underline{m}_n T+I)^{-1}TD_1^{-1}r_1r_1^*D_1^{-1}\\
&\qquad=b_n\mathsf{E}(1-\beta_1\gamma_1)r_1^*D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}TD_1^{-1}r_1.
\end{aligned}\tag{4.9}$$

From (3.2), (3.5) and (4.3) we get

$$\left|\mathsf{E}\beta_1\gamma_1r_1^*D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}TD_1^{-1}r_1\right|\le KN^{-1}.$$

Therefore

$$\left|(4.9)-N^{-1}b_n\mathsf{E}\operatorname{tr}D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}TD_1^{-1}T\right|\le KN^{-1}.$$

Since $\beta_1=b_n-b_n^2\gamma_1+\beta_1b_n^2\gamma_1^2$ we have

$$\begin{aligned}
&N\mathsf{E}\beta_1r_1^*D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}r_1-\mathsf{E}\beta_1\mathsf{E}\operatorname{tr}(\mathsf{E}\underline{m}_n T+I)^{-1}TD_1^{-1}\\
&\quad=-b_n^2N\mathsf{E}\gamma_1r_1^*D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}r_1\\
&\qquad+b_n^2\bigl(N\mathsf{E}\beta_1\gamma_1^2r_1^*D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}r_1\\
&\qquad\qquad-(\mathsf{E}\beta_1\gamma_1^2)\mathsf{E}\operatorname{tr}(\mathsf{E}\underline{m}_n T+I)^{-1}TD_1^{-1}\bigr)\\
&\quad=-b_n^2N\mathsf{E}\gamma_1r_1^*D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}r_1\\
&\qquad+b_n^2\bigl(\mathsf{E}\bigl[N\beta_1\gamma_1^2r_1^*D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}r_1\\
&\qquad\qquad-\beta_1\gamma_1^2\operatorname{tr}D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}T\bigr]\bigr)\\
&\qquad+b_n^2\operatorname{Cov}\bigl(\beta_1\gamma_1^2,\operatorname{tr}D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}T\bigr)
\end{aligned}$$

$[\operatorname{Cov}(X,Y)=\mathsf{E}XY-\mathsf{E}X\mathsf{E}Y]$. Using (3.2), (3.5) and (4.3), we have

$$\left|\mathsf{E}\bigl[N\beta_1\gamma_1^2r_1^*D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}r_1-\beta_1\gamma_1^2\operatorname{tr}D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}T\bigr]\right|\le K\delta_n^2.$$

Using (3.5), (3.6), (4.3) and (4.8) we have

$$\begin{aligned}
&\left|\operatorname{Cov}\bigl(\beta_1\gamma_1^2,\operatorname{tr}D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}T\bigr)\right|\\
&\quad\le(\mathsf{E}|\beta_1|^4)^{1/4}(\mathsf{E}|\gamma_1|^8)^{1/4}\\
&\qquad\times\bigl(\mathsf{E}\left|\operatorname{tr}D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}T-\mathsf{E}\operatorname{tr}D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}T\right|^2\bigr)^{1/2}\\
&\quad\le K\delta_n^3N^{-1/4}.
\end{aligned}$$

Since $\beta_1=b_n-b_n\beta_1\gamma_1$, we get from (3.5) and (3.6) $\mathsf{E}\beta_1=b_n+O(N^{-1/2})$.

Write

$$\begin{aligned}
&\mathsf{E}N\gamma_1 r_1^* D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}r_1\\
&\quad = N\mathsf{E}\big[(r_1^* D_1^{-1} r_1 - N^{-1}\operatorname{tr} D_1^{-1}T)\\
&\qquad\qquad \times \big(r_1^* D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}r_1 - N^{-1}\operatorname{tr} D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}T\big)\big]\\
&\qquad + N^{-1}\operatorname{Cov}\big(\operatorname{tr} D_1^{-1}T, \operatorname{tr} D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}T\big).
\end{aligned}$$

From (4.8) we see the second term above is $O(N^{-1})$. Therefore, we arrive at

$$\begin{aligned}
&N\left(c_n\int\frac{dH_n(t)}{1+t\mathsf{E}\underline{m}_n}+zc_n\mathsf{E}m_n\right)\\
&\quad = b_n^2 N^{-1}\mathsf{E}\operatorname{tr} D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}TD_1^{-1}T\\
&\qquad - b_n^2 N\mathsf{E}\big[(r_1^* D_1^{-1} r_1 - N^{-1}\operatorname{tr} D_1^{-1}T)\\
&\qquad\qquad \times \big(r_1^* D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}r_1\\
&\qquad\qquad\qquad - N^{-1}\operatorname{tr} D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}T\big)\big]\\
&\qquad + o(1).
\end{aligned}\tag{4.10}$$

Using (1.15) on (4.10) and arguing the same way (1.15) is used in Section 2 [below (2.7)], we see that under the assumptions in (iii) of Theorem 1.1, the CG case

$$N\left(c_n\int\frac{dH_n(t)}{1+t\mathsf{E}\underline{m}_n}+zc_n\mathsf{E}m_n\right)\to 0 \qquad \text{as } n\to\infty,\tag{4.11}$$

while under the assumptions in (ii) of Theorem 1.1, the RG case

$$\begin{aligned}
&N\left(c_n\int\frac{dH_n(t)}{1+t\mathsf{E}\underline{m}_n}+zc_n\mathsf{E}m_n\right)\\
&\quad = -b_n^2 N^{-1}\mathsf{E}\operatorname{tr} D_1^{-1}(\mathsf{E}\underline{m}_n T+I)^{-1}TD_1^{-1}T + o(1).
\end{aligned}$$

Let $A_n(z) = c_n\int\frac{dH_n(t)}{1+t\mathsf{E}\underline{m}_n(z)} + zc_n\mathsf{E}m_n(z)$. Using the identity

$$\mathsf{E}\underline{m}_n(z) = -\frac{(1-c_n)}{z} + c_n\mathsf{E}m_n$$

we have

$$\begin{aligned}
A_n(z) &= c_n\int\frac{dH_n(t)}{1+t\mathsf{E}\underline{m}_n(z)} - c_n + z\mathsf{E}\underline{m}_n(z)+1\\
&= -\mathsf{E}\underline{m}_n(z)\left(-z-\frac{1}{\mathsf{E}\underline{m}_n(z)}+c_n\int\frac{t\,dH_n(t)}{1+t\mathsf{E}\underline{m}_n(z)}\right).
\end{aligned}$$

It follows that

$$\mathsf{E}\underline{m}_n(z) = \left[-z + c_n \int \frac{t\,dH_n(t)}{1 + t\mathsf{E}\underline{m}_n(z)} + A_n/\mathsf{E}\underline{m}_n(z)\right]^{-1}.$$

From this, together with the analogous identity (4.4) we get

$$\begin{aligned}(4.12)\qquad &\mathsf{E}\underline{m}_n(z) - \underline{m}_n^0(z)\\ &\quad = -\underline{m}_n^0 A_n\left[1 - c_n \mathsf{E}\underline{m}_n \underline{m}_n^0 \int \frac{t^2\,dH_n(t)}{(1 + t\mathsf{E}\underline{m}_n)(1 + t\underline{m}_n^0)}\right]^{-1}.\end{aligned}$$

We see from (4.4) and the corresponding bound involving $\underline{m}_n^0(z)$, that the denominator of (4.12) is bounded away from zero.

Therefore from (4.11), in the CG case

$$\sup_{z\in\mathcal{C}_n} M_n^2(z) \to 0 \qquad \text{as } n \to \infty.$$

We now find the limit of $N^{-1}\mathsf{E}\,\mathrm{tr}\, D_1^{-1}(\mathsf{E}\underline{m}_n T + I)^{-1} T D_1^{-1} T$. Applications of (3.1)–(3.3), (3.6) and (4.3) show that both

$$\mathsf{E}\,\mathrm{tr}\, D_1^{-1}(\mathsf{E}\underline{m}_n T + I)^{-1} T D_1^{-1} T - \mathsf{E}\,\mathrm{tr}\, D^{-1}(\mathsf{E}\underline{m}_n T + I)^{-1} T D_1^{-1} T$$

and

$$\mathsf{E}\,\mathrm{tr}\, D^{-1}(\mathsf{E}\underline{m}_n T + I)^{-1} T D_1^{-1} T - \mathsf{E}\,\mathrm{tr}\, D^{-1}(\mathsf{E}\underline{m}_n T + I)^{-1} T D^{-1} T$$

are bounded. Therefore it is sufficient to consider

$$N^{-1}\mathsf{E}\,\mathrm{tr}\, D^{-1}(\mathsf{E}\underline{m}_n T + I)^{-1} T D^{-1} T.$$

Write

$$D(z) + zI - b_n(z)T = \sum_{j=1}^N r_j r_j^* - b_n(z)T.$$

It is straightforward to verify that $zI - b_n(z)T$ is nonsingular. Taking inverses we get

$$\begin{aligned}D^{-1}(z) &= -\big(zI - b_n(z)T\big)^{-1}\\ &\qquad + \sum_{j=1}^N \beta_j(z)\big(zI - b_n(z)T\big)^{-1} r_j r_j^* D_j(z)\\ (4.13)\qquad &\qquad - b_n(z)\big(zI - b_n(z)T\big)^{-1} T D^{-1}(z)\\ &= -\big(zI - b_n(z)T\big)^{-1} + b_n(z)A(z) + B(z) + C(z),\end{aligned}$$

where

$$A(z) = \sum_{j=1}^{N} \bigl(zI - b_n(z)T\bigr)^{-1}(r_j r_j^* - N^{-1}T)D_j^{-1}(z),$$

$$B(z) = \sum_{j=1}^{N} \bigl(\beta_j(z) - b_n(z)\bigr)\bigl(zI - b_n(z)T\bigr)^{-1} r_j r_j^* D_j^{-1}(z)$$

and

$$\begin{aligned} C(z) &= N^{-1} b_n(z)\bigl(zI - b_n(z)T\bigr)^{-1} T \sum_{j=1}^{N} \bigl(D_j^{-1}(z) - D^{-1}(z)\bigr) \\ &= N^{-1} b_n(z)\bigl(zI - b_n(z)T\bigr)^{-1} T \sum_{j=1}^{N} \beta_j(z) D_j^{-1}(z) r_j r_j^* D_j^{-1}(z). \end{aligned}$$

Since $\mathsf{E}\beta_1 = -z\mathsf{E}\underline{m}_n$ and $\mathsf{E}\beta_1 = b_n + O(N^{-1})$ we have $b_n \to -z\underline{m}$. From (4.3) it follows that $\|(zI - b_n(z)T)^{-1}\|$ is bounded.

We have by (3.5) and (3.6)

$$\mathsf{E}|\beta_1 - b_n|^2 = |b_n|^2 \mathsf{E}|\beta_1 \gamma_1|^2 \le K N^{-1}. \tag{4.14}$$

Let M be $n \times n$. From (3.1), (3.2), (3.6) and (4.14) we get

$$\begin{aligned} |N^{-1}\mathsf{E}\operatorname{tr} B(z)M| &\le K(\mathsf{E}|\beta_1 - b_n|^2)^{1/2}(\mathsf{E}|r_1^* r_1\|D_1^{-1}M\||^2)^{1/2} \\ &\le K N^{-1/2}(\mathsf{E}\|M\|^4)^{1/4} \end{aligned} \tag{4.15}$$

and

$$\begin{aligned} |N^{-1}\mathsf{E}\operatorname{tr} C(z)M| &\le K N^{-1}\mathsf{E}|\beta_1| r_1^* r_1 \|D_1^{-1}\|^2 \|M\| \\ &\le K N^{-1}(\mathsf{E}\|M\|^2)^{1/2}. \end{aligned} \tag{4.16}$$

For the following M, $n \times n$, is nonrandom, bounded in norm. Write

$$\operatorname{tr} A(z) T D^{-1} M = A_1(z) + A_2(z) + A_3(z), \tag{4.17}$$

where

$$A_1(z) = \operatorname{tr} \sum_{j=1}^{N} (zI - b_n T)^{-1} r_j r_j^* D_j^{-1} T (D^{-1} - D_j^{-1}) M,$$

$$A_2(z) = \operatorname{tr} \sum_{j=1}^{N} (zI - b_n T)^{-1} (r_j r_j^* D_j^{-1} T D_j^{-1} - N^{-1} T D_j^{-1} T D_j^{-1}) M$$

and

$$A_3(z) = \operatorname{tr} \sum_{j=1}^{N} (zI - b_n T)^{-1} N^{-1} T D_j^{-1} T (D_j^{-1} - D^{-1}) M.$$

We have $\mathsf{E}A_2(z) = 0$ and similarly to (4.16) we have

$$|\mathsf{E}N^{-1}A_3(z)| \le KN^{-1}. \tag{4.18}$$

Using (3.2) and (4.14) we get

$$\begin{aligned}
\mathsf{E}N^{-1}A_1(z) &= -\mathsf{E}\beta_1 r_1^* D_1^{-1} T D_1^{-1} r_1 r_1^* D_1^{-1} M(zI - b_n T)^{-1} r_1 \\
&= -b_n \mathsf{E}(N^{-1}\operatorname{tr} D_1^{-1} T D_1^{-1} T)\big(N^{-1}\operatorname{tr} D_1^{-1} M(zI - b_n T)^{-1} T\big) + o(1) \\
&= -b_n \mathsf{E}(N^{-1}\operatorname{tr} D^{-1} T D^{-1} T)\big(N^{-1}\operatorname{tr} D^{-1} M(zI - b_n T)^{-1} T\big) + o(1).
\end{aligned}$$

Using (3.1) and (4.7) we find

$$\begin{aligned}
&\big|\operatorname{Cov}\big(N^{-1}\operatorname{tr} D^{-1} T D^{-1} T, N^{-1}\operatorname{tr} D^{-1} M(zI - b_n T)^{-1} T\big)\big| \\
&\quad \le (\mathsf{E}|N^{-1}\operatorname{tr} D^{-1} T D^{-1} T|^2)^{1/2} N^{-1} \\
&\qquad \times \big(\mathsf{E}|\operatorname{tr} D^{-1} M(zI - b_n T)^{-1} T - \mathsf{E}D^{-1} M(zI - b_n T)^{-1} T|^2\big)^{1/2} \\
&\quad \le KN^{-1}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathsf{E}N^{-1}A_1(z) \\
&\quad = -b_n(\mathsf{E}N^{-1}\operatorname{tr} D^{-1} T D^{-1} T) \\
&\qquad \times \big(\mathsf{E}N^{-1}\operatorname{tr} D^{-1} M(zI - b_n T)^{-1} T\big) + o(1).
\end{aligned} \tag{4.19}$$

From (4.13), (4.15) and (4.16) we get

$$\begin{aligned}
&\mathsf{E}N^{-1}\operatorname{tr} D^{-1} T(zI - b_n T)^{-1} T \\
&\quad = N^{-1}\operatorname{tr}\big(-(zI - b_n T)^{-1} + \mathsf{E}B(z) + \mathsf{E}C(z)\big) T(zI - b_n T)^{-1} T \\
&\quad = -\frac{c_n}{z^2}\int \frac{t^2\,dH_n(t)}{(1 + t\mathsf{E}\underline{m}_n)^2} + o(1).
\end{aligned} \tag{4.20}$$

Similarly,

$$\begin{aligned}
&\mathsf{E}N^{-1}\operatorname{tr} D^{-1}(\mathsf{E}\underline{m}_n T + I)^{-1} T(zI - b_n T)^{-1} T \\
&\quad = -\frac{c_n}{z^2}\int \frac{t^2\,dH_n(t)}{(1 + t\mathsf{E}\underline{m}_n)^3} + o(1).
\end{aligned} \tag{4.21}$$

Using (4.13) and (4.15)–(4.20) we get

$$\begin{aligned}
&\mathsf{E}N^{-1}\operatorname{tr} D^{-1} T D^{-1} T \\
&\quad = -\mathsf{E}N^{-1} D^{-1} T(zI - b_n T)^{-1} T \\
&\qquad - b_n^2(\mathsf{E}N^{-1}\operatorname{tr} D^{-1} T D^{-1} T)\big(\mathsf{E}N^{-1}\operatorname{tr} D^{-1} T(zI - b_n T)^{-1} T\big) + o(1) \\
&\quad = \frac{c_n}{z^2}\int \frac{t^2\,dH_n(t)}{(1 + t\mathsf{E}\underline{m}_n)^2}\big(1 + z^2\mathsf{E}\underline{m}_n^2\mathsf{E}N^{-1}\operatorname{tr} D^{-1} T D^{-1} T\big) + o(1).
\end{aligned}$$

Therefore

$$\mathsf{E}N^{-1}\operatorname{tr}D^{-1}TD^{-1}T = \left[\frac{c_n}{z^2}\int\frac{t^2\,dH_n(t)}{(1+t\mathsf{E}\underline{m}_n)^2}\right]\left[1-c_n\int\frac{\mathsf{E}\underline{m}_n^2t^2\,dH_n(t)}{(1+t\mathsf{E}\underline{m}_n)^2}\right]^{-1}+o(1). \tag{4.22}$$

Finally we have from (4.13)–(4.19), (4.21) and (4.22)

$$\begin{aligned}
&N^{-1}\mathsf{E}\operatorname{tr}D^{-1}(\mathsf{E}\underline{m}_nT+I)^{-1}TD^{-1}T\\
&\quad=-\mathsf{E}N^{-1}D^{-1}(\mathsf{E}\underline{m}_nT+I)^{-1}T(zI-b_nT)^{-1}T\\
&\qquad-b_n^2(\mathsf{E}N^{-1}\operatorname{tr}D^{-1}TD^{-1}T)\\
&\qquad\quad\times\big(\mathsf{E}N^{-1}\operatorname{tr}D^{-1}(\mathsf{E}\underline{m}_nT+I)^{-1}T(zI-b_nT)^{-1}T\big)+o(1)\\
&\quad=\frac{c_n}{z^2}\int\frac{t^2\,dH_n(t)}{(1+t\mathsf{E}\underline{m}_n)^3}\\
&\qquad\times\left(1+z^2\mathsf{E}\underline{m}_n^2\left[\frac{c_n}{z^2}\int\frac{t^2\,dH_n(t)}{(1+t\mathsf{E}\underline{m}_n)^2}\right]\left[1-c_n\int\frac{\mathsf{E}\underline{m}_n^2t^2\,dH_n(t)}{(1+t\mathsf{E}\underline{m}_n)^2}\right]^{-1}\right)+o(1)\\
&\quad=\left[\frac{c_n}{z^2}\int\frac{t^2\,dH_n(t)}{(1+t\mathsf{E}\underline{m}_n)^3}\right]\left[1-c_n\int\frac{\mathsf{E}\underline{m}_n^2t^2\,dH_n(t)}{(1+t\mathsf{E}\underline{m}_n)^2}\right]^{-1}+o(1).
\end{aligned}$$

Therefore, from (4.12) we conclude that in the RG case

$$\sup_{z\in\mathcal{C}_n}M_n^2(z)\to c\int\frac{\underline{m}(z)^3t^2\,dH(t)}{(1+t\underline{m}(z))^3}\left(1-c\int\frac{\underline{m}(z)^2t^2\,dH(t)}{(1+t\underline{m}(z))^2}\right)^{-2}\qquad\text{as } n\to\infty,$$

which is (1.12).

Finally, for general standardized X_{11}, we see that in light of the above work, in order to show $\{M_n^2(z)\}$ for $z\in\mathcal{C}_n$ is bounded and equicontinuous, it is sufficient to prove $\{f_n'(z)\}$, where

$$\begin{aligned}
f_n(z)\equiv N\mathsf{E}\big[&(r_1^*D_1^{-1}r_1-N^{-1}\operatorname{tr}D_1^{-1}T)\\
&\times\big(r_1^*D_1^{-1}(\mathsf{E}\underline{m}_nT+I)^{-1}r_1-N^{-1}\operatorname{tr}D_1^{-1}(\mathsf{E}\underline{m}_nT+I)^{-1}T\big)\big]
\end{aligned}$$

is bounded. Using (2.3) we find

$$\begin{aligned}
|f'(z)|\le KN^{-1}\Big(&\big(\mathsf{E}(\operatorname{tr}D_1^{-2}T\overline{D}_1^{-2}T)\\
&\quad\times\mathsf{E}\big(\operatorname{tr}D_1^{-1}(\mathsf{E}\underline{m}_nT+I)^{-1}T(\mathsf{E}\,\overline{\underline{m}}_nT+I)^{-1}\overline{D}_1^{-1}T\big)\big)^{1/2}\\
&+\big(\mathsf{E}(\operatorname{tr}D_1^{-1}T\overline{D}_1^{-1}T)\\
&\quad\times\mathsf{E}\big(\operatorname{tr}D_1^{-2}(\mathsf{E}\underline{m}_nT+I)^{-1}T(\mathsf{E}\,\overline{\underline{m}}_nT+I)^{-1}\overline{D}_1^{-2}T\big)\big)^{1/2}
\end{aligned}$$

$$+ |\mathsf{E}\underline{m}'_n| \big(\mathsf{E}(\operatorname{tr} D_1^{-1} T \overline{D}_1^{-1} T) \\ \times \mathsf{E}\big(\operatorname{tr} D_1^{-1} (\mathsf{E}\underline{m}_n T + I)^{-2} T^3 \\ \times (\mathsf{E}\,\overline{\underline{m}}_n T + I)^{-2} \overline{D}_1^{-1} T)\big)^{1/2}\big).$$

Using the same argument resulting in (3.1) it is a simple matter to conclude that $\mathsf{E}\underline{m}'_n(z)$ is bounded for $z \in \mathcal{C}_n$. All the remaining expected values are $O(N)$ due to (3.1) and (4.3), and we are done.

5. Some derivations and calculations. This section contains proofs of formulas stated in Section 1. We begin with deriving some properties of $\underline{m}(z)$. We claim that for any bounded subset S of $\mathbb{C}$,

$$\inf_{z \in S} |\underline{m}(z)| > 0. \tag{5.1}$$

Suppose not. Then there exists a sequence $\{z_n\} \subset \mathbb{C}^+$ which converges to a number for which $\underline{m}(z_n) \to 0$. From (1.2) we must have

$$c \int \frac{t\underline{m}(z_n)}{1 + t\underline{m}(z_n)} \, dH(t) \to 1.$$

However, because H has bounded support, the limit of the left-hand side of the above is obviously 0. The contradiction proves our assertion.

Next, we find a lower bound on the size of the difference quotient $(\underline{m}(z_1) - \underline{m}(z_2))/(z_1 - z_2)$ for distinct $z_1 = x + iv_1$, $z_2 = y + iv_2$, $v_1, v_2 \neq 0$. From (1.2) we get

$$z_1 - z_2 = \frac{\underline{m}(z_1) - \underline{m}(z_2)}{\underline{m}(z_1)\underline{m}(z_2)} \left(1 - c \int \frac{\underline{m}(z_1)\underline{m}(z_2) t^2 \, dH(t)}{(1 + t\underline{m}(z_1))(1 + t\underline{m}(z_2))}\right).$$

Therefore, from (2.19) we can write

$$\frac{\underline{m}(z_1) - \underline{m}(z_2)}{z_1 - z_2} \\ = [\underline{m}(z_1)\underline{m}(z_2)] \left[1 - c \int \frac{\underline{m}(z_1)\underline{m}(z_2) t^2 \, dH(t)}{(1 + t\underline{m}(z_1))(1 + t\underline{m}(z_2))}\right]^{-1}$$

and conclude that

$$\left|\frac{\underline{m}(z_1) - \underline{m}(z_2)}{z_1 - z_2}\right| \geq \frac{1}{2} |\underline{m}(z_1)\underline{m}(z_2)|. \tag{5.2}$$

We proceed to show (1.17). Choose $f, g \in \{f_1, \dots, f_k\}$. Let S_F denote the support of $F^{c,H}$, and let $a \neq 0$, b be such that S_F is a subset of (a, b), on whose closure f and g are analytic. Assume the z_1 contour encloses the z_2 contour. Using integration by parts twice, first with respect to z_2 and then with respect to z_1, we get

$$\begin{aligned}
\text{RHS of (1.7)} &= \frac{1}{2\pi^2}\iint \frac{f(z_1)g'(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))}\frac{d}{dz_1}\underline{m}(z_1)\,dz_2\,dz_1\\
&= -\frac{1}{2\pi^2}\iint f'(z_1)g'(z_2)\log\bigl(\underline{m}(z_1)-\underline{m}(z_2)\bigr)\,dz_1\,dz_2\\
&\qquad\qquad\text{(where log is any branch of the logarithm)}\\
&= -\frac{1}{2\pi^2}\iint f'(z_1)g'(z_2)\bigl[\ln|\underline{m}(z_1)-\underline{m}(z_2)|\\
&\qquad\qquad + i\arg\bigl(\underline{m}(z_1)-\underline{m}(z_2)\bigr)\bigr]\,dz_1\,dz_2.
\end{aligned}$$

We choose the contours to be rectangles with sides parallel to the axes. The inside rectangle intersects the real axis at a and b, and the horizontal sides are a distance $v < 1$ away from the real axis. The outside rectangle intersects the real axis at $a-\varepsilon$, $b+\varepsilon$ (points where f and g remain analytic), with height twice that of the inside rectangle. We let $v \to 0$.

We need only consider the logarithm term and show its convergence, since the real part of the arg term disappears (f and g are real valued on $\mathbb{R}$) in the limit, and the sum (1.7) is real. Therefore the arg term also approaches zero.

We split up the log integral into 16 double integrals, each one involving a side from each of the two rectangles. We argue that any portion of the integral involving a vertical side can be neglected. This follows from (5.1), (5.2) and the fact that z_1 and z_2 remain a positive distance apart, so that $|\underline{m}(z_1)-\underline{m}(z_2)|$ is bounded away from zero. Moreover, at least one of $|\underline{m}(z_1)|$, $|\underline{m}(z_2)|$ is bounded, while the other is bounded by $1/v$, so the integral is bounded by $Kv\ln v^{-1} \to 0$.

Therefore we arrive at

$$\begin{aligned}
-\frac{1}{2\pi^2}\int_a^b\int_{a-\varepsilon}^{b+\varepsilon}\bigl[&\bigl(f'(x+i2v)g'(y+iv)+\bar f'(x+i2v)\bar g'(y+iv)\bigr)\\
&\times\ln|\underline{m}(x+i2v)-\underline{m}(y+iv)|\\
&-\bigl(f'(x+i2v)\bar g'(y+iv)\\
&\quad+\bar f'(x+i2v)g'(y+iv)\bigr)\\
&\qquad\times\ln|\underline{m}(x+i2v)-\overline{\underline{m}}(y+iv)|\bigr]\,dx\,dy.
\end{aligned}\tag{5.3}$$

Using subscripts to denote real and imaginary parts, we find

$$
\begin{aligned}
(5.3) = -\frac{1}{\pi^2}\int_a^b \int_{a-\varepsilon}^{b+\varepsilon} \bigl[\bigl(f_r'(x+i2v)g_r'(y+iv) - f_i'(x+i2v)g_i'(y+iv)\bigr) \\
\times \ln|\underline{m}(x+i2v) - \underline{m}(y+iv)| \\
- \bigl(f_r'(x+i2v)g_r'(y+iv) \\
+ f_i'(x+i2v)g_i'(y+iv)\bigr) \\
\times \ln|\underline{m}(x+i2v) - \overline{\underline{m}}(y+iv)|\bigr]\,dx\,dy
\end{aligned}
$$

$$
\begin{aligned}
(5.4)\quad &= \frac{1}{\pi^2}\int_a^b \int_{a-\varepsilon}^{b+\varepsilon} f_r'(x+i2v)g_r'(y+iv)\ln\left|\frac{\underline{m}(x+i2v) - \overline{\underline{m}}(y+iv)}{\underline{m}(x+i2v) - \underline{m}(y+iv)}\right|dx\,dy \\
&\quad + \frac{1}{\pi^2}\int_a^b \int_{a-\varepsilon}^{b+\varepsilon} f_i'(x+i2v)g_i'(y+iv) \\
(5.5)\quad &\qquad \times \ln\bigl|\bigl(\underline{m}(x+i2v) - \underline{m}(y+iv)\bigr) \\
&\qquad \times \bigl(\underline{m}(x+i2v) - \overline{\underline{m}}(y+iv)\bigr)\bigr|\,dx\,dy.
\end{aligned}
$$

We have for any real-valued h, analytic on the bounded interval $[\alpha,\beta]$ for all v sufficently small

$$
\sup_{x\in[\alpha,\beta]} |h_i(x+iv)| \le K|v|, \tag{5.6}
$$

where K is a bound on $|h'(z)|$ for z in a neighborhood of $[\alpha,\beta]$. Using this and (5.1), (5.2) we see that (5.5) is bounded in absolute value by $Kv^2\ln v^{-1} \to 0$.

For (5.4) we write

$$
\ln\left|\frac{\underline{m}(x+i2v) - \overline{\underline{m}}(y+iv)}{\underline{m}(x+i2v) - \underline{m}(y+iv)}\right| = \frac{1}{2}\ln\biggl(1 + \frac{4\underline{m}_i(x+i2v)\underline{m}_i(y+iv)}{|\underline{m}(x+i2v) - \underline{m}(y+iv)|^2}\biggr). \tag{5.7}
$$

From (5.2) we get

$$
\text{RHS of (5.7)} \le \frac{1}{2}\ln\biggl(1 + \frac{16\underline{m}_i(x+i2v)\underline{m}_i(y+iv)}{(x-y)^2|\underline{m}(x+i2v)\underline{m}(y+iv)|^2}\biggr).
$$

From (5.1) we have

$$
\sup_{\substack{x,y\in[a-\varepsilon,b+\varepsilon]\\ v\in(0,1)}} \frac{\underline{m}_i(x+i2v)\underline{m}_i(y+iv)}{|\underline{m}(x+i2v)\underline{m}(y+iv)|^2} < \infty.
$$

Therefore, there exists a $K > 0$ for which the right-hand side of (5.7) is bounded by

$$\frac{1}{2}\ln\left(1+\frac{K}{(x-y)^2}\right) \tag{5.8}$$

for $x, y \in [a-\varepsilon, b+\varepsilon]$. It is straightforward to show that (5.8) is Lebesgue integrable on bounded subsets of $\mathbb{R}^2$. Therefore, from (1.19) and the dominated convergence theorem we conclude that (1.20) is Lebesgue integrable and that (1.17) holds.

We now verify (1.18). From (1.2) we have

$$\frac{d}{dz}\underline{m}(z) = \underline{m}^2(z)\left[1 - c\int \frac{t^2\underline{m}^2(z)}{(1+t\underline{m}(z))^2}\,dH(t)\right]^{-1}.$$

In Silverstein and Choi (1995) it is argued that the only place where $\underline{m}'(z)$ can possibly become unbounded are near the origin and the boundary, ∂S_F, of S_F. It is a simple matter to verify

$$\begin{aligned}\mathsf{E}X_f &= \frac{1}{4\pi i}\int f(z)\frac{d}{dz}\log\left(1 - c\int\frac{t^2\underline{m}^2(z)}{(1+t\underline{m}(z))^2}\,dH(t)\right)dz\\ &= -\frac{1}{4\pi i}\int f'(z)\log\left(1 - c\int\frac{t^2\underline{m}^2(z)}{(1+t\underline{m}(z))^2}\,dH(t)\right)dz,\end{aligned}$$

where, because of (2.19), the arg term for log can be taken from $(-\pi/2, \pi/2)$. We choose a contour as above. From (3.17) of Bai and Silverstein (1998) there exists a $K > 0$ such that for all small v,

$$\inf_{x\in\mathbb{R}}\left|1 - c\int\frac{t^2\underline{m}^2(x+iv)}{(1+t\underline{m}(x+iv))^2}\,dH(t)\right| \geq Kv^2. \tag{5.9}$$

Therefore, we see the integrals on the two vertical sides are bounded by $Kv\ln v^{-1} \to 0$. The integral on the two horizontal sides is equal to

$$\begin{aligned}&\frac{1}{2\pi}\int_a^b f_i'(x+iv)\ln\left|1 - c\int\frac{t^2\underline{m}^2(x+iv)}{(1+t\underline{m}(x+iv))^2}\,dH(t)\right|dx\\ &\quad+\frac{1}{2\pi}\int_a^b f_r'(x+iv)\arg\left(1 - c\int\frac{t^2\underline{m}^2(x+iv)}{(1+t\underline{m}(x+iv))^2}\,dH(t)\right)dx.\end{aligned} \tag{5.10}$$

Using (2.19), (5.6) and (5.9) we see the first term in (5.10) is bounded in absolute value by $Kv\ln v^{-1} \to 0$. Since the integrand in the second term converges for all $x \notin \{0\} \cup \partial S_F$ (a countable set) we get, therefore, (1.18) from the dominated convergence theorem.

We now derive $d(c)$ ($c \in (0,1)$) in (1.1), (1.21) and the variance in (1.22). The first two rely on Poisson's integral formula

$$u(z) = \frac{1}{2\pi}\int_0^{2\pi} u(e^{i\theta})\frac{1-r^2}{1+r^2-2r\cos(\theta-\phi)}\,d\theta, \tag{5.11}$$

where u is harmonic on the unit disk in $\mathbb{C}$, and $z = re^{i\phi}$ with $r \in [0, 1)$. Making the substitution $x = 1 + c - 2\sqrt{c}\cos\theta$ we get

$$\begin{aligned} d(c) &= \frac{1}{\pi}\int_0^{2\pi} \frac{\sin^2\theta}{1 + c - 2\sqrt{c}\cos\theta} \ln\bigl(1 + c - 2\sqrt{c}\cos\theta\bigr)\, d\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi} \frac{2\sin^2\theta}{1 + c - 2\sqrt{c}\cos\theta} \ln\left|1 - \sqrt{c}e^{i\theta}\right|^2 d\theta. \end{aligned}$$

It is straightforward to verify that

$$f(z) \equiv -(z - z^{-1})^2\bigl(\log\bigl(1 - \sqrt{c}z\bigr) + \sqrt{c}z\bigr) - \sqrt{c}(z - z^3)$$

is analytic on the unit disk, and that

$$\Re f(e^{i\theta}) = 2\sin^2\theta \ln\left|1 - \sqrt{c}e^{i\theta}\right|^2.$$

Therefore from (5.11) we have

$$d(c) = \frac{f(\sqrt{c})}{1 - c} = \frac{c - 1}{c}\ln(1 - c) - 1.$$

For (1.21) we use (1.18). From (1.2), with $H(t) = I_{[1,\infty)}(t)$ we have for $z \in \mathbb{C}^+$

$$z = -\frac{1}{\underline{m}(z)} + \frac{c}{1 + \underline{m}(z)}. \tag{5.12}$$

Solving for $\underline{m}(z)$ we find

$$\begin{aligned} \underline{m}(z) &= \frac{-(z + 1 - c) + \sqrt{(z + 1 - c)^2 - 4z}}{2z} \\ &= \frac{-(z + 1 - c) + \sqrt{(z - 1 - c)^2 - 4c}}{2z}, \end{aligned}$$

the square roots defined to yield positive imaginary parts for $z \in \mathbb{C}^+$. As $z \to x \in [a(y), b(y)]$ [limits defined below (1.1)] we get

$$\begin{aligned} \underline{m}(x) &= \frac{-(x + 1 - c) + \sqrt{4c - (x - 1 - c)^2}\, i}{2x} \\ &= \frac{-(x + 1 - c) + \sqrt{(x - a(c))(b(c) - x)}\, i}{2x}. \end{aligned}$$

The identity (5.12) still holds with z replaced by x and from it we get

$$\frac{\underline{m}(x)}{1 + \underline{m}(x)} = \frac{1 + x\underline{m}(x)}{c},$$

so that

$$\begin{aligned}
1 - c\frac{\underline{m}^2(x)}{(1+\underline{m}(x))^2}
&= 1 - \frac{1}{c}\left(\frac{-(x-1-c) + \sqrt{4c-(x-1-c)^2}\, i}{2}\right)^2 \\
&= \frac{\sqrt{4c-(x-1-c)^2}}{2c}\left(\sqrt{4c-(x-1-c)^2} + (x-1-c)i\right).
\end{aligned}$$

Therefore, from (1.18)

$$\begin{aligned}
\mathsf{E}X_f &= \frac{1}{2\pi}\int_{a(c)}^{b(c)} f'(x)\tan^{-1}\left(\frac{x-1-c}{\sqrt{4c-(x-1-c)^2}}\right)dx \\
&= \frac{f(a(c))+f(b(c))}{4} - \frac{1}{2\pi}\int_{a(c)}^{b(c)} \frac{f(x)}{\sqrt{4c-(x-1-c)^2}}\,dx.
\end{aligned} \tag{5.13}$$

To compute the last integral when $f(x) = \ln x$ we make the same substitution as before, arriving at

$$\frac{1}{4\pi}\int_0^{2\pi} \ln\left|1-\sqrt{c}e^{i\theta}\right|^2 d\theta.$$

We apply (5.11) where now $u(z) = \ln|1-\sqrt{c}z|^2$, which is harmonic, and $r = 0$. Therefore, the integral must be zero, and we conclude

$$\mathsf{E}X_{\ln} = \frac{\ln(a(c)b(c))}{4} = \frac{1}{c}\ln(1-c).$$

To derive (1.22) we use (1.16). Since the z_1, z_2 contours cannot enclose the origin (because of the logarithm), neither can the resulting m_1, m_2 contours. Indeed, either from the graph of $x(\underline{m})$ or from $\underline{m}(x)$ we see that $x > b(c) \Leftrightarrow \underline{m}(x) \in (-(1+\sqrt{c})^{-1}, 0)$ and $x \in (0, a(y)) \Leftrightarrow \underline{m}(x) < (\sqrt{c}-1)^{-1}$. For our analysis it is sufficient to know that the m_1, m_2 contours, nonintersecting and both taken in the positive direction, enclose $(c-1)^{-1}$ and -1, but not 0. Assume the m_2 contour encloses the m_1 contour. For fixed m_2, using (5.12) we have

$$\begin{aligned}
\int \frac{\log(z(m_1))}{(m_1-m_2)^2}\,dm_1
&= \int \frac{1/m_1^2 - c/(1+m_1)^2}{-1/m_1 + c/(1+m_1)} \frac{1}{(m_1-m_2)}\,dm_1 \\
&= \int \frac{(1+m_1)^2 - cm_1^2}{cm_1(m_1-m_2)}\left(\frac{-1}{m_1+1} + \frac{1}{m_1 - 1/(c-1)}\right)dm_1 \\
&= 2\pi i\left(\frac{1}{m_2+1} - \frac{1}{m_2 - 1/(c-1)}\right).
\end{aligned}$$

Therefore

$$
\begin{aligned}
\operatorname{Var} X_{\ln} &= \frac{1}{\pi i}\int\left(\frac{1}{m+1}-\frac{1}{m-1/(c-1)}\right)\log(z(m))\,dm \\
&= \frac{1}{\pi i}\int\left[\frac{1}{m+1}-\frac{1}{m-1/(c-1)}\right]\log\left(\frac{m-1/(c-1)}{m+1}\right)dm \\
&\quad -\frac{1}{\pi i}\int\left[\frac{1}{m+1}-\frac{1}{m-1/(c-1)}\right]\log(m)\,dm.
\end{aligned}
$$

The first integral is zero since the integrand has antiderivative

$$
-\frac{1}{2}\left[\log\left(\frac{m-1/(c-1)}{m+1}\right)\right]^2,
$$

which is single valued along the contour. Therefore we conclude that

$$
\operatorname{Var} X_{\ln} = -2\big[\log(-1)-\log\big((c-1)^{-1}\big)\big] = -2\ln(1-c).
$$

Finally, we compute expressions for (1.23) and (1.24). Using (5.13) we have

$$
\begin{aligned}
\mathsf{E}X_{x^r} &= \frac{(a(c))^r+(b(c))^r}{4}-\frac{1}{2\pi}\int_{a(c)}^{b(c)}\frac{x^r}{\sqrt{4c-(x-1-c)^2}}\,dx \\
&= \frac{(a(c))^r+(b(c))^r}{4}-\frac{1}{4\pi}\int_0^{2\pi}\left|1-\sqrt{c}e^{i\theta}\right|^{2r}d\theta \\
&= \frac{(a(c))^r+(b(c))^r}{4}-\frac{1}{4\pi}\int_0^{2\pi}\sum_{j,k=0}^{r}\binom{r}{j}\binom{r}{k}(-\sqrt{c})^{j+k}e^{i(j-k)\theta}\,d\theta \\
&= \frac{1}{4}\left((1-\sqrt{c})^{2r}+(1+\sqrt{c})^{2r}\right)-\frac{1}{2}\sum_{j=0}^{r}\binom{r}{j}^2 c^j,
\end{aligned}
$$

which is (1.23).

For (1.24) we use (1.16) and rely on observations made in deriving (1.22). For $c\in(0,1)$ the contours can again be made enclosing -1 and not the origin. However, because of the fact that (1.7) derives from (1.14) and the support of $F^{c,I_{[1,\infty)}}$ on $\mathbb{R}^+$ is $[a(c),b(c)]$, we may also take the contours taken in the same way when $c>1$. The case $c=1$ simply follows from the continuous dependence of (1.16) on c.

Keeping m_2 fixed, we have on a contour within 1 of -1

$$
\begin{aligned}
&\int\frac{(-1/m_1+c/(1+m_1))^{r_1}}{(m_1-m_2)^2}\,dm_1 \\
&\quad = c^{r_1}\int\left(\frac{1}{m_1+1}+\frac{1-c}{c}\right)^{r_1}(1-(m_1+1))^{-r_1} \\
&\qquad\quad \times(m_2+1)^{-2}\left(1-\frac{m_1+1}{m_2+1}\right)^{-2}dm_1
\end{aligned}
$$

$$
\begin{aligned}
&= c^{r_1} \int \sum_{k_1=0}^{r_1} \binom{r_1}{k_1} \left(\frac{1-c}{c}\right)^{k_1} (1+m_1)^{k-r_1} \\
&\qquad \times \sum_{j=0}^{\infty} \binom{r_1+j-1}{j} (m_1+1)^j (m_2+1)^{-2} \sum_{\ell=1}^{\infty} \ell \left(\frac{m_1+1}{m_2+1}\right)^{\ell-1} dm_1 \\
&= 2\pi i c^{r_1} \sum_{k_1=0}^{r_1-1} \sum_{\ell=1}^{r_1-k_1} \binom{r_1}{k_1} \left(\frac{1-c}{c}\right)^{k_1} \binom{2r_1-1-(k_1+\ell)}{r_1-1} \ell (m_2+1)^{-\ell-1}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\operatorname{Cov}(X_{x^{r_1}}, X_{x^{r_2}}) \\
&\quad = -\frac{i}{\pi} c^{r_1+r_2} \sum_{k_1=0}^{r_1-1} \sum_{\ell=1}^{r_1-k_1} \binom{r_1}{k_1} \left(\frac{1-c}{c}\right)^{k_1} \binom{2r_1-1-(k_1+\ell)}{r_1-1} \ell \\
&\qquad\qquad \times \int (m_2+1)^{-\ell-1} \sum_{k_2=0}^{r_2} \binom{r_2}{k_2} \left(\frac{1-c}{c}\right)^{k_2} (m_2+1)^{k_2-r_2} \\
&\qquad\qquad\qquad \times \sum_{j=0}^{\infty} \binom{r_2+j-1}{j} (m_2+1)^j \, dm_2 \\
&\quad = 2c^{r_1+r_2} \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2} \binom{r_1}{k_1} \binom{r_2}{k_2} \left(\frac{1-c}{c}\right)^{k_1+k_2} \\
&\qquad\qquad \times \sum_{\ell=1}^{r_1-k_1} \ell \binom{2r_1-1-(k_1+\ell)}{r_1-1} \binom{2r_2-1-k_2+\ell}{r_2-1},
\end{aligned}
$$

which is (1.24), and we are done.

APPENDIX

We verify (1.9b) by modifying the proof in Bai and Yin (1993) [hereafter referred to as BY (1993)]. To avoid confusion we maintain as much as possible the original notation used in BY (1993).

THEOREM. *For* $Z_{ij} \in \mathbb{C}$, $i = 1, \dots, p$, $j = 1, \dots, n$ *i.i.d.* $\mathsf{E}Z_{11} = 0$, $\mathsf{E}|Z_{11}|^2 = 1$, *and* $\mathsf{E}|Z_{11}|^4 < \infty$; *let* $S_n = (1/n)XX^*$ *where* $X = (X_{ij})$ *is* $p \times n$ *with*

$$
X_{ij} = X_{ij}(n) = Z_{ij} I_{\{|Z|_{ij} \le \delta_n \sqrt{n}\}} - \mathsf{E}Z_{ij} I_{\{|Z|_{ij} \le \delta_n \sqrt{n}\}},
$$

where $\delta_n \to 0$ *more slowly than that constructed in the proof of Lemma* 2.2 *of Yin, Bai and Krishnaiah* (1988) *and satisfying* $\delta_n n^{1/3} \to \infty$. *Assume* $p/n \to y \in (0, 1)$

as $n \to \infty$. *Then for any* $\eta < (1-\sqrt{y})^2$ *and any* $\ell > 0$

$$\mathsf{P}\big(\lambda_{\min}(S_n) < \eta\big) = o(n^{-\ell}).$$

PROOF. We follow along the proof of Theorem 1 in BY (1993). The conclusions of Lemmas 1 and 3–8 need to be improved from "almost sure" statements to ones reflecting tail probabilities. We shall denote the augmented lemmas with primes ($'$) after the number. We remark here that the proof in BY (1993) assumes entries of Z_{11} to be real, but all the arguments can be easily modified to allow complex variables.

For Lemma 1 it has been shown that for the Hermitian matrices $T(l)$ defined in (2.2), and integers m_n satisfying $m_n/\ln n \to \infty$, $m_n\delta_n^{1/3}/\ln n \to 0$ and $m_n/(\delta_n\sqrt{n}) \to 0$

$$\mathsf{E}\,\mathrm{tr}\,T^{2m_n}(l) \le n^2\big((2l+1)(l+1)\big)^{2m_n}(p/n)^{m_n(l-1)}\big(1+o(1)\big)^{4m_n l}.$$

[(2.13) of BY (1993)]. Therefore, writing $m_n = k_n \ln n$, for any $\varepsilon > 0$ there exists an $a \in (0,1)$ such that for all n large,

$$\mathsf{P}\big(\mathrm{tr}\,T(l) > (2l+1)(l+1)y^{(l-1)/2} + \varepsilon\big) \le n^2 a^{m_n} = n^{2+k_n\log a} = o(n^{-\ell}) \tag{A.1}$$

for any positive ℓ. We call (A.1) Lemma $1'$.

We next replace Lemma 2 of BY (1993) with the following:

LEMMA $2'$. *Let for every* n $X_1, X_2, \dots, X_n$ *be i.i.d. with* $X_1 = X_1(n) \sim X_{11}(n)$. *Then for any* $\varepsilon > 0$ *and* $\ell > 0$,

$$\mathsf{P}\left(\left|n^{-1}\sum_{i=1}^n |X_i|^2 - 1\right| > \varepsilon\right) = o(n^{-\ell})$$

and for any $f > 1$,

$$\mathsf{P}\left(n^{-f}\sum_{i=1}^n |X_i|^{2f} > \varepsilon\right) = o(n^{-\ell}).$$

PROOF. Since as $n \to \infty$ $\mathsf{E}|X_1|^2 \to 1$,

$$n^{-f}\sum_{i=1}^n \mathsf{E}|X_i|^{2f} \le 2^{2f}\mathsf{E}|Z_{11}|^{2f}n^{1-f} \to 0 \qquad \text{for } f \in (1,2]$$

and

$$n^{-f}\sum_{i=1}^n \mathsf{E}|X_i|^{2f} \le 2^{2f}\mathsf{E}|Z_{11}|^4 n^{1-f+(2f-4)/2} = Kn^{-1} \to 0 \qquad \text{for } f > 2,$$

it is sufficient to show for $f \geq 1$,

$$\mathsf{P}\left(n^{-f}\left|\sum_{i=1}^{n}\left(|X_i|^{2f} - \mathsf{E}|X_i|^{2f}\right)\right| > \varepsilon\right) = o(n^{-\ell}). \tag{A.2}$$

For any positive integer m we have this probability bounded by

$$\begin{aligned}
&n^{-2mf}\varepsilon^{-2m}\mathsf{E}\left[\sum_{i=1}^{n}(|X_i|^{2f} - \mathsf{E}|X_i|^{2f})\right]^{2m} \\
&\quad = n^{-2mf}\varepsilon^{-2m}\sum_{\substack{i_1\geq 0,\ldots,i_n\geq 0\\ i_1+\cdots+i_n=2m}}\binom{2m}{i_1\cdots i_n}\prod_{t=1}^{n}\mathsf{E}(|X_t|^{2f} - \mathsf{E}|X_t|^{2f})^{i_t} \\
&\quad = n^{-2mf}\varepsilon^{-2m}\sum_{k=1}^{m}\binom{n}{k}\sum_{\substack{i_1\geq 2,\ldots,i_k\geq 2\\ i_1+\cdots+i_k=2m}}\binom{2m}{i_1\cdots i_k}\prod_{t=1}^{k}\mathsf{E}(|X_1|^{2f} - \mathsf{E}|X_1|^{2f})^{i_t} \\
&\quad \leq 2^{2m}n^{-2mf}\varepsilon^{-2m}\sum_{k=1}^{m}n^k\sum_{\substack{i_1\geq 2,\ldots,i_k\geq 2\\ i_1+\cdots+i_k=2m}}\binom{2m}{i_1\cdots i_k}\prod_{t=1}^{k}\mathsf{E}|X_1|^{2fi_t} \\
&\quad \leq 2^{2m}n^{-2mf}\varepsilon^{-2m}\sum_{k=1}^{m}n^k\sum_{\substack{i_1\geq 2,\ldots,i_k\geq 2\\ i_1+\cdots+i_k=2m}}\binom{2m}{i_1\cdots i_k}\prod_{t=1}^{k}(2\delta_n\sqrt{n})^{2fi_t-4}\mathsf{E}|Z_{11}|^4 \\
&\quad \leq 2^{2m}n^{-2mf}\varepsilon^{-2m}\sum_{k=1}^{m}k^{2m}(2\delta_n\sqrt{n})^{4fm-4k}n^k(\mathsf{E}|Z_{11}|^4)^k \\
&\quad = 2^{2m}\varepsilon^{-2m}\sum_{k=1}^{m}(2\delta_n)^{4fm}(\mathsf{E}|Z_{11}|^4)^k(4\delta_n^2 n)^{-k}k^{2m} \\
&\quad \leq (\text{for all } n \text{ large}) \quad m\left(\frac{(2\delta_n)^{2f}4m}{\varepsilon\ln(4\delta_n^2 n/\mathsf{E}|Z_{11}|^4)}\right)^{2m},
\end{aligned}$$

where we have used the inequality $a^{-x}x^b \leq (b/\ln a)^b$, valid for all $a > 1$, $b > 0$, $x \geq 1$. Choose $m_n = k_n \ln n$ with $k_n \to \infty$ and $\delta_n^{2f}k_n \to 0$. Since $\delta_n n^{1/3} \geq 1$ for n large we get for these $n\ln(\delta_n^2 n) \geq (1/3)\ln n$. Using this and the fact that $\lim_{x\to\infty}x^{1/x} = 1$, we have the existence of $a \in (0, 1)$ for which

$$m\left(\frac{(2\delta_n)^{2f}4m}{\varepsilon\ln(4\delta_n^2 n/\mathsf{E}|Z_{11}|^4)}\right)^{2m} \leq a^{2k_n\ln n} = n^{2k_n\ln a}$$

for all n large. Therefore (A.2) holds.

Redefining the matrix $X^{(f)}$ in BY (1993) to be $[|X_{uv}|^f]$, Lemma 3′ states for any positive integer f

$$\mathsf{P}\big(\lambda_{\max}\{n^{-f}X^{(f)}X^{(f)*}\} > 7+\varepsilon\big) = o(n^{-\ell}) \qquad \text{for any positive } \varepsilon \text{ and } \ell.$$

Its proof relies on Lemmas 1′, 2′ (for $f = 1, 2$) and on the bounds used in the proof of Lemma 3 in BY (1993). In particular we have the Gerŝgorin bound

$$\begin{aligned}
&\lambda_{\max}\{n^{-f}X^{(f)}X^{(f)*}\}\\
&\quad\le \max_i n^{-f}\sum_{j=1}^n |X_{ij}|^{2f} + \max_i n^{-f}\sum_{k\ne i}\sum_{j=1}^n |X_{ij}|^f|X_{kj}|^f \qquad \text{(A.3)}\\
&\quad\le \max_i n^{-f}\sum_{j=1}^n |X_{ij}|^{2f} + \left(\max_i n^{-1}\sum_{j=1}^n |X_{ij}|^f\right)\left(\max_j n^{-1}\sum_{k=1}^p |X_{kj}|^f\right).
\end{aligned}$$

We show the steps involved for $f = 2$. With $\varepsilon_1 > 0$ satisfying $(p/n+\varepsilon_1)(1+\varepsilon_1) < 7+\varepsilon$ for all n we have from Lemma 2′ and (A.3)

$$\begin{aligned}
&\mathsf{P}\big(\lambda_{\max}\{n^{-2}X^{(2)}X^{(2)*}\} > 7+\varepsilon\big)\\
&\quad\le p\mathsf{P}\left(n^{-2}\sum_{j=1}^n |X_{1j}|^4 > \varepsilon_1\right)\\
&\qquad + p\mathsf{P}\left(n^{-1}\sum_{j=1}^n |X_{1j}|^2 - 1 > \varepsilon_1\right) + n\mathsf{P}\left(p^{-1}\sum_{k=1}^p |X_{k1}|^2 - 1 > \varepsilon_1\right)\\
&\quad = o(n^{-\ell}).
\end{aligned}$$

The same argument can be used to prove Lemma 4′, which states for integer $f > 2$

$$\mathsf{P}\big(\|n^{-f/2}X^{(f)}\| > \varepsilon\big) = o(n^{-\ell}) \qquad \text{for any positive } \varepsilon \text{ and } \ell.$$

The proofs of Lemmas 4′–8′ are handled using the arguments in BY (1993) and those used above: each quantity L_n in BY (1993) that is $o(1)$ a.s. can be shown to satisfy $\mathsf{P}(|L_n| > \varepsilon) = o(n^{-\ell})$.

From Lemmas 1′ and 8′ there exists a positive C such that for every integer $k > 0$ and positive ε and ℓ,

$$\mathsf{P}\big(\|T - yI\|^k > Ck^4 2^k y^{k/2} + \varepsilon\big) = o(n^{-\ell}). \qquad \text{(A.4)}$$

For given $\varepsilon > 0$ let integer $k > 0$ be such that

$$\big|2\sqrt{y}\big(1 - (Ck^4)^{1/k}\big)\big| < \varepsilon/2.$$

Then

$$2\sqrt{y} + \varepsilon > 2\sqrt{y}(Ck^4)^{1/k} + \varepsilon/2 \ge \big(Ck^4 2^k y^{k/2} + (\varepsilon/2)^k\big)^{1/k}.$$

Therefore from (A.4) we get, for any $\ell > 0$,

$$\mathsf{P}\big(\|T - yI\| > 2\sqrt{y} + \varepsilon\big) = o(n^{-\ell}). \tag{A.5}$$

From Lemma $2'$ and (A.5) we get for positive ε and ℓ

$$\begin{aligned}
&\mathsf{P}\big(\|S_n - (1+y)I\| > 2\sqrt{y} + \varepsilon\big)\\
&\quad \le \mathsf{P}(\|S_n - I - T\| > \varepsilon/2) + o(n^{-\ell})\\
&\quad = \mathsf{P}\left(\max_{i \le p}\left|n^{-1}\sum_{j=1}^{n}|X_{ij}|^2 - 1\right| > \varepsilon/2\right) + o(n^{-\ell}) = o(n^{-\ell}).
\end{aligned}$$

Finally, for any positive $\eta < (1 - \sqrt{y})^2$ and $\ell > 0$

$$\begin{aligned}
\mathsf{P}\big(\lambda_{\min}(S_n) < \eta\big) &= \mathsf{P}\big(\lambda_{\min}\big(S_n - (1+y)I\big) < \eta - \big(1 - \sqrt{y}\big)^2 - 2\sqrt{y}\big)\\
&\le \mathsf{P}\big(\big(\|S_n - (1+y)I\|\big) > 2\sqrt{y} + \big(1 - \sqrt{y}\big)^2 - \eta\big) = o(n^{-\ell})
\end{aligned}$$

and we are done. $\square$

Acknowledgments. Part of this work was done while J. W. Silverstein visited the Department of Statistics and Applied Probability at National University of Singapore. He thanks the members of the department for their hospitality.

REFERENCES

BAI, Z. D. (1999). Methodologies in spectral analysis of large dimensional random matrices, A review. *Statist. Sinica* **9** 611–677.

BAI, Z. D. and SARANADASA, H. (1996). Effect of high dimension comparison of significance tests for a high dimensional two sample problem. *Statist. Sinica* **6** 311–329.

BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large dimensional random matrices. *Ann. Probab.* **26** 316–345.

BAI, Z. D. and SILVERSTEIN, J. W. (1999). Exact separation of eigenvalues of large dimensional sample covariance matrices. *Ann. Probab.* **27** 1536–1555.

BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York.

BURKHOLDER, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probab.* **1** 19–42.

DEMPSTER, A. P. (1958). A high dimensional two sample significance test. *Ann. Math. Statist.* **29** 995–1010.

DIACONIS, P. and EVANS, S. N. (2001). Linear functionals of eigenvalues of random matrices. *Trans. Amer. Math. Soc.* **353** 2615–2633.

JOHANSSON, K. (1998). On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.* **91** 151–204.

JONSSON, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix *J. Multivariate Anal.* **12** 1–38.

MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sb.* **1** 457–483.

SILVERSTEIN, J. W. (1985). The limiting eigenvalue distribution of a multivariate F matrix. *SIAM J. Math. Anal.* **16** 641–646.

SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal.* **55** 331–339.

SILVERSTEIN, J. W. and CHOI, S. I. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.* **54** 295–309.

SILVERSTEIN, J. W. and COMBETTES, P. L. (1992). Signal detection via spectral theory of large dimensional random matrices. *IEEE Trans. Signal Process.* **40** 2100–2105.

SINAI, YA. and SOSHNIKOV, A. (1998). Central limit theorem for traces of large symmetric matrices with independent matrix elements. *Bol. Soc. Brasil Mat. (N.S.)* **29** 1–24.

SOSHNIKOV, A. (2000). The central limit theorem for local linear statistics in classical compact groups and related combinatorial identities. *Ann. Probab.* **28** 1353–1370.

TITCHMARSH, E. C. (1939). *The Theory of Functions*, 2nd ed. Oxford Univ. Press.

YIN, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* **20** 50–68.

YIN, Y. Q., BAI, Z. D. and KRISHNAIAH, P. R. (1988). On limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521.

YIN, Y. Q. and KRISHNAIAH, P. R. (1983). A limit theorem for the eigenvalues of product of two random matrices. *J. Multivariate Anal.* **13** 489–507.

DEPARTMENT OF MATHEMATICS
NORTH EAST NORMAL UNIVERSITY
CHANGCHUN
CHINA 130024
E-MAIL: stabaizd@leonis.nus.edu.sg

DEPARTMENT OF MATHEMATICS
BOX 8205
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27695-8205
USA
E-MAIL: jack@math.ncsu.edu