Hacettepe Journal of Mathematics and Statistics
Volume 48 (1) (2019), 230 – 241

Calibration of the empirical likelihood for semiparametric varying-coefficient partially linear models with diverging number of parameters

Junwei He* and Xia Chen†‡

Abstract

This article is concerned with the calibration of the empirical likelihood for semiparametric varying-coefficient partially linear models with diverging number of parameters. However, there is always substantial lack-of-fit, when the empirical likelihood ratio is calibrated by a bias-corrected empirical likelihood, producing tests with type I errors much larger than nominal levels. So we consider an effective calibration method and study the asymptotic behavior of this bias-corrected empirical likelihood ratio function. Some simulation studies are conducted to illustrate our approach.

Keywords: Varying-coefficient partially linear models, Empirical likelihood, Bias correction, Asymptotic normality, Coverage accuracy.

2000 AMS Classification: 62G08, 62G15

Received : 15.06.2017 *Accepted* : 28.06.2018 *Doi* : 10.15672/HJMS.2018.604

1. Introduction

Consider the following semiparametric varying-coefficient partially linear models

$$Y = X^{\mathrm{T}}\alpha(U) + Z^{\mathrm{T}}\beta + \varepsilon \tag{1.1}$$

where $\alpha(\cdot) = (\alpha_1(\cdot), ..., \alpha_q(\cdot))^{\mathrm{T}}$ is a q-dimensional vector of unknown regression functions, $\beta = (\beta_1, ..., \beta_p)^{\mathrm{T}}$ is a p-dimensional of unknown regression coefficients, and ε is an independent random error with $E(\varepsilon|X, Z, U) = 0$ almost surely. Without loss of generality, we assume that the variable U is defined on the unit interval $[0, 1]$.

As the extension of the usual linear regression model and partially linear regression model, semiparametric varying-coefficient partially linear model (1.1) has attracted great

*School of Mathematics and Information Science, Shaanxi Normal University, Xi'an 710119, China, Email: junwei-he@snnu.edu.cn

†School of Mathematics and Information Science, Shaanxi Normal University, Xi'an 710119, China, Email: xchen80@snnu.edu.cn

‡Corresponding Author.

research interest. For example, Fan and Huang [4] proposed a profile-kernel inference and established the asymptotic normality of the profile least-square estimator for this model. You and Zhou [16] studied the model (1.1) using the empirical likelihood method when p is fixed. When dimensionality p of the parameters tends to infinity as the sample size $n \to \infty$, this generalized varying-coefficient partially linear model was considered by Lam and Fan [7]. More relevant works on the varying-coefficient partially linear model can be found in Huang and Zhang [6], Li et al. [8] and references therein.

Empirical likelihood method has taken much attention in literatures since it was introduced and developed by Owen [10,11]. One of the motivation is that the empirical likelihood-based confidence regions not only have natural shape and respect the range of the parameter, but also have the advantages of studentising automatically. In many cases, empirical likelihood-based confidence regions are shown to be Bartlett correctable(DiCiccio et al. [3], Chen and Cui [1]). Owen [12] and Xue and Zhu [15] are fairly comprehensive references.

However, in practical application, there is always lack-of-fit for the asymptotic normality distribution of empirical likelihood ratio with expectation p and variance $2p$ when we refer to the coverage probability, especially when p/n is not small. We find that this is mainly due to the underestimation of the expectation and variance of the empirical likelihood ratio, producing tests with type I errors much larger than the nominal level. And this inspires us to look for an effective estimation of the expectation and variance. Liu et al. [9] proposed a new method which is fitted for the calibration of empirical likelihood for high-dimensional data. Through the calibration of the expectation and variance of the empirical likelihood for the population mean, they got a considerable improvements for the coverage probabilities. Guo et al. [5] considered this calibration method for high-dimensional data in linear models and discussed the asymptotic behavior of the empirical likelihood ratio function in random and fixed design cases, respectively. Recently, Li et al. [8] showed that under some conditions, the bias-correction empirical likelihood for the semiparametric varying-coefficient partially linear models is asymptotic normal.

Taking these issues into account, in this paper, we consider a new calibration of empirical likelihood for semiparametric varying-coefficient partially linear models with diverging number of parameters and investigate the asymptotic behavior of this bias-corrected empirical likelihood ratio function. Numerical studies show that this new calibration method will have a great improvement.

The rest of this paper is organized as follows. In Section 2, we introduce the bias-corrected empirical likelihood(BCEL) for semiparametric varying-coefficient partially linear models. A new calibration of bias-corrected empirical likelihood is given in Section 3. In Section 4, some simulations are carried out to assess the performance of the proposed method. Technical proofs are stated in Section 5.

2. Bias-corrected Empirical Likelihood

Let $(Y_i; X_i^{\mathrm{T}}, Z_i^{\mathrm{T}}, U_i, 1 \le i \le n)$ be an independent identically distributed(i.i.d) random sample which come from the model (1.1) with the β and Z_i having the dimension $p \to \infty$ as $n \to \infty$. Then for any given β, we get

$$Y_i - Z_i^T \beta = X_i^T \alpha(U_i) + \varepsilon_i \tag{2.1}$$

Following Fan and Huang [4], we apply a local linear regression technique to estimate the varying-coefficient functions $\alpha_j(\cdot), j = 1, ..., q$. For v in a small neighborhood of u, one can approximate $\alpha_j(v)$ by

$$\alpha_j(v) \approx \alpha_j(u) + \alpha_j'(u)(v-u) \equiv a_j + b_j(v-u) \quad j = 1, ..., q \tag{2.2}$$

This leads to the following weighted local least squares problem: find $\{(a_j, b_j), j = 1, ..., q\}$ to minimize

$$\sum_{i=1}^{n}\{Y_i - X_i^T(a + b(U_i - u)) - Z_i^T\beta\}^2 K_h(U_i - u) \tag{2.3}$$

where $K(\cdot)$ is a kernel function, h is a bandwidth and $K_h(\cdot) = K_h(\cdot/h)/h$.

The solution of problem (2.3) is

$$\hat{\alpha}(u, \beta) = (I_q, O_q)(D_u^T W_u D_u)^{-1} D_u^T W_u (Y - Z^*\beta) \tag{2.4}$$

where I_q denotes a q-dimensional identity matrix, O_q is the $q \times q$ matrix with all the entries being 0 and Let

$$D_u = \begin{pmatrix} X_1^{\mathrm{T}} & \frac{U_1 - u}{h} X_1^{\mathrm{T}} \\ \vdots & \vdots \\ X_n^{\mathrm{T}} & \frac{U_n - u}{h} X_n^{\mathrm{T}} \end{pmatrix}, \quad Z^* = (Z_1, \ldots, Z_n) = \begin{pmatrix} Z_{11} & \ldots & Z_{1p} \\ \vdots & \ddots & \vdots \\ Z_{n1} & \ldots & Z_{np} \end{pmatrix}$$

$$Y = (Y_1, \ldots, Y_n), \quad W_u = \mathrm{diag}(K_h(U_1 - u), \ldots, K_h(U_n - u))$$

and

$$\mu(u) = (E(XX^{\mathrm{T}}|U = u))^{-1} E(XZ|U = u)$$

So we can write the auxiliary random vectors as follows

$$\hat{\eta}_i(\beta) = (Z_i - \hat{\mu}^{\mathrm{T}}(U_i) X_i)(Y_i - X_i^{\mathrm{T}} \hat{\mu}(U_i, \beta) - Z_i^{\mathrm{T}}\beta) \tag{2.5}$$

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

$$X_i^{\mathrm{T}} \hat{\mu}(U_i) = \sum_{k=1}^{n} S_{ik} Z_k \tag{2.6}$$

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

$$S = \begin{pmatrix} (X_1^{\mathrm{T}}, O)(D_{u_1}^{\mathrm{T}} W_{u_1} D_{u_1})^{-1} D_{u_1}^{\mathrm{T}} W_{u_1} \\ \vdots \\ (X_n^{\mathrm{T}}, O)(D_{u_n}^{\mathrm{T}} W_{u_n} D_{u_n})^{-1} D_{u_n}^{\mathrm{T}} W_{u_n} \end{pmatrix}$$

Thus, the bias-corrected auxiliary random vectors can be expressed as

$$\hat{\eta}_i(\beta) = (Z_i - \hat{\mu}^{\mathrm{T}}(U_i) X_i)(Y_i - X_i^{\mathrm{T}} \hat{\mu}(U_i, \beta) - Z_i^{\mathrm{T}}\beta) \triangleq \hat{Z}_i(\hat{Y}_i - \beta^{\mathrm{T}} \hat{Z}_i) \tag{2.7}$$

where $\hat{Z}_i = Z_i - \sum_{k=1}^{n} S_{ik} Z_k$, $\hat{Y}_i = Y_i - \sum_{k=1}^{n} S_{ik} Y_k$.

Therefore, a bias-corrected empirical log-likelihood ratio is defined as

$$l_n(\beta) = -2\max\left\{\sum_{i=1}^{n} \log(n\omega_i) \middle| \omega_i \geq 0, \sum_{i=1}^{n} \omega_i = 1, \sum_{i=1}^{n} \omega_i \hat{\eta}_i(\beta) = 0\right\} \tag{2.8}$$

By the Lagrange multiplier method, we can obtain

$$l_n(\beta) = 2\sum_{i=1}^{n} \log(1 + \lambda^T \hat{\eta}_i(\beta)) \tag{2.9}$$

where $\lambda = \lambda(\beta)$ is determined by

$$(2.10) \quad \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\eta}_i(\beta)}{1+\lambda^T\hat{\eta}_i(\beta)} = 0$$

According to Qin and Lawless [13], if $\alpha(u)$ is given and p is fixed, under some conditions, $l_n(\beta)$ is asymptotically χ^2 with p degree of freedom, which is a non-parametric version of Wilks' theorem. And when the number of p grows with the sample size n, Li et al. [8] showed that under some conditions, the conclusion below is valid.

$$(2.11) \quad \frac{l_n(\beta_0)-p}{\sqrt{2p}} \xrightarrow{d} N(0,1), \quad \text{as } n\to\infty$$

where β_0 is the true value of the parameter vector β.

3. A new Calibration method for BCEL

When testing hypotheses with the BCEL method, we would calculate the critical values based on normal approximation (2.11). However, these critical values often deviate from the true ones when p/n is not small. We find that this awkward fact is mainly due to the large difference between the true expectation and variance pair (E_n, V_n) of $l_n(\beta_0)$ and $(p, 2p)$. And our simulation also indicates that this method is not good. We know that the foundation of using (2.11) to calibrate the BCEL are that $l_n(\beta_0)$ is close to $K_n = n\bar{\hat{\eta}}_n^{\mathrm{T}}\Sigma^{-1}\bar{\hat{\eta}}_n$, and that $E(K_n) = p$, $Var(K_n) \approx 2p$. But in practice, we always use the moment estimation of K_n, which is, $T_n = n\bar{\hat{\eta}}_n^{\mathrm{T}}S_n^{-1}\bar{\hat{\eta}}_n$, whose expectation and variance are denoted as $(\hat{E}_{n1}, \hat{V}_{n1})$, for statistical inference and it can always get a better approximation to $l_n(\beta_0)$. But when (E_n, V_n) deviates from $(p, 2p)$ or $(\hat{E}_{n1}, \hat{V}_{n1})$, these calibration methods do not work any more.

We expect that replacing $(p, 2p)$ with $(\hat{E}_{n2}, \hat{V}_{n2})$, the expectation and variance of T_{nc}(see (3.2)), in (2.11), will improve the performance of the usual normal calibration. Let

$$f(\lambda) = 2\sum_{i=1}^{n}\log(1+\lambda^{\mathrm{T}}\hat{\eta}_i(\beta))$$

Obviously, $l_n(\beta_0) = \sup_\lambda f(\lambda) = f(\lambda_*)$, and λ_* is the maximum point of $f(\lambda)$. By second-order Taylor expansion, we have

$$(3.1) \quad f(\lambda) \approx g_1(\lambda) = 2\sum_{i=1}^{n}\left\{\lambda^{\mathrm{T}}\hat{\eta}_i - \frac{1}{2}(\lambda^{\mathrm{T}}\hat{\eta}_i)^2\right\}$$

provided $\lambda^{\mathrm{T}}\hat{\eta}_i$'s are small. So an approximation of $l_n(\beta_0)$ is

$$l_n(\beta_0) \approx \sup_\lambda f(\lambda) = \sup_\lambda g_1(S_n^{-1}\bar{\hat{\eta}}_n) = T_n$$

However, in the case of moderate n and large p, this approximation may not work any more. The remainder of each Taylor expansion in (3.1) is under control only for $\lambda^{\mathrm{T}}\hat{\eta}_i \in (-1,1)$. We find in our simulation that when p/n is not small, some of $\lambda^T\hat{\eta}_i$'s are greater than 1 with a large probability. Note that when

$$x \in (-1,1), \log(1+x) \approx x - \frac{x^2}{2}$$

while if

$$x > 1, \log(1+x) > \log(2) > x - \frac{x^2}{2}$$

Therefore, roughly we have $f(\lambda) \geq g_1(\lambda)$ in the neighborhood of 0. This finding also restrict us to approximate $l_n(\beta_0)$ by two terms Taylor expansion, because Taylor expansion of (3.1) would deviate from $l_n(\beta_0)$ if more terms are extracted and some of $\lambda_* \hat{\eta}_i$ are not small.

To reduce the approximation error of $g_1(\lambda)$, we add a high-order term $(\lambda^{\mathrm{T}}\hat{\eta}_i)^2$ to $g_1(\lambda)$. Intuitively $g_2(\lambda) = g_1(\lambda) + (\lambda^{\mathrm{T}}\hat{\eta}_i)^2$ is the better approximate to $f(\lambda)$. So is $\sup_\lambda g_2(\lambda)$ to $l_n(\beta_0) = \sup_\lambda f(\lambda)$. It can be verified

$$\sup_{\lambda} g_2(\lambda) = n\bar{\hat{\eta}}_n^{\mathrm{T}} S_{nc}^{-1} \bar{\hat{\eta}}_n = T_{nc} \tag{3.2}$$

with

$$S_{nc} = \frac{1}{n}\sum_{i=1}^{n} (\hat{\eta}_i - \bar{\hat{\eta}}_n)(\hat{\eta}_i - \bar{\hat{\eta}}_n)^{\mathrm{T}}$$

The following theorem establishes the asymptotic behavior of $l_n(\beta_0) - T_{nc}$.

3.1. Theorem. *Under Conditions* (C1) – (C9) *in Section* 5*, if* $p^{3+4/(k-2)}/n \to 0$*, for* $k \geq 4$*, then we have*

$$(l_n(\beta_0) - T_{nc})/p^{\frac{1}{2}} = o_p(1)$$

This theorem implies that using T_{nc} to approximate $l_n(\beta_0)$ is equivalent to using K_n or T_n from the asymptotic viewpoints. However, these approximations exhibit quite different finite-sample behaviors, especially when p/n is not small. Based on some simulations, we find that T_{nc} is amazingly close to $l_n(\beta_0)$ regardless of the choices of (n, p) in the sense that $(l_n(\beta_0) - T_{nc})/p^{\frac{1}{2}} = o_p(1)$ is always pretty small. To appreciate this, Fig.1 shows the scatter plots of 200 simulated values of $(l_n(\beta_0), T_n)$ and $(l_n(\beta_0), T_{nc})$ for the model(4.1) with the $\varepsilon_i \sim N(0,1)$. We choose p=10, 16 for n=200. From Fig.1, we can see that the value of $(l_n(\beta_0), T_{nc})$ are always around the line $y = x$, but T_n tends to under-approximate $l_n(\beta_0)$. See Sect.4 for more analysis and comparison.

Given the foregoing discussion and evidence, we expect that the expectation and variance of T_{nc} are good approximations of E_n and V_n, respectively. Let $(\hat{E}_{n2}, \hat{V}_{n2})$ be the moment estimation of (E_n, V_n). We may calculate critical values according to

$$l_n(\beta_0) - A_n/\sqrt{B_n} \xrightarrow{d} N(0,1) \tag{3.3}$$

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

4. Numerical Analysis

Here we report a simulation study designed to evaluate the performance of the proposed calibration method of BCEL. Throughout this section, we use the Epanechnikow kernel $K(u) = 0.75(1 - u^2)_+$, and use the "leave-one-out" cross-validation method to select the optimal bandwidth h_{opt}.

Consider the following semiparametric varying-coefficient partially linear model

$$Y_i = {X_i}^{\mathrm{T}}\alpha(U_i) + Z_i^{\mathrm{T}}\beta + \varepsilon_i, \quad i = 1, \ldots, n \tag{4.1}$$

In our simulations, $\beta = [0.5, 0.3, -0.5, 1, 0.1, -0.25, 0, \ldots, 0]^{\mathrm{T}}$, the covariate U_i is uniformly distributed on $[0, 1]$, the nonparametric component $\alpha(u) = (\alpha_1(u), \alpha_2(u))^{\mathrm{T}}$ with $\alpha_1(u) = 4 + \sin(2\pi u), \alpha_2(u) = 2u(1 - u)$, $X_i = (X_{i1}, X_{i2})^{\mathrm{T}}$ with $X_{i1} = 1$ and $X_{i2} \sim N(0, 1)$, the covariates Z_i is a p-dimensional normal random vector with mean zero and covariance matrix (σ_{ij}) with $\sigma_{ij} = 0.5^{|i-j|}$.

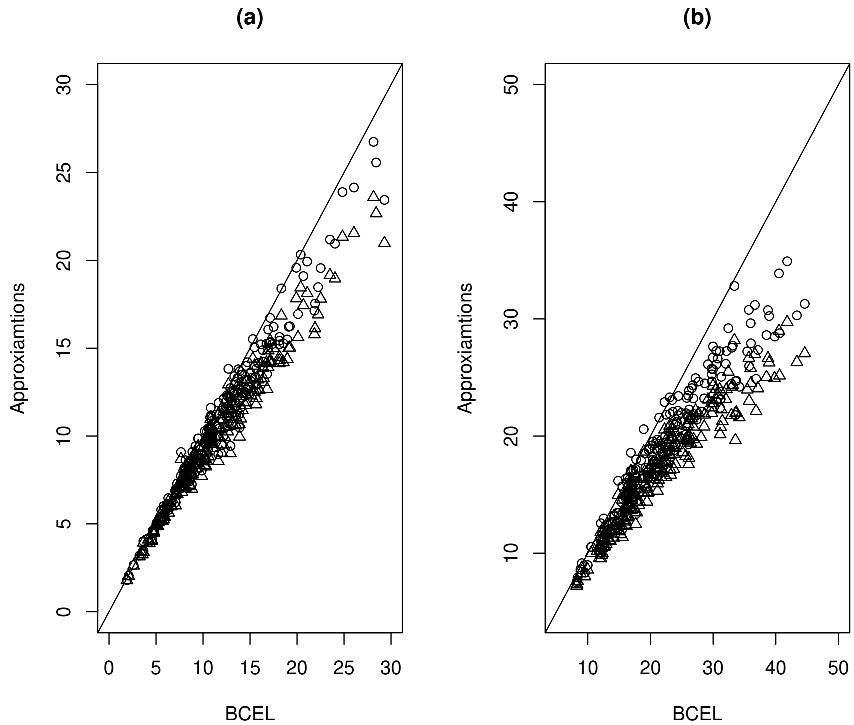

Fig.1. Scatter plots of n simulated values $(l_n(\beta_0), T_n)$ (triangles), $(l_n(\beta_0), T_{nc})$ (circles) for the model (4.1) with (a) n=200 and p=10, (b) n=200 and p=16. The solid line is $y = x$.

4.1. Simulation I. For this simulation, we evaluate the asymptotic normality of BCEL ratio using the following methods. The proposed method is based on the calibrated $l_n(\beta_0)$ with the sample mean and variance of T_{nc} obtained from 500 Bootstrap samples for each simulation data set(denoted as MEL). The normal calibration is based on the calibrated $l_n(\beta_0)$ with the sample mean and variance of T_n obtained from 500 Bootstrap samples for each simulation data set(denoted as SEL). And the standard normal calibration is base on the calibrated $l_n(\beta_0)$ with $(A_n, B_n) = (p, 2p)$(denoted as STEL). Through QQ-plots, we will demonstrate the advantages of MEL in different growth rates of p for each sample size. Here we only consider the case of noise $\varepsilon_i \sim N(0, 1)$.

We draw 1000 random samples of size 200, 400 or 600 from model (4.1). For comparison, we here take the dimensionality of the parametric component as $p = [cn^{1/3}]$. By assigning $c = 1.8, 2.8$ and 3.8, the corresponding dimensions $p = 10, 16$ and 22 for $n = 200; p = 13, 20$ and 27 for $n = 400; p = 15, 23$ and 32 for $n = 600$. The results are reported in Fig.2.

From Fig 2., we can observe from the QQ-plots that the MEL outperforms better than SEL and STEL as n increases or p decreases. Therefore, the MEL can be regarded as a reasonable alternative for the calibration of the BCEL in practice.

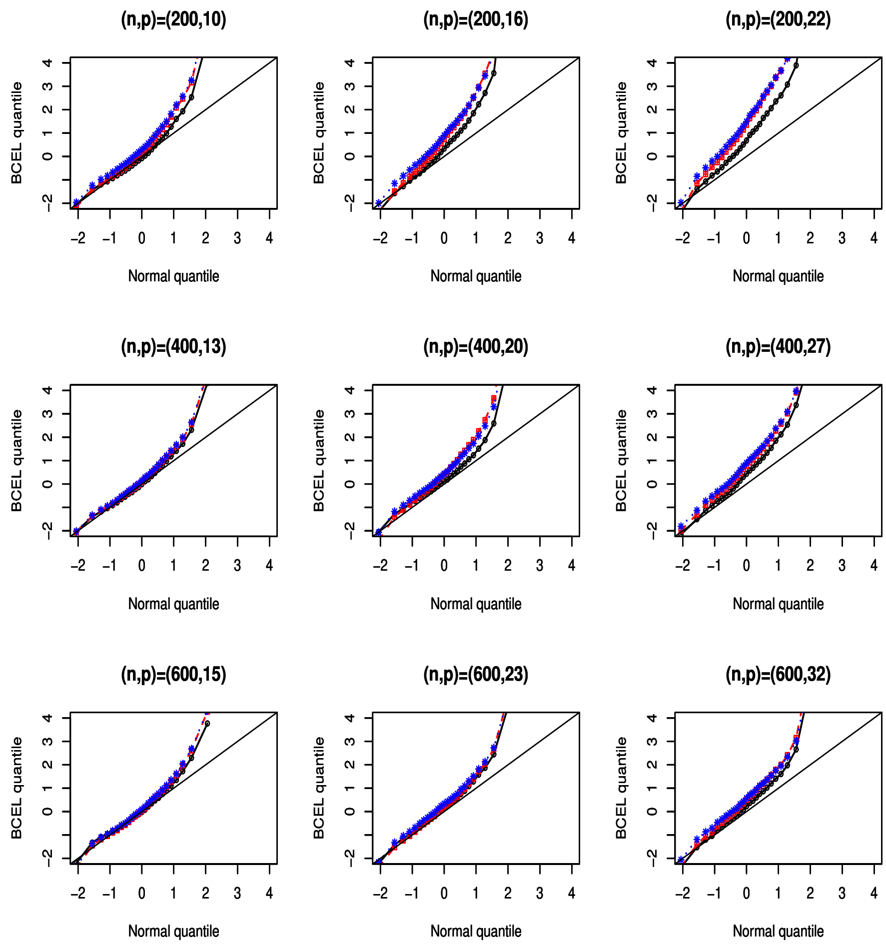

Fig.2. Normal QQ-plots of the BCEL ratio with $\epsilon_i \sim N(0,1)$: MEL(black and $-\circ-$), SEL(red and $\cdot\boxplus\cdot$), STEL(blue and $-\cdot*\cdot-$)

4.2. Simulation II. In this simulation, We draw 1000 random samples of size 200, 400 and 600, respectively. The choice of (n,p) is the same as Simulation I. As for noise, two error distributions were chosen: (i) the standard normal distribution; (ii) the chi-square distribution with freedom 3.

In this simulation, we will compare four calibration methods for the BCEL. Besides the MEL, SEL and STEL methods mentioned in Section 4.1, there also consider the ordinary χ_p^2 calibration(denoted as OEL). Talbes 1 and 2 report the coverage probability comparison for constructing confidence region on parameter β with nominal level 0.95.

It can be concluded from Tables 1 and 2 that the empirical coverage probabilities based on MEL are higher than that based on OEL, STEL and SEL. Especially for the case of $n=600$, $p=15$ and $\epsilon_i \sim N(0,1)$, the coverage probabilities of MEL is closed to the nominal level. Thus the calibration method of MEL is a good alternative. We can

Table 1. Coverage percentages for model (4.1) with the $\epsilon_i \sim N(0,1)$

n	p	MEL	OEL	SEL	STEL	$\hat{E}_{n1}$	$\hat{V}_{n1}$	$\hat{E}_{n2}$	$\hat{V}_{n2}$
200	10	0.920	0.838	0.846	0.854	10.67	19.15	11.69	25.80
	16	0.838	0.726	0.756	0.750	17.23	26.81	19.04	37.04
	22	0.764	0.552	0.593	0.615	23.46	38.75	26.79	56.09
400	13	0.937	0.899	0.910	0.914	13.24	26.39	13.81	29.77
	20	0.925	0.846	0.864	0.842	20.42	31.33	21.94	53.69
	27	0.841	0.741	0.777	0.789	27.98	51.72	30.24	58.93
600	15	0.936	0.898	0.904	0.911	15.36	29.69	15.62	37.41
	23	0.921	0.873	0.893	0.899	23.98	42.90	24.10	50.65
	32	0.896	0.836	0.872	0.849	32.88	54.57	34.72	64.91

Table 2. Coverage percentages for model (4.1) with the $\epsilon_i \sim \chi^2_3$

n	p	MEL	OEL	SEL	STEL	$\hat{E}_{n1}$	$\hat{V}_{n1}$	$\hat{E}_{n2}$	$\hat{V}_{n2}$
200	10	0.863	0.796	0.810	0.821	11.01	17.60	11.38	22.29
	16	0.803	0.694	0.721	0.698	16.73	25.31	18.27	34.76
	22	0.755	0.576	0.610	0.599	23.32	33.63	26.08	57.37
400	13	0.908	0.878	0.889	0.866	13.34	20.50	14.05	27.17
	20	0.844	0.772	0.798	0.763	20.45	31.61	22.07	41.41
	27	0.828	0.692	0.728	0.720	27.68	47.28	29.80	61.13
600	15	0.916	0.868	0.888	0.878	15.46	26.70	16.08	33.56
	23	0.890	0.852	0.869	0.871	23.83	43.75	24.62	47.74
	32	0.839	0.745	0.785	0.776	33.07	53.54	34.60	64.03

also observed from Table 1 and Table 2 that the MEL has improving coverage accuracy along with the increasing sample size. However, when the dimension p increases, the coverage probabilities of both MEL, OEL, STEL and SEL decrease. When $n = 200$ and $p = 22$, the performances of OEL, SEL and STEL are unacceptable. In comparison, our proposed method, MEL, can always attain the desired coverage percent and outperform the other three methods. The advantages get more remarkable when n decreases or p increases.

5. Proof of main results

Throughout the paper, we denote $\gamma_1(A) \leq \cdots \leq \gamma_p(A)$ as the eigenvalues and $tr(A)$ as the trace operator of a matrix A. To derive our main results, the following conditions required to be made.

(C1) The random variable U has a compact support Ω. The density function $f_U(u)$ of the U has a continuous second derivative and is uniformly bounded away from zero.
(C2) The $q \times q$ matrix $E(XX^{\mathrm{T}}|U = u)$ is non-singular for each $U \in \Omega$. Furthermore, $E(XX^{\mathrm{T}}|U = u)^{-1}$ and $E(XZ|U = u)$ are all Lipschitz continuous and each element of $E(XX^T|U = u)^{-1}$ and $E(XZ|U = u)$ is bounded.
(C3) $\{\alpha_i(\cdot), i = 1, \dots, q\}$ has continuous second derivatives in $u \in \Omega$.
(C4) The kernel $K(\cdot)$ is 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) $\Sigma = E[\varepsilon^2(Z - \mu^{\mathrm{T}}(U)X)(Z - \mu^{\mathrm{T}}(U)X)^{\mathrm{T}}]$ is a positive definite matrix with all the eigenvalues being uniformly bounded away from zero and infinity.

(C7) For some integer $k \geq 4$, $E(\| X\varepsilon \|^k) < \infty$, $E(\| X \|^k) < \infty$, $E(\| \varepsilon \|^k) < \infty$.
(C8) Let $\eta = \varepsilon(Z - \mu^{\mathrm{T}}(U)X)$, and η_j be the j-th component of $\eta, j = 1 \dots p$. For k of condition (C7), there is a positive constant c such that

$$E(\| \eta\sqrt{p} \|^k) < c, E(\| ZX^{\mathrm{T}}/\sqrt{p} \|^k) < c, E(\| \mu(U)XX^{\mathrm{T}}/\sqrt{p} \|^k) < c$$

and

$$\frac{1}{p}\sum_{l_1=1}^{p} E(| \eta_{l_1} |(\| ZX^{\mathrm{T}}/\sqrt{p} \|^4 + \| \mu(U)XX^{\mathrm{T}}/\sqrt{p} \|^4)) < c$$

.
(C9) $\max_{1\leq l_1,l_2,l_3\leq p} E(\eta_{l_1}\eta_{l_2}\eta_{l_3})^2$ is bounded, where η_{l_i} are the components of η.

In order to prove the main results, we introduce the following notations. Simple calculation yields that

$$\hat{\eta}_i(\beta) = \eta_i(\beta) + \sum_{k=1}^{3} M_{i,k} =: \eta_i(\beta) + R_i \tag{5.1}$$

where

$$\eta_i(\beta) = (Z_i - \mu^{\mathrm{T}}(U_i)X_i)(Y_i - X_i^{\mathrm{T}}\alpha(U_i) - Z_i^{\mathrm{T}}\beta) = (Z_i - \mu^{\mathrm{T}}(U_i)X_i)\epsilon_i$$
$$M_{i,1} = (Z_i - \mu^{\mathrm{T}}(U_i)X_i)X_i^{\mathrm{T}}(\alpha(U_i) - \hat{\alpha}(U_i,\beta))$$
$$M_{i,2} = (\mu(U_i) - \hat{\mu}(U_i))^{\mathrm{T}}X_i\epsilon_i$$
$$M_{i,3} = [(\mu(U_i) - \hat{\mu}(U_i))^{\mathrm{T}}X_i][X_i^{\mathrm{T}}(\alpha(U_i) - \hat{\alpha}(U_i,\beta))]$$

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

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

$$\sup_{u\in\Omega}\left|\frac{1}{n}\sum_{i=1}^{n} K_h(U_i - u)\left(\frac{U_i - u}{h}\right)^l X_{ij_1}X_{ij_2} - f(u)\mu_1\Gamma_{j_1j_2}(u)\right| = O_p(c_n)$$

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

where $j_1, j_2, j = 1, \dots, q, k = 1, \dots, p, l = 0, 1, 2, 4$, $\Gamma_{j_1j_2}(u)$ is the (j_1, j_2)-th the element of $\Gamma(u)$ and $\phi_{jk}(u)$ is the (j, k)-th element of $\phi(u)$.

We refer to Xia and Li [14] for details.

5.2. Lemma. *Under the Conditions of Lemma* 5.1, *we have,*

$$\| \hat{\alpha}(u,\beta) - \alpha(u) \| = O_p(c_n) \tag{5.2}$$

and

$$\max_{1\leq j\leq q}\sup_{u\in\Omega} | \hat{\alpha}_j(u,\beta) - \alpha_j(u) | = O_p(c_n) \tag{5.3}$$

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

Proof. We first give the proof of (5.2). Let

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

Then, we can rewrite

$$D_u^{\mathrm{T}} W_u D_u = \begin{pmatrix} S_{n,0} & S_{n,1} \\ S_{n,1} & S_{n,2} \end{pmatrix}$$

The elements of the above matrix are in the form of a kernel regression. From Lemma 5.1 and some simple calculation, we have

$$S_{n,l} = nf(u)\mu_l\Gamma(u)(1 + O_p(c_n)) \tag{5.4}$$

holds uniformly in $u \in \Omega$. So

$$\hat{\alpha}(u,\beta) = [nf(u)\Gamma(u)]^{-1}\sum_{i=1}^{n} K_h(U_i - u)X_i\{X_i^{\mathrm{T}}\alpha(U_i) + \varepsilon_i\} + O_p(c_n) \tag{5.5}$$

Applying Lemma 5.1 and (5.4), we can easily get

$$\frac{1}{n}\sum_{i=1}^{n} K_h(U_i - u)X_iX_i^{\mathrm{T}}\alpha(U_i) = f(u)\Gamma(u)\alpha(u)\{1 + O_p(c_n)\} \tag{5.6}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} K_h(U_i - u)X_i\varepsilon_i = o_p(1) \tag{5.7}$$

holds uniformly in $u \in \Omega$. From (5.5)-(5.7), $\hat{\alpha}(u,\beta) = \alpha(u) + O_p(c_n)$ holds uniformly in $u \in \Omega$. This completes the proof of (5.2).

By the similar method of Xia and Li [14], we can conclude the result (5.3), so we omit the details here. □

5.3. Lemma. *Under the Conditions of Lemma* 5.1, *we have*

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_i\right\| = O_p(n^{1/2}p^{1/2}c_n^2)$$

where $R_i = \sum_{k=1}^{3} M_{i,k}$ *can be found in* (5.1).

The proof of Lemma 5.3 is similar as that of Lemma B.3 in Li et al. [5].

5.4. Lemma. *Under conditions* (C1)-(C8), *we have*

$$tr[(S_{nc} - \Sigma)^2] = O_p(p^2(c_n^4 + 1/n)) \tag{5.8}$$

Proof. From the definition of η_i and S_{nc}, we can get

$$S_{nc} - \Sigma = \frac{1}{n}\sum_{i=1}^{n}\eta_i\eta_i^{\mathrm{T}} + \frac{1}{n}\sum_{i=1}^{n}(R_i\eta_i^{\mathrm{T}} + \eta_iR_i^{\mathrm{T}} + R_iR_i^{\mathrm{T}}) - \bar{\hat{\eta}}_n\bar{\hat{\eta}}_n^{\mathrm{T}} = J_1 + J_2 + J_3$$

It is easy to see that

$$\begin{aligned} tr[(S_{nc} - \Sigma)^2] =& tr[(J_1 + J_2 + J_3)^2] \le 4tr[(J_1)^2] + 4tr[(J_2)^2] + 2tr[(J_3)^2] \\ =& I_1 + I_2 + I_3 \end{aligned}$$

Thus, we know that

$$I_1 = O_p(p/n^2), \quad I_2 = O_p(p^2c_n^4)$$

For I_3, first we can get $\bar{\hat{\eta}}_n = O_p(\sqrt{p/n})$, then

$$I_3 = tr[(\bar{\hat{\eta}}_n\bar{\hat{\eta}}_n^{\mathrm{T}})^2] = O_p(p^3/n^2) = \frac{p}{n}O_p(p^2/n) = O_p(p^2/n)$$

Therefore, we have

$$tr[(S_{nc} - \Sigma)^2] = I_1 + I_2 + I_3 = O_p(p^2/n) + O_p(p^2c_n^4) + O_p(p^2/n) = O_p(p^2(c_n^4 + 1/n))$$

The proof is complete. □

5.5. Lemma. *Under conditions* (C1)-(C8), *if* $p^{3+4/(k-2)}/n \to 0$, *we have*

$$n\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i\right)^{\mathrm{T}}(S_{nc}^{-1}-\Sigma^{-1})\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i\right)\right\}=o_p(p^{1/2}) \tag{5.9}$$

Proof. Let $\hat{D}_n=\Sigma^{-1/2}S_{nc}\Sigma^{-1/2}-I_p$, similar arguments used in the proof of Lemma 6 in Chen et al. [2] yield

$$\begin{aligned}S_{nc}^{-1}-\Sigma^{-1}=&\Sigma^{-1/2}(\Sigma^{1/2}S_n^{-1}\Sigma^{1/2}-I_p)\Sigma^{-1/2}\\=&\Sigma^{-1/2}[-\hat{D}_n+\hat{D}_n^2+\hat{D}_n^2\{\Sigma^{1/2}S_n^{-1}\Sigma^{1/2}-I_p\}]\Sigma^{-1/2}\end{aligned}$$

Note that

$$\begin{aligned}tr((S_{nc}-\Sigma)^2)=&tr((\Sigma^{1/2}(\Sigma^{-1/2}S_n\Sigma^{-1/2}-I_p)\Sigma^{1/2})^2)\\=&tr(\hat{D}_n\Sigma\hat{D}_n\Sigma)\\\geq&\gamma_1^2(\Sigma)tr(\hat{D}_n^2)\end{aligned}$$

By Lemma 5.4 , we have

$$tr(\hat{D}_n^2)\leq\frac{1}{\gamma_1^2(\Sigma)}tr((S_{nc}-\Sigma)^2)=O_p(p^2(c_n^4+1/n))$$

Thus, we have

$$\begin{aligned}tr(S_{nc}^{-1}-\Sigma^{-1})^2&\leq 2tr\{\Sigma^{-2}(-\hat{D}_n+\hat{D}_n^2)^2\}+2tr\{\hat{D}_n^4(S_{nc}^{-1}-\Sigma^{-1})^2\}\\&\leq 2tr\{\Sigma^{-2}(-\hat{D}_n+\hat{D}_n^2)^2\}+2[tr\hat{D}_n^2]^2tr\{(S_{nc}^{-1}-\Sigma^{-1})^2\}\\&=2tr\{\Sigma^{-2}(-\hat{D}_n+\hat{D}_n^2)^2\}+o_p(tr\{(S_{nc}^{-1}-\Sigma^{-1})^2\})\\&=o_p(p^2(c_n^4+1/n))\end{aligned} \tag{5.10}$$

Then

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i^{\mathrm{T}}\right\|=O_p(\sqrt{p/n}) \tag{5.11}$$

This together with $p^{3+4/(k-2)}/n\to 0, c_n^2=o(1/\sqrt{n})$ and condition (C5), we can obtain

$$\begin{aligned}n\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i\right)^{\mathrm{T}}(S_{nc}^{-1}-\Sigma^{-1})\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i\right)\right\}&\leq n\left\|\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i^{\mathrm{T}}\right\|^2\sqrt{tr(S_{nc}^{-1}-\Sigma^{-1})^2}\\&=o_p(p^2(c_n^2+1/\sqrt{n}))\\&=o_p(p^{1/2})\end{aligned}$$

The proof is finished. □

Proof of Theorem 3.1 Applying the Taylor expansion to (2.9) and invoking Lemmas 5.3-5.5, we obtain that

$$\begin{aligned}l_n(\beta_0)=2\sum_{i=1}^{n}\log(1+\lambda^{\mathrm{T}}\hat{\eta}_i(\beta))=&n\left\{\bar{\hat{\eta}}_n^{\mathrm{T}}\Sigma^{-1}\bar{\hat{\eta}}_n\right\}+o_p(p^{1/2})\\=&n\left\{\bar{\hat{\eta}}_n^{\mathrm{T}}(\Sigma^{-1}-S_{nc}^{-1})\bar{\hat{\eta}}_n\right\}+n\left\{\bar{\hat{\eta}}_n^{\mathrm{T}}S_{nc}^{-1}\bar{\hat{\eta}}_n\right\}+o_p(p^{1/2})\end{aligned}$$

From Lemma 5.4, we have

$$n\left\{\bar{\hat{\eta}}_n^{\mathrm{T}}(\Sigma^{-1}-S_{nc}^{-1})\bar{\hat{\eta}}_n\right\}=o_p(p^{1/2})$$

So

$$\begin{aligned}\frac{l_n(\beta_0)-T_{nc}}{p^{1/2}} =& \frac{n\left\{\bar{\bar{\eta}}^{\mathrm{T}}{}_n(\Sigma^{-1}-S_{nc}^{-1})\bar{\bar{\eta}}_n\right\}+n\left\{\bar{\bar{\eta}}_n^{\mathrm{T}}S_{nc}^{-1}\bar{\bar{\eta}}_n\right\}+o_p(p^{1/2})-n\left\{\bar{\bar{\eta}}_n^{\mathrm{T}}S_{nc}^{-1}\bar{\bar{\eta}}_n\right\}}{p^{1/2}}\\ =&\frac{n\left\{\bar{\bar{\eta}}_n^{\mathrm{T}}(\Sigma^{-1}-S_{nc}^{-1})\bar{\bar{\eta}}_n\right\}}{p^{1/2}}\\ =&o_p(1)\end{aligned}$$

The proof is complete. □

Acknowledgments

This work is supported by the National Natural Science Foundation of China (11201276, 61772017), the Natural Science Basic Research Plan in Shaanxi Province of China (2014JQ1042), and the Fundamental Research Funds for the Central Universities (GK201503015).

References

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

[2] Chen, S. X., Peng, L. and Qin, Y. L. *Effects of data dimension on empirical likelihood*, Biometrika **9**, 711-722, 2009.

[3] DiCiccio, T. J., Hall, P. and Romano, J. P. *Empirical likelihood is bartlett-correctable*, Ann. Statist. **19**, 1053-1061, 1991.

[4] Fan, J. and Huang, T. *Profile likelihood inference on semiparametric varying-coefficient partially linear models*, Bernoulli **11**, 1031-1057, 2005.

[5] Guo, H., Zou, C. and Chen, B. *Empirical likelihood for high-dimensional linear regression models*, Metrika **77**, 921-945, 2014.

[6] Huang, Z. S. and Zhang, R. Q. *Empirical likelihood for nonparametric parts in semiparametric varying-coefficient partially linear models*, Statist. Probab. Lett. **79**, 1798-1808, 2009.

[7] Lam, C. and Fan, J. *Profile-kernel likelihood inference with diverging number of parameters*, Ann. Statist. **36**, 2232-2260, 2008.

[8] Li, G., Lin, L. and Zhu, L. *Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters*, J. Multivar. Anal. **105**, 85-111, 2012.

[9] Liu, Y., Zou, C. and Wang, Z. *Calibration of the empirical likelihood for high-dimensional data*, Ann. Inst. Stat. Math. **65**, 529-550, 2013.

[10] Owen, A. B. *Empirical likelihood ratio confidence intervals for single functionals*, Biometrika **75**, 237-249, 1988.

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

[12] Owen, A. B. *Empirical Likelihood*, Chapman & Hall, New York, 2001.

[13] Qin, J. and Lawless, J. *Empirical likelihood and general estimating equations*, Ann. Statist. **22**, 300-325, 1994.

[14] Xia, Y. and Li, W.K. *On the estimation and testing of functional-coefficient linear models*, Statistica Sinica **9**, 735-757, 1999.

[15] Xue, L. G. and Zhu, L. X. *Empirical Likelihood in Nonparametric and Semiparametric Models*, Science Press, Beijing, 2010.

[16] You, J. and Zhou, Y. *Empirical likelihood for semiparametric varying-coefficient partially linear regression models*, Statist. Probab. Lett. **76**, 412-422, 2006.