

Inference Based on Empirical Likelihood for Varying Coefficient Model with Random Effect

Wanbin Li^{1,2}, Liugen Xue¹

¹College of Applied Sciences, Beijing University of Technology, Beijing, China

²Department of Mathematics, Yancheng Teachers University, Yancheng, China

Email: lwb@emails.bjut.edu.cn, lgxue@bjut.edu.cn

Received November 14, 2013; revised December 14, 2013; accepted December 21, 2013

Copyright © 2013 Wanbin Li, Liugen Xue. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2013 are reserved for SCIRP and the owner of the intellectual property Wanbin Li, Liugen Xue. All Copyright © 2013 are guarded by law and by SCIRP as a guardian.

ABSTRACT

In this article, we develop a statistical inference technique for the unknown coefficient functions in the varying coefficient model with random effect. A residual-adjusted block empirical likelihood (RABEL) method is suggested to investigate the model by taking the within-subject correlation into account. Due to the residual adjustment, the proposed RABEL is asymptotically chi-squared distribution. We illustrate the large sample performance of the proposed method via Monte Carlo simulations and a real data application.

Keywords: Varying Coefficient Model; Random Effect; Empirical Likelihood; Longitudinal Data

1. Introduction

Varying coefficient model has been widely used to model all kinds of data. One popular application is the analysis of the longitudinal data (e.g. [1,2]). Although both [3] and [4] proposed effective inference procedure for the varying coefficient model and applied them to the analysis of CD4 count data, whose detailed information can be referred to [5], none of them considered the within-subject correlation of longitudinal data. To improve the efficiency of the inference by considering this kind of correlation, we consider the varying coefficient model with random effect

$$Y_{ij} = X_{ij}^T \theta(T_{ij}) + Z_{ij}^T \alpha_i + e_{ij}, \quad (1)$$

with $i = 1, \dots, n$, $j = 1, \dots, m$, where $\theta(\cdot)$ is an unknown smoothing $p \times 1$ function vector, α_i are independent $l \times 1$ vectors of random effect with mean 0 and covariance matrix D and e_{ij} are independent mean 0 random variables with variance $\sigma_e^2 > 0$. Let Y_{ij} and X_{ij} , Z_{ij} , T_{ij} denote the response and covariate variable, respectively, where the i and j are their associated j th measurement of the i th subject among all of the longitudinal data.

Random effect model is frequently employed to exploit the characteristics of longitudinal data over several time periods. Recently, there has been fruitful research

on it. (e.g. [6-9]) proposed an efficient estimation for the single index model with random effect and provided a further way to construct the confidence interval for the parameter of interest with the aid of an estimator for its asymptotic variance.

In this article, we address a general problem to construct the confidence interval for the varying coefficient model with random effect by using the empirical likelihood method (e.g. [10-12]). Thanks to the empirical likelihood method, we can construct the confidence interval without the estimation for the asymptotic variance, and the whole inference procedure is totally data-adaptive. For longitudinal data, except for [3,13] studied an empirical likelihood method for the varying coefficient error-in-variable models with longitudinal data. Both of them did not consider incorporating the within-subject correlation. [14] reported that it caused a loss of efficiency for empirical likelihood applications by ignoring the within-subject correlation. In this article, we propose a residual-adjusted block empirical likelihood (RABEL) method for the varying coefficient model with random effect to incorporate the within-subject correlation for longitudinal data. This approach is appealing in that it can not only construct the confidence interval for the unknown coefficient function, but also improve estimation efficiency through considering the within-subject

correlation of longitudinal data. Also the estimation procedure introduced in Section 2 makes the implementation much easier.

The rest of this article is organized as follows. In Section 2, we introduce the residual-adjusted block empirical likelihood method and provide details on constructing the confidence interval for the varying coefficient function of interest. Some asymptotic results are derived in Section 3 and several implementation issues are shown in Section 4. Simulations are reported in Section 5. Data arising from CD4 study is analyzed in Section 6. Proofs of the main results are relegated to Appendix.

2. Estimation Method

2.1. Empirical Likelihood Estimation

Assume that the observed data $\{(Y_{ij}, X_{ij}, T_{ij})\}$ are generated from model (1). Moreover, let $Y_i = (Y_{i1}, \dots, Y_{im})^T$, $X_i = (X_{i1}, \dots, X_{im})^T$, $Z_i = (Z_{i1}, \dots, Z_{im})^T$ and $e_i = (e_{i1}, \dots, e_{im})^T$. Based on the idea of GEE [15], we can construct an auxiliary random vector nonparametric component

$$\eta_i(\theta(t)) = X_i^T W_i \Omega_i^{-1} (Y_i - X_i \theta(t)) \quad (2)$$

where Ω_i is the within-subject covariance matrix. Denote $W_i = \text{diag}(K_h(t_{i1} - t), \dots, K_h(t_{im} - t))$, where $K_h(\cdot) = K(\cdot/h)$ with $K(\cdot)$ and h being a kernel function and proper bandwidth, respectively. Note that $E(\eta_i(\theta(t))) = 0$ if $\theta(t)$ is the true parameter. Therefore, we can introduce an estimating equation as

$\sum_{i=1}^n \eta_i(\theta(t)) = 0$, and a naive empirical log-likelihood ratio function for $\theta(t)$ can be derived as

$$l(\theta(t)) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \eta_i(\theta(t)) = 0 \right\}. \quad (3)$$

However, by the similar argument in [3], the proposed empirical log-likelihood ratio (3) is no longer a standard chi-square distribution unless an undersmoothing bandwidth is chosen. And some complicated techniques, such as Monte Carlo approximation or estimated transformation, have to be employed to make further inference. For its limit being chi-square in practice, we propose a RABEL method by taking the ideal of [14] and [16]. Under the framework of RABEL, the auxiliary random vector function is newly defined as

$$\tilde{\eta}_i(\theta(t)) = X_i^T W_i \Omega_i^{-1} [Y_i - X_i \theta(t) - X_i \{\tilde{\theta}(t_i) - \tilde{\theta}(t)\}] \quad (4)$$

where the estimator $\tilde{\theta}(t)$ is a preliminary estimator. And the RABEL ratio is further derived as

$$\tilde{l}(\theta(t)) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \tilde{\eta}_i(\theta(t)) = 0 \right\}. \quad (5)$$

By the Lagrange multiplier method, we have

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda^T \tilde{\eta}_i(\theta(t))}, \quad i = 1, \dots, n, \quad (6)$$

where λ is a $p \times 1$ vector satisfying

$$\frac{1}{n} \sum_{i=1}^n \frac{\tilde{\eta}_i(\theta(t))}{1 + \lambda^T \tilde{\eta}_i(\theta(t))} = 0, \quad i = 1, \dots, n. \quad (7)$$

Then, the empirical likelihood ratio function (5) can be represented as

$$\tilde{l}(\theta(t)) = 2 \sum_{i=1}^n \log(1 + \lambda^T \tilde{\eta}_i(\theta(t))), \quad i = 1, \dots, n. \quad (8)$$

2.2. Estimation of the Variance Component

The within-subject covariance Ω_i is assumed to be known in the proposed RABEL procedure in Section 2.1. However, in practice, we need to construct an estimator for it. Assume that the model (1) satisfies the variance covariance-variance model

$$\Omega_i = \sigma_\alpha^2 \mathbf{1}_m \mathbf{1}_m^T + \sigma_e^2 I_m, \quad (9)$$

where $\mathbf{1}_m$ represents an $m \times 1$ vector of ones and I_m is the $m \times m$ identity matrix. This proposed model (9) is called a variance component model and widely used in longitudinal analysis, see, for example (e.g. [8,9]). Let $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{im})^T$, with $\varepsilon_{ij} = \alpha_i + e_{ij}$. Hence, an estimator for the variance component $(\sigma_\alpha^2, \sigma_e^2)$ is derived as

$$\hat{\sigma}_\alpha^2 = \frac{1}{nm(m-1)} \sum_{i=1}^n \sum_{j=1}^m \sum_{j_1 \neq j_2} \hat{\varepsilon}_{ij_1} \hat{\varepsilon}_{ij_2},$$

and

$$\hat{\sigma}_e^2 = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \hat{\varepsilon}_{ij}^2 - \hat{\sigma}_\alpha^2, \quad (10)$$

where with the preliminary estimator $\tilde{\theta}(t)$ for $\theta(t)$, the residual $\hat{\varepsilon}_{ij}$ is defined as

$$\hat{\varepsilon}_{ij} = Y_{ij} - X_{ij}^T \tilde{\theta}(t_{ij}), \quad i = 1, \dots, n, \quad j = 1, \dots, m.$$

Therefore, we obtain the estimator for Ω_i , that is

$$\hat{\Omega}_i = \hat{\sigma}_\alpha^2 \mathbf{1}_m \mathbf{1}_m^T + \hat{\sigma}_e^2 I_m, \quad (11)$$

Furthermore, with the true covariance in (4) being interpolated by the estimator (11), we derive a new auxiliary random vector

$$\hat{\tilde{\eta}}_i(\theta(t)) = X_i^T W_i \hat{\Omega}_i^{-1} [Y_i - X_i \theta(t) - X_i \{\tilde{\theta}(t_i) - \tilde{\theta}(t)\}] \quad (12)$$

And, the newly estimated RABEL ratio function is

$$\hat{l}(\theta(t)) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) \middle| p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\eta}_i(\theta(t)) = 0 \right\}. \quad (13)$$

Additional implementation details of solving the preliminary estimator will be postponed to Section 4, after discussion of the asymptotic properties of the proposed method.

3. Asymptotic Result

In this section, we study the asymptotic properties of the estimators. Denote $f(\cdot)$ to be the density for covariate variable T_{ij} .

Theorem 1. *Suppose that conditions (C1)-(C6) in the Appendix hold, then*

$$\sqrt{n}(\hat{\sigma}_\alpha^2 - \sigma_\alpha^2) \xrightarrow{L} N(0, V_\alpha), \quad (14)$$

and

$$\sqrt{n}(\hat{\sigma}_e^2 - \sigma_e^2) \xrightarrow{L} N(0, V_e), \quad (15)$$

where $V_\alpha = \text{Var}(\alpha_1^2) + \frac{4\sigma_\alpha^2\sigma_e^2}{m} + \frac{2\sigma_e^4}{m(m-1)}$ and

$$V_e = \text{Var}(e_{11}^2) + \frac{2\sigma_e^4}{m-1}.$$

Theorem 2. *Suppose that conditions (C1)-(C7) in the Appendix hold. If $\theta(t)$ is the true value of the parameter, then*

$$\hat{l}(\theta(t)) \xrightarrow{L} \chi_p^2, \quad (16)$$

where χ_p^2 means a χ^2 distribution with freedom p .

Note that $\chi_p^2(1-\alpha)$ is the $(1-\alpha)$ quantile of χ_p^2 . Based on Theorem 2, we can derive the $(1-\alpha)$ confidence interval of $\theta(t)$, that is

$$C_\alpha(\theta(t)) = \left\{ \theta(t) \middle| \hat{l}(\theta(t)) \leq \chi_p^2(1-\alpha) \right\}.$$

4. Implementation Issue

4.1. Calculation of the Preliminary Estimator

For t in the small neighborhood of t_0 , a Taylor expansion for the k th component of the varying coefficient function $\theta(t)$ leads to

$$\theta_k(t) = a_k + b_k(t - t_0), \quad k = 1, \dots, p, \quad (17)$$

where $b_k = \frac{\partial \theta_k(t)}{\partial t} \Big|_{t=t_0}$.

By ignoring the within-subject correlation, we can ob-

tain the unknown vector a, b by minimizing

$$\sum_{i=1}^n \sum_{j=1}^m \left[Y_{ij} - X_{ij}^T \{ a - b(t_{ij} - t_0) \} \right]^2 K_h(t_{ij} - t_0), \quad (18)$$

where $a = (a_1, \dots, a_p)^T$, $b = (b_1, \dots, b_p)^T$. Then, the solution to the minimization of (18) is

$$(\hat{a}^T(t_0), \hat{b}^T(t_0))^T = (D_{t_0}^T W_{t_0} D_{t_0})^{-1} D_{t_0}^T W_{t_0} Y,$$

and

$$D_{t_0} = \begin{pmatrix} X_{11} & \dots & X_{nm} \\ X_{11}(t_{11} - t_0) & \dots & X_{nm}(t_{nm} - t_0) \end{pmatrix}.$$

Specially, by the idea of local linear estimation, the preliminary estimator $\tilde{\theta}(t_0)$ is followed by

$$\tilde{\theta}(t_0) = (I_p, 0_p) (D_{t_0}^T W_{t_0} D_{t_0})^{-1} D_{t_0}^T W_{t_0} Y. \quad (19)$$

4.2. Choice of Bandwidth

As is well-known, the choice of bandwidth h can affect both the bias and variance estimation and there is a trade-off between a proper bias and variance. A smaller variance arise with the choice of a large bandwidth value, whereas will increase the estimation bias. As pointed out by [3,13], an optimal bandwidth, which is selected by using the leave-one-subject-out cross validation, can satisfy the conditions. Therefore, in the following simulations in this article, we use kernel function

$K(u) = 0.75(1-u^2)_+$ and the optimal bandwidth is obtained by minimizing

$$CV(h) = \sum_{i=1}^n \sum_{j=1}^m \left[Y_{ij} - X_{ij}^T \tilde{\theta}^{(-i)}(t_{ij}, h) \right]^2,$$

where $\tilde{\theta}^{(-i)}(t_{ij}, h)$ is denoted to be the preliminary estimator (19) estimated with all over the measurements except the i th subject.

5. Empirical Study

In this section, we perform some Monte Carlo simulations to assess the finite sample performance of the proposed method. Assume the data is generated from the model

$$Y_{ij} = X_{ij}^T \theta(T_{ij}) + \alpha_i + e_{ij}, \quad i = 1, \dots, n; j = 1, \dots, m, \quad (20)$$

where $\theta(t) = (\theta_1(t), \theta_2(t))^T$ with $\theta_1(t) = \sin(\pi t/2)$ and $\theta_2(t) = 1 - \cos(\pi t)$. Moreover, we also assume that

$X_{ijk} \sim N(0, 4^2), k = 1, 2$ and T_{ij} are generated from an uniform distribution on interval $[0, 1]$. Meanwhile, denote that $\alpha_i \sim N(0, \sigma_\alpha^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$, where $(\sigma_\alpha, \sigma_e)$ is set to be $(0.2, 0.1)$. It is assumed that the number of observed subject and repeated measurement within subject are $n = 100$ and $m = 3$. For the purpose of intensive

comparison, in addition to the proposed procedure during the simulations, a “naive” approach, based on the working independence assumption, is also involved, assuming that the within cluster covariance matrices are identities. The simulation results were calculated by 100 runs.

Figure 1 reports the approximate 95% point-wise confidence intervals and their coverage probability curves for the coefficient function $\theta_1(t)$ and $\theta_2(t)$, calculated by the proposed method in this article and the “naive” method. Although from **Figure 1**, two methods construct close confidence intervals for the nonparametric component, the coverage probability curves in the right panels show a significant difference. The coverage probability curves, estimated with the proposed method in this article, are closer to the significance lever 95% and possess a more stable and superior performance.

By the two required methods, we construct the simultaneous confidence regions of $(\theta_1(t), \theta_2(t))$ at time point $t = 1$ in **Figure 2**. The two plot are designed with $(\sigma_\alpha, \sigma_e)$ being (0.1,0.1) and (0.2,0.1), respectively. Direct comparison of them illustrates the inference improvement of the proposed RABEL method on that “naive” one.

6. Real Application

To illustrate the effectiveness of the proposed inference

procedure in this article, we apply it to the analysis of a longitudinal AIDS data set, reported by [5]. Some inference methods are related in literatures of [3,4,17]. However, there is limited work of them that focused on the within-subject correlation under the random effect framework. In fact, when we use the Hausman test for the null hypothesis of the random effect, the random effect is approved at 5% level of significance with p -value 85.9%, which can be calculated by the R function `phtest()` in the `plm` package.

As to the j th measurement of the i th subject, let Y_{ij} be CD4 percentage, t_{ij} be the time in years after HIV infection, Z_{i1} be the centered age at HIV infection, Z_{i2} be the centered preCD4 percentage, and Z_{i3} be the smoking status, taking a value of 1 or 0 for smoker or nonsmoker. Hence, we consider the following varying coefficient model with random effect

$$Y_{ij} = \theta_0(t_{ij}) + Z_{ij,1}\theta_1(t_{ij}) + Z_{ij,2}\theta_2(t_{ij}) + Z_{ij,3}\theta_3(t_{ij}) + \alpha_i + e_{ij}, \tag{21}$$

where the baseline CD4 percentage curve $\theta_0(t)$ is used to represent the mean CD4 percentage of t years after the infection. By the proposed inference procedure in this article, we plot the curves of the unknown coefficient functions in model (21) and their approximate 95% confidence intervals in **Figure 3**. From the curve of baseline

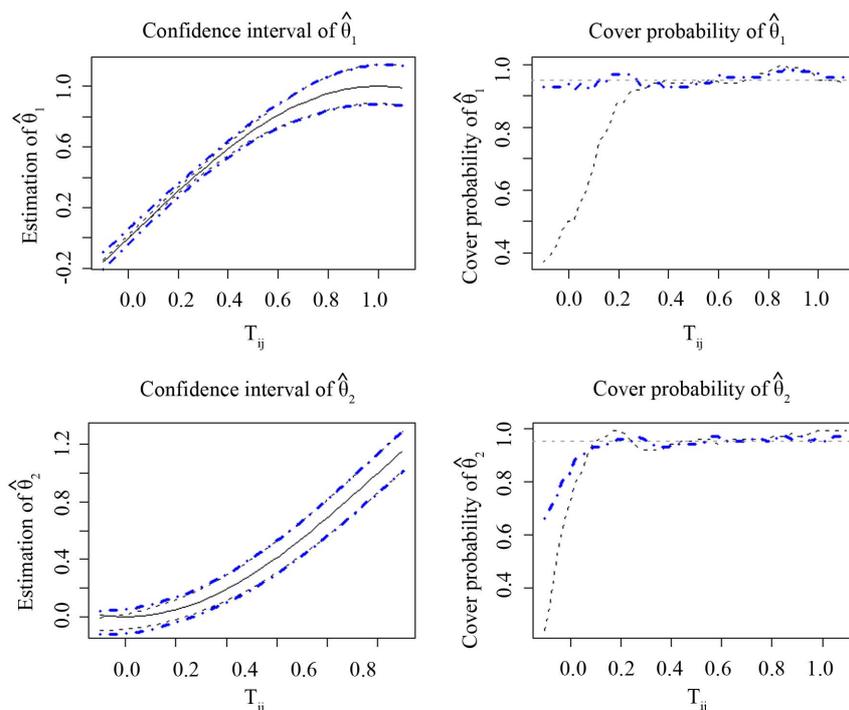


Figure 1. The true curve and its 95% confidence interval of the varying coefficient function are shown in the left two panels. On the other hand, the right two panels show the coverage probability curve of the corresponding confidence interval. And all the dashed and dotted-dashed curve are estimated by ignoring and considering the within-subject correlation, respectively.

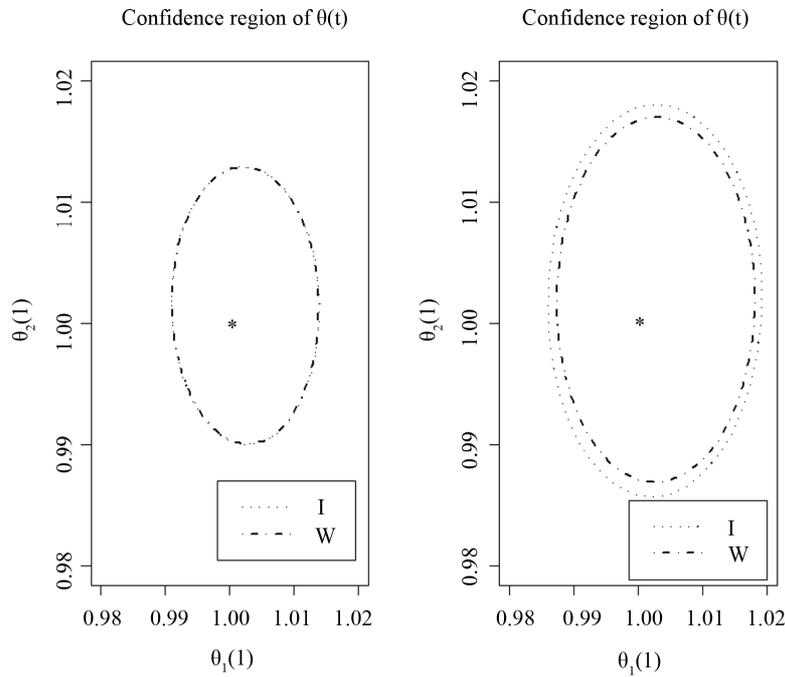


Figure 2. The simultaneous confidence region of the two varying coefficient functions at point $t = 1$, with dashed and dotted-dashed curve being estimated by ignoring and considering the within-subject correlation, respectively. The left and right panel are for $(\sigma_\alpha, \sigma_\epsilon) = (0.1, 0.1)$ and $(\sigma_\alpha, \sigma_\epsilon) = (0.2, 0.1)$, respectively.

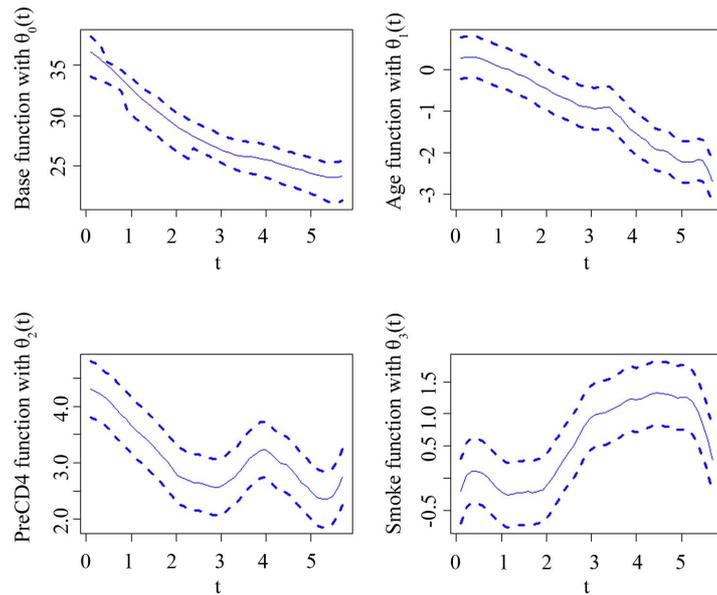


Figure 3. The four plots show the changing curve and their corresponding 95% confidence interval for the different covariate variables as baseline curve, Age, PreCD4, and Smoke status.

function in **Figure 3**, we can find that the estimated curve of the mean CD4 percentage depletion over time also indicate that after getting infected, the CD4 counts decreases sharply at the first 4 years and then the decreasing rate becomes slower although it sometimes changes a little, which is similar to the arguments in [3,

17,18]. Moreover, other estimated curves in **Figure 3** suggest the dependence of CD4 percentage on other variable, such as age, preCD4 and smoke status. Specially, the curve about the variable preCD4 is similar to that shown in Figure 4 in [4], although small nonlinear changing tendency was found in [3].

7. Acknowledgements

This work was partially supported by National Natural Science Foundation of China (11171012, Key program: 11331011), Science and Technology Project for the Supervisor of Excellent Doctoral Dissertation of Beijing (20111000503), Specialized Research Fund for the Doctoral Program of Higher Education of China (20121103110004) and the Beijing Municipal Key Disciplines (No.006000541212010).

REFERENCES

- [1] C. T. Chiang, J. A. Rice and C. O. Wu, "Smoothing Spline Estimation for Varying Coefficient Models with Repeatedly Measured Dependent Variables," *Journal of the American Statistical Association*, Vol. 96, No. 454, 2001, pp. 605-619.
- [2] A. Qu and R. Li, "Quadratic Inference Functions for Varying Coefficient Models with Longitudinal Data," *Biometrics*, Vol. 62, No. 2, 2006, pp. 379-391. <http://dx.doi.org/10.1111/j.1541-0420.2005.00490.x>
- [3] L. G. Xue and L. X. Zhu, "Empirical Likelihood for a Varying Coefficient Model with Longitudinal Data," *Journal of the American Statistical Association*, Vol. 102, No. 478, 2007, pp. 642-652. <http://dx.doi.org/10.1198/016214507000000293>
- [4] H. J. Wang, Z. Zhu and J. Zhou, "Quantile Regression in Partially Linear Varying Coefficient Models," *Annals of Statistics*, Vol. 37, No. 6B, 2009, pp. 3841-3866.
- [5] R. A. Kaslow, D. G. Ostrow, R. Detels, J. P. Phair, B. F. Polk and C. R. Rinaldo, "The Multicenter AIDS Cohort Study: Rationale, Organization and Selected Characteristics of the Participants," *American Journal of Epidemiology*, Vol. 126, No. 2, 1987, pp. 310-318. <http://dx.doi.org/10.1093/aje/126.2.310>
- [6] Q. Li and A. Ullah, "Estimating Partially Linear Panel Data Models with One-way Error Components," *Economic Reviews*, Vol. 17, No. 2, 1998, pp. 145-166. <http://dx.doi.org/10.1080/07474939808800409>
- [7] C. Gu and P. Ma, "Optimal Smoothing in Nonparametric Mixed Effect Models," *Annals of Statistics*, Vol. 33, No. 3, 2005, pp. 1357-1379. <http://dx.doi.org/10.1214/009053605000000110>
- [8] J. You, X. Zhou and Y. Zhou, "Statistical Inference for Panel Data Semiparametric Partially Linear Regression Models with Heteroscedastic Errors," *Journal of Multivariate Analysis*, Vol. 101, No. 5, 2010, pp. 1079-1101. <http://dx.doi.org/10.1016/j.jmva.2010.01.003>
- [9] Z. Pang and L. G. Xue, "Estimation for the Single-index Models with Random Effects," *Computational Statistics & Data Analysis*, Vol. 56, No. 6, 2012, pp. 1837-1853. <http://dx.doi.org/10.1016/j.csda.2011.11.007>
- [10] A. B. Owen, "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, Vol. 75, No. 2, 1988, pp. 237-249. <http://dx.doi.org/10.1093/biomet/75.2.237>
- [11] A. Owen, "Empirical Likelihood Ratio Confidence Regions," *Annals of Statistics*, Vol. 18, No. 1, 1990, pp. 90-120. <http://dx.doi.org/10.1214/aos/1176347494>
- [12] A. Owen, "Empirical Likelihood for Linear Models," *Annals of Statistics*, Vol. 19, No. 4, 1991, pp. 1725-1747. <http://dx.doi.org/10.1214/aos/1176348368>
- [13] P. Zhao and L. Xue, "Empirical Likelihood Inferences for Semiparametric Varying Coefficient Partially Linear Errors in Variables Models with Longitudinal Data," *Journal of Nonparametric Statistics*, Vol. 21, No. 7, 2009, pp. 907-923. <http://dx.doi.org/10.1080/10485250902980576>
- [14] G. R. Li, P. Tian and L. G. Xue, "Generalized Empirical Likelihood Inference in Semiparametric Regression Model for Longitudinal Data," *Acta Mathematica Sinica*, Vol. 24, No. 12, 2008, pp. 2029-2040. <http://dx.doi.org/10.1007/s10114-008-6434-7>
- [15] K. Y. Liang and S. L. Zeger, "Longitudinal Data Analysis Using Generalised Linear Models," *Biometrika*, Vol. 73, No. 1, 1986, pp. 12-22. <http://dx.doi.org/10.1093/biomet/73.1.13>
- [16] J. You, G. Chen and Y. Zhou, "Block Empirical Likelihood for Longitudinal Partially Linear Regression Models," *Canadian Journal of Statistics*, Vol. 34, No. 1, 2006, pp. 79-96. <http://dx.doi.org/10.1002/cjs.5550340107>
- [17] P. Zhao and L. Xue, "Variable Selection for Semiparametric Varying Coefficient Partially Linear Errors in Variables Models," *Journal of Multivariate Analysis*, Vol. 101, No. 8, 2010, pp. 1872-1883. <http://dx.doi.org/10.1016/j.jmva.2010.03.005>
- [18] J. Fan and R. Li, "New Estimation and Model Selection Procedures for Semiparametric Modeling in Longitudinal Data Analysis," *Journal of the American Statistical Association*, Vol. 99, No. 467, 2004, pp. 710-723. <http://dx.doi.org/10.1198/016214504000001060>
- [19] H. G. Müller and J. M. Chiou, "Nonparametric Quasi-Likelihood," *Annals of Statistics*, Vol. 27, No. 1, 1999, pp. 36-64. <http://dx.doi.org/10.1214/aos/1018031100>
- [20] D. A. Harville, "Matrix Algebra from a Statistician's Perspective," Springer, New York, 1997. <http://dx.doi.org/10.1007/b98818>
- [21] G. A. F. Seber, "A Matrix Handbook for Statisticians," John Wiley & Sons, Hoboken, 2007. <http://dx.doi.org/10.1002/9780470226797>
- [22] Y. Li, "Efficient Semiparametric Regression for Longitudinal Data with Nonparametric Covariance Estimation," *Biometrika*, Vol. 98, No. 2, 2011, pp. 355-370. <http://dx.doi.org/10.1093/biomet/asq080>

Appendix

Condition 1. The bandwidth satisfies $h = O(n^{-1/5})$.

Condition 2. The kernel $K(\cdot)$, a symmetric probability density function, is twice continuously differentiable at t_0 and satisfies $\int u^4 K(u) du < \infty$.

Condition 3. The intensity $f(t)$ of covariate variable T_{ij} is bounded away from 0 and infinity on $[0,1]$, and is continuously differentiable on $(0,1)$.

Condition 4. $\theta(t)$, $\Psi(t)$ are twice continuously differentiable on $(0,1)$, where

$$\Psi(t) = E \left[\left\{ X - E(X|T=t) \right\} \left\{ X - E(X|T=t) \right\}^T \right]$$

Condition 5. $E(\varepsilon^4(t)|t)$ and $E(X_r^4(t)|t)$ are twice continuous with t , and $\sup_{0 \leq t \leq 1} E(\varepsilon^4(t)|t) < \infty$, $\sup_{0 \leq t \leq 1} E(X_r^4(t)|t) < \infty$, where $X_r^4(t)$ is r th component of $X(t)$.

Condition 6. For given t , $\Psi(t)$ is positive definite matrix.

Condition 7. There exist two positive constants κ_1 and κ_2 , such that

$$0 < \kappa_1 \leq \min_{1 \leq i \leq n} \lambda_{r1} \leq \max_{1 \leq i \leq n} \lambda_{im} \leq \kappa_2 < \infty,$$

where λ_{r1} and λ_{im} denote to be the smallest and largest eigenvalues of Ω_i , respectively.

Lemma 1. Suppose that conditions (C1)-(C6) hold, $\tilde{\theta}(t)$ denote to be the preliminary estimators solved by

$$\begin{aligned} \frac{1}{\sqrt{nh}} \sum_{i=1}^n I_{i2} &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n X_i^T W_i \Omega_i^{-1} X_i \left\{ \theta(t_i) - \theta(t) - (\tilde{\theta}(t_i) - \tilde{\theta}(t)) \right\} \\ &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n X_{ij} \sigma_i^{jk} K \left(\frac{t - t_{ik}}{h} \right) X_{ik}^T \left\{ \theta(t_{ik}) - \theta(t) - (\tilde{\theta}(t_{ik}) - \tilde{\theta}(t)) \right\}, \end{aligned}$$

With several calculation and the Central Limit Theorem,

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^n I_{i1} \xrightarrow{D} N(0, B), \quad (\text{A.3})$$

and $\frac{1}{\sqrt{nh}} \sum_{i=1}^n I_{i2} \xrightarrow{P} 0$.

Therefore, Lemma 4 can be derived directly.

Lemma 5. Assume the conditions (C1)-(C7) hold, and $\theta(t)$ is the true parameter, then

$$\frac{1}{nh} \sum_{i=1}^n \tilde{\eta}_i(\theta(t)) \tilde{\eta}_i(\theta(t))^T \xrightarrow{P} B.$$

Proof. By the proof of Lemma 4, we can derive that

the estimation Equation (19), then we have

$$\sup_{t \in T} \|\tilde{\theta}(t) - \theta(t)\| = O_p \left\{ h^2 + \left(\frac{\log n}{nh} \right)^{1/2} \right\}. \quad (\text{A.1})$$

Proof. See the proof of Lemma 4.1 in [19].

Lemma 2. Let A_n be a sequence of random matrices converging to an invertible matrix A . Then

$$A_n^{-1} = A^{-1} - A^{-1}(A_n - A)A^{-1} + Op(\|A_n - A\|), \quad (\text{A.2})$$

where $\|A\|^2 = \text{tr}(A^T A)^{1/2}$.

Proof. See the proof in [20-22].

Lemma 4. Assume the conditions (C1)-(C7) hold, and $\theta(t)$ is the true parameter, then

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^n \tilde{\eta}_i(\theta(t)) \xrightarrow{D} N(0, B).$$

Proof. $\frac{1}{\sqrt{nh}} \sum_{i=1}^n \tilde{\eta}_i(\theta(t)) = \frac{1}{\sqrt{nh}} \sum_{i=1}^n I_{i1} + \frac{1}{\sqrt{nh}} \sum_{i=1}^n I_{i2}$,

where

$$\begin{aligned} \frac{1}{\sqrt{nh}} \sum_{i=1}^n I_{i1} &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n X_i^T W_i \Omega_i^{-1} \varepsilon_i \\ &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n X_{ij} \sigma_i^{jk} \varepsilon_{ik} K \left(\frac{t - t_{ik}}{h} \right), \end{aligned}$$

$$\begin{aligned} \frac{1}{\sqrt{nh}} \sum_{i=1}^n I_{i2} &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n X_i^T W_i \Omega_i^{-1} \varepsilon_i \\ &= \frac{1}{\sqrt{nh}} \sum_{i=1}^n X_{ij} \sigma_i^{jk} \varepsilon_{ik} K \left(\frac{t - t_{ik}}{h} \right), \end{aligned}$$

$$\begin{aligned} &\frac{1}{nh} \sum_{i=1}^n \tilde{\eta}_i(\theta(t)) \tilde{\eta}_i(\theta(t))^T \\ &= \frac{1}{nh} \sum_{i=1}^n I_{i1} I_{i1}^T + \frac{1}{nh} \sum_{i=1}^n I_{i2} I_{i1}^T + \frac{1}{nh} \sum_{i=1}^n I_{i1} I_{i2}^T + \frac{1}{nh} \sum_{i=1}^n I_{i2} I_{i2}^T \\ &= J_1 + J_2 + J_3 + J_4 \end{aligned}$$

According to the proof of Lemma 4, we know that $\frac{1}{nh} \sum_{i=1}^n I_{i1}^2 = O_p(1)$ and $\frac{1}{nh} \sum_{i=1}^n I_{i2}^2 = O_p(1)$ which are the r th or s th component of I_{i1} and I_{i2} . Some simple algebra calculation leads to that $J_2 \xrightarrow{P} 0$. And we can derive $J_l \xrightarrow{P} 0$, $l = 3, 4$. Moreover, by the law of large numbers, we can derive that $J_1 \xrightarrow{P} B$. So, the

proof of Lemma 5 is completed.

Proof of Theorem 1. By the definition, $\hat{\sigma}_\alpha^2$ can be written a

$$\begin{aligned} \hat{\sigma}_\alpha^2 &= \frac{1}{nm(m-1)} \sum_{i=1}^n \sum_{j_1=1}^m \sum_{j_2 \neq j_1} \hat{\varepsilon}_{ij_1} \hat{\varepsilon}_{ij_2} \\ &= \frac{1}{nm(m-1)} \sum_{i=1}^n \sum_{j_1=1}^m \sum_{j_2 \neq j_1} \varepsilon_{ij_1} \varepsilon_{ij_2} \\ &\quad + \frac{2}{nm(m-1)} \sum_{i=1}^n \sum_{j_1=1}^m \sum_{j_2 \neq j_1} \left[X_{ij_1}^\top \{ \theta(t_{j_1}) - \tilde{\theta}(t_{j_1}) \} \right] \varepsilon_{ij_2} \\ &\quad + \frac{1}{nm(m-1)} \sum_{i=1}^n \sum_{j_1=1}^m \sum_{j_2 \neq j_1} \left[X_{ij_1}^\top \{ \theta(t_{j_1}) - \tilde{\theta}(t_{j_1}) \} \right] \\ &\quad \times \left[X_{ij_2}^\top \{ \theta(t_{j_2}) - \tilde{\theta}(t_{j_2}) \} \right] \\ &= I_1 + I_2 + I_3. \end{aligned}$$

Using the Taylor formula and Lemma 1, we can derive

$$I_2 = o_p \left(n^{-\frac{1}{2}} \right) \text{ and } I_3 = o_p \left(n^{-\frac{1}{2}} \right).$$

Moreover, by the Lyapunov central limit theorem and that proof of Theorem 4 in [8], we can derive (14) in Theorem 1.

By the proof of the first part, we can easily prove (15) in Theorem 1.

Proof of Theorem 2. Here, we mainly provide the proof procedure by showing the evidence about the asymptotic equivalence between the auxiliary random vector (4) and that one (12) with the within-subject covariance being replaced by the estimator (11). That is the error caused by the use of plug-in estimation is negligible.

For the given estimator (11) for the within-subject covariance, the auxiliary random vector is

$$\begin{aligned} &X_i^\top W_i \hat{\Omega}_i^{-1} \left[Y_i - X_i \theta(t) - X_i \{ \tilde{\theta}(t_i) - \tilde{\theta}(t) \} \right] \\ &= X_i^\top W_i \Omega_i^{-1} \left[Y_i - X_i \theta(t) - X_i \{ \tilde{\theta}(t_i) - \tilde{\theta}(t) \} \right] \\ &\quad + X_i^\top W_i \Omega_i^{-1} \left(\hat{\Omega}_i^{-1} - \Omega_i^{-1} \right) \Omega_i^{-1} \left[Y_i - X_i \theta(t) \right. \\ &\quad \left. - X_i \{ \tilde{\theta}(t_i) - \tilde{\theta}(t) \} \right] \times \{ 1 + o_p(1) \}. \end{aligned} \tag{A.4}$$

The second term in (A.4) is high order of $o_p(1)$. Therefore, the empirical likelihood ratio based on the auxiliary random vector (12) with the estimated within subject covariance is asymptotic equivalent to that one (4) with a true one. By a Taylor expansion of (13) and following a similar lines as in the proof of Theorem (3.2) in

[3], we can show that $\hat{l}(\theta(t))$ is asymptotic equivalent to that one (5) with the true covariance

$$\begin{aligned} \tilde{l}(\theta(t)) &= -2 \max \left\{ \sum_{i=1}^n \log(np_i) \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \right. \\ &\quad \left. \sum_{i=1}^n p_i \eta_i(\theta(t)) = 0 \right\}. \end{aligned}$$

By the arguments in the proof of (2.14) in [11] and together with Lemma 3, we can derive that

$$\lambda = O \left((nh)^{\frac{1}{2}} \right), \tag{A.5}$$

where λ is defined by (7).

By Lemma 3-5 and (A.5), a Taylor expansion of (8) leads to

$$\tilde{l}(\theta(t)) = -2 \sum_{i=1}^n \left[\lambda^\top \tilde{\eta}_i(\theta(t)) - \{ \lambda^\top \tilde{\eta}_i(\theta(t)) \}^2 / 2 \right] + o_p(1). \tag{A.6}$$

Then by (7), it follows that

$$\begin{aligned} 0 &= \sum_{i=1}^n \frac{\tilde{\eta}_i(\theta(t))}{1 + \lambda^\top \tilde{\eta}_i(\theta(t))} \\ &= \sum_{i=1}^n \tilde{\eta}_i(\theta(t)) - \sum_{i=1}^n \tilde{\eta}_i(\theta(t)) \tilde{\eta}_i^\top(\theta(t)) \lambda \\ &\quad + \sum_{i=1}^n \frac{\tilde{\eta}_i(\theta(t)) \{ \lambda^\top \tilde{\eta}_i(\theta(t)) \}^2}{1 + \lambda^\top \tilde{\eta}_i(\theta(t))} \end{aligned} \tag{A.7}$$

Hence, by Lemma 3-5, and (A.7), we can derive that

$$\sum_{i=1}^n \{ \lambda^\top \tilde{\eta}_i(\theta(t)) \}^2 = \sum_{i=1}^n \lambda^\top \tilde{\eta}_i(\theta(t)) + o_p(1), \tag{A.8}$$

$$\lambda = \left\{ \sum_{i=1}^n \tilde{\eta}_i(\theta(t)) \tilde{\eta}_i^\top(\theta(t)) \right\}^{-1} \sum_{i=1}^n \tilde{\eta}_i(\theta(t)) + o(n^{-1/2}). \tag{A.9}$$

With the plug-in of (A.8) and (A.9) in (A.7), we can further derive that

$$\tilde{l}(\theta(t)) = \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\eta}_i(\theta(t)) \right\}^{-1} B^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\eta}_i(\theta(t)) \right\}. \tag{A.10}$$

Therefore, the asymptotic result of Theorem 2 follows from Lemma 4 and 5.