The Annals of Probability
1993, Vol. 21, No. 2, 649–672

CONVERGENCE RATE OF EXPECTED SPECTRAL DISTRIBUTIONS OF LARGE RANDOM MATRICES. PART II. SAMPLE COVARIANCE MATRICES

BY Z. D. BAI

Temple University

In the first part of the paper, we developed certain inequalities to bound the difference between distributions in terms of their Stieltjes transforms and established a convergence rate of expected spectral distributions of large Wigner matrices. The second part is devoted to establishing convergence rates for the sample covariance matrices, for the cases where the ratio of the dimension to the degrees of freedom is bounded away from 1 or close to 1, respectively.

1. Introduction. Basic concepts and literature review in this area have been given in Part I of this paper, and will not be repeated in this part. However, for convenience, a basic inequality needed in the proofs is cited in Section 2. Also, in Section 2, we shall establish some lemmas needed in the proofs of the main theorems. In Section 3, we shall establish the convergence rate for empirical spectral distributions of sample covariance matrices.

Note that the density function of the Marchenko–Pastur law [see (3.2)] is bounded when y, the ratio of the dimension to the degrees of freedom (or sample size), is different from 0 and 1. We may expect to have a similar result as that for Wigner matrices, that is, the order of $O(n^{1/4})$. We prove this result. However, when y is close to one, the density function is no longer bounded. The third term on the right-hand side of (2.12) of Theorem 2.2 in Part I is controlled only by $Av^{1/2}$. This shows that we can only get a rate of the order of $\sqrt{v}$, if we establish similar estimates for the integral of the difference of Stieltjes transforms of the empirical spectral distribution and the limiting spectral distribution. Moreover, its Stieltjes transform and the integral of the absolutely squared Stieltjes transform are not bounded. All these make it more difficult to establish an inequality at an ideal order when y is close to 1. In fact, the rate we actually established in this part is $O(n^{-5/48})$.

2. Preliminaries.

2.1. *A basic inequality from Part I.* We shall use Theorem 2.2 proved in Part I to prove our main results in this part. For reference, this theorem is now stated.

Received December 1990; revised January 1992.
AMS 1991 *subject classifications.* 60F15, 62F15.
Key words and phrases. Berry–Esseen inequality, convergence rate, large dimensional random matrix, Marchenko–Pastur distribution, sample covariance matrix, semicircular law, spectral analysis, Stieltjes transform, Wigner matrix.

THEOREM 2.2 OF PART I. *Let F be a distribution function and let G be a function of bounded variation satisfying $\int |F(x) - G(x)|\,dx < \infty$. Then we have*

$$\|F - G\| \le \frac{1}{\pi(1-\kappa)(2\gamma - 1)}\left[\int_{-A}^{A} |f(z) - g(z)|\,du \right.$$

$$(2.1) \qquad + 2\pi v^{-1}\int_{|x|>B} |F(x) - G(x)|\,dx$$

$$\left. + v^{-1}\sup_x \int_{|y| \ne 2v\tau} |G(x+y) - G(x)|\,dy\right],$$

where γ, κ, τ, A and B are positive constants satisfying $A > B$,

$$\kappa = \frac{4B}{\pi(A-B)(2\gamma - 1)} < 1 \tag{2.2}$$

and

$$\gamma = \frac{1}{\pi}\int_{|u|<\tau} \frac{1}{u^2 + 1}\,du > \frac{1}{2}.$$

2.2. *Stieltjes transforms of Marchenko–Pastur distribution.* Assume that $0 < y < 1$. In a similar manner as we did for the semicircular law in Part I, we obtain

$$s_y(z) = -\frac{1}{4\pi i y}\oint_{|\zeta|=1} \frac{(\zeta^2 - 1)^2\,d\zeta}{\zeta(\zeta^2 + 2\alpha'\zeta + 1)(\zeta^2 + 2\alpha\zeta + 1)},$$

where $\alpha = (1+y)/2\sqrt{y}$ and $\alpha' = \alpha - z/2\sqrt{y}$. The function

$$(\zeta^2 - 1)^2/\zeta(\zeta^2 + 2\alpha'\zeta + 1)(\zeta^2 + 2\alpha\zeta + 1)$$

has five singular points:

$$\zeta_1 = 0, \qquad \zeta_2 = -\sqrt{y}, \qquad \zeta_3 = -1/\sqrt{y}$$

and

$$\zeta_{4,5} = -\Big((1+y-z) \pm \sqrt{(1+y-z)^2 - 4y}\Big)/2\sqrt{y}.$$

It is obvious that $|\zeta_3| > 1$. By the convention (3.1) for the square root of a complex number given in Section 3.1 of Part I, both the real and imaginary parts of $\sqrt{(1+y-z)^2 - 4y}$ have the same signs as those of $z - y - 1$, respectively. Hence, the absolute value of ζ_5 is greater than that of ζ_4. Since $\zeta_4\zeta_5 = 1$, we have $|\zeta_5| > 1$. By elementary calculus, one finds that the residues of $\zeta_{1,2,4}$ are 1, $-(1-y)/z$ and $-\sqrt{(1+y-z)^2 - 4y}\,/z$, respectively. Hence, by the residue theorem, we obtain

$$s_y(z) = -\frac{y + z - 1 - \sqrt{(1+y-z)^2 - 4y}}{2yz}. \tag{2.3}$$

Because Marchenko–Pastur distributions are weakly continuous in y, letting $y \uparrow 1$, we obtain

$$s_1(z) = -\frac{z - \sqrt{z^2 - 4z}}{2z}. \tag{2.4}$$

Similarly, one may prove that (2.3) is still true when $y > 1$.

Now, we shall find bounds for $s_y(z)$ for both cases when $0 < \theta \le y \le \Theta < 1$ and when $\Theta \le y \le 1$, respectively. Note that

$$\left(s_y(z) + \frac{1}{z}\right)\left(s_y^*(z) + \frac{1}{z}\right) = \frac{1}{yz^2}$$

and

$$s_y(z)s_y^*(z) = \frac{1}{yz},$$

where

$$s_y^*(z) = -\frac{2}{y + z - 1 - \sqrt{(1 + y - z)^2 - 4y}}.$$

Recalling the convention for the square root of a complex number given in Part I, we find that the real part of $\sqrt{(1 + y - z)^2 - 4y}$ has the same sign as $u - 1 - y$. We conclude that

$$\left|s_y(z) + \frac{1}{z}\right| \le \left|s_y^*(z) + \frac{1}{z}\right|$$

for all z, and

$$|s_y(z)| \le |s_y^*(z)|$$

for all z such that $u \ge 1 + y$ or $u < 1 - y$. Therefore, we have

$$|s_y(z)| < \frac{1 + \sqrt{y}}{|z|\sqrt{y}}, \quad \text{for all } z \tag{2.5}$$

and

$$|s_y(z)| < \frac{1}{\sqrt{y|z|}}, \quad \text{for } z \text{ with } |u - 1| \ge y. \tag{2.6}$$

For the case of $\theta \le y \le \Theta < 1$, if $|u| \le (1 - y)(1 + \sqrt{y})/(1 + 3\sqrt{y})$, then by the fact that

$$s_y(z) = -\frac{2}{y + z - 1 + \sqrt{(1 + y - z)^2 - 4y}}, \tag{2.7}$$

we obtain

$$
\begin{aligned}
|s_y(z)| &< \frac{2}{1-y-|u|} \\
&\le \frac{1+3\sqrt{y}}{\sqrt{y}(1-y)}.
\end{aligned} \tag{2.8}
$$

From (2.5) it is easy to see that (2.8) is still true if $|u| \geq (1-y)(1+\sqrt{y})/(1+3\sqrt{y})$.

For the case of $\Theta \leq y \leq 1$, if $(1-\sqrt{y})^2 \leq u \leq 1+y$, then $(1+y-u)^2 \leq 4y$. Hence, by (2.7) we obtain

$$
\begin{aligned}
|s_y(z)| &< \frac{2}{\operatorname{Im}\left(\sqrt{(1+y-z)^2-4y}\right)} \\
&= \frac{2\sqrt{2}}{\sqrt{\sqrt{(w^2-v^2-4y)^2+4v^2w^2}-w^2+4y+v^2}} \\
&\le \frac{2\sqrt{2}}{\sqrt[4]{(w^2-4y)^2+4v^2w^2}} \\
&\le \frac{2\sqrt{2}}{\sqrt[4]{8yv^2}} \le \frac{2}{\sqrt{yv}},
\end{aligned} \tag{2.9}
$$

where $w = 1+y-u$. If $u < (1-\sqrt{y})^2$ or $u > 1+y$, the estimate (2.6) is true and hence the inequality (2.9) is still true.

2.3. *Integrals of the square of the absolute value of Stieltjes transforms.* Applying Lemma 3.1 of Part I to Marchenko–Pastur (for $0 < y < 1$) laws, we obtain

$$
\int |s_y(z)|^2 \, du \leq \frac{2\pi}{\sqrt{y}(1-y)}, \tag{2.10}
$$

since the density function has an upper bound $1/(\pi\sqrt{y}(1-y))$.

It should be noted that this bound tends to infinity when y tends to 0 or 1. This is reasonable for the case that $y \to 0$ because the distribution tends to be degenerate. For the case $y \to 1$ or even $y = 1$, the distribution is still continuous, although the density for $y = 1$ is unbounded. One may want to have a finite bound (probably depending on v). We have the following inequality.

LEMMA 2.1. *For any* $y \leq 1$, *we have*

$$\int |s_y(z)|^2\, du \leq \frac{2\pi(1+\sqrt{y})}{y\sqrt{v}}. \tag{2.11}$$

PROOF. Using the notation and going through the same lines of the proof of Lemma 3.1 of Part I, we find that

$$\begin{aligned} I &= 4\pi v \int_a^b \int_a^b \left(\frac{1}{(u-x)^2 + 4v^2} \right) \phi(x)\phi(u)\, dx\, du \\ &= 8\pi v \int_a^b \int_x^b \left(\frac{1}{(u-x)^2 + 4v^2} \right) \phi(x)\phi(u)\, dx\, du \quad \text{(by symmetry)} \\ &\leq 4vy^{-1}\sqrt{b} \int_0^\infty \int_a^b \phi(x) \left(\frac{1}{\sqrt{w}\,(w^2 + 4v^2)} \right) dw\, dx \\ &= 2y^{-1}\sqrt{\frac{2b}{v}} \int_0^\infty \left(\frac{1}{\sqrt{w}\,(w^2+1)} \right) dw \\ &= \frac{2\pi(1+\sqrt{y})}{\sqrt{v}\,y}. \end{aligned}$$

Here we used the fact that

$$\int_0^\infty \left(\frac{1}{\sqrt{w}\,(w^2+1)} \right) dw = \frac{\pi}{\sqrt{2}},$$

which may be computed by using the residue theorem and the equality

$$\int_0^\infty \left(\frac{1}{\sqrt{w}\,(w^2+1)} \right) dw = \frac{1}{1-i} \int_{-\infty}^\infty \left(\frac{1}{\sqrt{z}\,(z^2+1)} \right) dz.$$

This completes the proof of Lemma 2.1. □

LEMMA 2.2. *Let G be a function of bounded variation satisfying $\int |G(u)|\, du < \infty$. Let $g(z)$ denote its Stieltjes transform. When $z = u + iv$ with $v > 0$, we have*

$$\int |g(z)|^2\, du \leq 2\pi v^{-1} V(G) \|G\|, \tag{2.12}$$

where $V(G)$ denotes the total variation of G and $\|G\| = \sup_x(|G(x)|)$.

PROOF. Going through the same lines of the proof of Lemma 3.1 of Part I, we may obtain

$$\begin{aligned} I &= 4\pi v \int\int \left(\frac{1}{(u-x)^2 + 4v^2} \right) dG(x)\, dG(u) \\ &= 8\pi v \int \left[\int \frac{(u-x)G(x)\, dx}{\left((u-x)^2 + 4v^2\right)^2} \right] dG(u) \quad \text{(integration by parts)} \\ &\le 2\pi v^{-1} V(G) \|G\|. \end{aligned}$$

□

2.4. *Lemmas concerning Lévy distance.*

LEMMA 2.3. *Let $L(F,G)$ be the Lévy distance between the distributions F and G. Then we have*

$$L^2(F,G) \le \int |F(x) - G(x)|\, dx. \tag{2.13}$$

PROOF. Without loss of generality, assume that $L(F,G) > 0$. For any $r \in (0, L(F,G))$, there exists an x such that

$$F(x-r) - r > G(x) \quad [\text{or } F(x+r) + r < G(x)].$$

Then the square between the points $(x-r, F(x-r)-r)$, $(x, F(x-r)-r)$, $(x-r, F(x-r))$ and $(x, F(x-r))$ [or $(x, F(x+r))$, $(x+r, F(x+r))$, $(x, F(x+r)+r)$ and $(x+r, F(x+r)+r)$ for the latter case] is located between F and G (see Figure 1a, b). Then (2.13) follows from the fact that the right-hand side of (2.13) equals the area of the region between F and G. The proof is complete. □

LEMMA 2.4. *If G satisfies $\sup_x |G(x+y) - G(x)| \le D|y|$ for all y, then one may prove that*

$$L(F,G) \le \|F - G\| \le (D+1)L(F,G), \quad \textit{for all } F. \tag{2.14}$$

PROOF. The inequality on the left-hand side is actually true for all distributions F and G. It can follow easily from the argument in the proof of Lemma 2.3. To prove the right-hand-side inequality, let us consider the case where for some x,

$$F(x) > G(x) + \rho,$$

where $\rho \in (0, \|F - G\|)$. Since G satisfies the Lipschitz condition, we have

$$G(x + \rho/(D+1)) + \rho/(D+1) \le G(x) + \rho < F(x),$$

which implies that

$$L(F,G) \ge \frac{\rho}{D+1}.$$

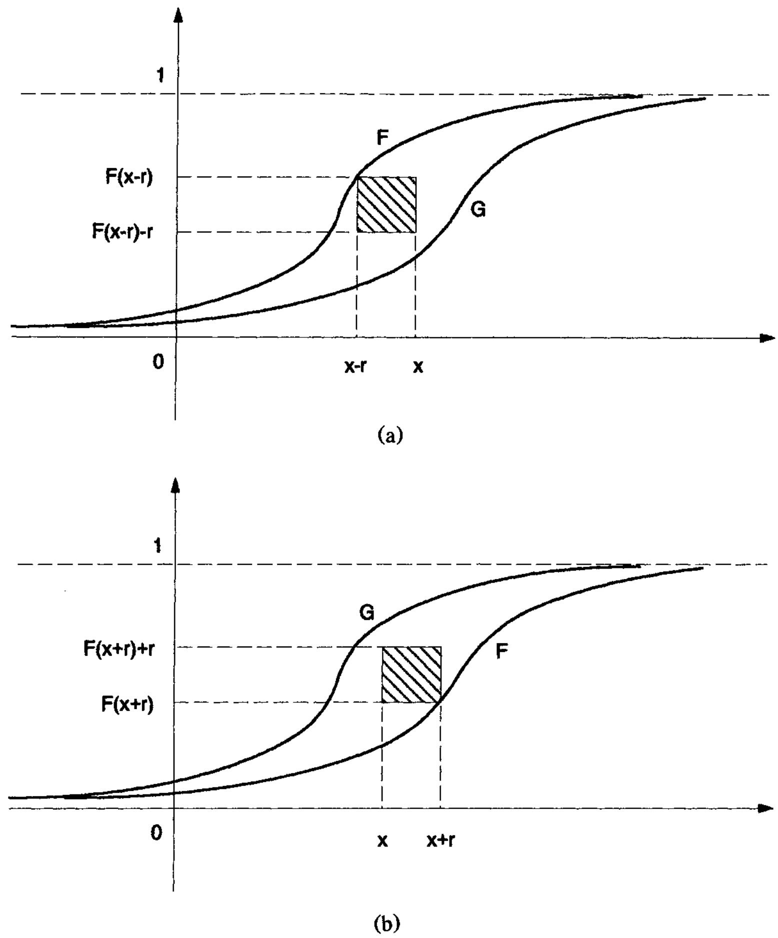

FIG. 1.

Then the right-hand-side inequality of (2.14) follows by making $\rho \to \|F - G\|$. The inequality for the other case, that is, $G(x) > F(x) + \rho$, can be similarly proved. □

LEMMA 2.5. *Let F_1, F_2 be distribution functions and let G satisfy $\sup_x|G(x+u) - G(x)| \le D|u|^\beta$, for all u and some $\beta \in (0,1]$. Then*

$$\begin{aligned}\|F_1 - G\|^{1+1/\beta} &\le 2\|F_1 - G\|^{1/\beta}\|F_2 - G\| \\ &\quad + 2(2D)^{1/\beta}\int |F_1(x) - F_2(x)|\,dx.\end{aligned} \tag{2.15}$$

PROOF. Let $0 < \rho < \|F_1 - G\|$. Then, we may find an x_0 such that $F_1(x_0) - G(x_0) > \rho$ [or $F_1(x_0) - G(x_0) < -\rho$, alternatively]. By the condition on G, for any $x \in [x_0 - (\rho/2D)^{1/\beta}, x_0]$ (or $[x_0, x_0 + (\rho/2D)^{1/\beta}]$ for the other case), we have $|F_1(x) - G(x)| \geq (1/2)\rho$. But for any x in this interval, we have

$$\tfrac{1}{2}\rho \leq |F_1(x) - G(x)| \leq \|F_2 - G\| + |F_1(x) - F_2(x)|.$$

Integrating the above inequality over this interval and then making $\rho \to \|F_1 - G\|$, we obtain (2.15). The proof is complete. □

3. Convergence rates of spectral distributions of sample covariance matrices. Let $W_p = (w_{ij}(n)) = n^{-1}X_pX_p'$: $p \times p$, where $X_p = (x_{ij}(n)$, $i = 1, \dots, p$, $j = 1, \dots, n)$. Throughout this section, we shall drop the index n from the entries of X_p and those of W_p and assume that X_{ij}'s are independent and the following conditions hold:

$$\text{(3.1)} \qquad \begin{aligned} &\text{(i)} \quad Ex_{ij} = 0, \qquad Ex_{ij}^2 = 1, \quad \text{for all } i, j; \\ &\text{(ii)} \quad \sup_n \sup_{i,j} Ex_{ij}^4 I_{[|x_{ij}| \geq M]} \to 0, \quad \text{as } M \to \infty. \end{aligned}$$

The matrix W_p is known as a sample covariance matrix. It should be noted here that the notation W_p is no longer the same as used in the last section. Denote by F_p its empirical spectral distribution. Under the conditions in (3.1), it is well known that $F_p \to_w F_y$ in probability, where $y = \lim_{n\to\infty}(p/n) \in (0, \infty)$ and F_y is the limiting spectral distribution of F_p, known as the Marchenko–Pastur (1967) distribution, which may have a mass of $1 - y^{-1}$ at the origin if $y > 1$ and has a density

$$\text{(3.2)} \qquad F_y'(x) = \frac{1}{2xy\pi}\sqrt{4y - (x - y - 1)^2}\, I_{[a,b]}(x),$$

with $a = a(y) = (1 - \sqrt{y})^2$, and $b = b(y) = (1 + \sqrt{y})^2$. If X_p is the $p \times n$ submatrix of the upper-left corner of an infinitely dimensional random matrix $[x_{ij},\ i, j = 1, 2, \dots]$, then the above convergence is true a.s. (almost surely) [e.g., Wachter (1978) who actually proved the a.s. version of the convergence under the uniform boundedness of the $(2 + \varepsilon)$th moments of the entries of X_p].

To consider the rate of the convergence, we shall establish the following theorem.

THEOREM 3.1. *Under the assumptions in* (3.1), *we have*

$$\text{(3.3)} \qquad \|EF_p - F_{y_p}\| = O(n^{-1/4}),$$

for $y_p = p/n \in (\theta, \Theta)$, *where* $0 < \theta < \Theta < 1$ *or* $1 < \theta < \Theta < \infty$.

THEOREM 3.2. *If* $0 < \theta \leq y \leq \Theta < \infty$, *then under the assumptions in* (3.1) *we have*

$$\text{(3.4)} \qquad \|EF_p - F_{y_p}\| = O(n^{-5/48}),$$

where $\theta < 1 < \Theta$.

REMARK 3.1. Because the convergence rate of $|y_p - y|$ can be arbitrarily slow, it is impossible to establish any rate for the convergence of $\|EF_p - F_y\|$ if we know nothing about the convergence rate of $|y_p - y|$. Conversely, if we know the convergence rate of $|y_p - y|$, then from (3.3) or (3.4), we can easily derive a convergence rate for $\|EF_p - F_y\|$. This is the reason why F_{y_p}, instead of the limit distribution F_y, is used in Theorem 3.1.

For simplicity, we shall drop the index p from y_p. We need only to prove the theorems for the cases where $\theta \le y \le \Theta < 1$ and $\Theta \le y \le 1$. For the case of $y > 1$, since the first n largest eigenvalues of the matrix $X_p X_p'$ are identical with those of $X_p' X_p$, the theorem then follows by considering the analogous result for the matrix $(1/p)X_p' X_p$. For simplicity, these two cases are referred to as $y \le \Theta$ and $y \le 1$, respectively.

To begin with, we shall establish two propositions applicable to both cases.

PROPOSITION 3.3. *Under condition* (3.1) *and the following additional assumption*:

$$|x_{ij}| < \sqrt{n}\,\eta, \quad \textit{with } \eta \to 0, \tag{3.5}$$

we have

$$\int_B^\infty |EF_p(x) - F_y(x)|\,dx = o(n^{-t}), \tag{3.6}$$

where $B = b + 1$, $b = b(y)$ *is defined below* (3.2) *and the constant* $t > 0$ *is fixed but can be arbitrarily large.*

REMARK 3.2. For any rate of $\eta = \eta_p \to 0$, Proposition 3.3 is true. In application of this proposition, the choice of η in (3.5) involves a truncation and normalization procedure described at the end of this section [see (3.62)].

PROOF. In Yin, Bai and Krishnaiah (1988), after truncation and normalization, it is actually proved (their arguments still work even without the assumption of identical distributions of the entries of X_p) that,

$$E(\lambda_p)^m \le (b + \zeta)^m, \tag{3.7}$$

for some sequence of constants $\zeta = \zeta_p \to 0$ and some sequence of integers $m = m_p$ such that $\log n/m \to 0$, under condition (i) of (3.1), (3.5) and

$$E|x_{ij}|^k \le dn^{(k-3)/2}, \quad \text{for all } k \ge 3, \tag{3.8}$$

where λ_p denotes the largest eigenvalue of $n^{-1}X_p X_p'$ and b is defined in (3.2). Note that (3.8) is implied by (3.5) and (ii) of (3.1).

Note that

$$1 - F_p(x) \le I_{[\lambda_p \ge x]}, \quad \text{for } x \ge 0. \tag{3.9}$$

We have

$$\int_B^{\infty}|EF_p(x) - F(x)|\,dx \le \int_B^{\infty} P(\lambda_p \ge x)\,dx$$

$$\le \int_B^{\infty}\left(\frac{b+\zeta}{x}\right)^m dx \tag{3.10}$$

$$= O\left(\left(\frac{b+\zeta}{B}\right)^{m-1}\right) = o(n^{-t}),$$

for any fixed $t > 0$. Proposition 3.3 is proved. □

Theorem 3.1 can be proved via Theorem 2.2. From the above results and the fact that $F_p(x) = F(x) = 0$ for all $x < 0$, we need only estimate $s_p(z) - s_y(z)$ for $z = u + iv$, $v > 0$, $|u| < A$, where A is a constant chosen according to (2.2).

From (2.3), the Stieltjes transform of the limiting spectral distribution F_y is given by

$$s_y(z) = -\frac{1}{2yz}\left\{z + y - 1 - \sqrt{(z+y-1)^2 - 4yz}\right\}. \tag{3.11}$$

Set

$$s_p(z) = \frac{1}{p} E\,\mathrm{tr}(W_p - zI_p)^{-1}. \tag{3.12}$$

PROPOSITION 3.4. *Choose* $v = (10C_0(A+1)/n)^{1/6}$, *where* C_0 *is a constant which will be specified in* (3.32). *Then, if* (3.1) *and* (3.5) *hold, we have that*

$$\int_{-A}^{A}|s_p(z) - s_y(z)|\,du \le Cv, \tag{3.13}$$

where C *is a positive constant.*

PROOF. By the inverse matrix formula [see (3.8) in Part I], we have

$$s_p(z) = \frac{1}{p}\sum_{k=1}^{p} E\frac{1}{w_{kk} - z - \alpha'(k)(W_p(k) - zI_{n-1})^{-1}\alpha(k)}$$

$$= \frac{1}{p}\sum_{k=1}^{p} E\frac{1}{\varepsilon_k + 1 - y - z - yzs_p(z)} \tag{3.14}$$

$$= -\frac{1}{z + y - 1 + yzs_p(z)} + \delta,$$

where $W_p(k)$ is the matrix obtained from W_p by deleting the kth row and kth column, $\alpha(k)$ denotes the vector obtained from the kth column of W_p by removing the kth element, and

$$\varepsilon_k = \frac{1}{n}\sum_{j=1}^{n}\left(x_{kj}^2 - 1\right) + y + yzs_p(z) - \alpha'(k)\left(W_p(k) - zI_{p-1}\right)^{-1}\alpha(k),$$

$$\begin{aligned}\delta = \delta_p &= -\frac{1}{p}\sum_{k=1}^{p} E\frac{\varepsilon_k}{\left(z + y - 1 + yzs_p(z)\right)\left(z + y - 1 + yzs_p(z) - \varepsilon_k\right)} \\ &= -\frac{1}{p}\sum_{k=1}^{p} E\frac{\varepsilon_k}{\left(z + y - 1 + yzs_p(z)\right)^2} \\ &\quad -\frac{1}{p}\sum_{k=1}^{p} E\frac{\varepsilon_k^2}{\left(z + y - 1 + yzs_p(z)\right)^2\left(z + y - 1 + yzs_p(z) - \varepsilon_k\right)}.\end{aligned} \tag{3.15}$$

Solving $s_p(z)$ from (3.14), we get two roots

$$s_{(1),(2)}(z) = -\frac{1}{2yz}\left(z + y - 1 - yz\delta \pm \sqrt{(z + y - 1 + yz\delta)^2 - 4yz}\right). \tag{3.16}$$

Comparing (3.16) with (3.11), it seems that the Stieltjes transform $s_p(z)$ should be the solution $s_{(2)}(z)$, for all values of z with $v > 0$, that is,

$$s_p(z) = -\frac{1}{2yz}\left(z + y - 1 - yz\delta - \sqrt{(z + y - 1 + yz\delta)^2 - 4yz}\right). \tag{3.17}$$

Now, we prove (3.17). First, we note that

$$\begin{aligned}\operatorname{Im}\left(z + y - 1 + yzs_p(z)\right) &= \operatorname{Im}\left(z - 1 + y\int_a^b \frac{x\,dF_p(x)}{x - z}\right) \\ &= v\left(1 + y\int_a^b \frac{x\,dF_p(x)}{(x - u)^2 + v^2}\right) > v.\end{aligned} \tag{3.18}$$

It follows immediately from (3.18) that

$$\frac{1}{\left|z + y - 1 + yzs_p(z)\right|} \le v^{-1}. \tag{3.19}$$

It is obvious that $|s_p(z)| \le v^{-1}$. Therefore,

$$|\delta| \le 2/v. \tag{3.20}$$

For any fixed u, when $v \to \infty$, we have $s_p(z) \to 0$, $s_{(2)}(z) \to 0$ and $s_{(1)}(z) \to -1/y$. This shows that (3.17) is true for all large v.

As in Part I, one may easily see that both $s_{(1)}(z)$ and $s_{(2)}(z)$ are continuous functions on the upper-half plane. Thus, to prove that (3.17) is true for all z on the upper-half plane, it is sufficient to show that $s_{(1)}(z) \ne s_{(2)}(z)$ for all z on the upper-half plane. Otherwise, there would be some z on the upper-half plane such that $s_p(z) = s_{(1)}(z) = s_{(2)}(z)$. Then the square root term in (3.16) would be zero. Hence, we have

$$s_p(z) = -\frac{y + z - 1 - yz\delta}{2yz}.$$

Substituting the expression derived for δ from (3.14) into the above expression, we obtain

$$s_p(z) = \frac{1 - y - z}{yz} + \frac{1}{y + z - 1 + yzs_p(z)}.$$

However, this is impossible, since the imaginary parts of the two terms are obviously negative [for the second term, see (3.18)]. This contradiction proves that (3.17) is true for all z on the upper-half plane.

Comparing (3.17) with (3.11), we need to show that both δ and the integral of the absolute value of δ with respect to u on a finite interval are "small." Then, we begin to find a bound for $|s_p(z) - s_y(z)|$ in terms of δ.

We now proceed to estimate $|\delta|$. First, we note that, by (3.11) of Part I,

$$\frac{1}{\left|z + y - 1 + yzs_p(z) - \varepsilon_k\right|} \le v^{-1}. \tag{3.21}$$

Then, by (3.15) and (3.21), we have

$$|\delta| \le \frac{1}{p}\sum_{k=1}^{p}\left(|E\varepsilon_k| + v^{-1}E|\varepsilon_k^2|\right)\left(\left|z + y - 1 + yzs_p(z)\right|\right)^{-2}. \tag{3.22}$$

Denote by $X_p(k)$ the $(p-1)\times n$ matrix obtained from X_p by eliminating the kth row, and denote by $x(k)$ the n vector of the kth row of X_p. Then, $\alpha(k) = (1/n)X_p(k)x(k)$ and $W_p(k) = (1/n)X_p(k)X_p'(k)$. Recalling the definition of ε_k, one finds that

$$\begin{aligned} &E^{(k)}\alpha'(k)\left(W_p(k) - zI_{p-1}\right)^{-1}\alpha(k)\\ &\quad= n^{-2}E^{(k)}x'(k)X_p'(k)\left(W_p(k) - zI_{p-1}\right)^{-1}X_p(k)x(k)\\ &\quad= n^{-2}\operatorname{tr} X_p'(k)\left(W_p(k) - zI_{p-1}\right)^{-1}X_p(k)\\ &\quad= n^{-1}\operatorname{tr}\left(W_p(k) - zI_{p-1}\right)^{-1}W_p(k)\\ &\quad= y - \frac{1}{n} + zn^{-1}\operatorname{tr}\left(W_p(k) - zI_{p-1}\right)^{-1}, \end{aligned} \tag{3.23}$$

where $E^{(k)}$ denotes the conditional expectation given $\{x_{ij},\ i \ne k\}$. Then by (3.10) of Part I, for all z with $|u| \le A$ we have

$$\begin{aligned} |E\varepsilon_k| &= \frac{|z|}{n}\left|E\left[\operatorname{tr}\left(W_p(k) - zI_{p-1}\right)^{-1} - \operatorname{tr}\left(W_p - zI_p\right)^{-1}\right]\right| + \frac{1}{n}\\ &\le \frac{C}{nv}, \end{aligned} \tag{3.24}$$

where the constant C may take the value $A + 1$.

Next, we proceed to estimate $E|\varepsilon_k|^2$. We have

$$E|\varepsilon_k^2| \le \frac{M}{n} + R_1 + R_2 + |E(\varepsilon_k)|^2, \tag{3.25}$$

where

$$R_1 = E\Big|\alpha'(k)\big(W_p(k) - zI_{p-1}\big)^{-1}\alpha(k) - E^{(k)}\Big(\alpha'(k)\big(W_p(k) - zI_{p-1}\big)^{-1}\alpha(k)\Big)\Big|^2,$$

$$R_2 = \frac{|z|^2}{n}E\Big|\operatorname{tr}\big(W_p(k) - zI_{p-1}\big)^{-1} - E\operatorname{tr}\big(W_p(k) - zI_{p-1}\big)^{-1}\Big|^2$$

and M is the constant in (3.1).

Let

$$\Gamma_k = \big(r_{ij}(k)\big) = \frac{1}{n}X_p'(k)\big(W_p(k) - zI_{p-1}\big)^{-1}X_p(k).$$

Then, we have

$$\begin{aligned} R_1 &\le \frac{2M}{n^2}\sum_{i,j}E\big|r_{ij}^2(k)\big| = \frac{2M}{n^2}E\operatorname{tr}\big(\Gamma_k\bar{\Gamma}_k\big) \\ &= \frac{2M}{n^2}E\operatorname{tr}\Big[\big(W_p(k) - zI_{p-1}\big)^{-1}W_p(k) \\ &\qquad\times\big(W_p(k) - \bar{z}I_{p-1}\big)^{-1}W_p(k)\Big] \\ &= \frac{2M}{n^2}E\operatorname{tr}W_p^2(k)\Big(\big(W_p(k) - uI_{p-1}\big)^2 + v^2I_{p-1}\Big)^{-1} \\ &\le \frac{4M}{n^2}E\operatorname{tr}\Big[\big(W_p(k) - uI_{p-1}\big)^2 + u^2I_{p-1}\Big] \\ &\qquad\times\Big(\big(W_p(k) - uI_{p-1}\big)^2 + v^2I_{p-1}\Big)^{-1} \\ &\le \frac{4M}{n^2}\Big[p - 1 + |u|^2E\operatorname{tr}\Big(\big(W_p(k) - uI_{p-1}\big)^2 \qquad (3.26)\\ &\qquad + v^2I_{p-1}\Big)^{-1}\Big] \\ &\le 4Mn^{-1} + 4MA^2n^{-1}v^{-2} \le Cn^{-1}v^{-2}. \qquad (3.27)\end{aligned}$$

Here, the constant C can be taken as $4M(A^2 + 1)$.

Define $\gamma_d(d) = 0$, and define for $d \ne k$,

$$\begin{aligned}\gamma_d(k) &= E_{d-1}\operatorname{tr}\big(W_p(k) - zI_{p-1}\big)^{-1} - E_d\operatorname{tr}\big(W_p(k) - zI_{p-1}\big)^{-1} \\ &= E_{d-1}\sigma_d(k) - E_d\sigma_d(k), \qquad d = 1, 2, \ldots, p,\end{aligned} \tag{3.28}$$

where

$$\sigma_d(k) = \operatorname{tr}\big(W_p(k) - zI_{p-1}\big)^{-1} - \operatorname{tr}\big(W(d,k) - zI_{p-2}\big)^{-1},$$

$W(d,k)$ is the matrix obtained from W_p by deleting the dth and kth rows and the dth and kth columns, $\alpha(d,k)$ is the vector obtained from the dth column vector of W_p by deleting the dth and kth elements and E_d denotes the conditional expectation given $\{x_{ij},\ d+1 \le i \le p,\ 1 \le j \le n\}$.

Again, by (3.11) of Part I, we have

$$|\sigma_d(k)| \le v^{-1}. \tag{3.29}$$

Therefore, we obtain

$$R_2 \le \frac{(A+1)^2}{n^2} \sum_{d=1}^{p} E|\gamma_d^2(k)| \le Cn^{-1}v^{-2}. \tag{3.30}$$

Then by the definition of ε_k and (3.24)–(3.30), we obtain that for some positive constant C,

$$E|\varepsilon_k|^2 \le \frac{C}{nv^2}. \tag{3.31}$$

Throughout the paper, the letter C denotes a generic positive constant which may take different values at different places. From (3.19), (3.22), (3.24) and (3.31), it follows that for some positive constant C_0,

$$|\delta| \le \frac{C_0}{nv^5}. \tag{3.32}$$

Choose $v = (10C_0(A+1)/n)^{1/6}$. By (3.31), we know that

$$|\delta| \le \frac{v}{10(A+1)^2}. \tag{3.33}$$

By (3.22) and (3.33), for large n, we have

$$\begin{aligned}
\int_{-A}^{A} |\delta|\,du &\le \frac{C}{nv^3}\int_{-A}^{A} |z+y-1+yzs_n(z)|^{-2}\,du \\
&\le \frac{C}{nv^3}\left[\int_{A}^{A} |s_n(z)|^2\,du + \int_{-A}^{A} |\delta|^2\,du\right] \\
&\le \frac{C}{nv^3}\left[\int_{A}^{A} |s_n(z)|^2\,du + v\int_{-A}^{A} |\delta|\,du\right] \\
&\le \frac{C}{nv^3}\int_{-A}^{A} |s_n(z)|^2\,du \\
&\le \frac{C}{nv^3}\int_{-A}^{A}\left[\int_{-\infty}^{\infty} \frac{1}{(x-u)^2+v^2}\,du\right] dF_n(x) \\
&\le \frac{C}{nv^4} \le Cv^2.
\end{aligned} \tag{3.34}$$

Here, in the derivation of the fourth inequality, we have used the fact that

$$a \le c + ba \Rightarrow c + ba \le c/(1-b) \tag{3.35}$$

for any positive a, c and $b < 1$.

Now, we are in position to estimate $|s_p(z) - s_y(z)|$. By (3.11) and (3.17), we have

$$\begin{aligned} &|s_p(z) - s_y(z)| \\ (3.36) \qquad &\le \left|\frac{\delta}{2}\right|\left[1 + \frac{|2(z+y-1) - yz\delta|}{\left|\sqrt{(z+y-1)^2 - 4yz} + \sqrt{(z+y-1+yz\delta)^2 - 4yz}\right|}\right]. \end{aligned}$$

As done in Part I, we need to find the condition guaranteeing that the real parts of $\sqrt{(z+y-1)^2 - 4yz}$ and $\sqrt{(z+y-1+yz\delta)^2 - 4yz}$ have the same sign. We claim this is true for $|u - y - 1| > y/[2(A+1)]$. In fact, if $|u - y - 1| > y/[2(A+1)]$, then (3.33) implies that

$$v + \mathrm{Im}(yz\delta) > 0,$$
$$|u - y - 1| > |\mathrm{Re}(yz\delta)|$$

and

$$\begin{aligned} &|(u - y - 1 - \mathrm{Re}(yz\delta))(v + \mathrm{Im}(yz\delta))| - 2y^2|\mathrm{Im}(z\delta)| \\ &\quad > \left(\frac{y}{2(A+1)} - \frac{vy}{10(A+1)}\right)\left(v - \frac{v}{10(A+1)}\right) - \frac{vy^2}{10(A+1)} > 0. \end{aligned}$$

From the above estimates it follows that the sign of the real part of $\sqrt{(z+y-1+yz\delta)^2 - 4yz}$ is

$$\begin{aligned} &\mathrm{sign}[2((u+y-1+\mathrm{Re}(yz\delta))(v+\mathrm{Im}(yz\delta)) - 4yv] \\ (3.37) \qquad &= \mathrm{sign}((u-y-1-\mathrm{Re}(yz\delta))(v+\mathrm{Im}(yz\delta)) + 2y\,\mathrm{Im}(yz\delta)) \\ &= \mathrm{sign}(u-y-1). \end{aligned}$$

Since the sign of the real part of $\sqrt{(z+y-1)^2 - 4yz}$ is

$$\mathrm{sign}(2v(u-y-1)) = \mathrm{sign}(u-y-1),$$

for $|u - y - 1| > y/[2(A+1)]$, by (3.37), the real parts of both $\sqrt{(z+y-1)^2 - 4yz}$ and $\sqrt{(z+y-1+yz\delta)^2 - 4yz}$ have a common sign. Hence for large p, (3.36) implies that

$$(3.38) \qquad |s_p(z) - s_y(z)| \le \frac{1}{2}|\delta|\left[1 + \frac{2A+2}{\sqrt{|(u-y-1)^2 - v^2 - 4y|}}\right].$$

If $|u - y - 1| \le y/[2(A+1)]$, then for all large p, we have

$$\begin{aligned} \left|\sqrt{(z-y-1)^2 - 4y} - 2i\sqrt{y}\right| &= \frac{|z-y-1|^2}{\left|\sqrt{(z-y-1)^2 - 4y} + 2i\sqrt{y}\right|} \\ &\le \frac{1}{2\sqrt{y}}\left[\frac{y^2}{4(A+1)^2} + v^2\right] \\ &\le \frac{1}{2}y \end{aligned}$$

and

$$\begin{aligned}
&\left|\sqrt{(z+y-1+yz\delta)^2-4yz}-2i\sqrt{y}\right| \\
&\quad = \frac{\left|(z-y-1+yz\delta)^2+4y^2z\delta\right|}{\left|\sqrt{(z+y-1+yz\delta)^2-4yz}+2i\sqrt{y}\right|} \\
&\quad \le \frac{1}{\sqrt{y}}\left[(u-y-1)^2+|v+yz\delta|^2+4y^2|z\delta|\right] \\
&\quad \le \frac{1}{\sqrt{y}}\left[\frac{y^2}{4(A+1)^2}+v^2\left(1+\frac{1}{10(A+1)}\right)^2+\frac{4y^2v}{10(A+1)}\right] \\
&\quad \le \frac{1}{2}y.
\end{aligned}$$

Therefore, for $|u-y-1| \le y/[2(A+1)]$, we have

$$\left|\sqrt{(z+y-1)^2-4yz}+\sqrt{(z+y-1+yz\delta)^2-4yz}\right| \ge 4\sqrt{y}-y > 2\sqrt{y}.$$

Combining the above and (3.38) together, for $|u| \le A$, we have

$$|s_n(z)-s_y(z)| \le \begin{cases} \frac{1}{2}|\delta|\left[1+\dfrac{2A+2}{\sqrt{\left|(u+y-1)^2-v^2-4yu\right|}}\right], & \\ \qquad\qquad \text{if } |u-y-1| > \dfrac{y}{2(A+1)}, \\ C_1|\delta|, \qquad \text{if } |u-y-1| \le \dfrac{y}{2(A+1)}, \end{cases} \tag{3.39}$$

where $C_1 = C_1(y)$ is a positive constant depending upon y; for example, here we may take $C = (A+2)/2\sqrt{y}$.

By (3.34) and (3.39), one finally gets

$$\begin{aligned}
&\int_{-A}^{A}|s_n(z)-s_y(z)|\,du \\
&\quad = \left\{\int_{[|u-y-1|>y/[2(A+1)],\,|u|\le A]} + \int_{[|u-y-1|\le y/[2(A+1)],\,|u|\le A]}\right\}|s_n(z)-s_y(z)|\,du \qquad (3.40)\\
&\quad \le Cv\int_{-A}^{A}\frac{1}{\sqrt{\left|(u+y-1)^2-v^2-4yu\right|}}\,du + C\int_{-A}^{A}|\delta|\,du \\
&\quad \le \eta v + Cv^2,
\end{aligned}$$

where

$$\eta = C \sup_{v<1} \int_{-A}^{A} \frac{du}{\sqrt{|(u+y-1)^2 - v^2 - 4yu|}}.$$

The proof of Proposition 3.4 is complete. □

PROOF OF THEOREM 3.1. Since the density function of the Marchenko–Pastur distribution is bounded when $\theta \le y \le \Theta < 1$, applying Theorem 2.2 and Proposition 3.4, we obtain a preliminary bound

$$\|EF_p - F_y\| = O(n^{-1/6}), \tag{3.41}$$

under the additional restriction of (3.5). Next, we shall improve the result as we did in Part I.

Assume that $\|EF_p - F_y\| \le \Delta_1 = \eta n^{-1/6}$, for some $\eta > 1$. We shall refine the estimates of $\sum_{ij} E|r_{ij}^2(k)|$ and $\sum_d E|\gamma_d^2(k)|$. Assume that $\Delta_1 \ge v \ge n^{-1/4}$. The exact value of v will be chosen later. Noticing that $y \le \Theta$, by (3.26), we get

$$\begin{aligned}
R_1 &\le \frac{4M(p-1)}{n^2}\left[1 + E\int_0^\infty \frac{(u^2 - v^2)\, dF_{p-1}^{(k)}(x)}{(x-u)^2 + v^2}\right] \\
&\le \frac{4MA^2}{n^2}\left[ny + E\int_0^\infty \frac{(p-1)\, dF_{p-1}^{(k)}(x)}{(x-u)^2 + v^2}\right] \\
&= \frac{4MA^2}{n^2}\left[ny + \int_0^\infty \frac{p\, dF_y(x)}{(x-u)^2 + v^2} + E\int_0^\infty \frac{p\, d\big(F_p(x) - F_y(x)\big)}{(x-u)^2 + v^2}\right. \\
&\qquad \left. + E\int_0^\infty \frac{d\big((p-1)F_{p-1}^{(k)}(x) - pF_p(x)\big)}{(x-u)^2 + v^2}\right] \qquad (3.42)\\
&\le \frac{4MA^2}{n^2}\left[ny + \frac{n\sqrt{y}}{v(1-y)} + E\int_0^\infty \frac{2p(x-u)\big(F_p(x) - F_y(x)\big)\, dx}{\big((x-u)^2 + v^2\big)^2}\right. \\
&\qquad \left. + E\int_0^\infty \frac{2(x-u)\big((p-1)F_{p-1}^{(k)}(x) - pF_p(x)\big)\, dx}{\big((x-u)^2 + v^2\big)^2}\right] \\
&\le \frac{4MA^2}{n^2}\left[ny + \frac{n\sqrt{y}}{v(1-y)} + 2ny\,\Delta_1 v^{-2} + 2v^{-2}\right] \\
&\le Cn^{-1}\Delta_1 v^{-2}. \qquad (3.43)
\end{aligned}$$

Here, in the derivation of the third inequality, the first integral in (3.42) was estimated by the upper bound $1/(\pi\sqrt{y}(1-y))$ of the density of F_y, the second by $\|F_p - F_y\| \le \Delta_1$ and the third by the fact that $|(p-1)F_{p-1}^{(k)}(x) - pF_p(x)| \le 1$, established in Lemma 3.3 in Part I.

Now, we estimate $E|\gamma_d(k)|^2$. Rewrite $\sigma_d(k)$ as

$$\begin{aligned}
\sigma_d(k) &= \frac{1 + \alpha'(d,k)\big(W(d,k) - zI_{p-2}\big)^{-2}\alpha(d,k)}{w_{dd} - z - \alpha'(d,k)\big(W(d,k) - zI_{p-2}\big)^{-1}\alpha(d,k)} \\
&= \frac{1 + (1/n)\operatorname{tr}\Gamma^{(2)}}{1 - z - (1/n)\operatorname{tr}\Gamma^{(1)}} + \frac{(1/n)x_d'\Gamma^{(2)}x_d - (1/n)\operatorname{tr}\Gamma^{(2)}}{1 - z - (1/n)\operatorname{tr}\Gamma^{(1)}} \\
&\quad + \frac{\big(1 + (1/n)x_d'\Gamma^{(2)}x_d\big)\big((1/n)x_d'\Gamma^{(1)}x_d - (1/n)\operatorname{tr}\Gamma^{(1)} - w_{dd} + 1\big)}{\big(1 - z - (1/n)\operatorname{tr}\Gamma^{(1)}\big)\big(w_{dd} - z - (1/n)x_d'\Gamma^{(1)}x_d\big)} \\
&:= \sigma_d'(k) + \sigma_d''(k) + \sigma_d'''(k),
\end{aligned}$$

where

$$\Gamma^{(1)} = \frac{1}{n}X'(d,k)\big(W(d,k) - zI_{p-2}\big)^{-1}X(d,k),$$

$$\Gamma^{(2)} = \frac{1}{n}X'(d,k)\big(W(d,k) - zI_{p-2}\big)^{-2}X(d,k),$$

$X(d,k)$ is the $(p-2)\times n$ matrix obtained from X_p by eliminating its dth and kth rows, and x_d is the n vector of the dth row of X_p. It is easy to see that

$$E_{d-1}\sigma_d'(k) - E_d\sigma_d'(k) = 0. \tag{3.44}$$

Similarly to the proof of (3.18), we may prove that

$$\left|\operatorname{Im}\left(1 - z - \frac{1}{n}\operatorname{tr}\Gamma^{(1)}\right)\right| > v.$$

We may also derive that

$$\begin{aligned}
\operatorname{tr}\Gamma^{(1)}\overline{\Gamma}^{(1)} &= \operatorname{tr}W^2(d,k)\Big(\big(W(d,k) - uI_{p-2}\big)^2 + v^2I_{p-2}\Big)^{-1} \\
&\le 2\operatorname{tr}\Big(\big(W(d,k) - uI_{p-2}\big)^2 + u^2I_{p-2}\Big) \\
&\quad\times\Big(\big(W(d,k) - uI_{p-2}\big)^2 + v^2I_{p-2}\Big)^{-1} \\
&\le 2\operatorname{tr}\Big[I_{p-2} + A^2\Big(\big(W(d,k) - uI_{p-2}\big)^2 + v^2I_{p-2}\Big)^{-1}\Big] \\
&\le C(p-2)v^{-2} < Cnv^{-2}
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{tr}\Gamma^{(2)}\overline{\Gamma}^{(2)} &= \operatorname{tr}W^2(d,k)\Big(\big(W(d,k) - uI_{p-2}\big)^2 + v^2I_{p-2}\Big)^{-2} \\
&\le v^{-2}\operatorname{tr}W^2(d,k)\Big(\big(W(d,k) - uI_{p-2}\big)^2 + v^2I_{p-2}\Big)^{-1} \\
&\le Cnv^{-4}.
\end{aligned}$$

Thus, we have

$$
\begin{aligned}
E\left|E_{d-1}\sigma_d''(k) - E_d\sigma_d''(k)\right|^2 &\le E\left|1 - z - \frac{1}{n}\operatorname{tr}\Gamma^{(1)}\right|^{-2}\left|\frac{1}{n}\left(x_d'\Gamma^{(2)}x_d - \operatorname{tr}\Gamma^{(2)}\right)\right|^2 \\
&\le Cn^{-2}E\left|1 - z - \frac{1}{n}\operatorname{tr}\Gamma^{(1)}\right|^{-2}\operatorname{tr}\Gamma^{(2)}\bar{\Gamma}^{(2)} \qquad (3.45)\\
&\le Cn^{-1}v^{-4}E\left|1 - z - \frac{1}{n}\operatorname{tr}\Gamma^{(1)}\right|^{-2}
\end{aligned}
$$

and by (3.29),

$$
\begin{aligned}
E\left|E_{d-1}\sigma_d'''(k) - E_d\sigma_d'''(k)\right|^2 &\le v^{-2}E\left|1 - z - \frac{1}{n}\operatorname{tr}\Gamma^{(1)}\right|^{-2}\left|\frac{1}{n}x_d'\Gamma^{(1)}x_d - \operatorname{tr}\Gamma^{(1)}\right|^2 \\
&\le n^{-2}v^{-2}E\left|1 - z - \frac{1}{n}\operatorname{tr}\Gamma^{(1)}\right|^{-2}\operatorname{tr}\Gamma^{(1)}\bar{\Gamma}^{(1)} \qquad (3.46)\\
&\le Cn^{-1}v^{-4}E\left|1 - z - \frac{1}{n}\operatorname{tr}\Gamma^{(1)}\right|^{-2}.
\end{aligned}
$$

Summing up (3.44), (3.45) and (3.46), we obtain

$$
E\left|\gamma_d(k)\right|^2 \le Cn^{-1}v^{-4}E\left|1 - z - \frac{1}{n}\operatorname{tr}\Gamma^{(1)}\right|^{-2}. \qquad (3.47)
$$

By (3.10) of Part I, we have

$$
\left|\operatorname{tr}\left(W_p(k) - zI_{p-1}\right)^{-1} - \operatorname{tr}\left(W_p - zI_p\right)^{-1}\right| < v^{-1} \qquad (3.48)
$$

and

$$
\left|\operatorname{tr}\left(W_p(k) - zI_{p-1}\right)^{-1} - \operatorname{tr}\left(W(d,k) - zI_{p-2}\right)^{-1}\right| < v^{-1}. \qquad (3.49)
$$

Thus, by (3.48) and (3.49), we have

$$
\begin{aligned}
\operatorname{tr}(\Gamma^{(1)}) &= p - 2 + z\operatorname{tr}\left(\left(W(d,k) - zI_{p-2}\right)^{-1}\right. \qquad (3.50)\\
&= ny + z\operatorname{tr}\left(\left(W_p - zI_p\right)^{-1}\right) + zR(d,k),
\end{aligned}
$$

with $|R(d,k)| < 4/v$. Consequently, we obtain that

$$
\begin{aligned}
E\left|\gamma_d(k)\right|^2 &\le Cn^{-1}v^{-4}E\left|1 - z - y - zn^{-1}\operatorname{tr}\left(W_p - zI_p\right)^{-1} + zn^{-1}R(d,k)\right|^{-2} \\
&= Cn^{-1}v^{-4}E\left|1 - z - y - zn^{-1}\operatorname{tr}\left(W_p - zI_p\right)^{-1}\right|^{-2} \\
&\quad\times\left|1 + \frac{zn^{-1}R(d,k)}{1 - z - (1/n)\operatorname{tr}\Gamma^{(1)}}\right|^2 \qquad (3.51)\\
&\le Cn^{-1}v^{-4}E\left|1 - z - y - zn^{-1}\operatorname{tr}\left(W_p - zI_p\right)^{-1}\right|^{-2}\left|1 + \frac{4(A+1)}{nv^2}\right|^2 \\
&\le Cn^{-1}v^{-4}E\left|1 - z - y - zn^{-1}\operatorname{tr}\left(W_p - zI_p\right)^{-1}\right|^{-2},
\end{aligned}
$$

for some positive constant C. Therefore, by (3.30) we obtain

$$R_2 \le Cn^{-2}v^{-4}E\left|1 - z - y - zn^{-1}\operatorname{tr}(W_p - zI_p)^{-1}\right|^{-2}. \tag{3.52}$$

Similarly to the proof of (3.30), we may prove that

$$E\left|\frac{1}{n}\operatorname{tr}(W_p - zI_p)^{-1} - ys_p(z)\right|^2 \le \frac{1}{n^2}\sum_{k=1}^{p} E|\gamma_k|^2 \le \frac{1}{nv^2}, \tag{3.53}$$

where

$$\begin{aligned}\gamma_k &= E_{k-1}\operatorname{tr}(W_p - zI_p)^{-1} - E_k\operatorname{tr}(W_p - zI_p)^{-1}\\ &= E_{k-1}\operatorname{tr}(\sigma_k) - E_k\operatorname{tr}(\sigma_k)\end{aligned}$$

and

$$\sigma_k = \frac{1 + \alpha_k'(W_p(k) - zI_{p-1})^{-2}\alpha_k}{\varepsilon_k + 1 - y - z - yzs_p(z)}$$

with $|\sigma_k| \le v^{-1}$. Therefore, from (3.51)–(3.53), we get

$$\begin{aligned}R_2 &\le Cn^{-2}v^{-4}\left|1 - z - y - yzs_p(z)\right|^{-2}\\ &\quad\times E\left|1 + \frac{zn^{-1}\operatorname{tr}(W_p - zI_p)^{-1} - yzs_p(z)}{1 - y - z - zn^{-1}\operatorname{tr}(W_p - zI_p)^{-1}}\right|^2\\ &\le Cn^{-2}v^{-4}\left|1 - z - y - yzs_p(z)\right|^{-2}\\ &\quad\times\left(1 + v^{-2}E\left|zn^{-1}\operatorname{tr}(W_p - zI_p)^{-1} - yzs_p(z)\right|^2\right)\\ &\le Cn^{-2}v^{-4}\left|1 - z - y - yzs_p(z)\right|^{-2}(1 + n^{-1}v^{-4})\\ &\le Cn^{-2}v^{-4}\left|1 - z - y - yzs_p(z)\right|^{-2},\end{aligned} \tag{3.54}$$

for some positive constant C.

By (3.14), we have

$$\left|1 - z - y - yzs_p(z)\right|^{-2} \le 3\left(|\delta|^2 + \left|s_p(z) - s_y(z)\right|^2 + \left|s_y(z)\right|^2\right).$$

Noting the bounds of $s_y(z)$ established in (2.8), we have

$$\begin{aligned}\left|1 - z - y - yzs_p(z)\right|^{-2} &\le C\left(|\delta|^2 + \frac{\pi^2\Delta_1^2}{v^2} + \frac{(1 + 3\sqrt{y})^2}{y(1 - y)^2}\right)\\ &\le C\left(|\delta|^2 + \frac{\Delta_1^2}{v^2}\right).\end{aligned} \tag{3.55}$$

Here we used the fact that $|s_p(z) - s_y(z)| \le \pi\Delta_1/v$, which can be easily obtained by integration by parts.

Substitute (3.43), (3.54) and (3.55) into (3.25) first and then substitute the result into (3.22). We obtain

$$|\delta| \le C\left[\frac{1}{nv} + \frac{\Delta_1}{nv^3} + \frac{1}{n^2v^5}\left|1-y-z-yzs_p(z)\right|^{-2} + \frac{1}{n^2v^2}\right]$$

$$\times\left|1-y-z-yzs_p(z)\right|^{-2}$$

$$\le C\left[\frac{\Delta_1}{nv^3} + \frac{1}{n^2v^5}\left|1-y-z-yzs_p(z)\right|^{-2}\right]\left|1-y-z-yzs_p(z)\right|^{-2} \tag{3.56}$$

$$\le C_0\left[\frac{\Delta_1}{nv^3} + \frac{1}{n^2v^5}|\delta|^2\right]\left[|\delta|^2 + \frac{\Delta_1^2}{v^2}\right]. \tag{3.57}$$

Choose $v = (40C_0\eta^3(A+1)^2)^{1/6}n^{-1/4}$. Then, by (3.57) and the fact that $|\delta| \le 2/v$ [see (3.20)], we obtain, for all large p,

$$|\delta| \le C_0\left[\frac{8}{n^2v^8} + \frac{2\Delta_1}{nv^4} + \frac{2\Delta_1^2}{n^2v^8}\right]|\delta| + \frac{C_0\Delta_1^3}{nv^5}$$

$$\le \frac{1}{2}|\delta| + \frac{C_0\Delta_1^3}{nv^5} \le \frac{2C_0\Delta_1^3}{nv^5} \tag{3.58}$$

$$\le \frac{v}{10(A+1)^2}.$$

Then, by (3.55), (3.56) and (3.58), we have

$$|\delta| \le C\left(\frac{\Delta_1}{nv^3} + \frac{1}{n^2v^5}\left(|\delta|^2 + \frac{\Delta^2}{v^2}\right)\right)\left|1-y-z-yzs_p(z)\right|^{-2}$$

$$\le C\left(\frac{\Delta_1}{nv^3}\left|1-y-z-yzs_p(z)\right|^{-2}\right)$$

$$\le \frac{C\Delta_1}{nv^3}\left[|\delta|^2 + \left|s_p(z)-s_y(z)\right|^2 + \left|s_y(z)\right|^2\right] \tag{3.59}$$

$$\le \frac{C\Delta_1}{nv^3}\left[\left|s_p(z)-s_y(z)\right|^2 + \left|s_y(z)\right|^2\right]$$

$$\le C\frac{\Delta_1}{nv^3}\left[\left|s_p(z)-s_y(z)\right|^2 + \left|s_y(z)\right|^2\right].$$

Hence, by Lemma 2.2 and (2.10),

$$\int_{-\infty}^{\infty}|\delta|\,du \le \frac{C\Delta_1}{nv^3}\left[\int_{-\infty}^{\infty}\left|s_p(z)-s_y(z)\right|^2du + \int_{-\infty}^{\infty}\left|s_y(z)\right|^2du\right]$$

$$\le \frac{C\Delta_1}{nv^3}\left[\frac{\Delta_1}{v}+1\right] \le Cv. \tag{3.60}$$

Recall that (3.39) holds for $|u| \le A$ provided that (3.33) is true. Therefore, by (3.58) we conclude that (3.39) is true for the newly chosen v and $|u| \le A$.

By (3.60), repeating the procedure of (3.40), we may prove that

$$\|EF_p - F_y\| = O(n^{-1/4}),$$

under the additional assumption (3.5).

To finish the proof of Theorem 3.1, we drop the restriction (3.5). Define

$$\hat{x}_{ij} = \sigma_{ij}^{-1}\left[x_{ij}I_{[|x_{ij}| \le \sqrt{n}\eta]} - Ex_{ij}I_{[|x_{ij}| \le \sqrt{n}\eta]}\right], \tag{3.61}$$

where σ_{ij}^2 is the variance of $x_{ij}I_{[|x_{ij}| \le \eta\sqrt{n}]}$, η is chosen such that $\eta \to 0$ and

$$\sup_{ij} E|x_{ij}^4|I_{[|x_{ij}| \ge \eta\sqrt{n}]} = o(\eta^2). \tag{3.62}$$

By the second condition in (3.1), it is easy to select η fulfilling the above condition.

Let $\hat{F}_p$ denote the spectral distribution of $(1/n)\hat{X}_p\hat{X}_p'$ with $\hat{X}_p = (\hat{x}_{ij})$. Then, by what we have proved under the restriction (3.5), we have

$$\|E\hat{F}_p - F_y\| = O(n^{-1/4}). \tag{3.63}$$

Note that, when $\theta \le y \le \Theta$, the density function of the Marchenko–Pastur distribution has an upper bound $D = 1/(\sqrt{y}(1-y))$. Applying Lemma 2.4 and the triangular inequality, we have

$$\begin{aligned}\|EF_p - F_y\| &\le (D+1)L(EF_p, F_y)\\ &\le (D+1)\left[L\left(EF_p, E\hat{F}_p\right) + L\left(E\hat{F}_p, F_y\right)\right]\\ &\le (D+1)\left[L\left(EF_p, E\hat{F}_p\right) + \|E\hat{F}_p - F_y\|\right],\end{aligned} \tag{3.64}$$

where $L(\cdot, \cdot)$ denotes the Lévy distance between distribution functions.

Denote the eigenvalues of the matrices $(1/n)X_pX_p'$ and $(1/n)\hat{X}_p\hat{X}_p'$ by $\lambda_1 \le \cdots \le \lambda_p$ and $\hat{\lambda}_1 \le \cdots \le \hat{\lambda}_p$, respectively. By Lemma 2.3, we have

$$L^2\left(EF_p, E\hat{F}_p\right) \le E\int\left|F_p(x) - \hat{F}_p(x)\right|dx = \frac{1}{p}\sum_{s=1}^{p} E|\lambda_s - \hat{\lambda}_s|. \tag{3.65}$$

Following the approach of Yin (1986), we prove that

$$\begin{aligned}L^2\left(EF_p, E\hat{F}_p\right) &\le \frac{1}{p}\sum_{s=1}^{p} E|\lambda_s - \hat{\lambda}_s|\\ &\le \left(\frac{1}{np}E\operatorname{tr}\left(X_p + \hat{X}_p\right)\left(X_p + \hat{X}_p\right)'\right)^{1/2}\\ &\quad\times\left(\frac{1}{np}E\operatorname{tr}\left(X_p - \hat{X}_p\right)\left(X_p - \hat{X}_p\right)'\right)^{1/2}\\ &\le C\left(\sup_{ij} Ex_{ij}^2 I_{[|x_{ij}| \ge \eta\sqrt{n}]}\right)^{1/2} = o(n^{-1/2}).\end{aligned} \tag{3.66}$$

Then (3.3) follows from (3.63), (3.64) and (3.66), and the proof of Theorem 3.1 is complete. □

PROOF OF THEOREM 3.2. Now, we consider the case of $0 < \Theta < y \le 1$. When applying Theorem 2.2 of Part I to this case, the reader should note that the density function is no longer bounded and hence the third term on the right-hand side of (2.1) does not have the order of $O(v)$. Therefore, we cannot get a preliminary estimate as good as (3.41), although the estimate of Proposition 3.4 is still true. However, we may obtain an estimate as follows:

$$\sup_x \int_{|u| \le 2va} |F_y(x+u) - F_y(x)|\, du \le Cv^{3/2}, \tag{3.67}$$

where C may be chosen as $\sqrt{2a}(1+\sqrt{y})/(\pi y)$ and the constant a is defined in Theorem 2.2 of Part I. By this and Proposition 3.4, applying Theorem 2.2 of Part I, we obtain the following preliminary estimate:

$$\|EF_p - F_y\| = O(n^{-1/12}). \tag{3.68}$$

Now, based on (3.68), we get an improved estimate by refining the estimates of $\Sigma_{ij} E|r_{ij}^2(k)|$ and $\Sigma_d E|\gamma_d^2(k)|$. Assume that $\|EF_p - F_y\| \le \Delta_1 = \eta_0 n^{-1/12}$ for some $\eta_0 > 1$ and assume that $\Delta_1 \ge v > n^{-5/24}$.

Corresponding to (3.43), applying Lemma 2.2 to the first integral in (3.42) [note that (3.42) is true for both the two cases], we find that (3.43) is still true for the newly defined Δ_1 and v, that is,

$$\begin{aligned} \sum_{i,j} E|r_{ij}^2(k)| &\le 2A^2\left[ny + \frac{\sqrt{2}n(1+\sqrt{y})}{\sqrt{v^3}} + 2ny\,\Delta_1 v^{-2} + 2v^{-2}\right] \\ &\le Cn\,\Delta_1 v^{-2}. \end{aligned} \tag{3.69}$$

We now refine the estimate of $\Sigma_d E|\gamma_d^2(k)|$. Using the same notation defined in the proof of Theorem 3.1 and checking the proof of formulae (3.44)–(3.54), we find that they are still true for the present case. Corresponding to (3.55), applying (2.9), we obtain

$$|1 - z - y - yzs_p(z)|^{-2} \le C\left(|\delta|^2 + \frac{\pi^2\Delta_1^2}{v^2} + \frac{4}{vy}\right) \le C\left(|\delta|^2 + \frac{\Delta_1^2}{v^2}\right). \tag{3.70}$$

This means that (3.55) is still formally true for the newly defined Δ_1 and v. Consequently, the expressions (3.56) and (3.57) are still formally true.

Choose $v = (40C_0\eta_0^3(A+1)^2)^{1/6} n^{-5/24}$. Corresponding to (3.58), for all large p, we may directly obtain from (3.57) that

$$\begin{aligned} |\delta| &\le 2C_0 n^{-1} v^{-3} \Delta_1 [2v^{-1}|\delta| + v^{-2}\Delta_1^2] \\ &\le 4C_0 n^{-1} v^{-5} \Delta_1^3 = \frac{v}{10(A+1)^2}. \end{aligned} \tag{3.71}$$

By (3.56) we may similarly prove that (3.59) is true for the present case. Hence, by Lemma 2.2 and (2.11),

$$\begin{aligned} \int_{-\infty}^{\infty} |\delta|\, du &\le \frac{C\Delta_1}{nv^3}\left[\int_{-\infty}^{\infty} |s_p(z) - s_y(z)|^2\, du + \int_{-\infty}^{\infty} |s_y(z)|^2\, du\right] \\ &\le \frac{C\Delta_1}{nv^3}\left[\frac{\Delta_1}{v} + \frac{1}{\sqrt{v}}\right] \le Cv. \end{aligned} \tag{3.72}$$

By (3.71) and (3.72), repeating the procedure of (3.40), one may prove that

$$\int_{-A}^{A} |s_p(z) - s_y(z)|\, du \le Cv.$$

Then applying Theorem 2.2 of Part I and (3.67), we obtain that

$$\|EF_p - F_y\| = O(n^{-5/48}),$$

under the additional assumption (3.5).

As done in the proof of Theorem 3.1, make the truncation and normalization for the entries of X_p in the same way. Use the same notation defined in the proof of Theorem 3.1. By what we have proved, we have

$$\|E\hat{F}_p - F_y\| = O(n^{-5/48}). \tag{3.73}$$

In the proof of Theorem 3.1 [see (3.65)], we have proved that

$$\int |EF_p(x) - E\hat{F}_p(x)|\, dx = o(n^{-1/2}). \tag{3.74}$$

Note that F_y satisfies the condition of Lemma 2.5 with $\beta = 1/2$ and $D = (1 + \sqrt{y})/(\pi y)$. Applying Lemma 2.5 and by (3.73) and (3.74), we obtain

$$\|EF_p - F_y\|^3 \le O(n^{-5/48})\|EF_p - F_y\|^2 + o(n^{-1/2}), \tag{3.75}$$

which implies (3.4). The proof of Theorem 3.2 is complete. □

Acknowledgment. The author would like to thank Professor J. W. Silverstein again for pointing out that the density of the Marchenko–Pastur distribution when $y = 1$ is unbounded, which led to the establishment of Theorem 3.2.

REFERENCES

Bai, Z. D. (1993). Convergence rate of expected spectral distributions of large-dimensional random matrices: Part I. Wigner matrices. *Ann. Probab.* **21** 625–648.

Marchenko, V. A. and Pastur, L. A. (1967). The distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* **72** 507–536 (*Math. USSR-Sb.* **1** 457–483).

Wachter, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* **6** 1–18.

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 the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521.

Department of Statistics
341 Speakman Hall
Temple University
Philadelphia, Pennsylvania 19122