Journal of Multivariate Analysis 105 (2012) 85–111

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters

Gaorong Li [a], Lu Lin [b], Lixing Zhu [c,d,*]

[a] College of Applied Sciences, Beijing University of Technology, Beijing 100124, China
[b] School of Mathematics, Shandong University, Jinan 250100, China
[c] Department of Mathematics, Hong Kong Baptist University, Hong Kong
[d] School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming 650221, China

ARTICLE INFO

Article history:
Received 6 November 2010
Available online 25 August 2011

AMS subject classifications:
primary 62G08
secondary 62G15
62G20

Keywords:
Varying coefficient partially linear model
Empirical likelihood
Bias correction
Asymptotic normality
Curse of dimensionality

ABSTRACT

The purpose of this paper is two-fold. First, for the estimation or inference about the parameters of interest in semiparametric models, the commonly used plug-in estimation for infinite-dimensional nuisance parameter creates non-negligible bias, and the least favorable curve or under-smoothing is popularly employed for bias reduction in the literature. To avoid such strong structure assumptions on the models and inconvenience of estimation implementation, for the diverging number of parameters in a varying coefficient partially linear model, we adopt a bias-corrected empirical likelihood (BCEL) in this paper. This method results in the distribution of the empirical likelihood ratio to be asymptotically tractable. It can then be directly applied to construct confidence region for the parameters of interest. Second, different from all existing methods that impose strong conditions to ensure consistency of estimation when diverging the number of the parameters goes to infinity as the sample size goes to infinity, we provide techniques to show that, other than the usual regularity conditions, the consistency holds under moment conditions alone on the covariates and error with a diverging rate being even faster than those in the literature. A simulation study is carried out to assess the performance of the proposed method and to compare it with the profile least squares method. A real dataset is analyzed for illustration.

1. Introduction

In many important statistical applications, the number of variables or parameters depends on the sample size and grows to infinity as sample size tends to infinity. Also, after variable selection in many, say "large p, small n", problems, the number of remaining variables still diverges with increasing sample size. On the other hand, semiparametric modeling is frequently employed to balance between modeling bias and "curse of dimensionality". The relevant references are, among others, [3,12,8,11,13,14,20,22,30]. In this paper, we consider a varying coefficient partially linear model (VCPLM) that is a special case of the model investigated by Lam and Fan [20].

Suppose that Y is a response variable and $(U, \mathbf{X}^T, \mathbf{Z}_n^T)$ is the associated covariates, where "T" is the transpose operator. The VCPLM takes the form:

$$Y = \mathbf{X}^T \alpha(U) + \mathbf{Z}_n^T \boldsymbol{\beta}_n + \varepsilon, \tag{1.1}$$

* Corresponding author at: Department of Mathematics, Hong Kong Baptist University, Hong Kong.
E-mail address: lzhu@hkbu.edu.hk (L. Zhu).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2011.08.010

where $\alpha(\cdot) = (\alpha_1(\cdot), \ldots, \alpha_q(\cdot))^T$ is a q-dimensional vector of unknown regression functions, $\boldsymbol{\beta}_n = (\beta_1, \ldots, \beta_{p_n})^T$ is a p_n-dimensional vector of unknown regression coefficients and ε is an independent random error with $E(\varepsilon|\mathbf{X}, \mathbf{Z}_n, U) = 0$ almost surely. Here the subscript is used to make it explicitly that both the covariates and parameters may change with n. Compared with the partially linear regression, model (1.1) permits the interaction between the covariates U and $\mathbf{X}$ in such a way that a different level of covariate U is associated with a different linear model. This allows statisticians to examine the extent to which the effects of covariate $\mathbf{X}$ vary over different levels of the variable U. Therefore, it is flexible because the VCPLM retains the flexibility of the nonparametric regression model and has the explanatory power of the linear regression model. The VCPLM is of course an extension of the partially linear model and the varying coefficient model [17].

When p_n does not increase with n, model (1.1) has been considered by Li et al. [21]. They used kernel smoother for estimation and applied this model to analyze China's nonmetal mineral manufacturing industry data. They further showed that the semiparametric varying coefficient model is more appropriate than either a parametric linear model or a semiparametric partially linear model for studying the production efficiency in China's nonmetal mineral manufacturing industry. More relevant work on the VCPLM can be found in [36,1]. Recently, Fan and Huang [10] have proposed a profile-kernel inference and established the asymptotic normality of the profile least-squares estimator for the VCPLM. You and Zhou [41] studied model (1.1) using the empirical likelihood method when p_n is fixed. When the number p_n of the parameters $\boldsymbol{\beta}_n$ grows with the sample size, the generalized varying coefficient partially linear model (GVCPLM) was considered by Lam and Fan [20]. They established the existence of the profile likelihood estimator and studied the asymptotic normality of the arbitrary linear combination $A_n\hat{\boldsymbol{\beta}}_n$ of the estimator $\hat{\boldsymbol{\beta}}_n$, where A_n is an $l \times p_n$ matrix for a fixed l. Hence in their paper the dimension l of $A_n\hat{\boldsymbol{\beta}}_n$ is fixed, and the quasi-likelihood condition in which the variance is a given function of mean, and the least favorable curve (see [33]) are required to eliminate the effect arising from plug-in estimation of nonparametric components. In this situation, their results can be used to construct confidence region for finite dimensional vector $A_n\boldsymbol{\beta}_n$ by normal approximation.

It is worth noticing that the least favorable curve plays a very important role to eliminate non-negligible bias that is created by plug-in estimation for infinite-dimensional nuisance parameter in the model, otherwise, the asymptotic behavior is difficult to investigate because the non-negligible bias makes the limiting distribution untractable. In the literature, other than the least favorable curve, undersmoothing or higher order kernel in local smoother are also often adopted in plug-in estimation to reduce the bias. However, the implementation of the estimation becomes inconvenient. In the present paper, we will propose a bias reduction approach combining with the empirical likelihood (EL) proposed first by Owen [27,28] to deal with this issue. But the bias reduction approach is generally applicable for any likelihood based on method.

It is well known that the EL for confidence region construction of $\boldsymbol{\beta}_n$ does not require a full specification of distribution from which data are drawn, but only an unbiased estimating function. Many other advantages of the EL over the normal approximation-based method have been shown in the literature. In particular, it does not impose prior constraints on the shape of the region, does not require the construction of a pivotal quantity and the region is range-preserving and transformation respecting (see [16]). The EL has been further developed by many statisticians such as [31,5,7,41,42,37–39,23,25], among others. Some comprehensive treatments can be found in [29,40].

Also the EL has been employed to investigate models with growing dimension p_n of parameter of interest. Hjort et al. [18] showed that when there is no nuisance parameter the EL ratio for β_n is asymptotically normal when $p_n = o(n^{1/3})$, a possibly fastest diverging rate under some conditions. Chen et al. [6] studied the asymptotic properties of the EL ratio and improved the growth rate of p_n. However, to achieve a faster rate, the required conditions are very strong, typically, it is assumed that all the components of observations are uniformly bounded. This rules out even normal cases. More challengingly, when there are infinite-dimensional nuisance parameters to be estimated nonparametrically, the plug-in estimation, if we use it, is of nonparametric nature with slower convergence rate than $\sqrt{n}$ and then asymptotic behavior becomes very different from those in the cases without these parameters. Furthermore, the condition of bounded support does not make sense in the case under study. We can see this in the next section. As such, there are no references in the literature for our problem when we consider mild conditions on the covariates and errors.

Therefore, the purpose of this paper is two-fold as follows.

1. *Bias reduction*. We first consider adopting a method to reduce the bias in terms of defining new residual based scores so that the limiting distribution of the EL ratio is either Chi-square when p_n is fixed or normal when p_n diverges. Our strategy is to introduce a conditional orthogonality approach. For any given u, by obtaining a conditional projection of the variable Z_n of interest onto the space spanned by X and using the centered Z_n being a weight in the estimating function, the residuals and the centered Z_n, both having mean zero, can be asymptotically conditionally orthogonal. This orthogonality will then help to get faster convergence rate of the bias to zero, and then to achieve a standard normal distribution for the EL ratio when p_n tends to infinity; for details see Section 3.1. The idea is from Zhu and Xue [42] who proposed a correction to reduce bias for a partial linear single-index model. It is worth mentioning that for the model under study, although our bias correction is a constructive method, it is similar to that with the profile empirical likelihood [41], which can correct the bias automatically. As the bias correction of Zhu and Xue [42] can be applied to other semiparametric models, our study described in this paper may also be useful for other models with large p_n scenarios which are worthy of further investigation.
2. *Weaker conditions on the involved variables and faster diverging rate of* p_n. Other than the usual regularity conditions on the smoothness and nonparametric smoothing for the involved nonparametric functions, we only need moment

conditions on the covariates and errors rather than the strong conditions required when other existing methods are applied, say bounded support of the variables involved; for details, see the remarks in Section 3.3. Unlike existing methods on relevant models having infinite dimensional nuisance parameter (e.g., [33,32,20]), it does not make sense in the case under study and actually we will not assume any thing else on the distribution of the involved variables, the quasi-likelihood framework, the higher order kernel and under-smoothing in reducing the bias, and the least favorable curve, which also try to reduce the bias.

For the problem with diverging number p_n of parameters, the obtained rate in this paper is $p_n^{3+2/(k-2)}/n \to 0$, where k is the order of moments of involved variables in the model. It is worth mentioning that the rate $p_n^{3+2/(k-2)}/n \to 0$ is likely to be the best rate for p_n in the context of the empirical likelihood when infinite dimensional nuisance parameters need to be estimated, and weaker conditions are required. Then the rate $p_n = o(n^{1/3})$ is still possible to reach. More details can be found in the remarks after Theorem 2.

Furthermore, it is of interest to know whether our bias correction can be applied to the moment-based estimation method such as profile least squares to improve estimation efficiency when plug-in estimation is needed for infinite dimensional nuisance parameters. From the proofs in Appendix, it seems possible. That is, the bias correction could be a general approach for many semiparametric settings. However, for confidence region construction, moment-based method involves one more plug-in estimation for the limiting variance of the estimating equation or the estimator of the parameter vector, which is often of complex structure. On the contrary, the EL-based method can avoid this plug-in. This is an advantage over the moment-based methods in the literature. On the other hand, it should also be mentioned that the EL method has its own limitation with heavier computational burden than moment-based estimation methods have.

The paper is organized as follows. In Section 2, the VCPLM with diverging number of parameters is considered. It is proved that the classical empirical likelihood does not possess the nonparametric version of the Wilks' theorem when plug-in estimators of the nonparametric coefficient functions are used and the number of parameters goes to infinity as sample size goes to infinity. Then the bias-corrected empirical log-likelihood ratio for $\boldsymbol{\beta}_n$ is proposed and the asymptotic normality of the BCEL ratio is provided in Section 3. In Section 4, simulation studies are carried out to assess the performance of the proposed method and to compare it with the profile least squares method. A real data example is used for illustration in Section 5. The technical proofs of the main results and some lemmas are relegated to Appendix.

2. Empirical likelihood

Let $\{(Y_{ni}; \mathbf{X}_i^T, \mathbf{Z}_{ni}^T, U_i), 1 \le i \le n\}$ be an independent identically distributed (i.i.d.) random sample which comes from the model in (1.1) with $\boldsymbol{\beta}_n$ and $\mathbf{Z}_n$ having the dimension $p_n \to \infty$ as $n \to \infty$.

To define the EL ratio, the auxiliary random vectors are introduced as

$$\eta_{ni}^0(\boldsymbol{\beta}_n) = \mathbf{Z}_{ni}(Y_{ni} - \mathbf{X}_i^T \alpha(U_i) - \mathbf{Z}_{ni}^T \boldsymbol{\beta}_n). \tag{2.1}$$

Note that $\{\eta_{ni}^0(\boldsymbol{\beta}_n), 1 \le i \le n\}$ are independent and $E[\eta_{ni}^0(\boldsymbol{\beta}_n)] = 0$ when $\alpha(u)$ and $\boldsymbol{\beta}_n$ are respectively the true coefficient function and parameter vector. Given $\alpha(u)$, the classical empirical log-likelihood ratio function for $\boldsymbol{\beta}_n$, defined as

$$\mathcal{R}(\boldsymbol{\beta}_n) = -2 \max \left\{ \sum_{i=1}^n \log(n\omega_i) | \omega_i \ge 0, \sum_{i=1}^n \omega_i = 1, \sum_{i=1}^n \omega_i \eta_{ni}^0(\boldsymbol{\beta}_n) = 0 \right\}. \tag{2.2}$$

According to Owen [29] and Qin and Lawless [31], if $\alpha(u)$ is given and $p_n = p$ is fixed, under some regularity conditions, $\mathcal{R}(\boldsymbol{\beta}_n)$ is asymptotically χ^2 with p degrees of freedom, which is a nonparametric version of Wilks' theorem. Recently, Hjort et al. [18] and Chen et al. [6] have extended the standard empirical likelihood method to general multivariate models in which the dimension p_n depends on the sample size and grows to infinity as the sample size tends to infinity. They showed that the EL ratio can be approximated well enough by a normal variable $N(p_n, 2p_n)$. For model (1.1), when $\alpha(u)$ is given and p_n grows with the sample size n, it is also shown that $\mathcal{R}(\boldsymbol{\beta}_n)$ for high dimensional parameter $\boldsymbol{\beta}_n$ can be approximated by the normal variable $N(p_n, 2p_n)$.

When $\alpha(u)$ is unknown, a plug-in nonparametric estimation is needed. However, it makes the plug-in empirical likelihood no longer asymptotically chi-square/normal. It makes the construction of the confidence region difficult. Usually, we need the Monte Carlo method to help. Clearly, it causes computational burden and accumulative errors for estimation. Hence, a bias correction is needed. To motivate our method, we give some analyses in the following. We first estimate the coefficient function by local polynomial smoother; see [9]. Rewrite model (1.1) as

$$Y_{ni} - \mathbf{Z}_{ni}^T \boldsymbol{\beta}_n = \mathbf{X}_i^T \alpha(U_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.3}$$

where $\alpha(u) = (\alpha_1(u), \ldots, \alpha_q(u))^T$. We linearly approximate each $\alpha_j(v)$ for v in a neighborhood of u by

$$\alpha_j(v) \approx \alpha_j(u) + \alpha_j'(u)(v - u) \equiv a_j + b_j(v - u), \quad j = 1, \ldots, q.$$

Denote $\mathbf{a} = (a_1, \ldots, a_q)^T$ and $\mathbf{b} = (b_1, \ldots, b_q)^T$. For any fixed $\boldsymbol{\beta}_n$, a local linear fit is defined as the following solution of the weighted least squares problems: finding $\mathbf{a}$ and $\mathbf{b}$ to minimize

$$\sum_{i=1}^{n} \{Y_{ni} - \mathbf{X}_i^T(\mathbf{a} + \mathbf{b}(U_i - u)) - \mathbf{Z}_{ni}^T\boldsymbol{\beta}_n\}^2 K_h(U_i - u), \tag{2.4}$$

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function and h is the bandwidth. Let $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ be the solutions to the minimization of (2.4). Then

$$[\hat{\mathbf{a}}^T, h\hat{\mathbf{b}}^T]^T = (\mathbf{D}_u^T\mathbf{W}_u\mathbf{D}_u)^{-1}\mathbf{D}_u^T\mathbf{W}_u(Y_n - \mathbf{Z}_n^*\boldsymbol{\beta}_n), \tag{2.5}$$

where

$$\mathbf{D}_u = \begin{pmatrix} \mathbf{X}_1^T & \dfrac{U_1 - u}{h}\mathbf{X}_1^T \\ \vdots & \vdots \\ \mathbf{X}_n^T & \dfrac{U_n - u}{h}\mathbf{X}_n^T \end{pmatrix}, \qquad \mathbf{Z}_n^* = (\mathbf{Z}_{n1}, \ldots, \mathbf{Z}_{nn})^T = \begin{pmatrix} Z_{n11} & \cdots & Z_{n1p_n} \\ \vdots & \ddots & \vdots \\ Z_{nn1} & \cdots & Z_{nnp_n} \end{pmatrix},$$
$$Y_n = (Y_{n1}, \ldots, Y_{nn})^T, \qquad \mathbf{W}_u = \mathrm{diag}(K_h(U_1 - u), \ldots, K_h(U_n - u)).$$

The solutions $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ depend on $\boldsymbol{\beta}_n$ implicitly. Then we can estimate $\alpha(u)$, when $\boldsymbol{\beta}_n$ is given, by

$$\hat{\alpha}(u, \boldsymbol{\beta}_n) = (I_q, \mathbf{0}_q)(\mathbf{D}_u^T\mathbf{W}_u\mathbf{D}_u)^{-1}\mathbf{D}_u^T\mathbf{W}_u(Y_n - \mathbf{Z}_n^*\boldsymbol{\beta}_n), \tag{2.6}$$

where I_q denotes a q-dimensional identity matrix, and $\mathbf{0}_q$ is the $q \times q$ matrix with all the entries being zero. As is shown in [4,26] and Lemma B.2 of the present paper in Appendix, the estimator $\hat{\alpha}(u, \boldsymbol{\beta}_n)$ has the following properties:

$$E(\hat{\alpha}(u, \boldsymbol{\beta}_n)) - \alpha(u) = O(h^2), \qquad \|\hat{\alpha}(u, \boldsymbol{\beta}_n) - \alpha(u)\| = O_P(c_n) \tag{2.7}$$

hold uniformly in the support of u, where $c_n = \left\{\frac{\log n}{nh}\right\}^{1/2} + h^2$. We plug the estimator $\hat{\alpha}(u, \boldsymbol{\beta}_n)$ into $\eta_{ni}^0(\boldsymbol{\beta}_n)$ of (2.1) and then get a plug-in estimating auxiliary random vectors as

$$\hat{\eta}_{ni}^0(\boldsymbol{\beta}_n) = \mathbf{Z}_{ni}(Y_{ni} - \mathbf{X}_i^T\hat{\alpha}(U_i, \boldsymbol{\beta}_n) - \mathbf{Z}_{ni}^T\boldsymbol{\beta}_n). \tag{2.8}$$

However, from the proof in Appendix, we can see that this plug-in results in a nonparametric bias because the convergence rate of the plug-in estimator $\hat{\alpha}(u, \boldsymbol{\beta}_n)$ is slower than $n^{-1/2}$. A simple calculation also yields

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\eta}_{ni}^0(\boldsymbol{\beta}_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_{ni}^0(\boldsymbol{\beta}_n) + \Delta_n, \tag{2.9}$$

where $\Delta_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{Z}_{ni}\mathbf{X}_i^T(\alpha(U_i) - \hat{\alpha}(U_i, \boldsymbol{\beta}_n))$. Similar to the proof of Lemma B.3 in Appendix, we can verify that, when an optimal bandwidth $h = O(n^{-1/5})$ is adopted,

$$\|\Delta_n\| = O_P(n^{1/2}p_n^{1/2}c_n) = O_P((p_n \log n)^{1/2}n^{1/10}). \tag{2.10}$$

Obviously, such a remainder is non-negligible for the asymptotic behavior of the EL even when the undersmoothing is used.

Thus, when p_n is fixed, the resulting plug-in empirical log-likelihood ratio function is asymptotically a weighted sum of independent standard chi-squared variables, each having one degree of freedom and an unknown weight. The more detailed discussions can be found in [37,38] about the partially linear single index model and single index model. When p_n grows with the sample size, the asymptotic properties of the plug-in empirical log-likelihood ratio will become much more complicated than the case with fixed p_n, and a bias correction to remove the effect from the bias term Δ_n becomes crucial for using the EL.

3. Bias-correction empirical likelihood

In this section, we construct the BCEL ratio and establish its asymptotic properties.

3.1. BCEL

We first adjust $\eta_{ni}^0(\boldsymbol{\beta}_n)$ in (2.1) and then the estimated auxiliary random vectors $\hat{\eta}_{ni}^0(\boldsymbol{\beta}_n)$ defined in (2.8) such that the effect of the bias Δ_n can be asymptotically eliminated. Define

$$\mu(u) = (E(\mathbf{X}\mathbf{X}^T|U = u))^{-1}E(\mathbf{X}\mathbf{Z}_n^T|U = u). \tag{3.1}$$

Note that for any given U, $\mu^T(U)\mathbf{X}$ is the projection of $\mathbf{Z}_n$ onto the space spanned by $\mathbf{X}$. Thus, $E[(\mathbf{Z}_n - \mu^T(U)\mathbf{X})\mathbf{X}^T|U] = 0$, and $\mathbf{Z}_n - \mu^T(U)\mathbf{X}$ is orthogonal to $\mathbf{X}^T$. This orthogonality will play a key role for the asymptotic normality of the bias-corrected empirical log-likelihood ratio when the following bias-corrected auxiliary random vectors are considered

$$\eta_{ni}(\boldsymbol{\beta}_n) = (\mathbf{Z}_{ni} - \mu^T(U_i)\mathbf{X}_i)(Y_{ni} - \mathbf{X}_i^T\alpha(U_i) - \mathbf{Z}_{ni}^T\boldsymbol{\beta}_n).$$

Their estimators are

$$\hat{\eta}_{ni}(\boldsymbol{\beta}_n) = (\mathbf{Z}_{ni} - \hat{\mu}^T(U_i)\mathbf{X}_i)(Y_{ni} - \mathbf{X}_i^T\hat{\alpha}(U_i, \boldsymbol{\beta}_n) - \mathbf{Z}_{ni}^T\boldsymbol{\beta}_n), \tag{3.2}$$

where $\hat{\mu}(u) = (\hat{E}(\mathbf{X}_i\mathbf{X}_i^T|U_i = u))^{-1}\hat{E}(\mathbf{X}_i\mathbf{Z}_{ni}^T|U_i = u)$ is the estimator of $\mu(u)$. $E(\mathbf{X}_i\mathbf{X}_i^T|U_i = u)$ and $E(\mathbf{X}_i\mathbf{Z}_{ni}^T|U_i = u)$ can be estimated easily by using the kernel smoothing method, respectively. For the convenience, we can also define the estimator of $\mathbf{X}_i^T\mu(U_i)$ directly as follows

$$\mathbf{X}_i^T\hat{\mu}(U_i) = \sum_{k=1}^{n} S_{ik}\mathbf{Z}_{nk}, \tag{3.3}$$

where S_{ik} is the (i, k)th element of the smoothing matrix $\mathbf{S}$, which depends only on the observations $\{(U_i, \mathbf{X}_i), i = 1, \ldots, n\}$, with

$$\mathbf{S} = \begin{pmatrix} (\mathbf{X}_1^T\mathbf{0}^T)(\mathbf{D}_{u_1}^T\mathbf{W}_{u_1}\mathbf{D}_{u_1})^{-1}\mathbf{D}_{u_1}^T\mathbf{W}_{u_1} \\ \vdots \\ (\mathbf{X}_n^T\mathbf{0}^T)(\mathbf{D}_{u_n}^T\mathbf{W}_{u_n}\mathbf{D}_{u_n})^{-1}\mathbf{D}_{u_n}^T\mathbf{W}_{u_n} \end{pmatrix}.$$

By Lemma B.1 in Appendix, one can show that $\hat{\mu}(u)$ is a consistent estimator of $\mu(u)$. From (3.2) and (3.3), the bias-corrected auxiliary random vectors can also defined by

$$\begin{aligned}\hat{\eta}_{ni}(\boldsymbol{\beta}_n) &= (\mathbf{Z}_{ni} - \hat{\mu}^T(U_i)\mathbf{X}_i)(Y_{ni} - \mathbf{X}_i^T\hat{\alpha}(U_i, \boldsymbol{\beta}_n) - \mathbf{Z}_{ni}^T\boldsymbol{\beta}_n) \\ &= \hat{\mathbf{Z}}_{ni}(\hat{Y}_{ni} - \boldsymbol{\beta}_n^T\hat{\mathbf{Z}}_{ni}),\end{aligned} \tag{3.4}$$

where $\hat{\mathbf{Z}}_{ni} = \mathbf{Z}_{ni} - \sum_{k=1}^{n} S_{ik}\mathbf{Z}_{nk}$, $\hat{Y}_{ni} = Y_{ni} - \sum_{k=1}^{n} S_{ik}Y_{nk}$. We now describe why the above procedure can reduce the bias and remove the effect of the remainder. By Lemma B.3 in Appendix, under some regularity conditions, we can verify that the remainder of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\eta}_{ni}(\boldsymbol{\beta}_n) - \eta_{ni}(\boldsymbol{\beta}_n))$ satisfies

$$\left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_i \right\| = o_P(p_n^{1/2}c_n^2\sqrt{n}) = o_P(1) \tag{3.5}$$

provided that the bandwidth h is of a convergence rate from $n^{-1/6}$ to $n^{-1/3}\log n$. Comparing the above result with (2.10), the bias-corrected auxiliary random vectors $\hat{\eta}_{ni}(\boldsymbol{\beta}_n)$ has, asymptotically, a smaller bias than $\hat{\eta}_{ni}^0(\boldsymbol{\beta}_n)$ does. Therefore, a bias-corrected empirical log-likelihood ratio is defined as

$$\hat{\mathcal{R}}(\boldsymbol{\beta}_n) = -2\max\left\{\sum_{i=1}^{n}\log(n\omega_i)|\omega_i \geq 0,\ \sum_{i=1}^{n}\omega_i = 1, \sum_{i=1}^{n}\omega_i\hat{\eta}_{ni}(\boldsymbol{\beta}_n) = 0\right\}. \tag{3.6}$$

For a given $\boldsymbol{\beta}_n$, a unique value for $\hat{\mathcal{R}}(\boldsymbol{\beta}_n)$ exists, provided that 0 is inside the convex hull of the point $(\hat{\eta}_{n1}(\boldsymbol{\beta}_n), \ldots, \hat{\eta}_{nn}(\boldsymbol{\beta}_n))$ [27,28]. By the Lagrange multiplier, $\hat{\mathcal{R}}(\boldsymbol{\beta}_n)$ can be represented as

$$\hat{\mathcal{R}}(\boldsymbol{\beta}_n) = 2\sum_{i=1}^{n}\log(1 + \lambda^T\hat{\eta}_{ni}(\boldsymbol{\beta}_n)), \tag{3.7}$$

where $\lambda \in R^{p_n}$ is the root of

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\eta}_{ni}(\boldsymbol{\beta}_n)}{1 + \lambda^T\hat{\eta}_{ni}(\boldsymbol{\beta}_n)} = 0. \tag{3.8}$$

We can obtain the *maximum bias-corrected empirical likelihood estimator* (MBCELE) $\widetilde{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}_n$ by minimizing the bias-corrected empirical log-likelihood ratio $\hat{\mathcal{R}}(\boldsymbol{\beta}_n)$. In Section 3.3, we will report that if $\boldsymbol{\beta}_n$ is the true parameter vector, the BCEL ratio $\hat{\mathcal{R}}(\boldsymbol{\beta}_n)$ is asymptotically $N(p_n, 2p_n)$ distributed.

It is interesting that when p_n is fixed, the proposed bias-corrected empirical likelihood method has the same formula as the profile empirical likelihood [41] for fixed p case, which can correct the bias automatically. However, the method in this paper is a constructive approach to make a conditional orthogonality [42] which can be used for diverse types of models. More specifically, for any given u, by obtaining a conditional projection of the variable Z_n of interest onto the space spanned by X and using the centered Z_n being a weight in the estimating function, the residuals and the centered Z_n, both having mean zero, can be asymptotically conditionally orthogonal. This orthogonality then helps to get faster convergence rate of the bias to zero, and then to achieve a standard normal distribution for the EL ratio when p_n tends to infinity.

3.2. Asymptotic properties of the BCEL ratio

In order to study the properties of $\hat{\mathcal{R}}(\boldsymbol{\beta}_n)$, we need to study the convergence rate of λ first. When p_n is fixed, $\|\lambda\| = O_P(n^{-1/2})$ has been the prevailing order for the Lagrange multiplier of the empirical likelihood except in the nonparametric curve estimation when n should be replaced by the "effective sample size" [5,39].

Theorem 1. *Under regularity conditions* (C1)–(C8) *in Appendix, if* $\boldsymbol{\beta}_n$ *is the true value of the parameter vector and* $p_n^{2+4/(k-2)}/n \to 0$ *as* $n \to \infty$, *we have*

$$\|\lambda\| = O_P(\sqrt{p_n/n}).$$

Remark 1. Theorem 1 implies that the magnitude of the Lagrange multiplier λ is dependent on the dimension p_n of the parametric component and the convergence rate of $\|\lambda\|$ will slow down with growing p_n. It can be regarded as a generalization of Owen's result on λ for a fixed p_n with $O_P(\sqrt{p_n/n}) = O_P(\sqrt{1/n})$.

Next we present the asymptotic behavior of the BCEL ratio.

Theorem 2. *In addition to regularity conditions* (C1)–(C9) *in Appendix, assume that either* $E(\varepsilon^3|U, \mathbf{X}, \mathbf{Z}_n) = 0$ *almost surely or* $k \geq 8$. *If* $\boldsymbol{\beta}_n$ *is the true value of the parameter vector,* $p_n^{3+2/(k-2)}/n \to 0$ *as* $n \to \infty$, *we have*

$$(2p_n)^{-1/2}(\hat{\mathcal{R}}(\boldsymbol{\beta}_n) - p_n) \xrightarrow{d} N(0, 1) \quad \text{as } n \to \infty, \tag{3.9}$$

where "$\xrightarrow{d}$*" denotes the convergence in distribution.*

Remark 2. Theorem 2 indicates that the dimension p_n can be increased more rapidly if we assume the existence of higher order moment for the covariates and error. By Theorem 2 and its proof, it is easy to see that, if all the components of $\mathbf{Z}_{ni}$ and $\mathbf{X}_i$ remain uniformly bounded or all the order moments exist, the rate $p_n = o(n^{1/3})$ is achievable.

Remark 3. Chen et al. [6] evaluated effects of data dimension on the asymptotic normality of the empirical likelihood ratio for high dimensional data under a general multivariate model and weak conditions. Although they further showed that $p_n = o(\sqrt{n})$ is achievable when any order of moment exists and there is not any nuisance parameter, they required that the components of the covariable can be expressed by independent variables through a linear transformation. Algebraically, it is almost to say that a rotation of the covariable can make this independent. However, there are few distributions being of this property and of course normal distribution is one of them. Thus, such a condition of independence is very restrictive.

Remark 4. Comparing the results with those of Lam and Fan [20], it is found that we significantly improve the diverging rate to infinity for the covariable $\mathbf{Z}_n$ under weaker conditions. They obtained the asymptotic normality of the arbitrary linear combination of the profile kernel estimator if $p_n = o(n^{1/5})$, but in their paper the dimension of the linear combination is fixed, and the quasi-likelihood condition and the least favorable curve are required to eliminate the effect arising from plug-in estimation of nonparametric components. In the simulations, we use the normal approximation that is based on the estimator of $A_n\boldsymbol{\beta}_n$ obtained by Lam and Fan [20] to construct the confidence regions of $A_n\boldsymbol{\beta}_n$. The comparison shows the advantage of the BCEL over the normal approximation although the new method is more computationally demanded.

As a conclusion of Theorem 2, $\hat{\mathcal{R}}(\boldsymbol{\beta}_n)$ can be used to construct a confidence region for $\boldsymbol{\beta}_n$. Let

$$\hat{I}_\alpha(\boldsymbol{\beta}_n) = \{\boldsymbol{\beta}_n : \hat{\mathcal{R}}(\boldsymbol{\beta}_n) \leq p_n + U_\alpha(2p_n)^{1/2}\}, \tag{3.10}$$

where U_α is the upper quantile of the standard normal distribution. Although $\hat{I}_\alpha(\boldsymbol{\beta}_n)$ gives a confidence region for $\boldsymbol{\beta}_n$ with asymptotically correct coverage probability $1 - \alpha$, the confidence regions of $\boldsymbol{\beta}_n$ are of little practical interest and the finite sample behavior could be particular unpredictable when p_n is very large. In practice, we are often confronted with the need of constructing confidence intervals for a particular regression coefficient or a certain linear combination of $\boldsymbol{\beta}_n$. For this, the profile empirical likelihood method can be used to construct confidence region for a linear combination $\theta = A_n\boldsymbol{\beta}_n$, where $A_n = (A_1, A_{n2})$ is an $l \times p_n$ matrix for a fixed l (independent of the sample size n), A_1 is a $l \times l$ matrix and A_{n2} is a $l \times (p_n - l)$ matrix. We always assume that the user-specified A_1^{-1} exists. For example, θ is the first l coordinates of $\boldsymbol{\beta}_n$ if we let $A_1 = I_l$ be the $l \times l$ identity matrix and $A_{n2} = \mathbf{0}$ be the $l \times (p_n - l)$ zeros matrix.

Suppose that $\gamma_n = (\theta^T, \boldsymbol{\beta}_{n(l)}^T)^T$, where $\boldsymbol{\beta}_{n(l)}$ denotes the column subvector of the last $p_n - l$ elements of $\boldsymbol{\beta}_n$. According to the partition of $\boldsymbol{\beta}_n$, write $\mathbf{Z}_{ni} = (\mathbf{Z}_{i1}^T, \mathbf{Z}_{ni2}^T)^T$, where $\mathbf{Z}_{i1}$ and $\mathbf{Z}_{ni2}$ are $l \times 1$ and $(p_n - l) \times 1$ subvectors, respectively. Let $\widetilde{\mathbf{Z}}_{ni} = (\widetilde{\mathbf{Z}}_{i1}^T, \widetilde{\mathbf{Z}}_{ni2}^T)^T = (\mathbf{Z}_{i1}^T A_1^{-1}, \mathbf{Z}_{ni2}^T - \mathbf{Z}_{i1}^T A_1^{-1} A_{n2})^T$. Then model (1.1) is reduced to the following model

$$Y_{ni} = \mathbf{X}_i^T \alpha(U_i) + \widetilde{\mathbf{Z}}_{ni}^T \gamma_n + \varepsilon_i, \quad i = 1, \ldots, n. \tag{3.11}$$

Similar to (3.4), the bias-corrected auxiliary random vectors for θ can be defined by

$$\widetilde{\eta}_i(\theta) = \left(\widetilde{\mathbf{Z}}_{i1} - \hat{\mu}_1^T(U_i)\mathbf{X}_i\right)\left(Y_{ni} - \mathbf{X}_i^T\hat{\alpha}(U_i;\theta,\hat{\boldsymbol{\beta}}_{n(l)}) - \widetilde{\mathbf{Z}}_{i1}^T\theta - \widetilde{\mathbf{Z}}_{ni2}^T\hat{\boldsymbol{\beta}}_{n(l)}\right)$$
$$= \hat{\mathbf{Z}}_{i1}\left(\hat{Y}_{ni} - \hat{\mathbf{Z}}_{i1}^T\theta - \hat{\mathbf{Z}}_{ni2}^T\hat{\boldsymbol{\beta}}_{n(l)}\right). \quad (3.12)$$

In (3.12), $\hat{\boldsymbol{\beta}}_{n(l)}$ denotes the subvector of the last $p_n - l$ elements of $\hat{\gamma}_n$, which is the profile least-squares estimator (PLSE) or MBCELE based on model (3.11). $\hat{\mu}_1(u)$ is the estimator of $\mu_1(u) = (E(\mathbf{X}_i\mathbf{X}_i^T|U_i = u))^{-1}E(\mathbf{X}_i\widetilde{\mathbf{Z}}_{i1}^T|U_i = u)$, and $\hat{\alpha}(u;\theta,\hat{\boldsymbol{\beta}}_{n(l)}) = (I_q, \mathbf{0}_q)(\mathbf{D}_u^T\mathbf{W}_u\mathbf{D}_u)^{-1}\mathbf{D}_u^T\mathbf{W}_u(Y_n - \widetilde{\mathbf{Z}}_1^T\theta - \widetilde{\mathbf{Z}}_{n2}^T\hat{\boldsymbol{\beta}}_{n(l)})$, $\hat{Y}_{ni} = Y_{ni} - \sum_{k=1}^n S_{ik}Y_{nk}$, $\hat{\mathbf{Z}}_{i1} = \widetilde{\mathbf{Z}}_{i1} - \sum_{k=1}^n S_{ik}\widetilde{\mathbf{Z}}_{k1}$, $\hat{\mathbf{Z}}_{ni2} = \widetilde{\mathbf{Z}}_{ni2} - \sum_{k=1}^n S_{ik}\widetilde{\mathbf{Z}}_{nk2}$. Therefore, the bias-corrected empirical log-likelihood ratio for θ is defined by

$$\hat{\mathfrak{R}}_l(\theta) = 2\sum_{i=1}^n \log(1 + \kappa^T\widetilde{\eta}_i(\theta)), \quad (3.13)$$

where κ satisfies $\frac{1}{n}\sum_{i=1}^n \widetilde{\eta}_i(\theta)/[1 + \kappa^T\widetilde{\eta}_i(\theta)] = 0$. Then $\hat{\mathfrak{R}}_l(\theta)$ has the following Chi-square limiting distribution.

Theorem 3. *Under regularity conditions* (C1)–(C8) *in Appendix and assuming that $E\|\mathbf{Z}_{i1}\|^k \le \infty$ for $k \ge 4$, if $\theta = A_n\boldsymbol{\beta}_n$ is the true value of the parameter vector, then as $n \to \infty$, we have*

$$\hat{\mathfrak{R}}_l(\theta) \xrightarrow{d} \chi_l^2, \quad (3.14)$$

where χ_l^2 denotes the Chi-square distribution with l degrees of freedom.

Based on Theorem 3, $\hat{\mathfrak{R}}_l(\theta)$ can be used to construct confidence regions for the linear combination $\theta = A_n\boldsymbol{\beta}_n$. For any given $0 < \alpha < 1$, there exists c_α such that $P(\chi_l^2 > c_\alpha) = \alpha$, then

$$I_\alpha(\theta) = \{\theta \in R^l \mid \hat{\mathfrak{R}}_l(\theta) \le c_\alpha\}$$

is the confidence region of the linear combination θ with asymptotically correct coverage probability $1 - \alpha$.

3.3. Partially linear model

When $q = 1$ and $\mathbf{X} \equiv 1$, model (1.1) is reduced to a partially linear model with diverging number of parameters. In this case, $\mu^T(u) = E(\mathbf{Z}_n|U = u)$. Then the random vector (3.2) is defined by

$$\widetilde{\eta}_{ni}(\boldsymbol{\beta}_n) = (\mathbf{Z}_{ni} - \hat{\mu}^T(U_i))(Y_{ni} - \hat{\alpha}(U_i,\boldsymbol{\beta}_n) - \mathbf{Z}_{ni}^T\boldsymbol{\beta}_n). \quad (3.15)$$

Let $\widetilde{\mathfrak{R}}(\boldsymbol{\beta}_n)$ denote $\hat{\mathfrak{R}}(\boldsymbol{\beta}_n)$ with $\hat{\eta}_{ni}(\boldsymbol{\beta}_n)$ being replaced by $\widetilde{\eta}_{ni}(\boldsymbol{\beta}_n)$. We state the following result.

Theorem 4. *Under regularity conditions* (C1)–(C8) *in Appendix, if $\boldsymbol{\beta}_n$ is the true value of the parameter vector, $p_n^{3+2/(k-2)}/n \to 0$, then*

$$(2p_n)^{-1/2}(\widetilde{\mathfrak{R}}(\boldsymbol{\beta}_n) - p_n) \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty. \quad (3.16)$$

When $q = 1$ and $\mathbf{X} \equiv 1$, model (3.11) is reduced to the following partially linear model

$$Y_{ni} = \alpha(U_i) + \widetilde{\mathbf{Z}}_{ni}^T\gamma_n + \varepsilon_i, \quad i = 1, \ldots, n. \quad (3.17)$$

Then the bias-corrected auxiliary random vector (3.12) for θ is reduced to

$$\widetilde{\eta}_i(\theta) = (\widetilde{\mathbf{Z}}_{i1} - \hat{\mu}_1^T(U_i))(Y_{ni} - \hat{\alpha}(U_i;\theta,\hat{\boldsymbol{\beta}}_{n(l)}) - \widetilde{\mathbf{Z}}_{i1}^T\theta - \widetilde{\mathbf{Z}}_{ni2}^T\hat{\boldsymbol{\beta}}_{n(l)}). \quad (3.18)$$

Let $\widetilde{\mathfrak{R}}_l(\theta)$ denote the bias-corrected empirical log-likelihood ratio with the auxiliary random vector (3.18).

Theorem 5. *Under regularity conditions* (C1)–(C8) *in Appendix and assuming that $E\|\mathbf{Z}_{i1}\|^k \le \infty$ for $k \ge 4$, if $\theta = A_n\boldsymbol{\beta}_n$ is the true value of the parameter vector, then, as $n \to \infty$, we have*

$$\widetilde{\mathfrak{R}}_l(\theta) \xrightarrow{d} \chi_l^2, \quad (3.19)$$

where χ_l^2 denotes the Chi-square distribution with l degrees of freedom.

Based on Theorem 5, a confidence region for $\theta = A_n\boldsymbol{\beta}_n$ is given by $\widetilde{I}_\alpha(\theta) = \{\theta : \widetilde{\mathfrak{R}}_l(\theta) \le \chi_l^2(\alpha)\}$ for $0 \le \alpha \le 1$.

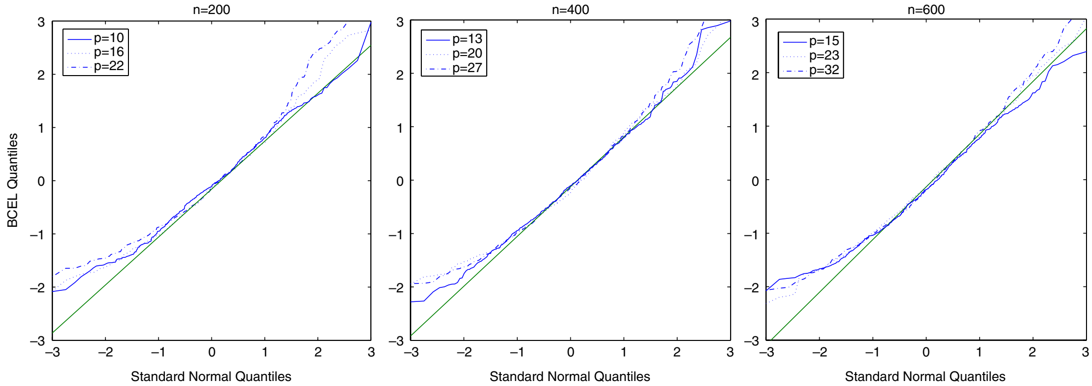

Fig. 1. BCEL ratio normal QQ-plots for the sample $n = 200$, 400 and 600 with the growth rate $p_n = \lfloor cn^{1/3} \rfloor$, where $c = 1.8$, 2.8 and 3.8.

4. Numerical studies

In this section, we present the results of Monte Carlo simulations to evaluate the asymptotic normality of BCEL ratio, and to compare BCEL with normal approximation that is based on the profile least-squares method (PLS) [20]. Throughout this section, we use the Epanechnikov kernel $K(u) = 0.75(1 - u^2)_+$, and use the "leave-one-out" cross-validation method to select the optimal handwidth h_{opt} satisfying condition (C5).

Consider the following varying coefficient partially linear model

$$Y_{ni} = \mathbf{X}_i^T \alpha(U_i) + \mathbf{Z}_{ni}^T \boldsymbol{\beta}_n + \varepsilon_i, \quad i = 1, \ldots, n. \tag{4.1}$$

In our simulation studies, the covariate U_i is uniformly distributed on [0,1], the nonparametric component $\alpha(u) = (\alpha_1(u), \alpha_2(u))^T$ with $q = 2$ in which $X_{i1} = 1$ and $X_{i2} \sim N(0, 1)$, the covariable $(\mathbf{Z}_{ni}^T, X_{i2})^T$ is a $(p_n + 1)$-dimensional normal random vector with mean zero and covariance matrix (σ_{ij}) with $\sigma_{ij} = 0.5^{|i-j|}$. The noise ε_i is generated from two different distributions: the standard normal and the t distribution with three degrees of freedom. Furthermore, $\boldsymbol{\beta}_n = [0.5, 0.3, -0.5, 1, 0.1, -0.25, 0, \ldots, 0]^T$, the coefficient functions are given as

$$\alpha_1(u) = 4 + \sin(2\pi u), \quad \text{and} \quad \alpha_2(u) = 2u(1 - u).$$

4.1. Simulation I

In this simulation, we evaluate the asymptotic normality of BCEL ratio through QQ-plots, and demonstrate the advantages of the bias-correction technique proposed in the different growth rates of p_n for each sample size. Here we only consider the case of the noise $\varepsilon \sim N(0, 1)$.

We draw 1000 random samples of sizes 200, 400 or 600 from model (4.1), respectively. For the dimension p_n of the parameter vector $\boldsymbol{\beta}_n$, we consider the growth rate $p_n = \lfloor cn^{1/3} \rfloor$. For comparison, we obtain three dimensions for each sample size by assigning $c = 1.8$, 2.8 and 3.8. The corresponding dimensions $p_n = 10$, 16 and 22 for $n = 200$, $p_n = 13$, 20 and 27 for $n = 400$, and $p_n = 15$, 23 and 32 for $n = 600$, respectively. Fig. 1 shows the QQ-plots for the BCEL ratio for three sample sizes with different growth rates.

Fig. 1 depicts the BCEL ratio $\hat{\mathcal{R}}(\boldsymbol{\beta}_n)$ asymptotically following $N(p_n, 2p_n)$ distribution, which is consistent with our asymptotic theory. From Fig. 1, we also see that the convergence of the standardized BCEL ratio to $N(0, 1)$ is faster for $c = 1.8$ case for each sample size than that of $c = 2.8$ and 3.8. In addition, the convergence becomes faster as the sample size increases, even when the growth rate of parameter dimensionality is the highest dimensionality $p_n = \lfloor 3.8n^{1/3} \rfloor$. This implies that the bias-correction technique plays an important role in the asymptotic normality of empirical log-likelihood ratio, specially for the case of much higher dimensionality. However, the QQ plots also show that for the chi-square variables, the standardizing ones still have non-negligible bias (both the tails are above the straight lines) unless the sample size is large enough.

4.2. Simulation II

For this simulation, we present the results of Monte Carlo simulations to compare BCEL with normal approximation that is based on the profile least-squares method (PLS) [20].

In the simulations, we draw 1000 random samples of sizes 200, 400 or 600 from model (4.1), respectively. Because the simulation results are similar for the three cases of $c = 1.8$, 2.8 and 3.8, we here take the dimensionality of the parametric component as $p_n = \lfloor 1.8n^{1/3} \rfloor$, then the corresponding dimensions of the parameter vector $\boldsymbol{\beta}_n$ are 10, 13 and 15, respectively. The confidence regions and their coverage probabilities, with the nominal level $1 - \alpha = 0.95$, are computed. When the

Table 1
The coverage probabilities (CP) and average lengths (AL) on β_1 when the nominal level is 0.95.

Method	n	$\varepsilon \sim N(0,1)$		$\varepsilon \sim t(3)$	
		CP	AL	CP	AL
	200	0.9420	0.2082	0.9290	0.2322
BCEL	400	0.9470	0.1019	0.9340	0.1391
	600	0.9490	0.0892	0.9410	0.1105
	200	0.9320	0.2426	0.9200	0.2703
PLS	400	0.9430	0.1245	0.9290	0.1618
	600	0.9480	0.1014	0.9360	0.1214

Table 2
The coverage probabilities (CP) on (β_1, β_2) and $(\beta_1, \beta_2, \beta_3)$, respectively, when the nominal level is 0.95.

Method	n	(β_1, β_2)		$(\beta_1, \beta_2, \beta_3)$	
		$\varepsilon \sim N(0,1)$	$\varepsilon \sim t(3)$	$\varepsilon \sim N(0,1)$	$\varepsilon \sim t(3)$
	200	0.9300	0.9230	0.9190	0.9120
BCEL	400	0.9410	0.9350	0.9250	0.9240
	600	0.9460	0.9410	0.9380	0.9350
	200	0.9310	0.9130	0.9180	0.9040
PLS	400	0.9390	0.9230	0.9220	0.9140
	600	0.9420	0.9310	0.9360	0.9270

noise $\varepsilon \sim N(0,1)$, the empirical coverage probabilities of $\boldsymbol{\beta}_n$ are 0.9020, 0.9140 and 0.9210, respectively, for the sample sizes $n = 200$, 400 and 600. From the simulations we can see that, as n increases, the coverage probabilities of $\boldsymbol{\beta}_n$ get bigger and approach the nominal level, even when the dimension of the parameter $\boldsymbol{\beta}_n$ grows with the sample size n.

In practice, we are often confronted with the need of constructing confidence intervals/regions for a particular regression coefficient or a certain linear combination of $\boldsymbol{\beta}_n$. For simplicity, the 95% confidence regions for the arbitrary linear combination $\theta = A_n\boldsymbol{\beta}_n$ are computed using the BCEL based on Theorem 3 and the profile least-squares (PLS), where A_n is an $l \times p_n$ matrix such that $A_nA_n^T \to G$ and G is an $l \times l$ nonnegative symmetric matrix. The normal approximation based confidence region is determined by the profile least-squares estimator (PLSE) $\boldsymbol{\beta}_n$: $\hat{\boldsymbol{\beta}}_n = \{\mathbf{Z}_n^T(I-\mathbf{S})^T(I-\mathbf{S})\mathbf{Z}_n\}^{-1}\mathbf{Z}_n^T(I-\mathbf{S})^T(I-\mathbf{S})Y_n$ that is obtained by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\{Y_{ni} - \mathbf{X}_i^T\hat{\alpha}(U_i, \boldsymbol{\beta}_n) - \mathbf{Z}_{ni}^T\boldsymbol{\beta}_n\}^2.$$

Similar to the results in [20], the PLSE can be proved to have the asymptotic normality as

$$\sqrt{n}A_nB\Sigma^{-1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n) \xrightarrow{d} N(0, G),$$

where $B = E[(\mathbf{Z}_n - \mu^T(U)\mathbf{X})(\mathbf{Z}_n - \mu^T(U)\mathbf{X})^T]$, Σ is defined in condition (C6). To construct confidence region, we also need to estimate the asymptotic variance that is of the form:

$$\hat{B} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{Z}_{ni} - \hat{\mu}^T(U_i)\mathbf{X}_i)(\mathbf{Z}_{ni} - \hat{\mu}^T(U_i)\mathbf{X}_i)^T = \frac{1}{n}\sum_{i=1}^{n}\hat{\mathbf{Z}}_{ni}\hat{\mathbf{Z}}_{ni}^T, \qquad \hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\hat{\mathbf{Z}}_{ni}\hat{\mathbf{Z}}_{ni}^T,$$

where $\hat{\varepsilon}_i = Y_{ni} - \mathbf{X}_i^T\alpha(U_i, \hat{\boldsymbol{\beta}}_n) - \mathbf{Z}_{ni}^T\hat{\boldsymbol{\beta}}_n$.

Now we consider the confidence regions for $A_n\boldsymbol{\beta}_n$, where A_n is any $1 \times p_n$ vector, $2 \times p_n$ and $3 \times p_n$ matrix, respectively. We evaluate the performance for the specific components β_1, (β_1, β_2) and $(\beta_1, \beta_2, \beta_3)$ (the results for other components are similar) and compare the BCEL with the PLS in terms of coverage accuracies of the confidence regions. The numerical results are reported in Tables 1 and 2. Figs. 2–3 show confidence regions for (β_1, β_2), (β_1, β_3) and (β_2, β_3), respectively. By the way, we also compute the MBCELE and PLSE, and present their bias and standard deviation for the case of $(n, p_n) = (400, 13)$ in Table 3 (the results for other cases are similar), where "Bias" represents the sample average over 1000 replications subtracting the true value of $\boldsymbol{\beta}_n$, "SD" represents the sample standard deviation over 1000 replications. We further give the average estimation errors $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|$ in L_2-norm (EE) to assess the performance of estimators.

Looking at Tables 1–3, and Figs. 2–3, we have the following results.

(1) From the 95% confidence intervals/regions and their coverage probabilities that are obtained by the two approaches, we find that BCEL consistently gives shorter lengths/smaller regions and achieves slightly higher coverage probabilities than those of PLS.

(2) The two approaches have better performance for the case of $\varepsilon \sim N(0,1)$ than that of $\varepsilon \sim t(3)$. From Fig. 3, it is easy to see that the confidence regions of the PLS are predetermined to be symmetric when the noise ε comes from the t distribution with three degrees of freedom. But the confidence regions of the BCEL are typically asymmetric for the case of $\varepsilon \sim t(3)$, and are determined entirely by the data.

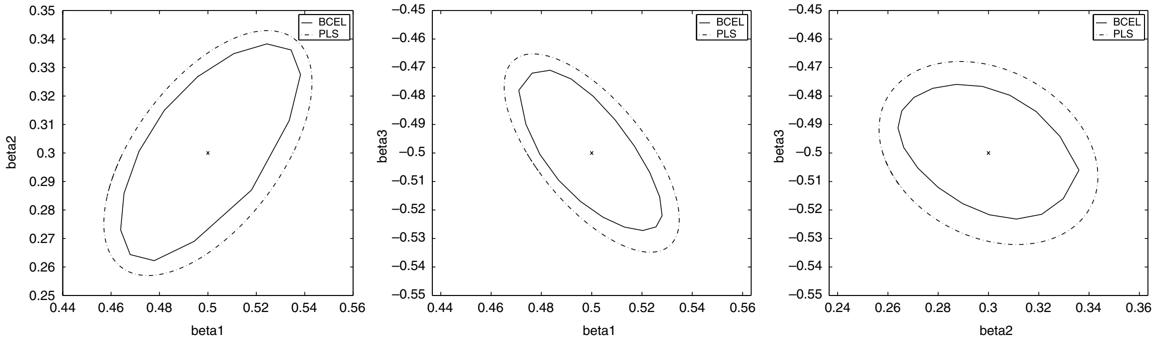

Fig. 2. 95% confidence regions for (β_1, β_2), (β_1, β_3) and (β_2, β_3) based on $n = 600$ when the noise comes from $\varepsilon \sim N(0, 1)$.

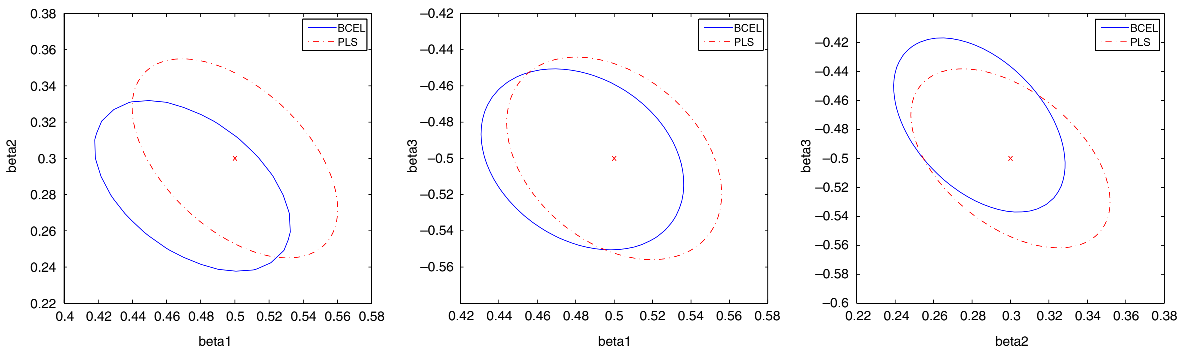

Fig. 3. 95% confidence regions for (β_1, β_2), (β_1, β_3) and (β_2, β_3) based on $n = 600$ when the noise comes from $\varepsilon \sim t(3)$.

Table 3
The biases and SD (in parentheses) for the parametric components based on 1000 replications for $(n, p_n) = (400, 13)$.

$\boldsymbol{\beta}_n$	MBCELE		PLSE	
	$\varepsilon \sim N(0, 1)$	$\varepsilon \sim t(3)$	$\varepsilon \sim N(0, 1)$	$\varepsilon \sim t(3)$
β_1	−0.0009(0.0388)	−0.0049(0.0763)	0.0008(0.0406)	−0.0046(0.0908)
β_2	0.0005(0.0447)	0.0039(0.0911)	0.0004(0.0438)	0.0086(0.1106)
β_3	0.0006(0.0449)	−0.0016(0.0974)	−0.0005(0.0443)	0.0014(0.1088)
β_4	−0.0003(0.0451)	−0.0087(0.0931)	−0.0019(0.0440)	−0.0085(0.1032)
β_5	0.0017(0.0416)	0.0097(0.0910)	0.0002(0.0441)	0.0112(0.1177)
β_6	−0.0039(0.0452)	−0.0050(0.1041)	0.0041(0.0431)	−0.0119(0.1184)
β_7	0.0015(0.0429)	0.0027(0.1080)	−0.0004(0.0436)	0.0131(0.1092)
β_8	0.0008(0.0443)	0.0039(0.1126)	0.0002(0.0438)	−0.0113(0.1126)
β_9	0.0001(0.0442)	−0.0136(0.1022)	−0.0003(0.0450)	−0.0159(0.1208)
β_{10}	−0.0010(0.0453)	0.0054(0.1065)	0.0012(0.0455)	0.0035(0.1082)
β_{11}	0.0008(0.0454)	0.0087(0.1008)	0.0010(0.0452)	−0.0083(0.1122)
β_{12}	0.0026(0.0445)	−0.0073(0.0930)	−0.0024(0.0444)	0.0077(0.1119)
β_{13}	−0.0009(0.0450)	−0.0030(0.0968)	0.0008(0.0439)	0.0102(0.1214)
EE	0.0056	0.0247	0.0055	0.0351

(3) From the biases, the sample standard deviations and the average estimation errors in Table 3, we observe that MBCEL and PLS are almost the same for the case of $\varepsilon \sim N(0, 1)$. When the noise is generated from $t(3)$ distribution, MBCEL has better performance than PLS. We think that the BCEL method can solve the optimal weights ω_i, $i = 1, \ldots, n$, in the estimate equation $\sum_{i=1}^{n} \omega_i \hat{\eta}_{ni}(\boldsymbol{\beta}_n) = 0$ for some long tail distributions.

5. Application: a real data example

In this section, we illustrate the bias-corrected empirical likelihood (BCEL) method and compare it with the profile least squares (PLS) method by using a real data set. The Fifth National Bank of Springfield faced a gender discrimination suit in which female employees received substantially smaller salaries than male employees. This example is based on a real case

Table 4
95% confidence intervals based on BCEL and PLS, and PLSE for Fifth National Bank data.

Method	Confidence intervals		PLSE
	BCEL	PLS	
Female	[−1.1459, −0.3982]	[−1.1970, −0.2742]	−0.7356
PCJob	[2.4908, 4.9531]	[2.2160, 5.6913]	3.9536
Edu1	[−3.2961, 0.1078]	[−3.5835, 0.0977]	−1.7429
Edu2	[−3.8672, −1.4957]	[−4.5518, −1.0181]	−2.7850
Edu3	[−3.3471, −1.5513]	[−3.6140, −1.0002]	−2.3071
Edu4	[−3.7249, 1.0115]	[−4.2487, 0.9479]	−1.6504
JobGrd1	[−24.0383, −21.2002]	[−25.2886, −20.6452]	−22.9669
JobGrd2	[−22.1956, −20.1524]	[−23.5001, −18.8935]	−21.1968
JobGrd3	[−18.1801, −16.4399]	[−19.6963, −15.2581]	−17.4772
JobGrd4	[−14.2885, −11.6168]	[−15.1970, −10.7037]	−12.9504
JobGrd5	[−9.8613, −6.1602]	[−9.8661, −5.0652]	−7.4656

with data dated 1995. Only the bank's name is changed. See Example 11.3 of [2]. These data consist of 208 employees with complete information on 8 recorded variables as follows.

- EduLev: educational level, a categorical variable with categories 1 (finished school), 2 (finished some college courses), 3 (obtained a bachelor's degree), 4 (took some graduate courses), 5 (obtained a graduate degree).
- JobGrade: job grade, a categorical variable indicating the current job level, the possible levels being 1–6 (6 the highest).
- YrHired: year that an employee was hired.
- YrBorn: year that an employee was born.
- Gender: a categorical variable with values "Female" and "Male", 1 for female employee and 0 for male employee.
- YrsPrior: number of years of work experience at another bank prior to working at the Fifth National Bank.
- PCJob: a dummy variable with value 1 if the employee's job is computer related and value 0 otherwise.
- Salary: current (1995) annual salary in thousands of dollars.

Fan and Peng [13] and Lam and Fan [20] conducted such a salary analysis using the additive model and the generalized varying coefficient partially linear model, respectively. In this subsection, we consider the following varying coefficient partially linear model (VCPLM)

$$\text{Salary} = \alpha_1(\text{Age}) + \alpha_2(\text{Age})\text{YrsExp} + \beta_1\text{Female} + \beta_2\text{PCJob} + \sum_{i=1}^{4}\beta_{2+i}\text{Edu}_i + \sum_{i=1}^{5}\beta_{6+i}\text{JobGrd}_i + \varepsilon,$$

where the variable YrsExp is total years of working experience, computed from the variables YrHired and YrsPrior.

Similar to the analysis of Fan and Peng [13], we deleted the samples with age over 60 or working experience over 30. They correspond mainly to the company executives who earned handsome salaries. As a result of this deletion, a sample of size 199 remains for our analysis. The "leave-one-out" cross-validation method is employed to select the bandwidth $h_{GCV} = 13.9$. Table 4 presents 95% confidence intervals for the parameters based on the BCEL and PLS methods, and provides the profile least squares estimator (PLSE) of the parameters. For comparison, we obtain the 95% confidence regions for the coefficients (β_1, β_2) of variables Female and PCJob that are shown in Fig. 4. We also can obtain similar results for other coefficients. However, we omit the presentation.

It is seen from Table 4 that the bias-corrected empirical likelihood confidence intervals are comparable to those of the profile least squares method. The profile least squares method gives larger intervals and imposes symmetry on the confidence intervals. Fan and Peng [13] analyzed the Fifth National Bank data set by using the variable selection method, and they found that the effect of EduLev with the Salary was insignificant, and the corresponding variables were excluded from the final fitted model. From Table 4, we find that the confidence intervals of Edu1 and Edu4 cover the zero point. In addition, the coefficient for Female is significantly negative and the coefficient for PCJob is significantly positive. The job grade (JobGrd) plays an important role for the employees' salaries. Those findings corroborate the results in [13] very well.

Fig. 4 indicates that, for this dataset, the BCEL-based confidence region is smaller than the one based on that of the PLS.

Acknowledgments

Gaorong Li's research was supported by the National Nature Science Foundation of China (11101014), Ph.D. Program Foundation of Ministry of Education of China (20101103120016), PHR (IHLB, PHR20110822), Training Programme Foundation for the Beijing Municipal Excellent Talents (2010D005015000002) and Doctor Foundation of BJUT (X0006013201101). Lu Lin's research was supported by NBRP (973 Program 2007CB814901) of China, the National Nature Science Foundation of China (11171188), NSF and SRRF projects (ZR2010AZ001 and BS2011SF006) of Shandong Province of China and K.C. Wong-HKBU Fellowship Programme for Mainland China Scholars 2010–11. Lixing Zhu's research was supported by a grant from the Research Grants Council of Hong Kong, and an FRG grant from Hong Kong Baptist University, Hong Kong. Special thanks go to Professor Fan, J.Q. and Dr. Lam, C. for allowing us to use the real dataset described in this paper, and to Professor Chen, S.X. for a constructive discussion.

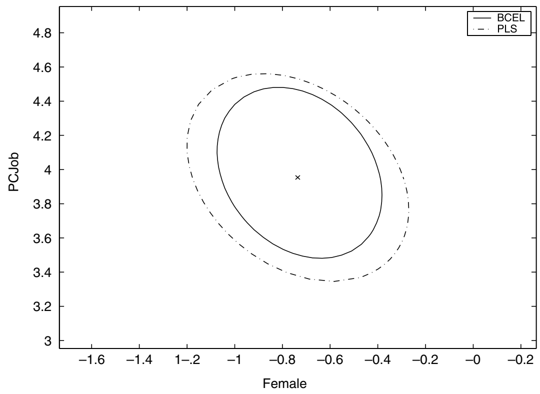

Fig. 4. 95% confidence regions for (β_{Female}, β_{PCJob}), based on BCEL and PLS, where "×" is the PLS estimator (−0.7356, 3.9536).

Appendix A. Proof of the theorems

Note. We present all the details of the proofs for review convenience, which makes the paper long. It will be much shorten afterward or will be put in a supplement if it is required.

Before we give the details of the proofs, we present some regularity conditions. Throughout the paper, we denote $\gamma_1(A) \le \cdots \le \gamma_{p_n}(A)$ as the eigenvalues and $\mathrm{tr}(A)$ as the trace operator of a matrix A.

(C1) The random variable U has a compact support Ω. The density function $f_U(u)$ of U has a continuous second derivative and is uniformly bounded away from zero.

(C2) The $q \times q$ matrix $E(\mathbf{X}\mathbf{X}^T|U)$ is non-singular for each $U \in \Omega$. $E(\mathbf{X}\mathbf{X}^T|U)$, $E(\mathbf{X}\mathbf{X}^T|U)^{-1}$ and $E(\mathbf{X}\mathbf{Z}_n^T|U)$ are all Lipschitz continuous and each element of $E(\mathbf{X}\mathbf{X}^T|U)^{-1}$ and $E(\mathbf{X}\mathbf{Z}_n^T|U)$ is bounded.

(C3) $\{\alpha_i(\cdot), i = 1, \ldots, q\}$ have continuous second derivatives in $u \in \Omega$.

(C4) The kernel $K(\cdot)$ is a bounded symmetric density function with bounded support.

(C5) The bandwidth h satisfies that $nh^6 \to 0$ and $nh^3/(\log n)^3 \to \infty$.

(C6) $\Sigma = E[\varepsilon^2(\mathbf{Z}_n - \mu^T(U)\mathbf{X})(\mathbf{Z}_n - \mu^T(U)\mathbf{X})^T]$ is a positive definite matrix with all eigenvalues being uniformly bounded away from zero and infinity.

(C7) $E(\varepsilon|U, \mathbf{X}, \mathbf{Z}_n) = 0$ almost surely. Furthermore, for some integer $k \ge 4$, $E(\|\mathbf{X}\varepsilon\|^k) < \infty$, $E\|\mathbf{X}\|^k < \infty$, $E(|\varepsilon|^k) < \infty$.

(C8) Let $\eta_n = \varepsilon(\mathbf{Z}_n - \mu^T(U)\mathbf{X})$, and η_{nj} be the j-th component of $\eta_n, j = 1, \ldots, p_n$. For k of condition (C7), there is a positive constant c such that as $n \to \infty$,

$$E(\|\eta_n(\boldsymbol{\beta}_n)/\sqrt{p_n}\|^k) < c, \qquad E(\|\mathbf{Z}_n\mathbf{X}^T/\sqrt{p_n}\|^k) < c, \qquad E(\|\mu(U)\mathbf{X}\mathbf{X}^T/\sqrt{p_n}\|^k) < c,$$

and

$$\frac{1}{p_n}\sum_{l_1=1}^{p_n} E(|\eta_{nl_1}|(\|\mathbf{Z}_n\mathbf{X}^T/\sqrt{p_n}\|^4 + \|\mu\mathbf{X}\mathbf{X}^T/\sqrt{p_n}\|^4)) < c.$$

(C9) $\max_{1\le l_1, l_2, l_3\le p_n} E(\eta_{nl_1}\eta_{nl_2}\eta_{nl_3})^2$ is bounded, where η_{nl_i} are the components of η_n.

Note that the above conditions are assumed to hold uniformly in $u \in \Omega$. Conditions (C1)–(C4) are also found in [10]. These conditions are actually quite mild and can be easily satisfied. Condition (C5) gives a range, from $O(n^{-1/3}\log n)$ to $O(n^{-1/6})$, of bandwidth that includes the optimal bandwidth. Condition (C6) ensures that there exists an asymptotic variance for the estimator of the growing parameters $\boldsymbol{\beta}_n$. Conditions (C7)–(C9) are technical conditions for the moments.

In this section, we present the proofs for the main results. Some lemmas and their proofs, that are needed for the proofs of the main theorems, are relegated to Appendix B.

In order to prove the main results, we introduce the following notations. By (3.2), simple calculation yields that

$$\hat{\eta}_{ni}(\boldsymbol{\beta}_n) = \eta_{ni}(\boldsymbol{\beta}_n) + \sum_{k=1}^{3} M_{i,k} =: \eta_{ni}(\boldsymbol{\beta}_n) + R_i, \tag{A.1}$$

where

$$\eta_{ni}(\boldsymbol{\beta}_n) = (\mathbf{Z}_{ni} - \mu^T(U_i)\mathbf{X}_i)(Y_{ni} - \mathbf{X}_i^T\alpha(U_i) - \mathbf{Z}_{ni}^T\boldsymbol{\beta}_n) = (\mathbf{Z}_{ni} - \mu^T(U_i)\mathbf{X}_i)\varepsilon_i,$$
$$M_{i,1} = (\mathbf{Z}_{ni} - \mu^T(U_i)\mathbf{X}_i)\mathbf{X}_i^T(\alpha(U_i) - \hat{\alpha}(U_i, \boldsymbol{\beta}_n)),$$
$$M_{i,2} = (\mu(U_i) - \hat{\mu}(U_i))^T\mathbf{X}_i\varepsilon_i,$$
$$M_{i,3} = [(\mu(U_i) - \hat{\mu}(U_i))^T\mathbf{X}_i][\mathbf{X}_i^T(\alpha(U_i) - \hat{\alpha}(U_i, \boldsymbol{\beta}_n))].$$

Note that when $\boldsymbol{\beta}_n$ is true value, $\eta_{ni}(\boldsymbol{\beta}_n)$ is actually the independent copies of η_n that is free of $\boldsymbol{\beta}_n$, where η_n is defined in condition (C8). However, in order to highlight the existence of this parameter, we still write it in $\eta_{ni}(\boldsymbol{\beta}_n)$.

Proof of Theorem 1. By (3.8), $\lambda \in R^{p_n}$ satisfies

$$0 = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\eta}_{ni}(\boldsymbol{\beta}_n)}{1+\lambda^T \hat{\eta}_{ni}(\boldsymbol{\beta}_n)} =: \varphi(\lambda). \tag{A.2}$$

Let $\lambda = \rho\theta$, where $\rho \geq 0$, $\theta \in R^{p_n}$ and $\|\theta\| = 1$. Introduce

$$\bar{\hat{\eta}}(\boldsymbol{\beta}_n) = \frac{1}{n}\sum_{i=1}^{n} \hat{\eta}_{ni}(\boldsymbol{\beta}_n), \qquad \hat{\eta}^*(\boldsymbol{\beta}_n) = \max_{1\leq i\leq n} \|\hat{\eta}_{ni}(\boldsymbol{\beta}_n)\|. \tag{A.3}$$

Substituting $1/(1+\lambda^T\hat{\eta}_{ni}(\boldsymbol{\beta}_n)) = 1 - \lambda^T\hat{\eta}_{ni}(\boldsymbol{\beta}_n)/(1+\lambda^T\hat{\eta}_{ni}(\boldsymbol{\beta}_n))$ into $\theta^T\varphi(\lambda) = 0$, we have

$$|\theta^T\bar{\hat{\eta}}(\boldsymbol{\beta}_n)| \geq \frac{\rho}{1+\rho\hat{\eta}^*(\boldsymbol{\beta}_n)}\theta^T S_n \theta,$$

where $S_n = \frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{ni}(\boldsymbol{\beta}_n)\hat{\eta}_{ni}^T(\boldsymbol{\beta}_n)$. Because $\hat{p}_i = \frac{1}{n}\frac{1}{1+\lambda^T\hat{\eta}_{ni}(\boldsymbol{\beta}_n)}$ is a probability mass, we have $0 < 1+\lambda^T\hat{\eta}_{ni}(\boldsymbol{\beta}_n) \leq 1+\rho\hat{\eta}^*(\boldsymbol{\beta}_n)$. Therefore,

$$\rho[\theta^T S_n\theta - \theta^T\bar{\hat{\eta}}(\boldsymbol{\beta}_n)\hat{\eta}^*(\boldsymbol{\beta}_n)] \leq |\theta^T\bar{\hat{\eta}}(\boldsymbol{\beta}_n)|. \tag{A.4}$$

From (A.1), it is easy to see that

$$\hat{\eta}^*(\boldsymbol{\beta}_n) \leq \eta^*(\boldsymbol{\beta}_n) + \max_{1\leq i\leq n}\|R_i\|, \tag{A.5}$$

where $\eta^*(\boldsymbol{\beta}_n) = \max_{1\leq i\leq n}\|\eta_{ni}(\boldsymbol{\beta}_n)\|$ and $\{\eta_{ni}(\boldsymbol{\beta}_n), i = 1,\ldots,n\}$ is a sequence of independent random variables with common distribution. From conditions (C7) and (C8), for any $\epsilon > 0$, then, again recalling the definition of $\eta_{ni}(\boldsymbol{\beta}_n)$ right below (A.1),

$$\begin{aligned}
P\left\{\eta^*(\boldsymbol{\beta}_n) \geq (p_n)^{1/2}n^{1/k}\epsilon\right\} &\leq \sum_{i=1}^{n} P\left\{\|\eta_{ni}(\boldsymbol{\beta}_n)\| \geq (p_n)^{1/2}n^{1/k}\epsilon\right\} \\
&\leq \frac{1}{np_n^{k/2}\epsilon^k}\sum_{i=1}^{n} E\|\eta_{ni}(\boldsymbol{\beta}_n)\|^k \\
&= \frac{1}{\epsilon^k}E\|\eta_{n1}(\boldsymbol{\beta}_n)/p_n^{1/2}\|^k.
\end{aligned} \tag{A.6}$$

Cauchy–Schwarz inequality yields that $\|\eta_{n1}(\boldsymbol{\beta}_n)/p_n^{1/2}\|^k \leq 1/p_n\sum_{j=1}^{p_n}|\eta_{n1j}(\boldsymbol{\beta}_n)|^k$, where $\eta_{n1j}(\boldsymbol{\beta}_n)$ are the components of $\eta_{n1}(\boldsymbol{\beta}_n)$. By (A.6), $\max_{1\leq i\leq n}\|\eta_{ni}(\boldsymbol{\beta}_n)\| = o_P(\sqrt{p_n}n^{1/k})$. Next we consider $\max_{1\leq i\leq n}\|R_i\|$. Note that each column vector of $\mu(U_i)$ is p_n-dimensional. By Lemma B.2 in Appendix B, and conditions (C7) and (C8), similar to (A.6) applied to the maxima on the right hand side of the second inequality below, we have, noting that the existence of k-th moments for $k \geq 4$,

$$\begin{aligned}
\max_{1\leq i\leq n}\|R_i\| &\leq \max_{1\leq i\leq n}\|M_{i,1}\| + \max_{1\leq i\leq n}\|M_{i,2}\| + \max_{1\leq i\leq n}\|M_{i,3}\| \\
&\leq \left(\max_{1\leq i\leq n}\|\mathbf{Z}_{ni}\mathbf{X}_i^T\| + \max_{1\leq i\leq n}\|\mu^T(U_i)\mathbf{X}_i\mathbf{X}_i^T\|\right)O_P(c_n) \\
&\quad + O_P(\sqrt{p_n}c_n)\max_{1\leq i\leq n}\|\mathbf{X}_i\varepsilon_i\| + O_P(\sqrt{p_n}c_n^2)\max_{1\leq i\leq n}\|\mathbf{X}_i\mathbf{X}_i^T\| \\
&\leq o_P(n^{1/k}\sqrt{p_n}c_n) + o_P(n^{1/k}\sqrt{p_n}c_n) + o_P(n^{1/k}\sqrt{p_n}c_n^2) \\
&= o_P(n^{1/k}\sqrt{p_n}c_n).
\end{aligned} \tag{A.7}$$

This, together with (A.5)–(A.7), gives $\hat{\eta}^*(\boldsymbol{\beta}_n) = o_P(n^{1/k}\sqrt{p_n})$. By the condition $p_n = o(n^{(k-2)/(2k)})$ in Theorem 1, it is easy to check that

$$\hat{\eta}^*(\boldsymbol{\beta}_n) = o_P(n^{1/k}\sqrt{p_n}) = o_P(\sqrt{n/p_n}n^{(2-k)/(2k)}p_n) = o_P(\sqrt{n/p_n}).$$

Because $|\theta^T\bar{\hat{\eta}}(\boldsymbol{\beta}_n)| \le \|\bar{\hat{\eta}}(\boldsymbol{\beta}_n)\| = O_P(\sqrt{p_n/n})$, then

$$\hat{\eta}^*(\boldsymbol{\beta}_n)|\theta^T\bar{\hat{\eta}}(\boldsymbol{\beta}_n)| = o_P(1).$$

This, together with (A.4), gives

$$|\rho[\theta^T S_n\theta + o_P(1)]| = O_P(\sqrt{p_n/n}). \tag{A.8}$$

Further, noting that by Lemma B.4 in Appendix B, $\theta^T S_n\theta$ converges to a positive constant in probability, $\rho = O_P(\sqrt{p_n/n})$, that is, $\|\lambda\| = \rho = O_P(\sqrt{p_n/n})$. □

Proof of Theorem 2. Let $W_i = \lambda^T\hat{\eta}_{ni}(\boldsymbol{\beta}_n)$. Applying Taylor expansion to (A.2), we obtain

$$\begin{aligned} 0 = \varphi(\lambda) &= \frac{1}{n}\sum_{i=1}^n \hat{\eta}_{ni}(\boldsymbol{\beta}_n)\left[1 - W_i + \frac{W_i^2}{1-W_i}\right] \\ &= \bar{\hat{\eta}}(\boldsymbol{\beta}_n) - S_n\lambda + \delta_n, \end{aligned} \tag{A.9}$$

where

$$\begin{aligned} \delta_n &= \frac{1}{n}\sum_{i=1}^n \frac{\hat{\eta}_{ni}(\boldsymbol{\beta}_n)W_i^2}{1-W_i} = \frac{1}{n}\sum_{i=1}^n \hat{\eta}_{ni}(\boldsymbol{\beta}_n)W_i^2 + \frac{1}{n}\sum_{i=1}^n \frac{\hat{\eta}_{ni}(\boldsymbol{\beta}_n)W_i^3}{1-W_i} \\ &=: \delta_{n1} + \delta_{n2}. \end{aligned} \tag{A.10}$$

For a preparation, we first present a bound for R_i. First, we note that $\max_{1\le i\le n}\|R_i\| = o_P(n^{1/k}\sqrt{p_n})$, and

$$\begin{aligned} \|R_i\| &\le \left(\|\mathbf{Z}_{ni}\mathbf{X}_i^T\| + \|\mu^T(U_i)\mathbf{X}_i\mathbf{X}_i^T\|\right)o_P(c_n) + o_P(\sqrt{p_n}c_n)\|\mathbf{X}_i\varepsilon_i\| + o_P(\sqrt{p_n}c_n^2)\|\mathbf{X}_i\mathbf{X}_i^T\| \\ &\le \left(\|\mathbf{Z}_{ni}\mathbf{X}_i^T\| + \|\mu^T(U_i)\mathbf{X}_i\mathbf{X}_i^T\|\right)o_P(c_n) + o_P(\sqrt{p_n}c_n)(\|\mathbf{X}_i\varepsilon_i\| + \|\mathbf{X}_i\mathbf{X}_i^T\|) \\ &=: b_{1ni}o_P(c_n) + b_{2i}o_P(\sqrt{p_n}c_n), \end{aligned} \tag{A.11}$$

where $o_P(c_n)$ is the convergence rate of the maximum difference between the nonparametric estimator $\hat{\alpha}$ (or $\hat{\mu}$) and the corresponding true function α (or μ). It is obtained by Lemma B.2 in Appendix B. These results can be used in the later steps of the proof.

From (A.6) in the proof of Theorem 1, we know that $\hat{\eta}^*(\boldsymbol{\beta}_n) = o_P(n^{1/k}\sqrt{p_n})$. Note that the condition $p_n = o(n^{(k-2)/(3k-4)})$ in Theorem 2. Thus,

$$\max_{1\le i\le n}|W_i| = \max_{1\le i\le n}\|\lambda^T\hat{\eta}_{ni}(\boldsymbol{\beta}_n)\| \le \|\lambda\|\hat{\eta}^*(\boldsymbol{\beta}_n) = o_P(p_n/n^{1/2-1/k}).$$

Then we have $\delta_{n2} = \delta_{n21}(1 + o_P(1))$, where $\delta_{n21} = n^{-1}\sum_{i=1}^n \|\hat{\eta}_{ni}(\boldsymbol{\beta}_n)W_i^3\|$. Further, we note that invoking C_r inequality, $\|\hat{\eta}_{ni}(\boldsymbol{\beta}_n)\|^k \le 2^{k/2}(\|\eta_{ni}(\boldsymbol{\beta}_n)\|^2 + \|R_i\|^2)^{k/2} \le C(\|\eta_{ni}(\boldsymbol{\beta}_n)\|^k + \|R_i\|^k)$ for some C depending on k.

We now deal with δ_{n21}. When $k \ge 4$, together with (A.11), Theorem 1 and the Hölder inequality, we have

$$\begin{aligned} \|\delta_{n21}\| &\le \hat{\eta}_{ni}^*(\boldsymbol{\beta}_n)\frac{1}{n}\sum_{i=1}^n |W_i|^3 = \hat{\eta}_{ni}^*(\boldsymbol{\beta}_n)\frac{1}{n}\sum_{i=1}^n |W_i|^{\frac{2(k-3)}{k-2}}|W_i|^{\frac{k}{k-2}} \\ &\le \hat{\eta}_{ni}^*(\boldsymbol{\beta}_n)\left(\frac{1}{n}\sum_{i=1}^n |W_i|^{\frac{2(k-3)}{k-2}\frac{k-2}{k-3}}\right)^{\frac{k-3}{k-2}}\left(\frac{1}{n}\sum_{i=1}^n |W_i|^{\frac{k}{k-2}(k-2)}\right)^{\frac{1}{k-2}} \\ &= \hat{\eta}_{ni}^*(\boldsymbol{\beta}_n)\left(\frac{1}{n}\sum_{i=1}^n W_i^2\right)^{\frac{k-3}{k-2}}\left(\frac{1}{n}\sum_{i=1}^n |W_i|^k\right)^{\frac{1}{k-2}} \\ &\le \hat{\eta}_{ni}^*(\boldsymbol{\beta}_n)(\lambda^T S_n\lambda)^{\frac{k-3}{k-2}}\left(\frac{1}{n}\sum_{i=1}^n \|\lambda\|^k\|\hat{\eta}_{ni}(\boldsymbol{\beta}_n)\|^k\right)^{\frac{1}{k-2}} \\ &\le \hat{\eta}_{ni}^*(\boldsymbol{\beta}_n)(\lambda^T S_n\lambda)^{\frac{k-3}{k-2}}\left(\frac{1}{n}\sum_{i=1}^n \|\lambda\|^k C(\|\eta_{ni}(\boldsymbol{\beta}_n)\|^k + \|R_i\|^k)\right)^{\frac{1}{k-2}} \\ &= o_P(p_n^{1/2}n^{1/k})o_P((p_n n^{-1})^{\frac{k-3}{k-2}})o_P\{(p_n^{k/2}n^{-k/2}p_n^{k/2})^{\frac{1}{k-2}}\} \\ &= o_P\left(p_n^{\frac{5(k-2)+2}{2(k-2)}}n^{-3/2+1/k}\right) = o_P(p_n^3 n^{-3/2}). \end{aligned} \tag{A.12}$$

The last equation is due to $p_n = o(n^{(k-2)/(3k-4)})$ and $k \geq 4$. When $k > 8$, without the condition that $E(\varepsilon^3|U, \mathbf{X}, \mathbf{Z}_n) = 0$, Lemma B.4 and the proof of Theorem 1 yield that

$$\|\delta_{n1}\| \leq \hat{\eta}^*(\boldsymbol{\beta}_n)\lambda^T S_n \lambda = o_P(p_n^{3/2} n^{1/k-1}). \tag{A.13}$$

Together with the bound for $\|\delta_{n21}\|$, we have $\|\delta_n\| = o_P(p_n^{3/2} n^{1/k-1})$. We now consider convergence rate of $\|\delta_{n1}\|$ in the case with the condition that $E(\varepsilon^3|U, \mathbf{X}, \mathbf{Z}_n) = 0$. From (A.1) and $W_i = \lambda^T \hat{\eta}_{ni}(\boldsymbol{\beta}_n)$, we have

$$\begin{aligned}\delta_{n1} = {} & \frac{1}{n}\sum_{i=1}^n \eta_{ni}(\boldsymbol{\beta}_n)\lambda^T \eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n)\lambda + \frac{2}{n}\sum_{i=1}^n \eta_{ni}(\boldsymbol{\beta}_n)\lambda^T \eta_{ni}(\boldsymbol{\beta}_n)R_i^T\lambda \\ & + \frac{1}{n}\sum_{i=1}^n \eta_{ni}(\boldsymbol{\beta}_n)\lambda^T R_i R_i^T \lambda + \frac{1}{n}\sum_{i=1}^n R_i \lambda^T \eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n)\lambda \\ & + \frac{2}{n}\sum_{i=1}^n R_i\lambda^T \eta_{ni}(\boldsymbol{\beta}_n)R_i^T\lambda + \frac{1}{n}\sum_{i=1}^n R_i\lambda^T R_i R_i^T\lambda \\ =: {} & I_1 + I_2 + I_3 + I_4 + I_5 + I_6. \end{aligned} \tag{A.14}$$

We now deal with all I_i, $i = 1, \ldots, 6$. Consider $\|I_1\|$. By (A.1), $\eta_{ni}(\boldsymbol{\beta}_n) = (\mathbf{Z}_{ni} - \mu^T(U_i)\mathbf{X}_i)\varepsilon_i =: \widetilde{\mathbf{Z}}_{ni}\varepsilon_i$, then

$$\begin{aligned} I_1 &= \frac{1}{n}\sum_{i=1}^n \eta_{ni}(\boldsymbol{\beta}_n)[\lambda^T \eta_{ni}(\boldsymbol{\beta}_n)]^2 \\ &= \frac{1}{n}\sum_{i=1}^n \varepsilon_i^3 \widetilde{\mathbf{Z}}_{ni}[\lambda^T \widetilde{\mathbf{Z}}_{ni}]^2. \end{aligned} \tag{A.15}$$

By condition $E(\varepsilon^3|U, \mathbf{Z}_n, \mathbf{X}) = 0$, we have $E(I_1) = 0$. Invoking the independence of ε_i from the other variables and using the Cauchy–Schwarz inequality, we have

$$\begin{aligned} \|I_1\| &= \left\| \frac{1}{n}\sum_{i=1}^n \varepsilon_i^3 \widetilde{\mathbf{Z}}_{ni}[\lambda^T \widetilde{\mathbf{Z}}_{ni}]^2 \right\| \\ &= \left(\sum_{l_1=1}^{p_n} \left(\frac{1}{n}\sum_{i=1}^n \varepsilon_i^3 \widetilde{Z}_{nil_1}[\lambda^T \widetilde{\mathbf{Z}}_{ni}]^2 \right)^2 \right)^{1/2} \\ &= \left(\sum_{l_1=1}^{p_n} \left(\frac{1}{n}\sum_{i=1}^n \varepsilon_i^3 \widetilde{Z}_{nil_1} \left[\sum_{l_2,l_3=1}^{p_n} \lambda_{l_2}\lambda_{l_3}\widetilde{Z}_{nil_2}\widetilde{Z}_{nil_3} \right] \right)^2 \right)^{1/2} \\ &= \left(\sum_{l_1=1}^{p_n} \left(\lambda^T H_{nl_1}\lambda\right)^2 \right)^{1/2} \leq \|\lambda\|^2 \left(\sum_{l_1=1}^{p_n} \|H_{nl_1}\|^2 \right)^{1/2}, \end{aligned} \tag{A.16}$$

where for any l_1, let H_{nl_1} be the $p_n \times p_n$ matrix whose elements are $\left(\frac{1}{n}\sum_{i=1}^n \varepsilon_i^3 \widetilde{Z}_{nil_1}\widetilde{Z}_{nil_2}\widetilde{Z}_{nil_3}\right)$. Note that the mean of each element of H_{nl_1} is equal to 0. To obtain its convergence rate, we can compute the mean of $\sum_{l_1=1}^{p_n} \|H_{nl_1}\|^2 = \sum_{l_1,l_2,l_3=1}^{p_n} \left(\frac{1}{n}\sum_{i=1}^n \varepsilon_i^3 \widetilde{Z}_{nil_1}\widetilde{Z}_{nil_2}\widetilde{Z}_{nil_3}\right)^2$. By the independence of $\varepsilon_i^3 \widetilde{Z}_{nil_1}\widetilde{Z}_{nil_2}\widetilde{Z}_{nil_3}$ and condition (C9), we immediately derive that

$$\frac{1}{n}\sum_{l_1,l_2,l_3=1}^{p_n} E\left(\varepsilon_i^3 \widetilde{Z}_{nil_1}\widetilde{Z}_{nil_2}\widetilde{Z}_{nil_3}\right)^2 \leq \frac{p_n^3}{n} \max_{l_1,l_2,l_3} E\left(\varepsilon_i^3 \widetilde{Z}_{nil_1}\widetilde{Z}_{nil_2}\widetilde{Z}_{nil_3}\right)^2 = O(p_n^3/n).$$

In other words, combining the rate of $\|\lambda\|$, $\|\lambda\|^2 \left(\sum_{l_1=1}^{p_n} \|H_{nl_1}\|^2\right)^{1/2} = o_P(p_n^{5/2} n^{-3/2})$. This is the convergence rate of $\|I_1\|$.

For I_2, by (A.11) and conditions (C7) and (C8), invoking the proof of Theorem 1 about the boundedness of eigenvalues of S_n and the Cauchy–Schwarz inequality, we have

$$\begin{aligned} \|I_2\| &= \left\| \frac{2}{n}\sum_{i=1}^n \eta_{ni}(\boldsymbol{\beta}_n)\lambda^T \eta_{ni}(\boldsymbol{\beta}_n)R_i^T\lambda \right\| \\ &\leq 2\sqrt{\left(\sum_{j=1}^{p_n} \left| \frac{1}{n}\sum_{i=1}^n (\lambda^T \eta_{ni}(\boldsymbol{\beta}_n))(\eta_{nij}(\boldsymbol{\beta}_n)R_i^T\lambda) \right|^2 \right)} \end{aligned}$$

$$
\begin{aligned}
&\leq 2\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(\lambda^T\eta_{ni}(\boldsymbol{\beta}_n))^2\right)\sum_{j=1}^{p_n}\left(\frac{1}{n}\sum_{i=1}^{n}(\eta_{nij}(\boldsymbol{\beta}_n)R_i^T\lambda)^2\right)} \\
&\leq 2\|\lambda\|^2\gamma_{p_n}^{1/2}(\tilde{S}_n)\left(\sum_{j=1}^{p_n}\frac{1}{n}\sum_{i=1}^{n}\|\eta_{nij}(\boldsymbol{\beta}_n)R_i\|^2\right)^{1/2} \\
&\leq o_P(c_n)\|\lambda\|^2\gamma_{p_n}^{1/2}(\tilde{S}_n)\left(\sum_{j=1}^{p_n}\frac{1}{n}\sum_{i=1}^{n}(\|\eta_{nij}(\boldsymbol{\beta}_n)b_{1ni}\|^2+p_n|\eta_{nij}(\boldsymbol{\beta}_n)b_{2i}|^2)\right)^{1/2} \\
&\leq o_P(c_n)o_P(p_nn^{-1})\gamma_{p_n}^{1/2}o_P(p_n)=o_P(p_n^2n^{-1}c_n),
\end{aligned}
\tag{A.17}
$$

where $\gamma_{p_n}(\tilde{S}_n)$ is the largest eigenvalue of $\tilde{S}_n$, and $\tilde{S}_n=\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n)$. For I_3, similar arguments above apply to R_i to achieve, invoking condition (C8),

$$
\begin{aligned}
\|I_3\| &= \left\|\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)\lambda^TR_iR_i^T\lambda\right\| \\
&\leq \|\lambda\|^2\sqrt{\sum_{j=1}^{p_n}\left(\frac{1}{n}\sum_{i=1}^{n}|\eta_{nij}(\boldsymbol{\beta}_n)|^2\|R_i\|^4\right)} \\
&\leq o_P(p_nn^{-1})O_P(p_n^{3/2}c_n^2) \\
&= o_P((p_n^{5/2}n^{-1})c_n^2).
\end{aligned}
\tag{A.18}
$$

For $\|I_4\|$, it is easy to see that it is smaller or equal to $\max\|R_i\|\,\|\lambda\|^2\gamma_{p_n}(\tilde{S}_n)=o_P(p_n^{3/2}n^{-1})c_n$. For $\|I_5\|$, we use R_i in the place of $\eta_n(\boldsymbol{\beta}_n)$ and follow the similar argument for $\|I_2\|$, we can derive that $\|I_5\|=o_P((p_n^2n^{-1})c_n^2)$. Following the argument for $\|I_3\|$ we can get $\|I_6\|=o_P((p_n^{5/2}n^{-1})c_n^3)$.

Summarizing the above results, and noting that the convergence rates about I_6 and I_5 (and I_4) are faster than those of I_3 and I_2 respectively, we then have

$$
\|\delta_n\|=o_P(p_n^{5/2}n^{-1}(n^{-1/2}+c_n^2))+o_P(p_n^2n^{-1}c_n).
\tag{A.19}
$$

We are now in the position to obtain the asymptotic representation of $\hat{\mathcal{R}}(\boldsymbol{\beta}_n)$. From (A.9), we obtain that

$$
\lambda=S_n^{-1}\bar{\hat{\eta}}(\boldsymbol{\beta}_n)+S_n^{-1}\delta_n.
\tag{A.20}
$$

Taylor expansion implies

$$
\log(1+W_i)=W_i-W_i^2/2+W_i^3/3(1+\xi_i)^4,
$$

for some ξ_i such that $|\xi_i|\leq|W_i|$. Therefore, combining (A.20) and some elementary calculations, we have

$$
\begin{aligned}
\hat{\mathcal{R}}(\boldsymbol{\beta}_n) &= 2\sum_{i=1}^{n}\log(1+W_i) \\
&= n\bar{\hat{\eta}}^T(\boldsymbol{\beta}_n)S_n^{-1}\bar{\hat{\eta}}(\boldsymbol{\beta}_n)-n\delta_n^TS_n^{-1}\delta_n+\frac{2}{3}\mathcal{R}_n\{1+o_P(1)\} \\
&= n\bar{\hat{\eta}}^T(\boldsymbol{\beta}_n)\Sigma^{-1}\bar{\hat{\eta}}(\boldsymbol{\beta}_n)+n\bar{\hat{\eta}}^T(\boldsymbol{\beta}_n)(S_n^{-1}-\Sigma^{-1})\bar{\hat{\eta}}(\boldsymbol{\beta}_n)-n\delta_n^TS_n^{-1}\delta_n+\frac{2}{3}\Re_n\{1+o_P(1)\},
\end{aligned}
\tag{A.21}
$$

where $\Re_n=\sum_{i=1}^{n}\{\lambda^T\hat{\eta}_{ni}(\boldsymbol{\beta}_n)\}^3$. Together with (A.19) and Lemma B.4, we have

$$
|n\delta_n^TS_n^{-1}\delta_n|\leq n\|\delta_n\|^2/\gamma_1(S_n)=o_P(p_n^5n^{-1}(n^{-1}+c_n^4))+o_P(p_n^4n^{-1}c_n^2)=o_P(p_n^{1/2}).
\tag{A.22}
$$

Furthermore, from the proof of (A.12), we can immediately derive that

$$
|\Re_n|=o_P\left((p_nn^{-1})^{\frac{k-3}{k-2}}\right)o_P\left\{(p_n^{k/2}n^{-k/2}p_n^{k/2})^{\frac{1}{k-2}}\right\}=o_P(1),
\tag{A.23}
$$

provided that $k\geq 4$. The proof of Theorem 2 is concluded from Lemmas B.5 and B.6 in Appendix B, together with expressions (A.21)–(A.23). □

Proof of Theorem 3. To prove the theorem, we first show that $\max_{1\le i\le n}\|\widetilde{\eta}_i(\theta)\| = o_P(n^{1/2})$. Note that $\widetilde{\mathbf{Z}}_{ni}^T\hat{\gamma}_n = \widetilde{\mathbf{Z}}_{i1}^T\hat{\theta} + \widetilde{\mathbf{Z}}_{ni2}^T\hat{\boldsymbol{\beta}}_{n(l)}$, where $\hat{\theta}$ is the subvector of the first l elements of $\hat{\gamma}_n$. By (3.12) and Lemma B.1, we can obtain that

$$\begin{aligned}\widetilde{\eta}_i(\theta) &= \left(\widetilde{\mathbf{Z}}_{i1} - \hat{\mu}_1^T(U_i)\mathbf{X}_i\right)\left(Y_{ni} - \mathbf{X}_i^T\hat{\alpha}(U_i;\theta,\hat{\boldsymbol{\beta}}_{n(l)}) - \widetilde{\mathbf{Z}}_{i1}^T\theta - \widetilde{\mathbf{Z}}_{ni2}^T\hat{\boldsymbol{\beta}}_{n(l)}\right)\\ &= \left(\widetilde{\mathbf{Z}}_{i1} - \mu_1^T(U_i)\mathbf{X}_i\{1+O_P(c_n)\}\right)\left(\varepsilon_i + \widetilde{\mathbf{Z}}_{ni}^T(\gamma_n-\hat{\gamma}_n) + \widetilde{\mathbf{Z}}_{i1}^T(\hat{\theta}-\theta) + \mathbf{X}_i^T\left(\alpha(U_i) - \hat{\alpha}(U_i;\theta,\hat{\boldsymbol{\beta}}_{n(l)})\right)\right)\\ &=: T_{i1}+T_{i2}+T_{i3}+T_{i4}. \end{aligned} \tag{A.24}$$

By (A.24), it is easy to show that

$$\max_{1\le i\le n}\|\widetilde{\eta}_i(\theta)\| \le \max_{1\le i\le n}\|T_{i1}\| + \max_{1\le i\le n}\|T_{i2}\| + \max_{1\le i\le n}\|T_{i3}\| + \max_{1\le i\le n}\|T_{i4}\|.$$

By condition (C7) and $E\|\mathbf{Z}_{i1}\|^k \le \infty$ for $k=4$, we have

$$\max_{1\le i\le n}\|T_{i1}\| \le \left(\max_{1\le i\le n}\|\widetilde{\mathbf{Z}}_{i1}\| + \max_{1\le i\le n}\|\mu_1^T(U_i)\mathbf{X}_i\| + O_P(c_n)\max_{1\le i\le n}\|\mathbf{X}_i\|\right)\max_{1\le i\le n}|\varepsilon_i| = o_P(n^{1/2}).$$

It is known from Theorem 1 in Lam and Fan (2008) that $\hat{\gamma}_n$ is a root-(n/p_n) consistent estimator of γ_n. Invoking the above argument and conditions (C7) and (C8), it can be shown that

$$\max_{1\le i\le n}\|T_{i2}\| = o_P(n^{1/2}), \qquad \max_{1\le i\le n}\|T_{i3}\| = o_P(n^{1/2}).$$

For $\max_{1\le i\le n}\|T_{i4}\|$. Recall the definition of $\hat{\alpha}(U_i;\theta,\hat{\boldsymbol{\beta}}_{n(l)})$ in Section 3.3, we have

$$\hat{\alpha}(U_i;\theta,\hat{\boldsymbol{\beta}}_{n(l)}) = (\mathbf{X}_i^T,\mathbf{0}_q)(D_{U_i}^TW_{U_i}D_{U_i})^{-1}D_{U_i}^TW_{U_i}\left(\widetilde{\mathbf{Z}}_{ni2}^T(\boldsymbol{\beta}_{n(l)} - \hat{\boldsymbol{\beta}}_{n(l)})\right) + \mathbf{X}_i^T\left(\alpha(U_i) - \hat{\alpha}(U_i,\gamma_n)\right).$$

By Lemmas B.1 and B.2, and again using the above argument, we can obtain that $\max_{1\le i\le n}\|T_{i4}\| = o_P(n^{1/2})$. Therefore, we have $\max_{1\le i\le n}\|\widetilde{\eta}_i(\theta)\| = o_P(n^{1/2})$.

For simplicity, we first introduce some notations. Let $\mu_1(u) = (E(\mathbf{X}\mathbf{X}^T|U=u))^{-1}E(\mathbf{X}\widetilde{\mathbf{Z}}_1^T|U=u)$ be a $q\times l$ matrix, $\mu_2(u) = (E(\mathbf{X}\mathbf{X}^T|U=u))^{-1}E(\mathbf{X}\widetilde{\mathbf{Z}}_{n2}^T|U=u)$ be a $q\times(p_n-l)$ matrix, and $\mu_n(u) = (E(\mathbf{X}\mathbf{X}^T|U=u))^{-1}E(\mathbf{X}\widetilde{\mathbf{Z}}_n^T|U=u)$ be $q\times p_n$ matrix.

Next we show that, as $n\to\infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\widetilde{\eta}_i(\theta) \xrightarrow{d} N(0,\Lambda(\theta)), \tag{A.25}$$

$$\frac{1}{n}\sum_{i=1}^n\widetilde{\eta}_i(\theta)\widetilde{\eta}_i^T(\theta) \xrightarrow{P} \Lambda(\theta), \tag{A.26}$$

where $\Lambda(\theta) = E[\xi_1\xi_1^T\varepsilon^2]$ is a positive defined matrix, and

$$\begin{aligned}\xi_i &= \left(\widetilde{\mathbf{Z}}_{i1} - \mu_1^T(U_i)\mathbf{X}_i\right) - E\left[(\widetilde{\mathbf{Z}}_1 - \mu_1^T(U)\mathbf{X})\widetilde{\mathbf{Z}}_n^T\right]\widetilde{\Psi}^{-1}\left(\widetilde{\mathbf{Z}}_{ni} - \mu_n^T(U_i)\mathbf{X}_i\right)\\ &\quad + E\left[(\widetilde{\mathbf{Z}}_1 - \mu_1^T(U)\mathbf{X})\widetilde{\mathbf{Z}}_1^T\right]\left(E\left[(\widetilde{\mathbf{Z}}_1 - \mu_1^T(U)\mathbf{X})(\widetilde{\mathbf{Z}}_1 - \mu_1^T(U)\mathbf{X})^T\right] - KP^{-1}K^T\right)^{-1}\\ &\quad \times\left\{\left(\widetilde{\mathbf{Z}}_{i1} - \mu_1^T(U_i)\mathbf{X}_i\right) - KP^{-1}(\widetilde{\mathbf{Z}}_{ni2} - \mu_2^T(U_i)\mathbf{X}_i)\right\},\end{aligned} \tag{A.27}$$

where $\widetilde{\Psi} = E\left[\left\{\widetilde{\mathbf{Z}}_n - \mu_n^T(U)\mathbf{X}\right\}\left\{\widetilde{\mathbf{Z}}_n - \mu_n^T(U)\mathbf{X}\right\}^T\right]$. From (3.12) and (A.24), we have

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^n\widetilde{\eta}_i(\theta) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\widetilde{\mathbf{Z}}_{i1} - \mu_1^T(U_i)\mathbf{X}_i\right)\varepsilon_i + \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\widetilde{\mathbf{Z}}_{i1} - \mu_1^T(U_i)\mathbf{X}_i\right)\widetilde{\mathbf{Z}}_{ni}^T(\gamma_n-\hat{\gamma}_n)\\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\widetilde{\mathbf{Z}}_{i1} - \mu_1^T(U_i)\mathbf{X}_i\right)\widetilde{\mathbf{Z}}_{i1}^T(\hat{\theta}-\theta) + \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\widetilde{\mathbf{Z}}_{i1} - \mu_1^T(U_i)\mathbf{X}_i\right)\mathbf{X}_i^T\left(\alpha(U_i) - \hat{\alpha}(U_i;\theta,\hat{\boldsymbol{\beta}}_{n(l)})\right)\\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\mu_1^T(U_i) - \hat{\mu}_1^T(U_i)\right)\mathbf{X}_i\varepsilon_i + \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\mu_1^T(U_i) - \hat{\mu}_1^T(U_i)\right)\mathbf{X}_i\widetilde{\mathbf{Z}}_{ni}^T(\gamma_n-\hat{\gamma}_n)\\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\mu_1^T(U_i) - \hat{\mu}_1^T(U_i)\right)\mathbf{X}_i\widetilde{\mathbf{Z}}_{i1}^T(\hat{\theta}-\theta)\end{aligned}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mu_1^T(U_i)-\hat{\mu}_1^T(U_i)\right)\mathbf{X}_i\mathbf{X}_i^T\left(\alpha(U_i)-\hat{\alpha}(U_i;\theta,\hat{\boldsymbol{\beta}}_{n(l)})\right)$$
$$=: T_1+T_2+T_3+T_4+T_5+T_6+T_7+T_8.$$

Note that the dimensionality of $\widetilde{\mathbf{Z}}_{i1}$ is l, which is a fixed constant and free of n. By Lemmas B.1 and B.2, and using the same argument similar to the proof of Lemma B.3, we can derive that $\left\|\sum_{k=4}^{8}T_k\right\| = O_P(n^{1/2}c_n^2)$, which implies that $T_k = o_P(1), k = 4,\ldots,8$. To prove (A.25), we consider T_k, $k = 1, 2, 3$, respectively. For T_2, note that $\hat{\gamma}_n = \{\widetilde{\mathbf{Z}}_n^{*T}(I-\mathbf{S})^T(I-\mathbf{S})\widetilde{\mathbf{Z}}_n^{*}\}^{-1}\widetilde{\mathbf{Z}}_n^{*T}(I-\mathbf{S})^T(I-\mathbf{S})Y_n$ is the profile least squares estimator of γ_n, where $\widetilde{\mathbf{Z}}_n^{*}=(\widetilde{\mathbf{Z}}_{n1},\ldots,\widetilde{\mathbf{Z}}_{nn})^T$. Similar to the proof of Theorem 4.1 in [10], we have

$$T_2 = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\widetilde{\mathbf{Z}}_{i1}-\mu_1^T(U_i)\mathbf{X}_i\right)\widetilde{\mathbf{Z}}_{ni}^T\{\widetilde{\mathbf{Z}}_n^{*T}(I-\mathbf{S})^T(I-\mathbf{S})\widetilde{\mathbf{Z}}_n^{*}\}^{-1}\widetilde{\mathbf{Z}}_n^{*T}(I-\mathbf{S})^T(I-\mathbf{S})(M+\varepsilon)$$
$$= -E\left[(\widetilde{\mathbf{Z}}_1-\mu_1^T(U)\mathbf{X})\widetilde{\mathbf{Z}}_n^T\right]\widetilde{\Psi}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\widetilde{\mathbf{Z}}_{ni}-\mu_n^T(U_i)\mathbf{X}_i\right)\varepsilon_i+o_P(1),$$

where $M=(\mathbf{X}_1^T\alpha(U_1),\ldots,\mathbf{X}_n^T\alpha(U_n))^T$, $\varepsilon=(\varepsilon_1,\ldots,\varepsilon_n)^T$, and $\widetilde{\Psi}$ is defined in (A.27).

For T_3, we first introduce some notations, let $K_n = \frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\mathbf{Z}}_{i1}-\sum_{k=1}^{n}S_{ik}\widetilde{\mathbf{Z}}_{k1}\right)\left(\widetilde{\mathbf{Z}}_{i1}-\sum_{k=1}^{n}S_{ik}\widetilde{\mathbf{Z}}_{k1}\right)^T$, $P_n = \frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\mathbf{Z}}_{ni2}-\sum_{k=1}^{n}S_{ik}\widetilde{\mathbf{Z}}_{nk2}\right)\left(\widetilde{\mathbf{Z}}_{ni2}-\sum_{k=1}^{n}S_{ik}\widetilde{\mathbf{Z}}_{nk2}\right)^T$, and $K = E\left[\left(\widetilde{\mathbf{Z}}_1-\mu_1^T(U)\mathbf{X}\right)\left(\widetilde{\mathbf{Z}}_{n2}-\mu_2(U)\mathbf{X}\right)^T\right]$, $P = E\left[\left(\widetilde{\mathbf{Z}}_{n2}-\mu_2^T(U)\mathbf{X}\right)\left(\widetilde{\mathbf{Z}}_{n2}-\mu_2^T(U)\mathbf{X}\right)^T\right]$. By Lemmas B.1 and B.2, it is to show that $K_n = K+o_P(1)$ and $P_n = P+o_P(1)$. Note that $\hat{\theta}$ is the subvector of the first l elements of the profile least squares estimator $\hat{\gamma}_n$, and similar to the argument of T_2 and the proof of Theorem 3.1 in [24], we have

$$T_3 = E\left[(\widetilde{\mathbf{Z}}_1-\mu_1^T(U)\mathbf{X})\widetilde{\mathbf{Z}}_1^T\right]\left(E\left[(\widetilde{\mathbf{Z}}_1-\mu_1^T(U)\mathbf{X})(\widetilde{\mathbf{Z}}_1-\mu_1^T(U)\mathbf{X})^T\right]-KP^{-1}K^T\right)^{-1}$$
$$\times\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\left(\widetilde{\mathbf{Z}}_{i1}-\mu_1^T(U_i)\mathbf{X}_i\right)-KP^{-1}(\widetilde{\mathbf{Z}}_{ni2}-\mu_2^T(U_i)\mathbf{X}_i)\right\}\varepsilon_i+o_P(1).$$

Therefore, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{\eta}_i(\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\varepsilon_i+o_P(1), \tag{A.28}$$

where ξ_i is defined by (A.27). Then, (A.25) follows directly from the proof of Theorem 3.1 in [24]. In addition, we also show that (A.26) holds by using the similar argument.

Applying the Taylor expansion to (3.13), and note that $\max_{1\le i\le n}\|\widetilde{\eta}_i(\theta)\| = o_P(n^{1/2})$, we can obtain that

$$\hat{\mathcal{R}}_l(\theta)=2\sum_{i=1}^{n}\left[\kappa^T\widetilde{\eta}_i(\theta)-\frac{1}{2}\{\kappa^T\widetilde{\eta}_i(\theta)\}^2\right]+o_P(1). \tag{A.29}$$

Invoking the proof of Theorem 1 in [28], we have

$$\sum_{i=1}^{n}[\kappa^T\widetilde{\eta}_i(\theta)]^2=\sum_{i=1}^{n}\kappa^T\widetilde{\eta}_i(\theta)+o_P(1), \tag{A.30}$$

$$\kappa=\left[\sum_{i=1}^{n}\widetilde{\eta}_i(\theta)\widetilde{\eta}_i^T(\theta)\right]^{-1}\sum_{i=1}^{n}\widetilde{\eta}_i(\theta)+o_P(n^{-1/2}). \tag{A.31}$$

From (A.29)–(A.31), we have

$$\hat{\mathcal{R}}_l(\theta)=\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{\eta}_i(\theta)\right]^T\left[\frac{1}{n}\sum_{i=1}^{n}\widetilde{\eta}_i(\theta)\widetilde{\eta}_i^T(\theta)\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{\eta}_i(\theta)\right]+o_P(1).$$

Together with (A.25) and (A.26), it is easy show that $\hat{\mathcal{R}}_l(\theta)$ is asymptotically chi-squared with l degrees of freedom. The proof is completed. □

The partially linear model is a special case of the varying-coefficient partially linear model. We can prove Theorems 4 and 5 by using the same arguments in the proofs of Theorems 2 and 3. We here omit the details.

Appendix B. Some lemmas

For the sake of convenience, let $c(0 < c < \infty)$ denote a constant not depending on n, but taking difference value at each appearance. The following notations will be used in the proof of lemmas. Let $\mu_i = \int u^i K(u)\mathrm{d}u,\ \nu_i = \int u^i K^2(u)\mathrm{d}u,\ \Gamma(u) = E(\mathbf{X}\mathbf{X}^T|U = u)$ and $\Phi(u) = E(\mathbf{X}\mathbf{Z}_n^T|U = u)$.

Lemma B.1. *Suppose that conditions* (C1)–(C5) *hold. If* $h \to 0$ *and* $nh \to \infty$ *as* $n \to \infty$, *then letting* $c_n = \left\{\frac{\log n}{nh}\right\}^{1/2} + h^2$ *and* $d_n = \left(\frac{\log n}{nh}\right)^{1/2}$,

$$\sup_{u\in\Omega} \frac{1}{n}\sum_{i=1}^{n} K_h(U_i - u)\left(\frac{U_i - u}{h}\right)^l X_{ij}\varepsilon_i = O_P(d_n),$$

$$\sup_{u\in\Omega} \left|\frac{1}{n}\sum_{i=1}^{n} K_h(U_i - u)\left(\frac{U_i - u}{h}\right)^l X_{ij_1}X_{ij_2} - f(u)\mu_l\Gamma_{j_1j_2}(u)\right| = O_P(c_n),$$

$$\sup_{u\in\Omega} \left|\frac{1}{n}\sum_{i=1}^{n} K_h(U_i - u)\left(\frac{U_i - u}{h}\right)^l X_{ij}Z_{nik} - f(u)\Phi_{jk}(u)\right| = O_P(c_n),$$

where $j_1, j_2, j = 1, \ldots, q,\ k = 1, \ldots, p_n,\ l = 0, 1, 2, 4,\ \Gamma_{j_1j_2}(u)$ *is the* (j_1, j_2)*th element of* $\Gamma(u)$ *and* $\Phi_{jk}(u)$ *is the* (j, k)*th element of* $\Phi(u)$.

The proof of Lemma B.1 *is similar to that of Lemma* A.2 *in* [35], *we then omit the details here.*

For any given parametric component $\boldsymbol{\beta}_n$, *the following lemma provides the consistency rate of the estimators of nonparametric functions. Let* $\alpha_j(u)$ *denote the jth component of* $\alpha(u), j = 1, \ldots, q$.

Lemma B.2. *Under the conditions of Lemma* B.1, *we have,*

$$\|\hat{\alpha}(u, \boldsymbol{\beta}_n) - \alpha(u)\| = O_P(c_n) \tag{B.1}$$

and

$$\max_{1\le j\le q}\sup_{u\in\Omega} |\hat{\alpha}_j(u, \boldsymbol{\beta}_n) - \alpha_j(u)| = O_P(c_n) \tag{B.2}$$

holds uniformly in $u \in \Omega$, *the support of* U.

Proof. We first present the proof of Eq. (B.1). Let

$$S_{n,l} = \sum_{i=1}^{n} K_h(U_i - u)\mathbf{X}_i\mathbf{X}_i^T\left(\frac{U_i - u}{h}\right)^l, \quad l = 0, 1, 2.$$

Note that

$$\mathbf{D}_u^T\mathbf{W}_u\mathbf{D}_u = \begin{pmatrix} S_{n,0} & S_{n,1} \\ S_{n,1} & S_{n,2}\end{pmatrix}.$$

Each element of the above matrix is in the form of a kernel regression. By Lemma B.1 and some elementary calculations, we have

$$S_{n,l} = nf(u)\mu_l\Gamma(u)(1 + O_P(c_n)) \tag{B.3}$$

holds uniformly in $u \in \Omega$. By (2.6), we have

$$\begin{aligned}\hat{\alpha}(u, \boldsymbol{\beta}_n) &= [nf(u)\Gamma(u)]^{-1}\sum_{i=1}^{n} K_h(U_i - u)\mathbf{X}_i(Y_{ni} - \mathbf{Z}_{ni}^T\boldsymbol{\beta}_n) + O_P(c_n)\\ &= [nf(u)\Gamma(u)]^{-1}\sum_{i=1}^{n} K_h(U_i - u)\mathbf{X}_i\left\{\mathbf{X}_i^T\alpha(U_i) + \varepsilon_i\right\} + O_P(c_n).\end{aligned} \tag{B.4}$$

Applying Lemma B.1, similar to the calculation of (B.3), we can easily show that

$$\frac{1}{n}\sum_{i=1}^{n} K_h(U_i - u)\mathbf{X}_i\mathbf{X}_i^T\alpha(U_i) = f(u)\Gamma(u)\alpha(u)\{1 + O_P(c_n)\} \tag{B.5}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} K_h(U_i - u)\mathbf{X}_i\varepsilon_i = o_P(1) \tag{B.6}$$

holds uniformly in $u \in \Omega$. From (B.4)–(B.6), $\hat{\alpha}(u, \boldsymbol{\beta}_n) = \alpha(u) + O_P(c_n)$ holds uniformly in $u \in \Omega$. This completes the proof of Eq. (B.1).

To prove (B.2), similar to Xia and Li [35], we further decompose $\hat{\alpha}_j(u, \boldsymbol{\beta}_n), j = 1, \ldots, q$. Here we only consider $\hat{\alpha}_1(u, \boldsymbol{\beta}_n)$ without loss of generality. For convenience, let $K_{ih}(u) = K_h(U_i - u), S_i = (X_{i2}, \ldots, X_{iq}), T_i = (X_{i1}, \ldots, X_{iq})$. Without confusion, we let $V_i = (S_i, (U_i - u)T_i)$ although it relates to u. Following Lemma 3 of [19], we have

$$\begin{aligned}\hat{\alpha}_1(u, \boldsymbol{\beta}_n) &= \alpha_1(u) + \frac{\sum_{i=1}^{n} K_{ih}(u)(X_{i1} - J_nH_n^{-1}V_i^T)\mathbf{X}_i^T(\alpha(U_i) - \alpha(u) - \alpha'(u)(U_i - u))}{\sum_{i=1}^{n} K_{ih}(u)(X_{i1} - J_nH_n^{-1}V_i^T)^2}\\ &\quad + \frac{\sum_{i=1}^{n} K_{ih}(u)(X_{i1} - J_nH_n^{-1}V_i^T)\varepsilon_i}{\sum_{i=1}^{n} K_{ih}(u)(X_{i1} - J_nH_n^{-1}V_i^T)^2}\\ &=: \alpha_1(u) + I_1 + I_2,\end{aligned} \tag{B.7}$$

where

$$\begin{aligned} H_n &= \sum_{i=1}^{n} K_{ih}(u)V_i^TV_i = \begin{pmatrix} \sum_{i=1}^{n} K_{ih}(u)S_i^TS_i & h\sum_{i=1}^{n} K_{ih}(u)\left(\frac{U_i-u}{h}\right)S_i^TT_i \\ h\sum_{i=1}^{n} K_{ih}(u)\left(\frac{U_i-u}{h}\right)T_i^TS_i & h^2\sum_{i=1}^{n} K_{ih}(u)\left(\frac{U_i-u}{h}\right)^2 T_i^TT_i \end{pmatrix}\\ &=: \begin{pmatrix} P_n & hR_n \\ hR_n^T & h^2Q_n \end{pmatrix},\\ J_n &= \sum_{i=1}^{n} K_{ih}(u)X_{i1}V_i = \left(\sum_{i=1}^{n} K_{ih}(u)X_{i1}S_i,\ h\sum_{i=1}^{n} K_{ih}(u)\left(\frac{U_i-u}{h}\right)X_{i1}T_i\right)\\ &=: (A_n, \quad hB_n). \end{aligned}$$

Let $A(u) = (\Gamma_{12}(u), \Gamma_{13}(u), \ldots, \Gamma_{1q}(u)), P(u) = (\Gamma_{ij}(u))_{i,j=2,\ldots,q}$ and $Q(u) = (\Gamma_{ij}(u))_{i,j=1,\ldots,q}$. From Lemma B.1 and condition (C4), we can easily show that

$$\begin{aligned} &\frac{1}{n}A_n = f(u)A(u) + O_P(c_n), \qquad \frac{1}{n}B_n = O_P(c_n)\mathbf{1}_q^T, \qquad \frac{1}{n}R_n = O_P(c_n)\mathbf{1}_{q-1}\mathbf{1}_q^T,\\ &\frac{1}{n}Q_n = f(u)\mu_2 Q(u) + O_P(c_n), \qquad \frac{1}{n}P_n = f(u)P(u) + O_P(c_n). \end{aligned} \tag{B.8}$$

Here $\mathbf{1}_q$ is the $q \times 1$ vector with 1 as all the elements. It can be seen that P_n is a symmetric matrix and its inverse exists, then

$$H_n^{-1} = \begin{pmatrix} P_n^{-1} + h^2P_n^{-1}R_n\mathcal{K}_n^{-1}R_n^TP_n^{-1} & -hP_n^{-1}R_n\mathcal{K}_n^{-1} \\ -h\mathcal{K}_n^{-1}R_n^TP_n^{-1} & \mathcal{K}_n^{-1} \end{pmatrix},$$

where $\mathcal{K}_n = h^2(Q_n - R_n^TP_n^{-1}R_n)$, and

$$\begin{aligned} J_nH_n^{-1} &= (A_nP_n^{-1} + h^2A_nP_n^{-1}R_n\mathcal{K}_n^{-1}R_n^TP_n^{-1} - h^2B_n\mathcal{K}_n^{-1}R_n^TP_n^{-1}, -hA_nP_n^{-1}R_n\mathcal{K}_n^{-1} + hB_n\mathcal{K}_n^{-1}),\\ J_nH_n^{-1}V_i^T &= A_nP_n^{-1}S_i^T + h^2A_nP_n^{-1}R_n\mathcal{K}_n^{-1}R_n^TP_n^{-1}S_i^T - h^2B_n\mathcal{K}_n^{-1}R_n^TP_n^{-1}S_i^T\\ &\quad - h(U_i - u)(A_nP_n^{-1}R_n\mathcal{K}_n^{-1}T_i^T - hB_n\mathcal{K}_n^{-1}T_i^T). \end{aligned}$$

From (B.8), we have

$$J_nH_n^{-1} = (A(u)(P(u))^{-1} + O_P(c_n),\ O_P(c_n)\mathbf{1}_q^T), \tag{B.9}$$

$$J_nH_n^{-1}J_n^T = nA(u)(P(u))^{-1}A^T(u)f(u) + O_P(c_n). \tag{B.10}$$

To deal with I_i for $i = 1, 2$, we consider their denominator first. Note that

$$\frac{1}{n}\sum_{i=1}^{n} K_{ih}(u)(X_{i1} - J_n H_n^{-1} V_i^T)^2 = \frac{1}{n}\sum_{i=1}^{n} K_{ih}(u) X_{i1}^2 - \frac{1}{n} J_n H_n^{-1} J_n^T$$
$$= \Gamma_{11}(u) f(u) - A(u)(P(u))^{-1} A^T(u) f(u) + O_P(c_n)$$
$$= f(u)\det(Q(u))/\det(P(u)) + O_P(c_n) \tag{B.11}$$

holds uniformly in $u \in \Omega$. Now we are in the position to handle I_1. Using the Taylor expansion, $\alpha_j(U_i) - \alpha_j(u) - \alpha_j'(u)(U_i - u) = \frac{1}{2}\alpha_j''(u^*)(U_i - u)^2, j = 1, \ldots, q$, where u^* is a point between U_i and u. By Cauchy–Schwarz inequality, Lemma B.1 and condition (C3), uniformly over $1 \le j \le q$, we have

$$\left| \frac{1}{n}\sum_{i=1}^{n} K_{ih}(u)(X_{i1} - J_n H_n^{-1} V_i^T) X_{ij}^T (\alpha_j(U_i) - \alpha_j(u)) \right|$$
$$\le \left\{ \frac{1}{n}\sum_{i=1}^{n} K_{ih}(u)(X_{i1} - J_n H_n^{-1} V_i^T)^2 \frac{1}{n}\sum_{i=1}^{n} K_{ih}(u) X_{ij}^2 (\alpha_j(U_i) - \alpha_j(u))^2 \right\}^{1/2}$$
$$= \left\{ \frac{1}{n}\sum_{i=1}^{n} K_{ih}(u)(X_{i1} - J_n H_n^{-1} V_i^T)^2 \frac{1}{4n}\sum_{i=1}^{n} K_{ih}(u) X_{ij}^2 \alpha_j''^2(u^*)(U_i - u)^4 \right\}^{1/2}$$
$$= c\{[f(u)\det(Q(u))/\det(P(u)) + O_P(c_n)] \cdot h^4[\mu_4 f(u)\Gamma_{11}(u) + O_P(c_n)]\}^{1/2}$$
$$= O_P(h^2). \tag{B.12}$$

From (B.11) and (B.12), we have $|I_1| = O_P(h^2)$. For I_2, we again apply Lemma B.1 to obtain

$$\frac{1}{n}\sum_{i=1}^{n} K_{ih}(u) T_i^T \varepsilon_i = O_P(d_n), \qquad \frac{1}{n}\sum_{i=1}^{n} K_{ih}(u)\left(\frac{U_i - u}{h}\right) T_i^T \varepsilon_i = O_P(d_n),$$

where d_n is defined in Lemma B.1. Therefore,

$$\frac{1}{n}\sum_{i=1}^{n} K_{ih}(u)(X_{i1} - J_n H_n^{-1} V_i^T)\varepsilon_i = (1, -J_n H_n^{-1})\left(\frac{1}{n}\sum_{i=1}^{n} K_{ih}(u) T_i \varepsilon_i, \frac{1}{n}\sum_{i=1}^{n} K_{ih}(u)(U_i - u) T_i \varepsilon_i\right)^T$$
$$= (1, -A(u)(P(u))^{-1})\frac{1}{n}\sum_{i=1}^{n} K_{ih}(u) T_i^T \varepsilon_i + O_P(h c_n d_n). \tag{B.13}$$

Combining (B.11) with (B.13) and invoking Lemma B.1 again, we have $|I_2| = O_P(d_n)$. Thus, this completes the proof of (B.2). □

The following lemma plays an important role in the proof of the other lemmas and theorems.

Lemma B.3. *Under the conditions of Lemma B.2, we have*

$$\left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_i \right\| = O_P(n^{1/2} p_n^{1/2} c_n^2), \tag{B.14}$$

where $R_i = \sum_{k=1}^{3} M_{i,k}$ can be found in (A.1).

Proof. Note that

$$\left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_i \right\| \le \left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_{i,1} \right\| + \left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_{i,2} \right\| + \left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_{i,3} \right\|. \tag{B.15}$$

Thus, to prove Lemma B.3, we only need to deal with the three sums about $M_{i,l} l = 1, 2, 3$, which are the p_n-dimensional column vectors, respectively. We first consider $\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} M_{i,1} \right\|$. For $k = 1, \ldots, p_n$ and $r = 1, \ldots, q$, we let

$$A_{kr}(U, \mathbf{X}, \mathbf{Z}_n) = (Z_{nk} - \mu_k^T(U)\mathbf{X}) X_r,$$

where $\mu_k(U)$ denotes the k-th column vector of $\mu(U)$. Note that $E[(\mathbf{Z}_n - \mu^T(U)\mathbf{X})\mathbf{X}^T|U] = 0$. Thus, we can easily see that $E(A_{kr}(U, \mathbf{X}, \mathbf{Z}_n)) = 0$ for each $k = 1, \ldots, p_n$ and $r = 1, \ldots, q$. Then

$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} M_{i,1} \right\| = \left(\sum_{k=1}^{p_n} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} M_{i,1}^{(k)} \right)^2 \right)^{1/2} = \left(\sum_{k=1}^{p_n} \frac{1}{n} \sum_{i,j=1}^{n} M_{i,1}^{(k)} M_{j,1}^{(k)} \right)^{1/2} \tag{B.16}$$

where

$$M_{i,1}^{(k)} = \sum_{r=1}^{q} A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni})(\hat{\alpha}_r(U_i, \boldsymbol{\beta}_n) - \alpha_r(U_i)). \tag{B.17}$$

Since q is a fixed constant, we then without loss of generality consider the sum over the term $A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni})(\hat{\alpha}_r(U_i, \boldsymbol{\beta}_n) - \alpha_r(U_i))$ of $M_{i,1}^{(k)}$ for $r = 1, \ldots, q$. If such sums are of the desired rate, we can then arrive the desired result. For $r = 1$, consider $\hat{\alpha}_1(u, \boldsymbol{\beta}_n)$. From (B.7) and (B.9), we have

$$\begin{aligned} \hat{\alpha}_1(U_i, \boldsymbol{\beta}_n) - \alpha_1(U_i) &= \hat{\mathcal{D}}_{n,l_1}^{-1}(U_i) \frac{1}{n} \sum_{l_1=1}^{n} K_{l_1 h}(U_i)(X_{l_1 1} - J_n H_n^{-1} V_{l_1}) g(\mathbf{X}_{l_1}, U_{l_1}, U_i, \varepsilon_{l_1}) \\ &= \mathcal{D}^{-1}(U_i) L_{nl_1}(U_i) + \mathcal{D}^{-1}(U_i) \frac{1}{n} \sum_{l_1=1}^{n} K_{l_1 h}(U_i) \zeta_{l_1}(\mathbf{X}_i) \varepsilon_{l_1} \\ &\quad + (\hat{\mathcal{D}}_{n,l_1}^{-1}(U_i) - \mathcal{D}^{-1}(U_i)) \frac{1}{n} \sum_{l_1=1}^{n} K_{l_1 h}(U_i)(X_{l_1 1} - J_n H_n^{-1} V_{l_1}) g(\mathbf{X}_{l_1}, U_{l_1}, U_i, \varepsilon_{l_1}), \end{aligned} \tag{B.18}$$

where

$$\begin{aligned} & g(\mathbf{X}_{l_1}, U_{l_1}, U_i, \varepsilon_{l_1}) = \mathbf{X}_{l_1}^T \left(\alpha(U_{l_1}) - \alpha(U_i) - \alpha'(U_i)(U_{l_1} - U_i) \right) + \varepsilon_{l_1}, \\ & \hat{\mathcal{D}}_{n,l_1}^{-1}(U_i) = \frac{1}{n} \sum_{l_1=1}^{n} K_{l_1 h}(U_i)(X_{l_1 1} - J_n H_n^{-1} V_{l_1})^2, \\ & \zeta_{l_1}(\mathbf{X}_i) = X_{l_1 1} - A(U_i)(P(U_i))^{-1} S_{l_1}^T, \\ & L_{nl_1}(U_i) = \frac{1}{n} \sum_{l_1=1}^{n} K_{l_1 h}(U_i) \zeta_{l_1}(\mathbf{X}_i) \mathbf{X}_{l_1}^T (\alpha(U_{l_1}) - \alpha(U_i) - \alpha'(U_i)(U_{l_1} - U_i)), \end{aligned}$$

and $\mathcal{D}(U_i) = f(U_i) \det(Q(U_i))/\det(P(U_i))$. From (B.11)–(B.13) and the proof of Theorem 4.1 in [35], we have

$$\frac{1}{n} \sum_{l_1=1}^{n} K_{l_1 h}(U_i)(X_{l_1 1} - J_n H_n^{-1} V_{l_1}) g(\mathbf{X}_{l_1}, U_{l_1}, U_i, \varepsilon_{l_1}) = O_P(c_n), \tag{B.19}$$

$$\hat{\mathcal{D}}_{n,l_1}^{-1}(U_i) - \mathcal{D}^{-1}(U_i) = O_P(c_n). \tag{B.20}$$

Furthermore, by Lemma B.1 and the Taylor expansion, following the expression of (4.1) in [35], together with (B.18)–(B.20), we further have

$$\begin{aligned} \hat{\alpha}_1(U_i, \boldsymbol{\beta}_n) - \alpha_1(U_i) &= \mathcal{D}^{-1}(U_i) L_{nl_1}(U_i) + \mathcal{D}^{-1}(U_i) \frac{1}{n} \sum_{l_1=1}^{n} K_{l_1 h}(U_i) \zeta_{l_1}(\mathbf{X}_i) \varepsilon_{l_1} \\ &= B_1(U_i) h^2 + C_1(U_i) \frac{1}{n} \sum_{l_1=1}^{n} K_{l_1 h}(U_i) T_{l_1}^T \varepsilon_{l_1} + O_P(c_n^2), \end{aligned}$$

where $B_1(U_i)$ is a function of U_i, $C_1(U_i)$ is a $1 \times q$ vector function of U_i. Similarly, we have,

$$\hat{\alpha}_r(U_i, \boldsymbol{\beta}_n) - \alpha_r(U_i) = B_r(U_i) h^2 + C_r(U_i) \frac{1}{n} \sum_{l_1=1}^{n} K_{l_1 h}(U_i) T_{l_1}^T \varepsilon_{l_1} + O_P(c_n^2). \tag{B.21}$$

Thus, $M_{i,1}^{(k)}$ can have the following asymptotic expression:

$$M_{i,1}^{(k)} = \sum_{r=1}^{q} A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni})(\hat{\alpha}_r(U_i, \boldsymbol{\beta}_n) - \alpha_r(U_i))$$
$$= \sum_{r=1}^{q} A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni}) \left(B_r(U_i)h^2 + C_r(U_i)\frac{1}{n}\sum_{l_1=1}^{n} K_{l_1 h}(U_i)T_{l_1}^T \varepsilon_{l_1} \right) + \sum_{r=1}^{q} A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni}) O_P(c_n^2).$$

Then, it is easy to see that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_{i,1}^{(k)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\sum_{r=1}^{q} A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni}) \left(B_r(U_i)h^2 + C_r(U_i)\frac{1}{n}\sum_{l_1=1}^{n} K_{l_1 h}(U_i)T_{l_1}^T \varepsilon_{l_1} \right)\right) + O_P(\sqrt{n}c_n^2)$$
$$= \sum_{r=1}^{q} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni}) \left(B_r(U_i)h^2 + C_r(U_i)\frac{1}{n}\sum_{l_1=1}^{n} K_{l_1 h}(U_i)T_{l_1}^T \varepsilon_{l_1} \right) + O_P(\sqrt{n}c_n^2)$$
$$=: \sum_{r=1}^{q} I_r + O_P(\sqrt{n}c_n^2). \tag{B.22}$$

To deal with the sum over I_r, we note that $A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni})B_r(U_i)$ are i.i.d. with mean zero because the conditional expectation of $A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni})B_r(U_i)$ given U_i is zero. Thus, the related sum is of the rate h^2. The sum of $A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni})C_r(U_i)\frac{1}{n}\sum_{l_1=1}^{n} K_{l_1 h}(U_i)T_{l_1}^T \varepsilon_{l_1}$ over i can be re-arranged to be, for any r with $1 \le r \le q$,

$$\frac{1}{\sqrt{n}}\sum_{l_1=1}^{n} K_{l_1 h}(U_i)T_{l_1}^T \varepsilon_{l_1} \frac{1}{n}\sum_{i=1}^{n} A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni})C_r(U_i). \tag{B.23}$$

We now compute its variance to get its convergence rate in probability. Invoking the independence among ε_{l_1} and the independence from the other variables, and the independence among $A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni})C_r(U_i)$, we can easily see that its variance equals

$$\frac{1}{n}\sum_{l_1=1}^{n} E\left([K_{l_1 h}(U_i)T_{l_1}^T \varepsilon_{l_1}]^2 \left[\frac{1}{n}\sum_{i=1}^{n} A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni})C_r(U_i)\right]^2 \right)$$
$$= \frac{1}{n}\sum_{l_1=1}^{n}\frac{1}{n^2}\sum_{i=1}^{n} E\left([K_{l_1 h}(U_i)T_{l_1}^T \varepsilon_{l_1}]^2 [A_{kr}(U_i, \mathbf{X}_i, \mathbf{Z}_{ni})C_r(U_i)]^2\right)$$
$$= O\left(\frac{1}{n}\right).$$

Altogether, $I_r = O_P\left(h^2 + \frac{1}{\sqrt{n}}\right)$. From this, together with (B.22), we have $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_{i,1}^{(k)} = O_P(\sqrt{n}c_n^2)$, and then $\left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_{i,1} \right\| = O_P(n^{1/2}p_n^{1/2}c_n^2)$.

Looking at the structure of $M_{i,2}$, we can see that it is very similar to that of $M_{i,1}$ and is even easier to handle because of the independence of ε_i from the other variables. Thus similar arguments for $M_{i,1}$ can obtain that

$$\left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_{i,2} \right\| = O_P(n^{1/2}p_n^{1/2}c_n^2).$$

Further, Lemma B.2 yields that any component of $M_{i,3}$ is of the order c_n^2, thus, $\left\| \frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_{i,3} \right\| = O_P(n^{1/2}p_n^{1/2}c_n^2)$. Thus, the proof of Lemma B.3 is completed. □

Lemma B.4. *Under regularity conditions* (C1)–(C8)*, when $\boldsymbol{\beta}_n$ is the true value of the parameter, we have*

$$\mathrm{tr}[(S_n - \Sigma)^2] = O_P(p_n^2(c_n^4 + 1/n)), \tag{B.24}$$

where $S_n = \frac{1}{n}\sum_{i=1}^{n} \hat{\eta}_{ni}(\boldsymbol{\beta}_n)\hat{\eta}_{ni}^T(\boldsymbol{\beta}_n)$, Σ is defined in condition (C6). Furthermore,

$$\max_{1\le k\le p_n} |\gamma_k(S_n) - \gamma_k(\Sigma)| = O_P(p_n(c_n^2 + n^{-1/2})). \tag{B.25}$$

Proof. Note that, recalling the definition of $\eta_{ni}(\boldsymbol{\beta}_n)$ right below (A.1), which is the independent copy of η_n in condition (C8),

$$S_n - \Sigma = \frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n) - \Sigma + \frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)R_i^T + \frac{1}{n}\sum_{i=1}^{n}R_i\eta_{ni}^T(\boldsymbol{\beta}_n) + \frac{1}{n}\sum_{i=1}^{n}R_iR_i^T.$$

Note every element of $R_iR_i^T$ and then $\frac{1}{n}\sum_{i=1}^{n}R_iR_i^T$ is of the rate c_n^2. We now deal with $\frac{1}{n}\sum_{i=1}^{n}R_i\eta_{ni}^T(\boldsymbol{\beta}_n)$. Note that $R_i = M_{i,1} + M_{i,2} + M_{i,3}$. We can then separately deal with $\frac{1}{n}\sum_{i=1}^{n}M_{i,j}\eta_{ni}^T(\boldsymbol{\beta}_n), j = 1, 2, 3$. We note that $M_{i,1}\eta_{ni}^T(\boldsymbol{\beta}_n)$ contains ε_i, and $M_{i,2}\eta_{ni}^T(\boldsymbol{\beta}_n)$ contains $(\mathbf{Z}_{ni} - \mu^T(U_i)\mathbf{X}_i)\mathbf{X}_i^T$, both are of zero means. Similar to the arguments used in proving the sums over $M_{i,1}$ and $M_{i,2}$ in Lemma B.3, we can obtain the rate c_n^2 for each element of these sums over $M_{i,1}\eta_{ni}^T(\boldsymbol{\beta}_n)$ and $M_{i,2}\eta_{ni}^T(\boldsymbol{\beta}_n)$. Also for $i = 3$, the sum over $M_{i,3}\eta_{ni}^T(\boldsymbol{\beta}_n)$ can be of the same rate. It is easy to see that

$$\mathrm{tr}[(S_n - \Sigma)^2] \le 2\mathrm{tr}\left\{\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n) - \Sigma\right\}^2 + 2\mathrm{tr}\left\{\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)R_i^T + R_i\eta_{ni}^T(\boldsymbol{\beta}_n) + R_iR_i^T\right\}^2.$$

Thus, $2\mathrm{tr}\left\{\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)R_i^T + R_i\eta_{ni}^T(\boldsymbol{\beta}_n) + R_iR_i^T\right\}^2$ is of the rate $O_P(p_n^2c_n^4)$. Consider the convergence rate of the first term. By the basic algebraic calculation, we have

$$\begin{aligned}\mathrm{tr}\left[\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n) - \Sigma\right]^2 &= \mathrm{tr}\left[\Sigma^2\left(\Sigma^{-1/2}\left\{\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n)\right\}\Sigma^{-1/2} - I_{p_n}\right)^2\right]\\ &\le \gamma_{p_n}^2(\Sigma)\mathrm{tr}(D_n^2),\end{aligned}$$

where $D_n = \Sigma^{-1/2}\left\{\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n)\right\}\Sigma^{-1/2} - I_{p_n}$. Since $\gamma_{p_n}(\Sigma)$ is bounded, then we only need to prove that $\mathrm{tr}(D_n^2) = O_P(p_n^2/n)$. From conditions (C7) and (C8), and noting the definition of η_n and $k \ge 4$ there, the Cauchy–Schwarz inequality yields

$$\begin{aligned}E(\|\eta_{n1}(\boldsymbol{\beta}_n)\|^4) &= p_n^2E\left(\left(1/p_n\sum_{j=1}^{p_n}|\eta_{n1j}|^2\right)^2\right)\\ &\le p_n^2E\left(\left(1/p_n\sum_{j=1}^{p_n}|\eta_{n1j}|^k\right)^{4/k}\right)\\ &\le p_n^2\left[E\left(1/p_n\sum_{j=1}^{p_n}|\eta_{n1j}|^k\right)\right]^{4/k} = O(p_n^2).\end{aligned}$$

Then

$$\begin{aligned}\mathrm{tr}(D_n^2) &= \mathrm{tr}\left[\Sigma^{-2}\left\{\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n)\right\}^2\right] - 2\mathrm{tr}\left[\Sigma^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n)\right\}\right] + p_n\\ &=: V_1 - 2V_2 + p_n,\end{aligned} \tag{B.26}$$

$$E(V_2) = \mathrm{tr}\left[E\left(\Sigma^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n)\right\}\right)\right] = \mathrm{tr}(I_{p_n}) = p_n, \tag{B.27}$$

and

$$\begin{aligned}E(V_1) &= \frac{1}{n^2}\mathrm{tr}\left[E\left(\Sigma^{-2}\left\{\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n)\|\eta_{ni}(\boldsymbol{\beta}_n)\|^2\right\}\right)\right]\\ &\quad + \frac{1}{n^2}\mathrm{tr}\left[\sum_{i\ne j}E\{\Sigma^{-1}\eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n)\}E\{\Sigma^{-1}\eta_{nj}(\boldsymbol{\beta}_n)\eta_{nj}^T(\boldsymbol{\beta}_n)\}\right]\\ &= \frac{1}{n^2}\mathrm{tr}\left[E\left(\Sigma^{-2}\left\{\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)\eta_{ni}^T(\boldsymbol{\beta}_n)\|\eta_{ni}(\boldsymbol{\beta}_n)\|^2\right\}\right)\right] + \frac{1}{n^2}\sum_{i\ne j}\mathrm{tr}(I_{p_n})\end{aligned}$$

$$\leq \frac{1}{n^2}\frac{1}{\gamma_1^2(\Sigma)}E\left\{\sum_{i=1}^{n}\|\eta_{ni}(\boldsymbol{\beta}_n)\|^4\right\}+\frac{n(n-1)}{n^2}p_n$$
$$= O(p_n^2/n)+p_n. \tag{B.28}$$

Thus $E\{\mathrm{tr}(D_n^2)\} = O(p_n^2/n)$ follows from (B.26)–(B.28), and then $\mathrm{tr}(D_n^2) = O_P(p_n^2/n)$. Together with this, we then obtain that $\mathrm{tr}[(S_n - \Sigma)^2] = O(p_n^2(c_n^4 + 1/n))$.

Prove (B.25). Note that

$$\begin{aligned}\|\gamma_k(S_n)-\gamma_k(\Sigma)\|^2 &= \|\gamma_k^{1/2}(S_n^2)-\gamma_k^{1/2}(\Sigma^2)\|^2\\ &\leq \sum_{k=1}^{p_n}\|\gamma_k^{1/2}(S_n^2)-\gamma_k^{1/2}(\Sigma^2)\|^2\\ &= \sum_{k=1}^{p_n}\gamma_k(S_n^2)+\sum_{k=1}^{p_n}\gamma_k(\Sigma^2)-2\sum_{k=1}^{p_n}\gamma_k^{1/2}(S_n^2)\gamma_k^{1/2}(\Sigma^2)\\ &= \mathrm{tr}(S_n^2)+\mathrm{tr}(\Sigma^2)-2\sum_{k=1}^{p_n}\gamma_k(S_n)\gamma_k(\Sigma).\end{aligned} \tag{B.29}$$

By Von Neumann's inequality [34], $\sum_{k=1}^{p_n}\gamma_k(S_n)\gamma_k(\Sigma) \geq \mathrm{tr}(S_n\Sigma)$. Hence

$$\max_{1\leq k\leq p_n}\|\gamma_k(S_n)-\gamma_k(\Sigma)\| \leq \sqrt{\mathrm{tr}((S_n-\Sigma)^2)}. \tag{B.30}$$

By (B.24) and (B.30), (B.25) is proved. □

This lemma implies that all the eigenvalues of S_n converge to those of Σ uniformly at the rate of $O_P(p_n(c_n^2 + 1/\sqrt{n}))$.

To prove Theorem 2, we need another two lemmas stated below.

Lemma B.5. *Under regularity conditions* (C1)–(C8), *and when* $p_n^{2+4/(k-2)}/n \to 0$, *then*

$$\frac{n\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{ni}^T(\boldsymbol{\beta}_n)\right)\Sigma^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{ni}(\boldsymbol{\beta}_n)\right)-p_n}{\sqrt{2p_n}} \xrightarrow{d} N(0,1).$$

Proof. By (A.1), a simple calculation yields that

$$n\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{ni}^T(\boldsymbol{\beta}_n)\right)\Sigma^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{ni}(\boldsymbol{\beta}_n)\right) =: K_1+K_2+2K_3,$$

where

$$K_1 = n\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}^T(\boldsymbol{\beta}_n)\right)\Sigma^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(\boldsymbol{\beta}_n)\right),$$
$$K_2 = n\left(\frac{1}{n}\sum_{i=1}^{n}R_i^T\right)\Sigma^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}R_i\right),$$
$$K_3 = n\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}^T(\boldsymbol{\beta}_n)\right)\Sigma^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}R_i\right).$$

From (B.14) and $\|\bar{\hat{\eta}}(\boldsymbol{\beta}_n)\| = O_P(\sqrt{p_n/n})$, we have $K_1 = O_P(p_n)$. By Lemma B.3, $K_2 \leq \frac{n}{\gamma_1(\Sigma)}\left\|\frac{1}{n}\sum_{i=1}^{n}R_i\right\|^2 = O_P(p_n n c_n^4) = o_P(\sqrt{p_n})$. Since $K_3 \leq K_1^{1/2}K_2^{1/2}$, we have $K_3 = o_P(\sqrt{p_n})$. Thus, Lemma B.5 can be proved by applying the martingale central limit theorem as given in [15] to $(K_1 - p_n)/\sqrt{2p_n}$. □

Lemma B.6. *Under regularity conditions* (C1)–(C8), *and when* $p_n^{3+2/(k-2)}/n \to 0$, *we have*

$$n\left\{\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{ni}^T(\boldsymbol{\beta}_n)\right\}(S_n^{-1}-\Sigma^{-1})\left\{\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{ni}(\boldsymbol{\beta}_n)\right\} = o_P(\sqrt{p_n}).$$

Proof. Let $\hat{D}_n = \Sigma^{-1/2} S_n \Sigma^{-1/2} - I_{p_n}$, similar arguments used in the proof of Lemma 6 in [6] yield

$$\begin{aligned} S_n^{-1} - \Sigma^{-1} &= \Sigma^{-1/2}(\Sigma^{1/2} S_n^{-1} \Sigma^{1/2} - I_{p_n})\Sigma^{-1/2} \\ &= \Sigma^{-1/2}[-\hat{D}_n + \hat{D}_n^2 + \hat{D}_n^2\{\Sigma^{1/2} S_n^{-1}\Sigma^{1/2} - I_{p_n}\}]\Sigma^{-1/2}. \end{aligned} \tag{B.31}$$

Note that

$$\begin{aligned} \mathrm{tr}((S_n - \Sigma)^2) &= \mathrm{tr}((\Sigma^{1/2}(\Sigma^{-1/2} S_n \Sigma^{-1/2} - I_{p_n})\Sigma^{1/2})^2) \\ &= \mathrm{tr}(\hat{D}_n \Sigma \hat{D}_n \Sigma) \geq \gamma_1^2(\Sigma)\mathrm{tr}(\hat{D}_n^2). \end{aligned}$$

By Lemma B.4, we have

$$\mathrm{tr}(\hat{D}_n^2) \leq \frac{1}{\gamma_1^2(\Sigma)}\mathrm{tr}((S_n - \Sigma)^2) = O_P(p_n^2(c_n^4 + 1/n)). \tag{B.32}$$

Again employing Lemma B.4 and the similar arguments used in the proof of Lemma 6 in [6], we have

$$\begin{aligned} \mathrm{tr}(S_n^{-1} - \Sigma^{-1})^2 &\leq 2\mathrm{tr}\{\Sigma^{-2}(-\hat{D}_n + \hat{D}_n^2)^2\} + 2\mathrm{tr}\{\hat{D}_n^4 (S_n^{-1} - \Sigma^{-1})^2\} \\ &\leq 2\mathrm{tr}\{\Sigma^{-2}(-\hat{D}_n + \hat{D}_n^2)^2\} + 2[\mathrm{tr}(\hat{D}_n^2)]^2 \mathrm{tr}\{(S_n^{-1} - \Sigma^{-1})^2\} \\ &= 2\mathrm{tr}\{\Sigma^{-2}(-\hat{D}_n + \hat{D}_n^2)^2\} + o_P(\mathrm{tr}\{(S_n^{-1} - \Sigma^{-1})^2\}) \\ &= o_P(p_n^2(c_n^4 + 1/n)). \end{aligned} \tag{B.33}$$

From the proof of Lemma B.5, we have that $\left\| \frac{1}{n}\sum_{i=1}^n \hat{\eta}_{ni}^T(\boldsymbol{\beta}_n) \right\| = O_P(\sqrt{p_n/n})$. Together with (B.33), $p_n^{3+2/(k-2)}/n \to 0$, and $c_n^2 = o(1/\sqrt{n})$ by condition (C5), we can obtain

$$\begin{aligned} & n\left\{\frac{1}{n}\sum_{i=1}^n \hat{\eta}_{ni}^T(\boldsymbol{\beta}_n)\right\}(S_n^{-1} - \Sigma^{-1})\left\{\frac{1}{n}\sum_{i=1}^n \hat{\eta}_{ni}(\boldsymbol{\beta}_n)\right\} \\ &\leq n\left\|\frac{1}{n}\sum_{i=1}^n \hat{\eta}_{ni}(\boldsymbol{\beta}_n)\right\|^2 \sqrt{\mathrm{tr}(S_n^{-1} - \Sigma^{-1})^2} \\ &= o_P(p_n^2(c_n^2 + 1/\sqrt{n})) = o_P(\sqrt{p_n}). \end{aligned}$$

The proof is finished. □

References

[1] I. Ahmad, S. Leelahanon, Q. Li, Efficient estimation of a semiparametric partially linear varying coefficient model, Ann. Statist. 33 (2005) 258–283.
[2] S.C. Albright, W.L. Winston, C.J. Zappe, Data Analysis and Decision Making with Microsoft Excel, Duxbury, Pacific Grove, CA, 1999.
[3] E. Candes, T. Tao, The Dantzig selector: statistical estimation when p is much large than n, Ann. Statist. 35 (2007) 2313–2351.
[4] R.J. Carroll, D. Ruppert, A.H. Welsh, Local estimating equations, J. Amer. Statist. Assoc. 93 (1998) 214–227.
[5] S.X. Chen, Empirical likelihood confidence intervals for nonparametric density estimation, Biometrika 83 (1996) 329–341.
[6] S.X. Chen, L. Peng, Y.L. Qin, Effects of data dimension on empirical likelihood, Biometrika 96 (2009) 711–722.
[7] T.J. DiCiccio, P. Hall, J.P. Romano, Bartlett adjustment for empirical likelihood, Ann. Statist. 19 (1991) 1053–1061.
[8] D.L. Donoho, High-dimensional data analysis: the curses and blessings of dimensionality, Lecture on August 8, 2000, to the American Mathematical Society on ''Math Challenges of the 21st Century'', 2000.
[9] J. Fan, I. Gijbels, Local Polynomial Modeling and its Applications, Chapman and Hall, London, 1996.
[10] J. Fan, T. Huang, Profile likelihood inferences on semiparametric varying-coefficient partially linear models, Bernoulli 11 (2005) 1031–1057.
[11] J. Fan, R. Li, Statistical challenges with high-dimensionality: feature selection in knowledge discovery, in: M. Sanz-Solé, J. Soria, J.L. Varona, J. Verdera (Eds.), Proceedings of International Congress of Mathematicians, vol. III, 2006, pp. 595–622.
[12] J. Fan, J. Lv, Sure independence screening for ultra-high dimensional feature space, J. R. Stat. Soc. Ser. B 70 (2008) 849–911. with discussion.
[13] J. Fan, H. Peng, Nonconcave penalized likelihood with a diverging number of parameters, Ann. Statist. 32 (2004) 928–961.
[14] J. Fan, H. Peng, T. Huang, Semilinear high-dimensional model for normalization of microarray data: a theoretical analysis and partial consistency (with discussion), J. Amer. Statist. Assoc. 100 (2005) 781–813.
[15] P.G. Hall, C.C. Hyde, Martingale Central Limit Theory and its Applications, Academic Press, New York, 1980.
[16] P. Hall, B. La Scala, Methodology and algorithms of empirical likelihood, Int. Statist. Rev. 58 (1990) 109–127.
[17] T.J. Hastie, R. Tibshirani, Varying-coefficient models, J. R. Stat. Soc. Ser. B 55 (1993) 757–796.
[18] N.L. Hjort, L.W. McKeague, I. Van Keilegom, Extending the scope of empirical likelihood, Ann. Statist. 37 (2009) 1079–1111.
[19] T.L. Lai, H. Robbins, C.Z. Wei, Strong consistency of least squares estimates in multiple regression II, J. Multivariate Anal. 9 (1979) 343–361.
[20] C. Lam, J. Fan, Profile-kernel likelihood inference with diverging number of parameters, Ann. Statist. 36 (2008) 2232–2260.
[21] Q. Li, C.J. Huang, D. Li, T.-T. Fu, Semiparametric smooth coefficient models, J. Bus. Econom. Statist. 20 (2002) 412–422.
[22] G.R. Li, H. Peng, L.X. Zhu, Nonconcave penalized M-estimation with a diverging number of parameters, Statist. Sinica 21 (2011) 391–419.
[23] G.R. Li, P. Tian, L.G. Xue, Generalized empirical likelihood inference in semiparametric regression model for longitudinal data, Acta Math. Sin. (Engl. Ser.) 24 (12) (2008) 2029–2040.
[24] G. Li, Q.H. Wang, Empirical likelihood regression analysis for right censored data, Statist. Sinica 13 (2003) 51–68.
[25] G.R. Li, L.X. Zhu, L.G. Xue, S.Y. Feng, Empirical likelihood inference in partially linear single-index models for longitudinal data, J. Multivariate Anal. 101 (2010) 718–732.

[26] Y.P. Mack, B.W. Silverman, Weak and strong uniform consistency of kernel regression estimates, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 61 (1982) 405–415.
[27] A. Owen, Empirical likelihood ratio confidence intervals for a single functional, Biometrika 75 (1988) 237–249.
[28] A. Owen, Empirical likelihood ratio confidence regions, Ann. Statist. 18 (1990) 90–120.
[29] A. Owen, Empirical Likelihood, Chapman and Hall, CRC, 2001.
[30] S. Portnoy, Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity, Ann. Statist. 16 (1988) 356–366.
[31] J. Qin, J. Lawless, Empirical likelihood and general estimating equations, Ann. Statist. 22 (1994) 300–325.
[32] T.A. Severini, J.G. Staniswalis, Quasi-likelihood estimation in semiparametric models, J. Amer. Statist. Assoc. 89 (1994) 501–511.
[33] T.A. Severini, W.H. Wong, Profile likelihood and conditionally parametric models, Ann. Statist. 20 (1992) 1768–1802.
[34] J. Von Neumann, Some matrix inequalities and metrization of matrix space, Tomsk Univ. Rev. 1 (1937) 286–300.
[35] Y. Xia, W.K. Li, On the estimation and testing of functional-coefficient linear models, Statist. Sinica 9 (1999) 737–757.
[36] Y. Xia, W. Zhang, H. Tong, Efficient estimation for semivarying-coefficient models, Biometrika 91 (2004) 661–681.
[37] L.G. Xue, L.X. Zhu, Empirical likelihood confidence regions of the parameters in a partially linear single-index model, Sci. China Ser. A 48 (2005) 1333–1348.
[38] L.G. Xue, L.X. Zhu, Empirical likelihood for single-index models, J. Multivariate Anal. 97 (2006) 1295–1312.
[39] L.G. Xue, L.X. Zhu, Empirical likelihood for a varying coefficient model with longitudinal data, J. Amer. Statist. Assoc. 102 (2007) 642–654.
[40] L.G. Xue, L.X. Zhu, Empirical Likelihood in Nonparametric and Semiparametric Models, Science Press, Beijing, 2010.
[41] J.H. You, Y. Zhou, Empirical likelihood for semiparametric varying-coefficient partially linear regression models, Statist. Probab. Lett. 76 (2006) 412–422.
[42] L.X. Zhu, L.G. Xue, Empirical likelihood confidence regions in a partially linear single-index model, J. R. Stat. Soc. Ser. B 68 (2006) 549–570.