Calibrated Empirical Likelihood for High-Dimensional Data in Regression Model

Hong Guo[†], Changliang Zou[†] and Zhaojun Wang[†*]
[†] *LPMC and Department of Statistics, School of Mathematical Sciences, Nankai University, Tianjin 300071, China*

January 3, 2013

Abstract

High-dimensional data is becoming prevalent, and many new methodology and accompanying theory for high-dimensional data analysis have emerged in response. Empirical likelihood, as a classical nonparametric method of statistical inference, has been proved to possess many advantages. So, in this paper, we apply the empirical likelihood method to high-dimensional data in regression model with random and fixed designs respectively, and investigate it's performance in high-dimensional setting. The work of this paper is two-fold. First, we investigate the asymptotic behavior of empirical likelihood for regression coefficients and give the regularity conditions under which the standard normal calibration of empirical likelihood is valid in high dimensions. Second, to reduce the lack-of-fit and improve the coverage accuracy, we apply the calibration method proposed by Liu et al. (2012) to empirical likelihood for linear model with random regressor. Our simulation study results in finite sample settings indicate that the proposed calibration method has the best performance in all situations designed in the present paper, compared with other calibration methods involved in this paper. For the fixed design setting, we give empirically a calibration of empirical likelihood ratio function, which improve the coverage accuracy to some extent. This can be verified from the simulation results.

Key words and phrases: High-dimensional data; Empirical likelihood; Wilk's phenomenon;

1 Introduction

Empirical likelihood (EL) method, proposed by Owen (1988, 1991), is a nonparametric tool for statistical inference. It possesses both the reliability of nonparametric methods and effectiveness of parametric likelihood approaches. The increasing interest in EL method is mainly due to its attractive properties: nonparametric version of Wilk's theorem (Wilks, 1938) and Bartlett correction (DiCiccio et al., 1991; Chen and Cui, 2006). Owen (2001) provided a comprehensive overview on the EL.

Recently, high-dimensional data, whose dimensionality p tends to infinity as the sample size $n \to \infty$, is becoming prevalent. Such as Hyperspectral Imagery, Internet portrals, Finance data, especially datasets in genomics and other areas of computational biology. So high-dimensional data analysis is a very significant and active area of research. However,

[**]Corresponding author, email: zjwang@nankai.edu.cn

when dimensionality p diverges, traditional statistical methods may not cope with this kind of growth of dimensionality. Thereby, restudying the performance of traditional statistical methods in high-dimensional settings and while establishing completely new approaches for high-dimensional data are very necessary and exigent. There has been a large body of interesting work going on in these areas. For example, Chen et al. (2010) considered the problem that testing certain structures of covariance matrix in the situations where p and n are of the same order or p can be larger order than n. Fan and Li (2001), Zou and Hastie (2005), Wang et al. (2009) and others studied the problem of variable selection in various high-dimensional setting.

At the same time, the method of EL has also been applied to some high-dimensional problems and there is an increasing interest in the asymptotic behavior of the empirical likelihood ratio (ELR) when the sample size n and the dimension of observation p both tend to infinity. The leading work on this research includes Tang and Leng (2010) for penalized EL with application to high-dimensional variable selection; Chen et al. (2009) for asymptotic properties of the EL for mean under a general multivariate model; Hjote et al. (2009) for a general investigation on the ELR based on plug-in estimation.

Liu et al. (2012) considered the EL for population mean in high-dimensional settings. They analyzed the asymptotic behavior of EL under a general multivariate model and provided weak restrictive conditions under which the best rate $p = o(n^{1/2})$ for the asymptotic normality was achieved. Furthermore, they proposed a new calibration method for ELR, which had better performance in most situations, compared with other existing calibration methods.

In this paper, motivated by Liu et al. (2012), we consider EL for high-dimensional data in linear model with random and fixed designs respectively. We investigate the asymptotic behaviors of the ELR function for these two cases and then propose a new calibration method of the ELR function, which has better performance in terms of coverage accuracy than other calibration methods presented in this paper, observed from simulation study. The rest paper is organized as follows. In Section 2, we investigate the asymptotic results of ELR in linear model with random design and present the sufficient conditions under which the asymptotic normality holds. Furthermore, the calibrations of ELR are given in this section. The asymptotic normality of ELR for fixed design case is considered in Section 3. In Section 4, some simulation studies are conducted. Section 5 concludes the paper and gives further discussion. The technical proof of main results and some lemmas are summarized in Appendix.

2 EL in linear model with random design

In this section, we consider the linear model $Y_i = {X_i}^t\beta + \epsilon_i$, $1 \le i \le n$, where $\beta \in R^p$ is a column vector of unknown coefficients, $X_i \in R^p$ is a column random vector with dimension $p \ge 1$, $Y_i \in R$ and ϵ_i is a random variable with mean zero and finite variance σ^2. For convenient, we assume, in this paper, that X_i and ϵ_i are independent. In the regression model, the data are of the form (x_i^t, y_i) for $1 \le i \le n$, which are observations of n *i.i.d.* random vectors (X_i^t, Y_i). Without loss of generality, we assume $\mathrm{E}X_i = 0$, the random vector can be centered to have mean zero if otherwise, and $\Sigma_n = \mathrm{Var}(X_i)$ is positive definite. Let $S_n = \frac{1}{n}\sum_{i=1}^n X_i X_i^t$.

According to Owen (1991), the ELR function for β is defined as

$$l_n(\beta) = -2\sup\{\sum_{i=1}^{n}\log(n\omega_i) : \omega_i \geq 0, \sum_{i=1}^{n}\omega_i = 1, \sum_{i=1}^{n}\omega_i Z_i(\beta) = 0\}, \tag{2.1}$$

where $Z_i = Z_i(\beta) = X_i(Y_i - X_i^t\beta)$ is column vector of p components. It is known that Z_i's are *i.i.d.* by construction. Let β_0 denote the true value of β, then $\beta = \beta_0$ if and only if $\mathrm{E}Z_i = 0$. Hence, to test $\beta = \beta_0$, we only test whether $\mathrm{E}Z_i = 0$. According to Owen (1990), the EL inference may be done with respect to the random vector Z_i's.

For fixed p, Owen (1991) showed the following Wilk's phenomenon:

$$l_n(\beta_0) \xrightarrow{\mathcal{L}} \chi_p^2 \quad \text{as } n \to \infty, \tag{2.2}$$

if the second moments of $X_iX_i^t$ and X_iY_i exist.

In high dimension settings, p diverges to infinity as $n \to \infty$, and then the asymptotic property (2.2) of $l_n(\beta_0)$ becomes invalid. It is known that χ_p^2 is asymptotically normal with mean p and variance $2p$, so we may expect that

$$(l_n(\beta_0) - p)/\sqrt{2p} \xrightarrow{\mathcal{L}} N(0,1) \quad \text{as } n \to \infty. \tag{2.3}$$

One aim of this paper is just to ensure the sufficient conditions under which the standard normal calibration of the ELR is still valid. Another one is to study the practical calibration of the ELR in finite sample settings.

In this paper p is allowed to grow with the sample size n and $p = o(n)$, which guarantee that the distribution of ELR $l_n(\beta_0)$ will not degenerate into a point mass at infinity, according to Tsao (2004, Theorem 2). Applying Lagrange multiplier method we get

$$l_n(\beta_0) = 2\sum_{i=1}^{n}\log(1 + \lambda^t Z_i), \tag{2.4}$$

where λ satisfies

$$\sum_{i=1}^{n}\frac{Z_i}{1 + \lambda^t Z_i} = 0. \tag{2.5}$$

Throughout the paper, unless otherwise simplicity stated, let $\gamma_1(A) \geq \gamma_2(A) \geq \cdots \geq \gamma_p(A)$ denote the eigenvalues of a symmetric matrix A. In this paper we assume that there exists positive constants $0 < c_1 < c_2 < \infty$ such that, uniformly in n, $c_1 < \gamma_p(\Sigma_n) \leq \gamma_1(\Sigma_n) < c_2$, which leads to $c_1/\sigma^2 < \gamma_p(\Sigma_n^*) \leq \gamma_1(\Sigma_n^*) < c_2/\sigma^2$ where $\Sigma_n^* = \mathrm{E}(Z_1Z_1^t) = \sigma^2\Sigma_n$. This is a basic assumption in the analysis of EL for high-dimensional data, similar condition was considered in Hjort et al. (2009).

For convenience, we define some notation. For a random vector ξ, we define $\|\xi\|_q = (\mathrm{E}\|\xi\|^q)^{1/q}$. Let $\alpha^{i_1i_2\cdots i_k} = \mathrm{E}(U_{1i_1}U_{1i_2}\cdots U_{1i_k})$ and $A^{i_1i_2\cdots i_k} = \frac{1}{n}\sum_{j=1}^{n}U_{ji_1}U_{ji_2}\cdots U_{ji_k}$ where U_{ij} is the j-th component of $U_i \equiv \Sigma_n^{*-1/2}Z_i$. In particular, $\alpha^i = 0$, $\alpha^{ij} = \delta^{ij}$, here δ^{ij} is the kronecker delta.

We give the following regularity conditions in this section, which are similar to Conditions C1-C6 presented in Liu et al. (2012):

Condition 1. $p^{-1}\sum_{j=1}^{p}\mathrm{E}|X_{1j}|^q < K$ for some $K > 0$ and $q \geq 4$.

Condition 2. $\mathrm{E}|\epsilon_1|^q < \infty$ for the same q as in Condition 1.

Condition 3. $\|X_1\|_q \gamma_1^{3/2}(\Sigma_n) = o(n^{q/(2q+4)})$.

Condition 4. $p^{2+4/q}/n \ \to 0$.

Condition 5. $p = o(n^{2/5})$.

Condition 6. $\sum_{i,j=1}^{p} \alpha^{iijj} = O(p^2)$.

Condition 7. $\sum_{i,j,k=1}^{p} \alpha^{ijk}\alpha^{ijk} = O(p^{5/2})$ and $\sum_{i,j,k=1}^{p} \alpha^{ijj}\alpha^{ikk} = O(p^{5/2})$.

Condition 1 is also assumed by Hjort et al. (2009). Conditions 1 and 4 guarantee that the eigenvalues of S_n are close to those of Σ_n, so that $0 < c_1 < \gamma_p(S_n) \leq \gamma_1(S_n) < c_2 < \infty$ holds with probability tending to one, when n is large. The similar statement holds for S_n^*. Condition 3, together with Conditions 1, 2 and 4, implies $\sup_{1\leq i\leq n} |\lambda_*^t Z_i| = o_p(1)$, which guarantees Taylor expansions of (2.4) and (2.5). The Condition 6 means that each fourth moment of U_i, like α^{iijj}, is bounded. Obviously, Condition 6 is weak. Chen et al. (2009) claimed that $p = o(n^{1/2})$ might be the best growth rate. For this rate, Condition 7 is needed, which is a relatively strong condition compared with Condition 6.

We give the following theorem, which may be reviewed as a corollary of Theorem 1 of Liu et al. (2012). The proofs are parallel to that presented in Appendix A of Liu et al. (2012) and thus we omit that here

Theorem 1. *Under Conditions 1-4 and 6, if Condition 5 or 7 holds, then*

$$(l_n(\beta_0) - p)/\sqrt{2p} \xrightarrow{d} N(0,1) \quad as\, n \to \infty.$$

Theorem 1 indicates that, under some regularity conditions, (2.3) is valid. Particularly, when $q = 4$ in Condition 4, the asymptotic normality of $l_n(\beta_0)$ holds for $p = o(n^{1/3})$, which achieves the the same rate to that of Liu et al. (2012) and improves the growth rate $p = o(n^{1/4})$ of Chen et al. (2009), without assumptions of a multivariate model structure and strong condition that all the components of observations are uniformly bounded, which rules out even normal cases. When $q \geq 8$, the rate for asymptotic normality of $l_n(\beta_0)$ is $p = o(n^{q/(2q+4)})$, which is close to $p = o(n^{1/2})$ when q is large enough. Chen et al. (2009) pointed out that $p = o(n^{1/2})$ may be the best rate for asymptotic normality of ELR.

We have shown previously that (2.3) is valid under certain conditions and may be used to conduct tests or construct confidence regions, where the critical values can be obtained from the normal approximation (2.3). However, our simulation study show that the empirical coverage percentage based on calibration (2.3) has a large deviation from the nominal coverage level when the ratio p/n is large, referring to Table 1. This fact is mainly due to that p and $2p$ do not agree with the true mean and variance of $\mathrm{EL}_n(\beta_0)$ respectively. Similar findings have also been revealed by Chen et al. (2009) and Liu et al. (2012).

According to the arguments of Liu et al. (2012), we similarly consider two approximations of ELR in (2.1) to improve the coverage accuracy, that is $T_{1n} \equiv n\bar{Z}_n^t S_n^{*-1}\bar{Z}_n$ and $T_{2n} \equiv n\bar{Z}_n S_{1n}^{*-1}\bar{Z}_n^t$ with $S_{1n}^* = \frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z}_n)(Z_i - \bar{Z}_n)^t$, the sample variance-covariance matrix of Z_i. Owen (2001, Chapter 3) pointed out that

$$T_{1n} = T_{2n}\left(1 + \frac{T_{2n} - 1}{n}\right)^{-1}.$$

Furthermore, we give the approximation expressions of mean and variance of T_{1n} and T_{2n} below,

$$\begin{aligned}
\mathrm{E}T_{1n} &\approx E_{1n} \equiv p + \frac{1}{n}(\sum_{i,j,k=1}^{p} \alpha^{ijk}\alpha^{ijk} + \sum_{i,j,k=1}^{p} \alpha^{ijj}\alpha^{ikk}), \\
\mathrm{Var}(T_{1n}) &\approx V_{1n} \equiv 2p + \frac{1}{n}(12\sum_{i,j,k=1}^{p} \alpha^{ijj}\alpha^{ikk} + 12\sum_{i,j,k=1}^{p} \alpha^{ijk}\alpha^{ijk} - 2\sum_{i,j=1}^{p} \alpha^{iijj}), \\
\mathrm{E}(T_{2n}) &\approx E_{2n} \equiv p + \frac{1}{n}(p^2 + 2p + \sum_{i,j,k=1}^{p} \alpha^{ijk}\alpha^{ijk} + \sum_{i,j,k=1}^{p} \alpha^{ijj}\alpha^{ikk}), \\
\mathrm{Var}(T_{2n}) &\approx V_{2n} \equiv 2p + \frac{1}{n}(8p^2 + 16p + 12\sum_{i,j,k=1}^{p} \alpha^{ijj}\alpha^{ikk} + 12\sum_{i,j,k=1}^{p} \alpha^{ijk}\alpha^{ijk} - 2\sum_{i,j=1}^{p} \alpha^{iijj}).
\end{aligned}$$

We find that the approximation expressions of mean and variance of T_{1n} and T_{2n} are the same as that of Liu et al. (2012). Indeed, the calculations are similar and refer to Appendix B of Liu et al. (2012) for details.

Denote $(\hat{E}_{in}, \hat{V}_{in})$ for $i = 1, 2$ the moment estimation of (E_{in}, V_{in}), which are obtained by substituting the unknown quantities in E_{in} and V_{in} for corresponding moments. In practice applications, we can calculate the critical values according to

$$(l_n(\beta_0) - a_n)/\sqrt{b_n} \xrightarrow{\mathcal{L}} N(0, 1) \quad \text{as } n \to \infty. \tag{2.6}$$

where (a_n, b_n) may be $(p, 2p)$ or $(\hat{E}_{in}, \hat{V}_{in})$ for $i = 1, 2$. In Section 4, we undertake a simulation study to compare five calibration methods in finite sample size settings, from the viewpoint of coverage accuracy. As expected, the new calibration method proposed by Liu et al. (2012) has the best performance for linear model with random regressor.

3 EL for linear model with fixed design

Now we consider the linear regression model with fixed design

$$Y_i = {x_i}^t\beta + \epsilon_i, \quad 1 \le i \le n, \tag{3.1}$$

where β is a p-dimensional column vector of unknown parameters and x_i is a $p \times 1$ vector of the ith fixed design point, for which Y_i is the response. Here we suppose that ϵ_i's are *i.i.d.* with mean zero and finite variance σ^2. This means that we assume homoscedasticity in the present paper. The data observed in the form $\{(x_i^t, Y_i) \mid 1 \le i \le n\}$.

For the linear regression model (3.1), we know that

$$\mathrm{E}(Y_i \mid x_i) = x_i^t\beta, \quad \mathrm{E}(\epsilon_i) = 0, \quad \mathrm{Var}(\epsilon_i) = \sigma^2.$$

Similarly, we define auxiliary variables $Z_i = Z_i(\beta) = x_i(Y_i - x_i^t\beta)$ for $1 \le i \le n$ and Let β_0 be the true value of β. We define $V_n = \frac{1}{n}\sum_{i=1}^n V^{(i)} = \frac{\sigma^2}{n}\sum_{i=1}^n x_i x_i^t$ with $V^{(i)} = \mathrm{Var}(Z_i)$. It can be seen that Z_i's are independent but not identically distributed due to the fixed design point and $\mathrm{E}\{Z_i(\beta_0)\} = 0$ for $1 \le i \le n$.

The ELR function for β is similar to that in (2.1), (2.4) and (2.5) except that Z_i's are independent but not identically distributed here.

Let X be the $p \times n$ design matrix whose n columns are the x_i's. We assume that X is of full rank p ($p \leq n$). Consider the following regularity conditions:

a. There exists positive absolute constants c_3 and c_4 such that

$$0 < c_3 < \inf_n \gamma_p(V^{(n)}) \leq \sup_n \gamma_1(V^{(n)}) < c_4 < \infty\,;$$

b. both $\mathrm{E}(\epsilon_1^4)$ and $\max_{1\leq i\leq n}\|x_i\|^4$ have a finite upper bound for all n, where $\|\cdot\|$ denotes the Euclidean norm.

Notice that, by Condition a, we have $\gamma_p(\frac{1}{n}XX^t) > c_3/\sigma^2$ and $\gamma_1(\frac{1}{n}XX^t) < c_4/\sigma^2$.

For fixed p, Owen (1991) proved that, under Conditions a and b,

$$l_n(\beta_0) \overset{\mathcal{L}}{\to} \chi^2_p \quad \text{as}\, n \to \infty, \tag{3.2}$$

Furthermore, Chen (1993) showed that the $l_n(\beta_0)$ is Bartlett correctable.

However, when p grows with the sample size n and $p = o(n)$, (3.2) is invalid. We can show that under some conditions,

$$\frac{l_n(\beta_0) - p}{w_n/n} \overset{\mathcal{L}}{\to} N(0,1), \quad \text{as}\, n \to \infty, \tag{3.3}$$

where $w_n^2 = \sum_{i=1}^n \sigma_i^2$ with

$$\begin{aligned} \sigma_1^2 &= (x_1^t V_n^{-1} x_1)^2 \mathrm{E}(\epsilon_1^4) - [\mathrm{tr}(V_n^{-1}V^{(1)})]^2, \\ \sigma_i^2 &= 4\sum_{k=1}^{i-1} \mathrm{tr}(V_n^{-1}V^{(i)}V_n^{-1}V^{(k)}) + (x_i^t V_n^{-1} x_i)^2 \mathrm{E}(\epsilon_1^4) - [\mathrm{tr}(V_n^{-1}V^{(i)})]^2, \quad 1 < i \leq n. \end{aligned}$$

Similar to the arguments of Liu et al. (2012), if we can show that

$$\mathrm{EL}_n(\beta_0) - n\bar{Z}_n^t \widetilde{S}_n^{-1} \bar{Z}_n = o_p(\sqrt{p}), \tag{3.4}$$

$$n\bar{Z}_n^t (V_n^{-1} - \widetilde{S}_n^{-1})\bar{Z}_n = o_p(\sqrt{p}) \tag{3.5}$$

and

$$\frac{n\bar{Z}_n^t V_n^{-1}\bar{Z}_n - p}{w_n/n} \overset{\mathcal{L}}{\to} N(0,1), \tag{3.6}$$

here $\widetilde{S}_n = \frac{1}{n}\sum_{i=1}^n Z_i Z_i^t$ and $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i Z_i^t$, then (3.3) is valid.

For this, we define some notation. Let $\xi_i = V_n^{-1} Z_i$ and define

$$\bar{\alpha}^{j_1\cdots j_k} = n^{-1}\sum_{i=1}^n \mathrm{E}(\xi_{ij_1}\cdots\xi_{ij_k}) \quad \text{and} \quad \bar{A}^{j_1\cdots j_k} = n^{-1}\sum_{i=1}^n \mathrm{E}(\xi_{ij_1}\cdots\xi_{ij_k} - \bar{\alpha}^{j_1\cdots j_k}),$$

where ξ_{ij} is the j-th component of ξ_i. In particular, it is easy to see that $\bar{\alpha}^i = 0$, $\bar{\alpha}^{ij} = \delta^{ij}$, δ^{ij} is the kronecker delta.

In addition to the usual regularity Conditions a and b, we still need the following regularity conditions in the present section, which is similar to the Conditions 1-7 in Section 2, except that replace Conditions 1, 2, 3, 6, and 7 by:

Condition $1'$. $p^{-1}\sum_{j=1}^p |x_{ij}|^q < K_1, 1 \leq i \leq n$, for some constant $K_1 > 0$ and $q \geq 4$;

Condition $2'$. $\mathrm{E}|\epsilon_1|^{2q} < K_2$ for some constant $K_2 > 0$;

Condition $3'$. $p = o(n^{\frac{q-2}{2q}})$;

Condition $6'$. $\sum_{i,j=1}^{p} \bar{\alpha}^{iijj} = O(p^2)$;

Condition $7'$. $\sum_{i,j,k=1}^{p} \bar{\alpha}^{ijk}\bar{\alpha}^{ijk} = O(p^{5/2})$ and $\sum_{i,j,k=1}^{p} \bar{\alpha}^{ijj}\bar{\alpha}^{ikk} = O(p^{5/2})$.

We have the following Propositions and Theorems, referring to Appendix for proofs.

Proposition 1. *Assume Conditions* $1'$, $2'$ *and* $3'$, *if Condition 5 or both Conditions 4 and* $7'$ *hold, then* (3.4) *is valid.*

Proposition 2. *Suppose Conditions* $1'$, $2'$, *4 and* $6'$, *then* (3.5) *holds.*

Proposition 3. *Under model* (3.1), *suppose that* $E(\xi_{ij_1}^{\alpha_1}\cdots\xi_{ij_l}^{\alpha_l}) \le B$ *for some positive absolute constant* $B < \infty$ *and any* $1 \le i \le n$, *whenever* $\sum_{i=1}^{l} \alpha_i \le 6$. *Then, if* $p^3/n \to 0$,

$$\frac{n\bar{Z}_n^t V_n^{-1}\bar{Z}_n - p}{w_n/n} \xrightarrow{\mathcal{L}} N(0,1).$$

Summarizing the results of Propositions 1, 2 and 3, and note that $\{2(\frac{c_3}{c_4})^2 pn^2 + O(np^2)\}^{1/2} \le w_n \le \{2(\frac{c_4}{c_3})^2 pn^2 + O(np^2)\}^{1/2}$, and $\frac{o_p(\sqrt{p})}{w_n/n} = o_p(1)$. We have the following conclusion.

Theorem 2. *Assume Conditions* $1' - 3'$, *4*, $6'$ *and the same conditions in Proposition 3, then*

$$\frac{l_n(\beta_0) - p}{w_n/n} \xrightarrow{\mathcal{L}} N(0,1), \quad as\, n \to \infty.$$

Remarks. (1) Obviously, a major difference between Theorem 1 and Theorem 2 is that the asymptotic variance of ELR function $l_n(\beta_0)$ depends on the design points and the forth moment of error in fixed regressor setting. This mainly because that the auxiliary variables Z_i, $1 \le i \le n$, are independent but not identically distributed and the variance of Z_i depends on the fixed design point x_i; **(2)** We have shown that the growth rate of p for asymptotic normality of $l_n(\beta_0)$can achieve to the order of $\sqrt{n}$ in random design setting. However, we investigate from the Proposition 3 that the rate is $p = o(n^{1/3})$ in the non-random case.

We hope to give a calibration method of $l_n(\beta_0)$ in fixed design setting, such as T_{2n}, and try to calculate its approximation mean and variance. However, it is very difficult when Z_i's are independent but not identically distributed. On the other hand, it has been empirically observed that an analogue of T_{2n}, say $T_n = n\bar{Z}_n^t \widetilde{S}_{1n}^{-1}\bar{Z}_n$, with $\widetilde{S}_{1n} = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z}_n)(Z_i - \bar{Z}_n)^t$, is possibly better approximation to the ELR function $l_n(\beta_0)$ than T_{2n}. Our simulation study also indicates that calibration of $l_n(\beta_0)$ with the sample mean and variance of T_n, which are obtained by resampling technique such as bootstrap, can improve the coverage accuracy to some extent for moderate sample size, compared with the calibration method (3.5).

4 Simulation study

In this section, we carried out some simulation studies to evaluate the finite-sample performance of the proposed calibration methods of $l_n(\beta_0)$ in random and fixed design settings,

Table 1: Coverage percentages when $p = cn^{0.4}$, $c = 3, 4, 5$

n	p	MEL	OEL	SEL	STEL	BEL	$\hat{E}_n$	$\hat{E}_{1n}$	$\hat{E}_{2n}$	$\hat{V}_n$	$\hat{V}_{1n}$	$\hat{V}_{2n}$
200	25	**95.76**	83.68	81.12	88.20	88.88	29.3	26.5	29.9	79.4	60.4	87.4
	33	**95.24**	73.36	70.08	85.56	82.32	41.3	36.6	42.3	137.5	95.4	141.6
	43	**94.56**	56.20	52.56	80.08	71.60	58.9	50.0	59.6	240.2	147.5	224.9
400	33	**95.40**	88.76	87.28	89.96	92.08	35.9	34.0	36.9	85.7	70.9	94.0
	44	**95.76**	82.80	80.68	86.96	88.68	49.8	45.9	51.0	132.6	99.6	140.1
	58	**96.20**	72.92	70.56	83.76	84.40	68.7	61.9	70.6	197.8	144.1	213.7
800	42	**94.92**	91.52	90.36	91.08	93.48	44.2	42.5	44.8	95.5	84.2	102.6
	55	**96.60**	89.48	87.44	89.76	93.32	58.7	55.9	59.8	127.6	111.8	143.2
	72	**96.84**	86.80	85.32	88.92	92.08	77.9	73.8	80.4	179.3	150.6	203.9

in terms of coverage accuracy. The nominal coverage level is selected as $1 - \alpha = 0.95$. We demonstrate the advantages of the proposed calibration methods in several different growth rates and sample sizes. Each result is based on 2000 replications.

First, we conducted simulations for random design setting. Similar to Liu et al. (2012), we compared the following five calibrations in our simulation studies:

(1) the proposed method (named as modified EL, MEL), i.e. the normal calibration (2.6) with $(a_n, b_n) = (\hat{E}_{1n}, \hat{V}_{1n})$. In our simulation $(\hat{E}_{1n}, \hat{V}_{1n})$ is the simulated value (i.e. sample mean and variance) of (E_{1n}, V_{1n});
(2) the ordinary χ^2_p calibration (2.2) (OEL);
(3) the standard normal calibration (2.3) (SEL);
(4) the normal calibration (2.6) with $(a_n, b_n) = (\hat{E}_{1n}, \hat{V}_{1n})$ (STEL);
(5) the χ^2_p calibration with Bartlett correction (BEL).

The regressors X_i's were generated from multi-normal distribution with zero mean vector and covariance matrix $\Sigma = (\sigma_{ij})$, where $\sigma_{ij} = 0.5^{|i-j|}, 1 \leq i, j \leq p$. The noise ϵ_i's were drawn from the chi-squared distribution $\chi^2_{p,3}$ with three degrees of freedom. We considered the growth rate $p = cn^{0.4}$, with $c = 3, 4$ and 5 for each sample size $n = 200, 400$ and 800, in the current simulation studies.

These simulation results presented in Table 1 basically agree with those that were discovered by Liu et al. (2012) and show that the MEL calibration method still works well in the linear regression with random regressors and outperforms the other four calibration methods in term of empirical coverage probability.

For the fixed design case, we performed another simulation study to compare the coverage accuracy of two types of confidence regions: one is based on the calibrated $l_n(\beta_0)$ with the sample mean and variance of T_n (denoted as CT in the tables) obtained from 500 Bootstrap samples for each simulated data set; and the other one is based on the calibration in (3.5) (denoted as CW in the tables). In this simulation, we generated the p-dimensional fixed design point x_i's from the standard multivariate normal distribution and considered the t-distribution with 5 degrees of freedom for the noise ϵ. Two types of growth rates were considered: the slower growth rate (i) $p = 10, 20$ and 30 for each value of $n = 200, 400$ and 600; the faster growth rate (ii) $p = 15, 30$ and 45 for $n = 200, 400$ and 600 respectively. Our

Table 2: Coverage percentages when $n = 200, 400, 600$ and $p = 10, 20, 30$

p	n	CW	CT	$\hat{E}_n$	$\hat{E}_2$	$\hat{V}_n$	V_1	$\hat{V}_2$
	200	84.85	**93.80**	12.3	10.5	33.5	23.6	195.3
10	400	89.25	**91.25**	11.2	10.3	25.0	21.8	92.6
	600	89.90	**87.80**	10.8	10.1	24.2	21.2	65.6
	200	66.25	**86.75**	28.6	22.4	100.8	53.1	438.0
20	400	81.45	**84.35**	24.4	21.1	69.8	46.6	216.4
	600	87.25	**85.10**	22.7	20.5	53.3	44.4	124.6
	200	44.05	**73.75**	49.6	35.4	227.9	88.4	750.6
30	400	70.75	**79.10**	39.2	32.2	122.1	74.3	319.2
	600	78.15	**77.80**	36.4	31.6	102.6	69.6	209.2

Table 3: Coverage percentages when $n = 200, 400, 600$ and $p = 15, 30, 45$

p	n	CW	CT	$\hat{E}_n$	$\hat{E}_2$	$\hat{V}_n$	V_1	$\hat{V}_2$
	200	77.75	**91.50**	20.0	16.3	59.8	37.6	309.1
15	400	84.90	**87.40**	17.9	15.9	47.7	33.8	156.3
	600	88.65	**86.65**	16.7	15.4	39.7	32.6	98.3
	200	44.20	**74.40**	49.5	35.3	226.0	88.6	712.3
30	400	70.75	**77.05**	39.3	32.4	123.3	74.3	317.9
	600	78.15	**77.80**	36.4	31.6	102.6	69.6	209.2
	200	12.05	**48.30**	93.1	58.0	655.6	153.1	1390.6
45	400	46.90	**63.15**	65.9	50.8	234.8	121.6	507.2
	600	64.80	**67.10**	58.4	48.4	179.7	111.1	313.4

simulation results are summarized in Tables 2 and 3, also involving the simulated means and variances of $l_n(\beta_0)$ (labeled as $(\hat{E}_n, \hat{V}_n)$) and T_n (labeled as $(\hat{E}_2, \hat{V}_2)$), and the simulated value of asymptotic variance in (3.3) (labeled as V_1).

It can be observed from the Tables 2 and 3 that for moderate sample sizes, the empirical coverage probabilities based on CT are higher than that based on CW, especially for the case of $n = 200$ and $p = 10$, the coverage percentage of CT is closer to the nominal coverage level, that is more accurate. Thus the calibration method CT is a good alternative in moderate sample size settings. We can observe from Tables 2 and 3 that the CW has improving coverage accuracy along with the increasing sample size, however the coverage accuracy of CT decreases somewhat when the sample sizes becomes large, even the coverage probability is lower than that of CW for large sample size, such as $n = 600$. This is mainly because that the Bootstrap samples, obtained by resampling technique with replacement, may not involve as more as information of data when the sample size increases. This problem may become severe as n is sufficiently large. It can also be seen that the coverage probabilities of both CT and CW decrease when the dimension p increases, especially when $n = 200$ and $p = 45$ in Table 3, the performance of CW becomes rather worse and is unacceptable.

In addition, to investigate the asymptotic normality of empirical likelihood ratio, we conducted further simulation study under the same simulation schemes to that in fixed design setting through Q-Q plots. Figures 1 and 2 display the Q-Q plots of $l_n(\beta_0)$ standardized with CW and CT (described above) against the normal distribution, for the two different growth rates (i) and (ii). It can be observed that as n increases for a fixed p or p decreases for a fixed n, there is a general convergence of standardized empirical likelihood ratio to $N(0, 1)$. Similar to Chen et al. (2009), the convergence in Figure 1 for the slower growth rate (i) is faster than that in Figure 2 for the faster growth rate (ii).

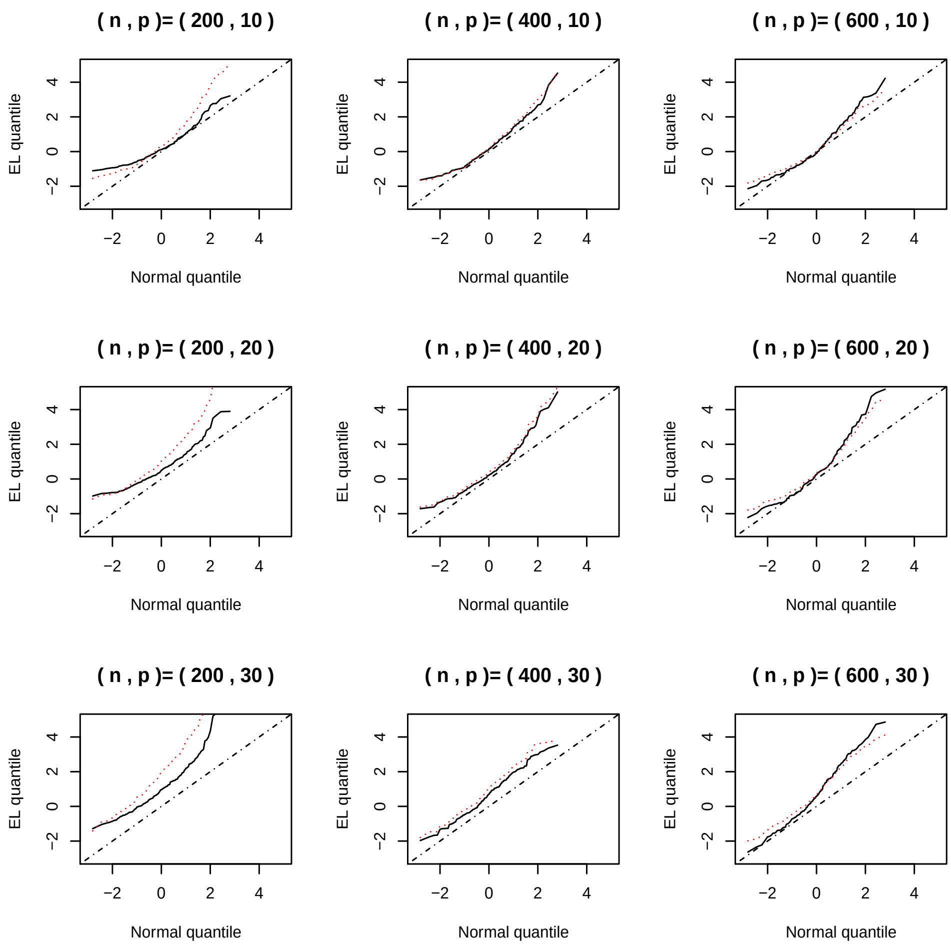

Figure 1: Normal Q-Q plots of two standardizations of $l_n(\beta_0)$ for data drawn from $t_{p,5}$: CT (solid) and CW (dotted).

5 Discussion

In this paper, we studied the asymptotic behavior and calibrations of EL for high-dimensional data in parametric regression and gave different calibration methods of EL for random regressors and fixed design settings respectively. It can be seen from the simulation results for the fixed design setting that, the coverage accuracy of the proposed calibration method based

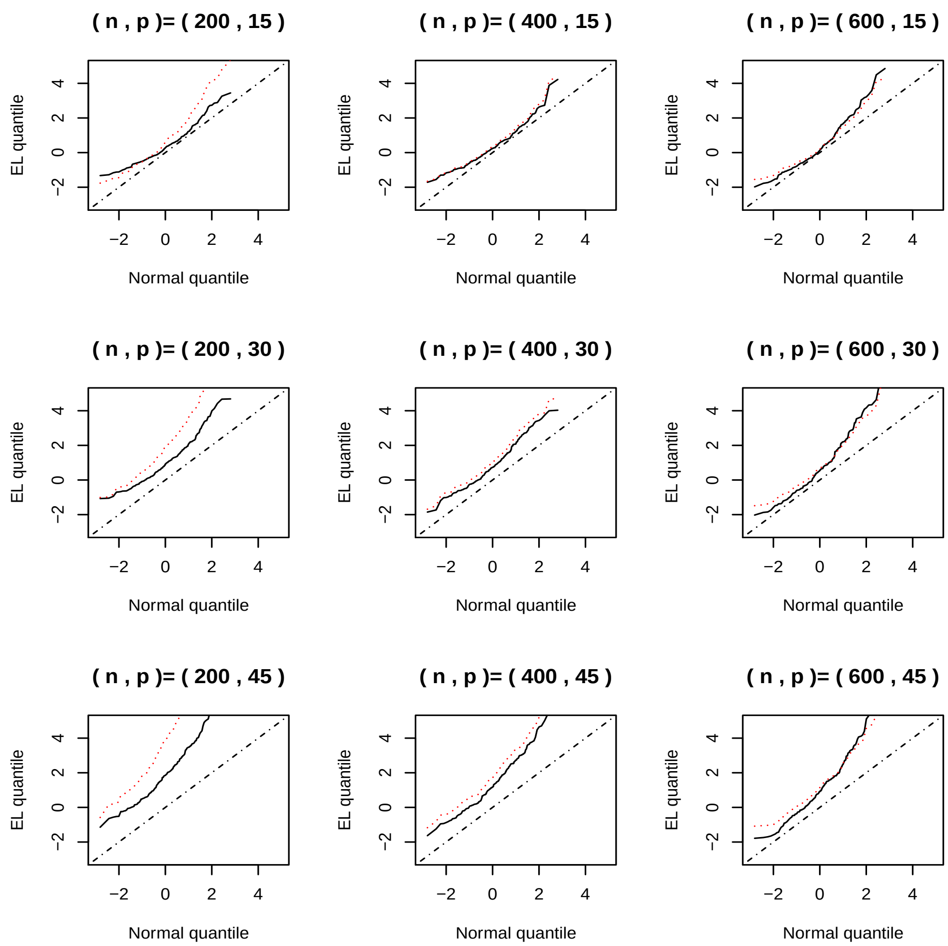

Figure 2: Normal Q-Q plots of two standardizations of $l_n(\beta_0)$ for data drawn from $t_{p,5}$: CT (solid) and CW (dotted).

on the T_n statistic becomes low for the large values of n, ever though it is higher than that based on (3.5) for moderate sample sizes. So it is still necessary to propose another calibration method with good coverage accuracy for large sample size. Besides, Chen and Van (2009) gave a review on EL in nonparametric regression, semiparametric regression, regression with missing response or covariates and regression with censored data for fixed dimensionality p. Exploring the properties, computations and asymptotic behaviors of EL in these regression models for high-dimensional data will be an interesting and challenging work. There has been some contribution to this area. Li et al. (2012) studied the EL for varying coefficient partially linear model in high-dimensional setting. They proposed a bias-corrected EL (BCEL) method for the estimation or inference about the parameters of interest in semiparametric models, which leads to the normal calibration of ELR statistics in high-dimensional setting. As shown in that paper, the standard normal calibration of ELR with $(p, 2p)$ is very rough and has large coverage error for high-dimensional data. So it is great interest and necessity to study the calibration of this corrected EL, which is a direction of our future work.

Acknowledgements

I like to thank professor Wang and Zou for questions and suggestions on this work.

Appendix: Proofs of Propositions 1, 2 and 3

Now our task is to prove the Propositions 1, 2 and 3 and we need auxiliary lemmas. Most of the lemmas here are analogues of those in Appendix A of Liu et al. (2012), but little difference. Actually, the large part of arguments here parallel that in Liu et al. (2012), also the manner of proof.

From the time being, let $L_n \equiv \max_{1\le i,j\le n} \mid \widetilde{S}_{n,i,j} - V_{n,i,j} \mid$ be the largest absolute element of $\widetilde{S}_n - V_n$ with $\widetilde{S}_n = \frac{1}{n}\sum_{i=1}^n Z_i Z_i^t$ and $V_n = \frac{1}{n}\sum_{i=1}^n \mathrm{Var}(Z_i) = \frac{\sigma^2}{n} XX^t$. Let $\mathcal{M}(A)$ denote the largest element of matrix A. For convenience, define $(A)_{ij}$ and $(a)_i$ as the (i,j) element of a matrix A and the i-th component of a vector a, respectively. To avoid confusion, we state here that the auxiliary variable Z_i appeared in the following arguments and formulas is that for fixed design case.

Lemma 1. *Assume Conditions $1'$ and $2'$. Then, for any given $\varepsilon > 0$ and some positive constant C_q only depending on q,*

$$\mathrm{P}(L_n \ge \varepsilon) < C_q K_1^2 K_2 \frac{p^2}{n^{q/2}\varepsilon^q}.$$

Proof. Let $R_n = \widetilde{S}_n - V_n \triangleq (R_{jk})_{p\times p}$ with $R_{jk} = \frac{1}{n}\sum_{i=1}^n x_{ij}x_{ik}(\epsilon_i^2 - \sigma^2)$, $j,k = 1,\cdots,p$. We have

$$\mathrm{P}(|R_{jk}| \ge \varepsilon) \le \frac{\mathrm{E}|\sqrt{n}R_{jk}|^q}{(\sqrt{n}\varepsilon)^q} = \frac{n^{-\frac{q}{2}}\mathrm{E}|\sum_{i=1}^n x_{ij}x_{ik}(\epsilon_i^2-\sigma^2)|^q}{n^{q/2}\varepsilon^q}.$$

Let $W_{ijk} = x_{ij}x_{ik}(\epsilon_i^2 - \sigma^2)$, $1 \le i \le n$ and $j,k = 1,2,\cdots,p$. Obviously, for any fixed j,k, W_{ijk}'s are independent but not identically distributed random variables with mean zero. By the Theorem 2 of S. W. Dharmadhikari et al. (1969), we have

$$\mathrm{E}|\sum_{i=1}^n W_{ijk}|^q \le C_q n^{q/2-1}\sum_{i=1}^n \mathrm{E}|W_{ijk}|^q \le C_q \mathrm{E}|\epsilon_1|^{2q} n^{q/2-1}\sum_{i=1}^n |x_{ij}|^q|x_{ik}|^q,$$

where C_q is a finite positive constant only depending on q. This leads to

$$\mathrm{P}(|R_{jk}| \ge \varepsilon) \le C_q \mathrm{E}|\epsilon_1|^{2q}\frac{\frac{1}{n}\sum_{i=1}^n |x_{ij}|^q|x_{ik}|^q}{n^{q/2}\varepsilon^q}.$$

Therefore, by Conditions $1'$ and $2'$, we obtain

$$\begin{aligned}\mathrm{P}(L_n \ge \varepsilon) &\le \sum_{j,k=1}^p \mathrm{P}(|R_{jk}| \ge \varepsilon) \le C_q \mathrm{E}|\epsilon_1|^{2q} p^2 \frac{\frac{1}{n}\sum_{i=1}^n (p^{-1}\sum_{j=1}^p |x_{ij}|^q)(p^{-1}\sum_{j=1}^p |x_{ik}|^q)}{n^{q/2}\varepsilon^q} \\ &< C_q K_1^2 K_2 \frac{p^2}{n^{q/2}\varepsilon^q}. \quad \blacksquare\end{aligned}$$

Lemma 2.

$$\max_{1\le i\le p} |\gamma_i(V_n) - \gamma_i(\widetilde{S}_n)| \le pL_n$$

Proof. The proof follows a similar line to that of Lemma 2 in Appendix A of Liu et al. (2012) and refer to that for details.■

According to Lemma 1 and Conditions $1'$ and $2'$, we have

$$\mathrm{P}(pL_n \geq \varepsilon) \leq C_q K_1^2 K_2 \frac{p^{2+q}}{n^{q/2}\varepsilon^q},$$

which, together with Condition 4, leads to $pL_n = o_p(1)$. This means that the eigenvalues of $\widetilde{S}_n$ are close to those of V_n.

Summarizing the results of Lemmas 1 and 2, we have the following Corollary:

Corollary 1. *Suppose Conditions $1'$, $2'$ and 4, then $c_3 < \gamma_p(\widetilde{S}_n) \leq \gamma_1(\widetilde{S}_n) < c_4$ holds with probability tending to one as $n \to \infty$.*

In order to prove $\sup_{1\leq i\leq n} |\lambda^t Z_i| = o_p(1)$ for the Taylor expansions of (2.4) and (2.5), we need the following lemma. Let $\rho_n \equiv \max_{1\leq i\leq n} \|Z_i\|$.

Lemma 3. *Assume Conditions $1'$ and $2'$, then*

$$\rho_n = o_p(n^{1/q}p^{1/2}).$$

Proof. It is easy to see that

$$\begin{aligned}\rho_n &= \max_{1\leq i\leq n} \big\{\|Z_i\|^{q/2} - \mathrm{E}\|Z_i\|^{q/2} + \mathrm{E}\|Z_i\|^{q/2}\big\}^{2/q} \\ &\leq \big\{\max_{1\leq i\leq n} \big|\|Z_i\|^{q/2} - \mathrm{E}\|Z_i\|^{q/2}\big| + \mathrm{E}|\epsilon_1|^{q/2} \max_{1\leq i\leq n} \|x_i\|^{q/2}\big\}^{2/q}.\end{aligned}$$

Applying C_p inequality twice, we have $\|x_i\|^{q/2} \leq p^{q/4}\big(\sum_{j=1}^p p^{-1}|x_{ij}|^q\big)^{1/2} \leq p^{q/4}K_1^{1/2}$. Thus, $\max_{1\leq i\leq n} \|x_i\|^{q/2} = O(p^{q/4})$.

On the other hand, according to Condition $2'$ and the Lemma 3 of Owen (1990), which continues to hold for independent but not identically distributed random variables, we get

$$\begin{aligned}\max_{1\leq i\leq n} \big|\|Z_i\|^{q/2} - \mathrm{E}\|Z_i\|^{q/2}\big| &\leq \max_{1\leq i\leq n} \Big\{\frac{\big|\|Z_i\|^{q/2} - \mathrm{E}\|Z_i\|^{q/2}\big|}{\big[\mathrm{Var}(\|Z_i\|^{q/2})\big]^{1/2}}\Big\} \max_{1\leq i\leq n} \big\{\mathrm{Var}(\|Z_i\|^{q/2})\big\}^{1/2} \\ &\leq \max_{1\leq i\leq n} \Big\{\frac{\big|\|Z_i\|^{q/2} - \mathrm{E}\|Z_i\|^{q/2}\big|}{\big[\mathrm{Var}(\|Z_i\|^{q/2})\big]^{1/2}}\Big\} \max_{1\leq i\leq n} \big\{\|x_i\|^{q/2}\big\}(\mathrm{E}|\epsilon_1|^q)^{1/2} \\ &= o_p(n^{1/2})O(p^{q/4}) = o_p(n^{1/2}p^{q/4}).\end{aligned}$$

The last second equality is due to Lemma 3 of Owen (1990) and Condition $2'$.

Summarizing the above, we obtain $\rho_n = \big\{o_p(n^{1/2}p^{q/4}) + O(p^{q/4})\big\}^{2/q} = o_p(n^{1/q}p^{1/2})$. ■

Lemma 4. *If Conditions $1'$, $2'$ and $3'$ hold, then*

$$\sup_{1\leq i\leq n} |\lambda^t Z_i| = o_p(1), \quad \lambda = \widetilde{S}_n^{-1}(\bar{Z}_n + \zeta_n),$$

where $\bar{Z}_n = O_p(p^{1/2}n^{-1/2})$ and $\zeta_n = o_p(n^{\frac{1-q}{q}}p^{3/2})$.

Proof. Let $\lambda = \|\lambda\|\vartheta$, where ϑ is a unit vector. Then we get from (2.5) that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\vartheta^t Z_i}{1+\|\lambda\|\vartheta^t Z_i} = 0.$$

It follows that

$$0 = \frac{1}{n}\sum_{i=1}^{n}\frac{\vartheta^t Z_i}{1+\|\lambda\|\vartheta^t Z_i} = \frac{1}{n}\sum_{i=1}^{n}\vartheta^t Z_i - \frac{\|\lambda\|\frac{1}{n}\sum_{i=1}^{n}\vartheta^t Z_i Z_i^t\vartheta}{1+\|\lambda\|\vartheta^t Z_i}.$$

Since

$$0 < 1+\|\lambda\|\vartheta^t Z_i \le 1+\|\lambda\|\max_{1\le i\le n}\|Z_i\| = 1+\|\lambda\|\rho_n,$$

hence

$$\frac{1}{n}\sum_{i=1}^{n}\vartheta^t Z_i - \frac{\|\lambda\|\frac{1}{n}\sum_{i=1}^{n}\vartheta^t Z_i Z_i^t\vartheta}{1+\|\lambda\|\vartheta^t Z_i} \le \vartheta^t\bar{Z}_n - \frac{\|\lambda\|}{1+\|\lambda\|\rho_n}\vartheta^t\widetilde{S}_n\vartheta,$$

which leads to $\vartheta^t\bar{Z}_n(1+\|\lambda\|\rho_n) \ge \vartheta^t\widetilde{S}_n\vartheta\|\lambda\|$ and then $\|\lambda\|(\vartheta^t\widetilde{S}_n\vartheta - \vartheta^t\bar{Z}_n\rho_n) \le \vartheta^t\bar{Z}_n$ follows.

Furthermore, we note that

$$\begin{aligned}
\mathrm{E}(\bar{Z}_n^t\bar{Z}_n) &= \frac{1}{n^2}\sum_{i,j=1}^{n}x_i^t x_j \mathrm{E}(\epsilon_i\epsilon_j) = \frac{\sigma^2}{n}\mathrm{tr}(\frac{1}{n}X^tX) = \frac{\sigma^2}{n}\mathrm{tr}(\frac{1}{n}XX^t)\\
&\le \frac{p}{n^2}\sum_{i=1}^{n}\gamma_1(x_i x_i^t\sigma^2) < \frac{p}{n}c_4 = O(\frac{p}{n}),
\end{aligned}$$

which implies that $\bar{Z}_n = O_p(p^{1/2}n^{-1/2})$.

Since $\|\vartheta^t\bar{Z}_n\| \le \|\bar{Z}_n\|$, hence $\vartheta^t\bar{Z}_n = O_p(p^{1/2}n^{-1/2})$ and $\vartheta^t\bar{Z}_n\rho_n = o_p(1)$. Notice that $\vartheta^t\widetilde{S}_n\vartheta \ge \gamma_p(\widetilde{S}_n) > c_3$ holds with probability tending to one as $n\to\infty$. Therefore, we have $\|\lambda\| = O_p(|\vartheta^t\bar{Z}_n|/c_3) = O_p(\|\bar{Z}_n\|) = O_p(p^{1/2}n^{-1/2})$ and $\sup_{1\le i\le n}|\lambda^t Z_i| = O_p(pn^{\frac{2-q}{2q}}) = o_p(1)$. This enables us to expand (2.5) to get $\bar{Z}_n - \widetilde{S}_n\lambda + \zeta_n = 0$ with $\zeta_n = \frac{1}{n}\sum_{i=1}^{n}Z_i(\lambda^t Z_i)^2(1+o_p(1))$. Then we derive $\lambda = \widetilde{S}_n^{-1}(\bar{Z}_n+\zeta_n)$. In addition, it is easy to see that

$$\begin{aligned}
\zeta_n &\le \frac{2}{n}\sum_{i=1}^{n}\|Z_i\||\lambda^t Z_i|^2 \le \frac{2}{n}\sum_{i=1}^{n}\rho_n\lambda^t Z_i Z_i^t\lambda \le 2\rho_n\|\lambda\|^2\gamma_1(\widetilde{S}_n) = \rho_n\|\lambda\|^2 O_p(\gamma_1(V_n))\\
&= O_p(\rho_n\|\lambda\|^2) = O_p(o_p(n^{1/q}p^{1/2})O_p(p/n)) = o_p(n^{\frac{1-q}{q}}p^{3/2}).
\end{aligned}$$

Furthermore, we get from the last second equality that

$$O_p(o_p(n^{1/q}p^{1/2})O_p(p/n)) = O_p(o_p(n^{1/q}p^{1/2})\bar{Z}_n p^{1/2}n^{-1/2}) = \bar{Z}_n o_p(1).$$

It follows that $\zeta_n = \bar{Z}_n o_p(1)$. ■

Lemma 5. *Suppose Conditions* $1'$, $2'$ *and* 4*, then*

$$(\widetilde{S}_n - V_n^{-1})\bar{Z}_n = (V_n^{-1}\bar{Z}_n)o_p(1).$$

Proof. Note firstly that $\gamma_1(\widetilde{S}_n^{-1}) < 1/c_3$ with probability tending to one as $n \to \infty$. Applying Lemma 2 we have

$$\begin{aligned}\|V_n(\widetilde{S}_n^{-1} - V_n^{-1})\bar{Z}_n\| &= \|(V_n - \widetilde{S}_n)\widetilde{S}_n^{-1}\bar{Z}_n\| = \Big[\bar{Z}_n\widetilde{S}_n^{-1}(V_n - \widetilde{S}_n)^2\widetilde{S}_n^{-1}\bar{Z}_n\Big]^{1/2} \\ &\le \|\bar{Z}_n\|\gamma_1(\widetilde{S}_n^{-1})\max_{1\le i\le n}|\gamma_i(V_n - \widetilde{S}_n)| \le \|\bar{Z}_n\|\gamma_1(\widetilde{S}_n^{-1})pL_n \\ &= \|\bar{Z}_n\|O_p(pL_n) = \|\bar{Z}_n\|o_p(1),\end{aligned}$$

which implies that $(\widetilde{S}_n^{-1} - V_n^{-1})\bar{Z}_n = (V_n^{-1}\bar{Z}_n)o_p(1)$. ■

The following result is the direct consequence of Lemmas 4 and 5.

Corollary 2. *Under Conditions* $1'$, $2'$, $3'$ *and 4, we have*

$$\lambda = V_n^{-1}\bar{Z}_n(1 + o_p(1)).$$

Proof of Proposition 1

Review that we have shown previously $\sup_{1\le i\le n}\lambda^t Z_i = o_p(1)$, hence we can expand (2.4) as follows:

$$\begin{aligned}\mathrm{EL}_n(\beta_0) &= 2\sum_{i=1}^n\left\{\lambda^t Z_i - \frac{1}{2}\lambda^t Z_i Z_i^t\lambda + \frac{1}{3}(Z_i^t\lambda)^3(1 + o_p(1))\right\} \\ &= n\{2\bar{Z}_n^t\lambda - \lambda^t\widetilde{S}_n\lambda\} + \Gamma_1,\end{aligned}$$

where $\Gamma_1 = \frac{2}{3}\sum_{i=1}^n(Z_i^t\lambda)^3(1 + o_p(1))$. By substituting λ for $\widetilde{S}_n^{-1}(\bar{Z}_n + \zeta_n)$ in the expression above, we get

$$\mathrm{EL}_n(\beta_0) = n\bar{Z}_n^t\widetilde{S}_n^{-1}\bar{Z}_n + \Gamma_1 - \Gamma_2,$$

where $\Gamma_2 = n\zeta_n^t\widetilde{S}_n^{-1}\zeta_n$. It is clear that the result of Proposition 1 is followed if we can prove that $\Gamma_1 = o_p(\sqrt{p})$ and $\Gamma_2 = o_p(\sqrt{p})$.

Indeed, according to Corollary 2 we have

$$Z_i^t\lambda = Z_i^t V_n^{-1}\bar{Z}_n(1 + o_p(1)).$$

This, together with the proof of Lemma 4, leads to $\Gamma_1 = \frac{2}{3}\sum_{i=1}^n(Z_i^t V_n^{-1}\bar{Z}_n)^3(1 + o_p(1))$ and $\zeta_n = \frac{1}{n}\sum_{i=1}^n Z_i(Z_i^t V_n^{-1}\bar{Z}_n)^2(1 + o_p(1))$. On the other hand, we note that $\Gamma_2 = n\zeta_n^t(\widetilde{S}_n^{-1} - V_n^{-1})\zeta_n + n\zeta_n^t V_n^{-1}\zeta_n$. Similar to Lemma 5, we can prove that $(\widetilde{S}_n^{-1} - V_n^{-1})\zeta_n = V_n^{-1}\zeta_n o_p(1)$. Therefore we get $\Gamma_2 = n\zeta_n^t V_n^{-1}\zeta_n(1 + o_p(1))$.

It is not difficult to see that we can rewrite Γ_1 and Γ_2 as

$$\Gamma_1 = \frac{2n}{3}\sum_{i,j,k=1}^p\bar{\alpha}^{ijk}\bar{A}^i\bar{A}^j\bar{A}^k(1 + o_p(1)),$$

$$\Gamma_2 = n\sum_{i,j,k,r,s=1}^p\bar{\alpha}^{ijk}\bar{\alpha}^{irs}\bar{A}^j\bar{A}^k\bar{A}^r\bar{A}^s(1 + o_p(1)).$$

With tedious calculation we further obtain

$$\mathrm{E}\Big(\frac{2n}{3}\sum_{i,j,k=1}^p\bar{\alpha}^{ijk}\bar{A}^i\bar{A}^j\bar{A}^k\Big) = \frac{4}{3n}\sum_{i,j,k=1}^p(3\bar{\alpha}^{ijj}\bar{\alpha}^{ikk} + 2\bar{\alpha}^{ijk}\bar{\alpha}^{ijk})(1 + o(1)),$$

$$\mathrm{E}\Big(n\sum_{i,j,k,r,s=1}^p\bar{\alpha}^{ijk}\bar{\alpha}^{irs}\bar{A}^j\bar{A}^k\bar{A}^r\bar{A}^s\Big) = \frac{1}{n}\sum_{i,j,k=1}^p(2\bar{\alpha}^{ijk}\bar{\alpha}^{ijk} + \bar{\alpha}^{ijj}\bar{\alpha}^{ikk})(1 + o(1)).$$

Recalling that $(A)_{ij}$ denotes the (i,j) element of a matrix A and $(a)_i$ the i-th component of a vector a. Since

$$\mathrm{E}\big|\xi_{rk}\big|^2 = \mathrm{E}\big|(V_n^{-1/2}Z_r)_k\big|^2 = \mathrm{E}\big[\big(V_n^{-1/2}x_rx_r^tV_n^{-1/2}\epsilon_r^2\big)_{kk}\big] = \big(V_n^{-1/2}x_rx_r^tV_n^{-1/2}\big)_{kk}\sigma^2,$$

Hence, Using C_p inequality, Cauchy-Schwartz inequality, Young inequality and Lemma 4 of Liu et al. (2012), we can verify that

$$|\sum_{i,j,k=1}^{p}\bar{\alpha}^{ijk}\bar{\alpha}^{ijk}| \le \frac{c_4}{c_3}p\sum_{i,j=1}^{p}\bar{\alpha}^{iijj} \quad \text{and} \quad \big|\sum_{i,j,k=1}^{p}\bar{\alpha}^{ijj}\bar{\alpha}^{ikk}\big| \le \frac{c_4}{c_3}p\sum_{i,j=1}^{p}\bar{\alpha}^{iijj}.$$

Summarizing the arguments above, together with Condition $6^{'}$, we get

$$\mathrm{E}\big(\frac{2n}{3}\sum_{i,j,k=1}^{p}\bar{\alpha}^{ijk}\bar{A}^i\bar{A}^j\bar{A}^k\big)^2 \le \frac{c_4}{c_3}\frac{20p}{3n}\sum_{i,j}^{p}\bar{\alpha}^{iijj}(1+o(1)) = O(p^3/n),$$

and

$$\mathrm{E}\big(n\sum_{i,j,k,r,s=1}^{p}\bar{\alpha}^{ijk}\bar{\alpha}^{irs}\bar{A}^j\bar{A}^k\bar{A}^r\bar{A}^s\big) \le \frac{3p}{n}\frac{c_4}{c_3}\bar{\alpha}^{iijj}(1+o(1)) = O(p^3/n),$$

which means that $\Gamma_1 = O_p(p^{3/2}n^{-1/2}) = o_p(\sqrt{p})$ and $\Gamma_2 = O_p(p^3/n)$ by Markov's inequality. Further, $p = o(n^{2/5})$ is sufficient for $\Gamma_2 = o_p(\sqrt{p})$, and under the Condition $7^{'}$, the order of p can achieve $p = o(n^{1/2})$. ■

The following Lemma is needed for the proof of Proposition 2.

Lemma 6. *Let $G_n = I_p - N_n$ with $N_n = V_n^{-1/2}\widetilde{S}_nV_n^{-1/2}$. Suppose Condition $6^{'}$, then*

$$tr(G_n^2) = O_p(p^2/n).$$

Proof. Note that

$$G_n^2 = \mathrm{I}_p - \frac{2}{n}\sum_{i=1}^{n}\xi_i\xi_i^t + \frac{1}{n^2}\sum_{i,j=1}^{n}\xi_i\xi_i^t\xi_j\xi_j^t.$$

Thus,

$$\begin{aligned}
\mathrm{E}\big[\mathrm{tr}(G_n^2)\big] &= \mathrm{E}\big[p - \frac{2}{n}\sum_{r=1}^{p}\sum_{i=1}^{n}\xi_{ir}^2 + \frac{1}{n^2}\sum_{i,j=1}^{n}\sum_{r,s=1}^{p}\xi_{ir}\xi_{is}\xi_{jr}\xi_{js}\big] \\
&= p - 2\sum_{r=1}^{p}\bar{\alpha}^{rr} + \frac{1}{n^2}\sum_{r,s=1}^{p}\big(\sum_{i=j}+\sum_{i\neq j}\big)\mathrm{E}(\xi_{ir}\xi_{is}\xi_{jr}\xi_{js}) \\
&\le -p + \frac{1}{n}\sum_{r,s=1}^{p}\bar{\alpha}^{rrss} + \frac{1}{n^2}\sum_{r,s=1}^{p}\sum_{i,j=1}^{n}\mathrm{E}(\xi_{ir}\xi_{is})\mathrm{E}(\xi_{jr}\xi_{js}) \\
&= \frac{1}{n}\sum_{r,s=1}^{p}\bar{\alpha}^{rrss} = O(p^2/n).
\end{aligned}$$

This implies that $\mathrm{tr}(G_n^2) = O_p(p^2/n)$. ■

Proof of Proposition 2

Define $e_n = n\bar{Z}_n^t(V_n^{-1} - \widetilde{S}_n^{-1})\bar{Z}_n$. It is easy to see that

$$e_n = n\bar{\xi}_n^t\bar{\xi}_n - n\bar{\xi}_n^t V_n^{1/2}\widetilde{S}_n^{-1}V_n^{1/2}\bar{\xi}_n = n\bar{\xi}_n^t(\mathrm{I}_p - N_n^{-1})\bar{\xi}_n.$$

Note that $\mathrm{I}_p - N_n^{-1} = -G_n - G_n^2 - \cdots - G_n^k + G_n^k(\mathrm{I}_p - N_n^{-1})$. Hence,

$$e_n = n[-\bar{\xi}_n^t G_n\bar{\xi}_n - \bar{\xi}_n^t G_n^2\bar{\xi}_n - \cdots - \bar{\xi}_n^t G_n^k\bar{\xi}_n] + n\bar{\xi}_n^t G_n^k(\mathrm{I}_p - N_n^{-1})\bar{\xi}_n. \quad \text{(A1)}$$

Now we consider the convergence of the right-hand side of (A1). First, we focus on the first term. By Lemmas 4 and 5, we have

$$\|\bar{\xi}_n\|^2 \le \frac{1}{c_3}\|\bar{Z}_n\|^2 = O_p(p/n),$$

and $|\gamma_i(G_n)| \le [tr(G_n^2)]^{1/2} = O_p(p/\sqrt{n})$. Therefore,

$$|\bar{\xi}_n^t G_n^k\bar{\xi}_n| \le \|\bar{\xi}_n\|^2 \max_{1\le i\le p}|\gamma_i(G_n^k)| \le \|\bar{\xi}_n\|^2[\mathrm{tr}(G_n^2)]^{k/2} = O_p(p^{k+1}/n^{k/2+1}),$$

which implies that the series $n\sum_{k=1}^{\infty}(-\bar{\xi}_n^t G_n^k\bar{\xi}_n)$ is convergent, as long as n is fixed and $p = o(n^{1/2})$. On the other hand, with lengthy algebra we have

$$\mathrm{E}|n\bar{\xi}_n^t G_n\bar{\xi}_n|^2 = n^2\mathrm{E}|\bar{Z}_n^t(V_n - \widetilde{S}_n)\bar{Z}_n|^2 = O(p^3/n),$$

which means that $n\bar{\xi}_n^t G_n\bar{\xi}_n = O_p(p^{3/2}/n^{1/2}) = o_p(\sqrt{p})$. Therefore we conclude from above that $n\sum_{k=1}^{\infty}(-\bar{\xi}_n^t G_n^k\bar{\xi}_n) = o_p(\sqrt{p})$.

For the second term in (A1), we will next show in the following arguments that the remainder term $n\bar{\xi}_n^t G_n^k(\mathrm{I}_p - N_n^{-1})\bar{\xi}_n$ is negligible as $k \to \infty$. For this, we firstly note that

$$\left|n\bar{\xi}_n^t G_n^k(\mathrm{I}_p - N_n^{-1})\bar{\xi}_n\right| \le \left|n\bar{\xi}_n^t G_n^k\bar{\xi}_n\right| + \left|n\bar{\xi}_n^t G_n^k N_n^{-1}\bar{\xi}_n\right|$$

and

$$\left|n\bar{\xi}_n^t G_n^k\bar{\xi}_n\right| = nO_p(p^{k+1}/n^{k/2+1}) = O_p(p^{k+1}/n^{k/2}).$$

Furthermore, applying the Lemma 4 of Liu et al. (2012), which remains valid for the case of any $n \times n$ symmetric matrix $A = (a_{ij})$, that is , $\mathcal{M}(A) \le \max_{1\le i\le n}|\gamma_i(A)|$, we have

$$\begin{aligned}
\left|n\bar{\xi}_n^t G_n^k N_n^{-1}\bar{\xi}_n\right| &\le n\|\bar{\xi}_n\|^2 \max_{1\le i\le n}\left\{|\gamma_i(G_n^k N_n^{-1})|\right\} \le np\|\bar{\xi}_n\|^2\mathcal{M}(G_n^k N_n^{-1}) \\
&\le np^2\|\bar{\xi}_n\|^2 \max_{1\le i\le p}|\gamma_i(G_n^k)|\gamma_1(N_n^{-1}) \le np^2\|\bar{\xi}_n\|^2\left[\mathrm{tr}(G_n^2)\right]^{k/2}\gamma_1(V_n)\gamma_1(\widetilde{S}_n) \\
&= O_p(p^{k+3}/n^{k+2}) = o_p(1).
\end{aligned}$$

Under Condition 4, the last equality holds as $k \to \infty$. Therefore, the consequence of Proposition 2 follows. ■

We give the following Lemma for the proof of Proposition 3.

Lemma 7. *Under model* (3.1)*, suppose that* $E(\xi_{ij_1}^{\alpha_1}\xi_{ij_2}^{\alpha_2}\cdots\xi_{ij_l}^{\alpha_l}) \le B$ *for some finite positive constant* $B < \infty$ *and any* $1 \le i \le n$*, whenever* $\sum_{i=1}^{l}\alpha_i \le 6$*, then for some finite positive constant* c*,*

$$E\big(\xi_n^t\sum_{i=1}^{n-1}\xi_i\big)^6 \le cp^6(n^3 + n^2 + n).$$

Proof. We state that, here and in the sequel, c stands for constant which may be different from line to line and even from formula to formula and whose value is not of interest.

Indeed, a direct calculation gives that

$$\begin{aligned} \mathrm{E}\big(\xi_n^t \sum_{i=1}^{n-1} \xi_i\big)^6 &= \mathrm{E}\big[(\xi_n^t \sum_{i_1=1}^{n-1} \xi_{i_1}) \cdots (\xi_n^t \sum_{i_6=1}^{n-1} \xi_{i_6})\big] = \sum_{i_1,\cdots,i_6=1}^{n-1} \sum_{j_1,\cdots,j_6=1}^{p} \mathrm{E}(\xi_{nj_1} \cdots \xi_{nj_6}) \mathrm{E}(\xi_{i_1 j_1} \cdots \xi_{i_6 j_6}) \\ &= B\big[5(n-1)(n-2)p^6 B^2 + 10(n-1)(n-2)p^2 B^2 + 45(n-1)(n-2)(n-3)p^6 B^2\big] \\ &\le cp^6(n^3 + n^2 + n). \quad \blacksquare \end{aligned}$$

Proof of Proposition 3

Let $J_n = \sum_{i=1}^n \xi_i$, it is clear that

$$\frac{n\bar{Z}_n^t V_n^{-1} \bar{Z}_n - p}{w_n/n} = \frac{n\|\bar{\xi}_n\|^2 - p}{w_n/n} = \frac{\|J_n\|^2 - np}{w_n}$$

To prove the Proposition 3, the strategy is to construct a martingale and then to apply the martingale central limit theorem based on it. Hence, it is technically convenient to define $M_n = \|J_n\|^2 - np$, $n \ge 1$, $M_0 = 0$, and $H_n = M_n - M_{n-1}, n \ge 1$. It is easy to see that

$$M_n = \|J_{n-1}\|^2 - (n-1)p + 2\xi_n^t J_{n-1} + \|\xi_n\|^2 - p,$$

and then

$$H_n = 2\xi_n^t J_{n-1} + \|\xi_n\|^2 - p, \quad \mathrm{E}H_n = \mathrm{E}\|\xi_n\|^2 - p.$$

In addition, denote by $\mathcal{F}_i = \sigma(\xi_1, \cdots, \xi_i) = \sigma(J_1, \cdots, J_i)$, for $i = 1, 2, \cdots, n$, the σ-fileds generated by $\xi_1, \cdots, \xi_i$. It can be seen that $\{M_n, F_n, n \ge 1\}$ is not a martingale, although it is true in the situation of Portnoy, S. (1988). In order to construct a martingale based on M_n for martingale limit theorem, we define some other notation.

Let $\phi_n = H_n - \mathrm{E}H_n$, $n \ge 1$, and $W_n = \sum_{i=1}^n \phi_i = M_n - \sum_{i=1}^n \mathrm{E}H_i$. It is easy to show that, for each $s < t$,

$$\mathrm{E}[W_n|\mathcal{F}_s] = \|J_s\|^2 - s\,p = W_s,$$

which indicates that $\{W_n, \mathcal{F}_n, n \ge 1\}$ is a martingale.

Recalling the definitions of σ_i's and w_n, it is easy to see that $\sigma_i^2 = \mathrm{E}\phi_i^2$ and $w_n^2 = \sum_{i=1}^n \sigma_i^2$. To apply the martingale central limit theorem of Chow and Teicher (1997, Theorem 1, P_{336}), it suffices to show that if $p^3/n \to 0$, then both

$$\sum_{i=1}^n \mathrm{E}|\phi_i|^3 / w_n^3 \to 0 \tag{A2}$$

and

$$\sum_{i=1}^n \mathrm{E}\big|\mathrm{E}(\phi_i^2|\mathcal{F}_{i-1}) - \sigma_i^2\big| / w_n^2 \to 0 \tag{A3}$$

hold.

Hence, in the following arguments, our task is to prove (A2) and (A3). For this, we calculate firstly σ_i's, for $i \ge 1$. It is easy to see that, for $i \ge 1$,

$$\sigma_i^2 = 4\mathrm{E}[J_{i-1}^t \xi_i \xi_i^t J_{i-1}] + \mathrm{E}(\|\xi_i\|^2 - p)^2 - (\mathrm{E}\|\xi_i\|^2 - p)^2.$$

For the first term of the expression above, we can verify that

$$\begin{aligned}\mathrm{E}[J_{i-1}^t\xi_i\xi_i^tJ_{i-1}] &= \mathrm{E}\big[J_{i-1}^t\mathrm{E}(\xi_i\xi_i^t)J_{i-1}\big] = \sum_{k=1}^{i-1}\mathrm{tr}(V_n^{-1}V^{(i)}V_n^{-1}V^{(k)})\\ &\geq \sum_{k=1}^{i-1}p\frac{\gamma_p(V^{(i)})\gamma_p(V^{(k)})}{\gamma_1^2(V_n)} \geq p(i-1)(\frac{c_3}{c_4})^2.\end{aligned}$$

Recall that $\mathrm{E}(\xi_{ij_1}^{\alpha_1}\xi_{ij_2}^{\alpha_2}\cdots\xi_{ij_l}^{\alpha_l}) \leq B$ for any $1 \leq i \leq n$ and all $j_1, j_2, \cdots, j_l = 1, \cdots, p$, whenever $\sum_{i=1}^k \alpha_i \leq 6$. Therefore we have, for $k \leq 3$,

$$\mathrm{E}\|\xi_i\|^{2k} = \mathrm{E}\big[(\sum_{j=1}^p \xi_{ij}^2)^k\big] \leq p^{k-1}\sum_{j=1}^p \mathrm{E}\xi_{ij}^{2k} \leq p^k B.$$

Then, it follows that $\mathrm{E}(\|\xi_i\|^2 - p)^2 = O(p^2)$, and $(\mathrm{E}\|\xi_i\|^2 - p)^2 = O(p^2)$. These lead to $\sigma_i^2 \geq 4(\frac{c_3}{c_4})^2p(i-1) + O(p^2)$ and then

$$w_n^2 = 2(\frac{c_3}{c_4})^2p\,n(n-1) + O(np^2) \geq 2(\frac{c_3}{c_4})^2n^2p(1 + O(p/n)). \tag{A4}$$

In order to prove (A2), we note that the martingale difference sequence $\phi_n = 2\xi_n^tJ_{n-1} + \|\xi_n\|^2 - \mathrm{E}\|\xi_n\|^2$. Hence, we have

$$\mathrm{E}|\phi_i|^3 \leq 32\big[\mathrm{E}(\xi_i^tJ_{i-1})^6\big]^{1/2} + 4Bp^3 \leq c\big[(\mathrm{E}|\xi_i^tJ_{i-1}|^6)^{1/2} + p^3\big].$$

From Lemma 7, we know that $\mathrm{E}(\xi_i^tJ_{i-1})^6 \leq cp^6(i^3 + i^2 + i)$. Hence we obtain that

$$\sum_{i=1}^n \mathrm{E}|\phi_i|^3 \leq c\sum_{i=1}^n p^3(i^{3/2} + i + \sqrt{i} + 1) \leq cp^3(n^{5/2} + n^2 + n^{3/2} + n).$$

Due to (A4), if $\frac{p^3}{n} \to 0$, then we derive

$$\frac{\sum_{i=1}^n \mathrm{E}|\phi_i|^3}{w_n^3} \leq cp^{3/2}(1/\sqrt{n} + 1/n + 1/n^{3/2} + 1/n^2) \to 0$$

and then (A2) follows.

We now focus on the proof of (A3). It is easy to see that

$$\mathrm{E}[\phi_i^2|\mathcal{F}_{i-1}] = 4J_{i-1}^t\mathrm{E}(\xi_i\xi_i^t)J_{i-1} + 4J_{i-1}^t\mathrm{E}\big[\xi_i(\|\xi_i\|^2 - \mathrm{E}\|\xi_i\|^2)\big] + O(p^2).$$

Therefore we have

$$\begin{aligned}\big|\mathrm{E}[\phi_i^2|\mathcal{F}_{i-1}] - \sigma_i^2\big|^2 \leq\ & 12\big\{\big|J_{i-1}^t\mathrm{E}(\xi_i\xi_i^t)J_{i-1} - \mathrm{E}(J_{i-1}^t\xi_i\xi_i^tJ_{i-1})\big|^2\\ &+\mathrm{E}\big[\xi_i^t(\|\xi_i\|^2 - \mathrm{E}\|\xi_i\|^2)\big]J_{i-1}J_{i-1}^t\mathrm{E}\big[\xi_i(\|\xi_i\|^2 - \mathrm{E}\|\xi_i\|^2)\big] + O(p^4)\big\}\end{aligned}$$

and then

$$\begin{aligned}\big[\mathrm{E}\big|\mathrm{E}(\phi_i^2|\mathcal{F}_{i-1}) - \sigma_i^2\big|^2\big]^{1/2} \leq \sqrt{12}\big\{&\mathrm{E}\big[J_{i-1}^t\mathrm{E}(\xi_i\xi_i^t)J_{i-1} - \mathrm{E}(J_{i-1}^t\xi_i\xi_i^tJ_{i-1})\big]^2\\ &+\mathrm{E}\big[\xi_i^t(\|\xi_i\|^2 - \mathrm{E}\|\xi_i\|^2)\big]\mathrm{E}(J_{i-1}J_{i-1}^t)\mathrm{E}\big[\xi_i(\|\xi_i\|^2 - \mathrm{E}\|\xi_i\|^2)\big] + O(p^4)\big\}^{1/2}.\end{aligned} \tag{A5}$$

For the first term in (A5), with tedious calculation we obtain that

$$
\begin{aligned}
&\mathrm{E}\big|J_{i-1}^t\mathrm{E}(\xi_i\xi_i^t)J_{i-1}-\mathrm{E}(J_{i-1}^t\xi_i\xi_i^tJ_{i-1})\big|^2\\
&=\sum_{k,l,s,t=1}^{i-1}\mathrm{E}[(\xi_k^tV_n^{-1/2}V^{(i)}V_n^{-1/2}\xi_l)(\xi_s^tV_n^{-1/2}V^{(i)}V_n^{-1/2}\xi_t)]+\big[\sum_{k=1}^{i-1}\mathrm{tr}(V_n^{-1}V^{(k)}V_n^{-1}V^{(i)})\big]^2\\
&\qquad -2\big[\sum_{k=1}^{i-1}\mathrm{tr}(V_n^{-1}V^{(k)}V_n^{-1}V^{(i)})\big]\mathrm{E}\big[\sum_{k,l=1}^{i-1}\xi_k^tV_n^{-1/2}V^{(i)}V_n^{-1/2}\xi_l\big] \qquad \text{(A6)}
\end{aligned}
$$

Recalling that $\xi_i, 1\le i\le n$ are independent with mean zero and noting that $\xi_i^tA\xi_j=\xi_j^tA\xi_i$ if A is a symmetric matrix, we can further simplify the terms of (A6) as follows. That is

$$
\begin{aligned}
&\sum_{k,l,s,t=1}^{i-1}\mathrm{E}\big[(\xi_k^tV_n^{-1/2}V^{(i)}V_n^{-1/2}\xi_l)(\xi_s^tV_n^{-1/2}V^{(i)}V_n^{-1/2}\xi_t)\big]\\
&=\sum_{k\neq l=1}^{i-1}\mathrm{E}\big[\xi_k^tV_n^{-1/2}V^{(i)}V_n^{-1/2}\xi_k\big]\mathrm{E}\big[\xi_l^tV_n^{-1/2}V^{(i)}V_n^{-1/2}\xi_l\big]\\
&\qquad +2\sum_{k\neq l=1}^{i-1}\mathrm{E}\big[\xi_k^tV_n^{-1/2}V^{(i)}V_n^{-1/2}\xi_k\xi_k^tV_n^{-1/2}V^{(i)}V_n^{-1/2}\xi_l\big]\\
&\qquad +\sum_{j=1}^{i-1}\mathrm{E}\big[\xi_j^tV_n^{-1/2}V^{(i)}V_n^{-1/2}\xi_j\big]^2\\
&\le\big[\sum_{k=1}^{i-1}\mathrm{tr}(V_n^{-1/2}V^{(k)}V_n^{-1/2}V_n^{-1/2}V^{(i)}V_n^{-1/2})\big]^2\\
&\qquad +2\sum_{k\neq l=1}^{i-1}\mathrm{E}\big[\xi_l^t(V_n^{-1/2}V^{(i)}V_n^{-1/2})\mathrm{E}(\xi_k\xi_k^t)(V_n^{-1/2}V^{(i)}V_n^{-1/2})\xi_l\big]\\
&\qquad +\sum_{j=1}^{i-1}\sum_{k,l,s,t=1}^{p}\big(V_n^{-1/2}V^{(i)}V_n^{-1/2}\big)_{kl}\big(V_n^{-1/2}V^{(i)}V_n^{-1/2}\big)_{st}\mathrm{E}(\xi_{jk}\xi_{jl}\xi_{js}\xi_{jt})\\
&\le\big[\sum_{k=1}^{i-1}\mathrm{tr}(V_n^{-1}V^{(k)}V_n^{-1}V^{(i)})\big]^2+2(i-1)(i-2)p\gamma_1^4(V_n^{-1})\gamma_1(V^{(l)})\gamma_1(V^{(k)})\gamma_1(V^{(i)})\\
&\qquad +(i-1)p^4\gamma_1^4(V_n^{-1/2})\gamma_1^2(V^{(i)})B\\
&\le\big[\sum_{k=1}^{i-1}\mathrm{tr}(V_n^{-1}V^{(k)}V_n^{-1}V^{(i)})\big]^2+2i^2p(\frac{c_4}{c_3})^4+ip^4B(\frac{c_4}{c_3})^2 \qquad \text{(A7)}
\end{aligned}
$$

and

$$
\big[\sum_{k=1}^{i-1}tr(V_n^{-1}V^{(k)}V_n^{-1}V^{(i)})\big]\big[\sum_{k,l=1}^{i-1}\mathrm{E}(\xi_k^tV_n^{-1/2}V^{(i)}V_n^{-1/2}\xi_l)\big]=\big[\sum_{k=1}^{i-1}tr(V_n^{-1}V^{(k)}V_n^{-1}V^{(i)})\big]. \qquad \text{(A8)}
$$

Finally, by (A6), (A7) and (A8), we derive that

$$
\mathrm{E}[J_{i-1}^t\mathrm{E}(\xi_i\xi_i^t)J_{i-1}-\mathrm{E}(J_{i-1}^t\xi_i\xi_i^tJ_{i-1})]^2\le 2i^2p(\frac{c_4}{c_3})^4+ip^4B(\frac{c_4}{c_3})^2 \qquad \text{(A9)}
$$

For the second term in (A5), it is clear that

$$\begin{aligned}
&\mathrm{E}\big[\xi_i^t(\|\xi_i\|^2-\mathrm{E}\|\xi_i\|^2)\big]\mathrm{E}(J_{i-1}J_{i-1}^t)\mathrm{E}\big[\xi_i(\|\xi_i\|^2-\mathrm{E}\|\xi_i\|^2)\big]\\
&=\mathrm{E}(\xi_i^t\|\xi_i\|^2)\big[\sum_{j,k=1}^{i-1}\mathrm{E}(\xi_k\xi_k^t)\big]\mathrm{E}(\xi_i\|\xi_i\|^2)\\
&=\sum_{k,l=1}^{p}\sum_{s,t=1}^{p}\big(\sum_{j=1}^{i-1}V_n^{-1/2}V^{(j)}V_n^{-1/2}\big)_{kl}\mathrm{E}(\xi_{ik}\xi_{is}^2)\mathrm{E}(\xi_{il}\xi_{it}^2)\\
&\le\sum_{k,l=1}^{p}\sum_{s,t=1}^{p}\mathcal{M}\big(\sum_{j=1}^{i-1}V_n^{-1/2}V^{(j)}V_n^{-1/2}\big)\mathrm{E}(\xi_{ik}\xi_{is}^2)\mathrm{E}(\xi_{il}\xi_{it}^2)\\
&\le p^4\gamma_1\big(\sum_{j=1}^{i-1}V_n^{-1/2}V^{(j)}V_n^{-1/2}\big)B^2\\
&\le p^4\gamma_1^2(V_n^{-1/2})\big(\sum_{j=1}^{i-1}\gamma_1(V^{(j)})\big)B^2<ip^4B\frac{c_3}{c_4} \qquad \text{(A10)}
\end{aligned}$$

Summarizing (A5), (A9), and (A10), we get

$$\big[\mathrm{E}\big|\mathrm{E}(\phi_i^2|\mathcal{F}_{i-1})-\sigma_i^2\big|^2\big]^{1/2}\le c\big\{i^2p+ip^4+O(p^4)\big\}^{1/2},$$

and further we obtain that

$$\begin{aligned}
&\sum_{i=1}^{n}\mathrm{E}\big|\mathrm{E}(\phi_i^2|\mathcal{F}_{i-1})-\sigma_i^2\big|/w_n^2\le\sum_{i=1}^{n}\big[\mathrm{E}\big|\mathrm{E}(\phi_i^2|\mathcal{F}_{i-1})-\sigma_i^2\big|^2\big]^{1/2}/w_n^2\\
&\le c\big(n^2p^{1/2}+n^{3/2}p^2+O(np^2)\big)/(n^2p)=o(1)
\end{aligned}$$

The last equality holds if $p^3/n\to 0$. Therefore, (A3) follows.

Applying the martingale central limit theorem of Chow and Teicher (1999), we get

$$\frac{W_n}{w_n}\xrightarrow{\mathcal{L}}N(0,1),\quad \text{as}\quad n\to\infty.$$

In addition, note that $\frac{n\bar{Z}_n^tV_n^{-1}\bar{Z}_n-p}{w_n/n}=\frac{W_n+\sum_{i=1}^n\mathrm{E}H_i}{w_n/n}$ and $\sum_{i=1}^n\mathrm{E}H_i=0$ whenever n is the sample size. Therefore, the consequence of Proposition 3 follows. ■

References

[1] Chen, S. X. (1993). On the accuracy of empirical likelihood confidence regions for linear regression model. *Ann. Inst. Statist. Math* **45**, 621-637.

[2] Chen, S. X. and Cui, H. J. (2006). On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika* **93**, 215-220.

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

[4] Chen, S. X. and Van Keilegom, I. (2009). A review on empirical likelihood methods for regression. *Test* **18**, 415-447.

[5] Chen, S. X., Zhang, L. -X. and Zhong, P. -L. (2010). Tests for high-dimensional covariance matrices. *J. Am. Statist. Assio.* **105**, 810-819.

[6] Chow, Y. S. and H. Teicher (1997). *Probability Theory: Independence, Interchangeability, Martingales.* Springer-Verlag, New York.

[7] Dharmadhikari, S. W. and Jogdeo, K. (1969). Bounds on moments of certain random variables. *Ann. Math. Statist.* **40**, 1506-1509.

[8] DiCiccio, T. J., Hall, P. and Romano, J. P. (1991). Empirical likelihood is Bartlett-correctible. *Ann. Statist.* **19**, 1053-1061.

[9] Hjort, N. L., Mckeague, I. W. and Van Keilegom I. (2009). Extending the scop of empirical likelihood. *Ann. Statist.* **37**, 1079-1111.

[10] Li, G. R., Lin, L. and Zhu, L. X. (2012). Empirical likelihood for a varying coefficients partially linear model with diverging number of parameters. *J. Mult. Anal.* **105**, 85-111.

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

[12] Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90-120.

[13] Owen, A. B. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19**, 1725-1747.

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

[15] Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential familyies when the numbers of parameters tends to infinity. *Ann. Statist.* **16**, 356-366.

[16] Tang, C. Y. and Leng, C. (2010). Penalized high-dimensional empirical likelihood. *Biometrika* **97**, 905-920.

[17] Tsao, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *Ann. Statist.* **32**, 1215-1221.

[18] Wang, H., Li, B. and Leng, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *J. R. Statist. Sco. B* **71**, 671-683.

[19] Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing component hypotheses. *Ann. Math. Statist.* **9**, 60-62.

[20] Liu, Y., Zou, C., and Wang, Z. (2012). Calibration of the empirical likelihood for high-dimensional data. *Ann. Inst. Statist. Math.* To appear.