第51卷第12期 中国科学技术大学学报 Vol. 51, No. 12
2021年12月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Dec. 2021
Received: 2021-02-22; Revised: 2021-03-20 doi:10.52396/JUST-2021-0054

A robust homogeneity pursuit algorithm for varying coefficient models with longitudinal data

TANG Heng, ZHENG Zhi, ZHANG Weiping*
Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China
* Corresponding author. E-mail: zwp@ustc.edu.cn

Abstract: This article explores the homogeneity of coefficient functions in varying coefficient models where individuals can be classified into different subgroups for each covariate where its varying coefficients are homogeneous in the same subgroup. With repeated measurements, we use B-spline function approximations and the change point detection algorithm to identify the homogeneity. To account for the potential outliers or heavy-tailedness of the observed distribution, we propose to estimate the coefficient functions under the framework of M-estimation, and use least absolute deviation (LAD) loss as an example. Numerical results show that our estimators outperform the commonly used least squares (LS) estimators when existing outliers and heavy-tailedness of observed distribution.

Keywords: varying coefficient model; M-estimator; B-spline functions; change point detection; homogeneity pursuit
CLC number: O212.1 **Document code**: A
2020 Mathematics Subject Classification: 62J02

1 Introduction

Homogeneity occurs in panel data analysis when the individuals in a subgroup share the same regression coefficients because of the similarity among them. For example, the relationship between the mean daily maximum temperature (TMAX), total rainfall (RAIN), and total sunshine duration (SUN) in neighboring geographical regions is expected to be approximately the same[1]. The traditional approaches assume all individuals share the same unknown parameters, e.g., Refs. [2,3]. It is important to pay attention to the subgroups because they may have very important practical meaning and lead to some important findings in practice. To identify subgroups, several methods have been proposed under the homogeneity setting. Ref. [4] proposed a two-step procedure that combines hierarchical clustering and the Lasso. Refs. [5,6] regarded the fused Lasso as an effort of exploring homogeneity, with the assistance of neighborhoods defined according to either time or location. Ref. [7] proposed the octagonal shrinkage and clustering algorithm for regression (OSCAR). Ref. [8] considered simultaneous grouping pursuit. Ref. [9] proposed a Bayesian framework by using shrinkage priors. Ref. [10] proposed a method called clustering algorithm in regression via data-driven segmentation (CARDS) to explore homogeneity. Ref. [11] designed an algorithm called clustering algorithm in regression via data-driven segmentation (CARD) to explore the homogeneity. Ref. [12] applied change point detection and a binary segmentation based algorithm to detect the subgroups. Ref. [1] proposed a modeling based on the single index models embedded with homogeneity. Ref. [13] detected change points for regression coefficients of each covariate separately by the binary segmentation based algorithm. Ref. [14] proposed a approach, referred to as spatially clustered coefficient (SCC) regression, to detect spatially clustered patterns in the regression coefficients.

Particularly, the aforementioned approaches assume that the estimation parameters are usually fixed as constants, which is not reasonable because in some cases the covariates may have dynamic effect on the response as time changing. For example, the influence of total rainfall (RAIN) and total sunshine duration (SUN) on temperature usually changes with the time, caused by, changes in the local environment, see Figure 1. The varying coefficient model is a useful extension of linear models and has many advantages in real applications. The varying coefficient model is often applied in epidemiological study for predicting CD4 (T-

Citation: Tang Heng, Zheng Zhi, Zhang Weiping. A robust homogeneity pursuit algorithm for varying coefficient models with longitudinal data. J. Univ. Sci. Tech. China, 2021, 51(12): 857-867.

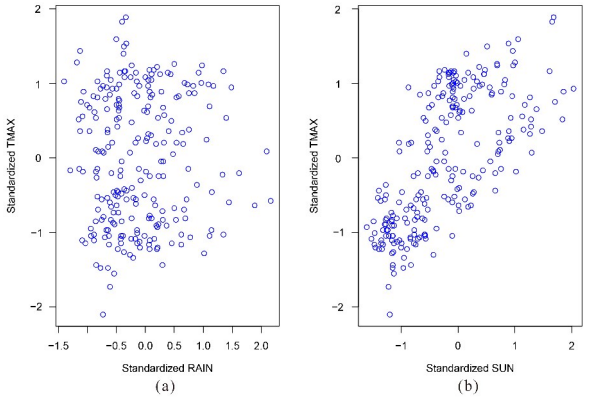

Figure 1. The scatter plot of standardize TMAX versus standardize RAIN (a) and standardize SUN (b) in Armagh from January 2001 to December 2020.

helper lymphocytes) cell changes among HIV (human immunodeficiency virus) -infected person[15]. This model can also be used in toxicity analysis in environmental studies[16]. Several methods have been proposed to estimate coefficient functions. Ref. [17] used the local linear regression and the kernel-based weights to estimate model. Ref. [18] estimated the trend function and the coefficient function without taking the first difference to eliminate the fixed effects. Ref. [19] proposed a new technique to estimate the unknown coefficient functions based on the first-order differences and the local linear regression. Ref. [20] proposed a penalised likelihood method with the Lasso penalty function.

Furthermore, the estimation methods have been proposed mostly based on least-squares (LS) estimation, see Refs. [1,13], among others. The least-squares (LS) estimates certainly have some nice properties, particularly when the random errors follow the normal distribution. However, it is well-known that the LS-estimation will not perform well when the dataset has outliers or is heavy-tailed distributed. For example, in the UK climate data, the distribution of TMAX in Armagh is obviously disobeying the normal distribution, see Figure 2. In this situation, robust estimation methods are desired. Based on the varying coefficient model, several M-estimation methods have been proposed. Ref. [21] presented the local linear LAD-estimation method. Ref. [22] established the asymptotic normality of proposed estimators for both the parametric and nonparametric parts. Ref. [23] proposed an efficient and robust penalized estimating procedure for varying coefficient single-index models based on modal regression and basis function approximations. Ref. [24] proposed a robust estimation procedure based on the exponential squared loss function for the varying coefficient partially nonlinear model. In this paper, we use B-spline function approximations to estimate the unknown functions under the framework of M-estimation. More details about M-estimation can be seen in Ref. [25].

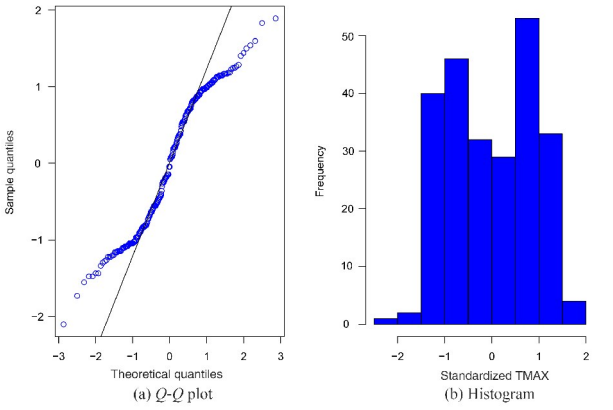

Figure 2. The distribution of standardized TMAX in Armagh.

Generally speaking, our method has three innovations. First, we identify the subgroup structure of every predictor separately, while traditional methods usually treat all predictors as only one estimation index to identify subgroups. Second, we assume that the estimated parameters are functions that changing with time, while the existing works generally treat the parameters as constants. At last, we estimate coefficient functions under the framework of M-estimation, which is more robust to the potential outliers and heavy-tailedness of observed distribution than widely-used LS estimation.

The rest of this article is organized as follows. Section 2 describes the varying coefficient model, presents the estimation method and the subgroup identification procedure. The results of simulation study are given in Section 3. In Section 4, we apply our method to estimate the subgroup structure of the UK climate data. Section 5 describes the conclusions.

2 Methodology

2.1 Varying coefficient model

Consider a time-varying response $Y(t)$ with p-dimensional covariate function $X(t)$, the observations are obtained from n subjects each repeatedly measured for n_i times over a set of distinct time points, i. e., (Y_{ij}, X_{ij}, t_{ij}) represents the jth observation of the ith individual at time t_{ij}, $i=1,\cdots, n$, $j=1,\cdots, n_i$, where $X_{ij} = (X_{ij1}, \cdots, X_{ijp})^{\mathrm{T}}$ is a p-dimensional vector of predictors.

Consider the following varying coefficient linear model:

$$Y_{ij} = X_{ij}^{\mathrm{T}}\beta_i(t_{ij}) + \epsilon_{ij} = \sum_{k=1}^{p} X_{ijk}\beta_{ik}(t_{ij}) + \epsilon_{ij} \quad (1)$$

where $\beta_i(t_{ij}) = (\beta_{i1}(t_{ij}), \beta_{i2}(t_{ij}), \cdots, \beta_{ip}(t_{ij}))^{\mathrm{T}}$ and

$$\beta_{ik}(t_{ij}) = \begin{pmatrix} \beta_{1k}(t_{ij}) & \text{when } i \in G_{1,k} \\ \beta_{2k}(t_{ij}) & \text{when } i \in G_{2,k} \\ \vdots & \\ \beta_{N_k k}(t_{ij}) & \text{when } i \in G_{N_k,k} \end{pmatrix}, k = 1, \cdots, p.$$

Without loss of generality, we assume that $t_{ij} \in [0, 1]$, X_{ij} and ϵ_{ij} are independent, $E(\epsilon_{ij} | X_{ij}) = 0$, $X_{ij1} = 1$. The set $\widehat{G}_k = \{G_{q,k}: q = 1, \cdots, N_k\}$ is a partition of set $\{1, \cdots, n\}$, which means there are N_k subgroups in the subjects on the kth predictor. The individuals in the same subgroup have the same coefficient function $\beta_{ik}(t)$. So that, $\beta_{uk}(t) = \beta_{qk}(t)$, for all $u \in G_{q,k}$. The number of subgroups N_k is unknown.

This allows us to get the homogeneity structure of individuals respect to each covariate. The individuals whose kth predictor have similar influence on response variables are more likely to be divided into the same subgroup. For kth predictor, we only have N_k unknown functions to estimate with N_k being much smaller than n. Therefore, the above assumption partitions the whole individuals into at least $\max_k(N_k)$ and at most $\prod_{k=1}^{p} N_k$ subgroups. The total number of functions to be estimated is $\sum_{k=1}^{p} N_k$.

2.2 Methods

In order to estimate the unknown functions $\{\beta_{ik}(t), i = 1, \cdots, n; k = 1, \cdots, p\}$, we use M-estimation with B-spline approximations. Our main purpose is to obtain the structure of subgroups, so we estimate all $\beta_{ik}(t)$ by the same B-spline basis. A spline is defined as a piecewise polynomial that is smoothly connected at its knots, $B(t) = (B_1(t), \cdots, B_L(t))^{\mathrm{T}}$ represents the B-spline basis functions of order d and knots $\tau = \{0 = \tau_0 < \tau_1 < \cdots < \tau_{L-d+1} = 1\}$ are formed by the equally spaced on the interval $[0, 1]$. We can approximate $\beta_{ik}(t)$ by $\beta_{ik}(t) = B(t)^{\mathrm{T}}\theta_{ik}$, $\theta_{ik} = (\theta_{ik1}, \cdots, \theta_{ikL})^{\mathrm{T}}$. Based on longitudinal observation $\{\langle Y_{ij}, X_{ij}, t_{ij}\rangle, i = 1, \cdots, n; j = 1, \cdots, n_i\}$, we can minimize the objective function:

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i} \omega_i \rho(Y_{ij} - \sum_{k=1}^{p} X_{ijk} B(t_{ij})^{\mathrm{T}}\theta_{ik}) \qquad (2)$$

where $\beta_i = (\beta_{i1}, \cdots, \beta_{ip})^{\mathrm{T}}$, then the M-estimator $\widehat{\beta}_{ik}(t)$ of $\beta_{ik}(t)$ is $\widehat{\beta}_{ik} = B(\cdot)^{\mathrm{T}}\theta_{ik}$, ω_i is a non-negative weight for the ith subject and ρ is a suitably chosen loss function. Usual choices of ω_i include $\omega_i \equiv 1/N = 1/\sum n_i$ and $\omega_i \equiv 1/(nn_i)$, which correspond to providing equal weight to each single observation and equal weight to each subject, respectively. Traditional estimation methods in subgroup identification are major in least-squares estimation $\rho(u) = u^2$, but we use least-absolute estimation $\rho(u) = |u|$ instead to account for the robustness consideration. The LAD loss is more robust than the LS loss in the cases that the data have some outliers and the random errors follow heavy-tail distributions. However, it is well known that LAD regression is also lack of robustness when the data include outliers in the covariates[26] (i. e., there exist leverage points). In this scenario, we can assign a weight to each observation in advance, and the final estimation is expected be robust to the outliers in the covariates if those pre-assigned weights correctly reflect the outlying information among all covariates, see Ref. [27] for more details. In literature, several robust estimation methods have been proposed and different kinds of ρ have been studied, like a ρ function with a bounded derivative $\rho'(\mu) = \max\{-1, \min\{\frac{u}{c}, 1\}\}$, $c > 0$, more details about the properties of M-estimators can be found in Refs. [28,29].

We will propose a three-step estimation procedure to identify the subgroups. In the first stage, we get an initial estimator as $\widetilde{\beta}_{ik}(t) = B(t)^{\mathrm{T}}\tilde{\theta}_{ik}$, which only depends on the observations for the ith individual. Second, we use the change point detection method to identify the subgroup structure via the initial estimated functions $\{\widetilde{\beta}_{ik}(t)\}$. At last, we estimate the $\beta_{ik}(t)$ under the identified subgroup structure. The following gives the details:

Step 1 (Initial estimation). For each individual i, $i = 1, \cdots, n$, based on the n_i observations, we can get an initial estimator $\widetilde{\beta}_{ik}$ by minimizing the objective function.

$$\sum_{j=1}^{n_i} \omega_i \rho(Y_{ij} - \sum_{k=1}^{p} X_{ijk} B(t_{ij})^{\mathrm{T}}\theta_{ik}) \qquad (3)$$

Step 2 (Homogeneity pursuit). For any $\widetilde{\beta}_{ik}$, $i = 1, \cdots, n$, $k = 1, \cdots, p$, $\tilde{\theta}_{ik} = (\tilde{\theta}_{ik1}, \cdots, \tilde{\theta}_{ikL})$. $\tilde{\theta}_{ikl}$ is the lth component of the $\tilde{\theta}_{ik}$. We can sort $\{\tilde{\theta}_{ikl}, i = 1, \cdots, n\}$ and denote them by $b_{(1)} \leqslant \cdots \leqslant b_{(n)}$.

For different individuals i_1, i_2, if their kth predictors belong to the same subgroup, the difference between $\tilde{\theta}_{i_1kl}$ and $\tilde{\theta}_{i_2kl}$ ($l = 1, \cdots, L$) is small. So identifying the subgroup structure among $\{\widetilde{\beta}_{ik}, i = 1, \cdots, n\}$ is equivalent to detecting the change points among $\{\tilde{\theta}_{ikl}, i = 1, \cdots, n\}$ for $l = 1, \cdots, L$. Then we use the binary segmentation algorithm, which has been applied in statistics, more details can be found in Refs. [30,31].

For any $1 \leqslant i_1 < i_2 \leqslant n$, let

$$\Delta_{i_1 i_2}(\kappa) = \sqrt{\frac{(i_2 - \kappa)(\kappa - i_1 + 1)}{i_2 - i_1 + 1}} \left(\frac{\sum_{l=\kappa+1}^{i_2} b_{(l)}}{i_2 - \kappa} - \frac{\sum_{l=i_1}^{\kappa} b_{(l)}}{\kappa - i_1 + 1} \right).$$

Given a threshold δ, which can be selected, the binary segmentation algorithm to detect the change points works as follows:

(A) Find $\widehat{k}_1$ such that

$$\Delta_{i,n}(\widehat{k}_1) = \max_{1\leqslant\kappa<n} \Delta_{1,n}(\kappa).$$

If $\Delta_{1,n}(\widehat{k}_1)\leqslant\delta$, there is no change point among $\{b_{(l)}, l=1, \cdots, n\}$, and the process of detection ends. Otherwise, add $\widehat{k}_1$ to the set of change points and divide the region $\{\kappa: 1\leqslant\kappa\leqslant n\}$ into two subregions: $\{\kappa: 1\leqslant\kappa\leqslant\widehat{k}_1\}$ and $\{\kappa: \widehat{k}_1+1\leqslant\kappa\leqslant n\}$.

(B) Detect the change points in the two subregions obtained in part (A), respectively. Consider the region $\{\kappa: 1\leqslant\kappa\leqslant\widehat{k}_1\}$ first. Find $\widehat{k}_2$ such that

$$\Delta_{1,\widehat{k}_1}(\widehat{k}_2) = \max_{1\leqslant\kappa\leqslant\widehat{k}_1} \Delta_{1,\widehat{k}_1}(\kappa).$$

If $\Delta_{1,\widehat{k}_1}(\widehat{k}_2)\leqslant\delta$, there is no change point in the region $\{\kappa: 1\leqslant\kappa\leqslant\widehat{k}_1\}$. Otherwise, add $\widehat{k}_2$ to the set of change points and divide the region $\{\kappa: 1\leqslant\kappa\leqslant\widehat{k}_1\}$ into two subregions: $\{\kappa: 1\leqslant\kappa\leqslant\widehat{k}_2\}$ and $\{\kappa: \widehat{k}_2+1\leqslant\kappa\leqslant\widehat{k}_1\}$. Similarly, for the region $\{\kappa: \widehat{k}_1+1\leqslant\kappa\leqslant n\}$, we find $\widehat{k}_3$ such that

$$\Delta_{\widehat{k}_1+1,n}(\widehat{k}_3) = \max_{\widehat{k}_1+1\leqslant\kappa< n} \Delta_{\widehat{k}_1+1,n}(\kappa).$$

If $\Delta_{\widehat{k}_1+1,n}(\widehat{k}_3)\leqslant\delta$, there is no change point in the region $\{\kappa: \widehat{k}_1+1\leqslant\kappa\leqslant n\}$. Otherwise, add $\widehat{k}_3$ to the set of change points and divide the region $\{\kappa: \widehat{k}_1+1\leqslant\kappa\leqslant n\}$ into two subregions: $\{\kappa: \widehat{k}_1+1\leqslant\kappa\leqslant\widehat{k}_3\}$ and $\{\kappa: \widehat{k}_3+1\leqslant\kappa\leqslant n\}$.

(C) For each subregion obtained in part (B), we continue using the same computational algorithm as that for the subregion $\{\kappa: 1\leqslant\kappa\leqslant\widehat{k}_1\}$ or $\{\kappa: \widehat{k}_1+1\leqslant\kappa\leqslant n\}$ in part (B) and keep doing so until there is no subregion containing any change point.

Finally, let all change points be $K_{ikl} = \{\widehat{k}_{(1)} < \widehat{k}_{(2)} < \cdots < \widehat{k}_{(N_{ikl})}\}$, then we can divide the kth predictor into $N_{ikl}+1$ subgroups by the lth component of the $\widehat{\theta}_{ik}$. We reuse the binary segmentation algorithm to get the change points for $l=1, \cdots, L$. We get all change points of the kth predictor $K_{ik} = \{K_{ik1}, \cdots, K_{ikL}\}$. According to the set of change points, we can identify the subgroup structure of the kth predictor $\widehat{G}_k = \{G_{1k}, \cdots, G_{N_k k}\}$, the individuals in the same subgroup have the same unknown functions $\widehat{\beta}_{zk}$, $z=1, \cdots, N_k$. For $k=1, \cdots, p$, we get the subgroup structure $\widehat{G}_k$ of $\{\widetilde{\beta}_{ik}, k=1, \cdots, p\}$, and let $\widehat{G} = \{\widehat{G}_1, \cdots, \widehat{G}_p\}$.

Step 3 (Final estimation) Based on the above calculation, we get the final subgroup structure. And minimize the following objective function

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i} \omega_i \rho(Y_{ij} - \sum_{k=1}^{p} X_{ijk} B(t_{ij})^{\mathrm{T}} \theta_{ik}),$$

we use $\widehat{\beta}_{zk}(t) = B(t)^{\mathrm{T}}\widehat{\theta}_{zk}$ ($k=1, \cdots, p, z=1, \cdots, N_k$) to replace the $\beta_{ik}(t) = B(t)^{\mathrm{T}}\theta_{ik}$ ($i=1, \cdots, n, k=1, \cdots, p$).

As we know, the choices of smoothing parameters strongly influence the adequacy of the estimators. In our paper, we use splines with equally spaced knots and fixed degrees. In the literature, d, the order of B-spline, is almost always fixed to be either $d=3$ (quadratic splines) or $d=4$ (cubic splines), see Refs. [1, 13]. The number of spline functions L, can be selected by 'leave-one-subject-out' cross-validation procedure, like Ref. [25]. We can also select a fixed L like Refs. [32, 33].

Apparently, the whole estimation procedure depends crucially on the threshold δ used. On the other hand, if δ is too small, the number of change points will increase, we will come up with too many groups, leading to inflated variances of the final estimators. On the other hand, if δ is too large, we will mistake different group parameters β_{ij}'s in the same region and treat them as the same parameter, leading to a biased final estimators. To identify the homogeneity pursuit accuracy, the δ can be selected by the CV procedure in Ref. [1], or by standard Bayesian information criterion (BIC) in Ref. [12]. We use the following BIC in our simulation studies and real data analysis

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i} (Y_{ij} - \sum_{k=1}^{p} X_{ijk} B(t_{ij})^{\mathrm{T}}\widehat{\theta}_{ik}) + \gamma \ln(\sum_{i=1}^{n}(n_i)),$$

where γ is the total number of distinct parameters in the estimated model.

We summarize the above method as the following Algorithm 2.1.

Algorithm 2.1 Subgroup identification algorithm in the varying coefficient model

Input: y_{ij}: the jth observation of the ith individual of response variable; x_{ij}: the jth observation of the ith individual of independent variables; L: the number of spline functions; a, b: the value range of δ

Output: $\widehat{\beta}_{ik}$: the estimation parameter of the kth independent variable of the ith individual

1 generate B-spline basis functions $B(t) = (B_1(t), \cdots, B_L(t))^{\mathrm{T}}$;
2 for i in 1:n do
3 generate covariance matrix $\tilde{x}_i$ by x_i and $B(t)$;
4 estimate $\widetilde{\beta}_i = (\widetilde{\beta}_{i1}, \cdots, \widetilde{\beta}_{ip})^{\mathrm{T}}$ by $\tilde{x}_i$ and y_i, $\widetilde{\beta}_{ik}$, $k=1, \cdots, p$, $\widetilde{\beta}_{ik}(t_{ij}) = B(t_{ij})^{\mathrm{T}}\tilde{\theta}_{ik}$, $\tilde{\theta}_{ik} = (\tilde{\theta}_{ik1}, \cdots, \tilde{\theta}_{ikL})$;
5 end for
6 for δ in a:b do
7 for k in 1:p do
8 for l in 1:L do
9 consider $\{\tilde{\theta}_{ikl}, i=1, \cdots, n\}$, use the binary segmentation algorithm to get all change points $K_{ikl} = \{\widehat{k}_{(1)} < \widehat{k}_{(2)} < \cdots < \widehat{k}_{(N_{ikl})}\}$;
10 end for
11 consider all change points of the kth predictor $K_{ik} = \{K_{ik1}, \cdots, K_{ikL}\}$, get the subgroup structure $\widehat{G}_k = \{G_{1k}, \cdots,$

$G_{N_k k}\}$;

end for

get the subgroup of every predictor $\widehat{G}=\{\widehat{G}_1,\cdots,\widehat{G}_p\}$;

the individuals have the estimation functions $\beta_{ik}=\widehat{\beta}_{iz}$, $i=1,\cdots,n, k=1,\cdots,p, z=1,\cdots,N_k$, while $(i,k)\in G_{z,k}$

get the $\widehat{\beta}_{ik}$ by minimizing the following objective function

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\omega_i\rho(Y_{ij}-\sum_{k=1}^{p}X_{ijk}\beta_{ik})$$

calculate the BIC by $\widehat{\beta}_{ik}$ and y_{ij}

end for

choose the right δ in order to make the BIC minimum

get the final estimation parameter $\{\beta_{ik}\}$ and the final subgroup structure $\widehat{G}$

3 Simulation

In this section, we investigate finite sample performance of proposed estimation method. At the same time, we compare the subgroup identification accuracy between the least-squares loss $\rho(u)=u^2$ and least-absolute loss $\rho(u)=|u|$. Two Models are conducted with the same model structure, while the coefficient parameters β_{i1}, β_{i2} in Model Ⅰ are segment linear functions and the coefficient parameters β_{i2} in Model Ⅱ are non-linear functions. We assume that the coefficient parameters are different kinds of smooth functions, in order to show that our subgroup identification algorithm has great accuracy in different situations.

Model Ⅰ

We generate the data from the following model:

$$y_{ij}=X_{ij1}\beta_{i1}(t_{ij})+X_{ij2}\beta_{i2}(t_{ij})+\varepsilon_{ij},\quad i=1,\cdots,n,\ j=1,\cdots,n_i,$$

where $t_{ij}=\frac{j}{n_i}$, $j=1,\cdots,n_i$ and $X_{ij}=(X_{ij1},X_{ij2})^{\mathrm{T}}=(1,0.1\times T\times t_{ij}+x_{ij2})^{\mathrm{T}}$. The coefficient functions are chosen as

$$\beta_{i1}(t)=\begin{cases}(t\times n_i-1), & 1\leqslant i\leqslant n/3;\\ 3(t\times n_i-1), & n/3<i\leqslant 2n/3;\\ 5(t\times n_i-1), & 2n/3<i\leqslant n;\end{cases}$$

$$\beta_{i2}(t)=\begin{cases}(t\times n_i-1), & 1\leqslant i\leqslant n/4;\\ 2(t\times n_i-1), & n/4<i\leqslant n/2;\\ 3(t\times n_i-1), & n/2<i\leqslant 3n/4;\\ 4(t\times n_i-1), & 3n/4<i\leqslant n.\end{cases}$$

We divide every coefficient function into different subgroups, the first coefficient function has three subgroups, the second coefficient function has four subgroups. In summary, all individuals have been divided into six subgroups,

$$G_1=\{i:1\leqslant i\leqslant\frac{n}{4}\},\ G_2=\{i:\frac{n}{4}<i\leqslant\frac{n}{3}\},$$

$$G_3=\{i:\frac{n}{3}<i\leqslant\frac{n}{2}\},\ G_4=\{i:\frac{n}{2}<i\leqslant\frac{2n}{3}\},$$

$$G_5=\{i:\frac{2n}{3}<i\leqslant\frac{3n}{4}\},\ G_6=\{i:\frac{3n}{4}<i\leqslant n\}.$$

Model Ⅱ

We generate the data from the model in Model Ⅰ but $X_{ij}=(X_{ij1},X_{ij2})^{\mathrm{T}}=(1,x_{ij2})$. The coefficient functions are chosen as

$$\beta_{i1}(t)=\begin{cases}3, & 1\leqslant i\leqslant\frac{n}{2};\\ 0, & \frac{n}{2}<i\leqslant n;\end{cases}$$

$$\beta_{i2}(t)=\begin{cases}3\mathrm{e}^{t-\frac{1}{n_i}}-3, & 1\leqslant i\leqslant\frac{n}{3};\\ 3\mathrm{e}^{\frac{n_i+1}{n_i}-t}-3\mathrm{e}, & \frac{n}{3}<i\leqslant n.\end{cases}$$

We divide every coefficient function into different subgroups, the first and the second coefficient functions have two subgroups. In summary, all individuals have been divided into three subgroups, $G_1=\{i:1\leqslant i\leqslant\frac{n}{3}\}$, $G_2=\{i:\frac{n}{3}<i\leqslant\frac{n}{2}\}$ and $G_3=\{i:\frac{n}{2}<i\leqslant n\}$.

Under the assumption of the model structure and coefficient functions in the above two models, we generate the x_{ij2} from the same distribution while the random errors are generated from three different distributions reflecting outliers.

Case Ⅰ: x_{ij2}'s are i. i. d. from a normal distribution with mean 0 and variance $\sigma^2=0.1$. The random errors $(\epsilon_{i1},\cdots,\epsilon_{in_i})^{\mathrm{T}}$ are independent random noise having the $N(0,1)$ distribution.

Case Ⅱ: x_{ij2}'s have the same distribution like Case Ⅰ. The random error $(\epsilon_{i1},\cdots,\epsilon_{in_i})^{\mathrm{T}}$ are independent random noise having the t distribution with $df=3$.

Case Ⅲ: x_{ij2}'s have the same distribution like Case Ⅰ. The random error $(\epsilon_{i1},\cdots,\epsilon_{in_i})^{\mathrm{T}}$ are independent random noise having the Cauchy distribution.

We apply three estimation approaches for comparison. Approach 1 (A1) is Under-fitting. In this estimation approach, we do not consider the individual characteristics. In other words, we treat all individuals as one whole group, so $N_K=1$, $\beta_{1k}=\cdots=\beta_{nk}$, $k=1,\cdots,p$. Approach 2 (A2) is Correct-fitting which is the method we proposed. In this estimation approach, we assume all individuals follow a potential subgroup structure. Approach 3 (A3) is Over-fitting. In the estimation approach, we do not consider the homogeneity pursuit between different individuals. In other words, we treat every individual as an one-member subgroup, so $N_k=n$, $\beta_{1k}=\widetilde{\beta}_{ik}$, which is estimated in Step 1 of Algorithm 2.1.

For each simulated data set, we consider the

balanced design where n_i's are all equal. We use cubic spines with 1 inner knots and without intercept; thus, the dimension of the B-spline basis is 4. We adopt standard BIC to select the number of subgroups. The choice of ω_i is $\omega_i \equiv 1/N = 1/\sum n_i$, representing $\omega_1 = \cdots = \omega_n$.

We use the following two metrics to compare the estimation accuracy and subgroup identification accuracy: The mean squared error (MSE) ($\mathrm{MSE}(\widehat{Y}_{ij}) = E(\| \widehat{Y}_{ij} - Y_{ij} \|^2)$), and the normalized mutual information (NMI). The NMI is proposed by Ref. [10] to evaluate the distance between two groups. Suppose $A = \{A_1, A_2, \cdots\}$ and $B = \{B_1, B_2, \cdots\}$ are two sets of disjoint group of $\{1, 2, \cdots, n\}$. The NMI is defined as

$$\mathrm{NMI}(A, B) = \frac{2I(A,B)}{H(A) + H(B)},$$

where

$$I(A, B) = \sum_{i,j} \frac{|A_i \cap B_j|}{n} \log\left(\frac{n|A_i \cap B_j|}{|A_i||B_j|}\right)$$

is the mutual information between the two groups, and

$$H(A) = \sum_i \frac{|A_i|}{n} \log\left(\frac{n}{|A_i|}\right).$$

NMI takes values in [0, 1], and a larger value indicates a higher degree of similarity between the two groups. NMI = 1 means group A is the same as group B. In our simulations, we compare the subgroup B which is estimated with the true subgroup A. NMI can assess the accuracy of subgroup identification.

We consider balanced design where T_i's are all equal. The number of repeated measurements is set to be $T = 30$ or 60, the number of individuals is set to be $n = 60$ or 120. The number of repeated measurements is smaller than the number of individuals, which is more common in our daily life. We calculate the results for estimation errors (MSE) and subgroup identification accuracy (NMI) for β_{ik}, $i = 1, \cdots, n$, $k = 1, \cdots, p$, based on the 100 simulation results. Normal distribution means the random error ϵ_{ij} comes from the normal distribution with mean 0, variance 1. t_3 distribution means the random error ϵ_{ij} comes from the t distribution with df is 3. Cauchy distribution means the random error ϵ_{ij} comes from the Cauchy distribution.

Table 1 reports the average of MSE and NMI, the data are generated from the Model Ⅰ. To A1, we consider all individuals as one group, so that the estimation subgroup structure $B = \{B_1\} = \{1, \cdots, n\}$, then

$$I(A, B) = \sum_i \frac{|A_i|}{n} \log\left(\frac{n|A_i|}{n|A_i|}\right) = 0.$$

So NMI = 0. At the same time, the MSE is much larger than A2 and A3 because A1 assumes $\beta_{1k} = \cdots = \beta_{nk}$, $k = 1, \cdots, p$. To A3, we consider every individual as on one-member subgroup, so that the estimation subgroup structure $B = \{B_1, \cdots, B_n\} = \{1, \cdots, n\}$, $B_i = \{i\}$, then

$$I(A, B) = \sum_i \frac{1}{n} \log\left(\frac{n}{|A_i|}\right) = \text{Constant}.$$

So NMI is constant and MSE is the smallest. To A2, the method we proposed, we can see that MSE is just a little larger than A3, representing that estimation error is small and our estimation functions are close to the true coefficient functions. When the random error obeys the normal distribution, our estimation method can recover the true subgroup structure well, and LS-estimation has a better effect on identification accuracy than LAD-estimation, but the difference is not great. However, when the random error obeys the t distribution, a common heavy-tailed distribution, LAD-estimation has obvious advantages over LS-estimation. For example, considering the situation $n = 120$, $T = 60$, when the random error obeys the normal distribution, NMI under LS-estimation is 0.9386 while NMI under LAD-estimation is 0.9299. When the random error obeys the t distribution, NMI under LS-estimation is 0.7382 while NMI under LAD-estimation is 0.7900. In addition, the identification accuracy will increase with the number of observation times, because the initial estimation functions $\tilde{\beta}_{ik}$, $i = 1, \cdots, n$, $k = 1, \cdots, p$, will be estimated more accurately. In the more extreme case, LS-estimation accuracy will further decline, we use the Cauchy distribution as an example. We can see that the NMI in A2 by the LS-estimation even smaller than the NMI in A3, it means many individuals are divided into wrong subgroups. Because MSE has a wide range of fluctuations, there is little significance in comparing MSE. But the LAD-estimation method still works, like NMI = 0.6772, when $T = 60$, $n = 120$.

Table 2 reports the average of MSE and NMI, the data are generated from the Model Ⅱ. The simulation results of simulation in Model Ⅱ show that in the case of different smooth coefficient functions, our proposed method can also work well. And the accuracy of subgroup identification is much better under the framework of M-estimation in the situation with outliers and heavy-tailed distribution. For example, also considering the situation $n = 120$, $T = 60$, when the random error obeys the normal distribution, NMI under LS-estimation is 0.9903 while NMI under LAD-estimation is 0.9625. When the random error obeys the t distribution, NMI under LS-estimation is 0.8922 while NMI under LAD-estimation is 0.9474. At the same time, the number of subgroups in Model Ⅱ is smaller than the number of subgroups in Model Ⅰ, so the NMI which represents the accuracy of subgroup identification in Model Ⅱ is a little higher than the NMI in Model Ⅰ in some cases, although the coefficient parameters β_{i2} in Model Ⅱ are non-linear functions.

Table 1. The average of MSE and NMI under Model Ⅰ.

		n	MSE(A1)		MSE(A2)		MSE(A3)		NMI(A1)		NMI(A2)		NMI(A3)	
			$T=30$	$T=60$	$T=30$	$T=60$	$T=30$	$T=60$	$T=30$	$T=60$	$T=30$	$T=60$	$T=30$	$T=60$
Normal	LAD	60	5095.55	54873.19	0.9779	0.9643	0.8646	0.9370	0	0	0.8872	0.9808	0.5879	0.5879
		120	5140.94	54679.69	0.9876	0.9942	0.8601	0.9373	0	0	0.8535	0.9299	0.5251	0.5251
	LS	60	5015.02	53423.40	0.9673	0.9895	0.7436	0.8676	0	0	0.9019	0.9956	0.5879	0.5879
		120	5014.44	53444.51	0.9817	0.9978	0.7475	0.8715	0	0	0.8734	0.9386	0.5251	0.5251
t_3	LAD	60	5094.90	54785.78	2.6892	2.8258	2.4760	2.7061	0	0	0.7792	0.9268	0.5879	0.5879
		120	5108.42	54699.71	3.1108	2.8940	2.6078	2.7547	0	0	0.7184	0.7900	0.5251	0.5251
	LS	60	5014.11	53458.75	2.7647	2.9043	2.1769	2.5650	0	0	0.7504	0.8891	0.5879	0.5879
		120	5022.01	53448.49	2.9186	2.9515	2.2380	2.6182	0	0	0.6842	0.7382	0.5251	0.5251
Cauchy	LAD	60	18969.32	92561.47	13939.85	37073.83	13929.79	37069.21	0	0	0.6819	0.7706	0.5879	0.5879
		120	52462.35	86469.29	45376.21	31825.07	44578.35	31821.30	0	0	0.5865	0.6772	0.5251	0.5251
	LS	60	7984.40	86433.64	2595.22	30657.09	1735.02	30081.34	0	0	0.5597	0.6287	0.5879	0.5879
		120	21368.28	108291.70	15988.22	79650.40	13043.86	79121.59	0	0	0.4729	0.5082	0.5251	0.5251

Table 2. The average of MSE and NMI under Model Ⅱ.

		n	MSE(A1)		MSE(A2)		MSE(A3)		NMI(A1)		NMI(A2)		NMI(A3)	
			$T=30$	$T=60$	$T=30$	$T=60$	$T=30$	$T=60$	$T=30$	$T=60$	$T=30$	$T=60$	$T=30$	$T=60$
Normal	LAD	60	7.3802	7.4374	1.0883	0.9992	0.9142	0.9627	0	0	0.9114	0.9710	0.3962	0.3962
		120	7.3219	7.4368	0.9919	0.9982	0.9202	0.9621	0	0	0.8886	0.9625	0.3488	0.3488
	LS	60	7.1616	7.2063	0.9916	0.9959	0.8392	0.9150	0	0	0.9428	0.9914	0.3962	0.3962
		120	7.1149	7.2007	0.9967	1.0002	0.8392	0.9150	0	0	0.9322	0.9903	0.3488	0.3488
t_3	LAD	60	9.3828	9.3059	3.0565	2.9444	2.8476	2.8542	0	0	0.8449	0.9506	0.3962	0.3962
		120	9.1949	9.3639	2.8974	2.9963	2.6878	2.8994	0	0	0.8117	0.9474	0.3488	0.3488
	LS	60	9.1182	9.1945	2.9926	2.9582	2.4037	2.7184	0	0	0.7874	0.9051	0.3962	0.3962
		120	9.1262	9.2380	3.0108	3.0382	2.4890	2.7848	0	0	0.6733	0.8922	0.3488	0.3488
Cauchy	LAD	60	4171.43	7926.31	4169.82	7926.29	4164.21	7925.71	0	0	0.6519	0.8319	0.3962	0.3962
		120	41098.36	46060.40	42086.33	45061.30	40076.19	40686.60	0	0	0.5851	0.7868	0.3488	0.3488
	LS	60	23996.50	19264.44	21052.26	18435.66	21023.77	18151.70	0	0	0.2717	0.2966	0.3962	0.3962
		120	48312.76	49427.15	44324.14	46035.53	40617.16	26766.37	0	0	0.2148	0.2335	0.3488	0.3488

Table 3. Subgroup structure based on the observations in 16 locations by the LAD-estimation approach.

	Armagh	Bradford	Camborne	Eastbourne	Eskdalemuir	Heathrow	Hurn	Lerwick
INTERCEPT	1	1	1	1	1	1	1	1
RAIN	1	2	2	3	1	1	3	3
SUN	4	1	2	1	3	1	1	2
	Leuchars	Oxford	Paisley	Ross-On-Wye	Shawbury	Sheffield	Waddington	Whitby
INTERCEPT	1	1	1	1	1	1	1	1
RAIN	2	1	2	1	2	2	2	3
SUN	1	1	1	1	1	1	1	1

4 Real data analysis

We apply the proposed algorithm to the UK climate dataset which is available from the UK Met Office website ①, and contains data of the mean daily maximum temperature (TMAX), the mean daily minimum temperature (TMIN), days of air frost (AF), total rainfall (RAIN), and total sunshine duration (SUN) collected from 37 stations. We first

① http://www.metoffice.gov.uk/public/weather/climatehistoric.

remove the missing values and thus select data during the period of January 2001 to December 2020, then observations from 16 locations can be used. Ref. [1] has used the dataset to identify the subgroup structure based on the single index models, and achieved good results. We assume unknown parameters $\beta_{ik}(t)$ are functions of time t while they consider β_{ik} as a constant. Let y_{ij} and $X_{ij} = (X_{ij1}, X_{ij2})^{\mathrm{T}}$ be the observations for TMAX, RAIN, and SUN, respectively, from the ith location at the jth month, $i = 1, \cdots, 16$, $j = 1, \cdots, 240$. We standardize the data and our estimation model is

$$y_{ij} = \beta_{i0} + X_{ij1}\beta_{i1}(t_{ij}) + X_{ij2}\beta_{i2}(t_{ij}),$$
$$i = 1, \cdots, 16, j = 1, \cdots, 240.$$

We use the estimation method by the LAD-estimation. Table 3 reports the subgroup structure based on the UK climate data. $\{1, 2, \cdots, 16\}$ represents the locations from Armagh to Whitby. To INTERCEPT, all locations follow one group, $\widehat{\beta}_{01} = -0.0358$. To the covariate RAIN, all locations are divided into three subgroups

$$\widehat{G}_1 = \{G_{1,1}, G_{2,1}, G_{3,1}\}, G_{1,1} = \{1,5,6,10,12\},$$
$$G_{2,1} = \{2,3,9,11,13,14,15\}, G_{3,1} = \{4,7,8,16\},$$

and $\beta_{i1} = \widehat{\beta}_{u1}$, when $i \in G_{u,1}$.

To the covariate SUN, all locations are divided into four subgroups,

$$\widehat{G}_2 = \{G_{1,2}, G_{2,2}, G_{3,2}, G_{4,2}\}, G_{1,2} = \{1\},$$
$$G_{2,2} = \{2,4,6,7,9,10,11,12,13,14,15,16\},$$
$$G_{3,2} = \{3,8\}, G_{4,2} = \{5\}.$$

$\beta_{i2} = \widehat{\beta}_{u2}$, when $i \in G_{u,2}$. Figure 3 shows the estimated coefficient functions.

Based on the subgroup structure of the predictor "INTERCEPT", "RAIN" and "SUN". We can divide all locations into seven subgroups,

$$\widehat{G} = \{G_1, G_2, G_3, G_4, G_5, G_6, G_7\}, G_1 = \{6,10,12\},$$
$$G_2 = \{2,9,11,13,14,15\}, G_3 = \{3\},$$
$$G_4 = \{4,7,16\}, G_5 = \{5\}, G_6 = \{8\}, G_7 = \{1\}.$$

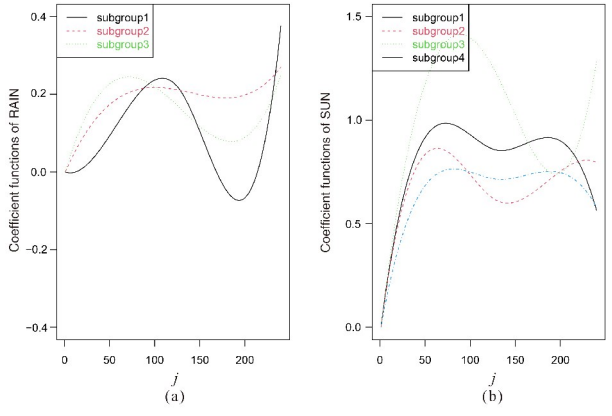

Figure 3. β_{i1} and β_{i2} in different subgroups by the LAD-estimation approach.

So total rainfall and total sunshine duration in two different locations in the same subgroup have the same impact on the mean daily maximum temperature.

Figure 4. UK map and subgroup structure by the LAD-estimation approach.

Figure 5. UK map and subgroup structure by the LS-estimation approach.

Table 4. Subgroup structures based on the observations in 16 locations by the LS-estimation approach.

	Armagh	Bradford	Camborne	Eastbourne	Eskdalemuir	Heathrow	Hurn	Lerwick
INTERCEPT	1	1	1	1	1	1	1	1
RAIN	1	1	1	1	1	1	1	1
SUN	1	2	3	4	5	2	6	3
	Leuchars	Oxford	Paisley	Ross-On-Wye	Shawbury	Sheffield	Waddington	Whitby
INTERCEPT	1	1	1	1	1	1	1	1
RAIN	1	1	1	1	1	1	1	1
SUN	6	6	5	6	6	2	2	6

From Figure 4, we can see that the similarity between the subgroup structures identified by our method and the geographical location in reality. Armagh, Camborne, Eskdalemuir, Lerwick are isolated and far away from each other and other 12 locations, so we divides them into different subgroups, every location as one-member subgroup. At the same time, Ross-On-Wye, Heathrow and Oxford are divided into the same subgroup while they are in close proximity. Eastbourne, Hurn and Whitby are divided into the same subgroup because they are all located by the sea.

For comparison, we use the estimation method by the LS-estimation while other steps are the same as the estimation method by the LAD-estimation. From the Table 4, based on the subgroup structure of the predictor "INTERCEPT", "RAIN" and "SUN", we can see that all locations are divided into six subgroups,

$$\widehat{G} = \{G_1, G_2, G_3, G_4, G_5, G_6\}, G_1 = \{1\},$$
$$G_2 = \{2,6,14,15\}, G_3 = \{3,8\}, G_4 = \{4\},$$
$$G_5 = \{5,11\}, G_6 = \{7,9,10,12,13,16\}.$$

From Figure 5, the LAD-estimation approach works more reasonable than the LS-estimation approach. For example, the estimation method divides Bradford, Heathrow, Sheffield and Waddington into one subgroup, while Heathrow is much different from other three locations.

5 Conclusions

In this paper, we proposed a subgroup identification algorithm to identify homogeneity in varying coefficient model for panel data. Instead of using the LS-estimation, we choose to use the LAD-estimation method to account for the potential outliers and heavy-tailedness of observed distribution. Numerical studies indicate our algorithm resulting in satisfactory performance. Additionally, the corresponding theoretical properties similar to Refs. [11,25] can also be investigated.

There are some issues can be studied in the future. First, in our proposed method, the threshold δ can be selected by the CV procedure or the standard BIC. Whether there is a better way to select a suitable δ to meet different scenarios or not, furthermore, varying δ to meet different kinds of predictors. Second, we use least-absolute estimation $\rho(u) = |u|$ to account for the robustness consideration. Other robust estimation functions may get a better estimation accuracy, and select different ρ to meet different scenarios.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Nos. 12171450, 71771203).

Conflict of interest

The authors declare no conflict of interest.

Author information

TANG Heng received his MS degree from University of Science and Technology of China. His research interests focus on longitudinal data analysis.

ZHANG Weiping (corresponding author) received his PhD degree in Statistics from University of Science and Technology of China (USTC). He is currently a Professor at USTC. His research interests include statistical methods in linear model and generalized linear model, Bayesian statistics, and longitudinal data analysis.

References

[1] Lian H, Qiao X, Zhang W. Homogeneity pursuit in single index models based panel data analysis. Journal of Business & Economic Statistics, 2019, 39(2): 386-401.

[2] Wang H, Xia Y. Shrinkage estimation of the varying coefficient model. Journal of the American Statistical Association, 2009, 104: 747-757.

[3] Xue L, Zhu L. Empirical likelihood for a varying coefficient model with longitudinal data. Journal of the American Statistical Association, 2007, 102: 642-654.

[4] Park M Y, Hastie T, Tibshirani R. Averaged gene expressions for regression. Biostatistics, 2007, 8(2): 212-227.

[5] Friedman J, Hastie T, Höfling H, et al. Pathwise coordinate optimization. The Annals of Applied Statistics, 2007, 1(2): 302-332.
[6] Tibshirani R, Saunders M, Rosset S, et al. Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2005, 67(1): 91-108.
[7] Bondell H D, Reich B J. Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. Biometrics, 2008, 64(1): 115-123.
[8] Zhu Y, Shen X, Pan W. Simultaneous grouping pursuit and feature selection over an undirected graph. Journal of the American Statistical Association, 2013, 108: 713-725.
[9] Yang Y, He X. Bayesian empirical likelihood for quantile regression. The Annals of Statistics, 2012, 40(2): 1102-1131.
[10] Ke Z T, Fan J, Wu Y. Homogeneity pursuit. Journal of the American Statistical Association, 2015, 110: 175-194.
[11] Wang W, Phillips P C, Su L. Homogeneity pursuit in panel data models: Theory and application. Journal of Applied Econometrics, 2018, 33(6): 797-815.
[12] Ke Y, Li J, Zhang W, et al. Structure identification in panel data analysis. The Annals of Statistics, 2016, 44(3): 1193-1233.
[13] Li J, Yue M, Zhang W. Subgroup identification via homogeneity pursuit for dense longitudinal/spatial data. Statistics in Medicine, 2019, 38: 3256-3271.
[14] Li F, Sang H. Spatial homogeneity pursuit of regression coefficients for large datasets. Journal of the American Statistical Association, 2019, 114: 1050-1062.
[15] Wu C O, Chiang C T, Hoover D R. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. Journal of the American Statistical Association, 1998, 93: 1388-1402.
[16] Cai Z, Fan J, Li R. Efficient estimation and inferences for varying-coefficient models. Journal of the American Statistical Association, 2000, 95: 888-902.
[17] Sun Y, Carroll R J, Li D. Semiparametric estimation of fixed-effects panel data varying coefficient models. In: Nonparametric Econometric Methods. Bingley, UK: Emerald Group Publishing Limited, 2009.
[18] Li D, Chen J, Gao J. Non-parametric time-varying coefficient panel data models with fixed effects. The Econometrics Journal, 2011, 14(3): 387-408.
[19] Rodriguez-Poo J M, Soberon A. Direct semi-parametric estimation of fixed effects panel data varying coefficient models. The Econometrics Journal, 2014, 17(1): 107-138.
[20] Li D, Ke Y, Zhang W, et al. Model selection and structure specification in ultra-high dimensional generalised semi-varying coefficient models. The Annals of Statistics, 2015, 43(6): 2676-2705.
[21] Tang Q, Wang J. L_1-estimation for varying coefficient models. Statistics, 2005, 39(5): 389-404.
[22] Zhang R, Zhao W, Liu J. Robust estimation and variable selection for semiparametric partially linear varying coefficient model based on modal regression. Journal of Nonparametric Statistics, 2013, 25(2): 523-544.
[23] Yang H, Lv J, Guo C. Robust estimation and variable selection for varying-coefficient single-index models based on modal regression. Communications in Statistics: Theory and Methods, 2016, 45(14): 4048-4067.
[24] Jiang Y, Ji Q, Xie B. Robust estimation for the varying coefficient partially nonlinear models. Journal of Computational and Applied Mathematics, 2017, 326: 31-43.
[25] Tang Q, Cheng L. M-estimation and B-spline approximation for varying coefficient models with longitudinal data. Journal of Nonparametric Statistics, 2008, 20(7): 611-625.
[26] Rousseeuw P J, Leroy A M. Robust Regression and Outlier Detection. Hoboken, NJ: Wiley, 2005.
[27] Ling S. Self-weighted least absolute deviation estimation for in finite variance autoregressive models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2005, 67(3): 381-393.
[28] He X, Zhu Z Y, Fung W K. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. Biometrika, 2002, 89(3): 579-590.
[29] Honda T. Quantile regression in varying coefficient models. Journal of Statistical Planning and Inference, 2004, 121(1): 113-125.
[30] Cho H, Fryzlewicz P. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. Journal of the Royal Statistical Society: Series B: Statistical Methodology, 2015, 77(2): 475-507.
[31] Fryzlewicz P. Wild binary segmentation for multiple change-point detection. Annals of Statistics, 2014, 42(6): 2243-2281.
[32] Fan J, Feng Y, Song R. Nonparametric independence screening in sparse ultra-high-dimensional additive models. Journal of the American Statistical Association, 2011, 106: 544-557.
[33] Huang J, Horowitz J L, Wei F. Variable selection in nonparametric additive models. Annals of Statistics, 2010, 38(4): 2282-2313.

纵向数据下变系数模型的一种稳健同质寻踪算法

汤恒,郑智,张伟平*

中国科学技术大学管理学院统计与金融系,安徽合肥 230026

* 通讯作者. E-mail:zwp@ ustc. edu. cn

摘要: 探讨了变系数模型中参系数函数的同质性,其中同一个子群中的个体的系数函数是相同的. 在重复观测的条件下,我们用 B 样条来拟合变系数模型的系数函数,同时用变点检测的方法来进行子群识别. 为了解释可能的异常值或重尾分布,我们在 M 估计的框架下拟合系数函数,在本文中以绝对值(LAD)损失为例. 模拟数据表明,当模拟数据集存在异常值或参数函数为重尾分布时,我们的估计方法优于常用的最小二乘(LS)估计.

关键词: 变系数模型;M 估计;B 样条;变点检测 ;同质寻踪