

Statistica Sinica Preprint No: SS-2018-0375

Title	A THRESHOLD VARYING-COEFFICIENT AUTOREGRESSIVE MODEL FOR ANALYZING THE INFLUENCE OF MEDIA REPORTS OF SUICIDE ON THE ACTUAL SUICIDES
Manuscript ID	SS-2018-0375
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0375
Complete List of Authors	Huazhen Lin Jiakun Jiang Binhuan Wang and Paul S. F. YIP
Corresponding Author	Binhuan Wang
E-mail	binhuan.wang@nyumc.org
Notice: Accepted version subject to English editing.	

A THRESHOLD AUTOREGRESSIVE MODEL FOR ANALYZING THE INFLUENCE OF MEDIA REPORTS OF SUICIDE ON THE ACTUAL SUICIDES

Huazhen Lin^{*}, Jiakun Jiang^{†,*}, Binhuan Wang[‡] and Paul S. F. YIP[§]

Southwestern University of Finance and Economics, China^{}*

Tsinghua University, China[†]

New York University School of Medicine, USA[‡]

University of Hong Kong, Hong Kong[§]

Abstract: The extensive coverage of suicides in media has long been thought to be responsible for triggering copycat suicides. However, the up-to-date evidence for a copycat suicide effect is indirect and inconclusive. To examine whether media reported suicides influence actual suicides and to identify a possible copycat effect, a flexible threshold autoregressive model is proposed to explore whether and how reporting of suicides in newspaper affecting the incidence of suicides. In particular, a penalized smoothing least squares estimator is proposed to conveniently estimate the parameters and unknown functions in the model. The performance of proposed method is confirmed by simulation studies, and the asymptotic behaviours of corresponding estimators are studied under mild regularity conditions. The proposed model is applied to investigate the relationship between the daily suicide incidence and the number of

1 INTRODUCTION

reported suicides in a top-selling tabloid newspaper in Hong Kong from January 2002 to December 2006. Our proposed model has identified a copycat suicide effect due to excessive media reporting and the threshold of the number of reports that may trigger a copycat effect.

Key words and phrases: Autoregressive model, Copycat suicide effect, Penalized smoothing least squares estimator, Time series data, Threshold model.

1. Introduction

The widespread suicide coverage in media has long been thought to be responsible for triggering copycat suicides and many scientific papers have discussed its impact (Phillips, 1974; Pirkis & Blood, 2001; Chen, Chen, & Yip, 2011; Niederkrotenthaler et al., 2012; Niederkrotenthaler & Stack, 2017). Particularly, suicides of celebrities have been found to exert a larger social impact (Yip et al., 2006; Fu & Yip, 2007; Chen et al., 2013). However, up-to-date evidence for a copycat suicide effect is indirect, inconclusive and not specific. In addition, the threshold for the number of reports that could trigger a copycat effect has not been investigated.

To examine whether the reported suicides in media is related to actual suicide incidence, coverage of the news reported suicides in a popular Hong Kong based tabloid newspaper, the Apple Daily (AD), are recorded. The

AD was on the top list of circulation in community, with a readership of more than two million (the total population of Hong Kong is about seven million) and a high penetration rate into Hong Kong household. Furthermore, the AD is famous for its reports on celebrities, gossip and scandals. Capturing a wide readership with sensationalism, exaggerated headlines, and attention-grabbing graphic images, the AD quickly became a top-selling newspaper in Hong Kong soon after its first issue in 1995. Such kind of exaggerated media reporting evidently has serious, adverse implications for the media industry (Chen et al, 2013).

The Hong Kong suicide rate was particularly high from 2002 to 2006 especially in 2003. The epidemic SARS during the month of March-May in 2003 had seriously affected Hong Kong, resulting the worst unemployment rate of 8.6 and the historically highest suicide rate. Furthermore, the suicide rate after the death of a celebrity, Mr. Leslie Cheung, on April 1, 2003, immediately surged by more than 20% in the following four to six weeks and stayed at a high level until the end of that year (Yip et al., 2006). The media coverage of Leslie's death was extensive and sensational. Also, the increasing charcoal-burning deaths rose to its historical level from its inception in 1997, about 320 people in 2003, which significantly contributed to the increase of suicide rate during that study period (Law et al., 2014).

1 INTRODUCTION

The spread of charcoal burning death was also shown to be linked with media report and Google search in Taiwan (Chang et al., 2015). The World Health Organization, the International Association of Suicide Prevention and many other organizations and press associations have issued guidelines for reporting suicide incidences (WHO 2014).

Based on the Poisson time series autoregression model, Chen et al. (2012) proposed a method to examine if widespread media reporting of the suicide of a young female singer by charcoal burning increased suicide rates in Taiwan, and they confirmed that a detailed description of a specific suicide method of the celebrity may incur strong copycat effect. Cheng et al. (2007) showed that there is a mutual causation between suicide reporting and suicide incidence; namely, greater number of reported suicide news triggers more actual suicides and an increased number of actual suicides results in more reported suicides. Consequently, the impact of media coverage on actual suicides is not linear, and instead multiplicative and interactive. However, the evidence to date for understanding the copycat suicide effect is still indirect and not clear (Cheng et al., 2012; Chen et al., 2011).

With online search tools, the information on daily number of articles in AD with headlines containing key words (in Chinese) related to suicidal behavior (e.g., “suicide,” “building jumping,” “charcoal burning,” or

“hanging”) during the period from January 2002 to December 2006 is collected. The daily numbers of suicides are obtained from the Coroner’s Court which is responsible for certifying any unnatural cause of death (including suicide). The main purpose of the study is to explore the relation between media reportage of suicides and suicide incidence.

Let Y_t and X_t denote the number of suicides and the number of reports in AD on day t ($t = 0, \dots, n$), respectively. Two characteristics are incorporated into the model. The first one is whether there is a copycat suicide effect. Researchers believe that the effect of Y_{t-j} on Y_t is amplified if suicides are extensively reported in the media, but it is not clear that how many reports can trigger the amplified effect of Y_{t-j} on Y_t . Second, the effect of previous media coverage X_{t-j} and previous suicides Y_{t-j} on Y_t may depend on the time gap j , such that the effects are stronger for recent media coverage on suicides, and may diminish as they become remote. If so, it is important to know when and how previous media coverage and previous suicides cease to have an effect. To address these issues, we propose the following model:

$$\begin{aligned} E \{Y_t | X_s, Y_s, s < t\} = & \mu + \sum_{j=1}^p \alpha_1(j) Y_{t-j} + \sum_{j=1}^q \alpha_2(j) X_{t-j} I(X_{t-j} \geq c_1) \\ & + \sum_{j=1}^w \alpha_3(j) X_{t-j} Y_{t-j} I(X_{t-j} \geq c_2), \end{aligned} \quad (1.1)$$

1 INTRODUCTION

where $\alpha_1(j)$, $\alpha_2(j)$ and $\alpha_3(j)$ quantify the correlation strength among observations over time and reflect when and how the previous media coverage and previous suicides cease to have an effect; c_1, c_2 are unknown threshold parameters that relate to the occurrence of a copycat suicide effect; p, q, w are the maximum time gaps for the second to the fourth term of the right-hand side of (1.1) to be non-zero. The goal of this model lies mainly in determining the threshold parameters c_k and in estimating the size of the effects $\alpha_2(j)$ and $\alpha_3(j)$, if the copycat effect occurs.

First of all, the proposed model (1.1) is a threshold model. The existing threshold models can be broadly grouped into two categories. First, only one threshold variable is included in the model. The threshold variable could be either actual variable (Hanse, 1999, 2000; Chan, 1993; Caner & Hansen, 2001, 2008; Qian, 1998; Koop & Potter, 1999; Delgado & Hidalgo, 2000; Li & Ling, 2012) or a combination of the multiple variables (Seo & Linton, 2007; Chen et al., 2006; Tsay, 1998). Second, multiple threshold variables are included in the model. Chen et al. (2012) proposed a two-threshold variable autoregressive (TTV-AR) model and applied grid search approach to estimate threshold values. Ni et al. (2018), from the Bayesian point of view, proposed a stochastic search variable selection method to study a subset selection of the TTV-AR model and estimate the parameters simultane-

ously. Wu & Chen (2007) proposed a threshold variable driven switching AR model where the threshold variable is a random latent (unobservable) indicator depending on covariates through link functions.

The statistical inference of the threshold model (1.1) cannot be tackled by a traditional regression problem as it involves the unknown threshold parameters c_k 's. A common practice is to estimate the thresholds by a simple grid search method: the threshold estimates are obtained from the point yielding the least squared error across an arbitrarily finite number of candidate points. The computational time for a grid search on G grid points is $O(G^2)$, which is computationally costly with a large G . The threshold parameters also raise a challenge for deriving the asymptotical distributions of resulting estimators because standard asymptotic methods require a smooth criterion function while our model is not the case. In this paper, we propose a smoothing technique to solve the problem: the computation with the proposed method is straightforward and can be accomplished using a standard Newton-Raphson algorithm. Furthermore, this smoothing technique helps us to establish the asymptotic theory and construct a sandwich formula to estimate the variances of estimators.

Another issue is on the estimations of $\alpha_k(j)$, $k = 1, 2, 3$. Since j takes a finite number of values, we can specify each $\alpha_k(j)$ as a separate parameter.

2 FTAR ESTIMATION

We call this a simple parametric method. This simple parametric approach may lose information because $\alpha_k(j)$ generally varies slowly over j ; that is, $\alpha_k(j)$ is smooth in some sense. A common method that incorporates the smoothness is a nonparametric smoothing technique. However, since the arguments of $\alpha_k(\cdot)$, $k = 1, 2, 3$ are discrete and finite, the traditional nonparametric method does not fit. In this paper, a penalized least squares method is proposed to incorporate the smoothness of $\alpha_k(\cdot)$ with discrete and finite argument.

The rest of this manuscript is organized as follows. We first introduce the flexible threshold autoregressive (FTAR) method in Section 2 and then establish asymptotic properties in Section 3. A brief discussion on the bandwidth and tuning parameter selection is summarized in Section 4. Numerical simulations and analyses of the Hong Kong suicide data with the FTAR procedure and other methods are provided in Sections 5 and 6, respectively. A concluding discussion is given in Section 7. All technical proofs are deferred to Appendix.

2. FTAR Estimation

For notational simplicity, we set $p = q = w$ by replacing p, q and w in model (1.1) with the maximum of p, q and w and setting some of $\alpha_k(j)$

to be zero. Let $\mathbf{V}_t = (V_{t1}, \dots, V_{tp})' \equiv (Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})'$ and $\mathbf{X}_t = (X_{t1}, \dots, X_{tp})' \equiv (X_{t-1}, \dots, X_{t-p})'$. Denote $\boldsymbol{\alpha}_k = (\alpha_k(1), \dots, \alpha_k(p))'$ for $k = 1, 2, 3$, and $\mathbf{c} = (c_1, c_2)'$ and $\boldsymbol{\Theta} = (\mu, \boldsymbol{\alpha}'_1, \boldsymbol{\alpha}'_2, \boldsymbol{\alpha}'_3, \mathbf{c}')'$. Here, $\boldsymbol{\Theta}$ represents all parameters defined in model (1.1).

First, we develop estimators for c_1 and c_2 . The objective least squares function is not continuous with respect to c_1 or c_2 . The discontinuity, which stems from the indicator functions $I(X_{tj} > c_k)$, raises a challenge for computation and derivation of the asymptotic distributions for estimators (Sherman, 1993; Han, 1987; and Faraggi & Simon, 1996) In this manuscript, we solve the discontinuity problem by using the kernel smoothing technique (Brown & Wang, 2005; Lin et al., 2011). Let Φ denote the standard normal distribution function. Note that if $X_{tj} > c_k$, $\Phi((X_{tj} - c_k)/h) \rightarrow 1$ as $h \rightarrow 0$, while if $X_{tj} < c_k$, $\Phi((X_{tj} - c_k)/h) \rightarrow 0$, where the bandwidth h goes to zero as the sample size increases; that is, $\Phi((X_{tj} - c_k)/h) \rightarrow I(X_{tj} > c_k)$. The inequality (7.3) in the proof in Appendix B shows that when h is small enough, the error from the approximation is negligible. Rather than a normal approximation, other approximations for $I(X_{tj} > c_k)$, such as a sigmoid approximation (Ma & Huang, 2007), can also be used. With such an approximation, the computation for $\boldsymbol{\Theta}$, especially for \mathbf{c} , is straightforward and can be accomplished through a standard Newton-Raphson iterative algo-

2 FTAR ESTIMATION

rithm. Finally, to incorporate the information that $\alpha_k(j), k = 1, 2, 3$ vary slowly over j , we estimate Θ by minimizing the following penalized least squares function:

$$L_n(\Theta) = l_n(\Theta) + \lambda J(\alpha_1, \alpha_2, \alpha_3), \quad (2.1)$$

with respect to Θ , where

$$l_n(\Theta) = \frac{1}{n} \sum_{t=1}^n \left[Y_t - \mu - \sum_{j=1}^p \alpha_1(j) V_{tj} - \sum_{k=1}^2 \sum_{j=1}^p \alpha_{k+1}(j) X_{tj} V_{tj}^{k-1} \Phi((X_{tj} - c_k)/h) \right]^2, \quad (2.2)$$

λ is a tuning parameter and $J(\alpha_1, \alpha_2, \alpha_3)$ is a penalty function to enforce the smoothness on $\alpha_k(\cdot), k = 1, 2, 3$. The choice of the penalty function $J(\alpha_1, \alpha_2, \alpha_3)$ is crucial. Note that $\alpha_k(j)$ varies slowly over j and the argument of $\alpha_k(\cdot)$ is an ordinal variable, and then we may assume that $\alpha_k(\cdot)$ changes smoothly between any two adjacent levels j and $j + 1$. This leads to a quadratic second order difference penalty

$$J(\alpha_1, \alpha_2, \alpha_3) = \sum_{k=1}^3 w_k \sum_{j=2}^{p-1} \{\alpha_k(j+1) - 2\alpha_k(j) + \alpha_k(j-1)\}^2, \quad (2.3)$$

where $w_k, k = 1, 2, 3$, are weights for each coefficient function. The purpose of introducing the weights w_k is to make the quadratic second order for $\alpha_k(\cdot)$ comparable by taking variations of corresponding variables into account. Consequently, we can avoid using separate tuning parameters

for each $\alpha_k(\cdot)$. In simulation studies and the real data analysis, we choose $w_1 = \text{median}\{\text{SD}(V_{tj}), j = 1, \dots, p\}$, $w_2 = \text{median}\{\text{SD}(X_{tj}), j = 1, \dots, p\}$ and $w_3 = \text{median}\{\text{SD}(X_{tj}V_{tj}), j = 1, \dots, p\}$. The simulation studies suggest a good performance for these choices. This penalty mimics the cubic spline by penalizing the L_2 -norm of the discrete version of the second-order derivatives for the coefficients $\alpha_k(\cdot)$ to encourage smoothness of coefficients (Guo, et al., 2015). Compared to the fused lasso penalty (Tibshirani et al., 2005), the above penalty (2.3) is computationally simple and captures smoothly varying features.

It is straightforward to develop a Newton-Raphson algorithm to solve the minimization problem (2.1). The following notations are necessary to present the gradient and Hessian matrix of $L_n(\Theta)$. Let $\alpha = (\alpha'_1, \alpha'_2, \alpha'_3)'$, $\phi_h(x) = \phi(x/h)/h$, where $\phi(\cdot)$ is the standard normal density function, and $\dot{\phi}_h(x) = \partial\phi_h(x)/\partial x$ is the derivative of $\phi_h(x)$; $\Upsilon_{tj1}(\Theta) = V_{tj}$, $\Upsilon_{tjk}(\Theta) = X_{tj}V_{tj}^{k-2}\Phi((X_{tj} - c_{k-1})/h)$, $k = 2, 3$, $\Upsilon_{tk}(\Theta) = (\Upsilon_{t1k}(\Theta), \dots, \Upsilon_{tpk}(\Theta))'$, $k = 1, 2, 3$, $\Upsilon_t(\Theta) = (\Upsilon_{t1}(\Theta)', \Upsilon_{t2}(\Theta)', \Upsilon_{t3}(\Theta)')'$; $\Upsilon_{tjk}(\Theta) = X_{tj}V_{tj}^{k-2}\phi_h(X_{tj} - c_{k-1})$, $\Upsilon_{tk}(\Theta) = (\Upsilon_{t1k}(\Theta), \dots, \Upsilon_{tpk}(\Theta))'$, $k = 2, 3$; $\dot{\Upsilon}_{tjk}(\Theta) = X_{tj}V_{tj}^{k-2}\dot{\phi}_h(X_{tj} - c_{k-1})$,

2 FTAR ESTIMATION

$\dot{\Upsilon}_{tk}(\Theta) = (\dot{\Upsilon}_{t1k}(\Theta), \dots, \dot{\Upsilon}_{tpk}(\Theta))'$, $k = 2, 3$; $\Omega = \mathbf{D}'\mathbf{D}$ and

$$\mathbf{D} = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \cdots & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{pmatrix}_{(p-2) \times p}.$$

First, we obtain

$$\begin{aligned} \frac{\partial L_n(\Theta)}{\partial \mu} &= -\frac{2}{n} \sum_{t=1}^n [Y_t - \mu - \alpha' \Upsilon_t(\Theta)], \\ \frac{\partial L_n(\Theta)}{\partial \alpha} &= -\frac{2}{n} \sum_{t=1}^n [Y_t - \mu - \alpha' \Upsilon_t(\Theta)] \Upsilon_t(\Theta) + 2(\lambda I_3 \otimes \Omega) \alpha, \\ \frac{\partial L_n(\Theta)}{\partial c_1} &= \frac{2}{n} \sum_{t=1}^n [Y_t - \mu - \alpha' \Upsilon_t(\Theta)] \alpha'_2 \Upsilon_{t2}(\Theta), \\ \frac{\partial L_n(\Theta)}{\partial c_2} &= \frac{2}{n} \sum_{t=1}^n [Y_t - \mu - \alpha' \Upsilon_t(\Theta)] \alpha'_3 \Upsilon_{t3}(\Theta), \end{aligned}$$

where I_3 is a 3-dimensional identity matrix, and \otimes means the Kronecker product. Then, the gradient of $L_n(\Theta)$ follows as

$$\mathbf{g}(\Theta) \triangleq \frac{\partial L_n(\Theta)}{\partial \Theta} = \left(\frac{\partial L_n(\Theta)}{\partial \mu}, \frac{\partial L_n(\Theta)}{\partial \alpha'}, \frac{\partial L_n(\Theta)}{\partial c_1}, \frac{\partial L_n(\Theta)}{\partial c_2} \right)'$$

The elements for the Hessian matrix $\mathbf{H}(\Theta) \triangleq \frac{\partial^2 L_n(\Theta)}{\partial \Theta \partial \Theta'}$ are given by:

$$\begin{aligned} \frac{\partial^2 L_n(\Theta)}{\partial \mu^2} &= 2, \quad \frac{\partial^2 L_n(\Theta)}{\partial \mu \partial \alpha'} = \frac{2}{n} \sum_{t=1}^n \Upsilon_t(\Theta)', \\ \frac{\partial^2 L_n(\Theta)}{\partial \mu \partial c_1} &= -\frac{2}{n} \sum_{t=1}^n \alpha'_2 \Upsilon_{t2,c_1}(\Theta), \quad \frac{\partial^2 L_n(\Theta)}{\partial \mu \partial c_2} = -\frac{2}{n} \sum_{t=1}^n \alpha'_3 \Upsilon_{t3,c_2}(\Theta), \end{aligned}$$

$$\begin{aligned}
 \frac{\partial^2 L_n(\Theta)}{\partial \alpha \partial \alpha'} &= \frac{2}{n} \sum_{t=1}^n \Upsilon_t(\Theta) \Upsilon_t(\Theta)' + 2(\lambda I_3 \otimes \Omega), \\
 \frac{\partial^2 L_n(\Theta)}{\partial \alpha \partial c_1} &= -\frac{2}{n} \sum_{t=1}^n \alpha'_2 \Upsilon_{t2,c_1}(\Theta) + \frac{2}{n} \sum_{t=1}^n [Y_t - \mu - \alpha' \Upsilon_t(\Theta)] \begin{pmatrix} 0 \\ \Upsilon_{t2,c_1}(\Theta) \\ 0 \end{pmatrix}, \\
 \frac{\partial^2 L_n(\Theta)}{\partial \alpha \partial c_2} &= -\frac{2}{n} \sum_{t=1}^n \alpha'_3 \Upsilon_{t3,c_2}(\Theta) + \frac{2}{n} \sum_{t=1}^n [Y_t - \mu - \alpha' \Upsilon_t(\Theta)] \begin{pmatrix} 0 \\ \Upsilon_{t3,c_2}(\Theta) \\ 0 \end{pmatrix}, \\
 \frac{\partial^2 L_n(\Theta)}{\partial c_1^2} &= \frac{2}{n} \sum_{t=1}^n [\alpha'_2 \Upsilon_{t2,c_1}(\Theta)]^2 - \frac{2}{n} \sum_{t=1}^n [Y_t - \mu - \alpha' \Upsilon_t(\Theta)] \alpha'_2 \Upsilon_{t2,c_1}(\Theta), \\
 \frac{\partial^2 L_n(\Theta)}{\partial c_1 \partial c_2} &= \frac{2}{n} \sum_{t=1}^n (\alpha'_2 \Upsilon_{t2,c_1}(\Theta)) (\alpha'_3 \Upsilon_{t3,c_2}(\Theta)), \\
 \frac{\partial^2 L_n(\Theta)}{\partial c_2^2} &= \frac{2}{n} \sum_{t=1}^n [\alpha'_3 \Upsilon_{t3,c_2}(\Theta)]^2 - \frac{2}{n} \sum_{t=1}^n [Y_t - \mu - \alpha' \Upsilon_t(\Theta)] \alpha'_3 \Upsilon_{t3,c_2}(\Theta),
 \end{aligned}$$

where $\Upsilon_{t2,c_1}(\Theta) = \partial \Upsilon_{t2}(\Theta) / \partial c_1$, $\Upsilon_{t3,c_2}(\Theta) = \partial \Upsilon_{t3}(\Theta) / \partial c_2$, $\Upsilon_{t2,c_1 c_1}(\Theta) = \partial^2 \Upsilon_{t2}(\Theta) / \partial c_1^2$ and $\Upsilon_{t3,c_2 c_2}(\Theta) = \partial^2 \Upsilon_{t3}(\Theta) / \partial c_2^2$. Finally, with an initial value $\Theta^{(0)}$ for Θ , we update the estimate of Θ at the $(k+1)$ -th iteration by

$$\Theta^{(k+1)} = \Theta^{(k)} - (\mathbf{H}(\theta^{(k)}))^{-1} \mathbf{g}(\Theta^{(k)})$$

until convergence.

To initialize the algorithm, we choose the initial values of c_1 and c_2 , for example, as $c_1 = c_2 = \text{median}_{t,j} X_{tj}$. Given c_1 and c_2 , we estimate the parameters μ and α by minimizing the squared errors without penalty, which is a standard least squares problem.

3 LARGE SAMPLE PROPERTIES OF THE ESTIMATORS

3. Large sample properties of the estimators

Now, we establish the consistency and asymptotic normality of the FTAR estimator. Without loss of generality, we assume the support of X_{tj} is $[0, 1]$. Some regularity conditions are stated in Appendix A. Denote $\Theta_1 \equiv (\mu, \alpha)'$, and the true values of Θ , Θ_1 and \mathbf{c} by Θ_0 , Θ_{10} and \mathbf{c}_0 , respectively. The consistency of $\hat{\Theta}$ is presented in Theorem 1.

Theorem 1. *Under Conditions A.1 to A.3 in Appendix A, it follows that*

$$\|\hat{\Theta}_1 - \Theta_{10}\| = O_p(n^{-1/2} + \lambda) \text{ and } \|\hat{\mathbf{c}} - \mathbf{c}_0\| = O_p(\sqrt{h/n} + \lambda\sqrt{h}).$$

It is expected that there exists a root- n consistent penalized estimator for the common regression coefficients Θ_1 with $\lambda = o(\frac{1}{\sqrt{n}})$. However, the estimator for \mathbf{c} converges to the true values is at a rate $O(\sqrt{h/n})$ with $\lambda = o(\frac{1}{\sqrt{n}})$, which is faster than root- n . Although a little surprising, this result is not new and has been observed by Seo & Linton (2007). It is due to the fact that the observed information for \mathbf{c} is through $I(X_{tj} > c_1)$ and $I(X_{tj} > c_2)$, which are indicator functions from zero to one. The jump implies an infinite derivative and brings a large amount of information for \mathbf{c} . From simulation experiments, we also observe that the mean squared errors for the estimator of \mathbf{c} are smaller than those for other regression coefficients. See Table 1 for details.

Furthermore, under some mild conditions, the penalized smoothing estimator is asymptotically normal.

Theorem 2. *Under Conditions A.1 to A.3 in Appendix A, it follows that*

$$\begin{aligned}\sqrt{n}(\widehat{\Theta}_1 - \Theta_{10} + \lambda V_{11}^{-1} \mathbf{b}) &\rightarrow N(0, V_{11}^{-1} \Sigma_2 V_{11}^{-1}), \\ \sqrt{\frac{n}{h}}(\widehat{\mathbf{c}} - \mathbf{c}_0 - h\lambda V_{22}^{-1} V'_{12} V_{11}^{-1} \mathbf{b}) &\rightarrow N(0, V_{22}^{-1} \Sigma_1 V_{22}^{-1}),\end{aligned}$$

where $V_{11}, V_{22}, V_{12}, \Sigma_1, \Sigma_2$ and \mathbf{b} are defined in Appendix B.

Therefore, both $\widehat{\Theta}_1$ and $\widehat{\mathbf{c}}$ can be asymptotic unbiased by choosing a small turning parameter $\lambda = o(\frac{1}{\sqrt{n}})$. Proofs of Theorems 1 and 2 are provided in Appendix C.

4. Selection of bandwidth and smoothing parameter

The estimation procedure requires selecting a bandwidth h . The leading terms for the estimators of the regression parameters, $\widehat{\Theta}_1$, are independent of the bandwidth h , indicating that the bandwidth h is not crucial for the asymptotic performance of $\widehat{\Theta}_1$. The asymptotic variance and bias of $\widehat{\mathbf{c}}$ are of order $O(h/n)$ and $\sqrt{nh}\lambda$ respectively, which are both decreases as h decreases. Thus, smaller h may lead to a better estimator. However, numerical studies show that for extremely small h , the proposed estimator may be unstable. Our extensive numeric results suggest that h can be the minimum

4 SELECTION OF BANDWIDTH AND SMOOTHING PARAMETER

difference between any two values of the X_t so that the $\Phi((X_{tj} - c_k)/h)$ can well-approximate the indicator function around the threshold parameters. In practice, we can generate a sequence of h around this minimum value and find an appropriate one which generates stable estimates.

Next, we consider the smoothing parameter λ for $\boldsymbol{\alpha}$. Most existing tuning parameter selection methods are designed for independent data. Cai, Fan & Yao (2000) proposed an analogue to the cross-validation (CV) method for the structure of time series data, and we use their method to select λ . Given an m smaller enough, we first use R subseries of lengths $n - r \times m (r = 1, \dots, R)$ from the beginning to estimate the unknown coefficient functions and parameters, and then compute the one-step forecasting errors for the following section of the time series with length m , based on the estimated model. Finally, we choose the value for λ which minimizes the average mean squared (AMS) error, $AMS(\lambda) = \sum_{r=1}^R AMS_r(\lambda)$, where

$$AMS_r(\lambda) = \frac{1}{m} \sum_{t=n-rm+1}^{n-rm+m} \left\{ Y_t - \hat{\mu}^r - \sum_{j=1}^p \hat{\alpha}_1^r(j) V_{tj} - \sum_{k=1}^2 \sum_{j=1}^p \hat{\alpha}_{k+1}^r(j) X_{tj} V_{tj}^{k-1} I(X_{tj} \geq \tilde{c}_k^r) \right\}^2, \quad (4.1)$$

and $\hat{\mu}^r$, $\hat{\alpha}_k^r(j)$, \tilde{c}_1^r and \tilde{c}_2^r are estimates for μ , $\alpha_k(j)$, c_1 and c_2 , based on the data $\{(Y_t, \mathbf{X}_t, \mathbf{V}_t), t = 1, \dots, n - rm\}$ given λ , $k = 1, 2, 3, j = 1, \dots, p$. Cai, Fan & Yao (2000) suggested using $m = [0.1n]$ and $R = 4$. Our simu-

lation studies and application show that this method can yield reasonable smoothing parameters.

5. Simulation studies

In this section, simulation studies are conducted to assess the finite-sample performance of the FTAR method. We evaluate the performance of the FTAR method by comparing it to the least squares method without penalty (termed LS-UNP), so we can determine how much efficiency the proposed method can obtain with the incorporated smoothness of $\alpha_k(\cdot)$. We are also interested in the effects of the bandwidth h and the smoothing parameter λ on the resulting estimators. Finally, we investigate the performance of FTAR in choosing λ_n with the formula (4.1). The performance of estimators is assessed via the empirical bias and standard deviation of resulting estimators. To be specific, for $\alpha_k = (\alpha_k(1), \dots, \alpha_k(p))'$, we assess empirical Bias = $\left[\frac{1}{p} \sum_{j=1}^p \{E^* \hat{\alpha}_k(j) - \alpha_k(j)\}^2 \right]^{1/2}$, SD = $\left[\frac{1}{p} \sum_{j=1}^p E^* \{\hat{\alpha}_k(j) - E^* \hat{\alpha}_k(j)\}^2 \right]^{1/2}$ and the root of MSE, RMSE = $\sqrt{\text{Bias}^2 + \text{SD}^2}$, where $E^*(\cdot)$ is the empirical expectation over 200 simulated data sets.

5 SIMULATION STUDIES

We simulate observations under the following model:

$$Y_t = \sum_{j=1}^p \alpha_1(j) Y_{t-j} + \sum_{j=1}^p \alpha_2(j) X_{t-j} I(X_{t-j} \geq c_1) + \sum_{j=1}^p \alpha_3(j) X_{t-j} Y_{t-j} I(X_{t-j} \geq c_2) + \varepsilon_t, \quad (5.1)$$

where $X_t \sim \text{Unif}(0, 4)$, $\varepsilon_t \sim \mathcal{N}(0, 1)$, $c_1 = 2$, and $c_2 = 3$. We choose $\alpha_k(j)$, $k = 1, 2, 3$ according to the following three cases:

(1) if $p = 15$,

- $\alpha_1(j) = -0.006(j - (p - 10)/2)^2 / ((p/2 - 5)^2 + 0.03)$,
- $\alpha_2(j) = -4.5(j - (p - 4)/2)^2 / ((p/2 - 2)^2 + 18)$,
- $\alpha_3(j) = -0.004(j - (p - 6)/2)^2 / ((p/2 - 3)^2 + 0.02)$;

(2) if $p = 30$,

- $\alpha_1(j) = -0.006(j - (p - 10)/2)^2 / ((p/2 - 5)^2 + 0.0165)$,
- $\alpha_2(j) = -4.5(j - (p - 4)/2)^2 / ((p/2 - 2)^2 + 14.4)$,
- $\alpha_3(j) = -0.003(j - (p - 6)/2)^2 / ((p/2 - 3)^2 + 0.01)$;

(3) if $p = 60$,

- $\alpha_1(j) = -0.006(j - (p - 10)/2)^2 / ((p/2 - 5)^2 + 0.0075)$,
- $\alpha_2(j) = -4.5(j - (p - 4)/2)^2 / ((p/2 - 2)^2 + 18)$,
- $\alpha_3(j) = -0.003(j - (p - 6)/2)^2 / ((p/2 - 3)^2 + 0.006)$.

The basic idea behind these settings is that for each p , $\alpha_1(j)$, $\alpha_2(j)$ and $\alpha_3(j)$ are selected to generate similar variances for each term in (5.1). In addition, we consider the fourth setting with $c_1 = 2$, $p = 10$ and $\varepsilon_t \sim \mathcal{N}(0, 0.3^2)$, in which $\alpha_1(j)$ and $\alpha_2(j)$ have similar shapes with that of the real data without the interaction term:

$$(4) \quad \begin{aligned} &\bullet \alpha_1(j) = 0.0008(j - 12)^2 - 0.0008, \\ &\bullet \alpha_2(j) = 0.004(j - 5)^2 + 0.025. \end{aligned}$$

For each setting, we use a sample size $n = 200$ with sequences with a length $n + p$ to accommodate the auto-regression structure. The summary statistics of estimated parameters are reported in Table 1 for $p = 15, 30, 60$ and 10, from which we can draw the following conclusions:

- (1) Both the FTAR and the LS-UNP methods are unbiased. The estimator for \mathbf{c} performs almost the same in most of cases because the penalty is not imposed on \mathbf{c} . However, the FTAR for $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_3$ generates much smaller standard deviations and hence has much smaller MSE's than the LS-UNP method, suggesting that the FTAR for $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_3$ is better than the LS-UNP in terms of the MSE.
- (2) Comparing the simulation results for $p = 15, 30, 60$ and 10, we can see that the differences on MSE between the proposed method and the

5 SIMULATION STUDIES

LS-UNP increases as p increases. This can be attributed to the fact that the degrees of freedom for the parameter space are controlled in our method, as a result of smoothing $\hat{\alpha}_k(\cdot), k = 1, 2, 3$. In contrast, the dimension of parameter space for the LS-UNP increases linearly with an increasing p .

- (3) The empirical standard deviations of the proposed estimators for c_1 and c_2 are smaller than those for α_1 and α_2 under all settings in Table 1, while α_3 has a small MSE due to its small scale. These results confirm the asymptotic result in Theorem 2 that $\hat{\mathbf{c}}$ has a faster converging rate than $\hat{\alpha}_k, k = 1, 2, 3$.

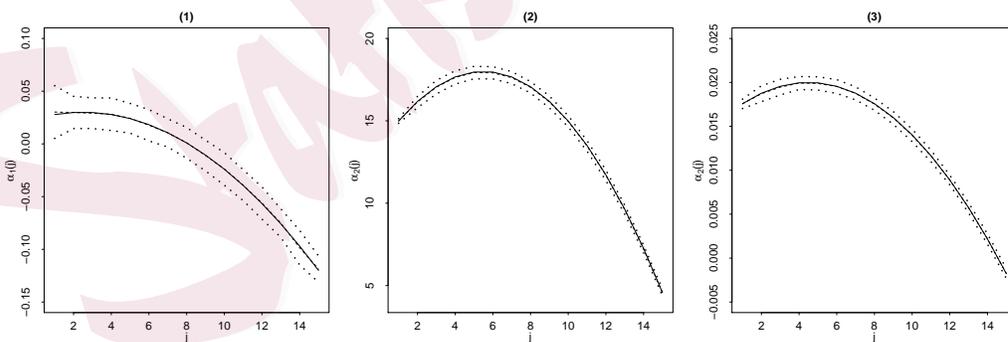
Figures 1, 2, 3 and 4 present the estimates and 95% point-wise confidence bands of $\alpha_1(j), \alpha_2(j)$ and $\alpha_3(j), j = 1, \dots, p$, under three settings ($p = 15, 30, 60$ and 10), using the proposed method with AMS-tuned λ . The results in Figures 1 to 4 suggest that the performance of the proposed method with AMS-tuned parameters is quite satisfactory.

Finally, we investigate the effect of varying h on the resulting estimates. We fix λ at 0.44, 6.66, 30 for $p = 15, 30, 60$, respectively. In order to show all RMSE's in the same figure, the sequence of RMSE over h is scaled to a one-unit variance for each coefficient function. The scaled RMSE's against h for each parameter are shown in Figure 5, which suggest that a smaller

TABLE 1
 Simulation results for $p = 15, 30, 60, 10$ using the proposed method with
 $\lambda = 0.007, 0.60, 4.88, 1.54$, respectively, and the LS-UNP method.

		Proposed			LS-UNP		
		Bias	SD	RMSE	Bias	SD	RMSE
p=15	α_1	0.00087	0.00793	0.00797	0.00077	0.01614	0.01616
	α_2	0.03005	0.16802	0.17068	0.01875	0.21356	0.21438
	α_3	0.00004	0.00036	0.00036	0.00003	0.00039	0.00039
	c_1	0.00001	0.00145	0.00145	0.00003	0.00147	0.00147
	c_2	0.00007	0.00108	0.00108	0.00004	0.00098	0.00098
p=30	α_1	0.00105	0.00524	0.00535	0.00184	0.01681	0.01691
	α_2	0.08210	0.42211	0.43002	0.08953	0.62154	0.62796
	α_3	0.00005	0.00033	0.00034	0.00007	0.00049	0.00050
	c_1	0.00011	0.00178	0.00178	0.00006	0.00165	0.00166
	c_2	0.00004	0.00186	0.00186	0.00022	0.00296	0.00297
p=60	α_1	0.000253	0.002078	0.002094	0.001629	0.021288	0.021350
	α_2	0.021940	0.216963	0.218070	0.033450	0.344298	0.345919
	α_3	0.000008	0.000112	0.000112	0.000014	0.000191	0.000191
	c_1	0.000019	0.000260	0.000261	0.000011	0.000131	0.000131
	c_2	0.000002	0.000056	0.000056	0.000002	0.000031	0.000032
p=10	α_1	0.00645	0.02048	0.02147	0.00671	0.06292	0.06328
	α_2	0.00284	0.00947	0.00989	0.00125	0.01537	0.01542
	c_1	0.00013	0.00224	0.00224	0.00079	0.01127	0.01129

FIG 1. Panels (1), (2) and (3) show the estimates for $\alpha_1(j), \alpha_2(j), \alpha_3(j), j = 1, \dots, 15$, and the associated 95% point-wise confidence bands for $p = 15$ and AMS-tuned $\lambda = 0.007$, respectively.



h being preferred. However, extremely small h causes sensitivity to initial values. Therefore, we fix $h = 0.001$ for $p = 10, 15, 30$ and $h = 0.0001$ for

5 SIMULATION STUDIES

FIG 2. Panels (1), (2) and (3) show the estimates for $\alpha_1(j), \alpha_2(j), \alpha_3(j), j = 1, \dots, 30$, and the associated 95% point-wise confidence bands for $p = 30$ and AMS-tuned $\lambda = 0.60$, respectively.

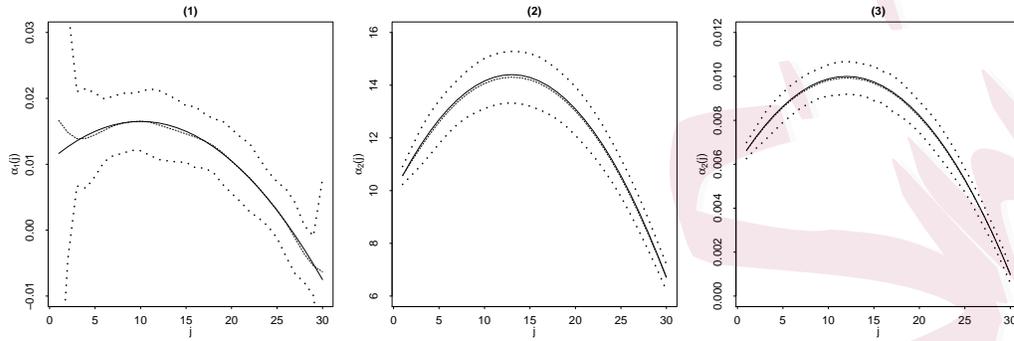
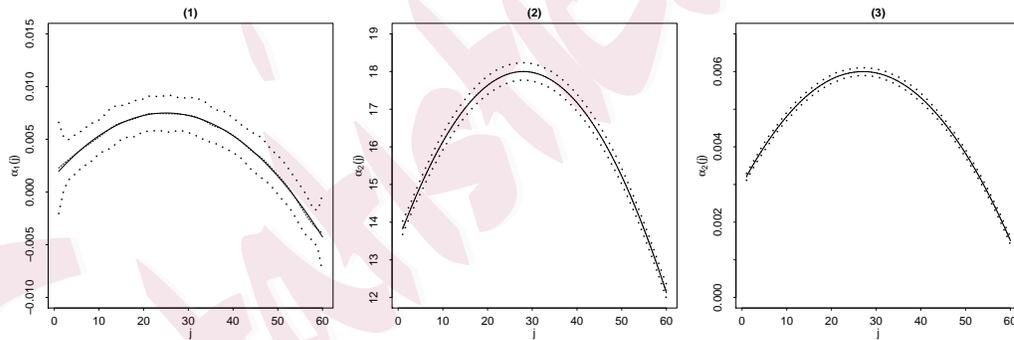


FIG 3. Panels (1), (2) and (3) show the estimates of $\alpha_1(j), \alpha_2(j), \alpha_3(j), j = 1, \dots, 60$, and the associated 95% point-wise confidence bands for $p = 60$ and AMS-tuned $\lambda = 4.88$, respectively.



$p = 60$, respectively.

FIG 4. Panels (1) and (2) show the estimates of $\alpha_1(j)$ and $\alpha_2(j)$, $j = 1, \dots, 10$, and the associated 95% point-wise confidence bands for $p = 10$ and AMS-tuned $\lambda = 1.54$, respectively.

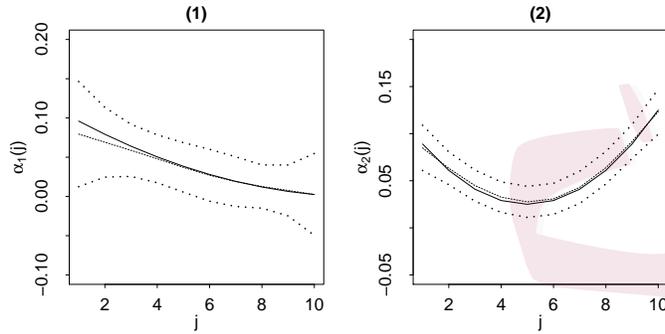
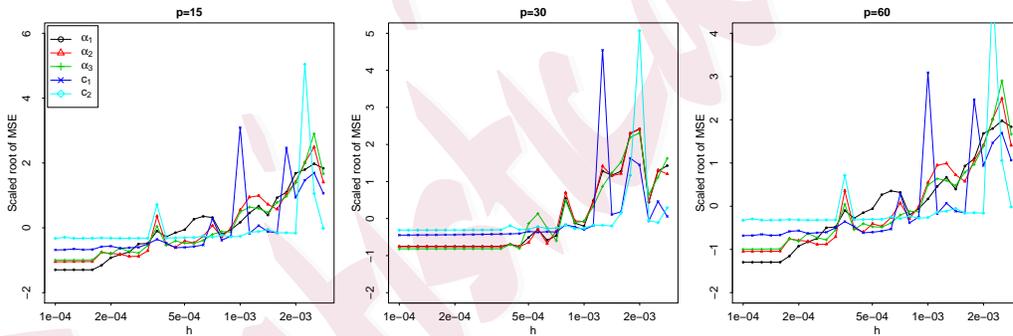


FIG 5. Root-MSE for $p = 15, 30, 60$ for various values of h given $\lambda = 0.007, 0.60, 4.88$, respectively.



6. Analyzing Hong Kong suicide data and media reportage

In order to examine whether the reported suicides in media influences actual suicides, we apply the proposed method to analyze the coverage in term of the number of reported suicides in a Hong Kong based tabloid newspaper (the AD) during the period from January 2002 to December 2006 via the

6 ANALYZING HONG KONG SUICIDE DATA AND MEDIA REPORTAGE

WISENEWS search. In total, there were 1,827 such reports. January 2002 to December 2006 is taken as the study period as Hong Kong's suicide rate increased to its historical maximum, with a rate of 18.6 per 100,000 (i.e., 1,264 suicide deaths in 2003, an average of about three deaths each day).

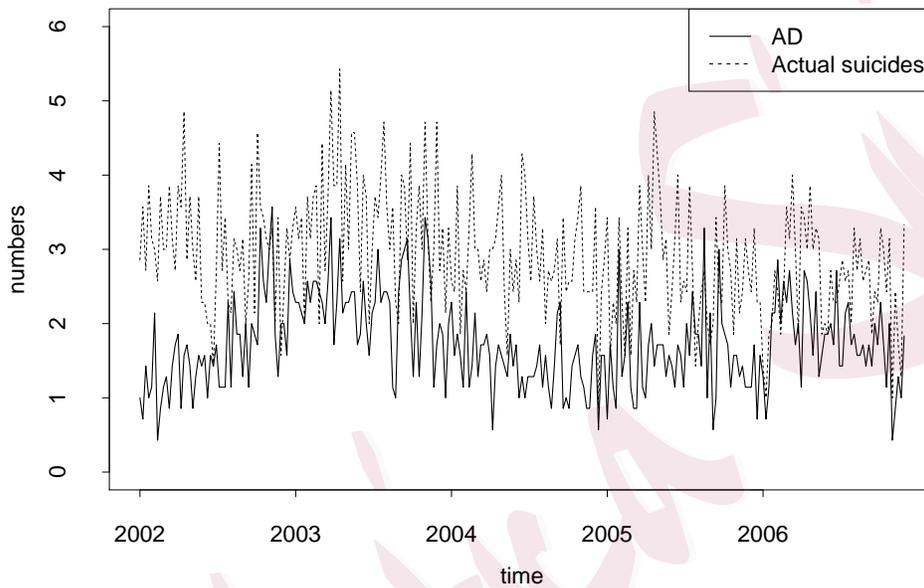
The number of reports is on a daily basis. Due to zero reporting in many days, we aggregate both daily number of suicides reported by AD and actual suicides with a weekly average. The outcome variable Y_t can be regarded as a continuous variable. The histogram and Q-Q plot of the weekly average show that its distribution is close to a normal one, which makes the developed model applicable. Figure 6 displays the aggregated daily numbers of reported and actual suicides in Hong Kong for each week from January 2002 to December 2006. The raw curves in Figure 6 show some lags of spikes between the AD reported and actual suicides.

As described in Section 1, we fit the following model on the data:

$$\begin{aligned} E \{Y_t | Y_s, s < t, X_s, s \leq t\} &= \mu + \sum_{j=1}^p \alpha_1(j) Y_{t-j} + \sum_{j=1}^p \alpha_2(j) X_{t-j} I(X_{t-j} \geq c_1) \\ &\quad + \sum_{j=1}^p \alpha_3(j) X_{t-j} Y_{t-j} I(X_{t-j} \geq c_2). \end{aligned}$$

First, we consider a relatively large order $p = 10$ to investigate the autoregression property. We take $h = 0.001$ which is around the minimum difference between any two values of the X_t , as well as can generate stable es-

FIG 6. Plots of the averaged daily numbers of suicide reported by AD and actual suicides in Hong Kong for each week from January 2002 to December 2006, respectively.



timates. The cross-validation method defined in Section 4 yields $\lambda = 18.71$.

The estimates of parameters and their standard deviations are shown in Table 2 and Figure 7. The calculation of standard deviation is done via the resampling method described in Fan & Yao (2003) with 500 bootstrapping samples. The results in Figure 7 show that $\hat{\alpha}_1(\cdot)$ is significantly different from zero and $\hat{\alpha}_3(\cdot)$ is not significantly different from zero.

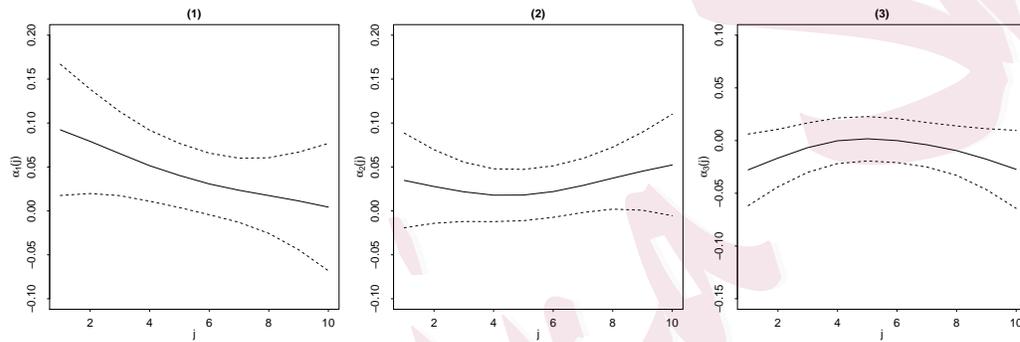
Therefore, we refit the model by removing the interaction term. Estimates for c_1 is also shown in Table 2, and estimates for $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ are

6 ANALYZING HONG KONG SUICIDE DATA AND MEDIA REPORTAGE

TABLE 2
Estimate, standard deviation and 95% CI for thresholds.

		Estimate	SD	95% CI
With interaction	c_1	2	0.07	(1.87, 2.13)
	c_2	3	0.05	(2.90, 3.10)
Without interaction	c_1	2	0.05	(1.90, 2.10)

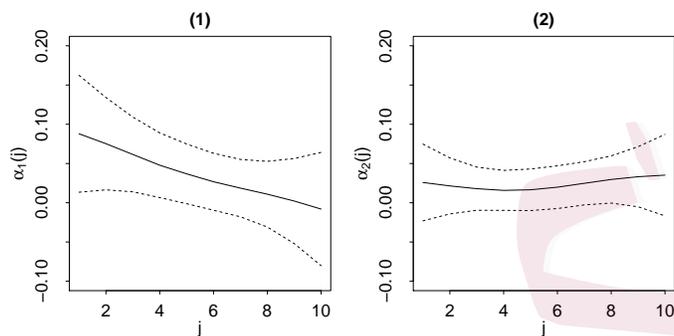
FIG 7. Panels (1), (2) and (3) show the estimates of $\alpha_1(j)$, $\alpha_2(j)$, $\alpha_3(j)$ and the associated 95% point-wise confidence bands, respectively.



plotted in Figure 8, which are based on $\lambda = 104.22$. The estimate for $\alpha_1(\cdot)$ implies the effect of the previous suicides declines as they become remote as expected. The $\hat{\alpha}_2(\cdot)$ and its 95% confidence bands suggest that the copycat effect of media coverage is at the borderline of significance, which can last for a long time up to 8 weeks. $\hat{c}_1 = 2$ implies that a copycat suicide effect occurs when the number of reports is more than two.

For comparison purposes, we also fit the data with a simple time series

FIG 8. Panels (1) and (2) show the estimates of $\alpha_1(j), \alpha_2(j)$ based on the model without the interaction term, and the associated 95% point-wise confidence bands, respectively.



regression with the form of

$$E \{Y_t | Y_s, s < t, X_s, s \leq t\} = \mu + \sum_{j=1}^p \alpha_1(j) Y_{t-j} + \sum_{j=1}^p \alpha_2(j) X_{t-j},$$

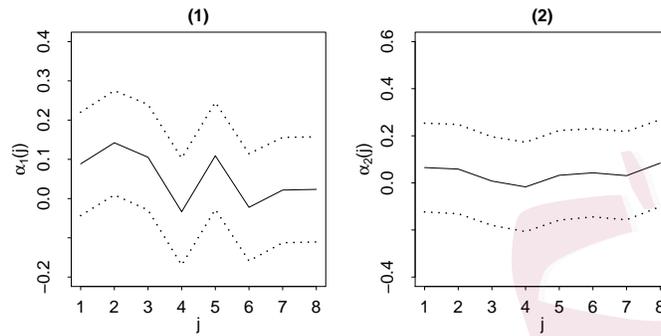
with $p = 8$. The estimates for $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ are plotted in Figure 9. The results show that $\hat{\alpha}_1(\cdot)$ is marginally different from zero and $\hat{\alpha}_2(\cdot)$ is not significantly different from zero. Comparing our results in Figure 8 with those in Figure 9, we can see our method implies a clearer trend and a narrower confidence band, and hence is more efficient.

7. Discussion

We have proposed a flexible threshold autoregressive (FTAR) model to explore whether and how reportage of suicide relating to the incidence of suicides. A penalized smoothing least squares estimator is adopted to estimate parameters and unknown functions. The proposed FTAR method

7 DISCUSSION

FIG 9. A simple time series regression model at lag-8.



yields an accurate estimate for the effect of reportage, which is confirmed by simulation studies. Theoretical properties, including uniform consistency and asymptotical normality, are proved under mild regularity conditions. Our model identifies a copycat suicide effect, which occurs when the number of reported cases is greater than two.

We also confirm an association between media reportage and the incidence of suicides. Although the effect diminishes in the beginning, remote media reportage can still trigger copycat effect. More importantly, in our model we set up threshold parameters to identify the occurrence of a copycat suicide effect. It is of further interest to investigate the pattern of copycat suicide effect, which will be reported in another paper.

References

Brown, B. M. & Wang, Y. G. (2005). Standard errors and covariance

-
- matrices for smoothed rank estimators. *Biometrika*, **92**, 149–158.
- Cai, Z., Fan, J. & Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series models. *Journal of the American Statistical Association*, **95**, 941–956.
- Caner, M. & Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica*, **69**, 1555-1596.
- Caner, M., & Hansen, B. E. (2008). Instrumental variable estimation of a threshold model. *Econometric Theory*, **20**, 813-843.
- Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics*, **21**, 520-533.
- Chang, S. S., Kwok, S. S. M., Cheng, Q., Yip, P. S. F., & Chen, Y. Y. (2015). The association of trends in charcoal-burning suicide with Google search and newspaper reporting in Taiwan: a time series analysis. *Social Psychiatry and Psychiatric Epidemiology*, **50**, 1451-1461.
- Chen, R. (1995). Threshold variable selection in open-loop threshold autoregressive models. *JOURNAL OF Time Series Analysis*. 16, 461-481.
- Chen, Y.-Y., Chen, F., Gunnell, D., & Yip, P. S. F., (2013). The impact of media reporting on the emergence of charcoal burning suicide in Taiwan. *PloS one*, 8(1).

7 DISCUSSION

Chen, Y. Y., Chen, F., & Yip, P. S. F. (2011). The impact of media reporting of suicide on actual suicides in taiwan, 2002-05. *Journal of Epidemiology and Community Health*, **65**, 934-940.

Chen, H., Chong, T. L., & Bai, J. (2012). Theory and applications of tar model with two threshold variables. *Econometric Reviews*, **31**, 142-170.

Chen, Y. Y., Liao, S. F., Teng, P. R., Tsai, C. W., Fan, H. F., Lee, W. C., Cheng, A. T. (2012). The impact of media reporting of the suicide of a singer on suicide rates in Taiwan. *Soc Psychiatry Psychiatr Epidemiol*, **47**, 215-221.

Cheng, A. T., Hawton, K., Chen, T. H., Yen, A. M., Chen, C. Y., Chen, L. C., et al. (2007). The influence of media coverage of a celebrity suicide on subsequent suicide attempts. *The Journal of Clinical Psychiatry*, **68**, 862-866.

Cheng, Q., Chang, S. S., & Yip, P. S. F. (2012). The opportunity and challenges of online data collection for suicide prevention. *Lancet*, **379**, 53-54.

Chen, C. W. S. & So, M. K. P. (2006). On a threshold heteroscedastic model. *International Journal of Forecasting*, **22**, 73-89.

Delgado, M. A., & Hidalgo, J. (2000). Nonparametric inference on struc-

-
- tural breaks. *Journal of Econometrics*, **96**, 113-144.
- Fan, J., & Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York, NY: Springer-Verlag.
- Faraggi, D., & Simon, R. (1996). A simulation study of cross-validation for selecting an optimal cutpoint in univariate survival analysis. *Statistics in medicine*, **15**, 2203–2213.
- Fu, K. W., & Yip, P. S. F. (2007). Long-term impact of celebrity suicide on suicidal ideation: Results from a population-based study. *Journal of Epidemiology and Community Health*, **61**, 540-546.
- Koop, G., & S. M. Potter. (1999). Dynamic Asymmetries in U.S. Unemployment, *Journal of Business and Economic Statistics*, **17**, 298-312.
- Guo, J., Hu, J., Jing, B., & Zhang, Z. (2015): Spline-Lasso in High-Dimensional Linear Regression. *Journal of the American Statistical Association*, **111**, 288–297.
- Han, A. K. (1987). Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics*, **35**, 303–316.
- Hanse, B. E. (1999). Threshold effects in non-dynamic panels: estimation, testing, and inference. *Journal of Econometrics*, **93**, 345-368.
- Hanse, B. E. (2000). Sample splitting and threshold estimation. *Econo-*

7 DISCUSSION

metrica, **68**, 575-603.

Law, Y. W., Yip, P. S., Zhang, Y., & Caine, E. D. (2014). The chronic impact of work on suicides and under-utilization of psychiatric and psychosocial services. *Journal of Affective Disorders*, **168**, 254-261.

Li, D., & Ling, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models. *Journal of Econometrics*, **167**, 240-253.

Lin, H. Z., Yip, P. S. F., & Huggins, R. M. (2011). A nonparametric estimation of the infection curve. *Science China Mathematics*, **54**, 1815.

Ni, S., Xia, Q., & Liu, J. (2018). Bayesian Subset Selection for Two-Threshold Variable Autoregressive Models. *Studies in Nonlinear Dynamics & Econometrics*.

Niederkrotenthaler, T., Fu, K. W., Yip, P. S., Fong, D. Y., Stack, S., Cheng, Q., & Pirkis, J. (2012). Changes in suicide rates following media reports on celebrity suicide: a meta-analysis. *Journal of Epidemiology and Community Health*, **66**, 1037-1042.

Niederkrotenthaler, T., & Stack, S. (2017). *Media and Suicide: International Perspectives on Research, Theory, and Policy*, New York, USA: Routledge.

-
- Phillips, D. (1974). The Influence of Suggestion on Suicide: Substantive and Theoretical Implications of the Werther Effect. *American Sociological Review*, **39**, 340-354.
- Pirkis, J., & Blood, R. W. (2001). Suicide and the media: Part 1. Reportage in nonfictional media. *Crisis-The Journal of Crisis Intervention and Suicide Prevention*, **22**, 146-154.
- Qian, L. (1998). On maximum likelihood estimators for a threshold autoregression. *Journal of Statistical Planning and Inference*, **75**, 21-46.
- Seo, M. H., & Linton, O. (2007). A smoothed least squares estimator for threshold regression models. *Journal of Econometrics*, **141**, 704-735.
- Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica*, **61**, 123-137.
- Tsay, R. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association*. **93**, 1188-1202.
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused Lasso. *Journal of the Royal Statistical Society: Series B*, **67**, 91-108.
- World Health Organization. (2014). *World Suicide Report*, Geneva: World Health Organization.
- Wu, S., & Chen, R. (2007). Threshold variable determination and thresh-

7 DISCUSSION

old variable driven switching autoregressive models. *Statistica Sinica*, **17**, 241-264.

Yip, P. S., Fu, K. W., Yang, K. C., Ip, B. Y., Chan, C. L., Chen, E. Y., Lee, D. T., Law, Y. W. & Hawton, K. (2006). The effects of a celebrity suicide on suicide rates in Hong Kong. *Journal of Affective Disorders*, **93**, 245-252.

Acknowledgement

Lin's research is supported by the Chinese Natural Science Foundation (No. 11571282, 11829101) and the Fundamental Research Funds for the Central Universities (14TD0046 and JBK141121). Yip's research is supported by a GRF grant (HKU-106140091) and a Strategic Public Policy Research grant (SPPR-HKU-12).

Appendix A

Condition A.1:

1. $f(\mathbf{x}_1, \mathbf{x}_r | \mathbf{v}_1, \mathbf{v}_r; r) \leq M \leq \infty$, for all $r \geq 1$, where $f(\mathbf{x}_1, \mathbf{x}_r | \mathbf{v}_1, \mathbf{v}_r; r)$ is the conditional density of $(\mathbf{X}_1, \mathbf{X}_r)$ given $(\mathbf{V}_1, \mathbf{V}_r)$, and $f(\mathbf{v} | \mathbf{x}) \leq M < \infty$, where $f(\mathbf{v} | \mathbf{x})$ is the conditional density of \mathbf{V}_t given $\mathbf{X}_t = \mathbf{x}$.

2. The process $\{\mathbf{X}_t, \mathbf{V}_t, Y_t\}$ is α -mixing with $\sum_k k^c [\alpha(k)]^{1-2/\delta} < \infty$ for some $\delta > 2$ and $c > 1 - 2/\delta$, where $\alpha(k) = \sup\{|Pr(A \cap B) - Pr(A)Pr(B)|; A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty\}$, \mathcal{F}_a^b is the σ -algebra generated by $\{(\mathbf{X}_t, \mathbf{V}_t, Y_t); a \leq t \leq b\}$.

3. $E|\mathbf{V}_t|^{2\delta} < \infty$, where δ is given in condition A.1.2.

4. \mathbf{X}_t is bounded with compact support $[0, 1]^p$.

Condition A.2:

1. Assume that $E\{Y_1^2 + Y_l^2 | \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_l = \mathbf{x}_2, \mathbf{V}_1 = \mathbf{v}_1, \mathbf{V}_l = \mathbf{v}_2\} \leq M < \infty$ for all $l > 1$.

2. Assume that $h \rightarrow 0$ and $nh \rightarrow \infty$. Further, assume that there exists a sequence of positive integers s_n such that $s_n \rightarrow \infty$, $s_n = o(\sqrt{nh})$, and $(n/h)^{1/2} \alpha(s_n) \rightarrow 0$, as $n \rightarrow \infty$.

3. There exists $\delta^* > \delta$, where δ is given in condition A.1.3, such that

$$E\{|Y_t|^{\delta^*} | \mathbf{V}_t = \mathbf{v}, \mathbf{X}_t = \mathbf{x}\} \leq M < \infty$$

for any \mathbf{v} and \mathbf{x} in supports of \mathbf{V}_t and \mathbf{X}_t , respectively, and

$$\alpha(n) = O(n^{-\theta^*}),$$

where $\theta^* \geq \delta\delta^*/\{2(\delta^* - \delta)\}$.

4. $E|\mathbf{V}_t|^{2\delta^*} < \infty$, and $n^{1/2-\delta/4} h^{\delta/\delta^*-1/2-\delta/4} = O(1)$.

Condition A.3:

7 DISCUSSION

1. Let f_j be the density function of X_{t_j} . The density function $f_j(\cdot)$ is positive and has continuous second derivatives on $[0, 1]$.

2. $\lambda \rightarrow 0$, $h^2 \log(n) \rightarrow 0$ and $nh \rightarrow \infty$ as $n \rightarrow \infty$.

The above conditions are used for deriving the convergence properties.

Conditions A.1 and A.2 are similar to those in Cai, Fan, & Yao (2000).

Appendix B: Notations and Lemma

Let f_j be the density function of X_{tj} , where $f_{j,s}(x_{tj}, x_{ts})$ is the joint density of X_{tj} and X_{ts} , Θ_0 is the true value of Θ , and $v_k = \int x^k \phi^2(x) dx$.

$$\sigma^2(\mathbf{V}_t, \mathbf{X}_t) = \text{var}(Y_t | \mathbf{V}_t, \mathbf{X}_t),$$

$$b_k(\Theta) = \sum_{j=1}^p v_0 \alpha_{k+1}^2(j) c_k^2 f_j(c_k) E(\sigma^2(\mathbf{V}_t, \mathbf{X}_t) V_{tj}^{2(k-1)} | X_{tj} = c_k),$$

$$\Upsilon_{tj1,0}(\Theta) = \Upsilon_{tj1}(\Theta), \Upsilon_{tjk,0}(\Theta) = X_{tj} V_{tj}^{k-1} I((X_{tj} > c_{k-1})/h), k = 2, 3,$$

$$C_{tk}(\Theta) = \sum_{\ell=1}^p \alpha_{k+1}(\ell) X_{t\ell} V_{t\ell}^{k-1} \phi((X_{t\ell} - c_k)/h),$$

$$F_k(\Theta) = 2 \sum_{j=1}^p \alpha_{k+1}^2(j) c_k^2 v_0 f_j(c_k) E[V_{tj}^2 | X_{tj} = c_k],$$

$$F(\Theta) = \sum_{j \neq l}^p \alpha_2(j) \alpha_3(l) c_1 c_2 f_{j,l}(c_1, c_2) E(V_{tl} | X_{tj} = c_1, X_{tl} = c_2),$$

$$\begin{aligned} \kappa_{kjl} = 2 \sum_{m=1}^p \alpha_{l+1,0}(m) c_{l,0} f_m(c_{l,0}) E \left[V_{tm}^{l-1} [V_{tj} I(k=1) + X_{tj} V_{tj}^{k-1} I(X_{tj} > c_{k-1,0}) I(k \neq 1, j \neq m) \right. \\ \left. + c_{l,0} V_{tj} I(c_{l,0} > c_{k-1,0}) I(k \neq 1) I(j=m)] \middle| X_{tm} = c_{l,0} \right], \end{aligned}$$

$$\varphi_l(\Theta) = 2 \sum_{j=1}^p \alpha_{l+1}(j) c_l f_j(c_l) E(V_{tj}^{l-1} | X_{tj} = c_l),$$

$$\delta_{kjm} = 4E\sigma^2(\mathbf{V}_t, \mathbf{X}_t) \Upsilon_{tjk,0}(\Theta_0) \Upsilon_{tvm,0}(\Theta_0),$$

$$\zeta_1(\Theta) = \sum_{j \neq s}^p \alpha_2(j) \alpha_3(s) c_1 c_2 f_{j,s}(c_1, c_2) E(\sigma^2(\mathbf{V}_t, \mathbf{X}_t) V_{tj} | X_{tj} = c_1, X_{ts} = c_2), \quad \zeta_2 = 4E\sigma^2(\mathbf{V}_t, \mathbf{X}_t),$$

$$\varpi_{kj} = 4E\sigma^2(\mathbf{V}_t, \mathbf{X}_t) \Upsilon_{tjk,0}(\Theta_0), \quad \vartheta_{kj} = 2E\Upsilon_{tjk,0}(\Theta_0), \quad \rho_{kjm} = 2E\Upsilon_{tjk,0}(\Theta_0) \Upsilon_{tvm,0}(\Theta_0),$$

$$\Sigma_{11} = \zeta_2, \quad \Sigma_{22} = (\delta_{1i1j})_{i \leq p, j \leq p}, \quad \Sigma_{33} = (\delta_{2i2j})_{i \leq p, j \leq p}, \quad \Sigma_{44} = (\delta_{3i3j})_{i \leq p, j \leq p}, \quad \Sigma_{12} = (\varpi_{1j})_{j \leq p},$$

$$\Sigma_{13} = (\varpi_{2j})_{j \leq p}, \quad \Sigma_{14} = (\varpi_{3j})_{j \leq p}, \quad \Sigma_{23} = (\delta_{1i2j})_{i \leq p, j \leq p}, \quad \Sigma_{24} = (\delta_{1i3j})_{i \leq p, j \leq p}, \quad \Sigma_{34} = (\delta_{2i3j})_{i \leq p, j \leq p},$$

$$A_{11} = 2, \quad A_{22} = (\rho_{1j1v})_{1 \leq j, v \leq p}, \quad A_{33} = (\rho_{2j2v})_{1 \leq j, v \leq p}, \quad A_{44} = (\rho_{3j3v})_{1 \leq j, v \leq p}, \quad A_{12} = (\vartheta_{1j})_{j \leq p},$$

$$A_{13} = (\vartheta_{2j})_{j \leq p}, \quad A_{14} = (\vartheta_{3j})_{j \leq p}, \quad A_{23} = (\rho_{1j2v})_{1 \leq j, v \leq p}, \quad A_{24} = (\rho_{1j3v})_{1 \leq j, v \leq p}, \quad A_{34} = (\rho_{2j3v})_{1 \leq j, v \leq p},$$

7 DISCUSSION

$$B_1 = (\varphi_l)_{l=1,2}, \quad B_2 = (\kappa_{1jl})_{1 \leq j \leq p, l=1,2}, \quad B_3 = (\kappa_{2jl})_{1 \leq j \leq p, l=1,2}, \quad B_4 = (\kappa_{3jl})_{1 \leq j \leq p, l=1,2},$$

$$\Sigma_1 = \text{diag}(b_1(\Theta_0), b_2(\Theta_0)), \quad V_{12} = (B'_1, B'_2, B'_3, B'_4)', \quad V_{22} = \text{diag}(F_1, F_2), \quad \mathbf{b} = (0, \boldsymbol{\alpha}'_0 A')',$$

$$\mathbf{A} = 2 \begin{pmatrix} \Omega & 0 & 0 \\ 0 & \Omega & 0 \\ 0 & 0 & \Omega \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\ \Sigma'_{12} & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} \\ \Sigma'_{13} & \Sigma'_{23} & \Sigma_{33} & \Sigma_{34} \\ \Sigma'_{14} & \Sigma'_{24} & \Sigma'_{34} & \Sigma_{44} \end{pmatrix}, \quad V_{11} = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A'_{12} & A_{22} & A_{23} & A_{24} \\ A'_{13} & A'_{23} & A_{33} & A_{34} \\ A'_{14} & A'_{24} & A'_{34} & A_{44} \end{pmatrix}.$$

Denote $U_k(\boldsymbol{\Theta}) = \frac{1}{n} \sum_{t=1}^n R_{t,k}$, $k = 1, 2$, where

$$R_{t,k} = \left[Y_t - \mu - \sum_{j=1}^p \alpha_1(j) V_{tj} - \sum_{k=1}^2 \sum_{j=1}^p \alpha_{k+1}(j) X_{tj} V_{tj}^{k-1} \Phi((X_{tj} - c_k)/h) \right] \\ \times \left\{ \sum_{j=1}^p \alpha_{k+1}(j) X_{tj} V_{tj}^{k-1} \phi_h(X_{tj} - c_k) \right\}.$$

Lemma A.1 Under Conditions A.1 and A.2, if $h \rightarrow 0$ and $nh \rightarrow \infty$

as $n \rightarrow \infty$, we have

- (a) $h \text{var}\{R_{t,k}(\boldsymbol{\Theta}_0)\} = b_k(\boldsymbol{\Theta}_0) + o(1)$;
- (b) $h \sum_{j=1}^{n-1} |\text{cov}(R_{1,k}(\boldsymbol{\Theta}_0), R_{1+j,k}(\boldsymbol{\Theta}_0))| = o(1)$;
- (c) $nh \text{var}\{U_k(\boldsymbol{\Theta}_0)\} = b_k(\boldsymbol{\Theta}_0) + o(1)$.

Proof. Denote

$$R_{t,k0} = \left[Y_t - \mu - \sum_{j=1}^p \alpha_1(j) V_{tj} - \sum_{k=1}^2 \sum_{j=1}^p \alpha_{k+1}(j) X_{tj} V_{tj}^{k-1} I(X_{tj} > c_k) \right] \\ \times \left\{ \sum_{j=1}^p \alpha_{k+1}(j) X_{tj} V_{tj}^{k-1} \phi_h(X_{tj} - c_k) \right\}, \quad k = 1, 2.$$

Suppose that Z is a standard normal variable. Then we have tail probability,

$$1 - \Phi(t) = P(Z \geq t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-\frac{z^2}{2}} dz \leq \frac{1}{\sqrt{2\pi}} \int_t^\infty \frac{z}{t} e^{-\frac{z^2}{2}} dz \leq \frac{1}{\sqrt{2\pi t}} e^{-t^2/2}, \quad (7.1)$$

for any $t > 0$. We first consider the case of $X_{tj} > c_k$. To simplify the notation, let $x = (X_{tj} - c_k)/h$ which is positive, since

$$|\Phi(x) - I(x > 0)| = |\Phi(x) - 1|I(X_{tj} \geq c_k + \sqrt{h}) + |\Phi(x) - 1|I(c_k < X_{tj} < c_k + \sqrt{h}),$$

then, $\forall \epsilon > 0$ and $\forall s \geq 1$,

$$\begin{aligned} & P\left(\frac{|\Phi(x) - I(x > 0)|}{h^s} > \epsilon | X_{tj} > c_k\right) \\ &= P\left(\frac{|\Phi(x) - 1|I(X_{tj} \geq c_k + \sqrt{h}) + |\Phi(x) - 1|I(c_k < X_{tj} < c_k + \sqrt{h})}{h^s} > \epsilon | X_{tj} > c_k\right) \\ &\leq P\left(\frac{|\Phi(x) - 1|I(X_{tj} \geq c_k + \sqrt{h})}{h^s} > \epsilon/2 | X_{tj} > c_k\right) \\ &\quad + P\left(\frac{|\Phi(x) - 1|I(c_k < X_{tj} < c_k + \sqrt{h})}{h^s} > \epsilon/2 | X_{tj} > c_k\right) \\ &\leq P\left(\frac{|\Phi(x) - 1|I(x \geq 1/\sqrt{h})}{h^s} > \epsilon/2 | X_{tj} > c_k\right) + P(c_k < X_{tj} < c_k + \sqrt{h} | X_{tj} > c_k) \\ &\equiv I_1 + I_2. \end{aligned}$$

By (7.1), we have

$$I_1 \leq P\left(\frac{\frac{1}{\sqrt{2\pi x}} e^{-x^2/2} I(x \geq 1/\sqrt{h})}{h^s} > \epsilon/2 | X_{tj} > c_k\right). \quad (7.2)$$

7 DISCUSSION

Noting that $f(y) = \frac{1}{y}e^{-y^2/2}$ is monotonically decreasing function for $y > 0$, we have $\frac{\frac{1}{\sqrt{2\pi}x}e^{-x^2/2}}{h^s} \leq \frac{1}{\sqrt{2\pi}h^{2s-1}}e^{-1/2h} \rightarrow 0$ when $x \geq 1/\sqrt{h}$. Then

$$I_1 \leq P\left(\frac{1}{\sqrt{2\pi}h^{2s-1}}e^{-1/2h}I(x \geq 1/\sqrt{h}) > \epsilon/2 | X_{tj} > c_k\right) \rightarrow 0.$$

Moreover, under condition A.3.1, we have

$$I_2 = P(c_k < X_{tj} < c_k + \sqrt{h} | X_{tj} > c_k) = O(\sqrt{h}).$$

Thus,

$$\lim_{n \rightarrow \infty} P\left(\frac{|\Phi(x) - I(x > 0)|}{h^s} > \epsilon | X_{tj} > c_k\right) = 0$$

which implies $|\Phi(\frac{X_{tj}-c_k}{h}) - I(X_{tj} > c_k)| = o_p(h^s)$ for any $s \geq 1$. Similarly, the same conclusion holds for $X_{tj} < c_k$. In summary, we have

$$|\Phi(\frac{X_{tj}-c_k}{h}) - I(X_{tj} > c_k)| = o_p(h^s), \text{ for any } s \geq 1. \quad (7.3)$$

Hence,

$$\begin{aligned} R_{t,k}(\Theta) &= R_{t,k0}(\Theta) + R_{t,k}(\Theta) - R_{t,k0}(\Theta) \\ &= R_{t,k0}(\Theta) + o_p(h^s). \end{aligned} \quad (7.4)$$

The rest of the proof is similar to that of Lemma A.1 in Cai, Fan & Yao (2000) and only give the proof of (a).

By conditioning on $(\mathbf{V}_t, \mathbf{X}_t)$, we have

$$\begin{aligned}
 & \text{Var}(R_{t,k0}(\Theta)) \\
 = & \sum_{j=1}^p E\sigma^2(\mathbf{V}_t, \mathbf{X}_t)\alpha_{k+1}^2(j)X_{tj}^2V_{tj}^{2(k-1)}\phi_h^2(X_{tj} - c_k) \\
 & + E\sigma^2(\mathbf{V}_t, \mathbf{X}_t)\sum_{j \neq j'}^p \alpha_{k+1}(j)\alpha_{k+1}(j')X_{tj}X_{tj'}V_{tj}^{k-1}V_{tj'}^{k-1}\phi_h(X_{tj} - c_k)\phi_h(X_{tj'} - c_k) \\
 = & I_1 + I_2.
 \end{aligned}$$

Firstly,

$$\begin{aligned}
 I_1 &= \sum_{j=1}^p \alpha_{k+1}^2(j)E \int \sigma^2(\mathbf{V}_t, \mathbf{X}_t)X_{tj}^2V_{tj}^{2(k-1)}\phi_h^2(X_{tj} - c_k)f_j(X_{tj})dX_{tj} \\
 &= \frac{1}{h^2}\sum_{j=1}^p \alpha_{k+1}^2(j)E \int \sigma^2(\mathbf{V}_t, \mathbf{X}_t)X_{tj}^2V_{tj}^{2(k-1)}\phi^2\left(\frac{X_{tj} - c_k}{h}\right)f_j(X_{tj})dX_{tj} \\
 &= \frac{1}{h}\sum_{j=1}^p \alpha_{k+1}^2(j)E \int \sigma^2(\mathbf{V}_t, \mathbf{X}_t)(zh + c_k)^2V_{tj}^{2(k-1)}\phi^2(z)f_j(zh + c_k)dz \\
 &= \frac{1}{h}\sum_{j=1}^p \alpha_{k+1}^2(j)E \int \sigma^2(\mathbf{V}_t, \mathbf{X}_t)\{c_k^2\}V_{tj}^{2(k-1)}\phi^2(z)\{f_j(c_k)\}dz \\
 &\quad + \frac{1}{h}\sum_{j=1}^p \alpha_{k+1}^2(j)E \int \sigma^2(\mathbf{V}_t, \mathbf{X}_t)\{c_k^2\}V_{tj}^{2(k-1)}\phi^2(z)\{f'_j(c_k)zh\}dz \\
 &\quad + \frac{1}{h}\sum_{j=1}^p \alpha_{k+1}^2(j)E \int \sigma^2(\mathbf{V}_t, \mathbf{X}_t)\{2zhc_k\}V_{tj}^{2(k-1)}\phi^2(z)\{f_j(c_k)\}dz \\
 &\quad + \frac{1}{h}\sum_{j=1}^p \alpha_{k+1}^2(j)E \int \sigma^2(\mathbf{V}_t, \mathbf{X}_t)\{2zhc_k\}V_{tj}^{2(k-1)}\phi^2(z)\{f'_j(c_k)zh\}dz + \dots \\
 &= \frac{1}{h}b_k(\Theta) + O(1) + O(1) + O(h) + o(h).
 \end{aligned}$$

7 DISCUSSION

Similarly, one can prove $I_2 = o(h)$. Thus, we can get

$$h\text{Var}(R_{t,k0}(\Theta_0)) = b_k(\Theta_0) + O(h) = b_k(\Theta_0) + o(1).$$

Appendix C: Proofs of Theorems

Proof of Theorem 1.

Let $\alpha_{1n} = n^{-1/2} + \lambda$, $\alpha_{2n} = \sqrt{h/n} + \sqrt{h}\lambda$. Denote $\Theta_0 = (\mu_0, \alpha'_{10}, \alpha'_{20}, \alpha'_{30}, \mathbf{c}'_0)'$ to be the true value of Θ . We wish to show that for any given $\varepsilon > 0$, there exists a large constant τ_1, τ_2 such that

$$\Pr \left[\min_{\|\mathbf{u}_1\|=\tau_1, \|\mathbf{u}_2\|=\tau_2} L_n\{\Theta_0 + (\alpha_{1n}\mathbf{u}_1', \alpha_{2n}\mathbf{u}_2')'\} > L_n(\Theta_0) \right] \geq 1 - \varepsilon,$$

where \mathbf{u}_1 has the same dimension as Θ_1 and \mathbf{u}_2 has the same dimension as \mathbf{c} . This implies with a probability of at least $1 - \varepsilon$, that there exists a local minimum in the ball $\{\Theta_0 + (\alpha_{1n}\mathbf{u}_1', \alpha_{2n}\mathbf{u}_2')' : \|\mathbf{u}_1\| \leq \tau_1, \|\mathbf{u}_2\| \leq \tau_2\}$. Hence, there exists a local minimum $(\Theta'_1, \mathbf{c}')'$ such that $\|\Theta_1 - \Theta_{10}\| = O_p(\alpha_{1n})$, $\|\mathbf{c} - \mathbf{c}_0\| = O_p(\alpha_{2n})$.

With the definition $\Theta^* = \Theta_0 + (\alpha_{1n}\mathbf{u}_1', \alpha_{2n}\mathbf{u}_2')' = (\mu^*, \alpha_{1n}^*, \alpha_{2n}^*, \alpha_{3n}^*, \mathbf{c}^{*'})'$, $\mathbf{u} = (\mathbf{u}'_1, \mathbf{u}'_2)'$, we have

$$D_n(\mathbf{u}) = L_n(\Theta^*) - L_n(\Theta_0) = l_n(\Theta^*) - l_n(\Theta_0) + \lambda \{J(\alpha_{1n}^*, \alpha_{2n}^*, \alpha_{3n}^*) - J(\alpha_{10}, \alpha_{20}, \alpha_{30})\}.$$

Since $\partial J(\alpha)/\partial \alpha = \mathbf{A}\alpha$, by the standard argument on Taylor expansion of

the likelihood function, we have

$$\begin{aligned}
 D_n(\mathbf{u}) &= \alpha_{1n} \mathbf{u}_1' \frac{\partial l_n(\boldsymbol{\Theta}_0)}{\partial \boldsymbol{\Theta}_1} + \alpha_{2n} \mathbf{u}_2' \frac{\partial l_n(\boldsymbol{\Theta}_0)}{\partial \mathbf{c}} + \frac{1}{2} \alpha_{1n}^2 \mathbf{u}_1' \frac{\partial^2 l_n(\boldsymbol{\Theta}_0)}{\partial \boldsymbol{\Theta}_1 \partial \boldsymbol{\Theta}_1'} \mathbf{u}_1 (1 + o_p(1)) \\
 &\quad + \frac{1}{2} \alpha_{2n}^2 \mathbf{u}_2' \frac{\partial^2 l_n(\boldsymbol{\Theta}_0)}{\partial \mathbf{c} \partial \mathbf{c}'} \mathbf{u}_2 (1 + o_p(1)) + \alpha_{1n} \alpha_{2n} \mathbf{u}_1' \frac{\partial^2 l_n(\boldsymbol{\Theta}_0)}{\partial \boldsymbol{\Theta}_1 \partial \mathbf{c}'} \mathbf{u}_2 (1 + o_p(1)) \\
 &\quad + \lambda \{(\boldsymbol{\alpha}^* - \boldsymbol{\alpha}_0)' \mathbf{A} \boldsymbol{\alpha}_0 + (\boldsymbol{\alpha}^* - \boldsymbol{\alpha}_0)' \mathbf{A} (\boldsymbol{\alpha}^* - \boldsymbol{\alpha}_0)\} \\
 &\equiv I_1 + I_2 + I_3 + I_4 + I_5 + I_6,
 \end{aligned}$$

where $\boldsymbol{\alpha}^* = (\boldsymbol{\alpha}_1^*, \boldsymbol{\alpha}_2^*, \boldsymbol{\alpha}_3^*)'$. Denote $\Delta_{tjk}(\boldsymbol{\Theta}) \equiv \Phi\{(X_{tj} - c_k)/h\} - I(X_{tj} > c_k)$. By (7.3), we have

$$\begin{aligned}
 \left| E \frac{\partial l_n(\boldsymbol{\Theta}_0)}{\partial \mu} \right| &= \left| \frac{2}{n} E \sum_{t=1}^n \sum_{r=1}^2 \sum_{\ell=1}^p \alpha_{r+1,0}(\ell) X_{t\ell} V_{t\ell}^{r-1} \Delta_{t\ell r}(\boldsymbol{\Theta}_0) \right| = o(h^s), \\
 \left| E \frac{\partial l_n(\boldsymbol{\Theta}_0)}{\partial \alpha_k(j)} \right| &= \left| \frac{2}{n} E \sum_{t=1}^n \sum_{r=1}^2 \sum_{\ell=1}^p \alpha_{r+1,0}(\ell) X_{t\ell} V_{t\ell}^{r-1} \Delta_{t\ell r}(\boldsymbol{\Theta}_0) \Upsilon_{tjk}(\boldsymbol{\Theta}_0) \right| = o(h^s), \\
 \left| E \frac{\partial l_n(\boldsymbol{\Theta}_0)}{\partial c_k} \right| &= \left| \frac{2}{nh} E \sum_{t=1}^n C_{tk}(\boldsymbol{\Theta}_0) \sum_{r=1}^2 \sum_{\ell=1}^p \alpha_{r+1,0}(\ell) X_{t\ell} V_{t\ell}^{r-1} \Delta_{t\ell r}(\boldsymbol{\Theta}_0) \right| = o(h^s),
 \end{aligned} \tag{7.5}$$

for $k = 1, 2, 3, j = 1, \dots, p$, where $\alpha_{r,0}(\ell)$ is the true value of $\alpha_r(\ell)$.

Denote $A_{kj} \equiv E \sigma^2(\mathbf{V}_t, \mathbf{X}_t) \Upsilon_{tjk,0}^2(\boldsymbol{\Theta}_0)$. Furthermore, similar to Lemma

7 DISCUSSION

1, we have

$$\begin{aligned}\Gamma_1 &\hat{=} \left[\text{var}\left\{\frac{\partial l_n(\Theta_0)}{\partial \alpha_1(1)}\right\}, \dots, \text{var}\left\{\frac{\partial l_n(\Theta_0)}{\partial \alpha_1(p)}\right\}, \text{var}\left\{\frac{\partial l_n(\Theta_0)}{\partial \alpha_2(1)}\right\}, \dots, \text{var}\left\{\frac{\partial l_n(\Theta_0)}{\partial \alpha_3(p)}\right\} \right]' \\ &= \frac{4}{n} \cdot (A_{11}, \dots, A_{1p}, \dots, A_{3p})' + o(h^s), \\ \Gamma_2 &\hat{=} \text{var}\left\{\frac{\partial l_n(\Theta_0)}{\partial \mu}\right\} = \frac{4}{n} \cdot E\sigma^2(\mathbf{V}_t, \mathbf{X}_t) + o(h^s), \\ \Gamma_3 &\hat{=} \left[\text{var}\left\{\frac{\partial l_n(\Theta_0)}{\partial c_1}\right\}, \text{var}\left\{\frac{\partial l_n(\Theta_0)}{\partial c_2}\right\} \right]' = \frac{1}{nh} \{b_1(\Theta_0), b_2(\Theta_0)\}' + O(1/n). \quad (7.6)\end{aligned}$$

Combining (7.5) and (7.6), we have

$$\begin{aligned}\frac{\partial l_n(\Theta_0)}{\partial \boldsymbol{\alpha}} &= E\left\{\frac{\partial l_n(\Theta_0)}{\partial \boldsymbol{\alpha}}\right\} + O_p(\Gamma_1^{1/2}) = O_p\left(\frac{1}{\sqrt{n}}\right), \\ \frac{\partial l_n(\Theta_0)}{\partial \mu} &= E\left\{\frac{\partial l_n(\Theta_0)}{\partial \mu}\right\} + O_p(\Gamma_2^{1/2}) = O_p\left(\frac{1}{\sqrt{n}}\right), \\ \frac{\partial l_n(\Theta_0)}{\partial \mathbf{c}} &= E\left\{\frac{\partial l_n(\Theta_0)}{\partial \mathbf{c}}\right\} + O_p(\Gamma_3^{1/2}) = O_p\left(\frac{1}{\sqrt{nh}}\right).\end{aligned}$$

then

$$I_1 = O_p(\alpha_{1n}\tau_1/\sqrt{n}), I_2 = O_p(\alpha_{2n}\tau_2/\sqrt{nh}). \quad (7.7)$$

Similar to Lemma 1, we also obtain

$$\begin{aligned}E\frac{\partial^2 l_n(\Theta_0)}{\partial \mu^2} &= \text{Var}\left(\frac{\partial^2 l_n(\Theta_0)}{\partial \mu^2}\right) = 2, \\ E\frac{\partial^2 l_n(\Theta_0)}{\partial \mu \partial \alpha_k(j)} &= 2E\Upsilon_{tjk,0}(\Theta_0) + o(h^s), \\ \text{Var}\frac{\partial^2 l_n(\Theta_0)}{\partial \mu \partial \alpha_k(j)} &= \frac{4}{n} \left[E\Upsilon_{tjk,0}^2(\Theta_0) - E^2\Upsilon_{tjk,0}(\Theta_0) \right] + o(1/n), \\ E\left(\frac{\partial^2 l_n(\Theta_0)}{\partial \alpha_k(j) \partial \alpha_m(v)}\right) &= 2E\Upsilon_{tjk,0}(\Theta_0)\Upsilon_{tvm,0}(\Theta_0) + o(h^s), \\ \text{Var}\left(\frac{\partial^2 l_n(\Theta_0)}{\partial \alpha_k(j) \partial \alpha_m(v)}\right) &= \frac{4}{n} \left[E\Upsilon_{tjk,0}^2(\Theta_0)\Upsilon_{tvm,0}^2(\Theta_0) - \{E\Upsilon_{tjk,0}(\Theta_0)\Upsilon_{tvm,0}(\Theta_0)\}^2 \right] + o(1/n).\end{aligned}$$

Hence,

$$I_3 = O_p(\alpha_{1n}^2 \tau_1^2). \quad (7.8)$$

Similarly, we have

$$\begin{aligned} E \frac{\partial^2 l_n(\Theta_0)}{\partial c_k^2} &= \frac{F_k(\Theta_0)}{h} + O(1), \quad E \left\{ \frac{\partial^2 l_n(\Theta_0)}{\partial c_k^2} \right\}^2 = \frac{F_k^2(\Theta_0)}{h^2} + O(1 + (nh^2)^{-1}), \\ E \frac{\partial^2 l_n(\Theta_0)}{\partial c_1 \partial c_2} &= F(\Theta_0) + O\left(\frac{1}{h} \exp \left\{ -\frac{(c_{1,0} - c_{2,0})^2}{4h^2} \right\}\right), \\ E \left\{ \frac{\partial^2 l_n(\Theta_0)}{\partial c_1 \partial c_2} \right\}^2 &= F^2(\Theta_0) + O\left(\exp \left\{ -\frac{(c_{1,0} - c_{2,0})^2}{2h^2} \right\}\right), \end{aligned}$$

then we obtain,

$$I_4 = O_p(\alpha_{2n}^2 \tau_2^2 / h). \quad (7.9)$$

Finally, by

$$\begin{aligned} E \frac{\partial^2 l_n(\Theta_0)}{\partial \alpha_k(j) \partial c_l} &= \kappa_{kjl} + O(h), \quad E \left\{ \frac{\partial^2 l_n(\Theta_0)}{\partial \alpha_k(j) \partial c_l} \right\}^2 = \kappa_{kjl}^2 + O(h^2), \\ E \frac{\partial^2 l_n(\Theta_0)}{\partial \mu \partial c_l} &= \varphi_l + O(h), \quad E \left\{ \frac{\partial^2 l_n(\Theta_0)}{\partial \mu \partial c_l} \right\}^2 = \varphi_l^2 + O(h^2), \end{aligned}$$

we get

$$I_5 = O_p(\alpha_{1n} \alpha_{2n} \tau_1 \tau_2). \quad (7.10)$$

By (7.7), (7.8), (7.9), (7.10) and coupling with $I_6 = O_p(\lambda \alpha_{1n} \tau_1 + \lambda \alpha_{1n}^2 \tau_1^2)$,

choosing large τ_1, τ_2 , then I_1, I_2, I_5 are dominated by I_3, I_4 . This completes

the proof of Theorem 1. \square

7 DISCUSSION

Proof of Theorem 2.

According to Theorem 1, when $\lambda = o(1/\sqrt{n})$, it can easily be shown that there exists a \sqrt{n} -consistent estimator $\widehat{\Theta}_1 = (\widehat{\mu}, \widehat{\alpha}')'$ and $\sqrt{n/h}$ -consistent estimator $\widehat{\mathbf{c}}$, satisfying the following equations

$$\begin{aligned} \frac{\partial l_n(\widehat{\Theta})}{\partial \Theta_1} + \lambda \widehat{\mathbf{b}} &= 0, \\ \frac{\partial l_n(\widehat{\Theta})}{\partial \mathbf{c}} &= 0, \end{aligned}$$

where $\widehat{\mathbf{b}} = (0, \widehat{\alpha}'A)'$ is a vector of $3p + 1$ dimension. By Taylor expansion,

$$\begin{aligned} -\sqrt{n} \frac{\partial l_n(\Theta_0)}{\partial \Theta_1} &= \frac{\partial^2 l_n(\Theta_0)}{\partial \Theta_1 \partial \Theta_1'} \sqrt{n} (\widehat{\Theta}_1 - \Theta_{10}) (1 + o_p(1)) \\ &\quad + \sqrt{h} \frac{\partial^2 l_n(\Theta_0)}{\partial \Theta_1 \partial \mathbf{c}'} \sqrt{\frac{n}{h}} (\widehat{\mathbf{c}} - \mathbf{c}_0) (1 + o_p(1)) + \sqrt{n} \lambda \widehat{\mathbf{b}}, \\ -\sqrt{nh} \frac{\partial l_n(\Theta_0)}{\partial \mathbf{c}} &= \sqrt{h} \frac{\partial^2 l_n(\Theta_0)}{\partial \mathbf{c} \partial \Theta_1'} \sqrt{n} (\widehat{\Theta}_1 - \Theta_{10}) (1 + o_p(1)) \\ &\quad + h \frac{\partial^2 l_n(\Theta_0)}{\partial \mathbf{c} \partial \mathbf{c}'} \sqrt{\frac{n}{h}} (\widehat{\mathbf{c}} - \mathbf{c}_0) (1 + o_p(1)). \end{aligned} \quad (7.11)$$

Since

$$\begin{aligned} E \left(\frac{\partial l_n(\Theta_0)}{\partial c_k} \right)^2 &= \frac{1}{nh} b_k(\Theta_0) + O(1/n), \\ E \frac{\partial l_n(\Theta_0)}{\partial c_1} \frac{\partial l_n(\Theta_0)}{\partial c_2} &= \frac{1}{n} \zeta_1 + O(h) \\ E \frac{\partial l_n(\Theta_0)}{\partial \alpha_k(j)} \frac{\partial l_n(\Theta_0)}{\partial \alpha_m(v)} &= \frac{1}{n} \delta_{kjm v} + o(h^s/n), \\ E \left(\frac{\partial l_n(\Theta_0)}{\partial \mu} \right)^2 &= \frac{1}{n} \zeta_2 + o(h^s/n), \\ E \frac{\partial l_n(\Theta_0)}{\partial \mu} \frac{\partial l_n(\Theta_0)}{\partial \alpha_k(j)} &= \frac{1}{n} \varpi_{kj} + o(h^s/n), \end{aligned}$$

we get $nh(\frac{\partial l_n(\Theta_0)}{\partial \mathbf{c}})^{\otimes 2} \rightarrow_p \Sigma_1$ and $n(\frac{\partial l_n(\Theta_0)}{\partial \Theta_1})^{\otimes 2} \rightarrow_p \Sigma_2$. Then by CLT we have

$$\sqrt{nh} \frac{\partial l_n(\Theta_0)}{\partial \mathbf{c}} \rightarrow N(0, \Sigma_1), \sqrt{n} \frac{\partial l_n(\Theta_0)}{\partial \Theta_1} \rightarrow N(0, \Sigma_2). \quad (7.12)$$

Furthermore, from the proof of Theorem 1, we know that $\frac{\partial^2 l_n(\Theta_0)}{\partial \Theta_1 \partial \Theta_1'} \rightarrow_p V_{11}$, $\frac{\partial^2 l_n(\Theta_0)}{\partial \Theta_1 \partial \mathbf{c}'} \rightarrow_p V_{12}$, $h \frac{\partial^2 l_n(\Theta_0)}{\partial \mathbf{c} \partial \mathbf{c}'} \rightarrow_p V_{22}$. Based on all these results coupled with (7.11), (7.12) and Slutsky's theorem, we obtain

$$\begin{aligned} \sqrt{n}(\widehat{\Theta}_1 - \Theta_{10} + \lambda V_{11}^{-1} \mathbf{b}) &\rightarrow N(0, V_{11}^{-1} \Sigma_2 V_{11}^{-1}), \\ \sqrt{\frac{n}{h}}(\widehat{\mathbf{c}} - \mathbf{c}_0 - h \lambda V_{22}^{-1} V_{12}' V_{11}^{-1} \mathbf{b}) &\rightarrow N(0, V_{22}^{-1} \Sigma_1 V_{22}^{-1}), \end{aligned}$$

where \mathbf{b} is defined in Appendix B. We complete the proof of Theorem 2. \square