Filomat 40:2 (2026), 397–413
https://doi.org/10.2298/FIL2602397T

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

On the asymptotic of likelihood ratio statistic in high-dimensional exploratory factor analysis

Ping Tang[a], **Junshan Xie**[a,*]

[a] *School of Mathematics and Statistics, Henan University, China*

Abstract. In this paper, the asymptotical properties on the likelihood ratio test in high-dimensional exploratory factor analysis are considered. When the dimension of the response variable p satisfies $p = p(N) \to \infty$ and $p/N \to c \in (0, 1)$ as the sample size $N \to \infty$, the Edgeworth expansion of the null distribution of the likelihood ratio test statistic and its uniform error bound are established. Some numerical simulations indicate that the proposed approximation is more accurate than the traditional chi-square approximate method on dealing with the high-dimensional test.

1. Introduction

Exploratory factor analysis is a useful statistical dimension reduction method, which achieves dimensionality reduction by exploring the low-dimensional latent structure underlying the observed data. In exploratory factor analysis, the true number of common latent factors is usually unknown, how to determine the number of latent common factors is a critical issue on this topic. There are lots of criteria and methods have been investigated for determining the number of the latent common factors. For instance, the most widely used procedure is the eigenvalues-greater-than one rule, that is Kaiser criterion (Guttman, 1954 [17]) ; Kaiser, 1960 [13]), the scree test (Cattell, 1966 [7]), the parallel analysis method (Horn, 1965 [11]; Keeling, 2000 [15]; Dobriban, 2020 [8]), the likelihood ratio test(Bartlett, 1950 [5]; 1951 [6]; Jöreskog, 1967 [12]; Anderson, 2003 [4], Kentaro, 2007 [16]), Akaike's information criterion (Akaike, 1973 [1], 1987 [2]), BIC (Schwarz, 1978 [20]) and so on.

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_N$ be the random sample of size N from the p-dimensional random vector $\boldsymbol{X}$. The exploratory factor analysis considers the following common factor model

$$\boldsymbol{X}_i = \mu + \Lambda F_i + U_i, \quad i = 1, 2, \cdots, N, \tag{1.1}$$

where μ is a p-dimensional mean value vector, Λ is a $p \times k$ loading matrix with $rank(\Lambda) = k < p$, F_i is a k-dimensional latent random vector containing the common factors, $E(F_i) = 0_k$, $Cov(F_i) = I_k$, where 0_k denotes a k-dimensional all-zero vector, I_k represents a $k \times k$ identity matrix, and k is the number of the

2020 *Mathematics Subject Classification*. Primary 62E20; Secondary 62H15.
Keywords. Edgeworth expansion, Error bound, Exploratory factor analysis, Likelihood ratio test.
Received: 20 December 2024; Accepted: 08 November 2025
Communicated by Dragan S. Djordjević
Research supported by Natural Science Foundation of Henan Province of China (Grant No.242300421379).
* Corresponding author: Junshan Xie
Email addresses: tangping8896@126.com (Ping Tang), junshan@henu.edu.cn (Junshan Xie)
ORCID iDs: https://orcid.org/0009-0007-5257-5833 (Ping Tang), https://orcid.org/0000-0003-3230-4933 (Junshan Xie)

common latent factors. U_i is a p-dimensional error vector with $E(U_i) = 0_p$, and $Cov(U_i) = \Psi$, where Ψ is a $p \times p$ positive definite diagonal matrix with $rank(\Psi) = p$. We assume that F_i and U_i are uncorrelated and $F_i \sim N_k(0_k, I_k)$ and $U_i \sim N_p(0_p, \Psi)$, then the population $X \sim N_p(\mu_p, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \Lambda\Lambda^\top + \Psi$ and $\Lambda^\top$ is the transpose of Λ.

The likelihood ratio test method can be used to estimate the true common factor number (see Kentaro, 2007 [16] and He et al. (2021) [10]). In particular, for each $k = 0, 1, \cdots, p$, we can consider the test

$$\mathrm{H}_{0,k} : \Sigma = \Lambda\Lambda^\top + \Psi \text{ with (at most) } k \text{ factors } \ \mathbf{vs} \ \ \mathrm{H}_{A,k} : \Sigma \ \text{ is any positive definite matrix.}$$

As a matter of convenience, we assume that the true factor number $k \geq 1$ is given. Under the k-factor model, we write $\Lambda = \Lambda_k$ and $\Psi = \Psi_k$, where Λ_k and Ψ_k are also given matrixes and $rank(\Lambda_k) = k < p$. Denote $\Sigma_k = \Lambda_k\Lambda_k^T + \Psi_k$. Then we will consider the following test

$$\mathrm{H'}_{0,k} : \Sigma = \Sigma_k \quad \mathbf{vs} \quad \mathrm{H'}_{A,k} : \Sigma \neq \Sigma_k. \tag{1.2}$$

A typical forward stepwise sequentially test procedure can be stated as follows. Firstly, we consider $k = 0$ and examine $\mathrm{H'}_{0,0}$ versus $\mathrm{H'}_{A,0}$ using the likelihood ratio test. If $\mathrm{H'}_{0,0}$ is rejected, we then consider $k = 1$ and examine the 1-factor model test $\mathrm{H'}_{0,1}$ versus $\mathrm{H'}_{A,1}$. If $\mathrm{H'}_{0,1}$ is rejected, we then consider $k = 2$ and examine the 2-factor model test $\mathrm{H'}_{0,2}$. The test procedure continues until we fail to reject $\mathrm{H'}_{0,\hat{k}}$ for some $\hat{k}$. Then $\hat{k}$ can be the estimation of the true number of factors. In this sense, the statistical asymptotic properties of the likelihood ratio test also have good research significance in the exploratory factor analysis.

By Muirhead (1982) [19], the likelihood ratio test statistic under $H'_{0,k}$ can be written as

$$T' = -(N-1)\log(|\hat{\Sigma}| \times |\Sigma_k|^{-1}) + (N-1)\Big[tr(\hat{\Sigma}\Sigma_k^{-1}) - p\Big], \tag{1.3}$$

where $\hat{\Sigma}$ is the unbiased sample covariance matrix of the observations $X_i, i = 1, 2, \cdots, N$.

When the dimension p is fixed and the sample size N is large, He et al. (2021) [10] obtain that

$$T' \xrightarrow{d} \chi^2_f, \tag{1.4}$$

where $\xrightarrow{d}$ represents the convergence in distribution and $f = p(p+1)/2$. The Bartlett correction provided a re-scaling strategy that further improves the infinite-sample accuracy of the chi-squared approximation of the likelihood ratio test statistic. When the dimension p is fixed and the sample size N is large, He et al. (2021) [10] also obtain the following chi-square approximation for the Bartlett correction test statistic $\rho T'$:

$$\rho T' \xrightarrow{d} \chi^2_f. \tag{1.5}$$

where the Bartlett correction coefficient $\rho = 1 - [6(N-1)(p+1)]^{-1}(2p^2 + 3p - 1)$.

However, in high-dimensional setting with the large dimension p and the large sample size n, researchers have found that the chi-squared approximation for the likelihood ratio statistic often becomes inaccurate, resulting in the failure of the corresponding likelihood ratio tests. To address this issue, some alternative approximations of the high-dimensional likelihood ratio statistic have been discussed. A remarkable work is due to He et al. (2021) [10] , who proved that the likelihood ratio statistic T' was asymptotically Gaussian distributed under the assumption of $N \geq p+2$, $p \to \infty$, $p/n \to c \in [0,1]$, as $n = N - 1 \to \infty$, which means the dimension p is allowed to diverge with the sample size N, their result can be listed as follows.

Lemma 1.1. *Suppose $N \geq p+2$, $p \to \infty$, $\frac{p}{n} \to c \in [0,1]$, as $n = N-1 \to \infty$, under $\mathbf{H}'_{0,k}$, we have*

$$\frac{T' + n\tilde{\mu}_n}{n\tilde{\sigma}_n} \xrightarrow{d} N(0,1),$$

where

$$\tilde{\mu}_n = -p + (p - n + \frac{1}{2})\log(1 - \frac{p}{n}), \quad \tilde{\sigma}_n^2 = -2\Big[\frac{p}{n} + \log(1 - \frac{p}{n})\Big].$$

The result only gives the asymptotic distribution of the likelihood ratio statistics in high-dimension data. In fact, the more precise results including the asymptotic expansion and convergence rate of the likelihood ratio statistics are also worth to study.

In this paper, we will not only obtain the asymptotic expansion of the likelihood ratio statistics T', but also obtain the convergence rate and the uniform error bound of the null distribution of the likelihood ratio statistic by Edgeworth expansion method, which is a popular method of dealing with the limiting distribution for the high-dimensional statistics, one can refer to Fujikoshi (2000) [9], Fujikoshi et al. (2010) [14], Wakaki (2010) [23], Wakaki et al. (2010) [3], Yamada (2012) [24], Mitsui et al. (2015) [18] and Sun and Xie (2021, 2022) [21][22].

The rest of the paper is organized as follows. Section 2 gives the edgeworth expansion on the null distribution of the likelihood ratio test statistic and its uniform error bound. The empirical performance of the proposed method by numerical simulations will be investigated in Section 3. At last, some technical proofs are listed in Section 4.

2. Main results

Denote $\psi(x) = \frac{d}{dx}\log\Gamma(x)$ and

$$\psi^{(s)}(a) = (\tfrac{d}{dx})^s\psi(x)|_{x=a} = \sum_{k=0}^{\infty}\frac{(-1)^{s+1}s!}{(a+k)^{s+1}}, \quad s = 1,2,\cdots. \tag{2.6}$$

Let

$$\Gamma(z,a) = \int_a^{\infty} t^{z-1}e^{-t}dt. \tag{2.7}$$

In this section, we provide the main results of this paper under the assumption $p = p(n) \to \infty$, $p/n \to c \in (0,1)$ as $n = N-1 \to \infty$. Now we will first calculate the high-dimensional Edgeworth expansion of the likelihood radio test statistic in the Exploratory factor analysis.

Theorem 2.1. *Assuming $p = p(n)$ is a series of positive integers depending on n such that $p > 1$and $p < n-1$, $p = p(n) \to \infty$ and $\frac{p}{n} \to c \in (0,1)$ as $n \to \infty$. For the test statistic T' defined in (1.2), let*

$$Z_{n,p} = \frac{T' + n\mu_{n,p}}{n\sigma_{n,p}}, \tag{2.8}$$

then we have

$$P(Z_{n,p} \le x) = \Phi_s(x) + O(\tfrac{1}{(n-p+1/2)^{s+1}}),$$

where

$$\mu_{n,p} = p\log\frac{2}{n} + \sum_{j=1}^{p}\psi(\tfrac{n-j+1}{2}), \tag{2.9}$$

$$\sigma_{n,p}^2 = -\frac{2p}{n} + \sum_{j=1}^{p}\psi^{(1)}(\tfrac{n-j+1}{2}), \tag{2.10}$$

$$\kappa_{n,p}^{(r)} = (-1)^r\sum_{j=1}^{p}\psi^{(r-1)}(\tfrac{n-j+1}{2}), \tag{2.11}$$

$$\gamma_{k,r,n,p} = \sum_{r_1+\cdots+r_k=r} \prod_{l=1}^{k} \frac{\kappa_{n,p}^{(r_l+3)}}{(r_l+3)!\sigma_{n,p}^{r_l+3}}, \tag{2.12}$$

and

$$\Phi_s(x) = \Phi(x) - \phi(x)\Big[\sum_{k=1}^{s} \frac{1}{k!} \sum_{r=0}^{s-k} \gamma_{k,r,n,p} h_{3k+r-1}(x)\Big], \tag{2.13}$$

for all $r \geq 3$. $\phi(x)$ and $\Phi(x)$ are the density function and the distribution function of the standard normal distribution, respectively. $h_r(x)$ is the rth order Hermite polynomial defined by

$$(\tfrac{d}{dx})^r \exp(-\tfrac{x^2}{2}) = (-1)^r h_r(x) \exp(-\tfrac{x^2}{2}), \quad r = 1, 2, \cdots.$$

To facilitate the proof of the following Theorem 2.1, we give the following result to certifies that $\kappa_{n,p}^{(r)}$ has a upper bound and $\sigma_{n,p}^2$ has a lower bound.

Proposition 2.1. *Under the assumptions of Theorem 2.1, there exists $\theta \in (0, 1)$ such that*

$$\kappa_{n,p}^{(r)} \leq \frac{2^{r-1}(r-2)!}{(n-p+\frac{1}{2})^{r-2}}\Big[1 + \frac{r-1}{n-p+\frac{1}{2}} + \frac{\theta(r-1)r}{3(n-p+\frac{1}{2})^2}\Big], \tag{2.14}$$

for all $r \geq 3$, and

$$\sigma_{n,p}^2 \geq 2\Big[-\frac{p}{n} + \log\frac{1+\frac{1}{n}}{1+\frac{1}{n}-\frac{p}{n}}\Big].$$

Moreover, $\sigma_{n,p}^2$ and $\lim_{n\to\infty}\sigma_{n,p}^2$ are both well-defined.

The uniform error bound on the asymptotical distribution function of $Z_{n,p}$ can be stated as follows.

Theorem 2.2. *Under the assumptions of Theorem 2.1, for any $0 < w < \sigma_{n,p}/2$, we have*

$$\sup_{x\in\mathbb{R}} |P(Z_{n,p} \leq x) - \Phi_s(x)| < \frac{1}{2\pi}[T_1(w) + T_2(w) + T_3(w)],$$

where $\Phi_s(x)$ is defined in (2.13), and

$$\begin{aligned}
T_1(w) \;=\; & \sum_{k=1}^{s} \frac{2^{3k/2}}{k!}\Big[\frac{4p}{nb\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^k \Big[\Gamma\Big(\frac{3k}{2}\Big) - \Gamma\Big(\frac{3k}{2}, \frac{b^2w^2}{2}\Big)\Big] \\
& - \sum_{k=1}^{s}\sum_{r=0}^{s-k} \frac{2^{(3k+r)/2}|\gamma_{k,r,n,p}|}{k!}\Big[\Gamma\Big(\frac{3k+r}{2}\Big) - \Gamma\Big(\frac{3k+r}{2}, \frac{b^2w^2}{2}\Big)\Big] \\
& + \Big[1 - \frac{8pw}{n\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^{-(s+1)} \Big[\frac{8pw}{n\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^{s+1} \\
& \times \Big\{\Gamma(s+1) - \Gamma\Big[s+1, \frac{b^2w^2}{2}\Big(1 - \frac{8pw}{n\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big)\Big]\Big\},
\end{aligned}$$

$$T_2(w) \;=\; \frac{n^2\sigma_{n,p}^2}{4b^2w^2\Big[\frac{pn}{4} + \frac{p(p-1)}{16} - 1\Big]}\Big(1 + \frac{4b^2w^2}{n^2\sigma_{n,p}^2}\Big)^{-\frac{pn}{4} - \frac{p(p-1)}{16} + 1},$$

$$T_3(w) \;=\; \frac{2}{b^2w^2} e^{-\frac{b^2w^2}{2}} + \sum_{k=1}^{s} \frac{1}{k!} \sum_{r=0}^{s-k} \Big|\gamma_{k,r,n,p}\Big| 2^{\frac{3k+r}{2}} \Gamma\Big(\frac{3k+r}{2}, \frac{b^2w^2}{2}\Big).$$

Here, $\Gamma(z, a)$ *is defined in (2.7) and*

$$G_n(x) = -\frac{1}{2x} - \frac{1}{x^2} - \frac{1}{x^3}\log(1-x) + \frac{1}{b(1-x)} + \frac{\theta}{3b^2(1-x)}.$$

Remark 2.3. *If taking* $s = 0$*, for any* $0 < w < \sigma_{n,p}/2$*, we have*

$$\sup_{x\in\mathbb{R}} |P(Z_{n,p} \leq x) - \Phi(x)| < \min_{0<w<\sigma_{n,p}/2} \frac{1}{2\pi}[T_1^0(w) + T_2^0(w) + T_3^0(w)], \tag{2.15}$$

where

$$\begin{aligned} T_1^0(w) &= \Big[1 - \frac{8pw}{n\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^{-1}\Big[\frac{8pw}{n\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]\Big\{1 - \Gamma\Big[1, \frac{b^2w^2}{2}\Big(1 - \frac{8pw}{n\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big)\Big]\Big\}, \\ T_2^0(w) &= \frac{n^2\sigma_{n,p}^2}{4b^2w^2[p(p-1)/16 + \frac{pn}{2} - 1]}\Big(1 + \frac{4b^2w^2}{n^2\sigma_{n,p}^2}\Big)^{-\frac{p(p-1)}{16} - \frac{pn}{2} + 1}, \\ T_3^0(w) &= \frac{2}{b^2w^2} e^{-\frac{b^2w^2}{2}}. \end{aligned}$$

Remark 2.4. *By Theorem 2.2, for any* $0 < w < \sigma_{n,p}/2$*, we have*

$$\sup_{x\in\mathbb{R}} |P(Z_{n,p} \leq x) - \Phi_s(x)| < \min_{0<w<\sigma_{n,p}/2} \frac{1}{2\pi}[T_1(w) + T_2(w) + T_3(w)].$$

3. Simulations

In this section, we first investigate the efficiency of our proposed Edgeworth expansional method with the high-dimensional covariance structure test in Exploratory Factor Analysis by numerical simulations; secondly, we compare our proposed more accurate high-dimensional Edgeworth expansion (AHEE) method in this paper with those of the traditional Chi-square approximation (CA) method and the proposed high-dimensional LR test method (HLRT) in Lemma 1.1; and finally we study the accuracy of the approximated distribution function $Z_{n,p}$ according to the numerical results on the uniform error bound of the $Z_{n,p}$.

Without loss of generality, We consider the likelihood ratio test under $H_{0,k}$. We can select $\mu = 0_{p\times 1}$,set $k = k_0 = 3$, $\Psi = (1-\rho^2)I_p$, and

$$\Lambda = \begin{bmatrix} \rho\times 1_{p_1} & 0_{p_1} & 0_{p_1} \\ 0_{p_1} & \rho\times 1_{p_1} & 0_{p_1} \\ 0_{p-2p_1} & 0_{p-2p_1} & \rho\times 1_{p-2p_1} \end{bmatrix},$$

where $p_1 = [p/3]$, $\rho = 0.6$, and I_{p_1} denotes a p_1 dimensional vector with all one entries, then $\Sigma = \Lambda\Lambda^T + \Psi$.

Now we begin to performed some simulations in order to exhibition the goodness of fit between the Edgeworth expansion of $Z_{n,p}$ and the standard normal distribution. We take $s = 0$ in Theorem 2.1, let the sample size $n = 1000$ and $p = 30, 60, 90, 180, 360, 720$, respectively. All the simulation results are based on $10,000$ independent replications, then we can plot the corresponding histograms in Figure 1, which shows that the histogram of $Z_{n,p}$ fits well with the standard normal density as p increases with n and becomes large relative to n.

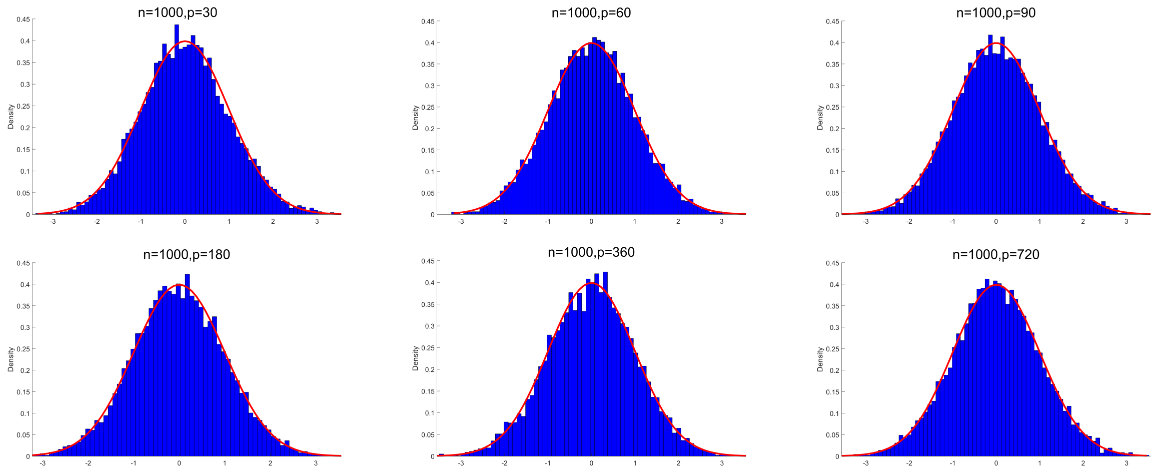

Figure 1: Comparison between the histograms of $Z_{n,p}$ and the standard normal density curve.

Moreover, we take $s = 1$ in Theorem 2.1, and compare the empirical sizes of the proposed more accurate high-dimensional Edgeworth expansion (AHEE) method with those of the traditional Chi-square approximation (CA) method in ([10]) and the proposed high-dimensional LR test method (HLRT) in Lemma 1.1. We still choose the sample size $n = 1000$ and perform 10000 repeated independent calculations. By means of the numerical simulations we can estimate the quantiles of the distribution of $\Phi_1(x)$. The following table 1 gives the simulation results. Table 1 shows that the empirical sizes of the CA method, the HLRT method and our proposed AHEE method are all very close to the given significance level α when the dimension p is small. But as p gets larger, the empirical sizes of the CA method moves away from the given significant level, it gets larger and larger and approaches to 1. Meanwhile, the sizes of the HLRT and our AHEE method are both still very close to the significance level α. Thus, both the proposed AHEE method and the HLRT method are efficient when dealing with the high-dimensional covariance structure test in exploratory factor analysis.

Table 1: Empirical sizes of CA , HLRT and AHEE methods

p	$\alpha = 0.05$			$\alpha = 0.1$			$\alpha = 0.5$		
	CA	AHEE	HLRT	CA	AHEE	HLRT	CA	AHEE	HLRT
30	0.0490	0.0543	0.0535	0.0977	0.1042	0.1068	0.4993	0.5021	0.5009
60	0.0510	0.0503	0.0498	0.1011	0.1003	0.1064	0.5018	0.5019	0.5027
90	0.0504	0.0509	0.0505	0.0997	0.1019	0.1035	0.5058	0.5020	0.5021
120	0.0582	0.0559	0.0533	0.1116	0.1035	0.1055	0.5242	0.4983	0.4985
180	0.0707	0.0485	0.0542	0.1321	0.0980	0.1015	0.5730	0.5010	0.4883
270	0.1740	0.0474	0.0543	0.2830	0.0982	0.1045	0.7627	0.5011	0.5013
360	0.5668	0.0510	0.0490	0.7020	0.1005	0.1022	0.9632	0.4942	0.5052
450	0.9856	0.0474	0.0496	0.9946	0.0980	0.0991	0.9997	0.5039	0.4924
540	1	0.0468	0.0532	1	0.0934	0.0997	1	0.4993	0.4990
630	1	0.0489	0.0490	1	0.1025	0.0964	1	0.4992	0.4973
720	1	0.0482	0.0548	1	0.1014	0.1035	1	0.5066	0.4980

Finally, for the uniform error bound of the asymptotical distribution of the statistic $Z_{n,p}$ in (2.15):

$$\Delta_{n,p} = \min_{0<w<\sigma_{n,p}/2} \left\{ \frac{1}{2\pi} [T_1^0(w) + T_2^0(w) + T_3^0(w)] \right\}.$$

we set $w = \frac{i}{50} \times \frac{\sigma_{n,p}}{2}, i = 1, 2, \cdots, 49$. If we choose the ratio of the dimension of p to n are $\frac{3}{10}$, $\frac{1}{2}$, $\frac{5}{8}$, then the corresponding values of (n,p) are $(240,800)$, $(400,800)$, $(500,800)$, $(300,1000)$, $(500,1000)$, $(625,1000)$, $(500,1500)$, $(750,1500)$, $(938,1500)$, respectively. When w increases from $\frac{1}{50}\times\frac{\sigma_{n,p}}{2}$ to $\frac{49}{50}\times\frac{\sigma_{n,p}}{2}$, we find that the corresponding minimum value of the error bound $\Delta_{n,p}$ decreases as n and p increase together for the same ratio p/n. In addition, when the ratio p/n increases, we get that the corresponding minimum value of the error bound $\Delta_{n,p}$ also decreases as p increases when n is fixed. The numerical results of the uniform error bound $\Delta_{n,p}$ are shown in the following Table 2, which shows that the uniform error bounds are very small when p and n are large in the high-dimensional case.

Table 2: Error bounds of the asymptotical distribution of $Z_{n,p}$.

p/n	(p,n)	$\Delta_{n,p}$	(p,n)	$\Delta_{n,p}$	(p,n)	$\Delta_{n,p}$
3/10	(240,800)	0.0249	(300,1000)	0.0149	(450,1500)	0.0060
1/2	(400,800)	0.0078	(500,1000)	0.0060	(750,1500)	0.0058
5/8	(500,800)	0.0057	(625,1000)	0.0041	(938,1500)	0.0038

4. Proofs

Proof of Theorem 2.1. By Corollary 8.4.8 in Muirhead (1982) [19], we know that under H_0, for any $t\in R$, the characteristic function of T' defined in (1.3) can be written as

$$\begin{aligned}\varphi_{T'}(t) &= E[\exp(itT')]\\ &= \Big(\frac{2e}{n}\Big)^{-itpn}\Big(1-2it\Big)^{-\frac{np}{2}(1-2it)}\times\frac{\Gamma_p\Big(\frac{n}{2}-int\Big)}{\Gamma_p\Big(\frac{n}{2}\Big)}.\end{aligned}\tag{4.16}$$

As the multivariate gamma function

$$\Gamma_p(\alpha)=\pi^{\frac{p(p-1)}{4}}\prod_{i=1}^{p}\Gamma\Big[\alpha-\frac{1}{2}(i-1)\Big]$$

for $Re(\alpha)>\frac{1}{2}(p-1)$, we can see that

$$\varphi_{T'}(t)=\Big(\frac{2e}{n}\Big)^{-itpn}\Big(1-2it\Big)^{-\frac{np}{2}(1-2it)}\times\prod_{j=1}^{p}\frac{\Gamma\Big(\frac{n-j+1}{2}-int\Big)}{\Gamma\Big(\frac{n-j+1}{2}\Big)}.\tag{4.17}$$

Taking the logarithm on both sides of the above equation, we get the cumulant generating function of T' by

$$\begin{aligned}&\log\varphi_{T'}(t)\\ =\ &-itnp\log\frac{2e}{n}-\frac{np}{2}(1-2it)\log(1-2it)+\sum_{j=1}^{p}\Big[\log\Gamma\Big(\frac{n-j+1}{2}-int\Big)-\log\Gamma\Big(\frac{n-j+1}{2}\Big)\Big].\end{aligned}$$

For any fixed real number a and b, using the Taylor expansion formula, we get that

$$\log\Gamma(a+b)=\log\Gamma(a)+\sum_{k=1}^{\infty}\frac{b^k}{k!}\psi^{(k-1)}(a).\tag{4.18}$$

When $t\in(0,\frac{1}{pn^2})$, some elementary calculations can lead to that

$$\log\varphi_T'(t) = -itpn[\log 2+1-\log n]-\frac{np}{2}(1-2it)\log(1-2it)+\sum_{j=1}^{p}\sum_{r=1}^{\infty}\frac{(-int)^r}{r!}\psi^{(r-1)}\Big(\frac{n-j+1}{2}\Big)$$

$$\begin{aligned} &= -itpn[\log 2 + 1 - \log n] - \frac{np}{2}\Big[-2it + 2(it)^2 + O(t^3)\Big] + \sum_{j=1}^{p}\sum_{r=1}^{\infty}\frac{(-int)^r}{r!}\psi^{(r-1)}\Big(\frac{n-j+1}{2}\Big) \\ &= it(-n)\Big[p\log\frac{2}{n} + \sum_{j=1}^{p}\psi\Big(\frac{n-j+1}{2}\Big)\Big] + \frac{(it)^2}{2}n^2\Big[-\frac{2p}{n} + \sum_{j=1}^{p}\psi^{(1)}\Big(\frac{n-j+1}{2}\Big)\Big] \\ &\quad + \sum_{j=1}^{p}\sum_{r=3}^{\infty}\frac{(-int)^r}{r!}\psi^{(r-1)}\Big(\frac{n-j+1}{2}\Big) + o(1) \\ &= it(-n)\mu_{n,p} + \frac{(it)^2}{2}n^2\sigma_{n,p}^2 + \sum_{r=3}^{\infty}\frac{(it)^r}{r!}n^r\kappa_{n,p}^{(r)} + o(1), \end{aligned} \tag{4.19}$$

where $\mu_{n,p}$, $\sigma_{n,p}^2$ and $\kappa_{n,p}^{(r)}$ are defined in (2.9), (2.10) and (2.11), respectively.

Then, we can get the characteristic function of $Z_{n,p}$ by

$$\begin{aligned} \varphi_{Z_{n,p}}(t) &= E[\exp(itZ_{n,p})] \\ &= \exp\Big(\frac{it\mu_{n,p}}{\sigma_{n,p}}\Big)\varphi_{T'}\Big(\frac{t}{n\sigma_{n,p}}\Big) \\ &= \exp\Big[\log\varphi_{T'}\Big(\frac{t}{n\sigma_{n,p}}\Big) + \frac{it\mu_{n,p}}{\sigma_{n,p}}\Big]. \end{aligned}$$

Inserting (4.3) into $\varphi_{Z_{n,p}}(t)$, we have

$$\begin{aligned} \varphi_{Z_{n,p}}(t) &= \exp\Big[\frac{it\mu_{n,p}}{\sigma_{n,p}} + \frac{(it)^2}{2} + \sum_{r=3}^{\infty}\frac{(it)^r\kappa_{n,p}^{(r)}}{r!\sigma_{n,p}^r} - \frac{it\mu_{n,p}}{\sigma_{n,p}} + o(1)\Big] \\ &= e^{-\frac{t^2}{2}}\Big\{1 + \sum_{k=1}^{\infty}\frac{1}{k!}\Big[\sum_{r=3}^{\infty}\frac{(it)^r\kappa_{n,p}^{(r)}}{r!\sigma_{n,p}^r}\Big]^k\Big\} \\ &= e^{-\frac{t^2}{2}}\Big\{1 + \sum_{k=1}^{\infty}\frac{(it)^{3k}}{k!}\Big[\sum_{r=0}^{\infty}\frac{(it)^r\kappa_{n,p}^{(r+3)}}{(r+3)!\sigma_{n,p}^{r+3}}\Big]^k\Big\} \\ &= e^{-\frac{t^2}{2}}\Big[1 + \sum_{k=1}^{\infty}\frac{(it)^{3k}}{k!}\sum_{r=0}^{\infty}\sum_{r_1+\cdots+r_k=r}\prod_{l=1}^{k}\frac{(it)^{r_l}\kappa_{n,p}^{(r_l+3)}}{(r_l+3)!\sigma_{n,p}^{r_l+3}}\Big] \\ &= e^{-\frac{t^2}{2}}\Big[1 + \sum_{k=1}^{\infty}\frac{1}{k!}\sum_{r=0}^{\infty}(it)^{3k+r}\gamma_{k,r,n,p}\Big], \end{aligned} \tag{4.20}$$

where $\gamma_{k,r,n,v}$ is defined in (2.12).

By the inverse formula of characteristic function, we know

$$P(Z_{n,p} \le x) = \Phi(x) + \sum_{k=1}^{\infty}R_k(x), \tag{4.21}$$

where $R_k(x)$ satisfies

$$\int_{-\infty}^{\infty} e^{itx}dR_k(x) = \frac{1}{k!}\sum_{r=0}^{\infty}(it)^{3k+r}e^{-\frac{t^2}{2}}\gamma_{k,r,n,p}. \tag{4.22}$$

In order to compute $R_k(x)$, we use integration by parts to get that

$$e^{-\frac{t^2}{2}} = \int_{-\infty}^{\infty} e^{itx}d\Phi(x) = \int_{-\infty}^{\infty}\Phi^{(1)}(x)d\Big(\frac{e^{itx}}{it}\Big)$$

$$\begin{aligned}&= (-it)^{-1}\int_{-\infty}^{\infty} e^{itx} d\Phi^{(1)}(x)\\&= (-it)^{-2}\int_{-\infty}^{\infty} e^{itx} d\Phi^{(2)}(x)\\&\vdots\\&= (-it)^{-(3k+r)}\int_{-\infty}^{\infty} e^{itx} d\Phi^{(3k+r)}(x),\end{aligned}$$

where $\Phi^{(k)}(x) = (\frac{d}{dx})^k\Phi(x)$. And we have

$$(it)^{(3k+r)}e^{-\frac{t^2}{2}} = (-1)^{-(3k+r)}\int_{-\infty}^{\infty} e^{itx} d\Phi^{(3k+r)}(x) = \int_{-\infty}^{\infty} e^{itx} d\Big[\Big(-\frac{d}{dx}\Big)^{3k+r}\Phi(x)\Big].$$

Thus

$$\begin{aligned}P(Z_{n,p} \le x) = &= \frac{1}{k!}\sum_{r=0}^{\infty}\int_{-\infty}^{\infty} e^{itx} d\Big[\Big(-\frac{d}{dx}\Big)^{3k+r}\Phi(x)\Big]\gamma_{k,r,n,p}\\&= \int_{-\infty}^{\infty} e^{itx} d\Big[\frac{1}{k!}\sum_{r=0}^{\infty}\gamma_{k,r,n,p}\Big(-\frac{d}{dx}\Big)^{3k+r}\Phi(x)\Big]\\&= \int_{-\infty}^{\infty} e^{itx} d\Big[-\frac{1}{k!}\sum_{r=0}^{\infty}\gamma_{k,r,n,p}h_{3k+r-1}(x)\phi(x)\Big]\\&= \Phi(x) - \phi(x)\Big[\sum_{k=1}^{\infty}\frac{1}{k!}\sum_{r=0}^{\infty}\gamma_{k,r,n,p}h_{3k+r-1}(x)\Big].\end{aligned} \tag{4.23}$$

Define

$$\varphi^{(s)}_{Z_{n,p}}(t) = \exp\Big(-\frac{t^2}{2}\Big)\Big[1+\sum_{k=1}^{s}\frac{1}{k!}\sum_{r=0}^{s-k}(it)^{3k+r}\gamma_{k,r,n,p}\Big], \tag{4.24}$$

and

$$\Phi_s(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi^{(s)}_{Z_{n,p}}(t)dt.$$

Similar to the calculations of (4.21), we can get that

$$\Phi_s(x) = \Phi(x) - \phi(x)\Big[\sum_{k=1}^{s}\frac{1}{k!}\sum_{r=0}^{s-k}\gamma_{k,r,n,p}h_{3k+r-1}(x)\Big]. \tag{4.25}$$

Inserting (4.25) into (4.23), we have

$$\begin{aligned}&P(Z_{n,p} \le x)\\&= \Phi_s(x) - \phi(x)\Big[\sum_{k=1}^{\infty}\frac{1}{k!}\sum_{r=0}^{\infty}\gamma_{k,r,n,p}h_{3k+r-1}(x) - \sum_{k=1}^{s}\frac{1}{k!}\sum_{r=0}^{s-k}\gamma_{k,r,n,p}h_{3k+r-1}(x)\Big]\\&= \Phi_s(x) - \phi(x)\Big[\sum_{k=1}^{s}\frac{1}{k!}\sum_{r=0}^{\infty}\gamma_{k,r,n,p}h_{3k+r-1}(x) + \sum_{k=s+1}^{\infty}\frac{1}{k!}\sum_{r=0}^{\infty}\gamma_{k,r,n,p}h_{3k+r-1}(x) - \sum_{k=1}^{s}\frac{1}{k!}\sum_{r=0}^{s-k}\gamma_{k,r,n,p}h_{3k+r-1}(x)\Big]\\&= \Phi_s(x) - \phi(x)\Big[\sum_{k=1}^{s}\frac{1}{k!}\sum_{r=s-k+1}^{\infty}\gamma_{k,r,n,p}h_{3k+r-1}(x) + \sum_{k=s+1}^{\infty}\frac{1}{k!}\sum_{r=0}^{\infty}\gamma_{k,r,n,p}h_{3k+r-1}(x)\Big].\end{aligned}$$

Note that $h_{3k+r-1}(x)$ is bounded for any fixed x. Then for any fixed x, there is a positive function $M_{3k+r-1}(x)$, such that $|h_{3k+r-1}(x)| \leq M_{3k+r-1}(x)$. Thus, we have

$$\begin{aligned} & |P(Z_{n,p} \leq x) - \Phi_s(x)| \\ \leq \ & \phi(x)M_{3k+r-1}(x)\Big|\sum_{k=1}^{s}\frac{1}{k!}\sum_{r=s-k+1}^{\infty}\gamma_{k,r,n,p} + \sum_{k=s+1}^{\infty}\frac{1}{k!}\sum_{r=0}^{\infty}\gamma_{k,r,n,p}\Big|. \end{aligned} \tag{4.26}$$

By Proposition 2.1, we know

$$\begin{aligned} & |\gamma_{k,r,n,p}| \\ \leq \ & \frac{1}{\sigma_{n,p}^{r+3k}}\sum_{r_1+\cdots+r_k=r}\Big\{\prod_{l=1}^{k}\frac{2^{r_l+2}p(r_l+1)!}{n(n-p+\frac{1}{2})^{r_l+1}(r_l+3)!}\Big[1+\frac{r_l+2}{n-p+\frac{1}{2}}+\frac{\theta(r_l+2)(r_l+3)}{3(n-p+\frac{1}{2})^2}\Big]\Big\} \\ \leq \ & \frac{2^r4^k}{\sigma_{n,p}^{r+3k}(n-p+\frac{1}{2})^{r+k}}\sum_{r_1+\cdots+r_k=r}\prod_{l=1}^{k}\Big[\frac{1}{(r_l+2)(r_l+3)}+\frac{1}{(n-p+\frac{1}{2})(r_l+3)}+\frac{\theta}{3(n-p+\frac{1}{2})^2}\Big], \end{aligned} \tag{4.27}$$

where $r_1, r_2, \cdots, r_k$ are all nonnegative integers which are not exceeding k.

Because of $p \to \infty$ and $p/n \to c \in (0,1)$ as $n \to \infty$, we have that $n-p+1/2 \to \infty$, and there exists $0 < a < n-p+1/2$ such that for any fixed r,

$$\begin{aligned} & \sum_{r_1+\cdots+r_k=r}\prod_{l=1}^{k}\Big[\frac{1}{(r_l+2)(r_l+3)}+\frac{1}{(n-p+\frac{1}{2})(r_l+3)}+\frac{\theta}{3(n-p+\frac{1}{2})^2}\Big] \\ \leq \ & \sum_{r_1+\cdots+r_k=r}\prod_{l=1}^{k}\Big[\frac{1}{(r_l+2)(r_l+3)}+\frac{1}{a(r_l+3)}+\frac{\theta}{3a^2}\Big], \end{aligned}$$

where $\sum_{r_1+\cdots+r_k=r}\prod_{l=1}^{k}\Big[\frac{1}{(r_l+2)(r_l+3)}+\frac{1}{a(r_l+3)}+\frac{\theta}{3a^2}\Big]$ is bounded.

Proposition 2.1 reveals that $\sigma_{n,p}$ is also bounded, therefore

$$\gamma_{k,r,n,p} = O\Big(\frac{1}{(n-p+\frac{1}{2})^{r+k}}\Big).$$

Note that $\phi(x)$ is also bounded for any fixed x, then we get that

$$\begin{aligned} & |P(Z_{n,p} \leq x) - \Phi_s(x)| \\ \leq \ & \phi(x)M_{3k+r-1}(x)\Big|\sum_{k=1}^{s}\frac{1}{k!}\sum_{r=s-k+1}^{\infty}\gamma_{k,r,n,p} + \sum_{k=s+1}^{\infty}\frac{1}{k!}\sum_{r=0}^{\infty}\gamma_{k,r,n,p}\Big| \\ = \ & O\Big(\frac{1}{(n-p+\frac{1}{2})^{s+1}}\Big). \end{aligned}$$

That is, for any $x \in \mathbb{R}$, we have

$$P(Z_{n,p} \leq x) = \Phi_s(x) + O\Big(\frac{1}{(n-p+\frac{1}{2})^{s+1}}\Big).$$

Proof of Theorem 2.2. By the inverse Fourier transformation, for any $x \in \mathbb{R}$,

$$P(Z_{n,p} \leq x) - \Phi_s(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{\varphi_{Z_{n,p}}(t) - \varphi_{Z_{n,p}}^{(s)}(t)}{-it}dt,$$

we have

$$\sup_x |P(Z_{n,p} \leq x) - \Phi_s(x)| \ \leq \ \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{|t|}|\varphi_{Z_{n,p}}(t) - \varphi_{Z_{n,p}}^{(s)}(t)|dt$$

$$\leq \frac{1}{2\pi}(I_1 + I_2 + I_3), \tag{4.28}$$

where

$$I_1 = \int_{|t|\leq bw} \tfrac{1}{|t|}|\varphi_{Z_{n,p}}(t) - \varphi^{(s)}_{Z_{n,p}}(t)|dt,$$
$$I_2 = \int_{|t|>bw} \tfrac{1}{|t|}|\varphi_{Z_{n,p}}(t)|dt,$$
$$I_3 = \int_{|t|>bw} \tfrac{1}{|t|}|\varphi^{(s)}_{Z_{n,p}}(t)|dt$$

with $0 < 2w/\sigma_{n,p} < 1$ and $b = n - p + \frac{1}{2}$.

We will next complete the upper bound of I_1, I_2 and I_3, respectively.

For the term I_1, we can get by (4.20) – (4.24) that

$$\begin{aligned}
&|\varphi_{Z_{n,p}}(t) - \varphi^{(s)}_{Z_{n,p}}(t)|\\
= {}& \exp(-\tfrac{t^2}{2})\Big|\sum_{k=1}^{s}\frac{1}{k!}\sum_{r=s-k+1}^{\infty}(it)^{3k+r}\gamma_{k,r,n,p} + \sum_{k=s+1}^{\infty}\frac{1}{k!}\sum_{r=0}^{\infty}(it)^{3k+r}\gamma_{k,r,n,p}\Big|\\
\leq {}& \exp(-\tfrac{t^2}{2})\Big\{\sum_{k=1}^{s}\frac{|t|^{3k}}{k!}\Big[\sum_{r=0}^{\infty}|t|^r|\gamma_{k,r,n,p}| - \sum_{r=0}^{s-k}|t|^r|\gamma_{k,r,n,p}|\Big] + \sum_{k=s+1}^{\infty}\frac{|t|^{3k}}{k!}\sum_{r=0}^{\infty}|t|^r|\gamma_{k,r,n,p}|\Big\}.
\end{aligned} \tag{4.29}$$

Under the assumption of $0 < 2w/\sigma_{n,v} < 1$, for $|t| \leq bw$, we can get from (4.27) that

$$\begin{aligned}
&\sum_{r=0}^{\infty}|t|^r|\gamma_{k,r,n,p}|\\
\leq {}& \Big(\frac{4p}{nb\sigma^3_{n,p}}\Big)^k\sum_{r=0}^{\infty}\Big(\frac{2|t|}{b\sigma_{n,p}}\Big)^r\sum_{r_1+\cdots+r_k=r}\prod_{l=1}^{k}\Big[\frac{1}{(r_l+2)(r_l+3)} + \frac{1}{b(r_l+3)} + \frac{\theta}{3b^2}\Big]\\
\leq {}& \Big(\frac{4p}{nb\sigma^3_{n,p}}\Big)^k\sum_{r=0}^{\infty}\Big(\frac{2\omega}{\sigma_{n,p}}\Big)^r\sum_{r_1+\cdots+r_k=r}\prod_{l=1}^{k}\Big[\frac{1}{(r_l+2)(r_l+3)} + \frac{1}{b(r_l+3)} + \frac{\theta}{3b^2}\Big]\\
= {}& \Big(\frac{4p}{nb\sigma^3_{n,p}}\Big)^k\sum_{r=0}^{\infty}\sum_{r_1+\cdots+r_k=r}\prod_{l=1}^{k}\Big(\frac{2w}{\sigma_{n,p}}\Big)^{r_l}\Big[\frac{1}{(r_l+2)(r_l+3)} + \frac{1}{b(r_l+3)} + \frac{\theta}{3b^2}\Big]\\
\leq {}& \Big(\frac{4p}{nb\sigma^3_{n,p}}\Big)^k\sum_{r=0}^{\infty}\sum_{r_1+\cdots+r_k=r}\prod_{l=1}^{k}\Big(\frac{2w}{\sigma_{n,p}}\Big)^{r_l}\Big(\frac{1}{r_l+3} + \frac{1}{b} + \frac{\theta}{3b^2}\Big)\\
\leq {}& \Big[\frac{4p}{nb\sigma^3_{n,p}}\sum_{r=0}^{\infty}\Big(\frac{2w}{\sigma_{n,p}}\Big)^r\Big(\frac{1}{r+3} + \frac{1}{b} + \frac{\theta}{3b^2}\Big)\Big]^k\\
= {}& \Big(\frac{4p}{nb\sigma^3_{n,p}}\Big)^k\Big[\sum_{r=0}^{\infty}\frac{1}{r+3}\Big(\frac{2w}{\sigma_{n,p}}\Big)^r + \Big(\frac{1}{b} + \frac{\theta}{3b^2}\Big)\sum_{r=0}^{\infty}\Big(\frac{2w}{\sigma_{n,p}}\Big)^r\Big]^k.
\end{aligned} \tag{4.30}$$

Note that for $0 < x < 1$,

$$\begin{aligned}
&\sum_{r=0}^{\infty}\frac{1}{r+3}x^r = \frac{1}{x^3}\sum_{r=0}^{\infty}\int_0^x t^{r+2}dt = \frac{1}{x^3}\int_0^x\sum_{r=0}^{\infty}t^{r+2}dt\\
= {}& -\frac{1}{2x} - \frac{1}{x^2} - \frac{1}{x^3}\log(1-x).
\end{aligned}$$

Define

$$G_n(x) = -\frac{1}{2x} - \frac{1}{x^2} - \frac{1}{x^3}\log(1-x) + \frac{1}{b(1-x)} + \frac{\theta}{3b^2(1-x)},$$

where $\theta \in (0, 1)$.

Thus we can rewrite (4.30) as

$$\sum_{r=0}^{\infty} |t|^r |\gamma_{k,r,n,p}| \le \Big[\frac{4p}{nb\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^k, \tag{4.31}$$

which means

$$\begin{aligned} &\sum_{k=1}^{s} \frac{|t|^{3k}}{k!} \sum_{r=s-k+1}^{\infty} |t|^r |\gamma_{k,r,n,p}| \\ \le\ & \sum_{k=1}^{s} \frac{|t|^{3k}}{k!} \Big\{\Big(\frac{4p}{nb\sigma_{n,p}^3}\Big)^k \Big[G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^k - \sum_{r=0}^{s-k} |t|^r |\gamma_{k,r,n,p}|\Big\} \\ =\ & \sum_{k=1}^{s} \frac{|t|^{3k}}{k!} \Big(\frac{4p}{nb\sigma_{n,p}^3}\Big)^k \Big[G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^k - \sum_{k=1}^{s} \frac{|t|^{3k}}{k!} \sum_{r=0}^{s-k} |t|^r |\gamma_{k,r,n,p}| \end{aligned} \tag{4.32}$$

and

$$\begin{aligned} &\sum_{k=s+1}^{\infty} \frac{|t|^{3k}}{k!} \sum_{r=0}^{\infty} |t|^r |\gamma_{k,r,n,p}| \le \sum_{k=s+1}^{\infty} \frac{|t|^{3k}}{k!} \Big[\frac{4p}{nb\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^k \\ \le\ & \Big[\frac{4p|t|^3}{nb\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^{s+1} \sum_{k=0}^{\infty} \frac{1}{k!} \Big[\frac{4p|t|^3}{nb\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^k \\ \le\ & \Big[\frac{4pwt^2}{n\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^{s+1} \sum_{k=0}^{\infty} \frac{1}{k!} \Big[\frac{4pwt^2}{n\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^k \\ =\ & \Big[\frac{4pwt^2}{n\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^{s+1} \exp\Big[\frac{4pwt^2}{n\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]. \end{aligned} \tag{4.33}$$

Combining (4.32) and (4.33), we have

$$\begin{aligned} &|\varphi_{Z_{n,p}}(t) - \varphi_{Z_{n,p}}^{(s)}(t)| \\ \le\ & \sum_{k=1}^{s} \frac{|t|^{3k}}{k!} \Big(\frac{4p}{nb\sigma_{n,p}^3}\Big)^k \Big[G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^k \exp\Big(-\frac{t^2}{2}\Big) - \sum_{k=1}^{s} \frac{|t|^{3k}}{k!} \sum_{r=0}^{s-k} |t|^r |\gamma_{k,r,n,p}| \exp\Big(-\frac{t^2}{2}\Big) \\ & + \Big[\frac{4pwt^2}{n\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^{s+1} \exp\Big[\frac{4pwt^2}{n\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big) - \frac{t^2}{2}\Big]. \end{aligned} \tag{4.34}$$

Define

$$H_{n,p,w} = \frac{4pw}{n\sigma_{n,p}^3} G_n\Big(\frac{2w}{\sigma_{n,p}}\Big).$$

Then we can rewrite (4.34) as

$$\begin{aligned} &|\varphi_{Z_{n,p,w}}(t) - \varphi_{Z_{n,p,w}}^{(s)}(t)| \\ \le\ & \sum_{k=1}^{s} \frac{|t|^{3k}}{k!} \Big(\frac{4p}{nb\sigma_{n,p}^3}\Big)^k \Big[G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^k \exp\Big(-\frac{t^2}{2}\Big)] - \sum_{k=1}^{s} \frac{|t|^{3k}}{k!} \sum_{r=0}^{s-k} |t|^r |\gamma_{k,r,n,p}| \exp\Big(-\frac{t^2}{2}\Big) \end{aligned}$$

$$+\Big[\big(H_{n,p,w}\big)t^2\Big]^{s+1}\exp\Big[\big(H_{n,p,w}-\frac{1}{2}\big)t^2\Big]. \tag{4.35}$$

Recall the definition of $\Gamma(\alpha)$ in (2.7). We can reach to

$$\begin{aligned}
I_1 &= 2\int_0^{bw}\tfrac{1}{t}|\varphi_{Z_{n,p}}(t)-\varphi^{(s)}_{Z_{n,p}}(t)|dt \\
&= 2\Big[\int_0^{\infty}\frac{1}{t}|\varphi_{Z_{n,p}}(t)-\varphi^{(s)}_{Z_{n,p}}(t)|dt-\int_{bw}^{\infty}\frac{1}{t}|\varphi_{Z_{n,p}}(t)-\varphi^{(s)}_{Z_{n,p}}(t)|dt\Big] \\
&< 2\sum_{k=1}^{s}\frac{1}{k!}\Big[\frac{4p}{nb\sigma^3_{n,p}}G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^k\Big[\int_0^{\infty}\frac{t^{3k}}{t}\exp\Big(-\frac{t^2}{2}\Big)dt-\int_{bw}^{\infty}\frac{t^{3k}}{t}\exp\Big(-\frac{t^2}{2}\Big)dt\Big] \\
&\quad -2\sum_{k=1}^{s}\sum_{r=0}^{s-k}\frac{|\gamma_{k,r,n,p}|}{k!}\Big[\int_0^{\infty}\frac{t^{3k+r}}{t}\exp\Big(-\frac{t^2}{2}\Big)dt-\int_{bw}^{\infty}\frac{t^{3k+r}}{t}\exp\Big(-\frac{t^2}{2}\Big)dt\Big] \\
&\quad +2\Big[\int_0^{\infty}\frac{\big(H_{n,p,w}t^2\big)^{s+1}}{t}\exp\Big(H_{n,p,w}t^2-\frac{t^2}{2}\Big)dt-\int_{bw}^{\infty}\frac{\big(H_{n,p,w}t^2\big)^{s+1}}{t}\exp\Big(H_{n,p,w}t^2-\frac{t^2}{2}\Big)dt\Big] \\
&< \sum_{k=1}^{s}\frac{2^{3k/2}}{k!}\Big[\frac{4p}{nb\sigma^3_{n,p}}G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^k\Big[\Gamma\Big(\frac{3k}{2}\Big)-\Gamma\Big(\frac{3k}{2},\frac{b^2w^2}{2}\Big)\Big] \\
&\quad -\sum_{k=1}^{s}\sum_{r=0}^{s-k}\frac{2^{(3k+r)/2}|\gamma_{k,r,n,p}|}{k!}\Big[\Gamma\Big(\frac{3k+r}{2}\Big)-\Gamma\Big(\frac{3k+r}{2},\frac{b^2w^2}{2}\Big)\Big] \\
&\quad +\Big[1-\frac{8pw}{n\sigma^3_{n,p}}G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^{-(s+1)}\Big[\frac{8pw}{n\sigma^3_{n,p}}G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big]^{s+1} \\
&\quad \times\Big\{\Gamma(s+1)-\Gamma\Big[s+1,\frac{b^2w^2}{2}\Big(1-\frac{8pw}{n\sigma^3_{n,p}}G_n\Big(\frac{2w}{\sigma_{n,p}}\Big)\Big)\Big]\Big\} \\
&:= T_1(w).
\end{aligned} \tag{4.36}$$

For the term I_2. By the well known fact $\varphi_{Z_{n,p}}(-t)=\overline{\varphi_{Z_{n,p}}(t)}$, then we have $|\varphi_{Z_{n,p}}(-t)|=|\varphi_{Z_{n,p}}(t)|$, and

$$I_2=\int_{|t|>bw}\frac{1}{|t|}|\varphi_{Z_{n,p}}(t)|dt=2\int_{t>bw}\frac{1}{t}|\varphi_{Z_{n,p}}(t)|dt. \tag{4.37}$$

Let $\tilde{t}=t/n\sigma_{n,p}$, By (4.16) and (4.19) , we have

$$\begin{aligned}
I_2 &= 2\int_{t>bw}\frac{1}{t}\Big|E\Big[\exp\Big(it\frac{T'+n\mu_{n,p}}{n\sigma_{n,p}}\Big)\Big]\Big|dt=2\int_{t>bw}\frac{1}{t}\Big|E[\exp(i\tilde{t}T')]\Big|dt \\
&= 2\int_{t>bw}\frac{1}{t}\Big|\Big(\frac{2e}{n}\Big)^{-ipn\tilde{t}}\Big|\Big|(1-2i\tilde{t})^{-\frac{np}{2}(1-2i\tilde{t})}\Big|\Big|\frac{\Gamma_p(\frac{n}{2}-in\tilde{t})}{\Gamma_p(\frac{n}{2})}\Big|dt \\
&= 2\int_{t>bw}\frac{1}{t}\Big|\exp\Big[-\frac{pn}{4}\log(1+4\tilde{t}^2)-pn\tilde{t}\arg(1-2i\tilde{t})\Big]\Big|\Big|\frac{\Gamma_p(\frac{n}{2}-in\tilde{t})}{\Gamma_p(\frac{n}{2})}\Big|dt.
\end{aligned} \tag{4.38}$$

By the formula in Fujikoshi et al. (2010) [14] that

$$\Big|\frac{\Gamma(x+iy)}{\Gamma(x)}\Big|^2=\prod_{k=0}^{\infty}\Big[1+\Big(\frac{y}{x+k}\Big)^2\Big]^{-1},$$

we can get

$$I_2 = 2\int_{t>bw}\frac{1}{t}\Big|\exp\Big[-\frac{pn}{4}\log(1+4\tilde{t}^2)-pn\tilde{t}\arg(1-2i\tilde{t})\Big]\Big|\prod_{j=1}^{p}\prod_{k=0}^{\infty}\Big[1+\Big(\frac{n\tilde{t}}{\frac{n-j+1}{2}+k}\Big)^2\Big]^{-\frac{1}{2}}dt$$

$$= \quad 2\int_{t>bw}\frac{1}{t}\Big|\exp[-pn\tilde{t}\arg(1-2i\tilde{t})]\Big|(1+4\tilde{t}^2)^{-\frac{np}{4}}\prod_{j=1}^{p}\prod_{k=0}^{\infty}\Big[1+\Big(\frac{n\tilde{t}}{\frac{n-j+1}{2}+k}\Big)^2\Big]^{-\frac{1}{2}}dt. \quad (4.39)$$

Here

$$\prod_{j=1}^{p}\prod_{k=0}^{\infty}\Big[1+\Big(\frac{-n\tilde{t}}{\frac{n-j+1}{2}+k}\Big)^2\Big]^{-\frac{1}{2}} \le \prod_{j=1}^{p}\prod_{k=0}^{[j/4]-2}\Big[1+\big(2\tilde{t}\big)^2\Big]^{-\frac{1}{2}} \le (1+4\tilde{t}^2)^{-\frac{p(p-1)}{16}}, \quad (4.40)$$

where $[j/4]$ denotes the integer part of $j/4$.

Note that $\arg(1-2i\tilde{t}) \in (0,2\pi]$ and $\sigma_{n,p}$ is bounded by Proposition 2.1. When $p/n \to \infty$ as $n\to\infty$, we know $pn\tilde{t}\arg(1-2i\tilde{t}) = p\arg(1-2i\tilde{t})t/\sigma_{n,p} \to +\infty$ and $\exp[-pn\tilde{t}\arg(1-2i\tilde{t})] \to 0$.

Combining (4.37)-(4.40), we have

$$\begin{aligned} I_2 &\le 2\int_{t>bw}\frac{1}{t}\big(1+4\tilde{t}^2\big)^{-\frac{pn}{4}-\frac{p(p-1)}{16}}dt \\ &= \int_{b^2w^2}^{\infty}\frac{1}{u}\Big(1+\frac{4u}{n^2\sigma_{n,p}^2}\Big)^{-\frac{pn}{4}-\frac{p(p-1)}{16}}du \\ &\le \frac{1}{b^2w^2}\int_{b^2w^2}^{\infty}\Big(1+\frac{4u}{n^2\sigma_{n,p}^2}\Big)^{-\frac{pn}{4}-\frac{p(p-1)}{16}}du \\ &= \frac{n^2\sigma_{n,p}^2}{4b^2w^2[\frac{pn}{4}+\frac{p(p-1)}{16}-1]}\Big(1+\frac{4b^2w^2}{n^2\sigma_{n,p}^2}\Big)^{-\frac{pn}{4}-\frac{p(p-1)}{16}+1} \\ &:= T_2(w). \end{aligned} \quad (4.41)$$

For the term I_3 in (4.24), we can similarly get that

$$\begin{aligned} I_3 &\le 2\int_{bw}^{\infty}\frac{1}{t}\exp\big(-\tfrac{t^2}{2}\big)\Big(1+\sum_{k=1}^{s}\frac{t^{3k}}{k!}\sum_{r=0}^{s-k}t^r|\gamma_{k,r,n,p}|\Big)dt \\ &= \int_{bw}^{\infty}\frac{2}{t}e^{-\frac{t^2}{2}}dt+\sum_{k=1}^{s}\frac{2}{k!}\sum_{r=0}^{s-k}|\gamma_{k,r,n,p}|\int_{bw}^{\infty}t^{3k+r-1}e^{-\frac{t^2}{2}}dt \\ &= \int_{\frac{b^2w^2}{2}}^{\infty}\frac{1}{u}e^{-u}du+\sum_{k=1}^{s}\frac{1}{k!}\sum_{r=0}^{s-k}|\gamma_{k,r,n,v}|2^{\frac{3k+r}{2}}\int_{\frac{b^2w^2}{2}}^{\infty}u^{\frac{3k+r}{2}-1}e^{-u}du \\ &\le \frac{2}{b^2w^2}e^{-\frac{b^2w^2}{2}}+\sum_{k=1}^{s}\frac{1}{k!}\sum_{r=0}^{s-k}|\gamma_{k,r,n,v}|2^{\frac{3k+r}{2}}\Gamma\Big(\frac{3k+r}{2},\frac{b^2w^2}{2}\Big) \\ &:= T_3(w). \end{aligned} \quad (4.42)$$

Collecting (4.28), (4.36), (4.41) and (4.42), we can complete the proof of Theorem 2.2. ■

Proof of Proposition 2.1. We first prove the upper bound of $\kappa_{n,p}^{(r)}$ in (2.11). By the Stirling's formula, there exists $\theta_0 \in (0,1)$ such that

$$\psi(z) = \log z - \frac{1}{2z} + \frac{\theta_0}{12z^2},$$

then, for any $r \ge 1$, the r-th derivative of $\psi(\frac{n-j+1}{2})$ can be written as

$$\psi^{(r)}\Big(\frac{n-j+1}{2}\Big) = (-1)^{r+1}\Big[\frac{2^r(r-1)!}{(n-j+1)^r}+\frac{2^r r!}{(n-j+1)^{r+1}}-\frac{2^r\theta_0(r+1)!}{3(n-j+1)^{r+2}}\Big].$$

Then we have

$$\begin{aligned}\kappa_{n,p}^{(r+3)} &= (-1)^{r+3}\sum_{j=1}^{p}\psi^{(r+2)}\Big(\frac{n-j+1}{2}\Big)\\ &= \sum_{j=1}^{p}\Big[\frac{2^{r+2}(r+1)!}{(n-j+1)^{r+2}}+\frac{2^{r+2}(r+2)!}{(n-j+1)^{r+3}}-\frac{2^{r+2}\theta_0(r+3)!}{3(n-j+1)^{r+4}}\Big] \qquad (4.43)\end{aligned}$$

for any $r \geq 0$.

According to the trigonometric inequality, we get that

$$\begin{aligned}|\kappa_{n,p}^{(r+3)}| &\leq \sum_{j=1}^{p}\Big|\Big[\frac{2^{r+2}(r+1)!}{(n-j+1)^{r+2}}+\frac{2^{r+2}(r+2)!}{(n-j+1)^{r+3}}-\frac{2^{r+2}\theta_0(r+3)!}{3(n-j+1)^{r+4}}\Big]\Big|\\ &\leq \sum_{j=1}^{p}(A_1+A_2+A_3), \qquad (4.44)\end{aligned}$$

where

$$A_1=\frac{2^{r+2}(r+1)!}{(n-j+1)^{r+2}},\quad A_2=\frac{2^{r+2}(r+2)!}{(n-j+1)^{r+3}},\quad A_3=\frac{2^{r+2}\theta_0(r+3)!}{3(n-j+1)^{r+4}}.$$

By the assumption $n-p>0$ and results in Mitsui et al. (2015) [18], we can obtain that

$$\begin{aligned}\sum_{j=1}^{p}A_1 &= \sum_{j=1}^{p}\frac{2^{r+2}(r+1)!}{(n-j+1)^{r+2}} = 2^{r+2}(r+1)!\sum_{j=1}^{p}\frac{1}{(n-j+1)^{r+2}}\\ &\leq 2^{r+2}r!\Big[\frac{1}{(n-p+\frac{1}{2})^{r+1}}-\frac{1}{(n+\frac{1}{2})^{r+1}}\Big]\\ &\leq \frac{2^{r+2}p(r+1)!}{(n-p+\frac{1}{2})^{r+1}(n+\frac{1}{2})}\\ &\leq \frac{2^{r+2}(r+1)!p}{(n-p+\frac{1}{2})^{r+1}n}. \qquad (4.45)\end{aligned}$$

And we can similarly get that

$$\sum_{j=1}^{p}A_2 \leq \frac{2^{r+2}(r+2)!p}{(n-p+\frac{1}{2})^{r+1}n} \qquad (4.46)$$

and

$$\sum_{j=1}^{p}A_3 \leq \frac{2^{r+2}\theta_0 p(r+3)!}{3(n-p+\frac{1}{2})^{r+3}n}. \qquad (4.47)$$

Combining (4.45)-(4.47), we know

$$|\kappa_{n,p}^{(r+3)}| \leq \frac{2^{r+2}p(r+1)!}{n(n-p+\frac{1}{2})^{r+1}}\Big[1+\frac{r+2}{n-p+\frac{1}{2}}+\frac{(r+2)(r+3)\theta_0}{3(n-p+\frac{1}{2})^2}\Big].$$

Let $b=n-p+1/2$, because $p\to\infty$ and $p/n\to c\in(0,1)$ as $n\to\infty$, we can get that $b\to\infty$ and $|\kappa_{n,p}^{(r+3)}|$ has a upper bound, that is,

$$|\kappa_{n,p}^{(r+3)}| \leq \frac{2^{r+2}(r+1)!p}{b^{r+1}n}\Big[1+\frac{r+2}{b}+\frac{(r+2)(r+3)\theta_0}{3b^2}\Big]\to 0.$$

Then we prove the lower bound of $\sigma^2_{n,p}$ in (2.10). Simultaneously, we also demonstrate that $\sigma^2_{n,p} > 0$ and $\lim_{n\to\infty} \sigma^2_{n,p} > 0$. By (2.10), we have

$$\sigma^2_{n,p} = -\frac{2p}{n} + \sum_{j=1}^{p} \psi^{(1)}(\tfrac{n-j+1}{2}) = -\frac{2p}{n} + \sum_{k=0}^{\infty}\sum_{j=1}^{p} \frac{1}{(\frac{n-j+1}{2}+k)^2}.$$

Let $f(x,y) = 1/(\frac{n+1-x}{2}+y)^2$. Then $f(x,y) > 0$ is a monotonically increasing function in regard to x for $0 < x < n$ and a decreasing function in regard to y for $y \geq 0$. Meanwhile, we have that

$$\sum_{k=0}^{\infty}\sum_{j=1}^{p} \frac{1}{(\frac{n-j+1}{2}+k)^2} \geq \int_0^{\infty}\int_0^{p} \frac{1}{(\frac{n+1-x}{2}+y)^2} dxdy = 2\log\frac{1+\frac{1}{n}}{1+\frac{1}{n}-\frac{p}{n}}.$$

Let $g(x) = -x + \log\frac{1+\frac{1}{n}}{1+\frac{1}{n}-x}$. We can see that $g(0) = 0$ and

$$g'(x) = \frac{x-\frac{1}{n}}{1+\frac{1}{n}-x} > 0,$$

which means that $g(x)$ is monotonically increasing for $0 < x < 1$ and $g(x) > 0$, that is

$$\sigma^2_{n,p} > 2\Big[-\frac{p}{n} + \log\frac{1+\frac{1}{n}}{1+\frac{1}{n}-\frac{p}{n}}\Big] > 0.$$

Let $h(y) = \log\frac{1}{1-y} - y, y \in (0,1), a > 1$, we can see that $h(0) = 0$ and $h'(y) = \frac{y}{1-y} > 0$, which means that $h(y)$ is a monotonically increasing function for $0 < y < 1$, then we have $h(y) > 0$, which means

$$\lim_{n\to\infty} \sigma^2_{n,p} = 2\lim_{n\to\infty}\Big[-\frac{p}{n} + \log\frac{1+\frac{1}{n}}{1+\frac{1}{n}-\frac{p}{n}}\Big] = -2y + 2\log\frac{1}{1-y} > 0.$$

And the proof of Proposition 2.1 is complete.■

Acknowledgements

The authors would like to thank the referee and the associate editor for many valuable comments and suggestions that have led to improvements in the paper.

References

[1] H. Akaike, *Information theory and an extension of the maximum likelihood principle*, Second International Symposium on Information Theory, (1973), 267–281.
[2] H. Akaike, *Factor analysis and AIC*, Psychometrika, **52(3)**(1987), 317-332.
[3] T. Akita, J.H. Jin and H. Wakaki, *High-dimensional Edgeworth expansion of a test statistic on independence and its error bound*, J. Multivariate Anal., **101(8)**(2010), 1806–1813.
[4] T. W. Anderson, *An introduction to multivariate statistical analysis*, John Wiley & Sons, New York, 2003.
[5] M. S. Bartlett, *Tests of significance in factor analysis*, Brit. J. Math. Stat. Psy., **3(2)**(1950), 77–85.
[6] M.S. Bartlett, *The effect of standardisation on an approximation in factor analysis*, Biometrika, **38**(1951), 337–344.
[7] R. B. Cattell, *The scree test for the number of factors*, Multivar. Behav. Res., **1(2)**(1966), 245–276.
[8] E. Dobriban, *Permutation methods for factor analysis and PCA*, Ann. Statist., **48(5)**(2020), 2824–2847.
[9] Y. Fujikoshi, *Error bounds for asymptotic approximations of the linear discriminant function when the sample size and dimensionality are large*, J. Multivariate Anal., **73**(2000), 1–17.
[10] Y. Q. He, Z. Wang and G. J. Xu, *A Note on the Likelihood Ratio Test in High-Dimensional Exploratory Factor Analysis*, Psychometrika, **86(2)**(2021), 442–463.
[11] J. L. Horn, *A rationale and test for the number of factors in factor analysis*, Psychometrika, **30(2)**(1965), 179–185.
[12] Jöreskog, K.G. *Some contributions to maximum likelihood factor analysis*, Psychometrika, **32(4)** (1967), 443–482.

[13] H. F. Kaiser, *The application of electronic computers to factor analysis*, Educ. Psychol. Meas., **20(1)**(1960), 141–151.
[14] N. Kato, T. Yamada and Y. Fujikoshi, *High-dimensional asymptotic expansion of LR statistic for testing intraclass correlation structure and its error bound*, J. Multivariate Anal., **101(1)**(2010), 101–112.
[15] K. B. Keeling, *A regression equation for determining the dimensionality of data*, Multivar. Behav. Res., **35(4)**(2000), 457–468.
[16] H. Kentaro, *On the likelihood ratio test for the number of factor in exploratory factor analysis*, Struct. Equ. Model., **14(3)**(2007), 505–526.
[17] G. Louis, *Some necessary conditions for common factor analysis*, Psychometrika, **19(2)**(1954), 149–162.
[18] M. Mitsui, K. Koizumi and T. Seo, *Likelihood ratio test statistic for block compound symmetry covariance structure and its asymptoic expansion*, Technical Report No.15-03, Statistical Research Group, Hiroshima University, Japan, 2015.
[19] R. J. Muirhead, *Aspects of multivariate statistical theory*, John Wiley & Sons Inc., New York, 1982.
[20] G. Schwarz, *Estimating the dimension of a model*, Ann. Statist., **6(2)**(1978), 461–464.
[21] G. M. Sun and J. S. Xie, *High-dimensional Edgeworth expansion of the determinant of sample correlation matrix and its error bound*, Stochastics, **93(3)**(2021), 428–446.
[22] G. M. Sun and J. S. Xie, *High-dimensional Edgeworth expansion of LR statistic for testing block circular symmetry covariance structure and its errors*, Commun. Stat.-Theor. M., **52(24)** (2023), 8636–8657.
[23] H. Wakaki, *Error bounds for high-dimensional Edgeworth expansions for some tests on covariance matrices*, Proceedings of International Conference on Multivariate Statistical Modeling and High Dimensional Data Mining, Kayeseri, Turkey, 2010.
[24] T. Yamada, *High-dimensional Edgeworth expansion of LR ststistic for testing circular symmetric covariance structure and its error bound*, Commun. Stat.-Theor. M., **41(11)** (2012), 1887–1910.