Statistical Methodology 32 (2016) 161–184

Contents lists available at ScienceDirect

Statistical Methodology

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

Efficient estimation of varying coefficient models with serially correlated errors

Xiaojuan Kang [a,*], Tizheng Li [b]

[a] *School of Economics and Management, Xi'an University of Technology, Xi'an 710048, China*

[b] *School of Science, Xi'an University of Architecture and Technology, Xi'an 710055, China*

HIGHLIGHTS

- A new estimation method is proposed for VCM with serially correlated errors.
- The proposed estimator is more efficient than the existing local linear estimator.
- A procedure is suggested to select the order of the AR error process.
- Simulation results show that significant gains can be achieved with our method.
- A real example is given to show the usefulness of the proposed estimation method.

ARTICLE INFO

Article history:
Received 19 September 2015
Accepted 21 May 2016
Available online 14 June 2016

MSC:
primary 62G08
secondary 62G05
62G20

Keywords:
Varying coefficient models
Autoregressive errors
Local linear smoothing
Profile least squares
SCAD

ABSTRACT

The varying coefficient model provides a useful tool for statistical modeling. In this paper, we propose a new procedure for more efficient estimation of its coefficient functions when its errors are serially correlated and modeled as an autoregressive (AR) process. We establish the asymptotic distribution of the proposed estimator and show that it is more efficient than the conventional local linear estimator. Furthermore, we suggest a penalized profile least squares method with the smoothly clipped absolute deviation (SCAD) penalty function to select the order of the AR error process. Simulation evidence shows that significant gains can be achieved in finite samples with the proposed estimation procedure. Moreover, a real data example is given to illustrate the usefulness of the proposed estimation procedure.

* Corresponding author.
E-mail address: lltz100@163.com (X. Kang).

http://dx.doi.org/10.1016/j.stamet.2016.05.005

1. Introduction

Investigating the relationship between a response variable and a set of covariates is a key issue in many statistical problems. Among others, mean regression models extract central trend of data by specifying the conditional mean function of a response variable given values of the covariates. A number of mean regression models have been proposed in the literature. The most traditional and simplest way to model the relationship is to employ the classical linear regression model. However, this model is often too restrictive to capture complicated characteristics which might exist in the data. From this point of view, the varying coefficient model is a very useful alternative. In general, a varying coefficient model assumes that the relationship between the response variable Y and the covariates $X_1, \ldots, X_p$ and U follows

$$Y_i = \mathbf{a}^{\mathrm{T}}(U_i)\mathbf{X}_i + \varepsilon_i, \tag{1}$$

where $Y_i, \mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$ and U_i are, respectively, the observations of the response variable Y and the covariates $X_1, \ldots, X_p$ and U, $\mathbf{a}(\cdot) = (a_1(\cdot), \ldots, a_p(\cdot))^{\mathrm{T}}$ is a vector of unknown coefficient functions, and ε_i's are random errors with $\mathrm{E}(\varepsilon_i|U_i, \mathbf{X}_i) = 0$ and $\mathrm{Var}(\varepsilon_i|U_i, \mathbf{X}_i) = \sigma^2(U_i, \mathbf{X}_i)$. By modeling the regression coefficients nonparametrically, the varying coefficient model is flexible and can be used to model nonlinear interaction effects among the covariates. The structure of the varying coefficient model is simple, since the conditional mean function is still a linear function of the covariates $X_1, \ldots, X_p$. If all coefficient functions are constant functions, the varying coefficient model reduces to the classical linear regression model.

The varying coefficient model was first introduced by Cleveland et al. [7] and further developed by Hastie and Tibshirani [18]. Due to its easy interpretation and its flexibility to explore dynamic pattern of a regression relationship, model (1) has attracted a great deal of attention over the past two decades. When its errors are independent and identically distributed or within-subject dependent, model (1) has been widely studied in estimation [6,4,14,18–20,24,27,35], hypothesis testing [4,15,16,33,37], and variable selection [8,12,23,30,28,34,38].

Except for those work on longitudinal data, almost all work mentioned above make an assumption that the errors ε_i's are not correlated. The work on longitudinal data usually makes a contemporaneous correlation assumption on the errors. This may be not true for the applications involving the use of time series data. There may exist significant serial correlation among the data when they are collected sequentially in time. As is well known, in nonparametric regression models with serially correlated errors, great efforts have been made to improve the efficiency of the conventional local polynomial estimator of the regression function. Under the assumption that the errors follow an invertible linear process, that is, a moving average (MA) process with infinite order, Xiao et al. [32] proposed a new estimation procedure that takes into account the correlation structure of the errors and showed that the proposed estimator is asymptotically more efficient than the conventional local polynomial estimator. Su and Ullah [26] assumed that the error process has a finite order nonparametric autoregressive (AR) structure, and proposed a three-step procedure for more efficient estimation of the regression function. Li and Li [21] investigated how to incorporate the correlation information of the errors into the local linear smoothing method. Under the assumption of the error process being an AR process, they proposed a new estimation procedure for the regression function by using local linear smoothing method and the profile least squares method. Liu et al. [22] assumed the error process to be an autoregressive moving average (ARMA) process and developed a new method to jointly estimate the transfer function nonparametrically and the ARMA parameters parametrically. Compared to the study on nonparametric regression models with serially correlated errors, less attention has been paid to varying coefficient models with serially correlated errors. Recently, Cai [3] studied the estimation problem of model (1) under the assumption that the design point $U_i = i/n$ and the series $\{(\varepsilon_i, \mathbf{X}_i)\}$ is strictly stationary α-mixing but $\{\varepsilon_i\}$ and $\{\mathbf{X}_i\}$ may not be independent. The local linear smoothing method was employed to estimate the coefficient functions and the asymptotic properties of the resulting estimators such as consistency and asymptotic normality were also established. However, as pointed out by Cai [3], the proposed estimation procedure may not be efficient because it does not take into account the correlation structure of the errors and hence