Delft University of Technology

# Log determinant of large correlation matrices under infinite fourth moment

Heiny, Johannes; Parolya, N.

**DOI**
10.1214/23-AIHP1368

**Publication date**
2024

**Document Version**
Accepted author manuscript

**Published in**
Annales de l'Institut Henri Poincar. (B) Probabilites et Statistiques

**Citation (APA)**
Heiny, J., & Parolya, N. (2024). Log determinant of large correlation matrices under infinite fourth moment. *Annales de l'Institut Henri Poincar. (B) Probabilites et Statistiques*, *60*(2), 1048-1076. https://doi.org/10.1214/23-AIHP1368

**Important note**
To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**
Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

*This work is downloaded from Delft University of Technology.*
*For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.*

Submitted to the Annales de l'Institut Henri Poincaré - Probabilités et Statistiques

# Log determinant of large correlation matrices under infinite fourth moment

Johannes Heiny[a] and Nestor Parolya[b]

[a]*Department of Mathematics, Ruhr University Bochum, E-mail:* johannes.heiny@rub.de
[b]*Delft Institute of Applied Mathematics, Delft Universtity of Technology. E-mail:* n.parolya@tudelft.nl

**Abstract.** In this paper, we show the central limit theorem for the logarithmic determinant of the sample correlation matrix $\mathbf{R}$ constructed from the $(p \times n)$-dimensional data matrix $\mathbf{X}$ containing independent and identically distributed random entries with mean zero, variance one and infinite fourth moments. Precisely, we show that for $p/n \to \gamma \in (0,1)$ as $n, p \to \infty$ the *logarithmic law*

$$\frac{\log\det \mathbf{R} - (p-n+\frac{1}{2})\log(1-p/n) + p - p/n}{\sqrt{-2\log(1-p/n) - 2p/n}} \xrightarrow{d} N(0,1)$$

is still valid if the entries of the data matrix $\mathbf{X}$ follow a symmetric distribution with a regularly varying tail of index $\alpha \in (3,4)$. The latter assumptions seem to be crucial, which is justified by the simulations: if the entries of $\mathbf{X}$ have the infinite absolute third moment and/or their distribution is not symmetric, the logarithmic law is not valid anymore. The derived results highlight that the logarithmic determinant of the sample correlation matrix is a very stable and flexible statistic for heavy-tailed big data and open a novel way of analysis of high-dimensional random matrices with self-normalized entries.

**Résumé.** Dans cet article nous démontrons le théorème de la limite centrale pour le déterminant logarithmique d'une matrice de corrélation $\mathbf{R}$ construite d'une matrice de données $\mathbf{X}$ de taille $(p \times n)$ contenante les entrées avec l'espérance 0, la variance 1 et le quatrième moment infini. Plus précisément, nous démontrons que dans le régime $p/n \to \gamma \in (0,1)$ quand $n, p \to \infty$ la loi logarithmique

$$\frac{\log\det \mathbf{R} - (p-n+\frac{1}{2})\log(1-p/n) + p - p/n}{\sqrt{-2\log(1-p/n) - 2p/n}} \xrightarrow{d} N(0,1)$$

est toujours valable si les entrées de la matrice de données $\mathbf{X}$ suivent une distribution symétrique avec une queue à variation régulière d'indice $\alpha \in (3,4)$. Ces dernières conditions semblent être cruciales, ce qui est justifié par les simulations : si les entrées de $\mathbf{X}$ n'ont pas de troisième moment et/ou si leur distribution n'est pas symétrique, la loi logarithmique n'est plus valable. Les résultats obtenus mettent en évidence que le déterminant logarithmique d'une matrice de corrélation est une statistique très stable et flexible pour les données massives à queue lourde et ouvrent une nouvelle voie pour analyser les grandes matrices aléatoires d'entrées auto-normalisées.

*MSC2020 subject classifications:* Primary 60B20; secondary 60F05 60G10 60G57 60G70
*Keywords:* sample correlation matrix, logarithmic determinant, random matrix theory, heavy tails, infinite fourth moment

## 1. Introduction

The analysis of the logarithmic determinant has always been of considerable interest in the large dimensional random matrix theory. The investigations of the moments of random determinants trace back to the 1950s (see, Dembo [11] and references therein). The central limit theorems (CLTs) for the logarithmic determinant of random Gaussian matrices, Wigner matrices and matrices with real independent and identically distributed (i.i.d.) entries with sub-exponential tails were proved by Goodman [21], Tao and Vu [37] and Nguyen and Vu [33], respectively. Girko [19] was the first to state that the result of Goodman [21] holds for general random matrices under the additional assumption that the fourth moment of the entries is equal to three (normal-like moments of order four). This CLT was named as Girko's logarithmic law or simply *logarithmic law*. Moreover, twenty years later Girko [20] using an elegant method of perpendiculars partially proved that the CLT for the logarithmic determinant holds in a very generic case under the existence of the $4+\varepsilon$ moments for some small $\varepsilon > 0$. Nguyen and Vu [33] provided a refined and more transparent proof of this claim assuming a much stronger condition of sub-exponential tails for the random matrix entries, analyzing additionally the rate of convergence of the logarithmic determinant of the sample covariance matrix. Both papers [20] and [33] rely on Girko's method of perpendiculars, whose starting point is the elementary fact that the magnitude of the determinant of a Gram matrix built of real vectors is equal to the volume of the parallelepiped spanned by those vectors. That is why by a simple "base times height" formula one can represent the determinant as a product of perpendiculars (see Bao et al. [8, p. 1602 ff.] for details). In cases where the determinant can be written as a product of independent random variables, for example independent beta distributions in the case of Gaussian matrix entries, the large/moderate deviation results are proved in [22], whereas fast Berry–Esseen bounds were recently provided by [25].

Consider a random sample $\mathbf{x}_1 \ldots, \mathbf{x}_n$ from a $p$-dimensional distribution collected into a $p \times n$ random data matrix $\mathbf{X}$. For statistical applications the logarithmic determinants of the sample covariance matrix $\mathbf{S} = n^{-1}\mathbf{X}\mathbf{X}^\top$ and the sample correlation matrix $\mathbf{R} = \{\mathrm{diag}(\mathbf{S})\}^{-1/2}\,\mathbf{S}\{\mathrm{diag}(\mathbf{S})\}^{-1/2}$ are of vital importance. They allow efficient inferential procedures on the structure of the true covariance/correlation matrices (see, the monographs of Anderson [3] and Yao, Zheng and Bai [42]). In particular, the determinant of the sample correlation matrix has numerous applications in stochastic geometry as it is proportional to the volume of the hyperellipsoid constructed from standardized vectors, see [34]. Furthermore, the determinant of $\mathbf{R}$ is the well-known likelihood ratio statistic for testing the independence of the elements of the random vector in case of multivariate normality of the columns of the data matrix, see, e.g., [10, 12, 13, 29] and references therein.

A wide variety of results have been obtained for the large dimensional sample covariance matrix $\mathbf{S}$, e.g., Marčenko–Pastur law/equation in [30, 36], CLT for linear spectral statistics in [6] and Tracy-Widom law in [14], to mention a few. For the sample correlation matrix $\mathbf{R}$, the situation gets more complicated because of the specific nonlinear dependence structure caused by the normalization $\{\mathrm{diag}(\mathbf{S})\}^{-1/2}$, which makes the analysis of this random matrix quite challenging. In case the elements of the data matrix $\mathbf{X}$ are i.i.d. with zero mean, variance equal to one and finite fourth moment it is shown by Jiang [28] (see, also [5],[15] and [26]) that the Marčenko–Pastur law is still valid for the sample correlation matrix $\mathbf{R}$. The asymptotic distribution of the largest eigenvalue of $\mathbf{R}$ is proved by [7] to obey the Tracy-Widom law. Moreover, the largest and smallest eigenvalues of $\mathbf{R}$ converge to the edges of the Marčenko–Pastur density almost surely [26]. Thus, the "first order" properties (almost sure convergence) of the eigenvalues of the sample covariance matrix $\mathbf{S}$ and sample correlation matrix $\mathbf{R}$ coincide in case the entries of the data matrix $\mathbf{X}$ possess at least finite second moments (see [27]). This observation changes if "second order" properties (such as CLTs) are of interest. To illustrate this fact, we compare the CLTs for the logarithmic determinants of $\mathbf{S}$ and $\mathbf{R}$ under finite fourth moment assumption.

The logarithmic law of the large sample covariance matrices can be deduced from the work of Bai and Silverstein [6] for the linear spectral statistics $\mathrm{tr}(f(\mathbf{S}))$ with a test function $f(x) = \log(x)$ in case $p$ the number of columns of the data matrix is smaller than $n$ the number of its rows and both tend to infinity such that their ratio tends to a constant, i.e., $p/n \to \gamma \in (0,1)$, as $n \to \infty$. More precisely, Wang and Yao [38] showed that if the i.i.d. entries of the data matrix $\mathbf{X} = (X_{ij})_{1\le i\le p; 1\le j\le n}$ satisfy $\mathbb{E}(X_{11}) = 0$, $\mathrm{Var}(X_{11}) = 1$ and $\mathbb{E}(X_{11}^4) < \infty$, the following logarithmic law for its corresponding sample covariance matrix $\mathbf{S}$ is valid

$$\frac{\log\det\mathbf{S} - (p-n+1/2)\log(1-p/n) + p - \frac{1}{2}\left[\mathbb{E}(X_{11}^4) - 3\right]p/n}{\sqrt{-2\log(1-p/n) + [\mathbb{E}(X_{11}^4) - 3]\,p/n}} \xrightarrow{d} N(0,1)\,, \quad \text{as } n \to \infty\,. \tag{1.1}$$

Later on, Bao, Pan and Wang [8] and Wang, Han and Pan [39] proved a similar CLT for the logarithmic determinant of the sample covariance matrices in case $p/n \to 1$ and $p \le n$ under finite fourth moments.

For the sample correlation matrix $\mathbf{R}$ the situation is more involved. The first generic result for the linear spectral statistics of $\mathrm{tr}(f(\mathbf{R}))$ for some test function $f(\cdot)$ was proved in [18] under existence of the fourth moment and it states that taking $f(x) = \log(x)$ for $p/n \to \gamma < 1$ one gets

$$\frac{\log\det\mathbf{R} - (p-n+\frac{1}{2})\log(1-p/n) + p - p/n}{\sqrt{-2\log(1-p/n) - 2p/n}} \xrightarrow{d} N(0,1)\,, \quad \text{as } n \to \infty\,. \tag{1.2}$$

Surprisingly, the latter logarithmic law is quite different from (1.1), especially the dependence on the fourth moment is not present in (1.2), which indicates that the fourth moment assumption can be eventually weakened (see also [35] and [41]).

In this paper, we contribute to the existing literature by showing that the logarithmic law (1.2) is valid for the sample correlation matrix even if the fourth moment of the entries of the data matrix $\mathbf{X}$ is infinite. To the best of our knowledge, this is the first result of this kind. We assume that the i.i.d. elements $X_{ij}$ of $\mathbf{X}$ possess regularly varying tails with index $\alpha \in (3,4)$ and $X_{ij} \stackrel{d}{=} -X_{ij}$ (symmetry). In particular, this implies that $\mathbb{E}X_{11}^4 = \infty$ and $\mathbb{E}|X_{11}|^3 < \infty$. Our proof relies on Girko's method of perpendiculars and a CLT for martingale differences together with the exact computation and asymptotics of the moments of the products of self-normalized variables.

The paper has the following structure: Section 2 contains notations, assumptions and the main result. In Section 3, more precisely in Theorem 3.3, we derive an exact formula for the fourth moment of a weighted sum of the components a random vector on the unit sphere. The latter result is of independent interest and can be considered as a first step to generalization of the key lemma for quadratic forms for correlated random vectors of unit length in case of an infinite fourth moment (c.f. [18, Lemma 5] and [32, Lemma 1]). Asymptotic formulas for the moments of self-normalized variables and the proof of the main theorem are presented in Section 4, while the appendix contains some additional auxiliary results.

## 2. Main result

Consider a $p$-dimensional population $\mathbf{x} = (X_1, \dots, X_p) \in \mathbb{R}^p$ where the coordinates $X_i$ are i.i.d. non-degenerated random variables with mean zero. For a sample $\mathbf{x}_1, \dots, \mathbf{x}_n$ from the population we construct the data matrix $\mathbf{X} = \mathbf{X}_n = (\mathbf{x}_1, \dots, \mathbf{x}_n) = (X_{ij})_{1\le i\le p; 1\le j\le n}$, the sample covariance matrix $\mathbf{S} = \mathbf{S}_n = n^{-1}\mathbf{X}\mathbf{X}^\top$ and the sample correlation matrix $\mathbf{R}$,

$$\mathbf{R} = \mathbf{R}_n = \{\mathrm{diag}(\mathbf{S}_n)\}^{-1/2}\,\mathbf{S}_n\{\mathrm{diag}(\mathbf{S}_n)\}^{-1/2} = \mathbf{Y}\mathbf{Y}^\top\,. \tag{2.1}$$

Here the standardized matrix $\mathbf{Y} = \mathbf{Y}_n = (Y_{ij})_{1\le i\le p; 1\le j\le n}$ for the sample correlation matrix has entries

$$Y_{ij} = Y_{ij}^{(n)} = \frac{X_{ij}}{\sqrt{X_{i1}^2 + \cdots + X_{in}^2}}\,, \tag{2.2}$$

which depend on $n$. Throughout the paper, we often suppress the dependence on $n$ in our notation. We consider the asymptotic regime

$$p = p_n \to \infty \quad \text{and} \quad \frac{p}{n} \to \gamma \in (0,1)\,, \quad \text{as } n \to \infty\,. \tag{$C_\gamma$}$$

We assume that $|X_{11}|$ has a regularly varying tail with index $\alpha > 0$, that is

$$\mathbb{P}(|X_{11}| > x) = L(x)\,x^{-\alpha}\,, \qquad x > 0\,, \tag{2.3}$$

for a function $L$ that is slowly varying at infinity. Thus, regularly varying distributions possess power-law tails and moments of $|X_{11}|$ of higher order than $\alpha$ are infinite. Typical examples include the Pareto distribution with parameter $\alpha$ and the $t$-distribution with $\alpha$ degrees of freedom.

Now we state the CLT for the logarithmic determinant of the sample correlation matrix $\mathbf{R}$ under infinite fourth moment which is the main result of this paper.

**Theorem 2.1.** *Assume ($C_\gamma$) and that the distribution of $X_{11}$ is symmetric and regularly varying with index $\alpha \in (3,4)$. Then, as $n \to \infty$, we have*

$$\frac{\log\det\mathbf{R} - (p - n + \frac{1}{2})\log(1 - \frac{p}{n}) + p - \frac{p}{n}}{\sqrt{-2\log(1-p/n) - 2p/n}} \xrightarrow{d} N(0,1)\,. \tag{2.4}$$

Theorem 2.1 is proved in Section 4. To numerically illustrate the role of the tail index parameter $\alpha$ and the effect of symmetry of $X_{11}$, we provide a small simulation in Figure 1 and Figure 2. First, we simulate the entries of the data matrix $X_{ij}$ independently from a $t$-distribution with different degrees of freedom smaller than four (infinite fourth moment). We observe a perfect fit of both the histogram and kernel density to the density of the standard normal distribution for all degrees of freedom except 2.5. Thus, the logarithmic law seems not to be valid in case the third absolute moment of the $t$-distribution is infinite, which is inline with our assumption $\alpha > 3$. In the latter case the kernel density still resembles the normal density but has a significantly larger variance, which indicates that the case $\alpha \in (2,3)$ should be investigated separately in the future. The effect of a larger variance becomes more pronounced if we decrease the tail parameter of the observations $X_{ij}$ even further.

Next, we generate the entries $X_{ij}$ from a non-symmetric distribution, namely inverse gamma with scale parameter 2 and varying shape parameter. Note that this distribution has a regularly varying tail with index $\alpha$ equal to the shape parameter and the function $L(x)$ from (2.3) behaving like a constant as $x \to \infty$. Thus, the shape parameter for inverse gamma distribution plays the same role as the degrees of freedom for $t$-distribution, namely if the shape coefficient is smaller than four then the moment of order four does not exist. Hence, the top row in Figure 2 represents the results when the fourth moment exists, while the pictures in the bottom row represent the case of an infinite fourth moment. One can clearly see that symmetry is vital for logarithmic law to be valid. Indeed, by a careful examination of the proof one can

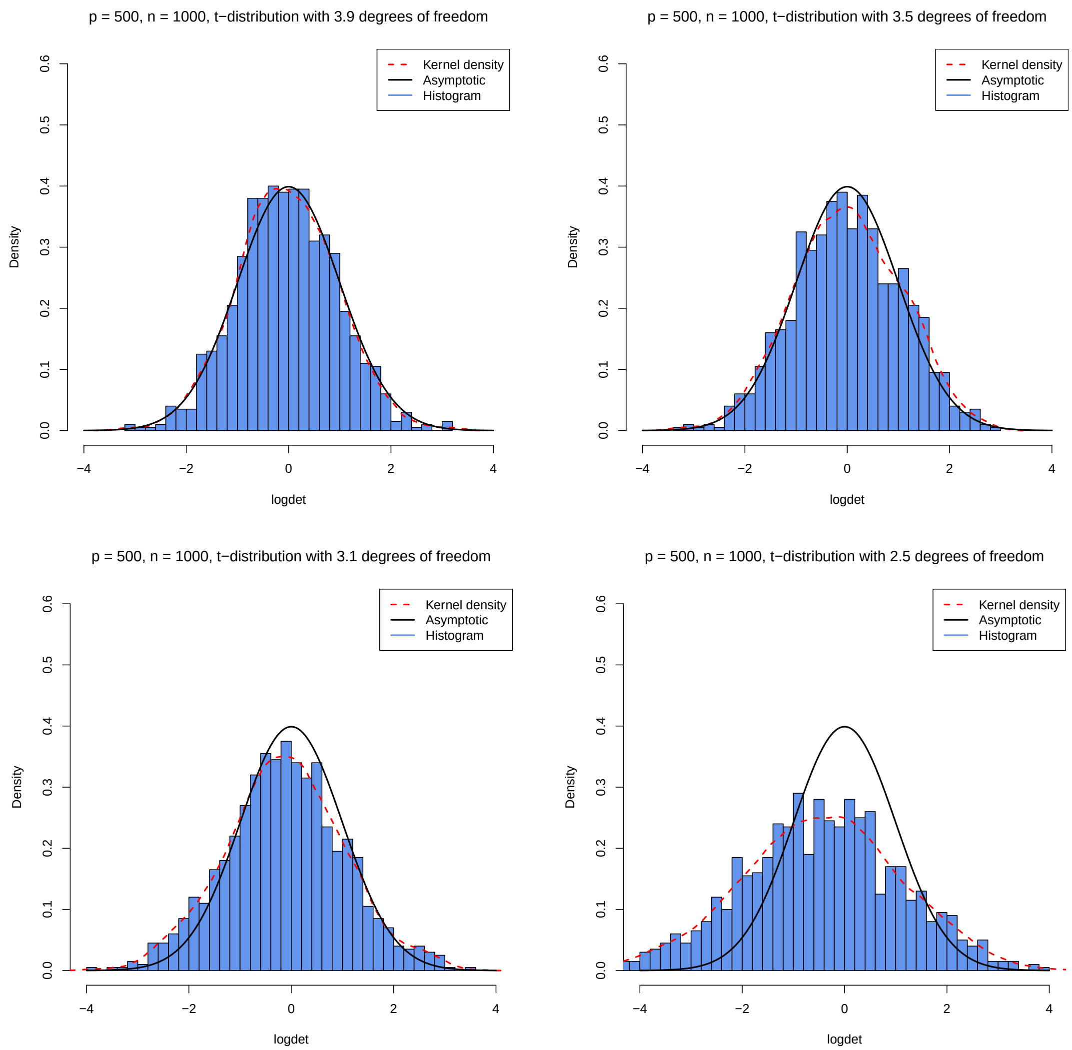


FIG 1. Logarithmic law for $t$ distribution with different degrees of freedom and $p = 500$, $n = 1000$ with 1000 repetitions.

see that asymmetric distribution of $X_{ij}$ as well as a tail parameter $\alpha < 3$ could possibly create additional terms in the asymptotic variance and, thus, the CLT in (2.4) might not be true anymore.

As a consequence, if our assumptions are violated, the limiting distribution of the logarithmic determinant of the sample correlation matrix still resembles the normal one but with a considerably larger variance. The asymmetry effects seem, however, to have a larger impact on the limiting distribution of $\log\det \mathbf{R}$ in the case of heavy tailed data. In case the distribution of heavy-tailed data is not symmetric, it might be beneficial to take an appropriate power transform of the data before using the derived logarithmic law for any testing procedures, for example, testing the uncorrelatedness.

Finally, we briefly comment on the extension of our result to $p$-dimensional observations with population covariance $\mathbf{\Sigma} \neq \mathbf{I}$, which amounts to replacing the data matrix $\mathbf{X}$ with $\mathbf{\Sigma}^{1/2}\mathbf{X}$, where $\mathbf{\Sigma}^{1/2}$ is the Hermitian square root of $\mathbf{\Sigma}$. In the sample covariance case, since

$$\log\det(\mathbf{\Sigma}^{1/2}\mathbf{S}\mathbf{\Sigma}^{1/2}) = \log\det \mathbf{S} + \log\det \mathbf{\Sigma}\,,$$

it is straightforward to obtain a CLT for $\log\det(\mathbf{\Sigma}^{1/2}\mathbf{S}\mathbf{\Sigma}^{1/2})$ from (1.1). Unfortunately, there seems to be no such simple relation for the logarithmic determinant of the sample correlation matrix

$$\widetilde{\mathbf{R}} = \{\mathrm{diag}(\mathbf{\Sigma}^{1/2}\mathbf{S}\mathbf{\Sigma}^{1/2})\}^{-1/2}\mathbf{\Sigma}^{1/2}\mathbf{S}\mathbf{\Sigma}^{1/2}\{\mathrm{diag}(\mathbf{\Sigma}^{1/2}\mathbf{S}\mathbf{\Sigma}^{1/2})\}^{-1/2}\,.$$

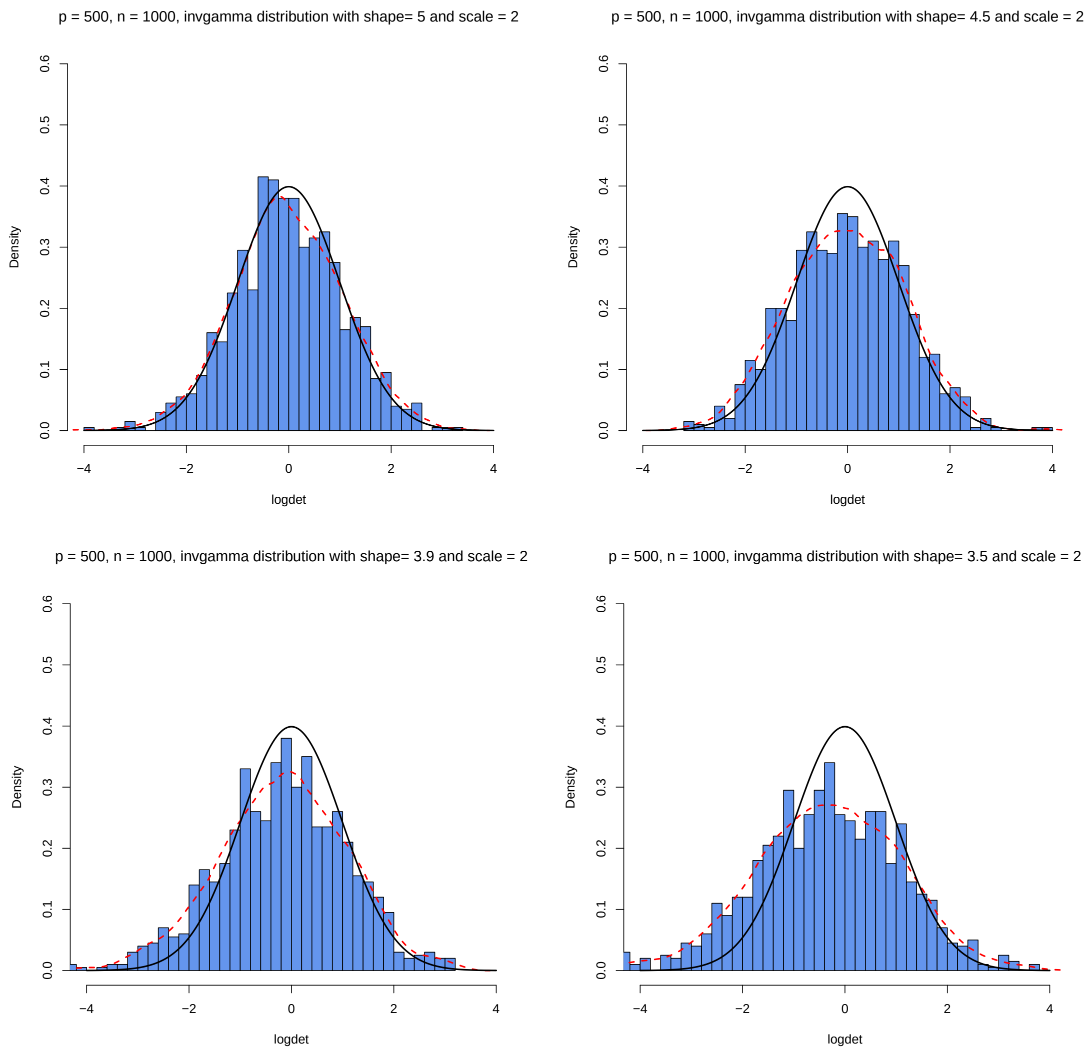


FIG 2. Logarithmic law for inverse gamma distribution with scale $\beta = 2$ and shape $\alpha \in \{5, 4.5, 3.9, 3.5\}$ for $p = 500$ and $n = 1000$ with 1000 repetitions.

Recently, [35] used the identity

$$\log\det\widetilde{\mathbf{R}} = \log\det(\mathbf{\Gamma}^{1/2}\mathbf{S}\mathbf{\Gamma}^{1/2}) - \log\det(\operatorname{diag}(\mathbf{\Gamma}^{1/2}\mathbf{S}\mathbf{\Gamma}^{1/2}))\,,$$

where $\mathbf{\Gamma} = \{\operatorname{diag}(\mathbf{\Sigma})\}^{-1/2}\,\mathbf{\Sigma}\{\operatorname{diag}(\mathbf{\Sigma})\}^{-1/2}$ is the associated population correlation matrix, to derive a CLT in the case of a finite fourth moment. It is an interesting topic for future research to figure out the dependence on $\mathbf{\Gamma}$ in the heavy-tailed case of infinite fourth moment.

### 2.1. *Outline of the proof*

The proof of Theorem 2.1 relies on novel identities and bounds for moments of quadratic forms in random vectors on the unit sphere. As the proof is quite long and technical, it will be split into several parts. First, using the method of perpendiculars from [39] we represent the log determinant of the sample correlation matrix $\mathbf{R} = \mathbf{Y}\mathbf{Y}^\top$ as

$$\log\det\mathbf{R} = c_n + \sum_{i=0}^{p-1}\log(1+\widetilde{Z}_{i+1}) \quad\text{with}\quad c_n = -p\log n + \log(n(n-1)\cdots(n-p+1))\,, \tag{2.5}$$

where, for $i = 0, \dots, p-1$, $\widetilde{Z}_{i+1} = \mathbf{y}_{i+1}^\top Q_i \mathbf{y}_{i+1} - 1$ and $\mathbf{y}_{i+1} = (Y_{i+1,1}, \dots, Y_{i+1,n})^\top$ denotes the $(i+1)$st row of the matrix $\mathbf{Y}$. Here $Y_{ij}$ are given in (2.2), and $Q_i$ is an $n \times n$ projection matrix normalized by its trace and only depends on the vectors $\mathbf{y}_1, \dots, \mathbf{y}_i$. Thus, $Q_i$ and the unit vector $\mathbf{y}_{i+1}$ are independent.

Writing $\operatorname{offdiag}(Q_i)$ for $Q_i - \operatorname{diag}(Q_i)$, the (centered) quadratic form $\widetilde{Z}_{i+1}$ is decomposed into diagonal and off-diagonal parts as follows

$$\widetilde{Z}_{i+1} = \Big[\mathbf{y}_{i+1}^\top \operatorname{diag}(Q_i)\mathbf{y}_{i+1} - 1\Big] + \Big[\mathbf{y}_{i+1}^\top \operatorname{offdiag}(Q_i)\mathbf{y}_{i+1}\Big], \qquad 0 \le i \le p-1.$$

The exact formula for the fourth moment of the diagonal part is derived in Section 3 (Corollary 3.4), while the off-diagonal part is studied in Appendix A. In fact the key results in Section 3, Lemma 3.2 and Theorem 3.3 in particular, are applicable to more general vectors than the $\mathbf{y}_{i+1}$ and therefore might be of independent interest.

In the next step, we perform a Taylor series expansion of the logarithms $\log(1+\widetilde{Z}_{i+1})$. Using the results on the diagonal and off-diagonal parts of $\widetilde{Z}_{i+1}$, we deduce that under ($C_\gamma$) it holds

$$\sum_{i=0}^{p-1} \log(1+\widetilde{Z}_{i+1}) = \sum_{i=0}^{p-1} \widetilde{Z}_{i+1} - \sum_{i=0}^{p-1} \frac{\widetilde{Z}_{i+1}^2}{2} + \text{ remainder}, \tag{2.6}$$

where the remainder term converges to zero in probability. A combination of (2.5) and (2.6) yields that, as $n \to \infty$,

$$\log\det \mathbf{R} + o_{\mathbb{P}}(1) = \left(\sum_{i=0}^{p-1} \widetilde{Z}_{i+1}\right) + \left(c_n - \sum_{i=0}^{p-1} \frac{\widetilde{Z}_{i+1}^2}{2}\right) =: T_1 + T_2 .$$

Finally, the statement of Theorem 2.1 can be deduced from $T_1 \xrightarrow{d} N(0, -2\log(1-\gamma) - 2\gamma)$, which is proven via martingale techniques, and $T_2 - \mu_n \to 0$, as $n \to \infty$, where $\mu_n = (p - n + \frac{1}{2})\log(1 - \frac{p}{n}) - p + \frac{p}{n}$.

The results about the diagonal and off-diagonal parts of $\widetilde{Z}_{i+1}$ are fundamental in analyzing $T_1$ and $T_2$, and in proving that the remainder term is $o_{\mathbb{P}}(1)$ as $n \to \infty$. For example, they allow obtaining the key Lemma 4.7, from which both the formula for the asymptotic variance and mean of $T_1$ can be deduced. Moreover, Proposition 4.3, which is important for justification of the Taylor series expansion of the logarithm, follows from both Proposition A.1 (fourth moment bound for off-diagonal part) and Corollary 3.4 (exact fourth moment for diagonal part). The latter result explains the necessity of the assumption $\alpha > 3$. At the same time Lemma 4.7 and Proposition 4.3 rely a lot on the technical results presented in Appendix B, which establish the rates for the moments of the diagonal entries of the normalized projection matrices $Q_i$.

The main challenge in proving the result of Theorem 2.1 lies in the non-applicability of the lemmas on the moments of centered quadratic forms; see, for example, Bai and Silverstein [4, Lemma B.26 and Lemma 9.1] or Gao et al. [18, Lemma 5 (supplement)]. Morales et al. [32] recently provided version of these lemmas suitable for sample correlation matrices which we state for the reader's convenience.

**Lemma 2.2.** *[32, Lemma 6] Let $B$ be an $n \times n$ non-random symmetric matrix, $x, y \in \mathbb{R}^n$ random vectors of i.i.d. entries with mean zero, variance one, $\mathbb{E}|x_i|^l, \mathbb{E}|y_i|^l \le \nu_l$ and $\mathbb{E}[x_i y_i] = \rho$. Then, for any $s \ge 1$,*

$$\mathbb{E}\left|\frac{x^\top B y}{\|x\|_2 \|y\|_2} - \frac{\rho}{n}\operatorname{tr} B\right|^s \le C_s \Big[n^{-s}\big(\nu_{2s}\operatorname{tr} B^s + (\nu_4 \operatorname{tr} B^2)^{s/2}\big) + \|B\|^s\big(n^{-s/2}\nu_4^{s/2} + n^{-s+1}\nu_{2s}\big)\Big], \tag{2.7}$$

*where $\|\cdot\|_2$ denotes the Euclidean norm on $\mathbb{R}^n$ and $C_s$ is a constant depending only on $s$.*

Note that the right-hand side in (2.7) is infinite whenever the fourth moment is infinite ($\alpha < 4$). Lemma 2.2 is proved by first replacing $\|x\|_2, \|y\|_2$ by their expectations and then applying the Bai Silverstein lemma. It is well-known that the replacement of $\|x\|_2, \|y\|_2$ is possible if and only if the fourth moment of the data is finite; see [24] for details.

Therefore, in the case of infinite fourth moment a more sophisticated approach is required. Our remedy lies in consideration of the entries $Y_{ij}$ directly. Even if the original entries of the data matrix $X_{ij}$ do not obey a finite fourth moment, the self-normalized $Y_{ij}$ in contrast are bounded in absolute value by 1 since $Y_{i1}^2 + \cdots Y_{in}^2 = 1$. Thus, to find the desired central limit theorem one needs to establish the exact limiting behavior of the moments of sums, which involve self-normalized random variables $Y_{ij}^2$ (diagonal part) and their mixed products $Y_{ik}Y_{il}$ (off-diagonal part). This requires a very delicate treatment of their convergence rates, which is achieved by providing exact formulas for those moments. Our rates are quite precise, such that one can clearly see where specific assumptions on the tail index $\alpha$ are important and where they can be weakened. We believe that the method used in this paper can be further extended to data matrices whose entries $X_{ij}$ are non-symmetric and/or obey infinite third moments. The result of the main theorem, however, will change and the desired CLT could be entirely different in that case. In Section 4, we provide a detailed proof of Theorem 2.1.

## 3. Diagonal part: Exact formula

In this section, we will derive an exact formula for the fourth moment of $\sum_{k=1}^n a_k(nZ_k^2-1)$, where $a_1,\dots,a_n$ are constants and $Z_1,\dots,Z_n$ are (essentially) exchangeable random variables satisfying $Z_1^2+\cdots+Z_n^2=1$. We start with the following lemma.

**Lemma 3.1.** *Let $Z_1,\dots,Z_n$ be random variables such that, for all positive integers $m_1,\dots,m_r$ with $m_1+\cdots+m_r\le 4$, $\beta_{2m_1,\dots,2m_r}:=\mathbb{E}[Z_{i_1}^{2m_1}Z_{i_2}^{2m_2}\cdots Z_{i_4}^{2m_4}]$ is finite and invariant under permutations of the indices. Then we have for any numbers $a_1,\dots,a_n$ with $a_1+\cdots+a_n=1$ that*

$$\begin{aligned}\mathbb{E}\Big[\Big(\sum_{k=1}^n a_k(Z_k^2-\mathbb{E}[Z_k^2])\Big)^4\Big] &= S_4\beta_8+4(S_3-S_4)\beta_{6,2}-4S_3\beta_2\beta_6+3(S_2^2-S_4)\beta_{4,4}\\ &+6(S_2-S_2^2-2S_3+2S_4)\beta_{4,2,2}+12(-S_2+S_3)\beta_2\beta_{4,2}+4(3S_2-2S_3-1)\beta_2\beta_{2,2,2}\\ &+(-6S_2+3S_2^2+8S_3-6S_4+1)\beta_{2,2,2,2}+6(1-S_2)\beta_2^2\beta_{2,2}+6S_2\beta_2^2\beta_4-3\beta_2^4\,,\end{aligned}\tag{3.1}$$

*where for $j\ge 1$, we define $S_j=a_1^j+\cdots+a_n^j$. Moreover, we have*

$$\begin{aligned}\mathbb{E}\Big[\Big(\sum_{k=1}^n a_kZ_k^2\Big)^4\Big] &= S_4\beta_8+4(S_3-S_4)\beta_{6,2}+6(S_2-S_2^2-2S_3+2S_4)\beta_{4,2,2}\\ &\quad+3(S_2^2-S_4)\beta_{4,4}+(1-6S_2+3S_2^2+8S_3-6S_4)\beta_{2,2,2,2}\,.\end{aligned}\tag{3.2}$$

**Proof.** We note that all sums in this proof run from 1 to $n$. Using $a_1+\cdots+a_n=1$, it is easy to check that

$$\sum_{k\neq\ell}a_ka_\ell=1-S_2\,,\qquad \sum_{k\neq\ell}a_k^2a_\ell=S_2-S_3\,,\qquad \sum_{k\neq\ell}a_k^3a_\ell=S_3-S_4\,,\tag{3.3}$$

$$\sum_{k\neq\ell}a_k^2a_\ell^2=S_2^2-S_4\,,\qquad \sum_{k\neq\ell\neq j}a_k^2a_\ell a_j=S_2-S_2^2-2S_3+2S_4\,,\tag{3.4}$$

$$\sum_{k\neq\ell\neq j}a_ka_\ell a_j=1-3S_2+2S_3\,,\qquad \sum_{k\neq\ell\neq j\neq h}a_ka_\ell a_ja_h=1-6S_2+3S_2^2+8S_3-6S_4\,.\tag{3.5}$$

For example, we shall show the second relation in (3.4),

$$\sum_{k\neq\ell\neq j}a_k^2a_\ell a_j=\sum_{k\neq\ell}a_k^2a_\ell(1-a_k-a_\ell)=\sum_{k\neq\ell}a_k^2a_\ell-\sum_{k\neq\ell}a_k^3a_\ell-\sum_{k\neq\ell}a_k^2a_\ell^2=S_2-S_2^2-2S_3+2S_4\,.$$

We have the decomposition

$$\begin{aligned}\mathbb{E}\Big[\Big(\sum_{k=1}^n a_kZ_k^2\Big)^4\Big] &= \mathbb{E}\Big[\Big(\sum_{k=1}^n a_k^2Z_k^4\Big)^2\Big]+2\mathbb{E}\Big[\sum_{j=1}^n a_j^2Z_j^4\sum_{k\neq\ell}a_ka_\ell Z_k^2Z_\ell^2\Big]+\mathbb{E}\Big[\Big(\sum_{k\neq\ell}a_ka_\ell Z_k^2Z_\ell^2\Big)^2\Big]\\ &=: I+II+III\,.\end{aligned}\tag{3.6}$$

For the first term, we get

$$I=\beta_8\sum_{k=1}^n a_k^4+\beta_{4,4}\sum_{k\neq\ell}a_k^2a_\ell^2=\beta_8S_4+\beta_{4,4}(S_2^2-S_4)\,,$$

where (3.3) was used for the last equality. In view of (3.3) and (3.4), we have

$$\begin{aligned}II&=2\beta_{6,2}\sum_{k\neq\ell}a_ka_\ell(a_k^2+a_\ell^2)+2\beta_{4,2,2}\sum_{k\neq\ell}a_ka_\ell(S_2-a_k^2-a_\ell^2)\\ &=\beta_{6,2}(4S_3-4S_4)+\beta_{4,2,2}(2S_2-2S_2^2-4S_3+4S_4)\,.\end{aligned}$$

Using (3.3)–(3.5) for the third equality, we find for the third term that

$$\begin{aligned} III &= \beta_{4,4}\,2\sum_{k\neq\ell} a_k^2 a_\ell^2 + \beta_{4,2,2} \sum_{\substack{k\neq\ell, j\neq h\\ \#\{k,\ell,j,h\}=3}} a_k a_\ell a_j a_h + \beta_{2,2,2,2}\sum_{k\neq\ell\neq j\neq h} a_k a_\ell a_j a_h \\ &= \beta_{4,4}\,2\sum_{k\neq\ell} a_k^2 a_\ell^2 + \beta_{4,2,2}\,4\sum_{k\neq\ell\neq j} a_k^2 a_\ell a_j + \beta_{2,2,2,2}\sum_{k\neq\ell\neq j\neq h} a_k a_\ell a_j a_h \\ &= 2\beta_{4,4}(S_2^2 - S_4) + 4\beta_{4,2,2}(S_2 - S_2^2 - 2S_3 + 2S_4) + \beta_{2,2,2,2}(1 - 6S_2 + 3S_2^2 + 8S_3 - 6S_4)\,. \end{aligned}$$

Simplifying $I + II + III$ establishes (3.2) by virtue of (3.6).

Next, we turn to (3.1). To this end, let $\mathbf{A}$ be the $n\times n$ diagonal matrix with diagonal entries $a_1,\dots,a_n$. By Lemma B.4, we have

$$\begin{aligned} \mathbb{E}\Big[\Big(\sum_{k=1}^n a_k Z_k^2\Big)^3\Big] &= \beta_{2,2,2}[(\operatorname{tr}\mathbf{A})^3 + 6\operatorname{tr}\mathbf{A}\operatorname{tr}(\mathbf{A}^2) + 8\operatorname{tr}(\mathbf{A}^3)] + (\beta_6 - 15\beta_{4,2} + 30\beta_{2,2,2})\operatorname{tr}(\mathbf{A}\circ\mathbf{A}\circ\mathbf{A}) \\ &\quad + (\beta_{4,2} - 3\beta_{2,2,2})[3\operatorname{tr}\mathbf{A}\operatorname{tr}(\mathbf{A}\circ\mathbf{A}) + 12\operatorname{tr}(\mathbf{A}\circ\mathbf{A}^2)] \\ &= \beta_{2,2,2}[1 + 6S_2 + 8S_3] + (\beta_6 - 15\beta_{4,2} + 30\beta_{2,2,2})S_3 + (\beta_{4,2} - 3\beta_{2,2,2})[3S_2 + 12S_3]\,, \end{aligned} \tag{3.7}$$

where $\circ$ denotes the Hadamard product. A simple calculation using $a_1 + \cdots + a_n = 1$ yields

$$\mathbb{E}\Big[\Big(\sum_{k=1}^n a_k Z_k^2\Big)^2\Big] = \beta_4 S_2 + \beta_{2,2}(1 - S_2)\,. \tag{3.8}$$

By the binomial theorem, we have

$$\mathbb{E}\Big[\Big(\sum_{k=1}^n a_k Z_k^2 - \beta_2\Big)^4\Big] = \sum_{t=0}^4 \binom{4}{t}\mathbb{E}\Big[\Big(\sum_{k=1}^n a_k Z_k^2\Big)^t\Big](-\beta_2)^{4-t}\,. \tag{3.9}$$

Plugging (3.8), (3.7) and (3.2) into (3.9) and then simplifying establishes (3.1). We omit details of this lengthy computation. $\square$

Additionally assuming $Z_1^2 + \cdots + Z_n^2 = 1$, the relation between the $\beta$'s is captured by the following crucial lemma.

**Lemma 3.2.** *Let $Z_1,\dots,Z_n$ be random variables such that $Z_1^2 + \cdots + Z_n^2 = 1$ and, for all positive integers $m_1,\dots,m_r$ with $m_1 + \cdots + m_r \le 4$, $\beta_{2m_1,\dots,2m_r} := \mathbb{E}[Z_{i_1}^{2m_1} Z_{i_2}^{2m_2}\cdots Z_{i_4}^{2m_4}]$ is invariant under permutations of the indices. Then it holds that $\beta_2 = 1/n$ and*

$$\beta_4 = \frac{1}{n} - (n-1)\beta_{2,2}\,, \qquad \beta_{4,2} = \frac{1}{2}\beta_{2,2} - \frac{n-2}{2}\beta_{2,2,2}\,, \tag{3.10}$$

$$\beta_6 = \frac{1}{n} - \frac{3(n-1)}{2}\beta_{2,2} + \frac{(n-1)(n-2)}{2}\beta_{2,2,2}\,, \tag{3.11}$$

$$\beta_{6,2} = \frac{1}{2}\beta_{2,2} - \frac{5(n-2)}{6}\beta_{2,2,2} + \frac{(n-2)(n-3)}{3}\beta_{2,2,2,2} - \beta_{4,4}\,, \tag{3.12}$$

$$\beta_{4,2,2} = \frac{1}{3}\beta_{2,2,2} + \frac{3-n}{3}\beta_{2,2,2,2}\,, \tag{3.13}$$

$$\begin{aligned} \beta_8 &= \frac{1}{n} + 2(1-n)\beta_{2,2} + \Big(\frac{4n^2}{3} - 4n + \frac{8}{3}\Big)\beta_{2,2,2} \\ &\quad + \Big(\frac{-n^3}{3} + 2n^2 - \frac{11n}{3} + 2\Big)\beta_{2,2,2,2} + (n-1)\beta_{4,4}. \end{aligned} \tag{3.14}$$

**Proof.** Since $Z_1^2 + \cdots + Z_n^2 = 1$, an application of the multinomial theorem shows that for $k \ge 1$,

$$1 = (Z_1^2 + \cdots + Z_n^2)^k = \sum_{r=1}^k \sum_{\substack{m_1+\cdots+m_r=k\\ m_j\ge 1}} \binom{n}{r}\binom{k}{m_1,\dots,m_r}\beta_{2m_1,\dots,2m_r}\,.$$

In particular, for $k=2,3,4$, one obtains

$$1 = n\beta_4 + n(n-1)\beta_{2,2}\,, \tag{3.15}$$

$$1 = n\beta_6 + 3n(n-1)\beta_{4,2} + n(n-1)(n-2)\beta_{2,2,2}\,, \tag{3.16}$$

$$\begin{aligned} 1 &= n\beta_8 + 4n(n-1)\beta_{6,2} + 3n(n-1)\beta_{4,4} + 6n(n-1)(n-2)\beta_{4,2,2} \\ &\quad + n(n-1)(n-2)(n-3)\beta_{2,2,2,2}\,. \end{aligned} \tag{3.17}$$

Since $Z_1^2+\cdots+Z_2^2=1$, it holds $Z_1^{2k}=Z_1^{2k}(Z_1^2+\cdots+Z_n^2)$. Taking expectation one obtains

$$\beta_{2k} = \beta_{2k+2} + (n-1)\beta_{2k,2}\,, \qquad k=1,2,3\,. \tag{3.18}$$

Using $Z_1^{2k}Z_2^2=Z_1^{2k}Z_2^2(Z_1^2+\cdots+Z_n^2)$, one analogously gets

$$\beta_{2k,2} = \beta_{2k+2,2} + \beta_{2k,4} + (n-2)\beta_{2k,2,2}\,, \qquad k=1,2\,, \tag{3.19}$$

and

$$\beta_{2,2,2} = 3\beta_{4,2,2} + (n-3)\beta_{2,2,2,2}\,. \tag{3.20}$$

The lemma now follows from equations (3.15)–(3.20) and some tedious but straightforward computations. □

We now state the main result of this section.

**Theorem 3.3.** *Let $Z_1,\dots,Z_n$ be random variables such that $Z_1^2+\cdots+Z_n^2=1$ and, for all positive integers $m_1,\dots,m_r$ with $m_1+\cdots+m_r\le 4$, $\beta_{2m_1,\dots,2m_r}:=\mathbb{E}[Z_{i_1}^{2m_1}Z_{i_2}^{2m_2}\cdots Z_{i_4}^{2m_4}]$ is invariant under permutations of the indices. Then we have for any numbers $a_1,\dots,a_n$ with $a_1+\cdots+a_n=1$ that*

$$\mathbb{E}\Big[\Big(\sum_{k=1}^n a_k(nZ_k^2-1)\Big)^4\Big] = K_{4,4}n^4\beta_{4,4} + K_{2,2}n^2\beta_{2,2} + K_{2,2,2}n^3\beta_{2,2,2} + K_{2,2,2,2}n^4\beta_{2,2,2,2} + K\,, \tag{3.21}$$

*where $S_j=a_1^j+\cdots+a_n^j$, $j\ge 1$ and*

$$\begin{aligned} K_{4,4} &= 3S_2^2 - 4S_3 + nS_4\,, \qquad K = 6nS_2 - 4n^2S_3 + n^3S_4 - 3\,, \\ K_{2,2} &= -12nS_2 + 8n^2S_3 - 2n^3S_4 + 6\,, \\ K_{2,2,2} &= 8nS_2 - 2nS_2^2 + \frac{8n(1-2n)}{3}S_3 + \frac{2n^2(2n-1)}{3}S_4 - 4\,, \\ K_{2,2,2,2} &= -2nS_2 + (2n-3)S_2^2 + \frac{4(n^2-2n+3)}{3}S_3 - \frac{n(n^2-2n+3)}{3}S_4 + 1\,. \end{aligned}$$

*In particular, we have*

$$K + K_{4,4} + K_{2,2} + K_{2,2,2} + K_{2,2,2,2} = 0\,. \tag{3.22}$$

**Proof.** We have

$$\mathbb{E}\Big[\Big(\sum_{k=1}^n a_k(nZ_k^2-1)\Big)^4\Big] = n^4\mathbb{E}\Big[\Big(\sum_{k=1}^n a_k(Z_k^2-1/n)\Big)^4\Big]\,.$$

The right-hand side can be explicitly computed using Lemma 3.1. Plugging in the formulas from Lemma 3.2, one can check, for example with mathematical software, that (3.21) holds.

Even though, equation (3.22) follows from the defintions of $K, K_{4,4}, K_{2,2}, K_{2,2,2}, K_{2,2,2,2}$. We will provide an additional proof which is more insightful. To this end, set $Z_1=\cdots=Z_n=n^{-1/2}$ which implies that the left-hand side in (3.21) is zero and that the right-hand side is $K+K_{4,4}+K_{2,2}+K_{2,2,2}+K_{2,2,2,2}$. □

While the main focus of this paper is on the sample correlation matrix $\mathbf{R}=\mathbf{Y}\mathbf{Y}^\top$ (see (2.1)), Theorem 3.3 might be of independent interest. We will apply Theorem 3.3 to the rows of $\mathbf{Y}$. Since $Y_{11},\dots,Y_{1n}$ are exchangeable random variables satisfying $Y_{11}^2+\cdots+Y_{1n}^2=1$, one obtains the following corollary.

**Corollary 3.4.** *Let $Y_{11},\dots,Y_{1n}$ be defined in (2.2) and for all positive integers $m_1,\dots,m_r$ with $m_1+\cdots+m_r\le 4$ set $\beta_{2m_1,\dots,2m_r}:=\mathbb{E}[Y_{11}^{2m_1}Y_{12}^{2m_2}\cdots Y_{1r}^{2m_4}]$. Then we have for any numbers $a_1,\dots,a_n$ with $a_1+\cdots+a_n=1$ that*

$$\mathbb{E}\Big[\Big(\sum_{k=1}^n a_k(nY_{1k}^2-1)\Big)^4\Big]=K_{4,4}n^4\beta_{4,4}+K_{2,2}n^2\beta_{2,2}+K_{2,2,2}n^3\beta_{2,2,2}+K_{2,2,2,2}n^4\beta_{2,2,2,2}+K\,, \tag{3.23}$$

*where $S_j=a_1^j+\cdots+a_n^j$, $j\ge 1$, and $K_{4,4},K_{2,2},K_{2,2,2},K_{2,2,2,2},K$ are defined in Theorem 3.3.*

## 4. Proof of the main result

### *4.1. Preliminaries*

Throughout this section, for integers $k_1,\dots,k_r$, we will use the notation

$$\beta_{2k_1,\dots,2k_r}:=\mathbb{E}[Y_{11}^{2k_1}\cdots Y_{1r}^{2k_r}]\,,$$

where we recall the definition of $Y_{ij}$ from (2.2). Since $\beta_{2k_1,\dots,2k_r}=\beta_{2k_{\pi(1)},\dots,2k_{\pi(r)}}$ for any permutation $\pi$ on $\{1,\dots,r\}$ we will typically write the indices in decreasing order. For example, instead of $\beta_{2,4}$ we prefer writing $\beta_{4,2}$. Now we compute the precise asymptotic behavior of $\beta_{2k_1,\dots,2k_r}$.

**Lemma 4.1.** *Let $\alpha\in(2,4)$ and assume that $\mathbb{E}[X_{11}^2]=1$ and $\mathbb{P}(|X_{11}|>x)=x^{-\alpha}L(x)$ for $x>0$ where $L$ is a slowly varying function. Define the $Y_{kn}$'s as in (2.2) and consider integers $k_1,\dots,k_r\ge 1$. Then it holds*

$$\lim_{n\to\infty}\frac{n^{N_1(1-\alpha/2)+r\alpha/2}}{L^{r-N_1}(n^{1/2})}\beta_{2k_1,\dots,2k_r}=\frac{(\alpha/2)^{r-N_1}\Gamma(N_1(1-\alpha/2)+r\alpha/2)\prod_{i:k_i\ge 2}\Gamma(k_i-\alpha/2)}{\Gamma(k_1+\cdots+k_r)}\,, \tag{4.1}$$

*where $N_1=\#\{1\le i\le r:k_i=1\}$. In particular, we have*

$$\lim_{n\to\infty}\frac{n^{\alpha/2}}{L(n^{1/2})}\beta_{2k}=\frac{\alpha\Gamma(\alpha/2)\Gamma(k-\alpha/2)}{2\Gamma(k)}\,,\qquad k\ge 1\,. \tag{4.2}$$

**Proof.** We remark that (4.1) was proved in [1] for $N_1=0$, that is $k_i\ge 2$. For the general case let $\beta=\alpha/2$, $X\stackrel{d}{=}X_{11}$ and consider $r\ge 1$, $k_1+\cdots+k_r=k\ge 1$ with $k_i\ge 1$. From Albrecher and Teugels [1], page 7, we have

$$\mathbb{E}[Y_{11}^{2k_1}\cdots Y_{1r}^{2k_r}]=\frac{(-1)^k}{n\Gamma(k)}\int_0^\infty\Big(\tfrac{t}{n}\Big)^{k-1}\varphi^{n-r}\Big(\tfrac{t}{n}\Big)\prod_{i=1}^r\varphi^{(k_i)}\Big(\tfrac{t}{n}\Big)\,\mathrm{d}t\,, \tag{4.3}$$

where $\varphi(s)=\mathbb{E}[\mathrm{e}^{-sX^2}]$, $s>0$, and $\varphi^{(m)}(s)=\frac{\mathrm{d}^m}{\mathrm{d}s^m}\varphi(s)$. By [1], we have

$$\lim_{n\to\infty}\varphi^{n-r}\Big(\tfrac{t}{n}\Big)=\mathrm{e}^{-t}\,,\qquad t>0\,, \tag{4.4}$$

and that the asymptotic behavior of $\varphi^{(m)}(s)$, $m\in\mathbb{N}$, at the origin is given by

$$(-1)^m\varphi^{(m)}(s)\sim\begin{cases}\beta\Gamma(m-\beta)s^{\beta-m}L(s^{-1/2})\,, & \text{if } m>\beta,\\ \mathbb{E}[X^{2m}]\,, & \text{if } m\le\beta,\end{cases}\qquad s\downarrow 0\,. \tag{4.5}$$

By (4.3), Potter's theorem and the dominated convergence theorem (for more details see [1] or [17]), we obtain in view of (4.4) and (4.5) that, as $n\to\infty$,

$$\begin{aligned}\mathbb{E}[Y_{11}^{2k_1}\cdots Y_{1r}^{2k_r}] &= \frac{(-1)^k}{n\Gamma(k)}\int_0^\infty \Big(\tfrac{t}{n}\Big)^{k-1}\varphi^{n-r}\Big(\tfrac{t}{n}\Big)\Big(\varphi^{(1)}\big(\tfrac{t}{n}\big)\Big)^{N_1}\prod_{i:k_1\ge 2}\varphi^{(k_i)}\Big(\tfrac{t}{n}\Big)\,\mathrm{d}t\\
&\sim \frac{1}{n\Gamma(k)}\int_0^\infty \Big(\tfrac{t}{n}\Big)^{k-1}\mathrm{e}^{-t}\Big(\mathbb{E}[X^2]\Big)^{N_1}\prod_{i:k_i\ge 2}\beta\Gamma(k_i-\beta)\big(\tfrac{t}{n}\big)^{\beta-k_i}\underbrace{L\Big(\big(\tfrac{t}{n}\big)^{-1/2}\Big)}_{\sim L(n^{1/2})}\,\mathrm{d}t\\
&\sim \Big(\prod_{i:k_i\ge 2}\Gamma(k_i-\beta)\Big)\frac{\beta^{r-N_1}L^{r-N_1}(n^{1/2})}{n^{N_1(1-\beta)+\beta r}\Gamma(k)}\int_0^\infty \mathrm{e}^{-t}\,t^{N_1(1-\beta)+\beta r-1}\,\mathrm{d}t\\
&= \frac{L^{r-N_1}(n^{1/2})}{n^{N_1(1-\beta)+\beta r}}\frac{\beta^{r-N_1}\Gamma(N_1(1-\beta)+\beta r)\prod_{i:k_i\ge 2}\Gamma(k_i-\beta)}{\Gamma(k)}\,.\end{aligned}$$

Rearranging yields (4.1). □

*Remark* 4.2. We mention that (4.1) per se does not tell us the speed of convergence of the left-hand side to the limit. For example, by (4.1) we (only) know that $n(n-1)\beta_{2,2}\sim 1$, as $n\to\infty$. Using the first identity in (3.10), we deduce that

$$1-n(n-1)\beta_{2,2}=n\beta_4\sim n^{1-\alpha/2}L(n^{1/2})(\alpha/2)\,\Gamma(\alpha/2)\Gamma(2-\alpha/2)\,,\quad n\to\infty\,,$$

where (4.2) was used in the last step. Thus, for certain cases, Lemma 4.1 in conjunction with Lemma 3.2 reveal the speed of convergence in (4.1).

### 4.2. *Proof of Theorem 2.1*

With some matrix algebra, Wang et al. [39, p. 85-86] derived for the log determinant of the sample covariance matrix $\mathbf{S}=n^{-1}\mathbf{X}\mathbf{X}^\top$ that

$$\log\det\mathbf{S}=-p\log n+\log((n(n-1)\cdots(n-p+1))+\sum_{i=0}^{p-1}\log(1+Z_{i+1})\,,\tag{4.6}$$

where

$$Z_{i+1}=\frac{b_{i+1}^\top P_i b_{i+1}-(n-i)}{n-i}\quad\text{and}\quad P_i=\mathbf{I}_n-B_{(i)}^\top(B_{(i)}B_{(i)}^\top)^{-1}B_{(i)}\,.$$

Here $P_0=\mathbf{I}_n$, $B_{(i)}=(b_1,\dots,b_i)^\top$, and $b_i=(x_{i1},\dots,x_{in})^\top$ denotes the $i$th row of the matrix $\mathbf{X}$, $i=1,\dots,p-1$.

Analogously to (4.6), we get for the log determinant of the sample correlation matrix $\mathbf{R}=\mathbf{Y}\mathbf{Y}^\top$ that

$$\log\det\mathbf{R}=-p\log n+\log(n(n-1)\cdots(n-p+1))+\sum_{i=0}^{p-1}\log(1+\widetilde{Z}_{i+1})\,,\tag{4.7}$$

where

$$\widetilde{Z}_{i+1}=\frac{n\,\widetilde{b}_{i+1}^\top\widetilde{P}_i\widetilde{b}_{i+1}-(n-i)}{n-i}\quad\text{and}\quad \widetilde{P}_i=\mathbf{I}_n-\widetilde{B}_{(i)}^\top(\widetilde{B}_{(i)}\widetilde{B}_{(i)}^\top)^{-1}\widetilde{B}_{(i)}\,.$$

Here $\widetilde{P}_0=\mathbf{I}_n$, $\widetilde{B}_{(i)}=(\widetilde{b}_1,\dots,\widetilde{b}_i)^\top$, and $\widetilde{b}_i=(Y_{i1},\dots,Y_{in})^\top$ denotes the $i$th row of the matrix $\mathbf{Y}$. An important observation is that

$$\widetilde{P}_i=P_i=\mathbf{I}_n-B_{(i)}^\top(B_{(i)}B_{(i)}^\top)^{-1}B_{(i)}$$

is the same projection matrix as in the sample covariance case. Moreover, due to [39, Proposition 2.1] all matrices $B_{(i)}B_{(i)}^\top$ are invertible with overwhelming probability.

We note that $P_i = P_i^2$ and $\operatorname{tr}(P_i) = n - i$, and define

$$Q_i = (q_{i,kl}) = P_i/(n-i)\,, \qquad 0 \le i \le p-1\,.$$

By [31, Lemma 2.1] and [31, Lemma 3.1], we have for $0 \le i \le p-1$ and $1 \le k, l \le n$ that

$$0 \le q_{i,kk} \le \frac{1}{n-i} \quad \text{and} \quad -\frac{1}{2(n-i)} \le q_{i,kl} \le \frac{1}{2(n-i)}\,. \tag{4.8}$$

In what follows we will need the notation

$$\begin{aligned} \mu_n &:= (p-n+\tfrac{1}{2})\log(1-\tfrac{p}{n}) - p + \tfrac{p}{n}\,,\\ c_n &:= -p\log n + \log(n(n-1)\cdots(n-p+1))\,,\\ \widetilde{Y}_{i+1} &:= \tfrac{1}{2}(\widetilde{Z}_{i+1}^2 - \mathbb{E}[\widetilde{Z}_{i+1}^2|\mathcal{F}_i])\,, \qquad 0 \le i \le p-1\,, \end{aligned}$$

where $\mathcal{F}_i = \mathcal{F}_i^{(n)}$ denotes the sigma algebra generated by the first $i$ rows of $\mathbf{X}$.

It is convenient to decompose $\widetilde{Z}_{i+1}$ as follows,

$$\widetilde{Z}_{i+1} = \sum_{j=1}^{n} q_{i,jj}(nY_{i+1,j}^2 - 1) + \sum_{k \ne l} q_{i,kl}\, nY_{i+1,k}Y_{i+1,l} =: U_{i+1} + V_{i+1}\,, \qquad 0 \le i \le p-1. \tag{4.9}$$

The next result is the key ingredient; it will be proved in Section 4.7.

**Proposition 4.3.** *In the setting of Theorem 2.1, if* $\alpha \in (2,4)$[1]*, we have for any* $\varepsilon \in (0, \alpha/2 - 1)$ *that*

$$\lim_{n\to\infty} n^{\varepsilon} \sum_{i=0}^{p-1} \mathbb{E}[V_{i+1}^4] = 0\,. \tag{4.10}$$

*Moreover, if* $\alpha \in (3,4)$*, there exists an* $\varepsilon > 0$ *such that*

$$\lim_{n\to\infty} n^{\varepsilon} \sum_{i=0}^{p-1} \mathbb{E}[U_{i+1}^4] = 0\,. \tag{4.11}$$

By Taylor's theorem, we get

$$\sum_{i=0}^{p-1} \log(1+\widetilde{Z}_{i+1}) = \sum_{i=0}^{p-1} (\widetilde{Z}_{i+1} - \frac{\widetilde{Z}_{i+1}^2}{2}) + \sum_{i=0}^{p-1} R_{i+1}\,, \tag{4.12}$$

where the remainder in Lagrange form is given by

$$R_{i+1} = \frac{1}{3}\Big(\frac{\widetilde{Z}_{i+1}}{1+\theta\widetilde{Z}_{i+1}}\Big)^3 \quad \text{for some } \theta = \theta(\widetilde{Z}_{i+1}) \in (0,1)\,. \tag{4.13}$$

This expansion is justified by

$$\max_{i=0,\ldots,p-1} |\widetilde{Z}_{i+1}| \stackrel{\mathbb{P}}{\to} 0\,, \qquad n \to \infty\,, \tag{4.14}$$

which is an immediate consequence of the following lemma.

**Lemma 4.4.** *Under the conditions of Theorem 2.1, we have for any* $\varepsilon > 0$ *that*

$$\lim_{n\to\infty} \sum_{i=0}^{p-1} \mathbb{P}(|\widetilde{Z}_{i+1}| > \varepsilon) = 0\,.$$

[1] We emphasize that some parts of our proof also work for $\alpha > 2$, which is the widest range of the tail parameter $\alpha$ for which the CLT for the log-determinant might hold. This is due to fact that for $\alpha \in (0,2)$ the limiting spectral distribution of the sample correlation matrix $\mathbf{R}$ is no longer the classical Marčenko–Pastur law but the so-called $\alpha$-heavy Marčenko–Pastur law; see [27] for details.

**Proof.** Using Markov's inequality, $|a+b|^4 \le 2^3(|a|^4+|b|^4)$ for $a,b\in\mathbb{R}$, and Proposition 4.3, we get for $\varepsilon>0$,

$$\sum_{i=0}^{p-1}\mathbb{P}(|\widetilde{Z}_{i+1}|>\varepsilon)\le\sum_{i=0}^{p-1}\frac{\mathbb{E}[\widetilde{Z}_{i+1}^4]}{\varepsilon}\le\frac{8}{\varepsilon}\sum_{i=0}^{p-1}(\mathbb{E}[U_{i+1}^4]+\mathbb{E}[U_{i+1}^4])\to 0\,,\qquad n\to\infty\,. \tag{4.15}$$

□

Recalling the definition of $\widetilde{Y}_{i+1}$, we have

$$\sum_{i=0}^{p-1}(\widetilde{Z}_{i+1}-\frac{\widetilde{Z}_{i+1}^2}{2})=\sum_{i=0}^{p-1}\widetilde{Z}_{i+1}-\sum_{i=0}^{p-1}\widetilde{Y}_{i+1}-\sum_{i=0}^{p-1}\tfrac{1}{2}\mathbb{E}[\widetilde{Z}_{i+1}^2|\mathcal{F}_i]\,. \tag{4.16}$$

In view of (4.7) and (4.12), one gets

$$\log\det\mathbf{R}-\mu_n=\sum_{i=0}^{p-1}\widetilde{Z}_{i+1}-\sum_{i=0}^{p-1}\widetilde{Y}_{i+1}+\sum_{i=0}^{p-1}R_{i+1}-\sum_{i=0}^{p-1}\tfrac{1}{2}\mathbb{E}[\widetilde{Z}_{i+1}^2|\mathcal{F}_i]+c_n-\mu_n\,. \tag{4.17}$$

By virtue of (4.17) and noting that $-2\log(1-p/n)-2p/n\to-2\log(1-\gamma)-2\gamma$, Theorem 2.1 follows from the next four limit relations by an application of the Slutsky lemma,

$$\frac{1}{\sqrt{-2\log(1-p/n)-2p/n}}\sum_{i=0}^{p-1}\widetilde{Z}_{i+1}\xrightarrow{d}N(0,1)\,, \tag{4.18}$$

$$\sum_{i=0}^{p-1}\widetilde{Y}_{i+1}\xrightarrow{\mathbb{P}}0\,, \tag{4.19}$$

$$\sum_{i=0}^{p-1}R_{i+1}\xrightarrow{\mathbb{P}}0\,, \tag{4.20}$$

$$\sum_{i=0}^{p-1}\tfrac{1}{2}\mathbb{E}[\widetilde{Z}_{i+1}^2|\mathcal{F}_i]-c_n+\mu_n\xrightarrow{\mathbb{P}}0\,. \tag{4.21}$$

Equations (4.18), (4.19), (4.20), (4.21) are proved in Sections 4.3, 4.4, 4.5 and 4.6, respectively. This completes the proof of Theorem 2.1.

### *4.3. Proof of* (4.18)

We will use the following CLT for martingale differences.

**Lemma 4.5** (e.g. Hall and Heyde [23])**.** *Let $\{S_{ni},\mathcal{F}_{ni},1\le i\le k_n,n\ge1\}$ be a zero-mean, square integrable martingale array with differences $Z_{ni}$. Suppose that $\mathbb{E}[\max_i Z_{ni}^2]$ is bounded in $n$ and that*

$$\max_i|Z_{ni}|\xrightarrow{\mathbb{P}}0\quad\text{and}\quad\sum_i Z_{ni}^2\xrightarrow{\mathbb{P}}1\,.$$

*Then we have $S_{nk_n}\xrightarrow{d}N(0,1)$.*

In view of $\mathbb{E}[\widetilde{Z}_{i+1}|\mathcal{F}_i]=0$, we observe that $(\widetilde{Z}_{i+1})_i$ is a martingale difference sequence with respect to the filtration $(\mathcal{F}_i)$. We apply Lemma 4.5 to the martingale differences $\sigma_n\widetilde{Z}_{i+1}$ with $\sigma_n=(-2\log(1-p/n)-2p/n)^{-1/2}$. From (4.14), we have $\max_{i=0,\dots,p-1}|\sigma_n\,\widetilde{Z}_{i+1}|\xrightarrow{\mathbb{P}}0$ as $n\to\infty$. In order to check the other conditions in Lemma 4.5, we need the following lemmas. The notation $S_j^{(i)}:=q_{i,11}^j+\cdots+q_{i,nn}^j$, $j\ge1$ will be useful.

**Lemma 4.6.** *Assume that the distribution of $X_{11}$ is symmetric, i.e., $X_{11}\stackrel{d}{=}-X_{11}$. Then it holds for $0\le i\le p-1$ that*

$$\mathbb{E}[U_{i+1}^2]=\frac{1-n\mathbb{E}[S_2^{(i)}]}{n-1}(1-n^2\beta_4)\,, \tag{4.22}$$

$$\mathbb{E}[V_{i+1}^2] = 2n^2\beta_{2,2}\left(\frac{1}{n-i} - \mathbb{E}[S_2^{(i)}]\right). \tag{4.23}$$

**Proof.** Let $0 \le i \le p-1$. By the binomial theorem, we have for $s \ge 1$,

$$\mathbb{E}\Big[\Big(n\sum_{k=1}^{n} q_{i,kk}Y_{i+1,k}^2 - 1\Big)^s \,\Big|\, \mathcal{F}_i\Big] = (-1)^s + s(-1)^{s-1} + \sum_{t=2}^{s}\binom{s}{t}(-1)^{s-t}n^t\mathbb{E}\Big[\Big(\sum_{k=1}^{n} q_{i,kk}Y_{i+1,k}^2\Big)^t \,\Big|\, \mathcal{F}_i\Big]. \tag{4.24}$$

A simple calculation using $\operatorname{tr}(Q_i) = 1$ yields

$$n^2\mathbb{E}\Big[\Big(\sum_{k=1}^{n} q_{i,kk}Y_{i+1,k}^2\Big)^2 \,\Big|\, \mathcal{F}_i\Big] = n^2\beta_4 S_2^{(i)} + n^2\beta_{2,2}(1 - S_2^{(i)}). \tag{4.25}$$

Combining (3.10) from Lemma 3.2 and (4.25), we obtain

$$n^2\mathbb{E}\Big[\Big(\sum_{k=1}^{n} q_{i,kk}Y_{i+1,k}^2\Big)^2 \,\Big|\, \mathcal{F}_i\Big] = n^2\beta_4\Big(S_2^{(i)}\Big(1 + \frac{1}{n-1}\Big) - \frac{1}{n-1}\Big) + \frac{n}{n-1}(1 - S_2^{(i)}).$$

In view of (4.24), this establishes (4.22).

By conditioning on $\mathcal{F}_i$ and using that $q_{i,kl} = q_{i,lk}$, one gets that

$$\mathbb{E}[V_{i+1}^2] = 2n^2\beta_{2,2}\,\mathbb{E}\sum_{k\neq l} q_{i,kl}^2 = 2n^2\beta_{2,2}\left(\frac{1}{n-i} - \mathbb{E}[S_2^{(i)}]\right),$$

where we used $\sum_{k,l} q_{i,kl}^2 = (n-i)^{-1}$ in the last step. □

**Lemma 4.7.** *Under the assumptions of Theorem 2.1, it holds that, as $n \to \infty$,*

$$\sum_{i=0}^{p-1}\mathbb{E}[U_{i+1}^2] = O\big(n^{(3-\alpha)/2+\varepsilon}\big) \qquad \textit{and} \qquad \sum_{i=0}^{p-1}\mathbb{E}[V_{i+1}^2] \sim -2\log(1 - \tfrac{p}{n}) - 2\tfrac{p}{n}$$

*for any $\varepsilon > 0$.*

**Proof.** From Lemma 4.6, equation (4.2) and an application of Lemma B.2, we get for any $\varepsilon > 0$,

$$\begin{aligned}\sum_{i=0}^{p-1}\mathbb{E}[U_{i+1}^2] &= \sum_{i=0}^{p-1}\frac{1 - n\mathbb{E}[S_2^{(i)}]}{n-1}(1 - n^2\beta_4) \le n^2\beta_4\sum_{i=0}^{p-1}\frac{n\mathbb{E}[S_2^{(i)}] - 1}{n-1} \\ &= n^2\beta_4\left(\frac{-p}{n-1} + \frac{n}{n-1}\Big(\sum_{i=0}^{p-1}\mathbb{E}[S_2^{(i)}] - \frac{p}{n}\Big) + \frac{p}{n-1}\right) \\ &= n^{2-\alpha/2+\varepsilon}O(n^{-1/2}) = O\big(n^{(3-\alpha)/2+\varepsilon}\big).\end{aligned}$$

Again from Lemma 4.6 and Lemma B.2, we get, as $n \to \infty$,

$$\begin{aligned}\sum_{i=0}^{p-1}\mathbb{E}[V_{i+1}^2] &= 2n^2\beta_{2,2}\left(\sum_{i=0}^{p-1}\frac{1}{n-i} - \Big(\sum_{i=0}^{p-1}\mathbb{E}[S_2^{(i)}] - \frac{p}{n}\Big) - \frac{p}{n}\right) \\ &= 2\underbrace{n^2\beta_{2,2}}_{\sim 1}\left(\sum_{i=0}^{p-1}\frac{1}{n-i} - O(n^{-1/2}) - \frac{p}{n}\right) \\ &\sim -2\log(1 - p/n) - 2p/n\end{aligned}$$

since $\sum_{i=0}^{p-1}\frac{1}{n-i} \sim -\log(1 - p/n)$. □

Recalling the definition of $\sigma_n^2$ and using Lemma 4.7, we see that

$$\sigma_n^2 \mathbb{E}\Big[\max_{i=0,\dots,p-1} \widetilde{Z}_{i+1}^2\Big] \le \sigma_n^2 \sum_{i=0}^{p-1} \mathbb{E}[\widetilde{Z}_{i+1}^2] = \sigma_n^2 \sum_{i=0}^{p-1} \Big(\mathbb{E}[U_{i+1}^2] + \mathbb{E}[V_{i+1}^2]\Big) = 1 + o(1)\,. \tag{4.26}$$

Due to $\sigma_n^2 \sum_{i=0}^{p-1} \mathbb{E}[\widetilde{Z}_{i+1}^2] = 1 + o(1)$, the condition $\sigma_n^2 \sum_{i=0}^{p-1} \widetilde{Z}_{i+1}^2 \stackrel{\mathbb{P}}{\to} 1$ is implied by

$$\sum_{i=0}^{p-1} (\widetilde{Z}_{i+1}^2 - \mathbb{E}[\widetilde{Z}_{i+1}^2 \,|\mathcal{F}_i]) \stackrel{\mathbb{P}}{\to} 0\,, \qquad n \to \infty\,, \tag{4.27}$$

and

$$\sum_{i=0}^{p-1} (\mathbb{E}[\widetilde{Z}_{i+1}^2 \,|\mathcal{F}_i] - \mathbb{E}[\widetilde{Z}_{i+1}^2]) \stackrel{\mathbb{P}}{\to} 0\,, \qquad n \to \infty\,. \tag{4.28}$$

Observe that (4.27) is equivalent to (4.19). Hence, it remains to show (4.28). To this end, recall that in Lemma 4.6 and its proof it was calculated that

$$\begin{aligned}
\sum_{i=0}^{p-1} \Big(\mathbb{E}[\widetilde{Z}_{i+1}^2 \,|\mathcal{F}_i] - \mathbb{E}[\widetilde{Z}_{i+1}^2]\Big) &= \sum_{i=0}^{p-1} \Big(\mathbb{E}[U_{i+1}^2 + V_{i+1}^2|\mathcal{F}_i] - \mathbb{E}[U_{i+1}^2 + V_{i+1}^2]\Big) \\
&= \sum_{i=0}^{p-1} \frac{n(S_2^{(i)} - \mathbb{E}[S_2^{(i)}])}{n-1}(1 - n^2\beta_4) + 2n^2\beta_{2,2}\,(S_2^{(i)} - \mathbb{E}[S_2^{(i)}]) \\
&\sim (3 - n^2\beta_4) \sum_{i=0}^{p-1} (S_2^{(i)} - \mathbb{E}[S_2^{(i)}]) \stackrel{\mathbb{P}}{\to} 0\,, \qquad n \to \infty\,,
\end{aligned}$$

where we used $n^2\beta_4 = o\big(n^{2-\alpha/2+\varepsilon}\big)$ for any $\varepsilon > 0$ by Lemma 4.1, and Lemma B.2 in the last step. Indeed, using Lemma B.2 for $k = 2$ we obtain

$$\begin{aligned}
\sum_{i=0}^{p-1} (S_2^{(i)} - \mathbb{E}[S_2^{(i)}]) &= \sum_{i=0}^{p-1} (S_2^{(i)} - \frac{1}{n}) - \sum_{i=0}^{p-1} (\mathbb{E}[S_2^{(i)}] - \frac{1}{n}) \\
&= \underbrace{\sum_{i=0}^{p-1}\sum_{\ell=1}^{n} \Big(q_{i,\ell\ell}^2 - \frac{1}{n^2}\Big)}_{=O_{\mathbb{P}}(n^{-1/2}),\ \text{Markov and Lemma B.2}} - \underbrace{\sum_{i=0}^{p-1}\sum_{\ell=1}^{n} \Big(\mathbb{E}[q_{i,\ell\ell}^2] - \frac{1}{n^2}\Big)}_{=O(n^{-1/2}),\ \text{Lemma B.2}} = O_{\mathbb{P}}(n^{-1/2})\,,
\end{aligned}$$

where for the first sum we have also used the fact that $0 \le \sum_{\ell=1}^{n} \Big(q_{i,\ell\ell}^2 - \frac{1}{n^2}\Big)$ by (B.2).

Thus, we have verified the conditions of Lemma 4.5 which now yields (4.18) and finishes the proof.

### 4.4. *Proof of* (4.19)

By Markov's inequality, one has for $\varepsilon > 0$,

$$\mathbb{P}\Big(\Big|\sum_{i=0}^{p-1} \widetilde{Y}_{i+1}\Big| > \varepsilon\Big) \le \varepsilon^{-1}\mathbb{E}\Big[\Big(\sum_{i=0}^{p-1} \widetilde{Y}_{i+1}\Big)^2\Big]\,. \tag{4.29}$$

If $j \neq i$ one can show by conditioning on $\mathcal{F}_{\max(i,j)}$ that $\mathbb{E}[\widetilde{Y}_{i+1}\widetilde{Y}_{j+1}] = 0$. This in conjunction with the inequality $(a+b)^2 \le 2(a^2+b^2)$ gives

$$\begin{aligned}
\mathbb{E}\Big[\Big(\sum_{i=0}^{p-1}\widetilde{Y}_{i+1}\Big)^2\Big] &= \sum_{i=0}^{p-1}\mathbb{E}[\widetilde{Y}_{i+1}^2] = \frac{1}{4}\sum_{i=0}^{p-1}\mathbb{E}\Big[(\widetilde{Z}_{i+1}^2 - \mathbb{E}[\widetilde{Z}_{i+1}^2|\mathcal{F}_i])^2\Big] \\
&\le \underbrace{\frac{1}{2}\sum_{i=0}^{p-1}\mathbb{E}[\widetilde{Z}_{i+1}^4]}_{=o(1) \text{ by } (4.15)} + \frac{1}{2}\sum_{i=0}^{p-1}\mathbb{E}[(\mathbb{E}[\widetilde{Z}_{i+1}^2|\mathcal{F}_i])^2]\Big) \\
&= o(1) + \frac{1}{2}\sum_{i=0}^{p-1}\mathbb{E}\Big[\Big\{\frac{1-nS_2^{(i)}}{n-1}(1-n^2\beta_4) + \underbrace{2n^2\beta_{2,2}\Big(\frac{1}{n-i} - S_2^{(i)}\Big)}_{\le c\,n^{-1} \text{ for some } c>0}\Big\}^2\Big] \\
&\le o(1) + (1-n^2\beta_4)^2\sum_{i=0}^{p-1}\mathbb{E}\Big[\Big\{\frac{1-nS_2^{(i)}}{n-1}\Big\}^2\Big] \\
&\le o(1) + O\big(n^{3-\alpha+2\varepsilon}\big)\,,
\end{aligned}$$

where we used Lemma 4.6 to obtain the third line, and Lemma 4.1 in the last step.

In view of (4.29) and since $\alpha > 3$, we have proved (4.19).

### 4.5. *Proof of* (4.20)

We need the following lemma.

**Lemma 4.8.** *[8, Lemma 4.1] For $R_{i+1}$ defined in (4.13) and $a > 0$, if $\widetilde{Z}_{i+1} \ge -1 + (\log n)^{-a}$ one has*

$$|R_{i+1}| \le C(U_{i+1}^2 + |V_{i+1}|^{2+\delta})\log\log n$$

*for any $0 \le \delta \le 1$. Here $C = C(a,\delta)$ is a positive constant that only depends on $a$ and $\delta$.*

A combination of Lemmas 4.4 and 4.8 yields that, with probability $1-o(1)$, one has

$$\sum_{i=0}^{p-1}|R_{i+1}| \le C\sum_{i=0}^{p-1}(U_{i+1}^2 + |V_{i+1}|^{2+\delta})\log\log n\,, \qquad 0 \le \delta \le 1\,. \tag{4.30}$$

By (4.30) and Markov's inequality, $\sum_{i=0}^{p-1} R_{i+1} \xrightarrow{\mathbb{P}} 0$ follows from

$$\lim_{n\to\infty}\log\log n\sum_{i=0}^{p-1}\Big(\mathbb{E}[U_{i+1}^2] + \mathbb{E}[|V_{i+1}|^{2+\delta}]\Big) = 0\,,$$

which in view of Lyapunov's inequality is implied by

$$\lim_{n\to\infty}\log\log n\sum_{i=0}^{p-1}\Big(\mathbb{E}[U_{i+1}^2] + (\mathbb{E}[V_{i+1}^4])^{(2+\delta)/4}\Big) = 0\,, \tag{4.31}$$

The $U$-part in (4.31) follows from Lemma 4.7.

Finally by Proposition A.1, we have, for any $\varepsilon > 0$ and $n$ sufficiently large, that $E[V_{i+1}^4] \le C\,n^{-\alpha/2+\varepsilon}$, $0 \le i \le p-1$, where the constant $C > 0$ does not depend on $n$ and $i$. Therefore,

$$\sum_{i=0}^{p-1}(\mathbb{E}[V_{i+1}^4])^{(2+\delta)/4} \le C^{(2+\delta)/4}\,p\,n^{(-\alpha/2+\varepsilon)(2+\delta)/4}$$

With $\delta = 1$ and using $p/n \to \gamma \in (0,1)$, the right-hand side is

$$C^{3/4}\frac{p}{n}n^{1-\frac{3\alpha}{8}+\frac{3\varepsilon}{4}} \to 0\,, \qquad n \to \infty\,,$$

for $\varepsilon > 0$ sufficiently small since $\alpha > 8/3$. This shows the $V$-part in (4.31) and completes the proof of (4.20).

*4.6. Proof of* (4.21)

In view of (4.28), equation (4.21) follows from

$$\sum_{i=0}^{p-1} \tfrac{1}{2}\mathbb{E}[\widetilde{Z}_{i+1}^2] - c_n + \mu_n \to 0\,, \qquad n \to \infty\,. \tag{4.32}$$

From Lemma 4.7, we have

$$\sum_{i=0}^{p-1} \tfrac{1}{2}\mathbb{E}[\widetilde{Z}_{i+1}^2] = \sum_{i=0}^{p-1} \tfrac{\mathbb{E}[U_{i+1}^2]}{2} + \sum_{i=0}^{p-1} \tfrac{\mathbb{E}[V_{i+1}^2]}{2} \sim -\log(1-p/n) - p/n\,.$$

Recalling the definitions $\mu_n = (p-n+\frac{1}{2})\log(1-\frac{p}{n}) - p + \frac{p}{n}$ and $c_n = -p\log n + \log(n(n-1)\cdots(n-p+1))$, (4.32) is thus equivalent to

$$(p-n-\frac{1}{2})\log(1-p/n) - p - \sum_{i=1}^{p-1}\log(1-i/n) \to 0\,, \qquad n \to \infty\,. \tag{4.33}$$

Taking the logarithm on both sides of the identity

$$\prod_{i=1}^{p-1}\Big(1-\frac{i}{n}\Big) = \frac{n!(1-(p-1)/n)}{(n-p+1)!\,n^{p-1}} = \frac{n!(n-p+1)}{n(n-p+1)!\,n^{p-1}} = \frac{(n-1)!}{(n-p)!\,n^{p-1}}\,,$$

we get

$$\sum_{i=1}^{p}\log(1-i/n) = \log(n-1)! - (p-1)\log n - \log(n-p)!\,.$$

We approximate these terms using Stirling's formula $\log(n!) = n\log n - n + \frac{1}{2}\log(2\pi n) + O(n^{-1})$ and obtain

$$\begin{aligned}
&(p-1)\log n - \log(n-1)! + \log(n-p)! = (p-1)\log n - (n-1)\log(n-1) + (n-1)\\
&\quad - \frac{\log(2\pi(n-1))}{2} + (n-p)\log(n-p) - (n-p) + \frac{\log(2\pi(n-p))}{2} + O(n^{-1})\\
&= p-1 + (p-1)\log n - (n-\tfrac{1}{2})\log(n-1) + (n-p+\tfrac{1}{2})\log(n-p) + O(n^{-1})\\
&= p-1 + (n-\tfrac{1}{2})\log(\tfrac{n}{n-1}) + (n-p+\tfrac{1}{2})\log(1-\tfrac{p}{n}) + O(n^{-1})\,.
\end{aligned}$$

Therefore, the left-hand side in (4.33) is $-1 + (n-\frac{1}{2})\log(\frac{n}{n-1}) + O(n^{-1})$ which converges to zero as $n \to \infty$.

This establishes (4.33) and thus finishes the proof of (4.21).

*4.7. Proof of Proposition 4.3*

First, we prove (4.10). Let $\alpha \in (2,4)$ and $\varepsilon \in (0, \alpha/2-1)$. By Proposition A.1 we have, for any $\delta > 0$ and $n$ sufficiently large, that $\mathbb{E}[V_{i+1}^4] \le C\,n^{-\alpha/2+\delta}$, $0 \le i \le p-1$, where the constant $C > 0$ does not depend on $n$. Therefore,

$$n^{\varepsilon}\sum_{i=0}^{p-1}\mathbb{E}[V_{i+1}^4] \le C\,p\,n^{-\alpha/2+\delta+\varepsilon}$$

and using $p/n \to \gamma \in (0,1)$, the right-hand side converges to zero for sufficiently small $\delta > 0$. This proves (4.10).

Next, we turn to the proof of (4.11). Let $\alpha \in (3,4)$ and $\varepsilon \in (0, \alpha-3)$. From Corollary 3.4, we know that for $0 \le i \le p-1$,

$$\begin{aligned}
\mathbb{E}[U_{i+1}^4] &= \mathbb{E}\mathbb{E}\Big[\Big(\sum_{k=1}^{n} q_{i,kk}(nY_{i+1,k}^2 - 1)\Big)^4\Big|\mathcal{F}_i\Big]\\
&= \mathbb{E}[K_{4,4}^{(i)}]n^4\beta_{4,4} + \mathbb{E}[K_{2,2}^{(i)}]n^2\beta_{2,2} + \mathbb{E}[K_{2,2,2}^{(i)}]n^3\beta_{2,2,2} + \mathbb{E}[K_{2,2,2,2}^{(i)}]n^4\beta_{2,2,2,2} + \mathbb{E}[K^{(i)}]\,,
\end{aligned} \tag{4.34}$$

where $S_j^{(i)} = q_{i,11}^j + \cdots + q_{i,nn}^j$, $j \ge 1$, and

$$K_{4,4}^{(i)} = 3(S_2^{(i)})^2 - 4S_3^{(i)} + nS_4^{(i)}\,, \qquad K^{(i)} = 6nS_2^{(i)} - 4n^2 S_3^{(i)} + n^3 S_4^{(i)} - 3\,,$$
$$K_{2,2}^{(i)} = -12nS_2^{(i)} + 8n^2 S_3^{(i)} - 2n^3 S_4^{(i)} + 6\,, \tag{4.35}$$

$$K_{2,2,2}^{(i)} = 8nS_2^{(i)} - 2n(S_2^{(i)})^2 + \frac{8n(1-2n)}{3} S_3^{(i)} + \frac{2n^2(2n-1)}{3} S_4^{(i)} - 4\,, \tag{4.36}$$

$$K_{2,2,2,2}^{(i)} = -2nS_2^{(i)} + (2n-3)(S_2^{(i)})^2 + \frac{4(n^2-2n+3)}{3} S_3^{(i)} - \frac{n(n^2-2n+3)}{3} S_4^{(i)} + 1\,. \tag{4.37}$$

By (3.22), we have

$$K^{(i)} + K_{4,4}^{(i)} + K_{2,2}^{(i)} + K_{2,2,2}^{(i)} + K_{2,2,2,2}^{(i)} = 0\,. \tag{4.38}$$

Plugging (4.38) into (4.34) gives

$$\mathbb{E}[U_{i+1}^4] = \mathbb{E}[K_{44}^{(i)}](n^4\beta_{4,4} - 1) + \mathbb{E}[K_{2,2}^{(i)}](n^2\beta_{2,2} - 1)$$
$$+ \mathbb{E}[K_{2,2,2}^{(i)}](n^3\beta_{2,2,2} - 1) + \mathbb{E}[K_{2,2,2,2}^{(i)}](n^4\beta_{2,2,2,2} - 1). \tag{4.39}$$

We will bound the right-hand side term by term.

Due to $p/n \to \gamma \in (0,1)$ it holds $(1-\gamma)n \sim n - p \le n - i \le n$, so that $n - i$ is of order $n$ for all $0 \le i \le p-1$. A combination of this fact with (4.8) yields that for sufficiently large $n$ there exists a constant $c > 1$ such that $|S_j^{(i)}| \le c^{1/2} n^{1-j}$. Thus we get

$$|K_{4,4}^{(i)}| = |3(S_2^{(i)})^2 - 4S_3^{(i)} + nS_4^{(i)}| \le \frac{3c}{n^2} + \frac{4c}{n^2} + \frac{c}{n^2} = \frac{8c}{n^2}\,. \tag{4.40}$$

Using (4.40), for any $\varepsilon > 0$ and $n$ sufficiently large the first term is bounded by

$$\Big|\sum_{i=0}^{p-1} \mathbb{E}[K_{44}^{(i)}](n^4\beta_{4,4} - 1)\Big| \le \underbrace{|(n^4\beta_{4,4} - 1)|}_{\le n^{4-\alpha+\varepsilon}} \sum_{i=0}^{p-1} \mathbb{E}[|K_{44}^{(i)}|] = O(n^{3-\alpha+\varepsilon})\,.$$

Note that $1 - n^2\beta_{2,2}, 1 - n^3\beta_{2,2,2}, 1 - n^4\beta_{2,2,2,2}$ are nonnegative. Thus,

$$\sum_{i=0}^{p-1} \mathbb{E}[U_{i+1}^4] \le O(n^{3-\alpha+\varepsilon}) + (1 - n^2\beta_{2,2})\Big|\sum_{i=0}^{p-1} \mathbb{E}[K_{2,2}^{(i)}]\Big|$$
$$+ (1 - n^3\beta_{2,2,2})\Big|\sum_{i=0}^{p-1} \mathbb{E}[K_{2,2,2}^{(i)}]\Big| + (1 - n^4\beta_{2,2,2,2})\Big|\sum_{i=0}^{p-1} \mathbb{E}[K_{2,2,2,2}^{(i)}]\Big|.$$

Next, we turn to the remaining terms. Since

$$n^2\beta_{2,2} \sim n^3\beta_{2,2,2} \sim n^4\beta_{2,2,2,2} \sim 1\,, \quad n \to \infty\,,$$

it holds for any $\varepsilon > 0$,

$$1 - n^2\beta_{2,2} = 1 - n(n-1)\beta_{2,2} + O(n^{-1}) = n\beta_4 + O(n^{-1}) = O(n^{1-\alpha/2+\varepsilon})\,,$$

where also (3.15) was used. Analogously, applying (3.16), (3.17) and Lemma 4.1, we get for any $\varepsilon > 0$ that

$$1 - n^3\beta_{2,2,2} = O(n^{1-\alpha/2+\varepsilon}) \quad \text{and} \quad 1 - n^4\beta_{2,2,2,2} = O(n^{1-\alpha/2+\varepsilon})\,.$$

Hence, (4.11) is proved if we can show that there exists an $\varepsilon > 0$ such that, as $n \to \infty$,

$$n^{1-\alpha/2+\varepsilon}\Big|\sum_{i=0}^{p-1} \mathbb{E}[K_{2,2}^{(i)}]\Big| \to 0\,, \qquad n^{1-\alpha/2+\varepsilon}\Big|\sum_{i=0}^{p-1} \mathbb{E}[K_{2,2,2}^{(i)}]\Big| \to 0\,, \qquad n^{1-\alpha/2+\varepsilon}\Big|\sum_{i=0}^{p-1} \mathbb{E}[K_{2,2,2,2}^{(i)}]\Big| \to 0\,.$$

Fortunately, Lemma B.1 verifies the latter. The proof is complete.

## Appendix A: Off-diagonal part of a quadratic form

**Proposition A.1.** *Let $\alpha \in (2,4)$. Under the assumptions of Theorem 2.1 we have for $\varepsilon > 0$ and $n$ sufficiently large,*

$$n^4 \mathbb{E}\Big[\Big(\sum_{k\neq l} q_{i,kl} Y_{i+1,k} Y_{i+1,l}\Big)^4\Big] \le C\, n^{-\alpha/2+\varepsilon}\,, \qquad i=0,1,\ldots,p-1\,, \tag{A.1}$$

*where the constant $C>0$ does not depend on $n$ and $i$.*

**Proof.** Let $0 \le i \le p-1$ and $s=4$. Throughout this proof, in the notation $\beta_{m_1,\ldots,m_r}$ we always assume $m_1+\cdots+m_r = s$. Using that $Y_{i+1,j} \stackrel{d}{=} -Y_{i+1,j}$, we have

$$\begin{aligned}
\mathbb{E}\Big[\Big(\sum_{k_1\neq k_2} q_{i,k_1k_2} Y_{i+1,k_1} Y_{i+1,k_2}\Big)^s \Big| \mathcal{F}_i\Big] &= \sum_{\substack{k_1\neq k_2,\ldots,k_{2s-1}\neq k_{2s}:\\ \sum_{t=1}^{2s} \delta_{k_jk_t} \text{ is even } \forall 1\le j\le 2s}} q_{i,k_1k_2}\cdots q_{i,k_{2s-1}k_{2s}}\, \mathbb{E}[Y_{i+1,k_1}\cdots Y_{i+1,k_{2s}}]\\
&= \sum_{r=2}^{s} \sum_{\mathbf{k}\in\mathcal{K}_{r,s}} q_{i,\mathbf{k}} \beta_{\mathbf{k}}\\
&\le \sum_{r=2}^{s} \max_{\mathbf{k}\in\mathcal{K}_{r,s}} \beta_{\mathbf{k}} \,\Big|\sum_{\mathbf{k}\in\mathcal{K}_{r,s}} q_{i,\mathbf{k}}\Big|\,,
\end{aligned}$$

where $\beta_{\mathbf{k}} = \mathbb{E}[Y_{i+1,k_1}\cdots Y_{i+1,k_{2s}}]$ and $q_{i,\mathbf{k}} = q_{i,k_1k_2}\cdots q_{i,k_{2s-1}k_{2s}}$ for $\mathbf{k}=(k_1,\ldots,k_{2s})$, and

$$\mathcal{K}_{r,s} = \Big\{(k_1,\ldots,k_{2s}) \in \{1,\ldots,n\}^{2s} : \substack{\#\{k_1,\ldots,k_{2s}\}=r;\ k_1\neq k_2,\ldots,k_{2s-1}\neq k_{2s};\\ \sum_{t=1}^{2s}\delta_{k_jk_t} \text{ is even } \forall 1\le j\le 2s}\Big\}.$$

Here $\delta_{k_jk_t}$ is the Kronecker-delta, i.e., $\delta_{k_jk_t} = \mathbb{1}_{\{k_j=k_t\}}$. We will bound $\max_{\mathbf{k}\in\mathcal{K}_{r,s}} \beta_{\mathbf{k}}$ and $|\sum_{\mathbf{k}\in\mathcal{K}_{r,s}} q_{i,\mathbf{k}}|$.

We start with the first term. By Lemma 4.1, we have for integers $k_1,\ldots,k_r\ge 1$ that

$$\lim_{n\to\infty} \frac{n^{N_1(1-\alpha/2)+r\alpha/2}}{L^{r-N_1}(n^{1/2})} \beta_{2k_1,\ldots,2k_r} \sim C(k_1,\ldots,k_r)\,, \tag{A.2}$$

where $N_1 := N_1(k_1,\ldots,k_r) := \#\{1\le j\le r : k_j = 1\}$ and

$$C(k_1,\ldots,k_r) := \frac{(\alpha/2)^{r-N_1}\Gamma(N_1(1-\alpha/2)+r\alpha/2)\prod_{i:k_i\ge 2}\Gamma(k_i-\alpha/2)}{\Gamma(k_1+\cdots+k_r)}.$$

Since $(\prod_j \Gamma(a_j))/\Gamma(\sum_j a_j) \le 1$ for $a_j\ge 0$ we observe that

$$C(k_1,\ldots,k_r) \le (\alpha/2)^{r-N_1} \le 2^{r-N_1}\,. \tag{A.3}$$

We recall the Potter bounds on the regularly varying function $L\ge 0$. For any $\varepsilon>0$ and sufficiently large $n$ it holds

$$n^{-\varepsilon} \le L((n^{1/2})) \le n^{\varepsilon}\,. \tag{A.4}$$

Choose $\varepsilon \in (0,\alpha/2-1)$. In view of (A.2)-(A.4), we have for sufficiently large $n$ that

$$\beta_{2k_1,\ldots,2k_r} \le n^{\alpha/2(N_1-r)-N_1} L^{r-N_1}(n^{1/2}) 2^{r-N_1} \le n^{-(\alpha/2-\varepsilon)(r-N_1)-N_1}\,. \tag{A.5}$$

Therefore, we obtain

$$\max_{\mathbf{k}\in\mathcal{K}_{r,s}} \beta_{\mathbf{k}} \le \max_{\substack{k_1,\ldots,k_r\ge 1:\\ k_1+\cdots+k_r=s}} \beta_{2k_1,\ldots,2k_r} \le n^{-r(\alpha/2-\varepsilon)} \max_{\substack{k_1,\ldots,k_r\ge 1:\\ k_1+\cdots+k_r=s}} n^{N_1(\alpha/2-1-\varepsilon)}\,.$$

Since $N_1 \le r - \mathbb{1}_{\{r<s\}}$, we conclude that for large $n$,

$$\max_{\mathbf{k}\in\mathcal{K}_{r,s}} \beta_{\mathbf{k}} \le n^{-r-(\alpha/2-1-\varepsilon)\mathbb{1}_{\{r<s\}}}\,. \tag{A.6}$$

This establishes a bound on $\max_{\mathbf{k}\in\mathcal{K}_{r,s}} \beta_{\mathbf{k}}$. For later reference, we note that $\alpha/2-1-\varepsilon>0$.

Next, we turn to the bound of $|\sum_{\mathbf{k}\in\mathcal{K}_{r,s}} q_{i,\mathbf{k}}|$. Let $(X_j)_{j\ge1}$ be an i.i.d. sequence (which is also independent of $\mathbf{X}$) with distribution $\mathbb{P}(X_j=1)=\mathbb{P}(X_j=-1)=1/2$. Using that $\mathbb{E}[X_j^t]=1$ if $t$ is even and zero otherwise, we have as above

$$\mathbb{E}\Big[\Big|\sum_{k_1\neq k_2} q_{i,k_1k_2}X_{k_1}X_{k_2}\Big|^s\Big|\mathcal{F}_i\Big]=\sum_{r=2}^{s}\sum_{\mathbf{k}\in\mathcal{K}_{r,s}} q_{i,\mathbf{k}}\,. \tag{A.7}$$

Applying Lemma B.3 with the sequence $(X_j)$, we get

$$\mathbb{E}\Big[\Big|\sum_{k_1\neq k_2} q_{i,k_1k_2}X_{k_1}X_{k_2}\Big|^s\Big|\mathcal{F}_i\Big]\le (Cs)^s\Big(\sum_{k\neq l} q_{i,kl}^2\Big)^{s/2}$$

In view of (A.7), we see that

$$\sum_{r=2}^{s}\sum_{\mathbf{k}\in\mathcal{K}_{r,s}} q_{i,\mathbf{k}}\le\sum_{r=2}^{s}\sum_{\mathbf{k}\in\mathcal{K}_{r,s}} |q_{i,\mathbf{k}}|\le (Cs)^s\Big(\sum_{k\neq l} q_{i,kl}^2\Big)^{s/2}, \tag{A.8}$$

where the last inequality follows from the fact that the right-hand side in (B.17) remains the same if we replace $a_{ij}$ with $|a_{ij}|$. Here $C$ is an absolute constant that does not depend on $s$.

Since the right-hand side in (A.6) depends on $r$, it is important find an upper bound on $|\sum_{\mathbf{k}\in\mathcal{K}_{r,s}} q_{i,\mathbf{k}}|$ that uses the value of $r=2,\dots,s$ as well. If $r=s$ we conclude from (A.8) and $\sum_{k,l} q_{i,kl}^2=(n-i)^{-1}$ that

$$\Big|\sum_{\mathbf{k}\in\mathcal{K}_{s,s}} q_{i,\mathbf{k}}\Big|\le (Cs)^s\Big(\sum_{k\neq l} q_{i,kl}^2\Big)^{s/2}\le (Cs)^s\Big(\frac{1}{n-i}\Big)^{s/2}. \tag{A.9}$$

Note that the term $(\sum_{k\neq l} q_{i,kl}^2)^{s/2}$ actually appears in $\sum_{\mathbf{k}\in\mathcal{K}_{s,s}} q_{i,\mathbf{k}}$. Indeed, this follows directly from the definition of the latter sum by setting $k_1=k_3, k_2=k_4,\dots,k_{2s-2}=k_{2s}$. Hence, the maximum number of distict indices $k_j$ in $q_{i,\mathbf{k}}$ and the maximum number of distinct indices in $(\sum_{k\neq l} q_{i,kl}^2)^{s/2}$ are both equal to $r$. From the definition of $\mathcal{K}_{r,s}$, recall that $\#\{k_1,\dots,k_{2s}\}=r$ if $(k_1,\dots,k_{2s})\in\mathcal{K}_{r,s}$.

If $r=s-1$, we may thus restrict ourselves to $s-1$ distinct indices. Due to $q_{kl}=q_{lk}$, this yields the bound

$$\Big|\sum_{\mathbf{k}\in\mathcal{K}_{s-1,s}} q_{i,\mathbf{k}}\Big|\le (Cs)^s\Big(\sum_{k\neq l} q_{i,kl}^2\Big)^{s/2-2}\sum_{k_1\neq k_2} q_{i,k_1k_2}^2\sum_{k_3=1;k_3\neq k_1}^{n} q_{i,k_1k_3}^2\le (Cs)^s\Big(\frac{1}{n-i}\Big)^{s/2+1}, \tag{A.10}$$

where the last inequality holds since $Q_i^2=Q_i/(n-i)$ and (4.8) imply

$$\sum_{l=1}^{n} q_{i,kl}^2=\frac{q_{i,kk}}{n-i}\le\Big(\frac{1}{n-i}\Big)^2.$$

From the definition of $\mathcal{K}_{r,s}$ and (4.8) it follows for $r=2$ that

$$\Big|\sum_{\mathbf{k}\in\mathcal{K}_{2,s}} q_{i,\mathbf{k}}\Big|=2^s\Big|\sum_{k<l} q_{i,kl}^s\Big|\le\frac{2}{(n-i)^{s-2}}\sum_{k\neq l} q_{i,kl}^2\le 2\Big(\frac{1}{n-i}\Big)^{s-1}.$$

In combination with (A.9) and (A.10), this yields that

$$\Big|\sum_{\mathbf{k}\in\mathcal{K}_{s-t,s}} q_{i,\mathbf{k}}\Big|\le (Cs)^s\Big(\frac{1}{n-i}\Big)^{s/2+\lceil t/2\rceil},\qquad t=0,\dots,s-2, \tag{A.11}$$

where $\lceil t/2\rceil$ is the smallest integer greater or equal to $t/2$ and $C>0$ is a constant.

Finally, we complete the proof of the proposition. In view of (A.6) and (A.11), we get for $s = 4$ and sufficiently large $n$,

$$
\begin{aligned}
n^s \Big|\mathbb{E}\Big[\Big(\sum_{k_1 \neq k_2} q_{i,k_1k_2} Y_{i+1,k_1} Y_{i+1,k_2}\Big)^s\Big]\Big| &= n^s \Big|\mathbb{E}\mathbb{E}\Big[\Big(\sum_{k_1 \neq k_2} q_{i,k_1k_2} Y_{i+1,k_1} Y_{i+1,k_2}\Big)^s \Big| \mathcal{F}_i\Big]\Big| \\
&\le n^s\, \mathbb{E} \sum_{r=2}^{s} \max_{\mathbf{k} \in \mathcal{K}_{r,s}} \beta_{\mathbf{k}} \Big| \sum_{\mathbf{k} \in \mathcal{K}_{r,s}} q_{i,\mathbf{k}} \Big| \\
&\le n^s \sum_{r=2}^{s} n^{-r-(\alpha/2-1-\varepsilon)\mathbb{1}_{\{r<s\}}} \cdot (Cs)^s \Big(\frac{1}{n-i}\Big)^{s/2+\lceil (s-r)/2 \rceil} \\
&\le (\widetilde{C}s)^s \Big(n^{-s/2} + \sum_{r=2}^{s-1} n^{s-r-(\alpha/2-1-\varepsilon)-s/2-\lceil (s-r)/2\rceil}\Big) \\
&= (\widetilde{C}s)^s \Big(n^{-s/2} + n^{s/2-(\alpha/2-1-\varepsilon)} \sum_{r=2}^{s-1} n^{-r-\lceil (s-r)/2\rceil}\Big) \\
&\le (\widetilde{C}s)^s \Big(n^{-s/2} + n^{s/2-(\alpha/2-1-\varepsilon)}\, s\, n^{-1-\lceil s/2\rceil}\Big) \\
&\le \widetilde{C}^s s^{s+1}\, n^{-1-(\alpha/2-1-\varepsilon)}
\end{aligned}
$$

with some constant $\widetilde{C} > 0$ that does not depend on $n$ or $s$. □

## Appendix B: Additional technical lemmas

The following lemmas are needed in the proof of our main result. Recall the matrix $Q_i = \{q_{i,kl}\}_{k,l=1}^n = P_i/(n-i)$, where the projection matrix $P_i = \mathbf{I}_n - B_{(i)}^\top (B_{(i)} B_{(i)}^\top)^{-1} B_{(i)}$ for $0 \le i \le p-1$ and $P_0 = \mathbf{I}_n$.

**Lemma B.1.** *Let $\alpha \in (3,4)$. Under the conditions of Theorem 2.1, there exists an $\varepsilon > 0$ such that, as $n \to \infty$,*

$$
n^{1-\alpha/2+\varepsilon}\Big|\sum_{i=0}^{p-1} \mathbb{E}[K_{2,2}^{(i)}]\Big| \to 0, \qquad n^{1-\alpha/2+\varepsilon}\Big|\sum_{i=0}^{p-1} \mathbb{E}[K_{2,2,2}^{(i)}]\Big| \to 0, \qquad n^{1-\alpha/2+\varepsilon}\Big|\sum_{i=0}^{p-1} \mathbb{E}[K_{2,2,2,2}^{(i)}]\Big| \to 0,
$$

*where $S_j^{(i)} = q_{i,11}^j + \cdots + q_{i,nn}^j$, $j \ge 1$, and $K_{2,2}^{(i)}, K_{2,2,2}^{(i)}, K_{2,2,2,2}^{(i)}$ are defined in (4.35), (4.36) and (4.37), respectively.*

**Proof.** Let's rewrite $\mathbb{E}[K_{2,2}^{(i)}]$, $\mathbb{E}[K_{2,2,2}^{(i)}]$ and $\mathbb{E}[K_{2,2,2,2}^{(i)}]$ in the following way

$$
\begin{aligned}
\mathbb{E}[K_{2,2}^{(i)}] &= -12n \sum_{\ell=1}^n \left(\mathbb{E}[q_{i,\ell\ell}^2] - \frac{1}{n^2}\right) + 8n^2 \sum_{\ell=1}^n \left(\mathbb{E}[q_{i,\ell\ell}^3] - \frac{1}{n^3}\right) - 2n^3 \sum_{\ell=1}^n \left(\mathbb{E}[q_{i,\ell\ell}^4] - \frac{1}{n^4}\right), \\
\mathbb{E}[K_{2,2,2}^{(i)}] &= 8n \sum_{\ell=1}^n \left(\mathbb{E}[q_{i,\ell\ell}^2] - \frac{1}{n^2}\right) - \frac{16n^2}{3} \sum_{\ell=1}^n \left(\mathbb{E}[q_{i,\ell\ell}^3] - \frac{1}{n^3}\right) + \frac{4n^3}{3} \sum_{\ell=1}^n \left(\mathbb{E}[q_{i,\ell\ell}^4] - \frac{1}{n^4}\right) + O(n^{-1}), \\
\mathbb{E}[K_{2,2,2,2}^{(i)}] &= -2n \sum_{\ell=1}^n \left(\mathbb{E}[q_{i,\ell\ell}^2] - \frac{1}{n^2}\right) + \frac{4n^2}{3} \sum_{\ell=1}^n \left(\mathbb{E}[q_{i,\ell\ell}^3] - \frac{1}{n^3}\right) - \frac{n^3}{3} \sum_{\ell=1}^n \left(\mathbb{E}[q_{i,\ell\ell}^4] - \frac{1}{n^4}\right) + O(n^{-1}),
\end{aligned}
$$

where we have used the fact that $|S_2^{(i)}| \le Cn^{-1}$ for some constant $C > 1$. The application of Lemma B.2 for $k = 2,3,4$ leads to

$$
n^{1-\alpha/2+\varepsilon}\Big|\sum_{i=0}^{p-1} \mathbb{E}[K_{2,2}^{(i)}]\Big| \le O\left(n^{1-\alpha/2+\varepsilon} n^{1/2}\right) = O\left(n^{(3-\alpha)/2+\varepsilon}\right).
$$

Similarly, we get

$$n^{1-\alpha/2+\varepsilon}\Big|\sum_{i=0}^{p-1}\mathbb{E}[K_{2,2,2}^{(i)}]\Big| = O\left(n^{(3-\alpha)/2+\varepsilon}\right),$$

$$n^{1-\alpha/2+\varepsilon}\Big|\sum_{i=0}^{p-1}\mathbb{E}[K_{2,2,2,2}^{(i)}]\Big| = O\left(n^{(3-\alpha)/2+\varepsilon}\right),$$

which verifies the statement of the lemma by noting that $\alpha > 3$.

□

**Lemma B.2.** *Under the conditions of Theorem 2.1, it holds for all $k \geq 2$ that*

$$0 \leq n^{k-2}\sum_{i=0}^{p-1}\sum_{\ell=1}^{n}\left(\mathbb{E}[q_{i,\ell\ell}^k] - \frac{1}{n^k}\right) \leq O(n^{-1/2}), \qquad n \to \infty. \tag{B.1}$$

**Proof.** First, using Jensen's inequality and the fact that $\sum_{\ell=1}^n q_{i,\ell\ell} = 1$ with $q_{i,ll} \geq 0$ we observe that

$$\frac{1}{n^k} = \left(\frac{1}{n}\sum_{\ell=1}^n q_{i,\ell\ell}\right)^k \leq \frac{1}{n}\sum_{\ell=1}^n q_{i,\ell\ell}^k,$$

which implies that

$$\frac{1}{n^{k-1}} \leq \sum_{\ell=1}^n q_{i,\ell\ell}^k. \tag{B.2}$$

Then, using this lower bound it follows that

$$\frac{p}{n^{k-1}} \leq \sum_{i=0}^{p-1}\sum_{\ell=1}^{n}\left(q_{i,\ell\ell} - \frac{1}{n} + \frac{1}{n}\right)^k = \frac{p}{n^{k-1}} + \sum_{i=0}^{p-1}\sum_{\ell=1}^{n}\sum_{j=0}^{k-1}\binom{k}{j}\left(q_{i,\ell\ell} - \frac{1}{n}\right)^{k-j}\frac{1}{n^j}$$

and, thus, taking expectations yields

$$\begin{aligned}
0 &\leq \sum_{i=0}^{p-1}\sum_{\ell=1}^{n}\mathbb{E}[q_{i,\ell\ell}^k] - \frac{p}{n^{k-1}} = \sum_{i=0}^{p-1}\sum_{\ell=1}^{n}\sum_{j=0}^{k-2}\binom{k}{j}\mathbb{E}\left(q_{i,\ell\ell} - \frac{1}{n}\right)^{k-j}\frac{1}{n^j} \\
&= \sum_{j=0}^{k-2}\sum_{i=0}^{p-1}\sum_{l=1}^{n}\binom{k}{j}\mathbb{E}\left(q_{i,\ell\ell} - \frac{1}{n}\right)^{k-j}\frac{1}{n^j},
\end{aligned}$$

where we have used for $j = k-1$ the property $\sum_{\ell=1}^n \mathbb{E}\left(q_{i,\ell\ell} - \frac{1}{n}\right) = 0$. Next we will show that for any $k \geq 2$

$$\sum_{i=0}^{p-1}\sum_{\ell=1}^{n}\mathbb{E}\left(q_{i,\ell\ell} - \frac{1}{n}\right)^k = O\left(n^{2-k-1/2}\right), \tag{B.3}$$

which will in fact imply that every term $\sum\limits_{i=0}^{p-1}\sum\limits_{\ell=1}^{n}\binom{k}{j}\mathbb{E}\left(q_{i,\ell\ell} - \frac{1}{n}\right)^{k-j}\frac{1}{n^j}$ will have the same order as the first one, i.e., for $j = 0$, and, thus, because $k$ is fixed, we will get

$$n^{k-2}\sum_{j=0}^{k-2}\sum_{i=0}^{p-1}\sum_{\ell=1}^{n}\binom{k}{j}\mathbb{E}\left(q_{i,\ell\ell} - \frac{1}{n}\right)^{k-j}\frac{1}{n^j} = O\left(n^{-1/2}\right).$$

We define for any $k \geq 2$

$$\delta_n := \delta_{n,k} := \sum_{i=0}^{p-1} \sum_{\ell=1}^{n} \mathbb{E}\left(q_{i,\ell\ell} - \mathbb{E}(q_{i,\ell\ell})\right)^k ,$$

where $\mathbb{E}(q_{i,\ell\ell}) = \frac{1}{n}$, which follows from the fact that $q_{i,\ell\ell}$ are identically distributed over $\ell$ and $q_{i,11} + \cdots + q_{i,nn} = 1$. Hence, it is enough to show that $\delta_n = O\big(n^{2-k-1/2}\big)$. Denote for $\ell = 1, \dots, n$ the vector $v_{\ell,i}$ as the $\ell$-th column of the matrix $B_{(i)}$. First, we note that for all $\ell = 1, \dots, n$ it holds

$$p_{i,\ell\ell} = 1 - v_{\ell,i}^\top (B_{(i)} B_{(i)}^\top)^{-1} v_{\ell,i} .$$

Denote now $\tilde{p}_{i,\ell\ell} = 1 - p_{i,\ell\ell}$ and use Minkowski's inequality to get

$$\begin{aligned}
\delta_n &= \sum_{i=0}^{p-1} \frac{n}{(n-i)^k} \mathbb{E}\left(p_{i,11} - \mathbb{E}(p_{i,11})\right)^k \leq \sum_{i=0}^{p-1} \frac{n}{(n-i)^k} \mathbb{E}\left|\tilde{p}_{i,11} - \mathbb{E}(\tilde{p}_{i,11})\right|^k \\
&= \sum_{i=0}^{p-1} \frac{n}{(n-i)^k} \mathbb{E}\left| v_{1,i}^\top (B_{(i)} B_{(i)}^\top)^{-1} v_{1,i} - \mathbb{E}(v_{1,i}^\top (B_{(i)} B_{(i)}^\top)^{-1} v_{1,i})\right|^k \overset{\text{Minkowski}}{\leq} C\left(\delta_n^{(1)} + \delta_n^{(2)} + \delta_n^{(3)}\right)
\end{aligned}$$

with some constant $C > 0$ possibly depending on $k$, whose value is not important and may change from line to line, and

$$\begin{aligned}
\delta_n^{(1)} &= \sum_{i=0}^{p-1} \frac{n}{(n-i)^k} \mathbb{E}\left| v_{1,i}^\top (B_{(i)} B_{(i)}^\top)^{-1} v_{1,i} - v_{1,i}^\top (B_{(i)} B_{(i)}^\top + \epsilon_n n \mathbf{I}_i)^{-1} v_{1,i}\right|^k , \\
\delta_n^{(2)} &= \sum_{i=0}^{p-1} \frac{n}{(n-i)^k} \mathbb{E}\left| v_{1,i}^\top (B_{(i)} B_{(i)}^\top + \epsilon_n n \mathbf{I}_i)^{-1} v_{1,i} - \frac{\mathbb{E}\operatorname{tr}(B_{(i,1)} B_{(i,1)}^\top + \epsilon_n n \mathbf{I}_i)^{-1}}{1 + \mathbb{E}\operatorname{tr}(B_{(i,1)} B_{(i,1)}^\top + \epsilon_n n \mathbf{I}_i)^{-1}}\right|^k , \\
\delta_n^{(3)} &= \sum_{i=0}^{p-1} \frac{n}{(n-i)^k} \left| \frac{\mathbb{E}\operatorname{tr}(B_{(i,1)} B_{(i,1)}^\top + \epsilon_n n \mathbf{I}_i)^{-1}}{1 + \mathbb{E}\operatorname{tr}(B_{(i,1)} B_{(i,1)}^\top + \epsilon_n n \mathbf{I}_i)^{-1}} - \mathbb{E}(v_{1,i}^\top (B_{(i)} B_{(i)}^\top)^{-1} v_{1,i})\right|^k ,
\end{aligned}$$

where $\epsilon_n$ is a sequence tending to zero arbitrarily slower than $1/n$ and $B_{(i,1)}$ denotes the matrix obtained from $B_{(i)}$ by deleting the 1st column $v_{1,i}$. Let's consider $\delta_n^{(1)}$ first. It holds

$$\begin{aligned}
&\left| v_{1,i}^\top (B_{(i)} B_{(i)}^\top)^{-1} v_{1,i} - v_1^\top (B_{(i)} B_{(i)}^\top + \epsilon_n n \mathbf{I}_i)^{-1} v_{1,i}\right| = \epsilon_n n \cdot v_{1,i}^\top (B_{(i)} B_{(i)}^\top)^{-1} (B_{(i)} B_{(i)}^\top + \epsilon_n n \mathbf{I}_i)^{-1} v_{1,i} \\
&\quad \leq \frac{\epsilon_n n}{\lambda_{min}(B_{(i)} B_{(i)}^\top + \epsilon_n n \mathbf{I}_i)} \underbrace{v_{1,i}^\top (B_{(i)} B_{(i)}^\top)^{-1} v_{1,i}}_{=\tilde{p}_{i,11} \leq 1} \\
&\quad \leq \frac{\epsilon_n n}{\lambda_{min}(B_{(i)} B_{(i)}^\top)} \sim \frac{\epsilon_n}{(1 - \sqrt{\frac{i}{n}})^2} \leq \frac{\epsilon_n}{(1 - \sqrt{\frac{p}{n}})^2} \leq C\epsilon_n .
\end{aligned} \tag{B.4}$$

Thus, for $\delta_n^{(1)}$ and sufficiently large $n$, we have

$$\delta_n^{(1)} \leq C^k \sum_{i=0}^{p-1} \frac{n}{(n-i)^k} \epsilon_n^k \leq \underbrace{\frac{np}{(n-p+1)^2}}_{=O(1)} O(\epsilon_n^k n^{-k+2}) = O(\epsilon_n^k n^{-(k-2)}) .$$

Now we proceed to $\delta_n^{(2)}$. Let's consider the following expression

$$\mathbb{E}\left| v_{1,i}^\top (B_{(i)} B_{(i)}^\top + \epsilon_n n \mathbf{I}_i)^{-1} v_{1,i} - \frac{\mathbb{E}\operatorname{tr}(B_{(i,1)} B_{(i,1)}^\top + \epsilon_n n \mathbf{I}_i)^{-1}}{1 + \mathbb{E}\operatorname{tr}(B_{(i,1)} B_{(i,1)}^\top + \epsilon_n n \mathbf{I}_i)^{-1}}\right|^k$$

$$= \mathbb{E}\left|\frac{v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{1,i}}{1 + v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{1,i}} - \frac{\mathbb{E}\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1}}{1 + \mathbb{E}\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1}}\right|^k$$

$$= \mathbb{E}\frac{\left|v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{1,i} - \mathbb{E}\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1}\right|^k}{(1 + \mathbb{E}\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1})^k (1 + v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{1,i})^k}$$

$$\leq \mathbb{E}\left|v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{1,i} - \mathbb{E}\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1}\right|^k$$

$$\leq C\left(\mathbb{E}\left|v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{1,i} - \operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1}\right|^k \right. \tag{B.5}$$

$$\left. + \mathbb{E}\left|\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} - \mathbb{E}\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1}\right|^k\right). \tag{B.6}$$

First, we analyze the term in (B.6) and define $\mathbb{E}_\ell := \mathbb{E}(\cdot|v_{\ell,i},\dots,v_{n,i})$ for $\ell = 1,\dots,n$ and $\mathbb{E}_{n+1} := \mathbb{E}$. It holds

$$\begin{aligned}
&\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} - \mathbb{E}\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1}\\
&= \sum_{\ell=2}^n (\mathbb{E}_\ell - \mathbb{E}_{\ell+1}) \operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1}\\
&= \sum_{\ell=2}^n (\mathbb{E}_\ell - \mathbb{E}_{\ell+1}) \left(\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} - \operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top - v_{\ell,i}v_{\ell,i}^\top + \epsilon_n n\mathbf{I}_i)^{-1}\right),
\end{aligned} \tag{B.7}$$

where the properties $\mathbb{E}_\ell(\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top - v_{\ell,i}v_{\ell,i}^\top + \epsilon_n n\mathbf{I}_i)^{-1}) = \mathbb{E}_{\ell+1}(\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top - v_{\ell,i}v_{\ell,i}^\top + \epsilon_n n\mathbf{I}_i)^{-1})$ and $\mathbb{E}_2(\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1}) = \operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1}$ were used. Together with the definition of the martingale differences sequence and Sherman-Morrison formula it implies

$$\begin{aligned}
&\mathbb{E}\left|\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} - \mathbb{E}\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1}\right|^k\\
&\leq C\sum_{\ell=2}^n \mathbb{E}\left|(\mathbb{E}_\ell - \mathbb{E}_{\ell+1})\left(\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} - \operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top - v_{\ell,i}v_{\ell,i}^\top + \epsilon_n n\mathbf{I}_i)^{-1}\right)\right|^k\\
&= C\sum_{\ell=2}^n \mathbb{E}\left|(\mathbb{E}_\ell - \mathbb{E}_{\ell+1})\frac{v_{\ell,i}^\top (B_{(i,1)}B_{(i,1)}^\top - v_{\ell,i}v_{\ell,i}^\top + \epsilon_n n\mathbf{I}_i)^{-2} v_{\ell,i}}{1 + v_{\ell,i}^\top (B_{(i,1)}B_{(i,1)}^\top - v_{\ell,i}v_{\ell,i}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{\ell,i}}\right|^k\\
&\leq \frac{C}{\epsilon_n^k n^k}\sum_{\ell=2}^n \mathbb{E}\left|(\mathbb{E}_\ell - \mathbb{E}_{\ell+1})\underbrace{\frac{v_{\ell,i}^\top (B_{(i,1)}B_{(i,1)}^\top - v_{\ell,i}v_{\ell,i}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{\ell,i}}{1 + v_{\ell,i}^\top (B_{(i,1)}B_{(i,1)}^\top - v_{\ell,i}v_{\ell,i}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{\ell,i}}}_{\leq 1}\right|^k\\
&\leq \frac{C(n-1)2^k}{\epsilon_n^k n^k} = O(\epsilon_n^{-k} n^{-(k-1)}).
\end{aligned} \tag{B.8}$$

Next, we turn to (B.5). To this end, we show that $v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{1,i}$ is bounded. Using the Sherman-Morrison formula we get

$$\begin{aligned}
&v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{1,i} = v_{1,i}^\top (B_{(i)}B_{(i)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{1,i} + \frac{(v_{1,i}^\top (B_{(i)}B_{(i)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{1,i})^2}{1 - v_{1,i}^\top (B_{(i)}B_{(i)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{1,i}}\\
&= \frac{v_{1,i}^\top (B_{(i)}B_{(i)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{1,i}}{1 - v_{1,i}^\top (B_{(i)}B_{(i)}^\top + \epsilon_n n\mathbf{I}_i)^{-1} v_{1,i}} \leq \frac{1}{1-\kappa}
\end{aligned}$$

because $v_{1,i}^\top(B_{(i)}B_{(i)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}\le\kappa$ for some $\kappa<1$, for which Lemma A.1 in the Appendix of [2] was used. Thus $0\le v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}\le C$, so it is enough to consider the case $k=2$. Indeed, let $k\ge 3$, then we get

$$\begin{aligned}
&\mathbb{E}\left|v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}-\operatorname{tr}(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}\right|^k\\
&=\mathbb{E}\left|v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}-\mathbb{E}_2v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}\right|^2\\
&\times\left|v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}-\mathbb{E}_2v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}\right|^{k-2}\\
&\le C\cdot\mathbb{E}\left|v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}-\mathbb{E}_2v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}\right|^2. 
\end{aligned}\tag{B.9}$$

Further we truncate the elements of the vector $v_{1,i}$ at the level $\zeta_n\sqrt{n}$, i.e., denote $\hat v_{1j,i}=v_{1j,i}\cdot\mathbb{1}\{|v_{1j,i}|\le\zeta_n\sqrt{n}\}$ with $\zeta_n$ arbitrarily slow converging to zero but no faster than $n^{-1/2}$, i.e., $\zeta_n\sqrt{n}\to+\infty$. Note that $\hat v_{1j,i}$ has mean zero since $v_{1j,i}$ is symmetrically distributed. Because the third absolute moment of $v_{1j,i}$ is finite it holds

$$\lim_{n\to\infty}\frac{\mathbb{E}|v_{1j,i}|^3\mathbb{1}\{|v_{1j,i}|>\zeta_n\sqrt{n}\}}{\zeta_n^3}=0\,.$$

Moreover, one can also check that the second moment of $\hat v_{1j,i}$ converges to 1, indeed let $\operatorname{Var}(\hat v_{11,i})=\sigma_n^2$ then

$$|\sigma_n^2-1|\le C\mathbb{E}(|v_{11,i}|^2\mathbb{1}\{|v_{11,i}|>\zeta_n\sqrt{n}\})\le C\zeta_n^2 n\frac{\mathbb{E}(|v_{11,i}|^3\mathbb{1}\{|v_{11,i}|>\zeta_n\sqrt{n}\})}{\zeta_n^3n^{3/2}}=o\left(n^{-1/2}\right).\tag{B.10}$$

Then the difference between the truncated and original quadratic forms is given by

$$\begin{aligned}
&\mathbb{E}\left|v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}-\hat v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}\hat v_{1,i}\right|^2\\
&=\mathbb{E}\left|(v_{1,i}-\hat v_{1,i})^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}(v_{1,i}+\hat v_{1,i})\right|^2\\
&\overset{\text{CS}}{\le} C\mathbb{E}(v_{1,i}-\hat v_{1,i})^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}(v_{1,i}-\hat v_{1,i})\cdot(v_{1,i}+\hat v_{1,i})^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}(v_{1,i}+\hat v_{1,i})\\
&\le C\mathbb{E}(v_{1,i}-\hat v_{1,i})^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}(v_{1,i}-\hat v_{1,i})\le C\mathbb{E}\frac{(v_{1,i}-\hat v_{1,i})^\top(v_{1,i}-\hat v_{1,i})}{\lambda_{min}(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)}\\
&\le\frac{C}{n}\mathbb{E}\left(\sum_{j=1}^n v_{1j,i}^2\mathbb{1}\{|v_{1j,i}|>\zeta_n\sqrt{n}\}\right)=C\mathbb{E}\left(|v_{11,i}|^2\mathbb{1}\{|v_{11,i}|>\zeta_n\sqrt{n}\}\right)\\
&\le C\zeta_n^2n\frac{\mathbb{E}\left(|v_{11,i}|^3\mathbb{1}\{|v_{11,i}|>\zeta_n\sqrt{n}\}\right)}{\zeta_n^3n^{3/2}}=o(n^{-1/2}).
\end{aligned}\tag{B.11}$$

Thus, we can safely replace $v_{1,i}$ by $\hat v_{1,i}$ in (B.9). We recall that $X_{11}$ is regularly varying with index $\alpha\in(3,4)$, i.e., $\mathbb{P}(|X_{11}>x)=x^{-\alpha}L(x)$ for a slowly varying function $L$. The following formula for truncated moments of $X_{11}$ is well–known (see, for instance, [9])

$$\hat\nu_4:=\mathbb{E}[|X_{11}|^4\mathbb{1}\{|X_{11}|>\zeta_n\sqrt{n}\}]\sim\frac{\alpha}{4-\alpha}\zeta_n^{4-\alpha}n^{(4-\alpha)/2}L(\zeta_n\sqrt{n}),\quad n\to\infty\,.$$

Consider now (B.9)

$$\begin{aligned}
&\mathbb{E}\left|v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}-\mathbb{E}_2v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}\right|^2\\
&\le C\underbrace{\mathbb{E}\left|v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}-\hat v_{1,i}^\top(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}\hat v_{1,i}\right|^2}_{o(n^{-1/2})\text{ by (B.11)}}
\end{aligned}$$

$$+\,C\,\mathbb{E}\underbrace{\left|\mathbb{E}_2 v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}-\mathbb{E}_2\hat v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}\hat v_{1,i}\right|^2}_{o(n^{-1})\text{ by (B.10)}}$$

$$+\,C\mathbb{E}\left|\hat v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}\hat v_{1,i}-\mathbb{E}_2\hat v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}\hat v_{1,i}\right|^2 \tag{B.12}$$

and apply Lemma B.26 from [4] on the last summand in (B.12)

$$\begin{aligned}&\mathbb{E}\left|\hat v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}\hat v_{1,i}-\mathbb{E}_2\hat v_{1,i}^\top (B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}\hat v_{1,i}\right|^2\\ &\le C\hat\nu_4\mathbb{E}\,\mathrm{tr}(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-2}\le \frac{C}{n}\hat\nu_4\\ &\sim \frac{C}{n}\zeta_n^{4-\alpha}n^{(4-\alpha)/2}L(\zeta_n\sqrt{n})=o(n^{-1/2}),\qquad n\to\infty,\end{aligned}$$

where in the last step we used the Potter bounds for the slowly varying function $L$.

Thus, similarly as for $\delta_n^{(1)}$ we get

$$n^{k-2}\delta_n^{(2)}=O(\epsilon_n^{-k}n^{-(k-1)})+o(n^{-1/2}). \tag{B.13}$$

Concerning $\delta_n^{(3)}$, we observe the following

$$\begin{aligned}&\left|\frac{\mathbb{E}\,\mathrm{tr}(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}}{1+\mathbb{E}\,\mathrm{tr}(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}}-\mathbb{E}(v_{1,i}^\top(B_{(i)}B_{(i)}^\top)^{-1}v_{1,i})\right|^k\\ \le\quad & C\left|\mathbb{E}(v_{1,i}^\top(B_{(i)}B_{(i)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}-\frac{\mathbb{E}\,\mathrm{tr}(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}}{1+\mathbb{E}\,\mathrm{tr}(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}})\right|^k\\ +\quad & C\underbrace{|\mathbb{E}(v_{1,i}^\top(B_{(i)}B_{(i)}^\top)^{-1}v_{1,i}-\mathbb{E}(v_{1,i}^\top(B_{(i)}B_{(i)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}|^k}_{=O(\epsilon_n^k)\text{ due to (B.4)}}\\ \overset{Jensen}{\le}\quad & \underbrace{C\,\mathbb{E}\left|(v_{1,i}^\top(B_{(i)}B_{(i)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}v_{1,i}-\frac{\mathbb{E}\,\mathrm{tr}(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}}{1+\mathbb{E}\,\mathrm{tr}(B_{(i,1)}B_{(i,1)}^\top+\epsilon_n n\mathbf{I}_i)^{-1}})\right|^k}_{=O(\epsilon_n^{-k}n^{-(k-1)})+o(n^{-1/2})\text{ due to (B.13)}}+O(\epsilon_n^k)\end{aligned}$$

and, as a result, we have

$$n^{k-2}\delta_n^{(3)}=O(\epsilon_n^k)+O(\epsilon_n^{-k}n^{-(k-1)})+o(n^{-1/2}) \tag{B.14}$$

and altogether we receive the following rate for $\delta_n$

$$n^{k-2}\delta_n=O(\epsilon_n^k)+O(\epsilon_n^{-k}n^{-(k-1)})+o(n^{-1/2}). \tag{B.15}$$

Now we need to specify the sequence $\epsilon_n$ such that $\delta_n$ converges to zero as fast as possible. Because $\epsilon_n$ can not vanish faster than $1/n$ we assume w.l.o.g. that $\epsilon_n=n^{-\varepsilon}$ for some $0<\varepsilon<1$, plug it into (B.15) and get

$$n^{k-2}\delta_n=O(n^{-k\varepsilon})+O(n^{k(\varepsilon-1)+1})+o(n^{-1/2}). \tag{B.16}$$

Choosing $\varepsilon=\frac{1}{2k}$ finishes the proof of the lemma. □

**Lemma B.3.** *[16, Lemma 7.10] Let $X_1,\dots,X_N$ be independent centered random variables and assume that*

$$(\mathbb{E}[|X_i|^s])^{1/s}\le\mu_s,\quad 1\le i\le N; s=2,3,\dots$$

*for some fixed constants $\mu_s$. Then we have for any deterministic complex numbers $a_{ij}, 1 \le i, j \le N$ that*

$$\Big(\mathbb{E}\Big[\Big|\sum_{i\neq j=1}^{N} a_{ij}X_iX_j\Big|^s\Big]\Big)^{1/s} \le C\, s\, \mu_s^2 \Big(\sum_{i\neq j=1}^{N} |a_{ij}|^2\Big)^{1/2}, \quad s = 2, 3, \ldots, \tag{B.17}$$

*where the constant $C$ does not depend on $s$.*

**Lemma B.4.** *[40, Theorem b) and d)] Let $\mathbf{z} = (Z_1, \ldots, Z_n)^\top$ be a random vector such that, for all nonnegative integers $m_1, \ldots, m_6$ with $m_1 + \cdots + m_6 \le 6$, $\mathbb{E}[Z_{i_1}^{m_1} Z_{i_2}^{m_2} \cdots Z_{i_6}^{m_6}]$ is (i) finite; (ii) zero if any $m_i$ is odd; and (iii) invariant under permutations of the indices. Let $\beta_2 = \mathbb{E}[Z_1^2], \beta_{2,2} = \mathbb{E}[Z_1^2 Z_2^2], \beta_4 = \mathbb{E}[Z_1^4], \beta_{4,2} = \mathbb{E}[Z_1^4 Z_2^2]$ and $\beta_6 = \mathbb{E}[Z_1^6]$. Then we have for any real-valued and symmetric $n \times n$ nonrandom matrix $\mathbf{A}$ that*

$$\begin{aligned}
\mathbb{E}[(\mathbf{z}^\top \mathbf{A}\mathbf{z} - \mathbb{E}[\mathbf{z}^\top \mathbf{A}\mathbf{z}])^3] &= 8\beta_{2,2,2}\operatorname{tr}(\mathbf{A}^3) + (\beta_{2,2,2} + 2\beta_2^3 - 3\beta_2\beta_{2,2})(\operatorname{tr}\mathbf{A})^3 \\
&+ 6(\beta_{2,2,2} - \beta_2\beta_{2,2})\operatorname{tr}\mathbf{A}\operatorname{tr}(\mathbf{A}^2) + 3(\beta_{4,2} - \beta_4\beta_2 + 3\beta_{2,2}\beta_2 - 3\beta_{2,2,2})\operatorname{tr}\mathbf{A}\operatorname{tr}(\mathbf{A}\circ\mathbf{A}) \\
&+ 12(\beta_{4,2} - 3\beta_{2,2,2})\operatorname{tr}(\mathbf{A}\circ\mathbf{A}^2) + (\beta_6 - 15\beta_{4,2} + 30\beta_{2,2,2})\operatorname{tr}(\mathbf{A}\circ\mathbf{A}\circ\mathbf{A}),
\end{aligned}$$

*where $\circ$ denotes the Hadamard product. Moreover,*

$$\begin{aligned}
\mathbb{E}[(\mathbf{z}^\top \mathbf{A}\mathbf{z})^3] &= \beta_{2,2,2}[(\operatorname{tr}\mathbf{A})^3 + 6\operatorname{tr}\mathbf{A}\operatorname{tr}(\mathbf{A}^2) + 8\operatorname{tr}(\mathbf{A}^3)] + (\beta_6 - 15\beta_{4,2} + 30\beta_{2,2,2})\operatorname{tr}(\mathbf{A}\circ\mathbf{A}\circ\mathbf{A}) \\
&\quad + (\beta_{4,2} - 3\beta_{2,2,2})[3\operatorname{tr}\mathbf{A}\operatorname{tr}(\mathbf{A}\circ\mathbf{A}) + 12\operatorname{tr}(\mathbf{A}\circ\mathbf{A}^2)].
\end{aligned}$$

*If $\mathbf{B}$ is another real-valued and symmetric $n \times n$ nonrandom matrix, one has*

$$\mathbb{E}[\mathbf{z}^\top \mathbf{A}\mathbf{z}\mathbf{z}^\top \mathbf{B}\mathbf{z}] = \beta_{2,2}[\operatorname{tr}\mathbf{A}\operatorname{tr}\mathbf{B} + 2\operatorname{tr}(\mathbf{A}\mathbf{B})] + (\beta_4 - 3\beta_{2,2})\operatorname{tr}(\mathbf{A}\circ\mathbf{B}).$$

## References


[1] ALBRECHER, H. and TEUGELS, J. L. (2007). Asymptotic analysis of a measure of variation. *Theory of Probability and Mathematical Statistics* **74** 1–10.

[2] ANATOLYEV, S. and YASKOV, P. (2017). Asymptotics of diagonal elements of projection matrices under many instruments/regressors. *Econometric Theory* **33** 717–738. MR3637973

[3] ANDERSON, T. W. (2003). *An introduction to multivariate statistical analysis*. John Wiley & Sons, New Jersey.

[4] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, second ed. *Springer Series in Statistics*. Springer, New York. MR2567175 (2011d:60014)

[5] BAI, Z. and ZHOU, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. MR2411613

[6] BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large dimensional sample covariance matrices. *Annals of Probability* **32** 553-605.

[7] BAO, Z., PAN, G. and ZHOU, W. (2012). Tracy-Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.* **17** 32 pp.

[8] BAO, Z., PAN, G. and ZHOU, W. (2015). The logarithmic law of random determinant. *Bernoulli* **21** 1600–1628.

[9] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation*. *Encyclopedia of Mathematics and its Applications* **27**. Cambridge University Press, Cambridge. MR898871 (88i:26004)

[10] BODNAR, T., DETTE, H. and PAROLYA, N. (2019). Testing for Independence of Large Dimensional Vectors. *Annals of Statistics* **47** 2977-3008.

[11] DEMBO, A. (1989). On random determinants. *Quarterly of Applied Mathematics* **47** 185–195.

[12] DETTE, H. and DÖRNEMANN, N. (2020). Likelihood ratio tests for many groups in high dimensions. *J. Multivariate Anal.* **178** 104605, 16. MR4079037

[13] DÖRNEMANN, N. (2023). Likelihood ratio tests under model misspecification in high dimensions. *J. Multivariate Anal.* **193** Paper No. 105122. MR4504582

[14] EL KAROUI, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714. MR2308592

[15] EL KAROUI, N. (2009). Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.* **19** 2362–2405. MR2588248

[16] ERDÖS, L. and YAU, H.-T. (2017). *A dynamical approach to random matrix theory*. *Courant Lecture Notes in Mathematics* **28**. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI; Available at http://www.math.harvard.edu/~htyau/RM-Aug-2016.pdf. MR3699468

[17] FUCHS, A., JOFFE, A. and TEUGELS, J. (2001). Expectation of the ratio of the sum of squares to the square of the sum: exact and asymptotic results. *Teor. Veroyatnost. i Primenen.* **46** 297–310. MR1968687

[18] GAO, J., HAN, X., PAN, G. and YANG, Y. (2017). High dimensional correlation matrices: the central limit theorem and its applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79** 677-693.

[19] GIRKO, V. (1979). Central limit theorem for random determinants. In *Theory of Probability & Its Applications* **23** 846–846. SIAM.

[20] GIRKO, V. L. (1998). A refinement of the central limit theorem for random determinants. *Theory of Probability & Its Applications* **42** 121–129.
[21] GOODMAN, N. R. (1963). The Distribution of the Determinant of a Complex Wishart Distributed Matrix. *The Annals of Mathematical Statistics* **34** 178 – 180.
[22] GROTE, J., KABLUCHKO, Z. and THÄLE, C. (2019). Limit theorems for random simplices in high dimensions. *ALEA Lat. Am. J. Probab. Math. Stat.* **16** 141–177. MR3903027
[23] HALL, P. and HEYDE, C. C. (1980). *Martingale limit theory and its application*. *Probability and Mathematical Statistics*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London. MR624435
[24] HEINY, J. (2022). Large sample correlation matrices: a comparison theorem and its applications. *Electron. J. Probab.* **27** Paper No. 94, 20. MR4456777
[25] HEINY, J., JOHNSTON, S. and PROCHNO, J. (2022). Thin-shell theory for rotationally invariant random simplices. *Electron. J. Probab.* **27** 1-41.
[26] HEINY, J. and MIKOSCH, T. (2018). Almost sure convergence of the largest and smallest eigenvalues of high-dimensional sample correlation matrices. *Stochastic Processes and their Applications* **128** 2779-2815.
[27] HEINY, J. and YAO, J. (2022). Limiting distributions for eigenvalues of sample correlation matrices from heavy-tailed populations. *Ann. Statist.* **50** 3249–3280. MR4524496
[28] JIANG, T. (2004). The Limiting Distributions of Eigenvalues of Sample Correlation Matrices. *Sankhyā: The Indian Journal of Statistics (2003-2007)* **66** 35–48.
[29] JIANG, T. and YANG, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.* **41** 2029–2074.
[30] MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)* **72** **(114)** 507–536. MR0208649
[31] MOHAMMADI, M. (2016). On the bounds for diagonal and off-diagonal elements of the hat matrix in the linear regression model. *REVSTAT Statistical Journal* **14** 75–87.
[32] MORALES-JIMENEZ, D., JOHNSTONE, I. M., MCKAY, M. R. and YANG, J. (2021). Asymptotics of eigenstructure of sample correlation matrices for high-dimensional spiked models. *Statist. Sinica* **31** 571–601. MR4286186
[33] NGUYEN, H. H. and VU, V. (2014). Random matrices: Law of the determinant. *Ann. Probab.* **42** 146–167.
[34] NIELSEN, J. (1999). The distribution of volume reductions induced by isotropic random projections. *Advances in Applied Probability* **31** 985–994.
[35] PAROLYA, N., HEINY, J. and KUROWICKA, D. (2021). Logarithmic law of large random correlation matrix. *arXiv preprint arXiv:2103.13900*.
[36] SILVERSTEIN, J. W. and CHOI, S. I. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *Journal of Multivariate Analysis* **54** 295-309.
[37] TAO, T. and VU, V. (2012). A central limit theorem for the determinant of a Wigner matrix. *Advances in Mathematics* **231** 74-101.
[38] WANG, Q. and YAO, J. (2013). On the sphericity test with large-dimensional observations. *Electron. J. Statist.* **7** 2164–2192.
[39] WANG, X., HAN, X. and PAN, G. (2018). The logarithmic law of sample covariance matrices near singularity. *Bernoulli* **24** 80–114.
[40] WIENS, D. P. (1992). On moments of quadratic forms in non-spherically distributed variables. *Statistics* **23** 265–270. MR1237804
[41] YANG, X., ZHENG, X. and CHEN, J. (2021). Testing high-dimensional covariance matrices under the elliptical distribution and beyond. *Journal of Econometrics* **221** 409-423.
[42] YAO, J., BAI, Z. and ZHENG, S. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis (No. 39)*. Cambridge University Press, New York.