Journal of Multivariate Analysis 127 (2014) 1–18

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Empirical likelihood of varying coefficient errors-in-variables models with longitudinal data

Yiping Yang [a], Gaorong Li [b], Heng Peng [c,*]

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

[b] College of Applied Sciences, Beijing University of Technology, Beijing 100124, PR China

[c] Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

ARTICLE INFO

Article history:
Received 22 July 2012
Available online 15 February 2014

AMS 2000 subject classifications:
primary 62G05
62G08
secondary 62G20

Keywords:
Varying coefficients model
Errors-in-variables
Longitudinal data
Empirical likelihood
Maximum empirical likelihood estimator

ABSTRACT

In this paper, we investigate the empirical likelihood inferences of varying coefficient errors-in-variables models with longitudinal data. The naive empirical log-likelihood ratios for the time-varying coefficient function based on the global and local variance structures are introduced. The corresponding maximum empirical likelihood estimators of the time-varying coefficients are derived, and their asymptotic properties are established. Wilks' phenomenon of the naive empirical log-likelihood ratio, which ignores the within subject correlation, is proven through the employment of undersmoothing. To avoid the undersmoothing, we recommend a residual-adjust empirical log-likelihood ratio and prove that its asymptotic distribution is standard chi-squared. Thus, this result can be used to construct the confidence regions of the time-varying coefficients. We also establish the asymptotic distribution theory for the corresponding residual-adjust maximum empirical likelihood estimator and find it to be unbiased even when an optimal bandwidth is used. Furthermore, we consider the construction of the pointwise confidence interval for a component of the time-varying coefficients and provide the simulation studies to assess the finite sample performance, while we conduct a real example to illustrate the proposed method.

1. Introduction

Varying coefficient models are often used as extensions of classical linear models (e.g. [23]: 245). Their appeals are that the modelling bias can be significantly reduced and the "curse of dimensionality" can also be avoided. These models have gained considerable attention due to their various applications in many areas, such as biomedical study, finance, econometrics, and environmental study. The estimation for the coefficient functions has been extensively discussed in the literatures, including the smoothing spline method [6], the locally weighted polynomial method [7], the two-step estimation procedure [4], and the basis function approximations [9].

Varying coefficient models are particularly appealing in longitudinal studies because they allow us to determine the extent to which covariates affect responses over time. For longitudinal data, Hoover et al. [7], considered the following varying coefficient model:

$$Y_i(t_{ij}) = X_i^T(t_{ij})\beta(t_{ij}) + \varepsilon_i(t_{ij}), \quad i = 1, \ldots, n, j = 1, \ldots, n_i, \tag{1}$$

* Corresponding author.
E-mail address: hpeng@math.hkbu.edu.hk (H. Peng).

http://dx.doi.org/10.1016/j.jmva.2014.02.004

where, for all $t \in \mathbb{R}$, $X_i(t) = (X_{i1}(t), \ldots, X_{ik}(t))^T$ are real-valued covariates at time t, $\beta(t) = (\beta_1(t), \ldots, \beta_k(t))^T$, $\beta_r(t)$ are smoothing functions of t for all $r = 1, \ldots, k$, $\varepsilon_i(t)$ $(i = 1, \ldots, n)$ are the stochastic processes with mean zero, and $\varepsilon_i(t)$ and $X_i(t)$ are independent. Throughout the remainder of this article, we assume that the designs $\{t_{ij}, i = 1, \ldots, n, j = 1, \ldots, n_i\}$ are independent and identically distributed according to an underlying density $f(\cdot)$, with respect to the Lebesgue measure. Hoover et al. [7] derived two nonparametric estimators of $\beta(t)$ using a class of smoothing splines and the locally weighted polynomial method. Wu et al. [30] obtained a kernel estimator of $\beta(t)$ by minimising a local least-squares criterion, and established the asymptotic normality. They further used the asymptotic distribution to develop approximate pointwise and simultaneous confidence bands for $\beta(t)$. Huang et al. [10] approximated each coefficient function by a polynomial spline and employed the least squares method to estimate the coefficient functions. Xue and Zhu [33] investigated the local empirical likelihood-based inference for $\beta(t)$, and proposed the naive, mean-corrected, and residual-adjusted empirical likelihood ratios. Tang and Cheng [25] discussed M-estimation and B-spline approximation for $\beta(t)$, and derived the asymptotic distributes of M-estimators.

For model (1), the covariates $X_i(t_{ij})$ are not always observable without error. If $X_i(t_{ij})$ is measured with error, instead of observing $X_i(t_{ij})$, we observe

$$W_i(t_{ij}) = X_i(t_{ij}) + U_i(t_{ij}), \tag{2}$$

where U is an independent error with mean zero and covariance matrix Σ_u. If some elements of X are measured without error, the corresponding elements of U and related variance components in Σ_u are set to zero. The varying coefficient model with measurement error is usually called the varying coefficient errors-in-variables (EV) model. It is will known that there exists the issue of identifiability associated with the error-in-variables models (see Fuller [5]). To deal with this problem, we need extra information that Σ_u is known or can be estimated. When Σ_u is unknown, the idea is to make replicate observations, enabling us to estimate Σ_u (see Remark 3). A careful study of such a model is often needed. You et al. [38] and Li and Greene [16] used the locally corrected method to estimate $\beta(t)$. Wang and Zhang [29] considered varying coefficient EV models with surrogate data and validation sampling. The asymptotic normality obtained by You et al. [38] or Li and Greene [16] can be used to construct the confidence regions for $\beta(t)$. Zhang et al. [39] studied the estimation and testing problems of partially linear varying-coefficient EV models under additional restricted conditions. However, for longitudinal data, an asymptotic normal distribution would involve bias. In addition, the asymptotic variance of the estimator has a very complex structure. To avoid taking these issues into account, we recommend using the empirical likelihood method to construct the confidence regions of $\beta(t)$. Xue and Zhu [34] and Zhu and Xue [42] investigated the empirical likelihood confidence regions of the parameters in the single-index model and the partially linear single-index model, respectively. Xue and Zhu [35] investigated the issue of estimation and confidence region construction for the partially linear model with longitudinal data. Cui and Chen [2] studied the empirical likelihood inference for the EV models. Zhao and Xue [40] investigated empirical likelihood inferences for semiparametric varying-coefficient partially linear EV models. Under missing data, Xue [31,32] studied the nonparametric regression model and the linear model. The other related works include Wang and Rao [28], You et al. [36], Stute et al. [24], Li et al. [14], Huang and Zhang [11], Hua et al. [8], Wang et al. [26], You and Zhou [37], Zhou and Liang [41], Li et al. [15], and Li et al. [13], among others.

In this article, we investigate estimation and confidence region construction issues using the empirical likelihood approach. First, we construct a naive empirical log-likelihood ratio (ELR) for $\beta(t)$ and show that its asymptotic distribution is standard chi-squared when undersmoothing is employed. Second, we define the maximum empirical likelihood estimator (MELE) and investigate its asymptotic properties. Third, we provide a modified ELR called the residual-adjusted ELR. The residual-adjusted ELR is asymptotically standard chi-squared. The confidence regions for $\beta(t)$ can be constructed through a standard chi-squared approximation. The corresponding residual-adjust MELE is obtained and the asymptotic normality is established. We also consider the construction of the pointwise confidence band for a component of the time-varying coefficients.

Our method offers the following improvements over existing methods. First, to incorporate serial correction into ELR functions, we introduce the naive ELRs, based on the global and local variance structures. Although Xue and Zhu [33] investigated model (1), they did not include errors in the covariates. When $X(t)$ is measured with the additive error, the method proposed by Xue and Zhu [33] leads to biased estimators. In addition, they did not discuss the correlation with longitudinal data. Second, the proposed method can not only attenuate the effect of bias and measurement errors, but also avoid undersmoothing. As a result of our bias and measurement error correction, the proposed ELR for $\beta(t)$ is asymptotically standard chi-squared, undersmoothing is avoided, and the optimal bandwidth can be used. Third, the MELEs, $\hat{\beta}_{\mathrm{EL}}(t)$ and $\hat{\beta}_{\mathrm{RAEL}}(t)$, are derived by the naive ELR function and the residual-adjusted ELR function, and the asymptotic properties are established. Note that the asymptotic normal distribution of the naive MELE, $\hat{\beta}_{\mathrm{EL}}(t)$, involves bias without undersmoothing. In contrast, the residual-adjusted MELE, $\hat{\beta}_{\mathrm{RAEL}}(t)$, is unbiased if an optimal bandwidth is used. Xue and Zhu [33] only investigated the naive MELE, they did not consider the residual-adjusted MELE for varying coefficient models with longitudinal data. It is worth mentioning that the residual-adjusted MELE is also unbiased without undersmoothing when the covariates are measured without error; that is, Σ_u is set to the zero matrix.

This article is organised as follows. In Section 2, a naive ELR function is defined. We also define the MELE and investigate its asymptotic properties. In Section 3, a residual-adjusted ELR and the corresponding residual-adjust MELE are proposed, and their asymptotic properties are derived. In Section 4, we discuss the pointwise confidence interval for a component of