

Jump detection in time series nonparametric regression models: a polynomial spline approach

Yujiao Yang · Qionxia Song

Received: 7 June 2012 / Revised: 19 April 2013 / Published online: 14 September 2013
© The Institute of Statistical Mathematics, Tokyo 2013

Abstract For time series nonparametric regression models with discontinuities, we propose to use polynomial splines to estimate locations and sizes of jumps in the mean function. Under reasonable conditions, test statistics for the existence of jumps are given and their limiting distributions are derived under the null hypothesis that the mean function is smooth. Simulations are provided to check the powers of the tests. A climate data application and an application to the US unemployment rates of men and women are used to illustrate the performance of the proposed method in practice.

Keywords B splines · Discontinuities · Jump detection · α -Mixing process · Time series · Nonparametric regression

1 Introduction

Regression analysis, as a major statistical tool, builds a functional relationship between response variables and explanatory variables. In certain applications, such a functional relationship has discontinuous points at some unknown positions, representing structural changes of a related process. In practice, structural changes may be caused by sudden events, abrupt policy changes and catastrophes among others. For example, the real estate price series would have a jump if new government policies are implemented (Hui et al. 2010). The temperature of the ocean off of Granite Canyon would

Y. Yang
Department of Statistics and Actuarial Science, East China Normal University,
Shanghai 200241, China

Q. Song (✉)
Department of Mathematical Sciences, University of Texas at Dallas,
Richardson, TX 75080, USA
e-mail: song@utdallas.edu

have a significant drop between February and April, called a spring transition, due to the change of ocean currents (Koo 1997). Jumps are an important part of the underlying regression function; accurate detection of them is crucial for understanding the structural changes of the process and for estimating the regression function properly. In addition, detecting jumps inspires one to investigate the underlying reasons for structural breaks.

Assume that observations $(X_t, Y_t)_{t=1}^n$ are from a strictly stationary bivariate stochastic process and consider the following nonparametric regression model,

$$Y_t = m(X_t) + \sigma(X_t)\varepsilon_t, \quad t = 1, \dots, n, \quad (1)$$

where $\{\varepsilon_t\}_{t=1}^n$ is a conditional white noise process with $E(\varepsilon_t|X_t = x) = 0$, $\text{Var}(\varepsilon_t|X_t = x) = 1$, $t = 1, \dots, n$. The conditional mean function $m(x) = E(Y_t|X_t = x)$ and the conditional variance function $\sigma^2(x) = \text{Var}(Y_t|X_t = x)$ are defined on a compact interval $[a, b]$. Assume $m(x)$ is continuous on $[a, b]$ except at a finite number of points, called jumps. Neither the number nor the locations of the jumps are known.

Jump regression analysis started in the early 1990s and has become an important research topic in statistics. There has been an increasing amount of literature on the detection of jumps in nonparametric regression models, tackled with different techniques. For instance, assuming the break points were known, Shiau (1987) applied partial splines to estimate the magnitude of discontinuities. Qiu (1991) proposed a trimmed spline estimate to deal with the second jump regression function. Müller (1992) and Qiu (1994) detected jumps based on kernel-type methods. For a fixed design model, Koo (1997) detected discontinuities by linear splines without theoretical justifications. Qiu and Yandell (1998) suggested a jump detection algorithm using local polynomial smoothing. More recently, under the i.i.d. assumption, Ma and Yang (2011) proposed a spline smoothing method to detect jumps. For more studies, refer to Qiu et al. (1991), Wu and Chu (1993), Müller and Stadtmüller (1999), Qiu (2003, 2005), Bowman et al. (2006), Gijbels et al. (2007), Joo and Qiu (2009) and references therein. A majority of references listed above focus on independent errors. However, this restriction would be problematic in time series analysis, and Wu and Zhao (2007) considered this problem in the dependent case. Lin et al. (2008) studied the nonparametric regression model with dependent observations by local polynomial smoothing. Another widely used approach is wavelet analysis. For details, see Wong et al. (2001), Chen et al. (2008), Hui et al. (2010), Zhou et al. (2010) and references therein.

In this paper, we propose a quick and direct method to detect possible jumps in nonparametric time series models through spline regression. The number, locations, as well as magnitudes of the jumps are all assumed unknown. Compared with the existing kernel/local polynomial smoothing-based method, the spline smoothing method has the advantages of simple implementation and fast computation; see Huang and Yang (2004) and Xue and Yang (2006). Our test statistics are based on the maximal difference of the spline estimators between neighboring knots. By applying the strong approximation results similarly as in Wang and Yang (2010), we obtain the limiting distributions of the test statistics in a conservative sense under the null hypothesis that $m(x)$ is continuous.

By investigating both theoretical properties and numerical performance of such detection procedure, certain practical guidelines are provided about their use. Basically, we use linear splines and we undersmooth the data in the first step to detect possible jumps with their locations. Then we apply a multiple-ordered regression spline procedure to refit the data and estimate the jump magnitudes.

The rest of the paper is organized as follows. Section 2 introduces the constant and linear spline smoothers, together with the test statistics. Section 3 describes the actual steps to implement the spline estimation and the test procedure. Numerical results are provided in Sects. 4 and 5. Section 6 is a brief discussion of our method. All technical proofs are included in the Appendix.

2 Estimation and main results

2.1 Spline estimation

In the first step, we use constant and linear splines to estimate $m(x)$ in model (1). To introduce spline functions, we divide the interval $[a, b]$ into $(N + 1)$ subintervals $J_j = [t_j, t_{j+1})$, $j = 0, \dots, N - 1$, $J_N = [t_N, b]$. $\{t_j\}_{j=1}^N$ is assumed to be a sequence of equally spaced interior knots, given as

$$t_0 = a < t_1 < \dots < t_N < b = t_{N+1}, \quad t_j = a + jh, \quad j = 0, 1, \dots, N + 1, \quad p = 1, 2,$$

where $h = (b - a)/(N + 1)$ is the distance between neighboring knots. We denote $C^{(p)}[a, b]$ as the space of functions that have p th-order continuous derivatives on the interval $[a, b]$ and $G_N^{(p-2)} = G_N^{(p-2)}[a, b]$ as the space of all $C^{(p-2)}[a, b]$ functions that are polynomials of degree $(p - 1)$ on each subinterval J_j . The B-spline basis of $G_N^{(p-2)}$ can be constructed recursively; see de Boor (2001). We denote the j th B-spline of order p as $b_{j,p}$. Specifically, the first-order (constant) spline and the second-order (linear) spline basis functions are given as

$$\begin{aligned} b_{j,1}(x) &= I_{J_j}(x), \quad j = 0, \dots, N, \\ b_{j,2}(x) &= K\{(x - t_{j+1})h^{-1}\}, \quad j = -1, \dots, N, \end{aligned}$$

where $I_{J_j}(\cdot)$ is the indicator function on J_j and $K(u) = (1 - |u|)_+$ is a triangular function. For any L^2 -integrable functions ϕ, φ on $[a, b]$, we define the empirical inner product $\langle \phi, \varphi \rangle_n = n^{-1} \sum_{t=1}^n \{\phi(X_t)\varphi(X_t)\}$ and the theoretical inner product $\langle \phi, \varphi \rangle = E\{\phi(X)\varphi(X)\}$. The corresponding empirical and theoretical L^2 norms are defined as $\|\phi\|_{2,n}^2 = n^{-1} \sum_{t=1}^n \phi^2(X)$ and $\|\phi\|_2^2 = E\{\phi^2(X)\}$, respectively. For the technical simplicity, we rescale the B-spline basis and denote $B_{j,p}(x) = \|b_{j,p}(x)\|_2^{-1} b_{j,p}(x)$. Based on the B-spline basis $\{B_{j,p}(x)\}_{j=-1}^N$, we denote $\mathbf{V}_{n,p}$ and \mathbf{V}_p as its empirical and theoretical inner product matrices, respectively, i.e.

$$\mathbf{V}_{n,p} = \langle \langle B_{j',p}, B_{j,p} \rangle_n \rangle_{j,j'=1-p}^N, \quad \mathbf{V}_p = \langle \langle B_{j',p}, B_{j,p} \rangle \rangle_{j,j'=1-p}^N.$$

$\mathbf{V}_{n,1}$ is a diagonal matrix and $\mathbf{V}_{n,2}$ is a tridiagonal matrix, which go to their deterministic versions \mathbf{V}_1 and \mathbf{V}_2 as n goes to infinity. Furthermore, denote \mathbf{S} as the inverse matrix of \mathbf{V}_2 and \mathbf{S}_j the 2×2 diagonal submatrices of \mathbf{S} , i.e.

$$\mathbf{S} = (s_{j',j})_{j,j'=-1}^N = \mathbf{V}_2^{-1}, \quad \mathbf{S}_j = \begin{pmatrix} s_{j-1,j-1} & s_{j-1,j} \\ s_{j,j-1} & s_{j,j} \end{pmatrix}, \quad j = 0, \dots, N. \quad (2)$$

The polynomial spline estimators are

$$\hat{m}_p(x) = \arg \min_{g \in G^{(p-2)}[a,b]} \sum_{t=1}^n \{Y_t - g(X_t)\}^2.$$

Solving the least squares problem under regular conditions, with the spline basis defined above, we write the spline estimators as

$$\hat{m}_p(x) = \{B_{j,p}(x)\}_{1-p \leq j \leq N}^T \mathbf{V}_{n,p}^{-1} \{(\mathbf{Y}, B_{j,p})_n\}_{j=1-p}^N, \quad (3)$$

where $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ is the response vector. In particular, $p = 1, 2$ correspond to constant and linear spline estimators, respectively.

2.2 Jump detection

For strictly stationary bivariate time series $\{(X_t, Y_t)\}_{t=1}^n$, to detect jumps in the conditional mean function $m(x)$, we test the hypotheses $\mathcal{H}_0 : m \in C^{(p)}[a, b]$ vs. $\mathcal{H}_1 : m \notin C[a, b]$. We list the following assumptions for our theoretical justification.

- (A1) There exist a function $m_0(x) \in C^{(p)}[a, b]$ and a vector $\mathbf{c} = (c_1, \dots, c_k)$ of jump magnitudes such that the conditional mean function $m(x) = c_l + m_0(x)$, $x \in [\tau_l, \tau_{l+1})$, for $l = 1, \dots, k - 1$, $m(x) = m_0(x)$, $x \in [\tau_0, \tau_1)$, $m(x) = c_k + m_0(x)$, $x \in [\tau_k, \tau_{k+1})$, where $\{\tau_i\}_{i=1}^k$ is a sequence of potential jump points, given as $a = \tau_0 < \tau_1 < \dots < \tau_k < \tau_{k+1} = b$.
- (A2) The density $f(x)$ of X is continuous and positive on its compact support $[a, b]$. The standard deviation $\sigma(x)$ is continuous and positive on $[a, b]$.
- (A3) There exist positive constants K_0 and λ_0 such that $\alpha(k) \leq K_0 \exp^{-\lambda_0 k}$ for all k , with the strong mixing coefficient of order k defined as

$$\alpha(k) = \sup_{B \in \sigma\{Y_s, s \leq t\}, C \in \sigma\{Y_s, s \geq k+t\}} |P(B \cap C) - P(B)P(C)|, \quad k \geq 1.$$

- (A4) The number of interior knots $N \sim n^{1/(2p+1)}$, i.e. $c_N n^{1/(2p+1)} \leq N \leq C_N n^{1/(2p+1)}$ for some positive constants c_N, C_N .
- (A5) The noise ε_t satisfies $E(\varepsilon_t | X_t = x) = 0$, $E(\varepsilon_t^2 | X_t = x) = 1$, and there exists an $M_0 > 0$ such that

$$\sup_{x \in [a,b]} E(|\varepsilon_t|^3 | X_t = x) \leq M_0.$$

Remark 1 Assumption (A1) is similar as that in Ma and Yang (2011) and Müller and Song (1997). Assumptions (A2–A5) are regular in the polynomial spline smoothing literature; see, for instance, Huang and Yang (2004), Xue and Yang (2006), Wang and Yang (2007), Song and Yang (2009) and Wang and Yang (2010). Assumption (A1) says our target curve is smooth enough with exception at a few jumping locations. In application, one can truncate the data set to satisfy Assumption (A2). Assumption (A3) assumes a weak dependence in the time series. Assumption (A4) is a technique assumption we use for the selection of number of knots. We achieve our asymptotics using knots number with the assumed rate, and the proof is in the appendix. Assumption (A5) is a moment assumption on the noise.

Under Assumption (A1), the hypotheses amount to $\mathcal{H}_0 : \|\mathbf{c}\|_2 = 0$ vs. $\mathcal{H}_1 : \|\mathbf{c}\|_2 > 0$, where $\|\mathbf{c}\|_2 = (c_1^2 + \dots + c_k^2)^{1/2}$ is the Euclidean norm of the vector \mathbf{c} of all the k jump magnitudes. Based on the polynomial spline estimators given in Eq. (3), we define the test statistics $T_{p,n}$ for $p = 1, 2$ as

$$T_{p,n} = \max_{0 \leq j \leq N-1} \hat{\delta}_{p,j}, \quad \hat{\delta}_{p,j} = \frac{|\hat{m}_p(t_{j+1}) - \hat{m}_p(t_j)|}{\sigma_{n,p,j}}, \tag{4}$$

where

$$\sigma_{n,1,j}^2 = \sigma^2(t_{j+1})(f(t_{j+1})nh)^{-1} + \sigma^2(t_j)(f(t_j)nh)^{-1}, \tag{5}$$

$$\sigma_{n,2,j}^2 = \sigma^2(t_{j+1}) \left(\frac{2f(t_{j+1})nh}{3} \right)^{-1} \boldsymbol{\zeta}_j^T \mathbf{S}_j \boldsymbol{\zeta}_j, \tag{6}$$

with \mathbf{S}_j defined in Eq. (2) and

$$\boldsymbol{\zeta}_j = (-q_{j-1}, q_j)^T, \quad q_j = \begin{cases} 1, & 0 \leq j \leq N-1, \\ \sqrt{2}, & j = -1, N. \end{cases}$$

Theorem 1 Under Assumptions (A1–A5) and the null hypothesis \mathcal{H}_0 , for any $\alpha \in [0, 1]$, we have, for $p = 1, 2$,

$$\limsup_{n \rightarrow \infty} P \left[T_{p,n} > \{4 \log(N+1)\}^{1/2} d_n(\alpha/2) \right] \leq \alpha,$$

where

$$d_n(\alpha) = 1 - \{2 \log(N+1)\}^{-1} \left[\log\left(\frac{\alpha}{2}\right) + \frac{1}{2} \{ \log \log(N+1) + \log 4\pi \} \right].$$

3 Implementation

3.1 Calculate test statistic for detection of jumps

In this section, we provide practical guidance to implement the jump detection procedure. Given any sample $\{(X_t, Y_t)\}_{t=1}^n$, we denote the minimum and maximum values of $\{X_t\}_{t=1}^n$ as the endpoints of interval $[a, b]$. The optimal order of N , by Assumption (A4), is $n^{1/(2p+1)}$. In practice, we select number of knots $C_p n^{1/(2p+1)}$ with a relatively large positive constant C_p to deliberately undersmooth the data to capture possible jump points. In our numerical analysis, we take C_p in a range of $[5, 10]$, and we optimize N from $[5n^{1/(2p+1)}, \min(10n^{1/(2p+1)}, n/2 - p - 1)]$ through the following BIC:

$$\text{BIC} = \log(\text{RSS}/n) + \log(n)(N + p)/n, \tag{7}$$

where $N + p$ is the number of parameters to be estimated and RSS is the residual sum of squares, i.e. $\text{RSS} = \sum_{t=1}^n \{Y_t - \hat{m}_p(X_t)\}^2$. Some other widely used selection criteria such as AIC can also be considered. We use equally spaced knots for technical simplicity.

To calculate $T_{p,n}$ in Eq. (4), we need to estimate the unknown functions $f(x)$ and $\sigma^2(x)$. We estimate $f(x)$ using the density estimator $\hat{f}(x)$ with Quartic kernel $\tilde{K}(u) = 15(1-u^2)^2 I_{\{|u| \leq 1\}}/16$ and the rule-of-thumb bandwidth of Silverman (1986), i.e. $h_{\text{rot},f} = (4\pi)^{1/10} (140/3)^{1/5} n^{-1/5} s_n$. For $\sigma^2(x)$, we take the spline estimator $\hat{\sigma}_p^2(x)$ based on the data $\{(X_t, Z_{t,p})\}_{t=1}^n$, where $Z_{t,p}$ is denoted as the square of the residual $(Y_t - \hat{m}_p(X_t))^2$, $p = 1, 2$. The calculation of \mathbf{S}_j in Eq. (6) is as follows. According to Lemma A.2 of Wang and Yang (2010), the inner product matrix \mathbf{V}_2 is approximated by the following distribution-free matrix

$$\mathfrak{J} = \begin{pmatrix} 1 & \sqrt{2}/4 & & & 0 \\ \sqrt{2}/4 & 1 & 1/4 & & \\ & 1/4 & 1 & \ddots & \\ & & \ddots & 1/4 & \\ 0 & & & 1/4 & 1 & \sqrt{2}/4 \\ & & & & \sqrt{2}/4 & 1 \end{pmatrix}_{(N+2) \times (N+2)}$$

Define its inverse matrix as $\mathbf{L} = \mathfrak{J}^{-1}$. Then \mathbf{S}_j , the diagonal submatrices the inverse matrix of \mathbf{V}_2 , can be approximated by

$$\mathbf{L}_j = \begin{pmatrix} l_{j-1,j-1} & l_{j-1,j} \\ l_{j,j-1} & l_{j,j} \end{pmatrix}, \quad j = 0, 1, \dots, N,$$

which are the 2×2 diagonal submatrices of \mathbf{L} .

The asymptotic p value, i.e. $p_{\text{value},p}$ is obtained by solving the equation $T_{p,n} = \{4 \log(N + 1)\}^{1/2} d_n(\alpha/2)$ with $T_{p,n}$ defined in Eq. (4) for $p = 1, 2$. Replacing

$f(x)$, $\sigma^2(x)$ and \mathbf{S}_j in Eqs. (5) and (6) with $\hat{f}(x)$, $\hat{\sigma}_p^2(x)$ and \mathbf{L}_j , respectively, we derive the p value

$$p_{\text{value},p} = 4 \exp \left\{ [2 \log(N + 1)] [1 - (4 \log(N + 1))^{-1/2} T_{p,n}] - 2^{-1} [\log \log(N + 1) + \log 4\pi] \right\}. \tag{8}$$

If the p value is below a pre-determined significant level α , we conclude that there exist jump points in $m(x)$ at the α level.

3.2 Construction of multiple-ordered spline space

We have discussed how to test whether a mean function contains jumps. Naturally, the next question is to locate these jumps if there are any. Our idea of locating jumps is to apply our test statistics to local time intervals between consecutive knots and then decide whether these time intervals contain jumps or not. We replace $T_{p,n}$ in Eq. (8) with $\hat{\delta}_{p,j}$, $j = 0, \dots, N - 1$ to obtain the corresponding $p_{\text{value},p,j}$ for each $\hat{\delta}_{p,j}$. If $p_{\text{value},p,j} < \alpha$, we deduce there exists a jump point around t_j . Since we deliberately use a large amount of knots to capture the possible jumps, it is not surprising that there could be two or three consecutive $p_{\text{value},p} < \alpha$ around where there is a true jump point. In this circumstance, we only count one jump. To be specific, suppose the i th jump point is detected with $p_{\text{value},p,j_{\text{left},i}} < \alpha, \dots, p_{\text{value},p,j_{\text{right},i}} < \alpha$ consecutively from left to right, the location of the i th jump point is then estimated as $\hat{\tau}_i = (t_{j_{\text{left},i-1}} + t_{j_{\text{right},i+1}}) / 2$. As n goes to infinity, one has $\hat{\tau}_i \rightarrow \tau_i$ for $\tau_i \in [t_{j_{\text{left},i-1}}, t_{j_{\text{right},i+1}}]$, $i = 1, \dots, k$.

Since using a large amount of knots may lead to overfitting, we refit the data using polynomial splines with number of knots by a data-driven BIC method in the next step. To accommodate the detected jumps, we will adjust the spline space by adding discontinuous basis functions to the spline basis. Basis functions $(x - t_j)_+$ at knot t_j or $(x - t_j)_+$ and $(x - t_{j+1})_+$ at two adjacent knots t_j and t_{j+1} are considered to be added to the spline basis if there is a jump there or in between. We use the similar idea of multiple-order regression splines by Koo (1997). The magnitude of the i th jump point is then estimated by $\hat{m}_p(t_{j_{\text{right},i+1}}) - \hat{m}_p(t_{j_{\text{left},i-1}})$.

4 Simulation

In this section, we conduct simulation study to investigate the finite sample properties of our test statistics described in Sect. 2. We generated data from model (1) with

$$m(x) = \sin(2\pi x) + c_1 I(\tau_1 \leq x < \tau_2) + c_2 I(\tau_2 \leq x \leq 1), \tag{9}$$

here $\tau_1 = \sqrt{2}/4$, $\tau_2 = \sqrt{2}/2$. We set $\sigma(x) = \sigma_0 [100 - \exp(x)] / [100 + \exp(x)]$, $\varepsilon \sim N(0, 1)$, where $\sigma_0 = 0.2, 0.5, 0.8$ are the noise levels. $\{X_i\}_{i=1}^n$ were generated as $X_i = \Phi(\eta_i)$, $i = 1, \dots, n$, where $\{\eta_i\}_{i=1}^n$ were simulated from a moving average sequence of order 4,

$$\eta_i = \frac{(\xi_i + 0.2\xi_{i-1} + 0.2\xi_{i-2} + 0.2\xi_{i-3} + 0.2\xi_{i-4})}{\sqrt{1 + 0.2^2 + 0.2^2 + 0.2^2}}.$$

Here, ξ_i 's are i.i.d. r.v.'s $\sim N(0, 1)$. The sample sizes were taken to be $n = 200, 600, 1000$ and the significant levels $\alpha = 0.05, 0.01$. In each simulation, a total of $200 + n$ observations were generated and the first 200 observations were discarded to ensure the asymptotic stationarity. For each setting, we generated 500 replications, and we let $T_{p,n,q}$ be the q th replication of the test statistic $T_{p,n}$ in Eq. (4). We denote the asymptotic detection power by $\hat{\beta}_p(c_1, c_2)$ for $p = 1, 2$, where $\hat{\beta}_p(c_1, c_2)$ is defined as

$$\hat{\beta}_p(c_1, c_2) = \sum_{q=1}^{500} I \left[T_{p,n,q} > \{4 \log(N+1)\}^{1/2} d_n(\alpha/2) \right] / 500.$$

Table 1 shows jump detection rate $\hat{\beta}_p(c_1, c_2)$ for function $m(x)$ in (9) for $c_1 = c_2 = 0$ (no jump), $c_1 = 2, c_2 = 0$ (one jump) and $c_1 = 2, c_2 = -2$ (multiple jumps). For comparison, we also list, in the last two columns, the detection rate using kernel smoothing by Qiu (1994) by $\hat{\beta}_{ks}$ under different nominal level $\alpha = 0.05, 0.01$. It is worth to point out that the method in Qiu (1994) was proposed for independent data set. While as our referee suggested, we believe that the method is robust to the independence assumption. From Table 1, we observe that as the sample size n increases, $\hat{\beta}_p(2, 0)$ and $\hat{\beta}_p(2, -2)$, $p = 1, 2$ approach 1 and $\hat{\beta}_p(0, 0)$, $p = 1, 2$ tend to below the pre-determined significant level α , in agreement with the asymptotic theory. Note that, when there are no jumps ($c_1 = c_2 = 0$) and the noise levels are high ($\sigma_0 = 0.5, 0.8$), $\hat{\beta}_{ks}(0, 0)$ s tend to be 1 instead of 0, which indicate a serious over-detecting. It seems that the kernel smoothing detection by Qiu (1994) tends to over-detection by reading noise as jumps when noise level is high. Therefore, we conclude that our method has an advantage for highly noisy data sets. In all our tables, we present the simulation results with the number of knots chosen by the BIC given in (7). We actually have tried other selection criterion like AIC, and the results are comparable with small advantage on BIC.

For 500 replications satisfying $p_{\text{value},p} < \alpha = 0.05$, $p = 1, 2$ in the settings when there is only one jump with size $c_1 = 2, c_2 = 0$ or when there are two jumps with sizes $c_1 = 2, c_2 = -2$, we calculated the frequencies of detection for more than one jump points (FD^O) or the frequencies for not detecting two jumps (FD^U). These results are displayed in Tables 2 and 3, respectively. In Tables 2 and 3, we also show the frequencies for τ_1, τ_2 falling between $t_{j_{\text{left}},1-1}$ and $t_{j_{\text{right}},1+1}$ described in Sect. 3. From Tables 2 and 3, we can observe that the frequencies of coverage for τ_1 or τ_1, τ_2 obtained by linear splines are higher than those got by constant splines, getting close to the nominal level 0.95. We do observe some situations with occurrence of underestimating number of jumps since our test statistics are conservative. While, the underestimating rates decrease as n goes large. Furthermore, as described in Sect. 3, we obtain the estimates of τ_1, τ_2, c_1, c_2 , and we plot their kernel density estimates in Fig. 1 with sample sizes $n = 600$ and $n = 1000$ at $\sigma_0 = 0.5, \alpha = 0.05$. The vertical lines are set at $\tau_1 = \sqrt{2}/4, \tau_2 = \sqrt{2}/2$ and magnitudes 2, -4 . As seen in Fig. 1, the centers of the density plots become narrower and closer to the vertical lines as the

Table 1 Asymptotic power of the proposed tests over 500 replications in model (9)

c_1, c_2	σ_0	Sample size (n)	$\hat{\beta}_2(c_1, c_2)$ $\alpha = 0.05$	$\hat{\beta}_2(c_1, c_2)$ $\alpha = 0.01$	$\hat{\beta}_1(c_1, c_2)$ $\alpha = 0.05$	$\hat{\beta}_1(c_1, c_2)$ $\alpha = 0.01$	$\hat{\beta}_{ks}(c_1, c_2)$ $\alpha = 0.05$	$\hat{\beta}_{ks}(c_1, c_2)$ $\alpha = 0.01$
$c_1 = 0$ $c_2 = 0$	0.2	200	0.128	0.102	0.270	0.134	0.568	0.922
		600	0.044	0.024	0.194	0.074	0.152	0.476
		1000	0.032	0.002	0.158	0.046	0.040	0.276
	0.5	200	0.194	0.086	0.054	0.024	0.998	1.000
		600	0.086	0.036	0.010	0.004	0.990	1.000
		1000	0.046	0.008	0.004	0.002	0.948	1.000
	0.8	200	0.140	0.068	0.038	0.012	1.000	1.000
		600	0.034	0.022	0.004	0.004	1.000	1.000
		1000	0.016	0.004	0.000	0.002	1.000	1.000
$c_1 = 0$ $c_2 = 2$	0.2	200	0.560	0.324	0.998	0.998	0.946	0.996
		600	0.972	0.786	1.000	1.000	0.576	0.942
		1000	1.000	0.974	1.000	1.000	0.410	0.850
	0.5	200	0.798	0.532	0.692	0.562	1.000	1.000
		600	0.998	0.988	0.956	0.894	0.984	1.000
		1000	1.000	1.000	0.996	0.984	0.968	1.000
	0.8	200	0.374	0.232	0.268	0.130	1.000	1.000
		600	0.832	0.616	0.566	0.352	1.000	1.000
		1000	0.978	0.924	0.790	0.656	1.000	1.000
$c_1 = 2$ $c_2 = -2$	0.2	200	1.000	0.996	1.000	1.000	1.000	1.000
		600	1.000	1.000	1.000	1.000	1.000	1.000
		1000	1.000	1.000	1.000	1.000	0.988	1.000
	0.5	200	1.000	0.988	0.996	0.980	1.000	1.000
		600	1.000	1.000	1.000	1.000	1.000	1.000
		1000	1.000	1.000	1.000	1.000	1.000	1.000
	0.8	200	0.964	0.850	0.922	0.860	1.000	1.000
		600	1.000	1.000	1.000	0.988	1.000	1.000
		1000	1.000	1.000	1.000	1.000	1.000	1.000

Table 2 Frequencies of detection for more than one jump point and frequencies of coverage for τ_1 in model (9) for $c_1 = 2, c_2 = 0$

	σ_0	Sample size (n)	FD^O $p = 2$	FC^τ $p = 2$	FD^O $p = 1$	FC^τ $p = 1$
	0.5	200	0.046	0.826	0.002	0.639
		600	0.006	0.901	0.000	0.462
		1000	0.006	0.911	0.000	0.468
	0.8	200	0.026	0.702	0.000	0.526
		600	0.002	0.896	0.000	0.508
		1000	0.002	0.932	0.000	0.464

FD^O : frequencies of detection for more than one jump point, FC^τ : frequencies of coverage for τ

Table 3 Frequencies for not detecting two jumps and frequencies of coverage for τ_1, τ_2 in model (9) for $c_1 = 2, c_2 = -2$

σ_0	Sample size (n)	$FD^U_{p=2}$	$FC^{\tau_1}_{p=2}$	$FC^{\tau_2}_{p=2}$	$FD^U_{p=1}$	$FC^{\tau_1}_{p=1}$	$FC^{\tau_2}_{p=1}$
0.5	200	0.462	0.743	0.807	0.582	0.522	0.507
	600	0.232	0.859	0.911	0.388	0.526	0.500
	1000	0.228	0.878	0.940	0.182	0.501	0.513
	2000	0.184	0.887	0.949	0.062	0.512	0.542
0.8	200	0.292	0.370	0.415	0.812	0.500	0.457
	600	0.242	0.858	0.921	0.660	0.518	0.506
	1000	0.102	0.898	0.953	0.478	0.529	0.467
	2000	0.036	0.927	0.971	0.196	0.537	0.532

FD^U : frequencies for not detecting two jumps,
 FC^{τ_1} : frequencies of coverage for τ_1 , FC^{τ_2} : frequencies of coverage for τ_2

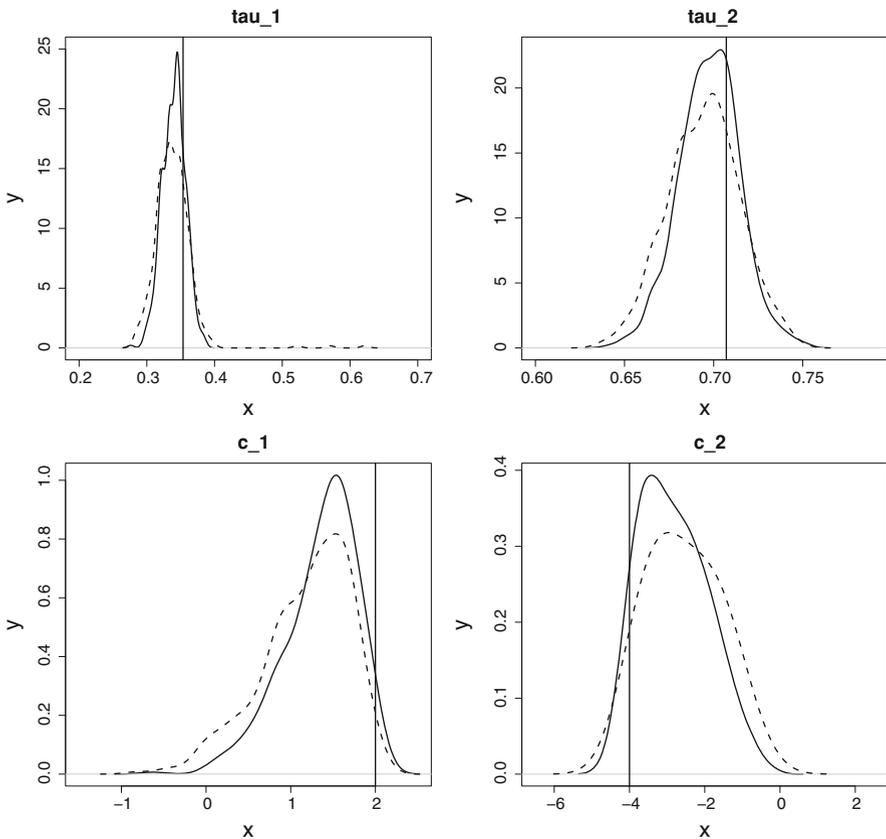


Fig. 1 Kernel density plots of $\hat{\tau}_1, \hat{\tau}_2$ and \hat{c}_1, \hat{c}_2 , $n = 600$ (dashed curve), $n = 1000$ (solid curve)

sample size n increases. Finally, to make an impression of the spline estimation for both smooth and discontinuous functions, at the noise level $\sigma_0 = 0.5$ with sample size $n = 1000$, we plot the linear spline estimate $\hat{m}_2(x)$ together with the true conditional mean function $m(x)$ in Fig. 2.

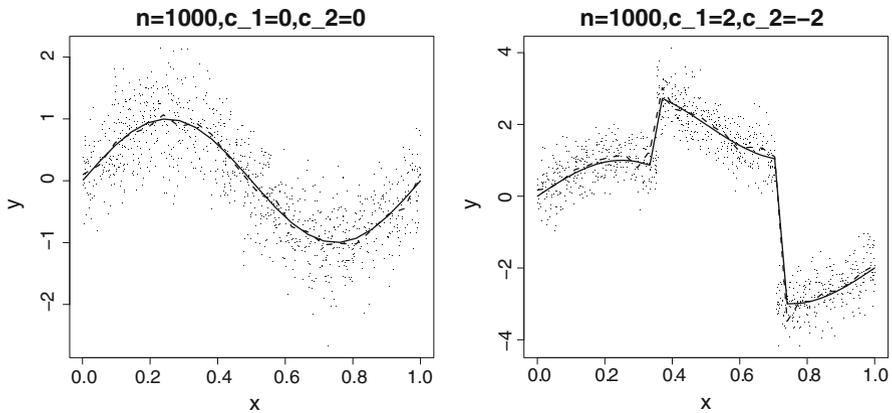


Fig. 2 Plots of the true function $m(x)$ (solid curve) and the linear spline estimates $\hat{m}_2(x)$ (dashed curve)

Table 4 Computing time (in seconds) per replication of generating and detecting jumps by local constant kernel smoothing, constant and linear splines

Sample size (n)	Local constant kernel smoothing	Constant splines	Linear splines
200	0.03	0.02	0.04
600	0.12	0.04	0.06
1000	0.26	0.07	0.08

In all our simulation experiments, the proposed spline method worked quickly, and we provide the time in seconds for all the methods in Table 4. The proposed spline method only needs to solve a moderate number of linear least squares; so in most cases one can see that the spline method worked much faster compared to its competitors such as kernel smoothing. We refer Xue and Yang (2006) and Wang and Yang (2007) for more computational time comparison results of the two methods.

5 Application

5.1 Global land-surface air temperature anomalies

The analysis of abrupt climate changes has recently found increasing interest; see, for instance, Alley et al. (2003), Ivanov and Evtimov (2010) and Matyasovszky (2011). Such abrupt changes may be caused by natural or human activities, such as solar and volcanic activities, greenhouse gases. Here, we consider the time series of global monthly land-surface air temperature anomalies from 1880 to 2011. The data set is available with the National Aeronautics and Space Administration Goddard Institute for Space Studies at http://data.giss.nasa.gov/gistemp/tabledata_v3/GLB.Ts.txt. Figure 3 displays the data together with the cubic spline estimator $\hat{m}_4(x)$. Given $\alpha = 0.05$, we detect five jump points in the year 1900, 1918, 1950, 1962, 1976, and the magnitudes ($^{\circ}\text{C}$) (p values) are 0.02 (0.008), -0.01 (0.000), -0.07 , (0.041), -0.11 (0.015), 0.11 (0.000),

Fig. 3 Global land-surface air temperature anomalies (*points*) and the cubic spline estimate $\hat{m}_4(x)$ (*solid curve*)

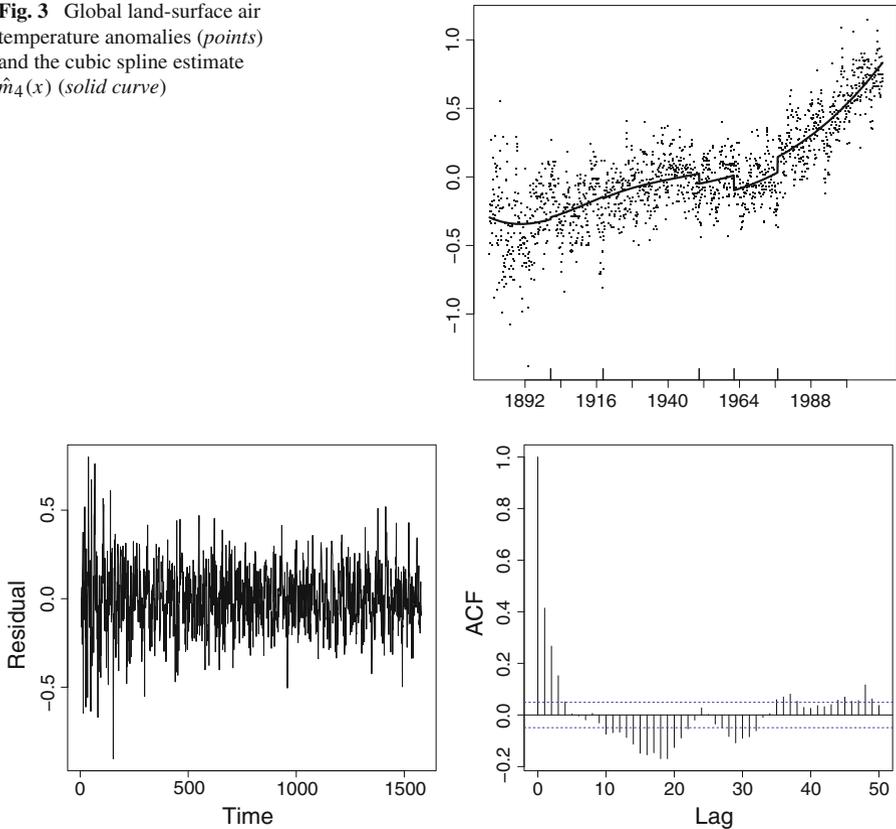


Fig. 4 Global land-surface air temperature: **a** plots of the residuals, **b** the estimated ACF with 95 % Bartlett intervals for the residuals

respectively. Besides these jumps, we see a warming tendency after the last positive jump in 1976.

For diagnostic purposes, we show in the Fig. 4 the residuals plot and the estimated autocorrelation function (ACF) plot of the residuals with 95 % Bartlett intervals. One sees that the temperature series is stable and the autocorrelation in the time series is very small.

5.2 Unemployment rates for men and women

We apply the proposed jump detection method to the monthly unemployment rates of men and women (20 years and over) in the United States. Each data set contains 565 observations from January 1965 to January 2012. Figure 5 displays these data points together with the linear spline estimates. Given the significant level $\alpha = 0.01$, we detect 4 jump points for men and 9 for women. We include in the Figs. 6 and 7 the residuals plots and the estimated ACF plots of the residuals with 95 % Bartlett intervals. An examination of these plots justifies the model assumptions.

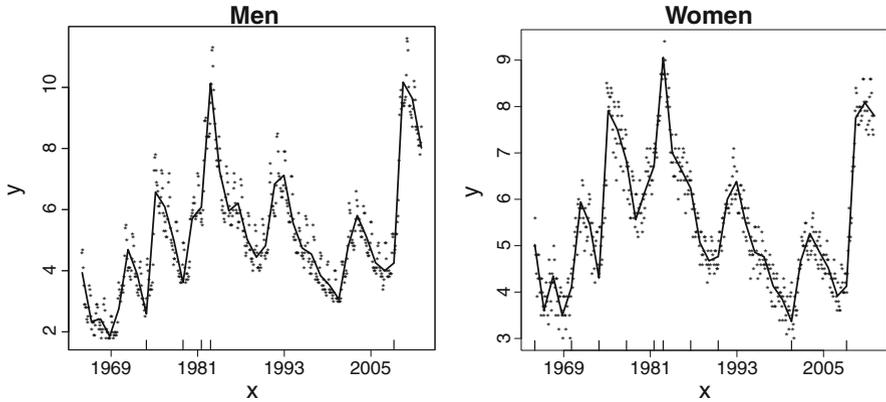


Fig. 5 Monthly unemployment rates (*points*) and the linear spline estimates $\hat{m}_2(x)$ (*solid curve*)

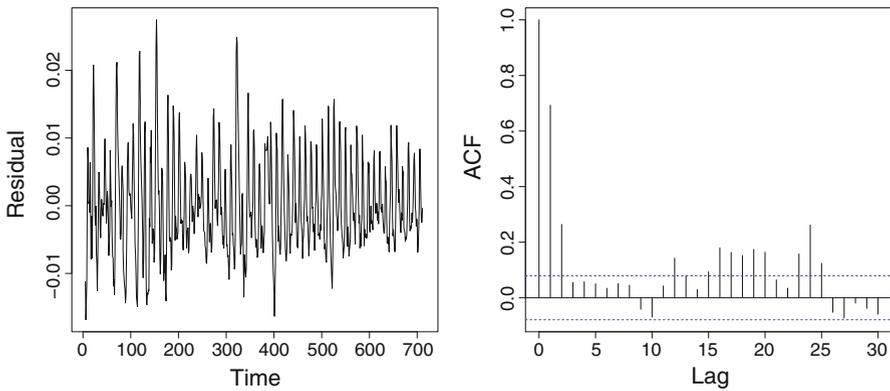


Fig. 6 Monthly unemployment rates for men: **a** plots of the residuals, **b** the estimated ACF with 95% Bartlett intervals for the residuals

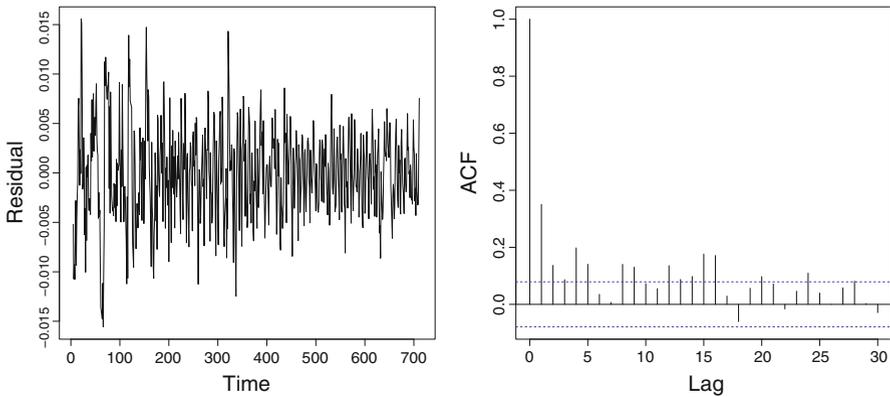


Fig. 7 Monthly unemployment rates for women: **a** plots of the residuals, **b** the estimated ACF with 95% Bartlett intervals for the residuals

Here and hereafter, we refer to the business cycle reference dates reported by NBER (2011) to compare the trends of unemployment rates of men and women. First, we present some jump points which are detected at the same time for men and women. Around December 1973, we detect a jump point for men with p value 0.0019 and magnitude 1.50, and a jump point for women with p value 0.0000 and magnitude 1.06. Then, for men, two consecutive p values 0.0019 and 0.0000 are detected to be below the pre-determined α around March 1982. Thus, we deduce there exists a jump point at this time of year and the magnitude is estimated to be 2.08. For women, we also detect two consecutive p values 0.0000 and 0.0000 around March 1982. So a jump point is detected and the estimated magnitude is 0.99. According to the business cycle reference dates, we can see the economy during these periods is in recession, resulting in sudden increases in the trends of unemployment rates. Undoubtedly, around March 2008, two jump points are detected in the financial crisis period; one for men with p value 0.0000 and magnitude 3.05, the other for women with p value 0.0000 and magnitude 1.63. Besides the above three jump points detected for men, around January 1979, there is one more jump point detected for men with p value 0.0026 and magnitude 0.20. Although the business cycle reference dates show a new recession starts from January 1980, from Fig. 5, we can see the trend of unemployment rate for men indeed begins to rise from January 1979. Next we list jump points that are only detected for women. Around January 1965, September 1977 and August 1986, three jump points are detected with p values 0.0011, 0.0073, 0.0004 and declined magnitudes 0.49, 1.44, 1.40, respectively. Since these years are all in the periods of expansion, the trends of unemployment rates fall. Another three jump points are detected around February 1970, June 1990 and August 2000. The p values of the three jumps are all 0.0000 and the magnitudes are 1.28, 1.78, 1.08. Similarly as stated before, these sudden increased trends occur either in the recession periods or several months before the recession periods.

In conclusion, the trends of unemployment rates for men and women both vary with the business cycles. With less jumps detected, the trend of unemployment rate for men is more stable than that of women. However, we observe that the magnitudes of positive jumps of men's unemployment rate is larger than (almost twice as) those of women's at the same period, showing that the crisis has more negative impact on men.

6 Concluding remarks

We have discussed jump detection and estimation of (dis)continuities of regression functions for time series data. The problem has wide applicability in empirical economic analysis and other areas. We have proposed a regression spline based algorithm that is intuitively appealing and simple to use. Test statistics for the existence of jumps are provided and their limiting distributions are derived under the null hypothesis that the mean function is smooth. Our idea of locating jumps is to apply our test statistic to local time intervals between consecutive knots and then decide whether these time intervals contain jumps or not. Simulations show that it has potential to work well in practice. Although we use equally spaced knots for the technical simplicity, it can be

proven that the same result will hold if the distance between neighboring knots are of the same order. Therefore, we can add some knots as needed.

Although our detection method is conservative, with proper selection of knot numbers, our simulation study shows that our method works quite well for data set with moderate size.

7 Appendix

The following notations are used throughout the proof. We define $\|\cdot\|_\infty$ as the supremum norm of a function r on $[a, b]$, i.e. $\|r\|_\infty = \sup_{x \in [a, b]} |r(x)|$. We will use c, C to denote some positive constants in a generic sense through the proof.

To prove Theorem 1, we decompose the estimation error $\hat{m}_p(x) - m_p(x)$ into a bias term and a noise term. Denoting $\mathbf{m} = (m(X_1), \dots, m(X_n))^T$ and $\mathbf{E}_\sigma = (\sigma(X_1)\varepsilon_1, \dots, \sigma(X_n)\varepsilon_n)^T$, we can rewrite \mathbf{Y} as $\mathbf{Y} = \mathbf{m} + \mathbf{E}_\sigma$. We project the response \mathbf{Y} onto the spline space $G_N^{(p-2)}$ spanned by $\{\mathbf{B}_{j,p}(\mathbf{X})\}_{j=1-p}^N$, where $\mathbf{B}_{j,p}(\mathbf{X})$ is denoted as

$$\mathbf{B}_{j,p}(\mathbf{X}) = \{B_{j,p}(X_1), \dots, B_{j,p}(X_n)\}^T, \quad j = 1 - p, \dots, N,$$

with $B_{j,p}(x)$ introduced in Sect. 2.1. We obtain the following decomposition

$$\hat{m}_p(x) = \tilde{m}_p(x) + \tilde{\varepsilon}_p(x),$$

where

$$\begin{aligned} \tilde{m}_p(x) &= \{B_{j,p}(x)\}_{1-p \leq j \leq N}^T \mathbf{V}_{n,p}^{-1} \{(\mathbf{m}, B_{j,p})_n\}_{j=1-p}^N, \\ \tilde{\varepsilon}_p(x) &= \{B_{j,p}(x)\}_{1-p \leq j \leq N}^T \mathbf{V}_{n,p}^{-1} \{(\mathbf{E}_\sigma, B_{j,p})_n\}_{j=1-p}^N. \end{aligned} \tag{10}$$

The bias term is $\tilde{m}_p(x) - m_p(x)$ and the noise term is $\tilde{\varepsilon}_p(x)$.

Lemma 1 As $n \rightarrow \infty$,

$$\begin{aligned} \|b_{j,1}\|_2^2 &= f(t_j)h(1+r_{j,n,1}), \quad \|b_{j,2}\|_2^2 = \frac{2f(t_{j+1})h}{3} \begin{cases} 1+r_{j,n,2}, & 0 \leq j \leq N-1, \\ \frac{1}{2}+r_{j,n,2}, & j = -1, N, \end{cases} \\ \langle b_{j,1}, b_{j',1} \rangle &= \begin{cases} 1, & j = j', \\ 0, & j \neq j', \end{cases} \quad \langle b_{j,2}, b_{j',2} \rangle = \frac{1}{6}f(t_{j+1})h \begin{cases} 1+\tilde{r}_{j,n,2}, & |j'-j|=1, \\ 0, & |j'-j|>1, \end{cases} \end{aligned}$$

where $\max_{0 \leq j \leq N} |r_{j,n,1}| + \max_{-1 \leq j \leq N} |r_{j,n,2}| + \max_{-1 \leq j \leq N-1} |\tilde{r}_{j,n,2}| \leq C\omega(f, h)$ and $\omega(f, h) = \max_{x, x' \in [a, b], |x-x'| \leq h} |f(x) - f(x')|$ is the moduli of continuity of a continuous function f on $[a, b]$. Furthermore,

$$\frac{1}{3}f(t_{j+1})h \{1 - C\omega(f, h)\} \leq \|b_{j,2}\|_2^2 \leq \frac{2}{3}f(t_{j+1})h \{1 + C\omega(f, h)\}.$$

Proof of Theorem 1 for $p = 1$ When $p = 1$, $\mathbf{V}_{n,1}^{-1}$ is a diagonal matrix, and $\tilde{\varepsilon}_1(x)$ in Eq. (10) can be rewritten as

$$\tilde{\varepsilon}_1(x) = \sum_{j=0}^N \varepsilon_j^* B_{j,1}(x) \|B_{j,1}\|_{2,n}^{-2}, \quad \varepsilon_j^* = n^{-1} \sum_{i=1}^n B_{j,1}(X_i) \sigma(X_i) \varepsilon_i, \quad x \in [a, b].$$

We define $\hat{\varepsilon}_1(x) = \sum_{j=0}^N \varepsilon_j^* B_{j,1}(x)$, and it is straightforward that $\hat{\varepsilon}_1(t_j) = B_{j,1}(t_j) \varepsilon_j^*$, $j = 0, \dots, N$. We treat the variance of $\hat{\varepsilon}_1(t_{j+1}) - \hat{\varepsilon}_1(t_j)$ as follows.

Lemma 2 *The variance of $\hat{\varepsilon}_1(t_{j+1}) - \hat{\varepsilon}_1(t_j)$, $j = 0, \dots, N - 1$, is $\sigma_{n,1,j}^2$ in Eq. (5), which satisfies*

$$\sigma_{n,1,j}^2 = E\{\hat{\varepsilon}_1(t_{j+1}) - \hat{\varepsilon}_1(t_j)\}^2 = \sigma^2(t_{j+1})(f(t_{j+1})nh)^{-1} + \sigma^2(t_j)(f(t_j)nh)^{-1}.$$

Accordingly, under Assumption (A2), one has $c(nh)^{-1/2} \leq \sigma_{n,1,j} \leq C(nh)^{-1/2}$ for any $j = 0, \dots, N - 1$ as n sufficiently large.

The proof can be easily obtained by Lemma A.1 combining with the fact that $(B_{j,1}, B_{j+1,1}) = 0$.

Denote, for $0 \leq j \leq N - 1$, $\tilde{\xi}_{n,1,j} = \sigma_{n,1,j}^{-1} \{\tilde{\varepsilon}_1(t_{j+1}) - \tilde{\varepsilon}_1(t_j)\}$ and $\hat{\xi}_{n,1,j} = \sigma_{n,1,j}^{-1} \{\hat{\varepsilon}_1(t_{j+1}) - \hat{\varepsilon}_1(t_j)\}$. The next lemma follows from Lemma A.6 of Wang and Yang (2010).

Lemma 3 *Under Assumptions (A2–A4), as $n \rightarrow \infty$,*

$$\left| \sup_{0 \leq j \leq N-1} |\hat{\xi}_{n,1,j}| - \sup_{0 \leq j \leq N-1} |\tilde{\xi}_{n,1,j}| \right| = O_p \left\{ (nh)^{-1/2} \log n \right\}.$$

Proof Rewrite $\hat{\varepsilon}_1(t_{j+1}) - \hat{\varepsilon}_1(t_j)$ as $\hat{\varepsilon}_1(t_{j+1}) - \hat{\varepsilon}_1(t_j) = \mathbf{D}_{j,1}^T \mathbf{\Lambda}_{j,1}$, $j = 0, \dots, N - 1$, where

$$\begin{aligned} \mathbf{D}_{j,1} &= (-n^{-1/2} B_{j,1}(t_j), n^{-1/2} B_{j+1,1}(t_{j+1}))^T, \\ \mathbf{\Lambda}_{j,1} &= \begin{pmatrix} n^{-1/2} \sum_{i=1}^n B_{j,1}(X_i) \sigma(X_i) \varepsilon_i \\ n^{-1/2} \sum_{i=1}^n B_{j+1,1}(X_i) \sigma(X_i) \varepsilon_i \end{pmatrix}. \end{aligned}$$

It follows that $\sigma_{n,1,j}^2 = \mathbf{D}_{j,1}^T \text{Cov}(\mathbf{\Lambda}_{j,1}) \mathbf{D}_{j,1}$ with

$$\text{Cov}(\mathbf{\Lambda}_{j,1}) = \begin{pmatrix} EB_{j,1}^2(X) \sigma^2(X) & 0 \\ 0 & EB_{j+1,1}^2(X) \sigma^2(X) \end{pmatrix}.$$

Let $\mathbf{Z}_j = (Z_{j1}, Z_{j2})^T = \mathbf{\Lambda}_{j,1}^T \{Cov(\mathbf{\Lambda}_{j,1})\}^{-1/2}$. Specifically, for $0 \leq j \leq N - 1$,

$$Z_{j1} = \left\{ nEB_{j,1}^2(X)\sigma^2(X)\varepsilon^2 \right\}^{-1/2} \left\{ \sum_{i=1}^n B_{j,1}(X_i)\sigma(X_i)\varepsilon_i \right\},$$

$$Z_{j2} = \left\{ nEB_{j+1,1}^2(X)\sigma^2(X)\varepsilon^2 \right\}^{-1/2} \left\{ \sum_{i=1}^n B_{j+1,1}(X_i)\sigma(X_i)\varepsilon_i \right\}.$$

By Lemmas 3.2, 3.3 and A.7 of Wang and Yang (2010), we have uniformly in j ,

$$P \left[|Z_{j\gamma}| \leq \{2 \log(N + 1)\}^{1/2} d_n \left(\frac{\alpha}{2} \right) \right] = 1 - \frac{\alpha}{2(N + 1)} + o(N^{-1}), \gamma = 1, 2.$$

Therefore, for $\gamma = 1, 2$,

$$\limsup_{n \rightarrow \infty} P \left[\max_{0 \leq j \leq N} Z_{j\gamma}^2 > 2 \log(N + 1) \left\{ d_n \left(\frac{\alpha}{2} \right) \right\}^2 \right] \leq \frac{\alpha}{2}.$$

Denote $\mathbf{Q}_{j,1} = \mathbf{\Lambda}_{j,1}^T \{Cov(\mathbf{\Lambda}_{j,1})\}^{-1} \mathbf{\Lambda}_{j,1} = \mathbf{Z}_j \mathbf{Z}_j^T = \sum_{\gamma=1,2} Z_{j\gamma}^2, j = 0, \dots, N - 1$. According to the maximization lemma of Johnson and Wichern (1992),

$$\left\{ \sigma_{n,1,j}^{-1} [\hat{\varepsilon}_1(t_{j+1}) - \hat{\varepsilon}_1(t_j)] \right\}^2 \leq \mathbf{\Lambda}_{j,1}^T \{Cov(\mathbf{\Lambda}_{j,1})\}^{-1} \mathbf{\Lambda}_{j,1} = \mathbf{Q}_{j,1}.$$

Hence,

$$\begin{aligned} & \liminf_{n \rightarrow \infty} P \left[\sup_{0 \leq j \leq N-1} \left| \sigma_{n,1,j}^{-1} [\hat{\varepsilon}_1(t_{j+1}) - \hat{\varepsilon}_1(t_j)] \right| \leq 2 \{ \log(N + 1) \}^{1/2} d_n \left(\frac{\alpha}{2} \right) \right] \\ & \geq \liminf_{n \rightarrow \infty} P \left[\max_{0 \leq j \leq N-1} \mathbf{Q}_{j,1} \leq 4 \log(N + 1) \left\{ d_n \left(\frac{\alpha}{2} \right) \right\}^2 \right] \\ & \geq 1 - \sum_{\gamma=1,2} \limsup_{n \rightarrow \infty} P \left[\max_{0 \leq j \leq N-1} Z_{j\gamma}^2 > 2 \log(N + 1) \left\{ d_n \left(\frac{\alpha}{2} \right) \right\}^2 \right] \\ & \geq 1 - \alpha. \end{aligned}$$

Note that $\hat{m}_1(t_{j+1}) - \hat{m}_1(t_j) = [\tilde{m}_1(t_{j+1}) - m(t_{j+1})] - [\tilde{m}_1(t_j) - m(t_j)] + [m(t_{j+1}) - m(t_j)] + [\tilde{\varepsilon}_1(t_{j+1}) - \tilde{\varepsilon}_1(t_j)]$. The theorem of de Boor (2001) on page 149 and Theorem 5.1 of Huang (2003) entail that under \mathcal{H}_0 the orders of the first three terms are all $Op(h)$, which makes

$$\sigma_{n,1,j}^{-1} [\log(N + 1)]^{-1/2} \|\tilde{m} - m\|_\infty = Op \left\{ (nh)^{1/2} h [\log(N + 1)]^{-1/2} \right\} = o_p(1).$$

We finally apply Lemma 3 to get

$$\begin{aligned} & \limsup_{n \rightarrow \infty} P \left[\sup_{0 \leq j \leq N-1} \sigma_{n,1,j}^{-1} |\hat{m}_1(t_{j+1}) - \hat{m}_1(t_j)| > \{4 \log(N+1)\}^{1/2} d_n(\alpha/2) \right] \\ &= \limsup_{n \rightarrow \infty} P \left[\sup_{0 \leq j \leq N-1} \sigma_{n,1,j}^{-1} |\hat{\varepsilon}_1(t_{j+1}) - \hat{\varepsilon}_1(t_j)| > \{4 \log(N+1)\}^{1/2} d_n(\alpha/2) \right] \\ &\leq \alpha. \end{aligned}$$

□

Proof of Theorem 1 for p = 2 For $p = 2$, we can rewrite the noise term $\tilde{\varepsilon}_2(x)$ in Eq. (10) as $\tilde{\varepsilon}_2(x) = \sum_{j=-1}^N \tilde{a}_j B_{j,2}(x)$, where

$$\tilde{\mathbf{a}} = (\tilde{a}_{-1}, \dots, \tilde{a}_N)^T = \mathbf{V}_{n,2}^{-1} \left\{ n^{-1} \sum_{i=1}^n B_{j,2}(X_i) \sigma(X_i) \varepsilon_i \right\}_{j=-1}^N.$$

Similarly as before, we denote $\hat{\varepsilon}_2(x) = \sum_{j=-1}^N \hat{a}_j B_{j,2}(x)$, where $\hat{\mathbf{a}} = (\hat{a}_{-1}, \dots, \hat{a}_N)^T$ is defined by replacing $\mathbf{V}_{n,2}^{-1}$ in the above formula with $\mathbf{S} = \mathbf{V}_2^{-1}$, i.e.

$$\begin{aligned} \hat{\mathbf{a}} &= \mathbf{S} \left\{ n^{-1} \sum_{i=1}^n B_{j,2}(X_i) \sigma(X_i) \varepsilon_i \right\}_{j=-1}^N \\ &= \left\{ \sum_{j'=-1}^N s_{j'j} n^{-1} \sum_{i=1}^n B_{j,2}(X_i) \sigma(X_i) \varepsilon_i \right\}_{j'=-1}^N. \end{aligned}$$

Thus, for any $x \in [a, b]$,

$$\hat{\varepsilon}_2(x) = \sum_{j'=-1}^N \hat{a}_{j'} B_{j',2}(x) = \sum_{j,j'=-1}^N s_{j'j} n^{-1} \sum_{i=1}^n B_{j,2}(X_i) \sigma(X_i) \varepsilon_i B_{j',2}(x).$$

Denote $\tilde{\xi}_{2,j} = \tilde{\varepsilon}_2(t_{j+1}) - \tilde{\varepsilon}_2(t_j)$, $\hat{\xi}_{2,j} = \hat{\varepsilon}_2(t_{j+1}) - \hat{\varepsilon}_2(t_j)$, and $\tilde{\xi}_{n,2,j} = \sigma_{n,2,j}^{-1} \tilde{\xi}_{2,j}$, $\hat{\xi}_{n,2,j} = \sigma_{n,2,j}^{-1} \hat{\xi}_{2,j}$. It follows that $\hat{\xi}_{2,j} = \mathbf{D}_{j,2}^T \mathbf{\Lambda}_{j,2}$, $j = 0, \dots, N - 1$, where

$$\begin{aligned} \mathbf{D}_{j,2} &= \left(-n^{-1/2} B_{j-1,2}(t_j), n^{-1/2} B_{j,2}(t_{j+1}) \right)^T, \\ \mathbf{\Lambda}_{j,2} &= \begin{pmatrix} n^{-1/2} \sum_{j'=-1}^N \sum_{i=1}^n B_{j',2}(X_i) \sigma(X_i) \varepsilon_i s_{j-1,j'} \\ n^{-1/2} \sum_{j'=-1}^N \sum_{i=1}^n B_{j',2}(X_i) \sigma(X_i) \varepsilon_i s_{j,j'} \end{pmatrix}. \end{aligned}$$

In the next lemma, we calculate the variance of $\hat{\xi}_{2,j}$.

Lemma 4 The variance of $\hat{\xi}_{2,j}$ is $\sigma_{n,2,j}^2$ in Eq. (6), which satisfies

$$\sigma_{n,2,j}^2 = \sigma^2(t_{j+1}) \left(\frac{2f(t_{j+1})nh}{3} \right)^{-1} \xi_j^T \mathbf{S}_j \xi_j, \quad j = 0, \dots, N-1.$$

And for large enough n , $c(nh)^{-1/2} \leq \sigma_{n,2,j} \leq C(nh)^{-1/2}$.

Proof Since $\sigma_{n,2,j}^2 = E\hat{\xi}_{2,j}^2 = \mathbf{D}_j^T \text{Cov}(\Lambda_j) \mathbf{D}_j$, by applying Lemma A.10 of Wang and Yang (2010), we can get the desired results. \square

Similar arguments used in Lemmas A.11 and A.12 of Wang and Yang (2010) yield that

$$\liminf_{n \rightarrow \infty} P \left[\sup_{0 \leq j \leq N-1} \left| \hat{\xi}_{n,2,j} \right| \leq 2 \{\log(N+1)\}^{1/2} d_n \left(\frac{\alpha}{2} \right) \right] \geq 1 - \alpha.$$

Then we can finish the proof similarly as for $p = 1$. \square

References

- Alley, R. B., Marotzke, J., Nordhaus, W. D., Overpeck, J. T., Peteet, D. M., Pielke, R., et al. (2003). Abrupt climate change. *Science*, 299, 2005–2010.
- Bowman, A. W., Pope, A., Ismail, B. (2006). Detecting discontinuities in nonparametric regression curves and surfaces. *Statistics and Computing*, 16, 377–390.
- Chen, G. M., Choi, Y. K., Zhou, Y. (2008). Detections of changes in return by a wavelet smoother with conditional heteroscedastic volatility. *Journal of Econometrics*, 143, 227–262.
- de Boor, C. (2001). *A practical guide to splines*. New York: Springer.
- Gijbels, I., Lambert, A., Qiu, P. H. (2007). Jump-preserving regression and smoothing using local linear fitting: a compromise. *Annals of the Institute of Statistical Mathematics*, 59, 235–272.
- Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *The Annals of Statistics*, 31, 1600–1635.
- Huang, J. Z., Yang, L. (2004). Identification of nonlinear additive autoregressive models. *Journal of the Royal Statistical Society Series B*, 66, 463–477.
- Hui, E. C. M., Yu, C. K. W., Ip, W. C. (2010). Jump point detection for real estate investment success. *Physica A*, 389, 1055–1064.
- Ivanov, M. A., Evtimov, S. N. (2010). 1963: The break point of the northern hemisphere temperature trend during the twentieth century. *International Journal of Climatology*, 30, 1738–1746.
- Johnson, R. A., Wichern, D. W. (1992). *Applied multivariate statistical analysis*. Englewood Cliffs: Prentice Hall.
- Joo, J., Qiu, P. H. (2009). Jump detection in a regression curve and its derivative. *Technometrics*, 51, 289–305.
- Koo, J. Y. (1997). Spline estimation of discontinuous regression functions. *Journal of Computational and Graphical Statistics*, 6, 266–284.
- Lin, Z. Y., Li, D., Chen, J. (2008). Change point estimators by local polynomial fits under a dependence assumption. *Journal of Multivariate Analysis*, 99, 2339–2355.
- Ma, S. J., Yang, L. J. (2011). A jump-detecting procedure based on spline estimation. *Journal of Nonparametric Statistics*, 23, 67–81.
- Matyasovszky, I. (2011). Detecting abrupt climate changes on different time scales. *Theoretical and Applied Climatology*, 105, 445–454.
- Müller, H. G. (1992). Change-points in nonparametric regression analysis. *The Annals of Statistics*, 20, 737–761.

- Müller, H. G., Song, K. S. (1997). Two-stage change-point estimators in smooth regression models. *Statistics and Probability Letters*, 34, 323–335.
- Müller, H. G., Stadtmüller, U. (1999). Discontinuous versus smooth regression. *The Annals of Statistics*, 27, 299–337.
- National Bureau of Economic Research (2011). US business cycle expansions and contractions. <http://www.nber.org/cycles/cyclesmain.html#announcements>.
- Qiu, P. H. (1991). Estimation of a kind of jump regression functions. *Systems Science and Mathematical Sciences*, 4, 1–13.
- Qiu, P. H. (1994). Estimation of the number of jumps of the jump regression functions. *Communications in Statistics-Theory and Methods*, 23, 2141–2155.
- Qiu, P. H. (2003). A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation. *Journal of Nonparametric Statistics*, 15, 437–453.
- Qiu, P. H. (2005). *Image processing and jump regression analysis*. New York: Wiley.
- Qiu, P. H., Yandell, B. (1998). A local polynomial jump detection algorithm in nonparametric regression. *Technometrics*, 40, 141–152.
- Qiu, P. H., Asano, C., Li, X. (1991). Estimation of jump regression function. *Bulletin of Informatics and Cybernetics*, 24, 197–212.
- Shiau, J. H. (1987). A note on MSE coverage intervals in a partial spline model. *Communications in Statistics Theory and Methods*, 16, 1851–1866.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.
- Song, Q., Yang, L. (2009). Spline confidence bands for variance function. *Journal of Nonparametric Statistics*, 21, 589–609.
- Wang, L., Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *The Annals of Statistics*, 35, 2474–2503.
- Wang, L., Yang, L. (2010). Simultaneous confidence bands for time-series prediction function. *Journal of Nonparametric Statistics*, 22, 999–1018.
- Wong, H., Ip, W., Li, Y. (2001). Detection of jumps by wavelets in a heteroscedastic autoregressive model. *Statistics and Probability Letters*, 52, 365–372.
- Wu, J. S., Chu, C. K. (1993). Kernel type estimators of jump points and values of a regression function. *The Annals of Statistics*, 21, 1545–1566.
- Wu, W. B., Zhao, Z. B. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society Series B*, 69, 391–410.
- Xue, L., Yang, L. (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica*, 16, 1423–1446.
- Zhou, Y., Wan, A. T. K., Xie, S. Y., Wang, X. J. (2010). Wavelet analysis of change-points in a non-parametric regression with heteroscedastic variance. *Journal of Econometrics*, 159, 183–201.