Statistics and Probability Letters 85 (2014) 6–14

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Simultaneous confidence band for single-index random effects models with longitudinal data

Suigen Yang[a,b,*], Liugen Xue[a], Gaorong Li[a]

[a] College of Applied Sciences, Beijing University of Technology, Beijing 100124, PR China
[b] College of Sciences, Tianjin University of Commerce, Tianjin 300134, PR China

ARTICLE INFO

Article history:
Received 16 April 2013
Received in revised form 19 October 2013
Accepted 19 October 2013
Available online 28 October 2013

Keywords:
Simultaneous confidence band
Single-index model
Random effects
Longitudinal data

ABSTRACT

We get the estimator of the link function, establish the asymptotic properties, and construct the simultaneous confidence band for single-index random effects models. Simulation studies and real data set are presented to evaluate the performance of the proposed method.

1. Introduction

With the increasing availability of longitudinal data, both theoretical and applied works in longitudinal data analysis have become more popular in recent years. Diggle et al. (2002) provided an excellent overview of the longitudinal data analysis. To avoid the so-called "curse of dimensionality" in the multivariate nonparametric regression with longitudinal data and to generate an association correlation structure between the repeated measurements, Pang and Xue (2012) considered the following single-index random effects models with longitudinal data:

$$Y_{ij} = g(X_{ij}^T\boldsymbol{\beta}_0) + Z_{ij}^T b_i + \varepsilon_{ij}, \quad i = 1, \ldots, n,\ j = 1, \ldots, m, \tag{1.1}$$

where $\boldsymbol{\beta}_0$ is a $p \times 1$ index coefficients vector associated with the covariates X_{ij}, b_i are independent $q \times 1$ vectors of random effects with mean zero and covariance matrix Φ, $g(\cdot)$ is an unknown link function, ε_{ij}'s are independent mean zero random variables with variance $\sigma_\varepsilon^2 > 0$. Here Φ is a positive definite matrix depending on a parameter vector ϕ; X_{ij} and Y_{ij} are the observable random variables, and Z_{ij} are $q \times 1$ known fixed design vector. We suppose that b_i and ε_{ij} are mutually independent and follow Gaussian distribution, and $\|\boldsymbol{\beta}_0\| = 1$ with the first nonzero element of $\boldsymbol{\beta}_0$ being positive to ensure identifiability.

Since the single-index models are popular and efficient modeling tools in multivariate nonparametric regression, the single-index models have recently received much attention, including that from Härdle et al. (1993), Carroll et al. (1997), Xia and Li (1999), Xia et al. (2002, 2004), Yu and Ruppert (2002), Zhu and Xue (2006), Wang et al. (2010), Liang et al. (2010), Cui et al. (2011) for cross-sectional data, and from Bai et al. (2009), Li et al. (2010), Pang and Xue (2012), Lai et al. (2012), Chen et al. (2013) for longitudinal/panel data. Further, random effects models have become very popular for the analysis of longitudinal or panel data, because they are flexible and widely applicable. Given the importance of the random effects models, it is not surprising that methodologies for random effects models have emerged in the extensive literature studies,

* Corresponding author at: College of Applied Sciences, Beijing University of Technology, Beijing 100124, PR China. Tel.: +86 18920068139.
E-mail address: yangsuigen@emails.bjut.edu.cn (S. Yang).

0167-7152/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spl.2013.10.014

such as Zeger and Diggle (1994), Ruckstuhl et al. (2000), Ke and Wang (2001), Wu and Zhang (2002), Hall and Maiti (2006), Jiang (2007) and Field et al. (2008), and among others. However, it has a lot of challenges for the studies and the applications of single-index models with longitudinal data when there exist the random effects in the models. Pang and Xue (2012) proposed an iterative estimation procedure to estimate the index parameter vector and the link function, and they proved the asymptotic properties of the resulting estimators.

Nonparametric simultaneous confidence band (SCB) is a powerful tool of global inference for functions, and one of the main motivations to construct the SCB is able to check the graphical representation of the nonparametric function in nonparametric or semiparametric regression models. Therefore, there is an extensive literature on the construction of the SCB; see, for example, Bickel and Rosenblatt (1973), Härdle and Bowman (1988), Sun and Loader (1994), Faraway and Sun (1995), Neumann and Polzehl (1998), Claeskens and Van Keilegom (2003), Li et al. (2013), and among others. Recently, Fan and Zhang (2000) and Zhang and Peng (2010) considered the simultaneous confidence bands (SCBs) for the coefficient functions in varying-coefficient models. Zhang et al. (2009) studied the semiparametric model with cluster data by accounting for within-cluster correlation. Krivobokova et al. (2010) constructed the SCBs for univariate smooth curves based on penalized spline estimators. Li et al. (in press) considered the SCB and hypothesis testing in single-index models under the independent data setting.

For the single-index random effects model (1.1), the fact that the uniform confidence bands for the link function and its derivative have not been established is certainly not due to the lack of interesting applications, but due to the greater technical difficulty in constructing such SCBs and establishing their theoretical properties, such as the introduction of the random effects and the estimation of the index parameter vector raise inferential challenges for the construction of SCBs. In addition, the variance estimation is often complicated for the semiparametric random effects models. In this paper, based on the estimation idea of Pang and Xue (2012), we first choose a suitable initial estimator of $\boldsymbol{\beta}_0$, and then estimate the link function and the index parameter vector using the iterative estimation procedure. The SCB for link function is obtained, and the asymptotic distributions of the maximum absolute deviation between the resulting estimators and the true link function and its derivative are established when there exists a $\sqrt{n}$-consistent estimator for the index parameter $\boldsymbol{\beta}_0$.

The paper is organized as follows. In Section 2, the iterative estimation procedure is given for model (1.1), and the asymptotic properties of the proposed estimator are established. The results can be used to construct the SCB for the link function. In Section 3 simulation studies are conducted to evaluate the performance of the proposed method, and a real dataset is analyzed to illustrate the proposed method. We conclude the paper in Section 4 with some remarks and present the technical proofs in the Supplementary materials.

2. Estimation methods and asymptotic properties

2.1. Estimation procedure

Suppose that the sample $\{(Y_{ij}, X_{ij}), i = 1, \ldots, n, j = 1, \ldots, m\}$ comes from model (1.1). Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im})^T$, $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{im})^T$, $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{im})^T$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im})^T$ and $\boldsymbol{G}(\boldsymbol{X}_i\boldsymbol{\beta}_0) = (g(X_{i1}^T\boldsymbol{\beta}_0), \ldots, g(X_{im}^T\boldsymbol{\beta}_0))^T$. Model (1.1) can be rewritten as

$$\boldsymbol{Y}_i = \boldsymbol{G}(\boldsymbol{X}_i\boldsymbol{\beta}_0) + \boldsymbol{Z}_i b_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, n. \tag{2.1}$$

It is easy to see that $E(\boldsymbol{Y}_i|\boldsymbol{X}_i) = \boldsymbol{G}(\boldsymbol{X}_i\boldsymbol{\beta}_0)$ and $\mathrm{Cov}(\boldsymbol{Y}_i|\boldsymbol{X}_i) = \boldsymbol{Z}_i\Phi\boldsymbol{Z}_i^T + \sigma_\varepsilon^2 I_m$, where I_m is the $m \times m$ identity matrix and $\boldsymbol{Z}_i$ is $m \times q$ known fixed design matrix.

Since $\|\boldsymbol{\beta}_0\| = 1$ means that the true value of $\boldsymbol{\beta}_0$ is the boundary point on the unit sphere, $g(X_{ij}^T\boldsymbol{\beta}_0)$ does not have derivative at the point $\boldsymbol{\beta}_0$. However, we must use the derivative of $g(X_{ij}^T\boldsymbol{\beta}_0)$ on $\boldsymbol{\beta}_0$ when constructing the estimating equation for $\boldsymbol{\beta}_0$. The "delete-one-component" method (Zhu and Xue, 2006; Wang et al., 2010) is used to solve this problem. The detail is as follows. Let $\boldsymbol{\beta}_0 = (\beta_1, \ldots, \beta_p)^T$ and $\boldsymbol{\beta}^{(r)} = (\beta_1, \ldots, \beta_{r-1}, \beta_{r+1}, \ldots, \beta_p)^T$ be a $p-1$ dimensional parameter vector deleting the rth component β_r. Without loss of generality, we may assume that the true vector $\boldsymbol{\beta}_0$ has a positive component β_r. Then, the true parameter $\boldsymbol{\beta}^{(r)}$ satisfies the constraint $\|\boldsymbol{\beta}^{(r)}\| < 1$. Thus, $\boldsymbol{\beta}_0$ is infinitely differentiable in a neighborhood of the true parameter $\boldsymbol{\beta}^{(r)}$, and the Jacobian matrix is

$$J_{\boldsymbol{\beta}^{(r)}} = (\gamma_1, \ldots, \gamma_p)^T,$$

where $\gamma_s (1 \le s \le p, s \ne r)$ is a $(p-1)$-dimensional unit vector with sth component 1, and $\gamma_r = -(1 - \|\boldsymbol{\beta}^{(r)}\|^2)^{-1/2}\boldsymbol{\beta}^{(r)}$.

Based on the estimation procedure in Pang and Xue (2012), we outline the iterative steps for estimating procedures for $\boldsymbol{\beta}_0$, $g(\cdot)$ and its derivative $g'(\cdot)$.

Step 0: We first give a consistent estimator of $\boldsymbol{\beta}_0$, which is denoted by $\widetilde{\boldsymbol{\beta}}$.

Step 1: Estimation of the link function g and its derivative g'. Given the initial estimator $\widetilde{\boldsymbol{\beta}}$, we apply the local linear regression technique in Fan and Gijbels (1996) to estimate the link function g and its derivative g'. The estimators of g and g' are obtained by minimizing the weighted sum of squares

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\{Y_{ij} - a - b(X_{ij}^T\widetilde{\boldsymbol{\beta}} - u)\}^2 K_h(X_{ij}^T\widetilde{\boldsymbol{\beta}} - u) \tag{2.2}$$