Journal of Computational and Applied Mathematics 321 (2017) 406–415

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

QR decomposition based orthogonality estimation for partially linear models with longitudinal data

Jiting Huang [a], Peixin Zhao [b,*]

[a] College of Mathematics and Statistics, Hechi University, Yizhou, Guangxi 546300, China

[b] College of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China

A R T I C L E I N F O

Article history:
Received 21 December 2016

MSC:
62G05
62G20
62G30

Keywords:
Partially linear model
Longitudinal data
Orthogonality estimation
QR decomposition

A B S T R A C T

This paper studies the estimation for a class of partially linear models with longitudinal data. By combining quadratic inference functions with QR decomposition technology, we propose a new estimation method for the parametric and nonparametric components. The resulting estimators for parametric and nonparametric components do not affect each other, and then it is easy for application in practice. Under some mild conditions, we establish some asymptotic properties of the resulting estimators. Some simulation studies are undertaken to assess the finite sample performance of the proposed estimation procedure.

1. Introduction

In practice, longitudinal data arise frequently in biomedical research where data are observed at irregular and possibly subject-specific time points. In order to explore possible dependent effects, partially linear models have been proposed for longitudinal data analysis. For example, Fan and Li [1] proposed a difference based method and a profile least squares method for estimating the regression coefficients. Xue and Zhu [2] proposed an empirical likelihood method for estimating the regression coefficients. Bai et al. [3] proposed a quadratic inference function based estimation method to obtain an efficient estimator for the parametric components. More works on inferences of partially linear models with longitudinal data can be found in Lin and Carroll [4], He et al. [5], Hu et al. [6], and among others. In most of the literature listed above, the estimation procedure of parametric component was obtained by involving the information of nonparametric components, such as the profile least squares estimation method proposed in [1] and the empirical likelihood method proposed in [2]. This motivates us to develop a new estimation method, which can separately estimate the parametric components and nonparametric components, and spends comparable computational cost.

More specifically, we propose an orthogonality estimation approach using QR decomposition technology and quadratic inference functions. The proposed estimation procedure allows us to directly incorporate correlation information of longitudinal data into model building for improving efficiency of regression coefficients. Furthermore, the proposed procedure can easily be implemented, and the resulting estimators of parametric and nonparametric components do not affect each other. Under some conditions, we investigate some asymptotic properties of the proposed estimators. In addition, to demonstrate the finite sample performance of the proposed estimator method, some simulation studies are conducted.

* Corresponding author.
E-mail address: zpx81@163.com (P. Zhao).

http://dx.doi.org/10.1016/j.cam.2017.02.024

Orthogonality based estimation is an efficient statistical inference method for semiparametric regression models. This estimation method can individually estimate the parametric and nonparametric components without losing efficiency, and has some excellent properties. Recently, Zhao et al. [7] proposed an orthogonality based estimation for semiparametric varying coefficient partially linear models with heteroscedastic errors. Wu and Zhu [8] proposed an orthogonality based estimation of moments for linear mixed models. Yang and Yang [9] proposed a statistical inference procedure for varying coefficient partially nonlinear models based on the orthogonality projection method and the smooth threshold estimating equations. In this article, we provide a new orthogonality estimation method based on QR decomposition. This is an additional positive result for the orthogonality technology in the semiparametric regression modeling for longitudinal data analysis, which extends the application of the orthogonality estimation method.

In addition, the quadratic inference functions (QIF) method, proposed in [10], can avoid estimating the nuisance correlation structure parameters by assuming that the inverse of the working correlation matrix can be approximated by a linear combination of several known basis matrices. The OIF method can efficiently take the within-cluster correlation into account and is more efficient than the GEE method if the working correlation is misspecified. Recently, Qu and Li [11] applied the QIF method to varying coefficient models with longitudinal data. Bai et al. [3] extended the QIF method to the semiparametric partial linear model. Dziak et al. [12] gave an overview on QIF method for longitudinal data. The more works on QIF method can be found in [13–15], and among others.

The rest of this paper is organized as follows. In Section 2, based on the quadratic inference functions and the QR decomposition technology, we propose an orthogonality estimation procedure for the parametric and nonparametric components, and present some asymptotic properties of the resulting estimators. In Section 3, we present some simulation studies to assess the finite sample performances of the proposed method. The technical proofs of all asymptotic results are provided in Section 4.

2. Estimation procedure and main results

Consider a longitudinal study with n subjects and m_i observations over time for the ith subject ($i = 1, \dots, n$). For each observation, we denote Y_{ij} is a response, $X_{ij} \in R^p$ and $U_{ij} \in R$ are covariates. Then, the partially linear model with longitudinal data has the following structure

$$Y_{ij} = X_{ij}^T\beta + \theta(U_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, n,\ j = 1, \dots, m_i, \tag{1}$$

where $\beta = (\beta_1, \dots, \beta_p)$ is a vector of unknown parameters, $\theta(u)$ is an unknown function of u, and ε_{ij} is the model error with $E(\varepsilon_{ij}|X_{ij}, U_{ij}) = 0$. In addition, we assume that covariate U_{ij} ranges over a nondegenerate compact interval, without loss of generality, which is assumed to be the unit interval [0, 1].

From Schumaker [16], it is known that the nonparametric function $\theta(u)$ can be approximated by some basis functions. More specifically, let $B(u) = (B_1(u), \dots, B_L(u))^T$ be B-spline basis functions with the order of M, where $L = K + M$, and K is the number of interior knots. Then, $\theta(u)$ can be approximated by

$$\theta(u) \approx B(u)^T\gamma, \tag{2}$$

where $\gamma = (\gamma_1, \dots, \gamma_L)^T$ is a vector of basis functions coefficients. For the sake of simplicity, we introduce some matrix notations. We denote $Z_{ij} = B(U_{ij}) = (B_1(U_{ij}), \dots, B_L(U_{ij}))^T, Z_i = (Z_{i1}, \dots, Z_{im_i})^T, X_i = (X_{i1}, \dots, X_{im_i})^T, Y_i = (Y_{i1}, \dots, Y_{im_i})^T$, and $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{im_i})^T$. Then, model (1) can be rewritten as

$$Y_i \approx X_i\beta + Z_i\gamma + \varepsilon_i, \quad i = 1, \dots, n. \tag{3}$$

Furthermore, we assume that Z_i, $i = 1, \dots, n$ are all column full rank matrices. Recalling the definition of the QR decomposition of a matrix, the matrix Z_i can be decomposed as

$$Z_i = Q_i \begin{pmatrix} R_i \\ \mathbf{0} \end{pmatrix},$$

where Q_i is a $m_i \times m_i$ orthogonal matrix, R_i is a $L \times L$ triangular matrix, and $\mathbf{0}$ is a $(m_i - L) \times L$ zero matrix. We further partition Q_i as $Q_i = (Q_{i1}, Q_{i2})$, where Q_{i1} is a $m_i \times L$ matrix, and Q_{i2} is a $m_i \times (m_i - L)$ matrix. It is easy to obtain that $Z_i = Q_{i1}R_i$ and $Q_{i2}^TQ_{i1} = 0$. Then we have $Q_{i2}^TZ_i = Q_{i2}^TQ_{i1}R_i = 0$. Based on this information, and left multiplying the both sides of Eq. (3) by Q_{i2}^T yields

$$Q_{i2}^TY_i = Q_{i2}^TX_i\beta + Q_{i2}^T\varepsilon_i, \quad i = 1, \dots, n. \tag{4}$$

Note that model (4) is constructed in the orthogonal column space of Z_i, and does not depend on the nonparametric components. Then, invoking model (4), the generalized estimating equations for the parametric components β can be defined as follows

$$\sum_{i=1}^{n} X_i^TQ_{i2}V_i^{-1}Q_{i2}^T(Y_i - X_i\beta) = 0, \tag{5}$$