Calibration of the empirical likelihood for high-dimensional data

Yukun Liu · Changliang Zou · Zhaojun Wang

Abstract   This article is concerned with the calibration of the empirical likelihood (EL) for high-dimensional data where the data dimension may increase as the sample size increases. We analyze the asymptotic behavior of the EL under a general multivariate model and provide weak conditions under which the best rate for the asymptotic normality of the empirical likelihood ratio (ELR) is achieved. In addition, there is usually substantial lack-of-fit when the ELR is calibrated by the usual normal in high dimensions, producing tests with type I errors much larger than nominal levels. We find that this is mainly due to the underestimation of the centralized and normalized quantities of the ELR. By examining the connection between the ELR and the classical Hotelling's T-square statistic, we propose an effective calibration method which works much better in most situations.

1 Introduction

With the rapid development of technology, various types of high-dimensional data have been generated in many areas, such as hyperspectral imagery, internet portals, microarray analysis and DNA. High-dimensional data refers to a data whose dimension p increases to infinity as the number of observations $n \to \infty$. Traditional statistical

Y. Liu
School of Statistics and Finance, East China Normal University, Shanghai 200241, China

C. Zou (✉) · Z. Wang
School of Mathematical Sciences, Nankai University, Tianjin 300071, China
e-mail: chlzou@yahoo.com.cn

methods may not work any more in this situation since they assume that p keeps unchanged as n increases. This challenge calls for new research on properties of traditional methods (Portnoy 1985; Hjort et al. 2009; Chen et al. 2009), and new statistical approaches to deal with high-dimensional data, see Bai and Saranadasa (1996) and Chen and Qin (2010) for two-sample test for means, Schott (2005) and Chen et al. (2010) for testing a specific covariance structure, Tang and Leng (2005) and the references therein for high-dimensional variable section.

Empirical likelihood (EL), introduced by Owen (1988, 1990), provides nonparametric analogs of parametric likelihood-based tests and confidence regions. Empirical likelihood methods have been proposed for many parameters of interest, such as a population mean, quantiles of a population distribution, regression parameters, multisample problems and estimating equations (c.f., Qin and Lawless 1994; Owen 2001). Given the interest in both high-dimensional data and empirical likelihood, the asymptotic and finite-sample properties of the EL for high-dimensional data are worth being carefully investigated.

In this paper, we focus on the EL for a population mean, which is one of its most important applications. Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables with dimension p, common mean μ and nonsingular covariance matrix Σ. The empirical likelihood ratio (ELR) function for μ is defined as

$$\mathrm{EL}_n(\mu) = -2 \sup \left\{ \sum_{i=1}^{n} \log(n w_i) : w_i \geq 0, \sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} w_i (X_i - \mu) = 0 \right\}. \quad (1)$$

Owen (1990) revealed the following Wilks phenomenon: for fixed p, if μ_0 is the true value of $\mu \mathrm{EL}_n(\mu_0) \xrightarrow{\mathcal{L}} \chi_p^2$ as $n \to \infty$.

In high-dimensional data, p diverges to infinity as $n \to \infty$, making the asymptotics of $\mathrm{EL}_n(\mu_0)$ a different story. To emphasize the dependence on n, we rewrite μ, Σ as μ_n, Σ_n. Intuitively, since χ_p^2 is asymptotically normal with mean p and variance $2p$, we may expect that

$$(\mathrm{EL}_n(\mu_{0n}) - p)/\sqrt{2p} \xrightarrow{\mathcal{L}} N(0, 1), \quad \text{as} \quad n \to \infty, \quad (2)$$

where μ_{0n} is the true value of μ_n.

As pioneers, Hjort et al. (2009) and Chen et al. (2009) have made excellent contributions to the problem under what conditions the ELR for the population mean has an asymptotically normal distribution. Hjort et al. (2009) proved (2) holds by assuming boundedness of all components of $X_i = (X_{i1}, \ldots, X_{ip})^T$ and $p = o(n^{1/3})$. Alternatively, Chen et al. (2009) assumed that the data has certain linear structure, i.e., $X_i = \Gamma Z_i + \mu$, where Γ is a $p \times m$ matrix, $m \geq p$, and $Z_i = (Z_{i1}, \ldots, Z_{im})^\mathsf{T}$ is a random vector such that

$$E(Z_i = 0), \ \mathrm{Var}(Z_i) = I_m, \ E(Z_{il}^{4k}) = m_{4k} \in (0, \infty),$$
$$E(Z_{ik_1}^{\alpha_1} Z_{ik_2}^{\alpha_2} \cdots Z_{ikq}^{\alpha_q}) = E(Z_{ik_1}^{\alpha_1}) E(Z_{ik_2}^{\alpha_2}) \cdots E(Z_{ikq}^{\alpha_q}), \quad (3)$$

whenever $\sum_{k=1}^{q} \alpha_k \leq 4k$ and $k_1 \neq k_2 \cdots \neq k_q$. Here k is some positive integer. Under this model, they showed that the effect of the dimensionality and the dependence among components of X_i on the EL are leveraged through the trace of the covariance matrix $\Sigma_n = \Gamma\Gamma^\mathsf{T}$ and its largest eigenvalue. In particular, if k could be sufficiently large, they argued that the best rate for p_n would be $p = o(n^{1/2})$. The data structure (3) generates a rich collection of X_i from Z_i with a given covariance Σ_n. However, it is difficult to justify this model since the condition that power transformations of different components of Z_i are uncorrelated, is approximately saying that they are independent and thus not easily met in practice. For instance, it can be verified that a random vector with the standard multivariate t distribution does not satisfy this condition.

This paper has two objectives. One is to thoroughly investigate the conditions for the validity of (2) without the data structure assumption. On the foundation established by Hjort et al. (2009) and Chen et al. (2009), we analyze the asymptotic behavior of the EL under a general multivariate model and provide more general conditions such that the best dimension growth rate for the asymptotic normality of the ELR is achieved. We shall see that these conditions include Chen et al. (2009) assumption as a special case. The other objective of the paper is to study the practical calibration of the ELR in finite sample settings. The main point of Hjort et al. (2009) and Chen et al. (2009) is that in the high-dimensional data setting, one is suggested to calculate critical values according to (2). However, there is usually substantial lack-of-fit when the ELR is calibrated by the usual normal, yielding tests with type I errors much larger than nominal levels. See the simulation results in Chen et al. (2009) and this paper. We find that the lack-of fit is mainly due to the deviation of the expectation and variance of $\mathrm{EL}_n(\mu_0)$ from p and $2p$, respectively, and the gaps get more prominent as p increases. We propose to replace p and $2p$ in (2) with better approximations of the expectation and variance of $\mathrm{EL}_n(\mu_0)$. In doing so, the resulting critical values get much more accurate and the lack-of-fit in the original normal-based calibration largely disappears.

The rest of this paper is organized as follows: In Sect. 2 we investigate the asymptotical normality of the ELR under very general conditions. Section 3 provides a new method to approximate the finite-sample distribution of the ELR. The performance of this new calibration method is evaluated by a simulation study in Sect. 4. All proofs are postponed to the Appendix.

2 Asymptotic results

In this section, we study under what conditions the ELR, $\mathrm{EL}_n(\mu_{0n})$, has an asymptotically normal distribution. Without loss of generality, we assume $\mu_{0n} = 0$ henceforth. As pointed out by Tsao (2004), when $p/n \geq 1/2$, $\mathrm{EL}_n(0)$ has no definition with probability tending to one and (2) will never be valid. In this paper we assume that p is an increasing function of n and $p = o(n/\log n)$, so that $\mathrm{EL}_n(0)$ is almost surely well-defined as n is sufficiently large (see Lemma 8 in Appendix D). According to (1), if $\mathrm{EL}_n(0)$ has definition, by Lagrange multiplier method we get

$$\mathrm{EL}_n(0) = 2 \sum_{i=1}^{n} \log\left(1 + \lambda_*^\mathsf{T} X_i\right), \tag{4}$$

where λ_* is the solution to

$$\sum_{i=1}^{n} \frac{X_i}{1 + \lambda^{\mathsf{T}} X_i} = 0. \tag{5}$$

For exposition convenience, we define some notation. Let $\gamma_k(A)$ denote the kth largest eigenvalue of a symmetric matrix A. For a random vector X, we define $\|X\|_q = [\mathrm{E}(\|X\|^q)]^{1/q}$. Throughout this paper we assume that there exists $c_0 > 0$ such that $\inf_n \gamma_p(\Sigma_n) > c_0$. This is a basic assumption in the analysis of the EL for high-dimensional data and is adopted in both Hjort et al. (2009) and Chen et al. (2009). Moreover, let $\alpha^{i_1 i_2 \dots i_k} = \mathrm{E}(Y_{1i_1} Y_{1i_2} \dots Y_{1i_k})$ with $Y_r = (Y_{r1}, \dots, Y_{rp})^{\mathsf{T}} = \Sigma_n^{-1/2} X_r$. Similar to Hjort et al. (2009) and Chen et al. (2009), we separately secure conditions under which

$$\mathrm{EL}_n(0) - n\bar{X}^{\mathsf{T}} S_n^{-1} \bar{X} = o_p\left(\sqrt{p}\right), \tag{6}$$

$$n\bar{X}^{\mathsf{T}}\left(\Sigma_n^{-1} - S_n^{-1}\right)\bar{X} = o_p\left(\sqrt{p}\right), \tag{7}$$

hold, where $S_n = n^{-1}\sum_{i=1}^{n} X_i X_i^{\mathsf{T}}$ and $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. If these two conclusions hold, then

$$\mathrm{EL}_n(0) = n\bar{X}^{\mathsf{T}}\Sigma_n^{-1}\bar{X} - n\bar{X}^{\mathsf{T}}\left(\Sigma_n^{-1} - S_n^{-1}\right)\bar{X} + o_p\left(\sqrt{p}\right) = n\bar{X}^{\mathsf{T}}\Sigma_n^{-1}\bar{X} + o_p\left(\sqrt{p}\right).$$

By applying the martingale central limit theorem (Hall and Hyde 1992), we have

$$(2p)^{-1/2}\left\{n\bar{X}^{\mathsf{T}}\Sigma_n^{-1}\bar{X} - p\right\} \xrightarrow{\mathcal{L}} N(0, 1), \tag{8}$$

when $p/n \to c \geq 0$. See Lemma 5 of Chen et al. (2009) or Bai and Saranadasa (1996) for a detailed discussion. Hence, the validity of (6) and (7) leads to (2).

For the validity of (6) and (7), we consider the following conditions.

Condition (C1) $p^{-1}\sum_{j=1}^{p} \mathrm{E}(|X_{1j}|^q) < K$ *for some $K > 0$ and $q \geq 4$;*
Condition (C2) $\|X_1\|_q \gamma_1^{3/2}(\Sigma_n) = o(n^{(q-2)/(2q)} p^{-1/2})$;
Condition (C3) $p^{2+4/q}/n \to 0$;
Condition (C4) $p = o(n^{2/5})$;
Condition (C5) $\alpha^{iijj} = O(p^2)$;
Condition (C6) $\alpha^{ijk}\alpha^{ijk} = O(p^{5/2})$ *and* $\alpha^{ijj}\alpha^{ikk} = O(p^{5/2})$.

Here and henceforth, without otherwise statement, if one index occurs twice in one term like $\alpha^{i_1 i_2 \dots i_k}$, it means a summation over the range of this index. Conditions (C1) and (C3) guarantee that the eigenvalues of S_n are close to those of Σ_n, so that S_n is nonsingular and $\gamma_p(S_n) > c_0 > 0$ when n is large. See Lemmas 1, 2. Condition (C1) is also assumed by Hjort et al. (2009) and is essentially analogous to assuming $E(Z_{il}^{4k}) = m_{4k} \in (0, \infty)$ for $k \geq 1$ in (3). With similar techniques given in Appendix B, we can verify that the linear structure assumption (3) implies condition (C1). Condition

(C2), together with (C1) and (C3), implies $\sup_{1 \le i \le n} |\lambda_*^{\mathsf{T}} X_i| = o_p(1)$ which leads to Taylor expansions of (13) and (14). Condition (C5) is on the fourth moments of the standardized observations Y_i's. This condition is very weak since it requires a sum of p^2 terms to be $O(p^2)$. Roughly speaking, it means that each α^{iijj}-type fourth moment of Y_i is bounded. To achieve the best rate $p = o(n^{1/2})$ claimed by Chen et al. (2009), we need condition (C6) which seems to be a relatively strong condition since it requires a sum of p^3 terms to be $O(p^{5/2})$. It can be verified that if the data has the linear structure (3) as assumed by Chen et al. (2009), then both conditions (C5) and (C6) are satisfied since (See the Appendix)

$$\alpha^{iijj} = O\left(p^2\right), \quad \alpha^{ijj}\alpha^{ikk} = O\left(p^2\right), \quad \alpha^{ijk}\alpha^{ijk} = O\left(p^2\right). \qquad (9)$$

We present the main result of this section in the following theorem. It is actually a corollary of Propositions 1 and 2 given in the appendix, which provide conditions for the validity of (6) and (7), respectively.

Theorem 1 *Assume conditions* (C1)*,* (C2)*,* (C3) *and* (C5)*. If condition* (C4) *or* (C6) *holds, then* $(\mathrm{EL}_n(\mu_0) - p)/\sqrt{2p} \xrightarrow{\mathcal{L}} N(0, 1)$*, as* $n \to \infty$*.*

When $\gamma_1(\Sigma_n)$ is bounded and $q = 4$, $\mathrm{EL}_n(\mu_0)$ has an asymptotically normal distribution if $p = o(n^{1/3})$. This improves the rates $p = o(n^{1/4})$ of Chen et al. (2009) and $p = o(n^{1/6})$ of Hjort et al. (2009). If more rigorous conditions are assumed, that is, condition (C6) holds and more moments ($q > 8$) of the data are finite, then the rate of p which guarantees the asymptotical normality of $\mathrm{EL}_n(\mu_0)$ is as large as $p = o(n^{q/(2q+4)})$ and it is close to $p = o(n^{1/2})$ when q is large enough, which improves the rate $p = o(n^{1/3})$ of Hjort et al. (2009). Chen et al. (2009) also obtained this rate and they further pointed out that $p = o(n^{1/2})$ may be the best rate for the asymptotical normality of $\mathrm{EL}_n(\mu_0)$. Unlike Chen et al. (2009), we obtain this best rate with conditions (C5) and (C6) not the linear structure assumption (3) on the data. The former conditions are more general in certain sense as indicated earlier.

3 A new calibration method for empirical likelihood

In the previous section, we have shown that (2) holds under some weak conditions. When testing hypotheses or constructing confidence regions with the EL method, we would calculate critical values based on the normal approximation (2). However, these critical values often deviate severely from the true ones when p/n is not small. Please refer to Tables 1, 2, 3 in Sect. 4. Similar findings have also been revealed by Chen et al. (2009). We find that this awkward fact is mainly due to the large difference between the true expectation and variance pair (E_n, V_n) of $\mathrm{EL}_n(0)$ and $(p, 2p)$. See also Tables 1, 2, 3. In addition, Chen et al. (2009) pointed out that the χ_p^2 calibration proves to be an improvement of the calibration based on (2). Since the EL is Bartlett correctable for any fixed p (DiCiccio et al. 1991), it is natural to expect the χ_p^2 calibration with Bartlett correction for the EL will

Table 1 Coverage percentages for $n = 200$

p	Scenario	MEL	OEL	SEL	STEL	BEL	E_n	$\hat{E}_{n1}$	$\hat{E}_{n2}$	V_n	$\hat{V}_{n1}$	$\hat{V}_{n2}$
25	(I)	94.52	87.36	84.92	89.12	91.32	28.2	25.4	28.8	70.5	48.8	75.8
	(II)	93.84	79.00	75.12	82.84	85.36	30.6	26.8	30.2	87.3	63.1	90.1
	(III)	95.60	85.88	83.36	88.04	90.08	28.4	26.1	29.5	76.1	56.1	83.1
	(IV)	94.84	87.48	84.68	88.84	90.84	28.1	25.4	28.8	70.6	48.8	75.8
	(V)	93.84	76.68	74.00	81.20	83.04	31.2	28.4	31.8	104	81.2	108
33	(I)	94.00	79.24	77.00	84.56	86.56	39.4	36.3	39.7	117	86.3	113
	(II)	91.28	66.24	63.76	74.48	77.48	43.4	38.7	42.1	157	110	137
	(III)	93.64	76.12	73.60	81.84	83.92	40.2	37.2	40.6	133	95.4	122
	(IV)	93.72	78.68	76.24	83.76	85.44	39.6	36.3	39.7	114	86.5	113
	(V)	91.24	64.00	60.92	70.72	73.48	44.6	40.2	43.6	184	128	155
43	(I)	90.04	61.96	58.76	72.08	75.16	56.5	51.4	54.8	209	143	170
	(II)	89.12	46.04	43.16	57.80	61.12	62.5	55.0	58.4	262	181	208
	(III)	88.96	59.56	56.84	70.04	72.72	57.7	52.5	55.9	244	155	182
	(IV)	89.44	61.64	59.00	72.12	74.36	56.7	51.4	54.8	234	143	170
	(V)	87.52	43.76	41.32	55.76	58.08	64.0	56.2	59.6	327	196	223

Table 2 Coverage percentages for $n = 400$

p	Scenario	MEL	OEL	SEL	STEL	BEL	E_n	$\hat{E}_{n1}$	$\hat{E}_{n2}$	V_n	$\hat{V}_{n1}$	$\hat{V}_{n2}$
33	(I)	95.44	90.00	88.44	91.56	93.20	35.2	32.7	36.1	78.2	59.4	86.4
	(II)	93.64	84.00	81.88	87.24	89.08	37.8	34.0	37.4	96.3	71.5	98.5
	(III)	95.52	88.48	86.44	90.20	91.80	35.8	33.2	36.6	84.9	64.7	91.7
	(IV)	94.96	89.44	87.84	91.00	92.24	35.4	32.7	36.1	83.2	59.4	86.4
	(V)	95.92	84.32	81.96	86.84	89.28	37.7	35.4	38.8	95.5	88.4	115
44	(I)	95.04	86.24	84.48	90.04	91.08	48.5	46.2	49.6	122	98.1	125
	(II)	93.84	75.64	73.00	82.24	84.52	52.8	48.5	51.9	143	120	147
	(III)	95.88	86.36	84.56	89.52	91.20	48.6	46.9	50.3	119	105	132
	(IV)	96.28	86.36	84.64	89.88	91.32	48.3	46.2	49.6	114	98.2	125
	(V)	95.68	78.40	75.88	82.92	84.84	51.8	49.8	53.2	141	136	163
58	(I)	95.04	78.60	76.40	85.20	87.04	66.9	64.6	68.0	178	157	184
	(II)	92.84	61.96	59.24	73.80	76.36	73.5	68.3	71.7	221	194	221
	(III)	95.12	75.64	73.52	83.64	84.88	67.6	65.4	68.8	189	166	193
	(IV)	94.27	77.23	75.34	84.75	86.49	66.7	64.6	68.0	190	157	184
	(V)	94.12	65.12	62.84	75.28	77.68	71.9	69.2	72.6	230	207	234

have better precision. Our simulation indicates that these two methods are still not good.

The foundations of using (2) to calibrate the EL are that $EL_n(0)$ is close to $K_n \equiv n\bar{X}^{\mathsf{T}}\Sigma_n^{-1}\bar{X}$, and that $E(K_n) = p$, $Var(K_n) \approx 2p$. When (E_n, V_n) deviates severely from $(p, 2p)$, this calibration does not work any more. Even so, it motivates us to

Table 3 Coverage percentages for $n = 800$

p	Scenario	MEL	OEL	SEL	STEL	BEL	E_n	$\hat{E}_{n1}$	$\hat{E}_{n2}$	V_n	$\hat{V}_{n1}$	$\hat{V}_{n2}$
42	(I)	95.84	92.88	91.48	93.60	94.88	43.4	41.0	44.4	89.0	72.4	99.4
	(II)	93.36	88.08	86.04	89.56	91.36	46.0	42.1	45.5	107	81.3	108
	(III)	95.45	93.15	91.68	93.52	94.44	43.4	41.3	44.7	88.5	76.4	103
	(IV)	95.20	92.48	91.04	92.80	93.96	43.7	41.0	44.4	92.2	72.4	99.4
	(V)	96.12	90.24	88.40	91.80	93.20	45.0	43.2	46.6	103	95.0	122
55	(I)	95.65	91.25	89.80	92.60	93.50	57.6	55.8	59.2	128	109	136
	(II)	94.25	84.50	82.55	88.30	89.80	61.4	57.6	61.0	142	125	152
	(III)	96.15	91.65	90.40	92.80	94.10	57.7	56.2	59.6	126	114	141
	(IV)	96.00	91.50	90.05	92.90	93.70	57.9	55.8	59.2	120	109	136
	(V)	95.60	86.15	85.15	88.65	90.05	60.0	58.6	62.0	148	139	166
72	(I)	96.55	88.35	86.85	91.75	92.70	76.5	75.9	79.3	178	164	191
	(II)	94.85	77.60	75.10	84.15	86.20	82.8	78.9	82.3	201	193	220
	(III)	96.40	87.15	85.55	90.65	91.80	77.1	76.4	79.8	183	170	197
	(IV)	96.45	88.05	86.70	91.35	92.55	77.4	75.9	79.3	164	164	191
	(V)	95.85	80.40	78.15	86.15	87.50	80.7	79.5	82.9	211	203	230

consider a possibly better approximation to $\mathrm{EL}_n(0)$, say $T_n \equiv n\bar{X}^\mathsf{T} S_n^{-1} \bar{X}$, which plays a critical role in the standard large-sample EL theorem (Owen 1990). By Proposition 2 in Appendix A, we have established that under certain conditions T_n is the dominant part of $\mathrm{EL}_n(0)$, so it can be expected that the expectation and variance of T_n should be a better approximation of (E_n, V_n) than $(p, 2p)$. It can be verified that

$$\mathrm{E}(T_n) \approx E_{n1} \equiv p + \frac{1}{n} \left(\alpha^{ijk}\alpha^{ijk} + \alpha^{kii}\alpha^{kjj} \right),$$

$$\mathrm{Var}(T_n) \approx V_{n1} \equiv 2p + \frac{1}{n} \left(12\alpha^{jjk}\alpha^{iik} + 12\alpha^{ijk}\alpha^{ijk} - 2\alpha^{iijj} \right).$$

We expect that replacing $(p, 2p)$ with (E_{n1}, V_{n1}) in (2) will improve the performance of the usual normal calibration. However, our simulation results show that T_n is still not a very good approximation as it is always less than $\mathrm{EL}_n(0)$ (see Fig. 1). An intuitive explanation for this phenomenon is as follows. Let $f(\lambda) = 2\sum_{i=1}^n \log(1 + \lambda^\mathsf{T} X_i)$. Obviously, $\mathrm{EL}_n(0) = \sup_\lambda f(\lambda) = f(\lambda_*)$, and λ_* is the maximum point of $f(\lambda)$. By second-order Taylor expansion, we have

$$f(\lambda) \approx g_1(\lambda) \equiv 2\sum_{i=1}^n \left\{ \lambda^\mathsf{T} X_i - \frac{1}{2} \left(\lambda^\mathsf{T} X_i \right)^2 \right\}, \tag{10}$$

provided $\lambda^\mathsf{T} X_i$'s are small. So an approximation of $\mathrm{EL}_n(0) = \sup_\lambda f(\lambda)$ is naturally $\sup_\lambda g_1(\lambda) = g_1(S_n^{-1}\bar{X}) = T_n$. However, in the case of moderate n and large p, this approximation may not work any more. The remainder of each Taylor expansion in (10) is under control only for $\lambda^\mathsf{T} X_i \in (-1, 1)$. We find in our simulations that when
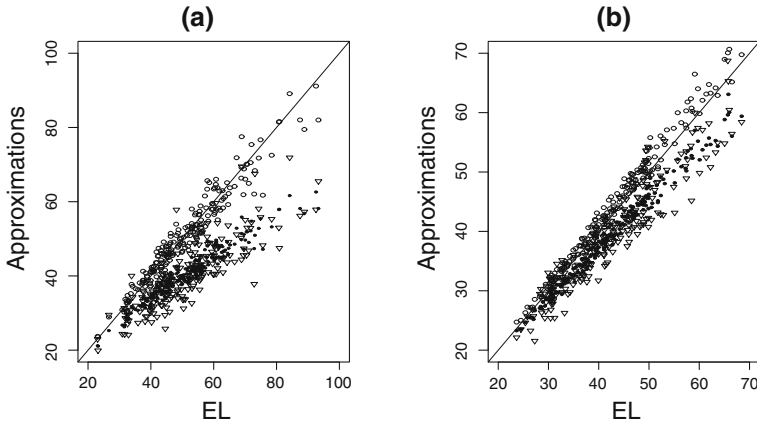
Fig. 1 Scatter plots of n simulated values of $(EL_n(0), K_n)$ (*triangles*), $(EL_n(0), T_n)$ (*dots*) and $(EL_n(0), T_{nc})$ (*circles*) for the standard p-dimensional normal distribution with (**a**) $n = 200$ and $p = 40$, (**b**) $n = 400$ and $p = 40$. The *solid line* is $y = x$

p/n is not small, some of $\lambda_*^\mathsf{T} X_i$'s are greater than 1 with quit a large probability, but none of them will be less than -1 due to the constraint $1 + \lambda_*^\mathsf{T} X_i > 0$. When λ lies in a neighborhood of 0, $\lambda^\mathsf{T} X_i$'s may have similar behaviors, that is some of them are greater than 1 but almost none of them is less than -1. Note that when $x \in (-1, 1)$, $\log(1+x) \approx x - x^2/2$; while if $x > 1$, $\log(1+x) > \log(2) > x - x^2/2$. Therefore, roughly we have $f(\lambda) \geq g_1(\lambda)$ in a neighborhood of 0. This finding also restricts us to approximate $EL_n(0)$ by two terms Taylor expansion, because Taylor expansions of (13) would deviate from $EL_n(0)$ more and more severely if more terms are extracted and some of $\lambda_*^\mathsf{T} X_i$'s are not small.

To reduce the approximation error of $g_1(\lambda)$, we propose to add a high-order term $(\lambda^\mathsf{T} \bar{X})^2$ to $g_1(\lambda)$. Intuitively $g_2(\lambda) = g_1(\lambda) + (\lambda^\mathsf{T} \bar{X})^2$ is a better approximate to $f(\lambda)$; so is $\sup_\lambda g_2(\lambda)$ to $EL_n(0) = \sup_\lambda f(\lambda)$. It can be verified that $\sup_\lambda g_2(\lambda) = n\bar{X}^\mathsf{T} S_{nc}^{-1} \bar{X} \equiv T_{nc}$ with $S_{nc} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^\mathsf{T}$. As a matter of fact, T_{nc} happens to be the classical Hotelling's T-square statistic. As indicated by Brown and Chen (1998), T_{nc} is also the so-called least-square empirical likelihood ratio $LEL_n(0)$. In Chapter 3 of Owen (2001), the author suggests using the distribution of T_{nc} to calibrate the ELR in low-dimensional settings. The following proposition, which can be proved in a similar fashion to Proposition 3, establishes the asymptotic behavior of $EL_n(0) - T_{nc}$.

Proposition 1 *Suppose conditions* (C1)*,* (C2) *and* (C5) *hold. Then*

$$(EL_n(0) - T_{nc})/p^{1/2} = o_p(1),$$

provided (C4) *holds or both* (C3) *and* (C6) *hold.*

This proposition implies that using T_{nc} to approximate EL is equivalent to using T_n or K_n from asymptotic viewpoints. However, these three approximations exhibit quite different finite-sample behaviors, especially when p/n is not small. Based on a large amount of simulation results, we find that T_{nc} is amazingly close to $EL_n(0)$ regardless

of the choice of (n, p) in the sense that $(\text{EL}_n(0) - T_{nc})/p^{1/2}$ is always pretty small (p of course should not be too large so that the EL is well defined). To appreciate this, Fig. 1 shows scatter plots of 200 simulated values of $(\text{EL}_n(0), K_n)$, $(\text{EL}_n(0), T_n)$ and $(\text{EL}_n(0), T_{nc})$ for the p-dimensional standard multivariate normal distribution. We choose $p = 40$ and $n = 200, 400$. From Fig. 1 we can see that the values of $(\text{EL}_n(0), T_{nc})$ are always around the line $y = x$, but both K_n and T_n tend to under-approximate $\text{EL}_n(0)$. When n increases from 200 to 400, all the three approximations improve much, while T_{nc} is still the best approximation to $\text{EL}_n(0)$. See Sect. 4 for more thorough analysis and comparison.

Given the foregoing discussion and evidence, we expect that the expectation and variance of T_{nc}, i.e.,

$$\text{E}(T_{nc}) \approx E_{n2} \equiv p + \frac{1}{n}\left(p^2 + 2p + \alpha^{ijk}\alpha^{ijk} + \alpha^{kii}\alpha^{kjj}\right),$$

$$\text{Var}(T_{nc}) \approx V_{n2} \equiv 2p + \frac{1}{n}\left(8p^2 + 16p + 12\alpha^{jjk}\alpha^{iik} + 12\alpha^{ijk}\alpha^{ijk} - 2\alpha^{iijj}\right)$$

are good approximations of E_n and V_n, respectively. Let $(\hat{E}_{ni}, \hat{V}_{ni})(i = 1, 2)$ be the moment estimate of (E_{ni}, V_{ni}). We may calculate critical values according to

$$(\text{EL}_n(0) - A_n)/\sqrt{B_n} \xrightarrow{\mathcal{L}} N(0, 1), \tag{11}$$

where (A_n, B_n) could be chosen as $(p, 2p)$, or $(\hat{E}_{ni}, \hat{V}_{ni})(i = 1, 2)$. In the next section, we will compare the three methods and show that the method based on $(\hat{E}_{n2}, \hat{V}_{n2})$ is the best. Hence, it is our final recommendation.

4 Simulation study

Here we report a simulation study designed to evaluate the performance of the proposed calibration method of empirical likelihood. The number of variety of multivariate distributions and parameters are too large to allow a comprehensive, all-encompassing comparison. We choose certain representative examples for illustration. Three well-known multivariate distributions are considered: (i) multi-normal (scenario I); (ii) multivariate t with five degrees of freedom, denoted by $t_{p,5}$ (scenario II); (iii) multi-variate chi-square distribution with three degrees of freedom (see Mardia et al. 1979, p. 92), denoted by $\chi^2_{p,3}$ (scenario III). The covariance matrix $\Sigma = (\sigma_{ij})$ in these three distributions is fixed with $\sigma_{ij} = 0.5^{|i-j|}$ for $1 \le i, j \le p$.

In addition, the moving average model studied in Chen et al. (2009) is also included here, allowing us to have a more broad picture of the efficiency of the proposed method since it satisfies the data structure (3). To be more specific, the $p \times 1$ i.i.d. data vectors $\{X_i\}_{i=1}^n$ are generated from

$$X_{ij} = Z_{ij} + \rho Z_{i,j+1}, \ i = 1, \ldots, \ j = 1, \ldots, p,$$

where for each i, Z_{ij} ($j = 1, 2, \cdots, p$) are i.i.d. random variables with zero mean and unit variance. Two distributions for the Z_{ij} are considered. One is the standard normal distribution (labeled as scenario (IV)), and the other is a standardized Pareto distribution with distribution function $(1 - x^{-4.5})I(x \geq 1)$ (labeled as scenario (V)). As the Pareto distribution has only four finite moments, we had $q = 4$ in the conditions of Theorem 1, whereas $q = \infty$ for all the other four scenarios. Note that neither $t_{p,5}$ nor $\chi^2_{p,3}$ distribution belongs to model (3). It can be also seen that except for scenario (II), the other four scenarios satisfy conditions (C5) and (C6). Hence, for scenarios (I), (III) and (V), the best rate $p = o(n^{1/2})$ for the asymptotical normality of the EL could be achieved, while for scenarios (II) and (V), the rates are $o(n^{2/5})$ and $o(n^{1/3})$, respectively.

The settings of the combinations of p and n in Chen et al. (2009) are adopted here. Say, we consider the growth rate for p with respect to n as $p = cn^{0.24}$. The sample size n is chosen as 200, 400 and 800. By assigning $c = 3, 4$ and 5, we obtain three dimensions for each sample size, i.e. $p = 25, 33$ and 43 for $n = 200$; $p = 33$, 44 and 58 for $n = 400$; and $p = 42, 55$ and 72 for $n = 800$, respectively. We compare five calibration methods for the EL: (a) the proposed method (denoted as modified EL, MEL), i.e. the normal calibration (11) with $(A_n, B_n) = (\hat{E}_{n2}, \hat{V}_{n2})$; (b) the ordinary χ^2_p calibration (OEL); (c) the usual standardized normal calibration (2) (SEL); (d) the normal calibration (11) with $(A_n, B_n) = (\hat{E}_{n1}, \hat{V}_{n1})$ (STEL); (e) the χ^2_p calibration after Bartlett correction (BEL; DiCiccio et al. 1991). Tables 1, 2, 3 report coverage percentage comparisons for constructing confidence regions with nominal coverage level 0.95 when $n = 200, 400$ and 800 respectively. Each coverage percentage is obtained based on 2,500 simulation repetitions. The averages of the stimulated (E_n, V_n) and ($\hat{E}_{ni}, \hat{V}_{ni}$), $i = 1, 2$ are also tabulated.

As we expect, the performances of all the five methods get better when n increases or p decreases. The SEL calibration seems the worst among the five methods. When $n = 200$ and 400, it gives rather unsatisfactory coverage probabilities in most cases. For instance, when $n = 400$ and $p = 58$, its coverage probabilities for scenarios (II) and (V) are as low as 60 % around. Even for the normal case in which the moment conditions in the asymptotic analysis are much easier to met, its coverage percentage still stays below 80 %. The ordinary χ^2_p calibration method works uniformly better than SEL in terms of coverage percentage. This is consistent with the findings of Chen et al. (2009) which pointed out that χ^2_p can be regarded as an intermediate convergence of the ELR. However, we also observe that the improvement of OEL over SEL is limited and in most situations their coverage percentages are unacceptably low. It is partly because the expectation and variance of the ELR are far away from p and $2p$ when p is large, as shown in these tables. We can also clearly see that the BEL and STEL methods improve the efficiency of OEL and SEL. Unfortunately, in the cases that n is not large enough such as $n = 200$ or 400, there are still considerable coverage errors with these two methods. In comparison, our proposed method, MEL, can always attain (or approximately) the desired coverage percent and outperform the rest four methods in most cases. The advantage gets more remarkable when n decreases or p increases.

To check whether the above conclusions would change for other choices of nominal coverage level, in Fig. 2, we display Q–Q plots of the standardized ELR (11) with
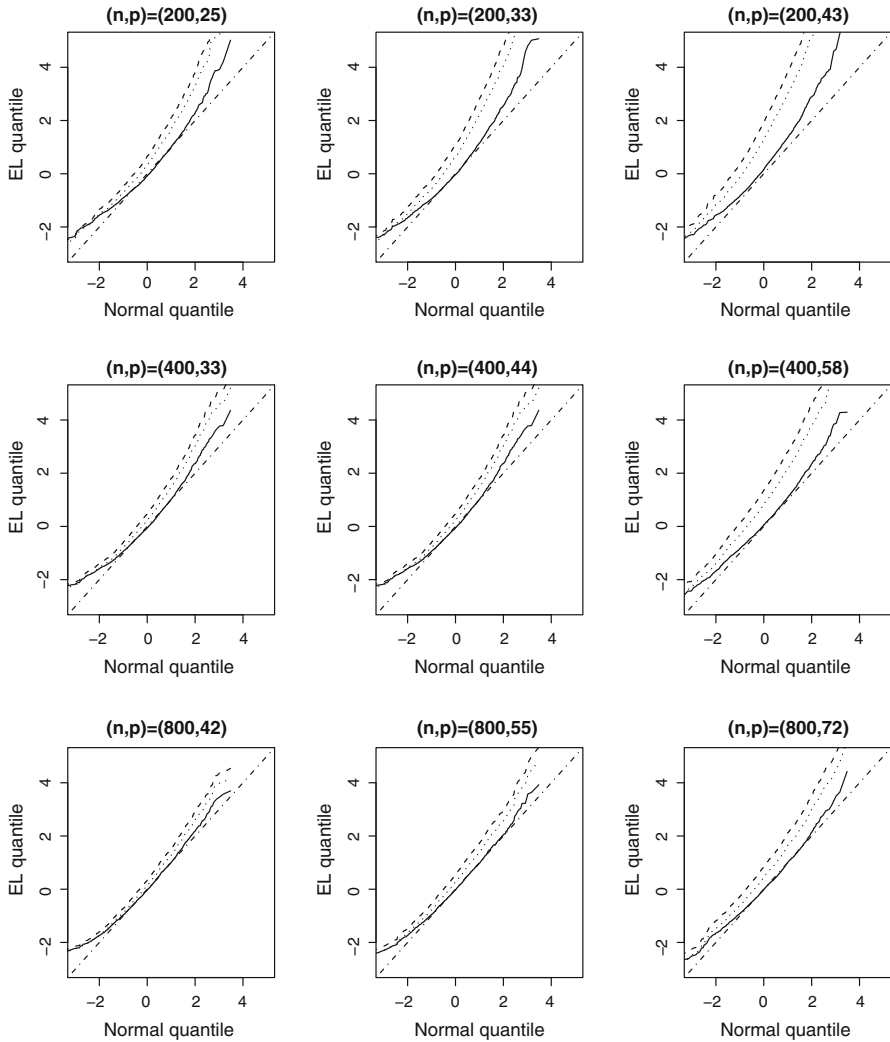
Fig. 2 Normal Q–Q plots of the standardized empirical likelihood ratios with $t_{p,5}$: MEL (*solid line*), SEL (*dashed lines*) and STEL (*dotted lines*)

$(p, 2p)$, and $(\hat{E}_{ni}, \hat{V}_{ni})$, $i = 1, 2$, which reflect the convergence of SEL, STEL and MEL to $N(0, 1)$ respectively. Here we only present the results of $t_{p,5}$ (Scenario II) since the results for other scenarios are similar. There is a general convergence of all three standardized ELRs to $N(0, 1)$ as n and p increase simultaneously. The MEL has the best goodness-of fit in all cases and outperform the other two methods by quite a large margin when n is small. We also observe from the Q–Q plots that the MEL does not improve the lack-of-fit at lower quantiles. This is not surprising. Analogous phenomenon has been mentioned in many applications of the Edgeworth expansion where the improvements of coverage accuracies at lower quantiles are usually not as substantial as those at upper quantiles (Hall 1992). However, in the use of EL for

constructing confidence regions or testing, upper quantiles are of our major concern. Therefore, the MEL can be regarded as a reasonable alternative for the calibration of the ELR in practice.

5 Discussion

An issue of using the EL is that for each given μ_{0n}, the ELR function $EL_n(\mu_{0n})$ is well defined only if the convex hull of $\{X_1 - \mu_{0n}, \ldots, X_n - \mu_{0n}\}$ contains the p-dimensional zero vector. When n is not large, or when a good candidate (vector) value of μ_{0n} is not available, this convex hull often fails to contain 0 (Owen 2001). Chen et al. (2008) indicated that when the maximum EL solution does not exist, blindly setting $EL_n(\mu_{0n}) = \infty$ as suggested in the literature fails to provide any information on whether μ_{0n} is grossly unfit to the data or is in fact only slightly off an appropriate value. Also, it is numerically difficult to determine that there is no solution. This problem is particularly critical in testing problems since μ_n may be far away from μ_{0n}. The situation would be definitely more serious for high-dimensional data streams as Tsao (2004) pointed out when p is moderately large but fixed, the distribution of $EL_n(\mu_{0n})$ has an atom at infinity for fixed n. This makes the practical implementation of EL methods infeasible and inconvenient. Chen et al. (2008), Liu and Chen (2010), Emerson and Owen (2009) and Liu and Yu (2010) propose several remedies, through adding some artificial data points to the observed sample. It is of great interest and also necessity to study the calibration of these adjusted EL for high-dimensional data.

6 Appendix

6.1 Appendix A: Proof of Theorem 1

We prove Theorem 1 by proving the following two propositions.

Proposition 2 *Assume conditions* (C1), (C2) *and* (C5) *are valid. If condition* (C4) *holds or both conditions* (C3) *and* (C6) *hold, then (7) is valid.*

Proposition 3 *Under conditions* (C1), (C3) *and* (C5), *(8) is valid.*

To prove the propositions, we begin with several necessary lemmas. The first lemma comes from Hjort et al. (2009) and we omit the proof. Throughout the following proofs, let $L_n = \max_{1 \leq i,j \leq p} |\Sigma_{n,i,j} - S_{n,i,j}|$ where $\Sigma_{n,i,j}$ and $S_{n,i,j}$ denote the (i,j) components of the theoretical and sample covariance matrices Σ_n and S_n.

Lemma 1 *Let* $A_n(p,q) = p^{-1} \sum_{j=1}^{p} E(|X_{1j}|^q)$. *For any* $\varepsilon > 0$, *there exists a constant* $c(q) > 0$ *such that* $P(L_n \geq \varepsilon) \leq \frac{c(q)p^2}{n^{q/2}\varepsilon^q} A_n(p,q)^2$.

Lemma 2 *Let* $\gamma_i(\Sigma)$ *denote the i-th largest eigenvalue of a symmetric matrix* Σ. *It holds that* $\max_{1 \leq i \leq p} |\gamma_i(\Sigma_n) - \gamma_i(S_n)| \leq pL_n$.

Proof The fact that Σ_n and S_n are both non-negative matrix implies that $\gamma_i(\Sigma_n) = \gamma_i^{1/2}(\Sigma_n^2)$ and $\gamma_i(S_n) = \gamma_i^{1/2}(S_n^2)$. It then follows that

$$\max_{1 \le i \le p} |\gamma_i(\Sigma_n) - \gamma_i(S_n)|^2 \le \sum_{i=1}^{p} |\gamma_i^{1/2}\left(\Sigma_n^2\right) - \gamma_i^{1/2}\left(S_n^2\right)|^2$$

$$= \operatorname{tr}\left(\Sigma_n^2\right) + \operatorname{tr}\left(S_n^2\right) - 2\sum_{i=1}^{p} \gamma_i(S_n)\gamma_i(\Sigma_n).$$

By Von Neymann's inequality, $\sum_{i=1}^{p} \gamma_i(\Sigma_n)\gamma_i(S_n) \ge \operatorname{tr}(S_n \Sigma_n)$, we have

$$\max_{1 \le i \le p} |\gamma_i(\Sigma_n) - \gamma_i(S_n)| \le \{\operatorname{tr}(\Sigma_n - S_n)^2\}^{1/2} \le pL_n,$$

where the last inequality holds because the absolute value of each element of $(\Sigma_n - S_n)^2$ is no greater than L_n^2. $\qquad \square$

Under condition (C1) we have

$$P(pL_n \ge \varepsilon) \le \frac{c(q)p^{2+q}}{n^{q/2}\varepsilon^q} K^2. \tag{12}$$

Furthermore if condition (C3) also holds, then $pL_n = o_p(1)$. Based on Lemmas 1 and 2, we have the following corollary.

Corollary 1 *Under conditions* (C1) *and* (C3)*, the inequality* $\gamma_p(S_n) \ge c_0$ *holds with probability tending to one as* $n \to \infty$.

By Lagrange multiplier method we get

$$\mathrm{EL}_n(0) = 2\sum_{i=1}^{n} \log\left(1 + \lambda_*^\mathsf{T} X_i\right), \tag{13}$$

where λ_* is the solution to

$$\sum_{i=1}^{n} \frac{X_i}{1 + \lambda^\mathsf{T} X_i} = 0. \tag{14}$$

To expand (14), we need to guarantee that $\sup_{1 \le i \le n} |\lambda_*^\mathsf{T} X_i| = o_p(1)$, which requires studying the magnitude of $B_n \equiv \max_{1 \le i \le n} \|X_i\|$.

Lemma 3 *If* X_i *has finite* q-th *absolute moment, then* $B_n = o_p(n^{1/q}\|X_i\|_q)$.

Proof It is clear that

$$\max_{1 \le i \le n} \|X_i\| \le \left\{\max_{1 \le i \le n} \left|\|X_i\|^{q/2} - \mathrm{E}\left(\|X_i\|^{q/2}\right)\right| + \mathrm{E}\|X_i\|^{q/2}\right\}^{2/q},$$

and

$$\max_{1 \leq i \leq n} \left| \|X_i\|^{q/2} - \mathrm{E}\|X_i\|^{q/2} \right| \left\{ \mathrm{Var}\left(\|X_i\|^{q/2} \right) \right\}^{-1/2} = o_p\left(n^{1/2} \right).$$

Therefore we have

$$B_n = \max_{1 \leq i \leq n} \|X_i\| = o_p\left(n^{1/q} \left[\mathrm{Var}\left(\|X_1\|^{q/2} \right) \right]^{1/q} \right) + O_p\left(\left[\mathrm{E}\|X_1\|^{q/2} \right]^{2/q} \right).$$

Since $\mathrm{Var}(\|X_1\|^{q/2}) \leq \mathrm{E}\|X_1\|^q$ and $[\mathrm{E}\|X_1\|^{q/2}]^{2/q} \leq \|X_1\|_q$, we get

$$B_n = \max_{1 \leq i \leq n} \|X_i\| = o_p\left(n^{1/q} \|X_i\|_q \right).$$

$\square$

Lemma 4 *Let $A = (a_{ij})$ denote a $p \times p$ matrix and $Ma(A) = \max_{1 \leq i, j \leq n} |a_{ij}|$. If A is non-negative definite, then $Ma(A) \leq \gamma_1(A)$.*

Proof Suppose A has eigen decomposition $A = Q\Gamma Q^{\mathsf{T}}$, where $\Gamma = \mathrm{diag}\{\gamma_1, \cdots, \gamma_p\}$ and $Q = (q_{ij})$ is an orthogonal matrix. It is clear that Q^{T} is also an orthogonal matrix. Therefore $\sum_{k=1}^p q_{ik}^2 = \sum_{i=1}^p q_{ik}^2 = 1$. Then the result $Ma(A) \leq \lambda_1$ can be concluded from

$$|a_{ij}| = \left| \sum_{k=1}^p q_{ik} \gamma_k(A) q_{jk} \right| \leq \left(\sum_{k=1}^p q_{ik}^2 \right)^{1/2} \left(\sum_{k=1}^p [\gamma_k(A)]^2 q_{jk}^2 \right)^{1/2}$$

$$\leq \gamma_1(A) \left(\sum_{k=1}^p q_{jk}^2 \right)^{1/2} = \gamma_1(A).$$

$\square$

Lemma 5 *Assume conditions* (C1), (C3) *and* (C2). *Then* $\sup_{1 \leq i \leq n} |\lambda_*^{\mathsf{T}} X_i| = o_p(1)$ *and* $lambda_* = S_n^{-1}(\bar{X} + \beta_n)$, *where* $\bar{X} = O_p(n^{-1/2} p^{1/2} [\gamma_1(\Sigma_n)]^{1/2})$ *and* $\beta_n = O_p(n^{(1-q)/q} p \|X_1\|_q [\gamma_1(\Sigma_n)]^2)$.

Proof Let $\lambda_* = \|\lambda_*\| u$ where u is a vector with length one. By the same argument as Owen (1990)(page 101) we have

$$\frac{1}{n} \sum_{i=1}^n \frac{u^{\mathsf{T}} X_i}{1 + \lambda_*^{\mathsf{T}} X_i} = 0 \Rightarrow u^{\mathsf{T}} \bar{X}(1 + \|\lambda_*\| B_n) \geq u^{\mathsf{T}} S_n u \|\lambda_*\|$$

$$\Rightarrow \|\lambda_*\|(u^{\mathsf{T}} S_n u - u^{\mathsf{T}} \bar{X} B_n) \leq u^{\mathsf{T}} \bar{X}.$$

It can be seen that $\bar{X} = O_p(p^{1/2} n^{-1/2} [\gamma_1(\Sigma_n)]^{1/2})$ from $\mathrm{E}(\bar{X}^{\mathsf{T}} \bar{X}) = \mathrm{tr}(\Sigma_n)/n \leq p\gamma_1(\Sigma_n)/n$. Since $|u^{\mathsf{T}} \bar{X}| \leq \|\bar{X}\|$ we get $u^{\mathsf{T}} \bar{X} = O_p(p^{1/2} n^{-1/2} [\gamma_1(\Sigma_n)]^{1/2})$. It follows that

$$u^{\mathsf{T}} \bar{X} B_n = o_p\left(p^{1/2} n^{-1/2} [\gamma_1(\Sigma_n)]^{1/2} \cdot n^{1/q} \|X_i\|_q \right) = o_p(1).$$

where the last equality holds because of condition (C2). According to Corollary 1, $u^\mathsf{T} S_n u > c_0$ holds with probability tending to one as $n \to \infty$. Therefore we have $\|\lambda_*\| = O(|u^\mathsf{T} \bar{X}|/c_0) = O(\|\bar{X}\|) = O_p(n^{-1/2} p^{1/2} [\gamma_1(\Sigma_n)]^{1/2})$ and

$$\sup_{1 \leq i \leq n} \left| \lambda_*^\mathsf{T} X_i \right| = \|\lambda_*\| \cdot B_n = O \left(\left| u^\mathsf{T} \bar{X} \right| B_n \right) = o_p(1).$$

Now expanding (14) gives $\bar{X} - S_n \lambda_* + \beta_n = 0$, where $\beta_n = \frac{1}{n} \sum_{i=1}^n X_i (X_i^\mathsf{T} \lambda_*)^2 (1 + o_p(1))$. So we have $\lambda_* = S_n^{-1}(\bar{X} + \beta_n)$, and

$$\|\beta_n\| \leq \frac{2}{n} \sum_{i=1}^n \|X_i\| \left| X_i^\mathsf{T} \lambda_* \right|^2 \leq \frac{2}{n} \sum_{i=1}^n B_n \left(\lambda_*^\mathsf{T} X_i X_i^\mathsf{T} \lambda_* \right)$$
$$= O \left(\|\lambda_*\|^2 \gamma_1(\Sigma_n) B_n \right) = O \left(\|\bar{X}\|^2 \gamma_1(\Sigma_n) B_n \right)$$
$$= O_p \left(n^{(1-q)/q} p \|X_1\|_q [\gamma_1(\Sigma_n)]^2 \right).$$

It can be seen from the last but one equality that $\beta_n = \bar{X} \cdot o_p(1)$. □

Lemma 6 *If conditions (C1) and (C3) hold, then* $(S_n^{-1} - \Sigma_n^{-1})\bar{X} = (\Sigma_n^{-1} \bar{X}) o_p(1)$.

Proof By $\gamma_1(\Sigma_n^{-1}) \leq 1/c_0$, we know $\gamma_1(S_n^{-1}) \leq 1/c_0$ with probability tending to one. Using Lemma 2, we have

$$\|\Sigma_n(S_n^{-1} - \Sigma_n^{-1})\bar{X}\| = \|(\Sigma_n - S_n) S_n^{-1} \bar{X}\|$$
$$\leq \max_{1 \leq i \leq p} |\gamma_i(\Sigma_n - S_n)| \gamma_1(S_n^{-1}) \|\bar{X}\|$$
$$\leq p L_n \|\bar{X}\|/c_0 = \|\bar{X}\| \cdot o_p(1),$$

which implies $(S_n^{-1} - \Sigma_n^{-1})\bar{X} = (\Sigma_n^{-1} \bar{X}) o_p(1)$. □

Combining Lemmas 5 and 6 results in the following corollary:

Corollary 2 *Assume conditions* (C1), (C3) *and* (C2). *Then* $\lambda_* = \Sigma_n^{-1} \bar{X}(1 + o_p(1))$.

Proof of Proposition 1 Noting that $\max_{1 \leq i \leq p} \lambda_*^\mathsf{T} X_i = o_p(1)$, by Taylor's expansion of (13), we have

$$\text{EL}_n(0) = 2 \sum_{i=1}^n \left\{ \lambda_*^\mathsf{T} X_i - \frac{1}{2} \lambda_*^\mathsf{T} X_i X_i^\mathsf{T} \lambda_* + \frac{1}{3} \left(X_i^\mathsf{T} \lambda_* \right)^3 (1 + o_p(1)) \right\}$$
$$= n \left(2 \bar{X}^\mathsf{T} \lambda_* - \lambda_*^\mathsf{T} S_n \lambda_* \right) + \delta_1$$

where $\delta_1 = \frac{2}{3} \sum_{i=1}^n (X_i^\mathsf{T} \lambda_*)^3 (1 + o_p(1))$. Replacing λ_* with $S_n^{-1}(\bar{X} + \beta_n)$ in the above expression leads to $\text{EL}_n(0) = n \bar{X}^\mathsf{T} S_n^{-1} \bar{X} - \delta_2 + \delta_1$ where $\delta_2 = n \beta_n^\mathsf{T} S_n^{-1} \beta_n$. The theorem is proved by showing that $\delta_1 = o_p(\sqrt{p})$ and $\delta_2 = o_p(\sqrt{p})$.

According to Corollary 2 and Lemma 5,

$$\delta_1 = \frac{2}{3}\sum_{i=1}^{n}\left(X_i^\mathsf{T}\Sigma_n^{-1}\bar{X}\right)^3(1+o_p(1)), \quad \beta_n = \frac{1}{n}\sum_{i=1}^{n}X_i\left(X_i^\mathsf{T}\Sigma_n^{-1}\bar{X}\right)^2(1+o_p(1)).$$

Along the line of proving Lemma 6, we can show that $\delta_2 = n\beta_n^\mathsf{T}\Sigma_n^{-1}\beta_n(1+o_p(1))$. Recall the definition of Y_r and $\alpha^{i_1i_2...i_k}$. Let

$$A^{i_1i_2...i_k} = \frac{1}{n}\sum_{j=1}^{n}Y_{j,i_1}\ldots Y_{j,i_k} - \alpha^{i_1i_2...i_k}. \tag{15}$$

Then δ_1 and δ_2 can be rewritten as

$$\delta_1 = \frac{2n}{3}\alpha^{ijk}A^iA^jA^k(1+o_p(1)), \tag{16}$$
$$\delta_2 = n\alpha^{ijk}\alpha^{irs}A^jA^kA^rA^s(1+o_p(1)),$$

where we adopt the summation convention employed in conditions (C5) and (C6). With tedious calculation we obtain

$$\mathrm{E}\left(\frac{2n}{3}\alpha^{ijk}A^iA^jA^k\right)^2 = \frac{4}{3n}\left(3\alpha^{ijj}\alpha^{ikk}+2\alpha^{ijk}\alpha^{ijk}\right)(1+o(1));$$
$$\mathrm{E}\left(n\alpha^{ijk}\alpha^{irs}A^jA^kA^rA^s\right) = n^{-1}\left(2\alpha^{ijk}\alpha^{ijk}+\alpha^{ijj}\alpha^{ikk}\right)(1+o(1)).$$

It can be proved that $|\alpha^{ijk}\alpha^{ijk}| \le \alpha^{iijj}\alpha^{kk}$ and $|\alpha^{ijj}\alpha^{ikk}| \le \frac{1}{2}(\alpha^{iijj}\alpha^{kk}+\alpha^{iikk}\alpha^{kk})$. Therefore

$$\mathrm{E}\left(\frac{2n}{3}\alpha^{ijk}A^iA^jA^k\right)^2 \le \frac{20}{3n}p\alpha^{iijj} = O\left(p^3/n\right),$$
$$\mathrm{E}\left(n\alpha^{ijk}\alpha^{irs}A^jA^kA^rA^s\right) \le \frac{3}{n}p\alpha^{iijj} = O\left(p^3/n\right).$$

Thus we have $\delta_1 = O_p(p^{3/2}/n^{1/2}) = o_p(\sqrt{p})$. Since $\alpha^{ijk}\alpha^{irs}A^jA^kA^rA^s \ge 0$, it can be seen that $\delta_2 = O(p^3/n)$ by Markov's inequality.

It is clear that $p = o(n^{2/5})$ is sufficient for $\delta_2 = o_p(\sqrt{p})$; while if both $\alpha^{ijk}\alpha^{ijk} = O(p^{5/2})$ and $\alpha^{ijj}\alpha^{ikk} = O(p^{5/2})$, then the rate of p can be relaxed to $p = o(n^{1/2})$. □

Lemma 7 Denote $D_n = I - V_n$ with $V_n = \Sigma_n^{-1/2}S_n\Sigma_n^{-1/2}$. If condition (C5) holds, then $\mathrm{tr}(D_n^2) = O_p(p^2/n)$.

Proof It is easy to see that $\alpha^{ij} = 1$ for $i = j$ and 0 otherwise. Note that

$$D_n^2 = \left(I - \frac{1}{n}\sum_{r=1}^{n}Y_rY_r^\mathsf{T}\right)^2 = I - \frac{2}{n}\sum_{r=1}^{n}Y_rY_r^\mathsf{T} + \frac{1}{n^2}\sum_{r,s=1}^{n}Y_rY_r^\mathsf{T}Y_sY_s^\mathsf{T}.$$

Therefore we have

$$
\begin{aligned}
\text{Etr}\left(D_n^2\right) &= p - \frac{2}{n}\sum_{i=1}^{p}\sum_{r=1}^{n}\alpha_{ii} + \frac{1}{n^2}\sum_{i,j=1}^{p}\sum_{r,s=1}^{n}\text{E}Y_{ri}Y_{rj}^{\mathsf{T}}Y_{sj}Y_{si}^{\mathsf{T}} \\
&= -p + \frac{1}{n^2}\sum_{i,j=1}^{p}\left(\sum_{r\neq s}+\sum_{r=s}\right)\text{E}Y_{ri}Y_{rj}^{\mathsf{T}}Y_{sj}Y_{si}^{\mathsf{T}} \\
&= -p + \frac{n(n-1)}{n^2}\sum_{i,j=1}^{p}\alpha^{ij}\alpha^{ji} + \frac{n}{n^2}\sum_{i,j=1}^{p}\alpha^{iijj} \\
&= \left(\alpha^{iijj}-p\right)/n = O\left(p^2/n\right).
\end{aligned}
$$

$\square$

Proof of Proposition 2 Let $\bar{Y} = \Sigma_n^{-1/2}\bar{X}$. Then $\eta_n = n\bar{Y}^{\mathsf{T}}(I - V_n^{-1})\bar{Y}$. Note that $I - V_n^{-1} = D_n - D_n^2 + \cdots - (-1)^k D_n^k - (-1)^k D_n^k(I - V_n^{-1})$. Hence

$$
\eta_n = n\left\{\bar{Y}^{\mathsf{T}}D_n\bar{Y} - \bar{Y}^{\mathsf{T}}D_n^2\bar{Y} + \cdots - (-1)^k\bar{Y}^{\mathsf{T}}D_n^k\bar{Y}\right\} + (-1)^k n\bar{Y}^{\mathsf{T}}D_n^k(I - V_n^{-1})\bar{Y}.
\tag{17}
$$

Now we study the convergence of this expansion. Note that

$$
\|\bar{Y}\|^2 = \bar{X}^{\mathsf{T}}\Sigma^{-1}\bar{X} \leq \gamma_1\left(\Sigma_n^{-1}\right)\|\bar{X}\|^2 \leq c_0\|\bar{X}\|^2 = O_p\left(\gamma_1(\Sigma)n^{-1}p\right),
$$

and that $|\gamma_j(D_n)| \leq [\text{tr}(D_n^2)]^{1/2} = O_p(p/\sqrt{n})$ according to Lemma 7. So we have

$$
\left|\bar{Y}^{\mathsf{T}}D_n^k\bar{Y}\right| \leq \|\bar{Y}\|^2 \max_{1\leq j\leq p}\left|\gamma_j\left(D_n^k\right)\right| \leq \|\bar{Y}\|^2\left[\text{tr}\left(D_n^2\right)\right]^{k/2} \leq O_p\left(\gamma_1(\Sigma)p^{k+1}/n^{(k+2)/2}\right),
$$

which means the series $n\sum_{k=1}^{\infty}\bar{Y}^{\mathsf{T}}D_n^k\bar{Y}(-1)^{k-1}$ is convergent for fixed n when $p = o(n^{1/2})$. With lengthy algebra we have $\text{E}|n\bar{Y}^{\mathsf{T}}D_n\bar{Y}|^2 = n^2\text{E}|\bar{X}^{\mathsf{T}}(\Sigma_n - S_n)\bar{X}|^2 = O(p^3/n)$, which implies $n\bar{Y}^{\mathsf{T}}D_n\bar{Y} = O_p(p^{3/2}/\sqrt{n}) = o_p(\sqrt{p})$. Thus, $n\sum_{k=1}^{\infty}\bar{Y}^{\mathsf{T}}D_n^k\bar{Y}(-1)^{k-1} = o_p(\sqrt{p})$.

The remaining task is to prove that the remainder term in (17) is negligible as $k \to \infty$. It is clear that

$$
|(-1)^k n\bar{Y}^{\mathsf{T}}D_n^k\left(I - V_n^{-1}\right)\bar{Y}| \leq \left|n\bar{Y}^{\mathsf{T}}D_n^k\bar{Y}\right| + \left|n\bar{Y}^{\mathsf{T}}D_n^k V_n^{-1}\bar{Y}\right|,
$$

where the first term has been shown to be $O_p(\gamma_1(\Sigma)p^{k+1}/n^{k/2})$. Note that $|n\bar{Y}^{\mathsf{T}}D_n^k V_n^{-1}\bar{Y}| \leq np\text{Ma}(D_n^k V_n^{-1})\|\bar{Y}\|^2$, where the operator $\text{Ma}(\cdot)$ is defined in Lemma 4.

It is clear from Corollary 1 that $\gamma_1(V^{-1}) = \gamma_1(S_n^{-1}\Sigma_n) \leq \gamma_1(\Sigma_n)/c_0$ with probability tending to one. According to Lemma 4, we have

$$\mathrm{Ma}(D_n^k V_n^{-1}) \leq p\mathrm{Ma}(D_n^k)\mathrm{Ma}(V_n^{-1}) \leq p \max_{1 \leq i \leq p} |\gamma_i(D_n^k)| \max_{1 \leq i \leq p} |\gamma_i(V_n^{-1}))$$

$$= O_p(p^{k+1} n^{-k/2} \gamma_1(\Sigma_n)).$$

This leads to $|n\bar{Y}^\mathsf{T} D_n^k V_n^{-1} \bar{Y}| = O_p(\gamma_1^2(\Sigma) p^{k+3}/n^{k/2})$.

Finally, we have $|(-1)^k n \bar{Y}^\mathsf{T} D_n^k (I - V_n^{-1}) \bar{Y}| = O_p(\gamma_1^2(\Sigma) p^{k+3}/n^{k/2})$, which converges to zero as $k \to \infty$ because of conditions (C3). □

6.2 Appendix B: calculation of E_{ni} and V_{ni} for $i = 1, 2$

In what follows, we adopt the notation defined in (15) and the summation convention employed in condition (C6) for convenience. It is clear that $\alpha^{ij} = \delta^{ij}$, where δ^{ij} is the Kronecker delta, that is $\delta^{ij} = 1$ if $i = j$; otherwise, $\delta^{ij} = 0$. Let B^{ij} denote the (i, j) element of a matrix B.

Note that $T_n = n\bar{X}^\mathsf{T} S_n^{-1} \bar{X} = n\bar{Y}^\mathsf{T}(\Sigma_n^{1/2} S_n^{-1} \Sigma_n^{1/2})\bar{Y}$. Since $(\Sigma_n^{-1/2} S_n \Sigma_n^{-1/2})^{ij} = \delta^{ij} + A^{ij}$, it can be verified that

$$\left(\Sigma_n^{1/2} S_n^{-1} \Sigma_n^{1/2}\right)^{ij} = \delta^{ij} - A^{ij} + A^{ik} A^{kj} - A^{ik} A^{kl} A^{lj} + \cdots.$$

Consequently, we know

$$T_n = n(A^i A^i - A^i A^j A^{ij} + A^i A^j A^{ik} A^{kj} - A^i A^j A^{ik} A^{kl} A^{lj} + \cdots),$$

and its expectation $ET_n = EnA^i A^i - EnA^i A^j A^{ij} + EnA^i A^j A^{ik} A^{kj}(1 + o(1))$. Note that $E(A^i A^i) = p/n$, $E(A^i A^j A^{ij}) = n^{-2}(\alpha^{iijj} - p)(1 + o(1))$ and

$$E(A^i A^j A^{ik} A^{kj}) = \frac{1}{n^2}(\alpha^{iijj} - p + \alpha^{kii}\alpha^{kjj} + \alpha^{ijk}\alpha^{ijk})(1 + o(1)),$$

which result in $E(T_n) = p + \frac{1}{n}(\alpha^{kii}\alpha^{kjj} + \alpha^{ijk}\alpha^{ijk})(1 + o(1))$.

To get V_{n1}, we need to calculate ET_n^2:

$$ET_n^2 = n^2 E(A^i A^i A^j A^j - 2A^i A^j A^{ij} A^r A^r + A^i A^j A^{ij} A^r A^s A^{rs}$$
$$+ 2A^i A^j A^{ik} A^{kj} A^r A^r + \cdots)$$
$$\equiv I_1 - 2I_2 + I_3 + 2I_4 + \cdots,$$

where

$$I_1 = p^2 + 2p + \frac{1}{n}\left(\alpha^{iijj} - p^2 - 2p\right)(1 + o(1));$$

$$I_2 = \frac{1}{n}[(p + 5)\alpha^{iijj} - 2p^2 - 4p + 2\alpha^{iij}\alpha^{rrj} + 2\alpha^{rij}\alpha^{rij}](1 + o(1));$$

$$I_3 = \frac{1}{n}\{3\alpha^{iijj} - p^2 - 2p + 8\alpha^{jjk}\alpha^{iik} + 4\alpha^{ijk}\alpha^{ijk}\}(1 + o(1));$$

$$I_4 = \frac{1}{n}\{(p+2)\alpha^{iijj} - p^2 - 2p + (p+6)\alpha^{ijk}\alpha^{ijk} + (p+4)\alpha^{jjk}\alpha^{iik}\}(1+o(1)).$$

Finally, we get

$$\mathrm{E}T_n^2 = p^2 + 2p - \frac{2}{n}\alpha^{iijj} + \frac{2p+12}{n}(\alpha^{jjk}\alpha^{iik} + \alpha^{ijk}\alpha^{ijk})(1+o(1)),$$

$$\mathrm{Var}(T_n) = \mathrm{E}T_n^2 - (\mathrm{E}T_n)^2$$

$$= p^2 + 2p - \frac{2}{n}\alpha^{iijj} + \frac{2p+12}{n}(\alpha^{jjk}\alpha^{iik} + \alpha^{ijk}\alpha^{ijk})$$

$$- \left\{ p + \frac{1}{n}\alpha^{ijk}\alpha^{ijk} + \frac{1}{n}\alpha^{kii}\alpha^{kjj} \right\}^2 + \cdots$$

$$= 2p - \frac{2}{n}\alpha^{iijj} + \frac{12}{n}(\alpha^{jjk}\alpha^{iik} + \alpha^{ijk}\alpha^{ijk}) + \cdots.$$

To calculate E_{n2} and V_{n2}, we note that $T_{nc} = n\bar{Y}^\mathsf{T}(\Sigma_n^{1/2}S_{nc}^{-1}\Sigma_n^{1/2})\bar{Y}$ and $(\Sigma_n^{-1/2}S_n\Sigma_{nc}^{-1/2})^{ij} = \delta^{ij} + A^{ij} - A^iA^j$. Similarly,

$$(\Sigma_n^{1/2}S_n^{-1}\Sigma_n^{1/2})^{ij} = \delta^{ij} - A^{ij} + A^iA^j + A^{ik}A^{kj} - 2A^{ik}A^kA^j - A^{ik}A^{kl}A^{lj} + \cdots.$$

Thus, we have

$$T_{nc} = n(A^iA^i - A^iA^jA^{ij} + A^iA^jA^iA^j + A^iA^jA^{ik}A^{kj} - 2A^iA^jA^{ik}A^kA^j) + \cdots$$

$$T_{nc}^2 = n^2(A^iA^iA^rA^r - 2A^rA^rA^iA^jA^{ij} + A^rA^sA^{rs}A^iA^jA^{ij} + 2A^rA^rA^iA^jA^iA^j$$

$$+ 2A^rA^rA^iA^jA^{ik}A^{kj}) + \cdots.$$

Some algebra leads to $\mathrm{E}nA^iA^jA^iA^j = \frac{1}{n}(p^2 + 2p)$ and $n^2\mathrm{E}(A^rA^rA^iA^jA^iA^j) = \frac{1}{n}(p^3 + 6p^2 + 8p)$. By comparing the expansion of T_n and T_{nc}, it is straightforward to see

$$\mathrm{E}T_{nc} = p + \frac{1}{n}(p^2 + 2p + \alpha^{kii}\alpha^{kjj} + \alpha^{ijk}\alpha^{ijk})(1+o(1)),$$

$$\mathrm{E}T_{nc}^2 = p^2 + 2p - \frac{2}{n}\alpha^{iijj} + \frac{1}{n}(2p^3 + 12p^2 + 16p)$$

$$+ \frac{2p+12}{n}(\alpha^{jjk}\alpha^{iik} + \alpha^{ijk}\alpha^{ijk})(1+o(1)),$$

and therefore

$$\mathrm{Var}(T_{nc}) = 2p - \frac{2}{n}\alpha^{iijj} + \frac{8p^2 + 16p}{n} + \frac{12}{n}(\alpha^{jjk}\alpha^{iik} + \alpha^{ijk}\alpha^{ijk}) + \cdots.$$

6.3 Appendix C: Proof of claim (9)

Suppose the model (3) is valid for $k \geq 1$. Let $\beta^{ijkl} = \mathrm{E}(X_{1i}X_{1j}X_{1k}X_{1l})$, $\xi^{ijkl} = \mathrm{E}(Z_{1i}Z_{1j}Z_{1k}Z_{1l})$, $\Sigma^{-1/2} = (H^{ij})$, $\Gamma = (\Gamma^{ij})$, $\Sigma^{-1} = (B^{ij})$. Apparently, $\Sigma^{ij} = \Gamma^{ir}\Gamma^{jr}$ and for fixed $r \neq s$, $\xi^{rrss} = 1$ and $\xi^{rrrr} = m_4 \in (0, \infty)$. Then, we have $\beta^{ijkl} = \Sigma^{ij}\Sigma^{kl} + \Sigma^{ik}\Sigma^{jl} + \Sigma^{il}\Sigma^{kj} + \Gamma^{ir}\Gamma^{jr}\Gamma^{kr}\Gamma^{lr}(m_4 - 3)$. Now, we separately calculate α^{iijj}, $\alpha^{ijk}\alpha^{ijk}$ and $\alpha^{ijj}\alpha^{ijk}$. First,

$$\alpha^{iijj} = H^{ir}H^{is}H^{ju}H^{jv}\beta^{rsuv} = p^2 + 2p + B^{rs}B^{uv}\Gamma^{rt}\Gamma^{st}\Gamma^{ut}\Gamma^{vt}(m_4 - 3).$$

Let $a_t = B^{rs}\Gamma^{rt}\Gamma^{st}$ and $E = B^{rs}B^{uv}\Gamma^{rt}\Gamma^{st}\Gamma^{ut}\Gamma^{vt}$. Then $E = \sum_{t=1}^{m} a_t^2$ and $\sum_{t=1}^{m} a_t = p$. Note that Σ^{-1} is nonsingular, so $a_t > 0$ and furthermore

$$E = \sum_{t=1}^{m} a_t^2 \leq \max_{1 \leq i \leq m} a_i \sum_{t=1}^{m} a_t \leq \left(\sum_{t=1}^{m} a_t \right)^2 = p^2.$$

Thus we have $\alpha^{iijj} = O(p^2)$.

Note that $\beta^{ijk} = \Gamma^{ir}\Gamma^{js}\Gamma^{kt}\xi^{rst} = \Gamma^{ir}\Gamma^{jr}\Gamma^{kr}\xi^{rrr}$. Then, we have

$$\alpha^{ijk}\alpha^{ijk} = H^{ir}H^{is}H^{ju}\beta^{rsu}H^{ia}H^{ib}H^{jc}\beta^{abc} = C^{gh}C^{gh}C^{gh}\xi^{ggg}\xi^{hhh}$$

where $C^{gh} = B^{ra}\Gamma^{rg}\Gamma^{ah}$. Since for fixed r, $|\xi^{rrr}| \leq (\xi^{rrrr})^{3/4} = m_4^{3/4}$, it follows that

$$|\alpha^{ijk}\alpha^{ijk}| \leq \sum_{g,h} |C^{gh}|^3 m_4^{3/2} \leq m_4^{3/2} \max_{g,h} |C^{g,h}| \cdot \sum_{g,h} |C^{gh}|^2. \tag{18}$$

It can be seen that $\max_{g,h} |C^{g,h}| \leq \max_g C^{h,h} \leq \sum_{h=1}^{p} C^{h,h} = p$ and $\sum_{g,h} (C^{gh})^2 = p$. This together with (18) leads to $|\alpha^{ijk}\alpha^{ijk}| = O(p^2)$. Similarly, $\alpha^{ijj}\alpha^{ikk} = O(p^2)$.

6.4 Appendix D: On the assertion that $\mathrm{EL}_n(0)$ is almost well-defined.

Lemma 8 *Let $X_1, \cdots, X_n$ be i.i.d. random variables with dimension p, common mean μ_0 and nonsingular covariance matrix Σ. The event that μ_0 is inside the convex hull of point $X_1, \ldots, X_n$ happens with probability tending to 1 provided $p = o(n/\log n)$.*

Proof First of all, by Lemma 2 of Owen (1990), for any p there exists a constant $\varepsilon_0 > 0$ not depending on n such that

$$\inf_{\theta \in \Omega_p} P((X^{\top} - \mu_0)\theta > 0) \geq \varepsilon_0 \tag{19}$$

where Ω_p denotes the unit p-dimensional sphere in $\mathbb{R}^p$.

Next, we will show

$$\sup_{\theta \in \Omega_p} |P((X - \mu_0)^\top \theta > 0) - P_n((X - \mu_0)^\top \theta > 0)| \to 0, \quad a.s. \qquad (20)$$

where P_n denotes the empirical probability measure based on $X_1, \ldots, X_n$. In fact, this can be proved by using a generalization of the Glivenko–Cantelli theorem on Ω_p. By Theorem 12.12 in DasGupta (2008) (or Theorem 2 in Vapnik and Chervonenkis (1971)), for any $\epsilon > 0$, we have

$$P\left(\sup_{\theta \in \Omega_p} |P((X - \mu_0)^\top \theta > 0) - P_n((X - \mu_0)^\top \theta > 0)| > \epsilon\right)$$

$$\leq 8S(n, \mathbb{HS}^p)e^{-n\epsilon^2/32}$$

$$\leq 8 \sum_{i=0}^{\mathrm{VC}(\mathbb{HS}^p)} C_n^i e^{-n\epsilon^2/32}$$

$$= 8 \sum_{i=0}^{p+1} C_n^i e^{-n\epsilon^2/32} \leq 8n^{p+1} e^{-n\epsilon^2/32},$$

where $\mathbb{HS}^p$ is the collection of all linear half-spaces in $\mathbb{R}^p$, $S(n, \mathbb{HS}^p)$ is the nth shattering coefficient of $\mathbb{HS}^p$ and $\mathrm{VC}(\mathbb{HS}^p)$ is the Vapnik–Chervonenkis (VC) dimension of $\mathbb{HS}^p$. The second inequality above comes from Proposition 12. 1 in DasGupta (2008) and the equality holds by the classical result that $\mathrm{VC}(\mathbb{HS}^p) = p + 1$ (Vapnik and Chervonenkis (1971)). For $p = o(n/\log(n))$, we have $8n^{p+1}e^{-n\epsilon^2/32} \to 0$ as n tends to infinity, which means (20) holds.

It then follows from (20) and (19) that for any p

$$\inf_{\theta \in \Omega_p} P_n((X - \mu_0)^\top \theta > 0) \geq \frac{\varepsilon_0}{2}, \quad a.s.$$

This immediately leads to

$$P\left(\inf_{\theta \in \Omega_p} P_n((X - \mu_0)^\top \theta \geq 0) > 0\right) \to 1, \quad n \to \infty,$$

from which we know the convex hull of $X_1, \ldots, X_n$ contains μ_0 with probability tending to 1. Therefore, the empirical likelihood function $\mathrm{EL}(\mu_0)$ is almost well-defined for large n as long as $p = o(n/\log(n))$. $\qquad \square$

References

Bai, Z., Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistics Sinica*, *6*, 311–329.

Brown, B. M., Chen, S. X. (1998). Combined and least squares empirical likelihood. *Annals of the Institue of Statistical Mathematics*, *50*, 697–714.

Chen, J. H., Variyath, A. M., Abraham, B. (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, *17*, 426–443.

Chen, S. X., Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, *38*, 808–835.

Chen, S. X., Peng, L., Qin, Y.-L. (2009). Effects of data dimension on empirical likelihood. *Biometrika*, *96*, 1–12.

Chen, S. X., Zhang, L.-X., Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices. *Journal of American Statistian Assocciation*, *105*, 810–815.

DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. New York: Springer.

DiCiccio, T. J., Hall, P., Romano, J. P. (1991). Empirical likelihood is Bartlett correctable. *The Annals of Statistics*, *19*, 1053–1061.

Emerson, S. C., Owen, A. B. (2009). Calibration of the empirical likelihood method for a vector mean. *Econometrical Journal of Statisitcs*, *3*, 1161–1192.

Hall, P. (1992). *The bootstrap and Edgeworth expansion*. New York: Springer.

Hall, P., Hyde, C. C. (1992). *Martingale central limit theory and its applications*. New York: Academic Press.

Hjort, H. L., Mckeague, I. W., Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *The Annals of Statistics*, *37*, 1079–1115.

Liu, Y., Chen, J. (2010). Adjusted Empirical Likelihood with High-Order Precision. *The Annals of Statistics*, *38*, 1341–1362.

Liu, Y., Yu, C. W. (2010). Bartlett Correctable Two-Sample Adjusted Empirical Likelihood. *Journal of Multivariate Analysis*, *101*, 1701–1711.

Mardia, K. V., Kent, J. T., Bibby, J. M. (1979). *Multivariate analysis*. London: Academic Press.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*, 237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, *18*, 90–120.

Owen, A. B. (2001). *Empirical likelihood*. New York: Chapman & Hall/CRC.

Portnoy, S. (1985). Asymptotic behavior of M-estimations of p regression parameters with p^2/n is large. II. normal approximation. *The Annals of Statistics*, *13*, 1403–1417.

Qin, J., Lawless, J. (1994). Empirical likelihood and general equations. *The Annals of Statistics*, *22*, 300–325.

Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika*, *92*, 951–956.

Tang, C. Y., Leng, C. (2005). Penalized high-dimensional empirical likelihood. *Biometrika*, *97*, 905–920.

Tsao, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *The Annals of Statistics*, *32*, 1215–1221.

Vapnik, V., Chervonenkis, A. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, *16*, 264–280.