Statistics in Biosciences
https://doi.org/10.1007/s12561-019-09248-0

Nonparametric Time-Varying Coefficient Models for Panel Data

Huazhen Lin[1] · Hyokyoung G. Hong[2] · Baoying Yang[3] · Wei Liu[1] · Yong Zhang[4] · Gang-Zhi Fan[5] · Yi Li[6]

Received: 19 April 2018 / Revised: 18 March 2019 / Accepted: 18 June 2019

Abstract

The collection rate of contributions to public pension (CRCP), expressed as the ratio of the actual contributions to the expected contributions from insurers, is a key component of the public pension system in China. Recent years have seen various patterns of change in CRCPs at the provincial level. In order to study the drastic changes in a short time and understand their underlying implications, we propose a nonparametric time-varying coefficients model for longitudinal data with pre-specified finite time points, also known as panel data. By utilizing a penalized least squares method, the proposed method enables estimation of a large number of parameters, which can exceed the sample size. The resulting estimator is shown to be efficient, robust, and computationally feasible. Furthermore, it possesses desirable theoretical properties such as $n^{1/2}$-consistency, asymptotic normality, and the oracle property.

Keywords Collection rate of public pension contributions · Nonparametric time-varying coefficients model · Panel data · Penalized least squares estimation

✉ Yi Li
yili@umich.edu

Huazhen Lin
linhz@swufe.edu.cn

[1] Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, China

[2] Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA

[3] Department of Statistics, College of Mathematics, Southwest Jiaotong University, Chengdu, China

[4] School of Insurance, Southwestern University of Finance and Economics, Chengdu, China

[5] Department of Real Estate Studies, Konkuk University, Seoul 143-701, Korea

[6] Department of Biostatistics, University of Michigan, Ann Arbor, USA

Published online: 27 June 2019

1 Introduction

Over the last decade, a massive number of rural workers in China have migrated to major cities for higher paying jobs. As a result, public pension arrears, as reported by the Chinese Ministry of Social Security (CMSS), increased from 28.06 billion Chinese yuan in 1998 to 43.06 billion in 2006. While contribution evasion rates are a global issue impacting many developing countries [50], the CMSS has also reported in a special audit that the situation has degenerated in recent years. In order to maintain and improve the public pension system, it is critical to understand the system and identify risk factors that may impact contribution evasion rates. A key measure for a sustainable public pension system for urban employees in China is the collection rate of contributions to public pension (CRCP), i.e., the ratio of the actual contributions to the expected contributions from insurers. A high CRCP indicates a healthy and stable public pension system, whereas a low CRCP often implies contribution evasions and may result in a system breakdown [43].

The literature has identified the following six factors that may play an important role in influencing CRCPs: per capita disposable income, the consumer price index, the enterprise scale, the unemployment rate, the proportion of state-owned enterprise workers' salary in the total society wages, and the choice of collection institutions [14,37]. However, conflicting results were reported, mostly due to the unavailability of data and lack of proper statistical methods [14,17,35–37]. This paper systemically explores the relationship of the aforementioned factors with the CRCP. We use the relevant data of 30 provinces in China available from 2002 to 2015 from the Chinese Statistical Yearbook (http://www.stats.gov.cn/tjsj/ndsj/2015/indexeh.htm).

The collected data are longitudinal with some pre-specified finite time points, also known as panel data. Figure 1 shows strong geographic and temporal variations of the CRCP: the eastern region, which is more developed, presents lower CRCPs than the middle and western regions, which are typically less developed; CRCPs keep increasing over time in many regions. The variations and the increasing trend may be due to the imbalanced distributions of the relevant risk factors as well as their potential time-varying changes. Moreover, Fig. 2, which depicts the CRCP against each of the six factors over time, hints at possible time-varying effects as well. To elucidate such complex relationships between the risk factors and CRCPs, we resort to a time-varying coefficients model for the panel data:

$$Y_{it} = f_t + \sum_{j=1}^{p} \beta_{t,j} X_{it,j} + \alpha_i + \varepsilon_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{1.1}$$

where Y_{it} is the CRCP of province i at time t, $X_{it} = (X_{it,1}, \ldots, X_{it,p})'$ are p risk factors in province i at time t, $\beta_t = (\beta_{t1}, \ldots, \beta_{tp})'$ and f_t are unknown functions of t, α_i reflects the heterogeneity of province i, and ε_{it} is independent random errors. In our data, $T = 13$ and $n = 30$. For identifiability, we assume that $f_1 = 0$. Model (4.1) is termed a fixed-effects time-varying coefficients model when α_i is allowed to be correlated with X_{it}, or a random-effects time-varying coefficients model when α_i is uncorrelated with X_{it}.

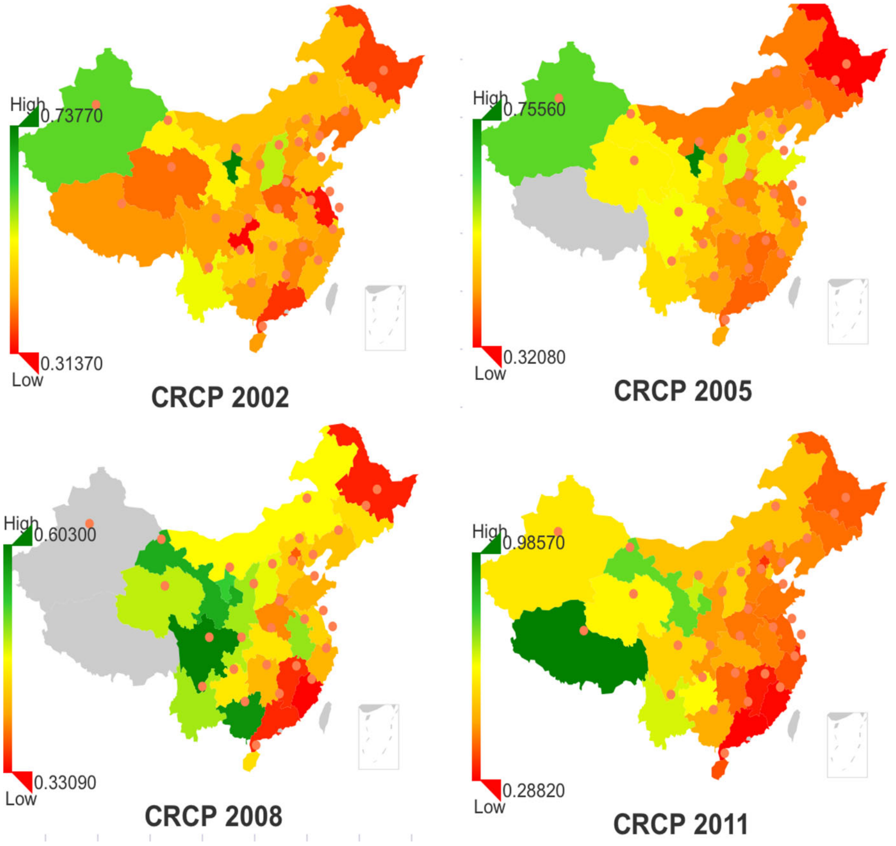
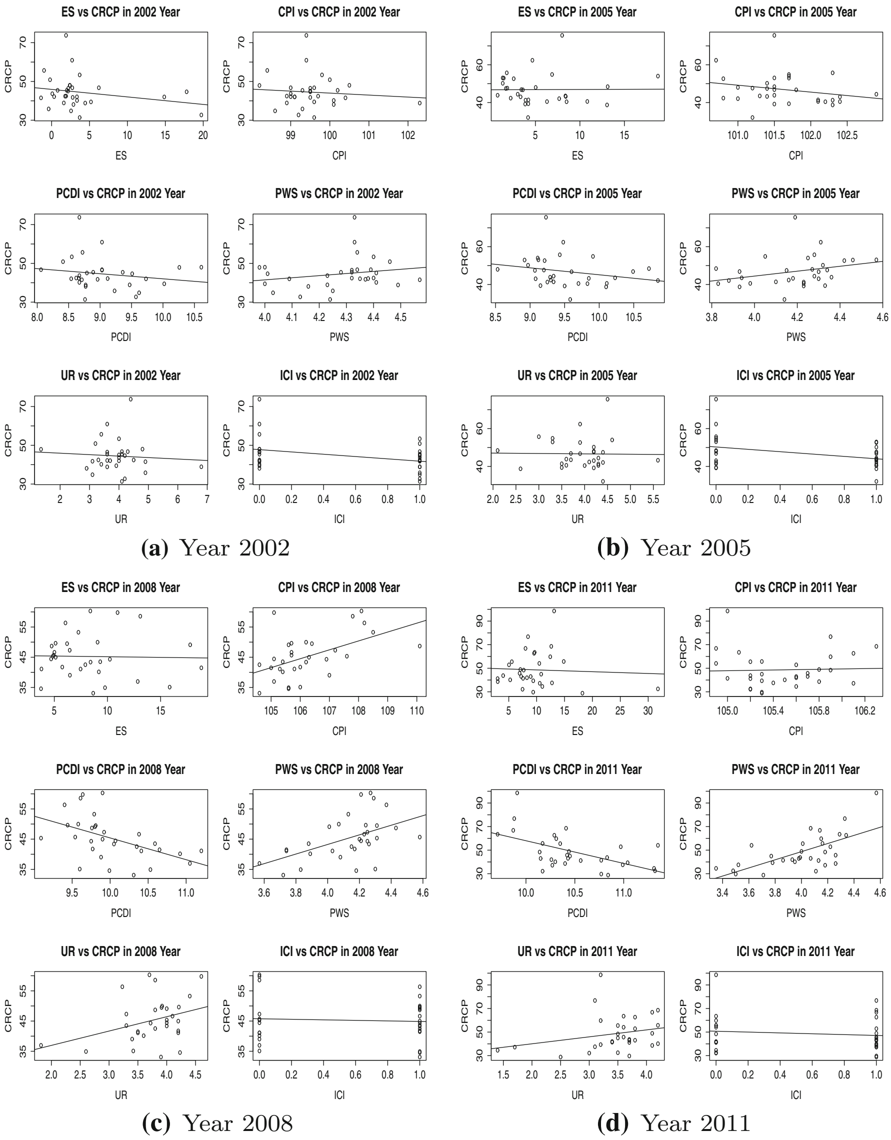


Fig. 1 The plot of the collection rate of contributions to public pension (CRCP) in 2002, 2005, 2008, and 2011

The seminal work of Hastie and Tibshirani [20] on time-varying coefficients models has motivated applications in longitudinal data analysis [11,13,22,26,28,55], time series analysis [3,4,23], survival analysis [2,12,16,31–34,52,60]; and [5], and functional data analysis [44].

Limited work, however, has been conducted for time-varying coefficient models with panel data. Robinson [48] first introduced (4.1) without the explanatory variables X_{it} for large T and small n. Li et al. [25] considered (4.1) for T tending to infinity. When T is sufficiently large, Robinson [48] and Li et al. [25] applied the local linear technique [7] to estimate f_t and β_t with $f_t = f(t/T)$ and $\beta_t = (\beta_1(t/T), \ldots, \beta_p(t/T))$ [1,46]. When T goes to infinity, the distance between any two points in $\{t/T;\ t = 1, \ldots, T\}$ tends to 0, the local linear approximation biases can be ignored and asymptotic results of the resulting estimator can be easily established. In this case, the kernel smoother methods, e.g., those developed by Qian and Wang [42] and Rodriguez-Poo and Soberon [49] in a similar context, may also be applicable. However, when T is fixed, the distance between two time points does not tend to 0 even when the sample size n goes to infinity, resulting in large local linear approximation biases and difficulties in



(a) Year 2002

(b) Year 2005

(c) Year 2008

(d) Year 2011

Fig. 2 The scatter plots of CRCP versus six factors at selected years

establishing asymptotic theories. Except for some ad hoc methods that present f_t and β_t as distinct parameters at each t, little progress has been made with fixed values of T. Without any constraints on f_t and β_t, the number of parameters to be estimated will be more than $T(p+1)$, which typically exceeds the sample size n.

We propose a flexible and computationally feasible approach to estimating f_t and β_t for panel data with a fixed T. As our focus is not on predicting the CRCP for each province, we treated α_i as a nuisance parameter. We do not impose any parametric assumptions on f_t or β_t. Instead, we only assume that f_t and β_t vary smoothly with t [7], Hastie et al. 2001. In general, such an assumption is expressed as a smoothing property of some unknown functions. However, when t is discrete, the ordinary definition of smoothness does not apply. We therefore modify the smoothness assumption by assuming that f_t and f_s and β_t and β_s remain close to each other when t and s are contiguous. Consequently, if we can identify time intervals within which f_t or β_t are constant, we effectively achieve dimension reduction for the parameter space.

We propose a penalized least squares method to identify the jump points of f_t and β_t and estimate the time-varying regression functions f_t and β_t simultaneously. Our method is flexible and data-driven as it does not require *a priori* specifications of the number and locations of the jump points of f_t and β_t. In addition, the estimator has the desirable $n^{1/2}$-consistency, asymptotic normality, and the oracle property, i.e., the resulting estimator is as efficient as if the jump points of f_t and β_t were known. The proposed method is applied to the CRCP data, and we have identified two significant factors for the CRCP, the per capita disposable income and the unemployment rate, which were not detected by the available methods.

The remainder of this paper is organized as follows. Section 2 introduces the nonparametric time-varying effects model for panel data and proposes a penalized least squares method for inference. Section 3 compares the proposed method with the existing ones, and Sect. 4 details the analysis of the CRCP study. We conclude the paper with some remarks in Sect. 5. Appendix details the proofs of the theorems.

2 The Proposed Model and Method

Consider n independent subjects, each with observations (X_{it}, Y_{it}), $t = 1, \ldots, T$, $i = 1, \ldots, n$, generated by (4.1). Let $g_t = f_t - f_{t-1}$ and $\gamma_{tj} = \beta_{t,j} - \beta_{t-1,j}$ with $f_0 = 0$ and $\beta_{0,j} = 0$ for $j = 1, \ldots, p$. Thus $f_t = f_{t-1}$ and $\beta_{t,j} = \beta_{t-1,j}$ if $g_t = 0$ and $\gamma_{tj} = 0$, respectively. Therefore, the number and the location of the jump points of f_t and β_t can be determined by the number and the location of nonzero parameters in $\{g_t, \gamma_{tj}, t = 1, \ldots, T, j = 1, \ldots, p\}$, leading to a reformulation of (4.1) as

$$Y_{it} = \sum_{d=1}^{t} g_d + \sum_{j=1}^{p} \sum_{d=1}^{t} \gamma_{dj} X_{it,j} + \alpha_i + \varepsilon_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T. \tag{2.1}$$

Taking the difference of Y_{it} and Y_{is} for $t > s$, we have

$$Y_{it} - Y_{is} = \sum_{d=s+1}^{t} g_d + \sum_{j=1}^{p} \left(\sum_{d=1}^{t} \gamma_{dj} X_{it,j} - \sum_{d=1}^{s} \gamma_{dj} X_{is,j} \right) + \varepsilon_{i,ts}, \tag{2.2}$$

where $\varepsilon_{i,ts} = \varepsilon_{it} - \varepsilon_{is}$, $i = 1, \ldots, n$, $0 \le s < t$, and $1 \le t \le T$. To identify and estimate the nonzero parameters in $\{g_t, \gamma_{tj}, t = 1, \ldots, T, j = 1, \ldots, p\}$ while encouraging the "smoothness" of f_t and β_t, we propose the following penalized least squares estimation,

$$L_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{t>s}\left\{Y_{it} - Y_{is} - \sum_{d=s+1}^{t} g_d - \sum_{j=1}^{p}\left(\sum_{d=1}^{t}\gamma_{dj}X_{it,j} - \sum_{d=1}^{s}\gamma_{dj}X_{is,j}\right)\right\}^2 + \sum_{t=2}^{T} p_\lambda(|g_t|) + \sum_{t=1}^{T}\sum_{j=1}^{p} p_\lambda(|\gamma_{tj}|), \tag{2.3}$$

with $g_1 = f_1 - f_0 = 0$ for identifiability purposes, where $\theta = (g_2, \ldots, g_T, \gamma_1', \ldots, \gamma_T')'$ and $\gamma_t = (\gamma_{t1}, \ldots, \gamma_{tp})'$. Here, $p_\lambda(\cdot)$ is a penalty function, such as $p_\lambda(|\beta|) = \lambda|\beta|^q$ with $\lambda > 0$, $q > 0$, which yields the well-known ridge regression with $q = 2$ and Lasso penalized regressions [53] with $q = 1$. Another popular choice is the smoothly clipped absolute deviation (SCAD) penalty function [9].

By minimizing $L_n(\theta)$, we show that there exists a positive probability that some g_t and γ_{tj} can be estimated to be exactly zero, lending support to the usage of the automated selection of the jump points f_t and β_t and the estimation of the time-varying regression, simultaneously.

To proceed, let $\theta = (\theta^{(1)'}, \theta^{(2)'})'$, where $\theta^{(1)} = (\theta_1, \ldots, \theta_m)'$, with m being the number of nonzero parameters, and $\theta^{(2)} = (\theta_{m+1}, \ldots, \theta_{Tp+T-1})'$. Throughout the paper, the subscript "0" represents the true value. Without loss of generality, we assume that $\theta_0^{(2)} = 0$. We consider a general nonconcave penalty function $p_\lambda(\cdot)$ for (2.3). Let $a_n = \max_{1\le j\le m} \dot{p}_\lambda(|\theta_{j0}|)$, $\Sigma = \text{diag}\{\ddot{p}_\lambda(|\theta_{10}|), \ldots, \ddot{p}_\lambda(|\theta_{m0}|)\}$, $b = (\dot{p}_\lambda(|\theta_{10}|)\text{sgn}(\theta_{10}), \ldots, \dot{p}_\lambda(|\theta_{m0}|)\text{sgn}(\theta_{m0}))'$, $\Lambda = E\left(\sum_{t>s} W_{i,ts}(\theta)^{\otimes 2}|_{\theta=\theta_0}\right)$, and $\Upsilon = E\left[\sum_{t>s}\varepsilon_{i,ts}W_{i,ts}(\theta)\right]^{\otimes 2}|_{\theta=\theta_0}$, where

$$W_{i,ts}(\theta) = \frac{\partial}{\partial\theta^{(1)}}\left\{\sum_{d=s+1}^{t} g_d + \sum_{j=1}^{p}\left(\sum_{d=1}^{t}\gamma_{dj}X_{it,j} - \sum_{d=1}^{s}\gamma_{dj}X_{is,j}\right)\right\}, \tag{2.4}$$

and $b^{\otimes 2} = bb'$ for any vector b. We have the following theorems, whose proofs are given in the appendix.

Theorem 1 *If* $\max_{1\le j\le m} \ddot{p}_\lambda(|\theta_{j0}|) \to 0$, *then there exists a minimizer* $\widehat{\theta}$ *of* $L_n(\theta)$ *such that*

$$\|\widehat{\theta} - \theta_0\| = O_p(n^{-1/2} + a_n).$$

Consider, for example, the SCAD penalized function. An $n^{1/2}$- consistent estimator is obtained based on a proper λ and $a_n = 0$. Its oracle property will be stated in the following theorem.

Theorem 2 (Oracle Property) *Assume* $\liminf_{n\to\infty,\theta\to 0^+} \dot{p}_\lambda(\theta)/\lambda > 0$. *If* $\lambda \to 0$ *and* $\sqrt{n}\lambda \to \infty$ *as* $n \to \infty$, *the* $n^{1/2}$*−consistent local minimizer,* $\widehat{\theta} = (\widehat{\theta^{(1)'}}, \widehat{\theta^{(2)'}})'$, *satisfies*

(a) *Sparsity:* $\widehat{\theta}^{(2)} = 0$ *with probability going to 1.*
(b) $\sqrt{n}(2\Lambda + \Sigma)\{\widehat{\theta}^{(1)} - \theta_0^{(1)} + (2\Lambda + \Sigma)^{-1}b\} \rightarrow N(0, 4\Upsilon)$.

Theorem 2 has important implications. For the SCAD penalty, when $\lambda \rightarrow 0$, it follows that $b = 0$ and $\Sigma = 0$. Therefore,

$$\sqrt{n}\left(\widehat{\theta}^{(1)} - \theta_0^{(1)}\right) \rightarrow N(0, \Lambda^{-1}\Upsilon\Lambda^{-1})$$

in distribution. That is, the penalized least squares estimator for $\theta^{(1)}$ performs as well as the least squares estimator for estimating $\theta^{(1)}$ when $\theta^{(2)} = 0$ is known *a priori*.

Because the SCAD penalty function satisfies (1) continuity, (2) sparsity, and (3) unbiasedness and the estimators based on the penalized function using the penalty SCAD have the “oracle property,” we use the SCAD penalty as our penalty function in the numerical studies and real data analysis. However, as the SCAD penalty function $p_\lambda(\cdot)$ is nonconcave, we use a local quadratic algorithm [9] to accommodate time-varying regression models for panel data.

Given an initial value $\theta_j^{[0]}$, for example, an un-penalized estimate, we consider the following local quadratic approximation for $p_\lambda(|\theta_j|)$:

$$p_\lambda(|\theta_j|) \approx p_\lambda\left(|\theta_j^{[0]}|\right) + \frac{1}{2}\dot{p}_\lambda\left(|\theta_j^{[0]}|\right)\left(\theta_j^2 - (\theta_j^{[0]})^2\right)/|\theta_j^{[0]}|.$$

Define $S(\theta) = \frac{1}{n}\sum_{i=1}^n \sum_{t>s}\Big\{Y_{it} - Y_{is} - \sum_{d=s+1}^t g_d - \sum_{j=1}^p \Big(\sum_{d=1}^t \gamma_{dj} X_{it,j} - \sum_{d=1}^s \gamma_{dj} X_{is,j}\Big)\Big\}^2$, $U_\lambda(\theta) = \Big\{\dot{p}_\lambda(|\theta_1|)\frac{\theta_1}{|\theta_1|}, \ldots, \dot{p}_\lambda(|\theta_{T(p+1)-1}|)\frac{\theta_{T(p+1)-1}}{|\theta_{Tp+T-1}|}\Big\}'$, and $\Sigma_\lambda(\theta) = \mathrm{diag}\left\{\dot{p}_\lambda(|\theta_1|)/|\theta_1|, \ldots, \dot{p}_\lambda(|\theta_{Tp+T-1}|)/|\theta_{Tp+T-1}|\right\}$. We then conduct the following iterative estimation.

Step 1 Compute $\theta^{[1]}$ by the Newton–Raphson algorithm

$$\theta^{[1]} = \theta^{[0]} - \{\ddot{S}(\theta^{[0]}) + \Sigma_\lambda(\theta^{[0]})\}^{-1}\{\dot{S}(\theta^{[0]}) + U_\lambda(\theta^{[0]})\},$$

where $\dot{S}$ and $\ddot{S}$ are the first and second derivative of S, respectively.
Step 2 Repeat Step 1 until convergence.

For the selection of λ, Lin and Peng [27] showed that the BIC criterion performs well. By treating the number of nonzero estimates of parameters as an approximation to the generalized degree of freedom, DF_λ, we propose to select λ by minimizing

$$\mathrm{BIC}_\lambda = \log\{S(\widehat{\theta})\} + 2^{-1}\mathrm{DF}_\lambda n^{-1}\log n. \tag{2.5}$$

In practice, to approximate distribution and construct the confidence interval for $\theta^{(1)}$, we need to estimate the variances of $\widehat{\theta}^{(}1)$, which can be approximated by $\frac{1}{n}\widehat{\Lambda}^{-1}\widehat{\Upsilon}\widehat{\Lambda}^{-1}$, where

$$\widehat{\Lambda} = \frac{1}{n}\sum_{i=1}^n\sum_{t>s}\left\{W_{i,ts}(\widehat{\theta})\right\}^{\otimes 2}, \ \widehat{\Upsilon} = \frac{1}{n}\sum_{i=1}^n\left\{\sum_{t>s}\widehat{\varepsilon}_{i,ts}W_{i,ts}(\widehat{\theta})\right\}^{\otimes 2},$$

$\widehat{\varepsilon}_{i,ts} = Y_{it} - Y_{is} - \sum_{d=s+1}^{t} \widehat{g}_d - \sum_{j=1}^{p} \left(\sum_{d=1}^{t} \widehat{\gamma}_{dj} X_{it,j} - \sum_{d=1}^{s} \widehat{\gamma}_{dj} X_{is,j} \right)$, and $W_{i,ts}(\theta)$ is defined by (2.4).

3 Numerical Study

We evaluate the performance of the proposed method via extensive simulations. Data are generated from

$$Y_{it} = f_t + \beta_{t,1} X_{it,1} + \beta_{t,2} X_{it,2} + \alpha_i + \epsilon_{it}, \quad i = 1, \ldots, n \text{ and } t = 1, \ldots, T,$$

where $f_t = F(t)$, $\beta_{t,1} = F(T + t)$, and $\beta_{t,2} = F(T + 2t)$ with $F(x) = I(x < 3) + 2I(3 \leq x < 6) + 3I(6 \leq x < 8) + 4I(8 \leq x < 11) + 2I(11 \leq x < 17) + 3.5I(17 \leq x < 23) + 2.5I(x \geq 23)$. The random effect, α_i, is assumed to follow $N(0, 1)$, and the random error is generated from $\epsilon_{it} = 0.5\epsilon_{i,t-1} + \eta_{it}$, where the η_{it}'s are independently and identically distributed standard normal variables and $\epsilon_{i0} = 0$. We generate the covariates $\{X_{it,j}\}$ via following two settings.

Setting 1 For each $j = 1, 2$, $\{X_{it,j}\}$ is generated from the AR(1) process with $X_{it,j} = 0.5X_{i(t-1),j} + e_{it,j}$, $t = 1, \ldots, T$ and $X_{i0,j} = 0$, where the $e_{it,j}{}'s$ are independent standard normal random variable.

Setting 2 $\{X_{it,1}\}$ is generated from the AR(1) process with $X_{it,j} = 0.5X_{i(t-1),j} + e_{it,j}$ and $X_{it,2}$ from $X_{it,2} = r \times X_{it,1} + \sqrt{1 - r^2} \times e_{it,1}$. In this case, the correlation coefficient between $X_{it,1}$ and $X_{it,2}$ at a given t is approximately equal to $r = 0.5$ or $r = 0.8$, where the $e_{it,j}{}'s$, $j = 1, 2$ are independent standard normal random variables.

We consider two different sample sizes, $n = 100$ and $n = 200$, and two different total number of time points, $T = 5$ and $T = 10$. For each simulation configuration, a total of 500 data sets are generated. Given the subset $\vartheta \equiv \{\vartheta_k, k = 1, \ldots, K\} \subset \{f_t, \beta_{t,j} : t = 1, \ldots, T, j = 1, 2\}$, the performance of the estimators for ϑ is assessed based on Bias $= \{K^{-1} \sum_{k=1}^{K} (E\widehat{\vartheta}_k - \vartheta_k)^2\}^{1/2}$, SD $= \{K^{-1} \sum_{k=1}^{K} E(\widehat{\vartheta}_k - E\widehat{\vartheta}_k)^2\}^{1/2}$, and the mean squared error (MSE), which is defined as MSE $=$ Bias^2 $+$ SD^2, where $E\widehat{\vartheta}_k$ is approximated by the sample mean of $\widehat{\vartheta}_k$ based on the 500 simulations.

For each simulation configuration, we use (2.5) to select λ. The proposed method is compared with the ordinary least squares estimator without penalty (termed "Naive") and the least squares estimator with the known jump points for f_t and β_t (termed "Oracle"). To our knowledge, the naive estimator is the only available method to analyze time-varying panel data as distinct parameters at various time points t. The results of these methods are reported in Tables 1, 2, and 3. The performances of all of the method tend to improve when the sample sizes increase or the correlations between the covariate processes decrease. However, in most cases, the proposed estimator presents considerably smaller MSE than the Naive estimator.

Moreover, the SDs and MSEs of the proposed estimator are comparable to those of the oracle estimator, especially when the sample size is large, which hints the oracle property of the proposed estimator.

Furthermore, we examine the performance of the proposed criterion for selecting λ as in (2.5). Figure 3 depicts BIC versus $\|\widehat{\theta}-\theta_0\|$ for a "typical" sample of Setting 1 with $T=10$ and λ ranges from 0.05 to 1 for two different sample sizes $n=100$ and $n=200$. Figure 3 reveals that BIC tends to increase as $\|\widehat{\theta}-\theta_0\|$ increases. Therefore, it is expected that the λ selected by BIC would minimize $\|\widehat{\theta}-\theta_0\|$.

Finally, we test the accuracy of our standard error formula given in Sect. 2. The standard deviations, denoted by SD in Table 2, of 500 estimated f_t and β_{tj}, based on 500 simulations, can be regarded as the true standard errors. The average and the standard deviation of 500 estimated standard errors, denoted by SE_{ave} and SE_{std}, summarize the overall performance of the standard error formula. The performance of the standard error formula is quite satisfactory.

4 Analyzing Risk Factors of the Collection Rate of Contributions to Public Pension

The collection rate of public pension contributions (CRCP), the response variable, along with the explanatory variables, including per capita disposable income (PCDI), consumer price index (CPI), enterprise scale (ES), unemployment rate (UR), proportion of state-owned enterprise workers' salary to the total society wages (PWS), and insurance collection institutions (ICI), which is a binary indicator with ICI $=0$ if social security agencies; 1 if local tax agencies, were collected from 30 provinces from the years 2002 to 2015. In particular, PCDI has been suggested to reduce contribution evasions [40]. The rising cost of living is reflected in CPI, which might negatively affect the proportion of pension contributions [18]. We included ES since it represents the proportion of employees working for large- or medium-sized enterprises, which may affect the CRCP since employees in small enterprises are more likely to practice evasion due to their general lack of oversight and short survival times [45]. A high UR often implies a high sense of insecurity among insurers, motivating their pension contributions. We included PWS because state-owned enterprises typically have strict pension plans in China to prevent their employees from escaping the pension contributions [45]. The inclusion of ICI, identifying social security agencies versus local tax agencies, is due to the conjecture that local tax agencies are more efficient and forceful in collecting contributions [30].

We consider the following model:

$$\begin{aligned} CRCP_{it} = {} & f_t+\alpha_i+\beta_{t,1}PCDI_{it}+\beta_{t,2}CPI_{it}+\beta_{t,3}ES_{it}+\beta_{t,4}UR_{it}+\beta_{t,5}PWS_{it} \\ & +\beta_{t,6}ICI_{it}+\epsilon_{it}, \end{aligned}$$

where $i=1,\ldots,30,\ t=2002,\ldots,2015$. The tuning parameter λ was chosen by (2.5). The plot of the BIC versus λ is presented in Fig. 4a, which suggests that $\lambda=0.07$ is the best choice. The resulting time-varying coefficient functions along with their 95% confidence intervals are shown in Fig. 4b–h. The SDs are estimated by the method described in Sect. 2.

Table 1 Simulation results for Setting 1 with $T = 5, 10$

		$n = 100$			$n = 200$		
		Proposed	Naive	Oracle	Proposed	Naive	Oracle
$T = 5$							
$\{f_1, f_2\}$	Bias	0.00004	0.00146	0.00000	0.00000	0.00216	0.00000
	SD	0.00678	0.08109	0.00000	0.00000	0.05847	0.00000
	MSE	0.00005	0.00658	0.00000	0.00000	0.00342	0.00000
$\{f_3, f_4, f_5\}$	Bias	0.00834	0.00478	0.00299	0.00182	0.00415	0.00232
	SD	0.13253	0.14868	0.11009	0.08828	0.10797	0.07951
	MSE	0.01763	0.02213	0.01213	0.00780	0.01168	0.00633
$\{\beta_{1,1}, \beta_{2,1}\}$	Bias	0.00231	0.00277	0.00231	0.00201	0.00212	0.00223
	SD	0.08238	0.10394	0.08234	0.05941	0.07302	0.05934
	MSE	0.00679	0.01081	0.00679	0.00353	0.00534	0.00353
$\{\beta_{3,1}, \beta_{4,1}, \beta_{5,1}\}$	Bias	0.00045	0.00126	0.00046	0.00065	0.00164	0.00066
	SD	0.06722	0.09844	0.06720	0.04833	0.06832	0.04836
	MSE	0.00452	0.00969	0.00452	0.00234	0.00467	0.00234
$\{\beta_{1,2}, \beta_{2,2}, \beta_{3,2}, \beta_{4,2}, \beta_{5,2}\}$	Bias	0.00395	0.00471	0.00386	0.00097	0.00213	0.00093
	SD	0.05354	0.09935	0.05249	0.03703	0.06892	0.03704
	MSE	0.00288	0.00989	0.00277	0.00137	0.00475	0.00137
$T = 10$							
$\{f_1, f_2\}$	Bias	0.00000	0.00356	0.00000	0.00000	0.00011	0.00000
	SD	0.00000	0.08450	0.00000	0.00000	0.05700	0.00000
	MSE	0.00000	0.00715	0.00000	0.00000	0.00325	0.00000
$\{f_3, f_4, f_5\}$	Bias	0.01509	0.00290	0.00358	0.00925	0.00361	0.00312
	SD	0.13206	0.15684	0.11389	0.09014	0.10953	0.07940
	MSE	0.01767	0.02461	0.01298	0.00821	0.01201	0.00631
$\{f_6, f_7\}$	Bias	0.02096	0.00383	0.00616	0.01076	0.00417	0.00336
	SD	0.15257	0.16093	0.13630	0.10979	0.11812	0.10129
	MSE	0.02372	0.02591	0.01862	0.01217	0.01397	0.01027
$\{f_8, f_9, f_{10}\}$	Bias	0.02960	0.00929	0.00814	0.01294	0.00402	0.00297
	SD	0.14075	0.15616	0.12731	0.10004	0.11383	0.09358
	MSE	0.02069	0.02447	0.01627	0.01018	0.01297	0.00877
$\{\beta_{1,1}, \beta_{2,1}, \beta_{3,1}, \beta_{4,1}, \beta_{5,1}, \beta_{6,1}\}$	Bias	0.00035	0.00169	0.00011	0.00075	0.00197	0.00073
	SD	0.05317	0.09899	0.05262	0.03595	0.06837	0.03593
	MSE	0.00283	0.00980	0.00277	0.00129	0.00468	0.00129
$\{\beta_{7,1}, \beta_{8,1}, \beta_{9,1}, \beta_{10,1}\}$	Bias	0.00257	0.00303	0.00250	0.00107	0.00294	0.00107
	SD	0.06075	0.10000	0.06074	0.04258	0.07059	0.04257
	MSE	0.00370	0.01001	0.00370	0.00181	0.00499	0.00181
$\{\beta_{1,2}, \beta_{2,2}\}$	Bias	0.00178	0.00425	0.00369	0.00236	0.00126	0.00119
	SD	0.08225	0.10229	0.07947	0.05771	0.07295	0.05768
	MSE	0.00677	0.01048	0.00633	0.00334	0.00532	0.00333

Table 1 continued

		$n = 100$			$n = 200$		
		Proposed	Naive	Oracle	Proposed	Naive	Oracle
$\{\beta_{3,2}, \beta_{4,2}, \beta_{5,2},$ $\beta_{6,2}, \beta_{7,2}, \beta_{8,2},$ $\beta_{9,2}, \beta_{10,2}\}$	Bias	0.00036	0.00408	0.00024	0.00138	0.00386	0.00137
	SD	0.04242	0.09880	0.04147	0.02991	0.06907	0.02990
	MSE	0.00180	0.00978	0.00172	0.00090	0.00479	0.00090

Table 2 True and estimated standard errors for Setting 1 with $T = 10$ and $n = 200$

	SD	SE_{ave}	SE_{std}		SD	SE_{ave}	SE_{std}
$g_2 = 0.0$	0.00000	0.00000	0.00000	$g_7 = 0.0$	0.00000	0.00000	0.00000
$g_3 = 1.0$	0.10350	0.07843	0.00427	$g_8 = 1.0$	0.07930	0.07724	0.00412
$g_4 = 0.0$	0.00000	0.00000	0.00000	$g_9 = 0.0$	0.00000	0.00000	0.00000
$g_5 = 0.0$	0.00000	0.00000	0.00000	$g_{10} = 0.0$	0.00000	0.00000	0.00000
$g_6 = 1.0$	0.09693	0.07815	0.00349				
$\gamma_{1,1} = 2.0$	0.03104	0.03532	0.00225	$\gamma_{2,1} = 3.5$	0.05917	0.05720	0.00421
$\gamma_{1,2} = 0.0$	0.00000	0.00000	0.00000	$\gamma_{2,2} = 0.0$	0.00000	0.00000	0.00000
$\gamma_{1,3} = 0.0$	0.00000	0.00000	0.00000	$\gamma_{2,3} = -1.0$	0.06171	0.06191	0.00428
$\gamma_{1,4} = 0.0$	0.00000	0.00000	0.00000	$\gamma_{2,4} = 0.0$	0.00000	0.00000	0.00000
$\gamma_{1,5} = 0.0$	0.00000	0.00000	0.00000	$\gamma_{2,5} = 0.0$	0.00000	0.00000	0.00000
$\gamma_{1,6} = 0.0$	0.00000	0.00000	0.00000	$\gamma_{2,6} = 0.0$	0.00000	0.00000	0.00000
$\gamma_{1,7} = 1.5$	0.04948	0.05435	0.00440	$\gamma_{2,7} = 0.0$	0.00000	0.00000	0.00000
$\gamma_{1,8} = 0.0$	0.00000	0.00000	0.00000	$\gamma_{2,8} = 0.0$	0.00000	0.00000	0.00000
$\gamma_{1,9} = 0.0$	0.00000	0.00000	0.00000	$\gamma_{2,9} = 0.0$	0.00000	0.00000	0.00000
$\gamma_{1,10} = 0.0$	0.00000	0.00000	0.00000	$\gamma_{2,10} = 0.0$	0.00000	0.00000	0.00000

As suggested by the reviewer, we also fit a conventional model (without time-dependent effects) as a comparison for the panel data:

$$Y_{it} = f_t + \sum_{j=1}^{p} \beta_j X_{it,j} + \alpha_i + \varepsilon_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T. \tag{4.1}$$

Taking the difference of Y_{it} and Y_{is} for $t > s$ to avoid the effect of α_i , we have

$$Y_{it} - Y_{is} = \sum_{d=s+1}^{t} g_d + \sum_{j=1}^{p} \beta_j (X_{it,j} - X_{is,j}) + \varepsilon_{i,ts}, \tag{4.2}$$

where $\varepsilon_{i,ts} = \varepsilon_{it} - \varepsilon_{is}, i = 1, \ldots, n,\ 0 \leq s < t,$ and $1 \leq t \leq T$. Thus, we compared the proposed approach with the naive approach and the model (4.1) in Fig. 4b–h.

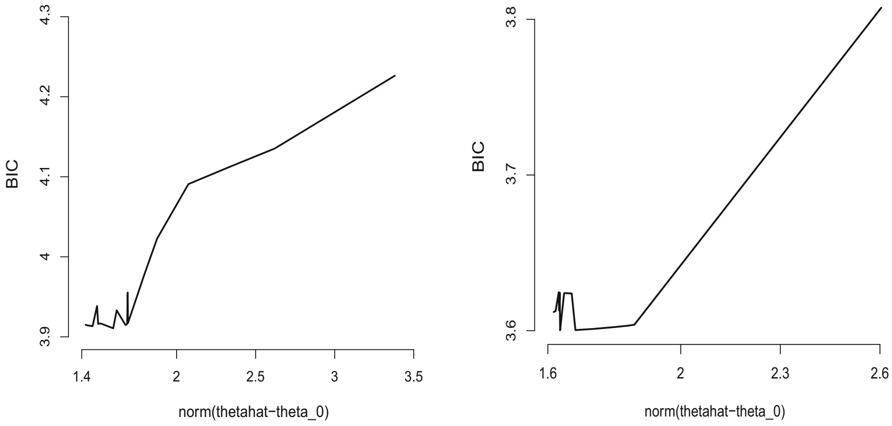
The results for the proposed and the naive methods showed similar trends, but the naive estimates had much wider 95% confidence intervals. Consequently, the naive

Table 3 Simulation results for Setting 2 with $T = 5$ and $r = 0.5, 0.8$

		$n = 100$			$n = 200$		
		Proposed	Naive	Oracle	Proposed	Naive	Oracle
$r = 0.5$							
$\{f_1, f_2\}$	Bias	0.00013	0.00021	0.00000	0.00000	0.00142	0.00000
	SD	0.00290	0.08216	0.00000	0.00000	0.06087	0.00000
	MSE	0.00001	0.00675	0.00000	0.00000	0.00371	0.00000
$\{f_3, f_4, f_5\}$	Bias	0.00864	0.00842	0.00638	0.00069	0.00775	0.00623
	SD	0.13344	0.14908	0.11136	0.08652	0.10678	0.07883
	MSE	0.01788	0.02230	0.01244	0.00749	0.01146	0.00625
$\{\beta_{1,1}, \beta_{2,1}\}$	Bias	0.00335	0.00236	0.00329	0.00214	0.00044	0.00201
	SD	0.08272	0.12133	0.08195	0.05923	0.08526	0.05914
	MSE	0.00685	0.01473	0.00673	0.00351	0.00727	0.00350
$\{\beta_{3,1}, \beta_{4,1}, \beta_{5,1}\}$	Bias	0.00189	0.00489	0.00183	0.00076	0.00326	0.00077
	SD	0.06972	0.11612	0.06942	0.05122	0.08326	0.05118
	MSE	0.00487	0.01351	0.00482	0.00262	0.00694	0.00262
$\{\beta_{1,2}, \beta_{2,2}, \beta_{3,2}, \beta_{4,2}, \beta_{5,2}\}$	Bias	0.00176	0.00406	0.00169	0.00034	0.00366	0.00035
	SD	0.05599	0.13014	0.05549	0.04021	0.09197	0.04022
	MSE	0.00314	0.01695	0.00308	0.00162	0.00847	0.00162
$r = 0.8$							
$\{f_1, f_2\}$	Bias	0.00013	0.00021	0.00000	0.00000	0.00142	0.00000
	SD	0.00290	0.08216	0.00000	0.00000	0.06087	0.00000
	MSE	0.00001	0.00675	0.00000	0.00000	0.00371	0.00000
$\{f_3, f_4, f_5\}$	Bias	0.00868	0.00842	0.00638	0.00069	0.00775	0.00623
	SD	0.13352	0.14908	0.11136	0.08652	0.10678	0.07883
	MSE	0.01790	0.02230	0.01244	0.00749	0.01146	0.00625
$\{\beta_{1,1}, \beta_{2,1}\}$	Bias	0.00618	0.00019	0.00219	0.00233	0.00180	0.00178
	SD	0.10686	0.18532	0.10105	0.07275	0.13055	0.07269
	MSE	0.01146	0.03434	0.01022	0.00530	0.01705	0.00529
$\{\beta_{3,1}, \beta_{4,1}, \beta_{5,1}\}$	Bias	0.00057	0.00676	0.00073	0.00045	0.00568	0.00100
	SD	0.09137	0.17687	0.08821	0.06534	0.12603	0.06531
	MSE	0.00835	0.03133	0.00778	0.00427	0.01592	0.00427
$\{\beta_{1,2}, \beta_{2,2}, \beta_{3,2}, \beta_{4,2}, \beta_{5,2}\}$	Bias	0.00311	0.00586	0.00244	0.00008	0.00529	0.00051
	SD	0.08393	0.18784	0.08009	0.05804	0.13275	0.05805
	MSE	0.00705	0.03532	0.00642	0.00337	0.01765	0.00337

method failed to detect the significance of f_t on CRCP, whereas the proposed method identified f_t significantly negative from 2009 to 2013. The estimates from the fixed effect model were significantly different from the proposed and the naive methods, especially for β_1 and β_5. This suggests that the fixed effect model may be not suitable for fitting the data.

Moreover, Fig. 4c–h illustrate the estimated curves for the effect functions corresponding to the six risk factors. The effect estimates of CPI, ES, and PWS were not



(a) $n = 100$ **(b)** $n = 200$

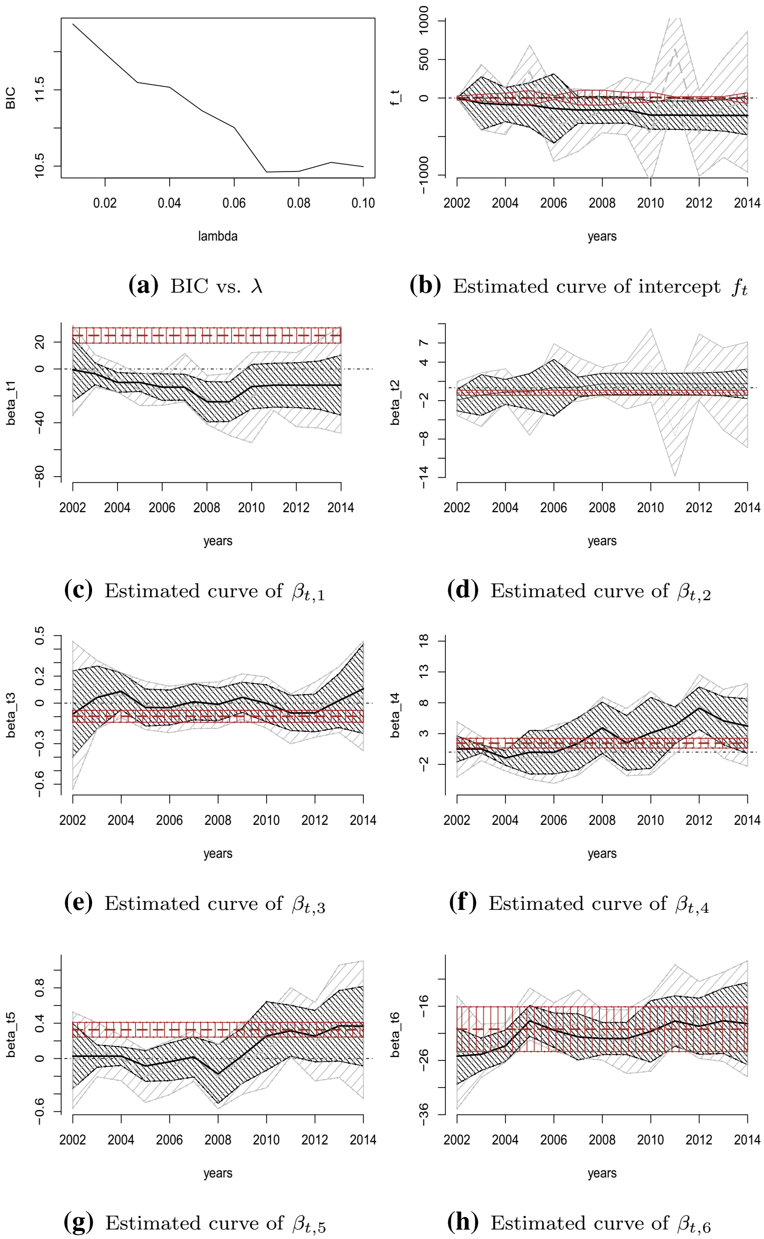
Fig. 3 BIC versus $\|\widehat{\theta} - \theta_0\|$ when λ changes over the interval (0.05, 1) for the generated data with Setting 1 with $T = 10$

significantly different from zero, suggesting that they might not influence the collection rates from 2002 to 2015.

However, the effect coefficients of PCDI ($\beta_{t,1}$), UR ($\beta_{t,4}$), and ICI ($\beta_{t,6}$) were significantly different from zero. The estimate of $\beta_{t,1}$ was significantly negative in 2008. This change might be attributed to a change in employees' perspectives on public pensions, as the soaring real estate market might have caused high-income employees to invest more in real estate than in public pension. The financial crisis in 2008 might have led to a positive estimate of $\beta_{t,4}$. This resulted in a high unemployment rate, forcing the employed insurers to increase the level of pension contributions. The estimate of $\beta_{t,6}$ was significantly negative from 2002 to 2014 and it has an increase from 2002 to 2005, with a peak in 2005. This might be because in 2005, the government decreased the individual payment ratio from 11 to 8% to temporarily stimulate public pension contributions. The result has largely confirmed the previous conjectures [30, 45].

5 Conclusion

We propose a flexible and computationally feasible approach to draw inferences on time-varying coefficients models for panel data with fixed time points. The proposed estimator is shown to have desirable theoretical properties, such as the $n^{1/2}$-consistency, asymptotic normality, and oracle property, indicating that this estimator is as efficient as if the jump points of f_t and β_t were known. Simulation studies validate the finite sample performance. We applied the proposed method to identify influential factors among PCDI, CPI, ES, UR, PWS, and ICI for the endowment insurance payment rate. The proposed method sheds light on mechanisms of the pension contribution system, which were unknown before.



(a) BIC vs. λ

(b) Estimated curve of intercept f_t

(c) Estimated curve of $\beta_{t,1}$

(d) Estimated curve of $\beta_{t,2}$

(e) Estimated curve of $\beta_{t,3}$

(f) Estimated curve of $\beta_{t,4}$

(g) Estimated curve of $\beta_{t,5}$

(h) Estimated curve of $\beta_{t,6}$

Fig. 4 The estimators of the time-varying coefficient functions for the CRCP. The solid black line represents the proposed estimator and the dark shadow shows the 95% confidence limit (CL) of the proposed estimator. The dashed gray line represents the Naive estimator and the light shadow indicates the 95% CL of the Naive estimator. The dashed brown line represents the no time-independent effect estimator and the brown shadow indicates the 95% CL of the Naive estimator (Color figure online)

As a future research direction, designing an efficient method to estimate province-specific effects would allow for proper predictions of the CRCP for each province.

Acknowledgements We thank the Editor, the AE, and two referees for the helpful suggestions that helped improve much the manuscript. The research is partially supported by the National Natural Science of China (No. 11829101, 11571282) and the Fundamental Research Funds for the Central Universities (JBK120509, JBK140507).

Appendix

Proof of Theorem 1

Let $\alpha_n = n^{-1/2} + a_n$. Denote by θ_0 the true value of θ. We want to show that for any given $\varepsilon > 0$, there exists a large constant C such that

$$\Pr\left\{ \inf_{\|u\|=C} L_n(\theta_0 + \alpha_n \cdot u) > L_n(\theta_0) \right\} \geq 1 - \varepsilon. \tag{5.1}$$

This implies with a probability larger than $1 - \varepsilon$ that there exists a local minimum in the ball $\{\theta_0 + \alpha_n \cdot u : \|u\| \leq C\}$. Hence, there exists a local minimizer such that $\|\widehat{\theta} - \theta_0\| = O_p(\alpha_n)$.

Define $\theta^* = \theta_0 + \alpha_n \cdot u = (\theta_1^*, \ldots, \theta_{Tp+T-1}^*)'$, using $p_\lambda(0) = 0$, we have

$$D_n(\theta^*) = L_n(\theta^*) - L_n(\theta_0) \geq S(\theta^*) - S(\theta_0) + \sum_{j=1}^{m} \{p_\lambda(|\theta_j^*|) - p_\lambda(|\theta_{j0}|)\},$$

where $S(\theta) = \frac{1}{n} \sum_{i=1}^{n} \sum_{t>s} \Big\{ Y_{it} - Y_{is} - \sum_{d=s+1}^{t} g_d - \sum_{j=1}^{p} \Big(\sum_{d=1}^{t} \gamma_{dj} X_{it,j} - \sum_{d=1}^{s} \gamma_{dj} X_{is,j} \Big) \Big\}^2$, m is the number of components of $\theta_0^{(1)}$. Let $\dot{S}$ be the gradient vector of S; by the standard argument of the Taylor expansion, we have

$$\begin{aligned} D_n(\theta^*) &\geq \dot{S}(\theta_0)'(\theta^* - \theta_0) + (\theta^* - \theta_0)' \ddot{S}(\theta_0)(\theta^* - \theta_0)\{1 + o_p(1)\} \\ &\quad + \sum_{j=1}^{m} [\dot{p}_\lambda(|\theta_{j0}|) \mathrm{sgn}(\theta_{j0})(\theta_j^* - \theta_{j0}) + \ddot{p}_\lambda(|\theta_{j0}|) \left(\theta_j^* - \theta_{j0}\right)^2 \{1 + o(1)\}] \\ &\widehat{=} I_1 + I_2 + I_3. \end{aligned} \tag{5.2}$$

Noting that $E\{\dot{S}(\theta_0)\} = 0$ and $Var\{\dot{S}(\theta_0)\} = O(n^{-1})$, by the central limit theory we have

$$I_1 = \left\{ E\{\dot{S}(\theta_0)\} + O_p\left(\sqrt{\mathrm{Var}(\dot{S}(\theta_0))}\right) \right\} (\theta^* - \theta_0) = O_p\left(\frac{\alpha_n}{\sqrt{n}}\right). \tag{5.3}$$

Similarly, we get

$$I_2 = O(\alpha_n^2 C^2). \tag{5.4}$$

For I_3, it is easy to see that it is bounded by

$$m\alpha_n a_n C + \alpha_n^2 \max\{|\ddot{p}_\lambda(|\theta_{j0}|)| : |\theta_{j0}| \neq 0\} C^2. \tag{5.5}$$

From (5.3), (5.4), and (5.5), I_1 and I_3 are dominated by I_2. Hence, by choosing a sufficiently large C, (5.1) holds. □

Proof of Theorem 2

We first show that with a probability tending to 1, for any given $\theta^{(1)}$ satisfying $\|\theta^{(1)} - \theta_0^{(1)}\| = O_p(n^{-1/2})$ and any constant C,

$$L_n((\theta^{(1)'}, 0')') = \min_{\|\theta^{(2)}\| \leq Cn^{-1/2}} L_n((\theta^{(1)'}, \theta^{(2)'})'). \tag{5.6}$$

To show (5.6), by Taylor's expansion, we have

$$\begin{aligned}\frac{\partial L_n(\theta)}{\partial \theta_r} &= -\frac{2}{n}\sum_{i=1}^{n}\sum_{t>s}\left\{Y_{it} - Y_{is} - \sum_{d=s+1}^{t} g_d - \sum_{j=1}^{p}\left(\sum_{d=1}^{t}\gamma_{dj}X_{it,j} - \sum_{d=1}^{s}\gamma_{dj}X_{is,j}\right)\right\} \\ &\quad \times \frac{\partial}{\partial \theta_r}\left\{\sum_{d=s+1}^{t} g_d + \sum_{j=1}^{p}\left(\sum_{d=1}^{t}\gamma_{dj}X_{it,j} - \sum_{d=1}^{s}\gamma_{dj}X_{is,j}\right)\right\} \\ &\quad + \dot{p}_\lambda(|\theta_r|)\mathrm{sgn}(\theta_r).\end{aligned}$$

By the central limit theorem, we have

$$\frac{\partial L_n(\theta)}{\partial \theta_r} = \lambda\{-\lambda^{-1}\dot{p}_\lambda(|\theta_r|)\mathrm{sgn}(\theta_r) + O_p(n^{-1/2}/\lambda)\},$$

where $\liminf_{n\to\infty}\liminf_{\theta\to 0^+} \lambda^{-1}\dot{p}_\lambda(\theta) > 0$ and $n^{-1/2}/\lambda \to 0$. The sign of the derivative is completely determined by that of θ_r. Hence (5.6) follows.

By (5.6), Part (a) follows. Now we prove Part (b). It can be shown that there exists $\widehat{\theta}^{(1)}$ in Theorem 1 that is a $n^{1/2}$- consistent local maximizer of $L_n((\theta^{(1)'}, 0')')$, which is regarded as a function of $\theta^{(1)}$, and that satisfies the following equation:

$$\left.\frac{\partial L_n(\theta)}{\partial \theta_r}\right|_{\theta=(\theta^{(1)},0)'} = 0, \qquad \text{for} \quad r = 1, \ldots, m.$$

Note that $\widehat{\theta}^{(1)}$ is a constant estimator. Thus, we have

$$\begin{aligned}0 = \frac{\partial L_n(\theta)}{\partial \theta_r}\bigg|_{\theta=(\theta^{(1)},0)'} &= \frac{S(\theta)}{\partial \theta_r}\bigg|_{\theta=(\theta^{(1)},0)'} + \dot{p}_\lambda(|\widehat{\theta_r}|)\mathrm{sgn}(\widehat{\theta_r}) \\ &= \frac{\partial S(\theta_0)}{\partial \theta_r} + \sum_{l=1}^{m}\left\{\frac{\partial^2 S(\theta_0)}{\partial\theta_r\theta_l} + o(1)\right\}(\widehat{\theta_l} - \theta_{l0}) \\ &\quad + \dot{p}_\lambda(|\theta_{r0}|)\mathrm{sgn}(\theta_{r0}) + \{\ddot{p}_\lambda(|\theta_{r0}|) + o_p(1)\}(\widehat{\theta_r} - \theta_{r0}).\end{aligned}$$

Furthermore, we have

$$\frac{\partial^2 S(\theta_0)}{\partial\theta^{(1)}\partial\theta^{(1)'}} = 2\varLambda(1 + o_p(1)),$$

where

$$\varLambda = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{t>s}\left[\frac{\partial}{\partial\theta^{(1)}}\left\{\sum_{d=s+1}^{t} g_d + \sum_{j=1}^{p}\left(\sum_{d=1}^{t}\gamma_{dj}X_{it,j} - \sum_{d=1}^{s}\gamma_{dj}X_{is,j}\right)\right\}\right]^{\otimes 2}\bigg|_{\theta=\theta_0}.$$

Hence following by Slutsky's theorem we have

$$\sqrt{n}(2\varLambda + \varSigma)\left\{\widehat{\theta}^{(1)} - \theta_0^{(1)} + (2\varLambda + \varSigma)^{-1}b\right\} = \sqrt{n}\frac{\partial S(\theta_0)}{\partial\theta^{(1)}} + o_p(1).$$

This completes the proof of Part (b). □

References

1. Cai Z (2007) Trending time-varying coefficient time series models with serially correlated errors. J Economet 136:163–188
2. Cai Z, Sun Y (2003) Local linear estimation for time-dependent coefficients in Cox's regression models. Scand J Stat 30:93–111
3. Cai Z, Fan J, Yao Q (2000) Functional-coefficient regression models for nonlinear time series models. J Am Stat Assoc 95:941–956
4. Chen R, Tsay RS (1993) Functional-coefficient autoregressive models. J Am Stat Assoc 88:298–308
5. Chen K, Lin H, Zhou Y (2012) Efficient estimation for the Cox model with varying coefficients. Biometrika 99:379–392
6. Efron B, Hastie T, Johnstone I, Tibshirani R (2004) Least angle regression (with discussion). Ann Stat 32:407–499
7. Fan J, Gijbels I (1996) Local polynomial modeling and its applications. Chapman and Hall, London
8. Fan J, Zhang W (1999) Statistical estimation in varying coefficient models. Ann Stat 27:1491–1518
9. Fan J, Li R (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc 96:1348–1360
10. Fan J, Yao Q (2003) Nonlinear time series: nonparametric and parametric methods. Springer, New York
11. Fan J, Zhang JT (2000) Two-Step estimation of functional linear models with applications to longitudinal data. J R Stat Soc B 62:303–322
12. Fan J, Lin H, Zhou Y (2006) Local partial likelihood estimation for life time data. Ann Stat 34:290–325

13. Fan J, Huang T, Li R (2007) Analysis of longitudinal data with semiparametric estimation of covariance function. J Am Stat Assoc 102:632–641
14. Feng J, He L, Satob H (2011) Public pension and household saving: evidence from urban China. J Comp Econ 39:470–485
15. Friedman J, Hastie T, Tibshirani R (2001) The elements of statistical learning. Springer series in statistics. Springer, New York
16. Gamerman D (1991) Markov chain Monte Carlo for dynamic generalized linear models. Biometrika 85:215–227
17. Gao Q (2010) Redistributive nature of the Chinese social benefit system: progressive or regressive? China Q 201:1–19
18. Gillion C (2000) Social security pensions: development and reform. International Labour Organisation, Geneva
19. Hastie T, Tibshirani R (1990) Generalized additive models. Chapman and Hall, London
20. Hastie T, Tibshirani R (1993) Varying-coefficient models (with discussion). J R Stat Soc B 55:757–796
21. Hess W, Persson M, Rubenbauer S, Gertheiss J (2013) Using lasso-type penalties to model time-varying covariate effects in panel data regressions-a novel approach illustrated by “Death of Distance” in international trade. Working Paper
22. Hoover DR, Rice JA, Wu CO, Yang LP (1998) Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika 85:809–822
23. Huang J, Shen H (2004) Functional coefficient regression models for non-linear time series: a polynomial spline approach. Scand J Stat 31:515–534
24. Hunter DR, Li R (2005) Variable selection using MM algorithms. Ann Stat 33:1617–1642
25. Li DG, Chen J, Gao JT (2011) Non-parametric time-varying coefficient panel data models with fixed effects. Econ J 14:387–408
26. Lin DY, Ying Z (2001) Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). J Am Stat Assoc 96:103–113
27. Lin H, Peng H (2013) Smoothed rank correlation of the linear transformation regression model. Comput Stat Data Anal 57(1):615–630
28. Lin H, Song XK, Zhou Q (2007) Varying-coefficient marginal models and applications in longitudinal data analysis. Sankhya 69:581–614
29. Lin H, Zhou L, Peng H, Zhou XH (2011) Selection and combination of biomarkers using ROC method for disease classification and prediction. Can J Stat 39(2):324–343
30. Liu J (2011) Resources, incentives and sectoral interests: a longitudinal study of the system of collecting social insurance contributions in China (1999–2008). Soc Sci China 3:9
31. Martinussen T, Scheike TH, Skovgaard IM (2000) Efficient estimation of fixed and time-varying covariates effects in multiplicative intensity models. Scand J Stat 29:57–74
32. Marzec L, Marzec P (1997) On fitting Cox’s regression model with time-dependent coefficients. Biometrika 84:901–908
33. Murphy SA (1993) Testing for a time dependent coefficient in Cox’s regression model. Scand J Stat 20:35–50
34. Murphy SA, Sen PK (1991) Time-dependent coefficients in a Cox-type regression model. Stoch Process Appl 39(1):153–180
35. Nielsen I, Smyth R (2008a) Job satisfaction and response to incentives among China’s urban workforce. J Socio Econ 37:1921–1936
36. Nielsen I, Smyth R (2008b) Who bears the burden of employer compliance with social security contributions? Evidence from Chinese firm level data. China Econ Rev 19:230–244
37. Nyland C, Smyth R, Zhu J (2006) What determines the extent to which employers will comply with their social security obligations? Evidence from Chinese firm-level data. Soc Policy Admin 40:196–214
38. Olsen MK, Schafer J (2001) A two-part random-effects model for semi-continuous longitudinal data. J Am Stat Assoc 96:730–745
39. Orbe S, Ferreira E, Rodriguez-Poo J (2005) Nonparametric estimation of time varying parameters under shape restrictions. J Economet 126:53–57
40. Palacios R, Pallares-Miralles M (2000) International patterns of pension provision. Social Protection Discussion Paper Series No. 0009. The World Bank, Washington, DC
41. Phillips P (2001) Trending time series and macroeconomic activity: some present and future challenges. J Economet 100:21–27

42. Qian J, Wang L (2012) Estimating semiparametric panel data models by marginal integration. J Economet 167:483–493
43. Queisser M, Reilly A, Hu Y (2016) China's pension system and reform: an OECD perspective. Econ Political Stud 4:345–367
44. Ramsay JO, Silverman BW (1997) Functional data analysis. Springer, New York
45. Roberts S, Stafford B, Ashworth, K (2004) Assessing the coverage gap: A synopsis of the ISSA Initiative study. ISSA Initiative—Findings and Opinions No. 12. The International Social Security Association, Geneva, Switzerland
46. Robinson PM (1989) Nonparametric estimation of time-varying parameters. Statistical analysis and forecasting of economic structural change. Springer, Berlin, pp 253–264
47. Robinson PM (1991) Time-varying nonlinear regression. Economic structure change. Springer, Berlin, pp 179–190
48. Robinson PM (2012) Nonparametric trending regression with cross-sectional dependence. J Economet 169(1):4–14
49. Rodriguez-Poo J, Soberon A (2014) Direct semi-parametric estimation of fixed effects panel data varying coefficient models. Econ J 17:107–138
50. Stanovnik T, Bejakovic P, Chlon-Dominczak A (2015) The collection of pension contributions: a comparative review of three Central European countries. Econ Res Ekon Istraz 28:1149–1161
51. Sun YG, Carroll RJ, Li DD (2009) Semiparametric estimation of fixed effects panel data varying coeffcient models. Adv Econom 25:101–129
52. Tian L, Zucker D, Wei LJ (2005) On the Cox model with time-varying regression coefficients. J Am Stat Assoc 100:172–183
53. Tibshirani R (1996) Regression shrinkage and selection via the LASSO. J R Stat Soc B 58:267–288
54. Wang H, Li R, Tsai C (2007) Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika 94:553–568
55. Wu CO, Chiang CT, Hoover DR (1998) Asymptotic confidence regions for kernel smoothing of a varying coefficient model with longitudinal data. J Am Stat Assoc 93:1388–1402
56. Zhang CH (2010) Nearly unbiased variable selection under minimax concave penalty. Ann Stat 38:894–942
57. Zou H (2006) The adaptive Lasso and its oracle properties. J Am Stat Assoc 101:1418–1429
58. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J R Stat Soc B 67(2):301–320
59. Zou H, Li R (2008) One-step sparse estimates in nonconcave penalized likelihood models. Ann Stat 36:1509–1533
60. Zucker DM, Karr AF (1990) Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. Ann Stat 18:329–353