Statistics and Its Interface Volume 0 (2023) 1

Simultaneous Change-point Detection and Curve Estimation *

Zhaoying Lu, Ning Hao and Hao Helen Zhang†

In this work, we focus on a nonparametric regression model that accounts for discontinuities. We propose a method called Simultaneous CHange-point detection And Curve Estimation (SCHACE) for effectively detecting jumps in a data sequence and accurately capturing nonlinear trends between these jumps in the mean curve. The SCHACE is a unified regularization framework that incorporates two statistical tools: the normalized fused Lasso for change-point detection and B-splines for curve estimation. Notably, this approach is a single-step method that does not require iteration and is straightforward to implement. We demonstrate the advantages of the SCHACE by simulated and real-world data examples.

AMS 2000 subject classifications: Primary 62G08, 62J07.
Keywords and phrases: B-splines, Jump regression, Nonparametric regression, Normalized fused Lasso.

1. INTRODUCTION

Change-point detection, also known as jump detection, is a critical component in the analysis of sequential data across various fields, including biology, climatology, computer science, econometrics, and engineering. Its primary objective is to identify sudden and significant changes or jumps within a sequence of observations. This technique holds great significance in practical applications such as monitoring heart rates, daily temperature fluctuations, greenhouse gas emissions, and analyzing streaming video data. Over the years, numerous change-point detection techniques have been developed to effectively identify multiple changes within potentially long and noisy sequences. Some popular approaches include binary segmentation and variants [37, 27, 11], the screening and rank algorithm [26, 14], SMUCE [10], and penalized regression [16, 15, 35, 28] among many others. While change-point detection methods are effective in scenarios where the true model structure follows a piecewise constant pattern, they may not provide the necessary flexibility to accurately capture gradual mean changes and nonlinear trends in data sequences. In such cases, a nonparametric regression model with jumps becomes beneficial. Estimating regression curves that incorporate jumps has gained considerable attention and has been applied across a diverse range of fields. This approach allows for a more comprehensive analysis of data sequences, enabling the modeling of both abrupt changes and smooth, nonlinear trends. For example, in economics, data analysis to the exchange rate data can detect the sharp change in the exchange rate of the Icelandic Króna (ISK) per US dollar (USD) in September 2008 that was likely caused by the subprime mortgage crisis in the United States [18]; in climatology, a jump detection method [39] can identify two jumps in sea-level pressure using a data set of December sea-level pressures in the period 1921–1992 by the Bombay weather station in India; in ecology, piecewise-smooth function estimation displays the trends of the relative light transmittance with discontinuities, which is helpful to understand the forest dynamics in the ecosystem [1].

As a classical topic, nonparametric regression aims to estimate a regression function that is continuously differentiable up to a certain order. Various nonparametric regression methods have been developed, e.g., local polynomial regression [7], splines [38, 6], and wavelet methods [2] among many others. Nonparametric regression with change points was studied by McDonald and Owen [22], which considered the split linear smoother by the weighted average of the left, right, and central linear fits for each point. Similar approaches using statistics constructed from the one- or two-sided observations of any design point have been explored in the literature. For example, a variety of methods have been developed based on locally defined diagnostic statistics [41, 13, 32, 24]. Besides aforementioned method, other approaches include curve fitting with change points using splines [34, 17, 40, 20], kernel smoother [39, 30, 12], and Bayesian methods [4, 8]. See [31] for an in-depth review of the jump regression analysis.

Two main goals in jump regression analysis are jump detection and curve estimation. In the literature, a popular two-step strategy involves determining the jump positions first, followed by curve estimation [32, 24, 33]. When the jump points are accurately estimated, curve estimation in smooth regions is straightforward. While these methods perform well when the jump locations are easily detectable, they heavily rely on the accuracy of jump detection, which can be challenging when dealing with noisy data. Consequently, the

*This project is partially supported by National Science Foundation Grant DMS-1722691 (Hao), CCF 1740858 (Zhang), National Institutes of Health Grant CA 260399 (Zhang), and Simons Foundation Grant 524432 (Hao).
†Corresponding author.

precision of curve estimation is significantly influenced by the effectiveness of jump detection, particularly in scenarios where the data is prone to noise. Another useful technique is jump-preserving curve estimation [30, 12] which directly estimates the regression curve while adapting at each point to a possible discontinuity. The major goal of the second approach is to achieve an accurate estimation of the mean function, which may not necessarily uncover jump points well. Moreover, some recent methods include penalized regression with truncated power spline basis [18] and Piecewise Constant plus Smooth Regression Estimator (PCpluS) [29].

In this paper, we introduce a novel approach named Simultaneous CHange-point detection And Curve Estimation (SCHACE) for detecting change points in data and capturing the underlying nonlinear and smooth trends between these change points. One main advantage of SCHACE is its unified regression framework, seamlessly integrating the normalized fused Lasso and B-spline regression into a single-step process. By incorporating these techniques, the SCHACE method enables the automatic accommodation of change points in the nonparametric regression curve fitting. Moreover, it simultaneously identifies the number and locations of the change points, determines jump magnitudes, and estimates the regression curve. This comprehensive approach sets SCHACE apart in effectively addressing change point detection and curve estimation in a unified manner.

The remainder of this paper is organized as follows. Section 2 introduces the regression model with discontinuities and the proposed SCHACE method. In Section 3, we demonstrate numerical experiments based on both synthetic and real data examples. We conclude the paper with a short discussion in Section 4.

2. METHODOLOGY

2.1 Model

Throughout the paper, we use $\mathbf{y} = (y_1, ..., y_N)^\top$ to denote a sequence of observations generated by

$$y_i = f(i/N) + \epsilon_i \quad \text{for} \quad i = 1, \cdots, N, \tag{1}$$

where $f(\cdot)$ is a piecewise continuous function on $[0, 1]$, and $\{\epsilon_i\}_{i=1}^N$ are independent and identically distributed (IID) errors with mean zero and variance σ^2. We assume that $f(\cdot)$ is a left continuous function with J discontinuities at $0 < \tau_1 < \cdots < \tau_J < 1$. That is,

$$\lim_{x \to \tau_j^-} f(x) = f(\tau_j) \neq \lim_{x \to \tau_j^+} f(x).$$

We call τ_j a change point, or a jump point. We aim to estimate the piecewise function $f(\cdot)$ and identify its jump points.

Two important special cases of model (1) have been intensively studied in the literature. First, without discontinuity, model (1) is reduced to the classic nonparametric regression model. Local polynomial regression [7] and spline method [38] are two popular approaches to nonparametric curve estimation. Second, if $f(\cdot)$ is a piecewise constant function, it becomes a standard multiple change-point model. See [25] for a recent review. Among many change-point detection methods, the fused Lasso approach can identify change points and estimate the mean function simultaneously. As illustrated in section 2.2, our approach is closely related to the spline method and the fused Lasso.

2.2 Simultaneous Change-Point Detection and Curve Estimation

Our proposed framework combines the B-splines and the normalized fused Lasso into a penalized regression. Before introducing the new method, we briefly review relevant statistical tools for curve estimation and change-point detection.

The B-spline method utilizes a linear combination of the B-spline basis functions of order n to fit the data. To elaborate, given the set of nondecreasing knot points $t_0 \leq t_1 \leq \cdots \leq t_{s+2n-1}$ with s internal knots, the basis functions of order j are defined recursively by

$$B_{i,1}(x) = \begin{cases} 1, & t_i \leq x \leq t_{i+1}, \\ 0, & \text{otherwise;} \end{cases}$$

$$B_{i,j}(x) = \frac{x - t_i}{t_{i+j-1} - t_i} B_{i,j-1}(x) + \frac{t_{i+j} - x}{t_{i+j} - t_{i+1}} B_{i+1,j-1}(x),$$

for $i = 1, ..., 2n - j$. For a fixed order n (that corresponds to degree $n - 1$ polynomials) and properly chosen s interior knots, we regress the response $\mathbf{y}$ on B-spline basis $\{B_{i,n}\}$. Thus, the estimation of f can be expressed as

$$\sum_{i=1}^{s+n} \gamma_i B_{i,n}(x). \tag{2}$$

We fix $n = 4$ in this paper. Define a matrix $\mathbf{Z}$ with $Z_{ij} = B_{j,n}(i/N)$.

The fused Lasso [35] is designed for detecting the change points from perturbed observations of piecewise constant signals. Assume that $\mathbf{y}$ is modeled as

$$y_i = m_i + \epsilon_i \quad \text{for} \quad i = 1, \cdots, N, \tag{3}$$

where $\mathbf{m} = (m_1, ..., m_N)^\top$ is the piecewise constant mean vector. The fused Lasso aims at striking a trade-off between total variation, $\sum_{i=2}^N |m_i - m_{i-1}|$, and the residual sum of squares (RSS), $\sum_{i=1}^N (y_i - m_i)^2$. By combining these two cost functions, the fused Lasso estimator is obtained by solving the convex optimization problem

$$\min_{\mathbf{m} \in \mathbb{R}^N} \quad \frac{1}{2} \sum_{i=1}^N (y_i - m_i)^2 + \lambda \sum_{i=2}^N |m_i - m_{i-1}|, \tag{4}$$

where $\lambda > 0$ is a regularization parameter that controls the balance between the fit and model complexity. After a transformation $\beta_i = m_{i+1} - m_i$ for $i = 1, ..., N-1$, we can rewrite (4) as a standard Lasso problem

$$(5) \qquad \min_{\boldsymbol{\beta} \in \mathbb{R}^{N-1}} \quad \frac{1}{2}\|\mathbf{y} - m_1\mathbf{1} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1,$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_{N-1})^\top$, $\mathbf{1} = (1, ..., 1)^\top \in \mathbb{R}^N$, $\mathbf{X}$ is an $N \times (N-1)$ matrix with $X_{ij} = 1$ when $i > j$ and $X_{ij} = 0$ otherwise.

Owrang et al. [28] proposed a variant of the fused Lasso called normalized fused Lasso, which replaces $\mathbf{X}$ by a design matrix with standardized columns. Let $\widetilde{\mathbf{X}}$ be the standardized version of $\mathbf{X}$, and $\tilde{\mathbf{y}}$ be the centered vector of $\mathbf{y}$. The normalized fused Lasso is equivalent to the following Lasso optimization with abuse of notation of $\boldsymbol{\beta}$.

$$(6) \qquad \min_{\boldsymbol{\beta} \in \mathbb{R}^{N-1}} \quad \frac{1}{2}\|\tilde{\mathbf{y}} - \widetilde{\mathbf{X}}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1.$$

The main difference between the fused Lasso (5) and the normalized fused Lasso (6) is that they use different relative penalty levels among the features (i.e., columns of the design matrix). Intuitively (6) is better as it treats all the features fairly by standardizing the columns of the design matrix. It was reported in [28] that the normalized fused Lasso improves the fused Lasso for change-point detection. The computation for (5) can be faster if making use of the special design matrix $\mathbf{X}$. There is little difference in computation time if one uses R package `glmnet` to calculate (5) and (6).

Our framework, the SCHACE method, incorporates the B-splines and the normalized fused Lasso in a unified penalized regression scheme, where the normalized fused Lasso is used to detect the discontinuities of the mean function, and B-spline allows the estimation of the nonparametric regression model. To elaborate, we combine the normalized design matrix $\widetilde{\mathbf{X}}$ with the B-spline basis matrix $\mathbf{Z} \in \mathbb{R}^{N \times df}$, where $df = s + n$ is the degrees of freedom of the B-spline basis matrix. Consider the optimization problem:

$$(7) \qquad \min_{\boldsymbol{\beta} \in \mathbb{R}^{N-1}, \boldsymbol{\gamma} \in \mathbb{R}^{df}} \quad \frac{1}{2}\|\tilde{\mathbf{y}} - \widetilde{\mathbf{X}}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1.$$

The optimization (7) leads to a sparse solution for $\boldsymbol{\beta}$, which gives change point locations. For example, for a solution $\hat{\boldsymbol{\beta}}$ and an index i with $\hat{\beta}_i \neq 0$, the location i/N is an estimated jump point. The B-spline part is used to estimate the continuous component of $f(\cdot)$ so we do not apply penalty on $\boldsymbol{\gamma}$ in estimation.

Although (7) can identify change point locations well, it is not ideal for estimating the mean function $f(\cdot)$ due to the over-shrinkage of the Lasso estimator. Therefore, we suggest a refit by ordinary least squares after variable selection, which follows the idea of the relaxed Lasso [23]. That is, we regress $\mathbf{y}$ on $\mathbf{Z}$ and the submatrix of $\mathbf{X}$ corresponding to the nonzero component of $\hat{\boldsymbol{\beta}}$ with an intercept. Note that it is equivalent but more convenient to use $\mathbf{X}$ instead of $\widetilde{\mathbf{X}}$ in this refit step because the coefficients from ordinary least squares directly give the jump magnitudes at each jump location. The refit step without penalization can reduce the shrinkage of the Lasso estimation and give better estimations of the jump sizes. Figure 1 illustrates the improvement of the curve estimation after the refit step (blue curve) over the original Lasso solution (green curve) from (7) using one of the real data examples presented in section 3.4. Both blue and green curves identify two jump points in 2008 and one in 2015. But the jump magnitudes increase from 2.745, -0.005, and 2.7888 to 26.087, -17.360, and 16.989, respectively, after the refit step. The two curves are similar in most regions, but the blue curve is more faithful to the data in the region near the jump points.

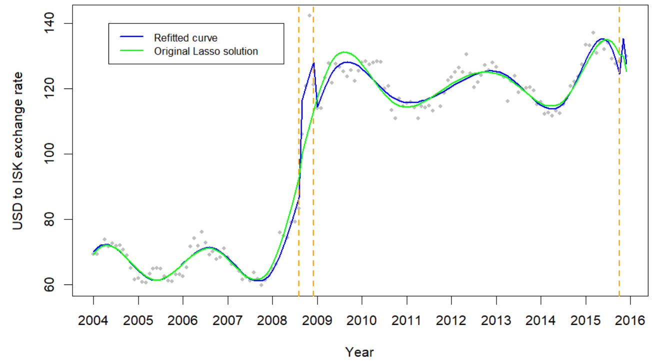

Figure 1. Plot of the monthly average of USD to ISK exchange rate from January 2004 to December 2015 (gray dots), the estimated curve by the penalized regression (7) (green curve), the refitted regression curve (blue curve), and the detected change points (orange dashed lines).

2.3 Choices of Tuning Parameters

Our procedure involves two tuning parameters, the number of quantile interior knots s for B-splines of order 4 and the regularization parameter λ. We consider a range from 0 to 12 for s. A λ sequence between 0 and $\lambda_{\max}$ is given by R package `glmnet` automatically once the length of the sequence is specified. We fix the length of the sequence as 100 in all numerical experiments. The choices of tuning parameters are critical to the performance of the SCHACE. We suggest two tuning procedures, the extended Bayesian Information Criterion (eBIC) proposed in [3] and cross-validation (CV) for choosing parameters.

The eBIC is one of the popular information criteria which determines the tuning parameter λ and s by minimizing the score

$$\text{eBIC}(\lambda, s) \quad = N \cdot \log(\text{RSS}(\lambda, s)) + p \cdot \log(N)$$

$$+ 4\nu \cdot p \cdot \log(s + n + N - 1),$$

where p is the model size (the number of nonzero components of the estimated regression coefficients) and $\text{RSS}(\lambda, s)$ denotes the residual sum of squares with the refitted least squares coefficients. The parameter $\nu \in [0, 1]$ controls the model complexity. We take $\nu = 0.5$ as suggested by Foygel and Drton [9].

Note that special care is needed when conducting a CV procedure for sequential data in the change-point analysis. For K-fold CV, we divide the index set $\{1, ..., N\}$ into K subsets based on the reminders modulo K. For example, with $K = 3$, the three subsets will be $\{1, 4, ...\}$, $\{2, 5, ...\}$ and $\{3, 6, ...\}$. This is in line with the order-preserved splitting principle proposed in [42]. Moreover, the commonly used L_2-error is inferior in calculating prediction error for data points around jump positions because only one of the folds contains a particular jump location. Therefore, we suggest the trimmed L_2-error for CV, which can ignore the big errors near jump points and gives a better choice of tuning parameters.

3. NUMERICAL STUDIES

3.1 Experiment Setup

In this section, we conduct synthetic experiments and real-world data analysis to assess the performance of the proposed SCHACE method in jump detection and curve fitting. For tuning parameter selection, we consider two procedures, the eBIC and 3-fold CV with trimmed L_2-error, which are labeled by SCHACE-eBIC and SCHACE-CV in the tables, respectively. Besides the SCHACE, we include two recent algorithms, the MultiDegree Spline Estimator (MDSE) [18] and Piecewise Constant plus Smooth regression estimator (PCpluS) [29] for comparison. For MDSE, we choose the quadratic programming variant MDSE-QP with degree-3 truncated power basis and 20 interior knots. For PCpluS, the tuning parameters are determined by 5-fold cross-validation, following the suggestion in the paper.

3.2 Evaluation Metrics

We compare the performance of different methods based on four evaluation metrics. The first is the number of detected change points. The second one is the number of true positives, which are informative points that are close enough to true jump positions. The definition of the informative point was proposed by [19] to evaluate different change-point detection procedures. Roughly speaking, an informative point corresponding to a true change-point location, say τ_i is the closest estimated change-point location to τ_i. In this paper, we regard an information point as a true positive if and only if there is a true change point within a distance $h = 0.01$. The third measure is the false discovery rate (FDR) which is calculated by the ratio between the number of false positives and the total number of detected change points, $\hat{J}$. That is

$$\text{FDR} = \frac{\#\text{False Positives}}{\hat{J} \vee 1}.$$

Last but not least, the mean squared error (MSE) is reported to reflect the accuracy of the fitted regression curve in synthetic experiments.

3.3 Synthetic data

We consider six mean functions with discontinuities including four well-known benchmark functions **burt** and **cosine** from Abramovich et al. [1], **heavisine** from Donoho and Johnstone [5], **blip** from Marron et al. [21], and two additional examples **cubic** and **step**. Figure 2 provides an illustration of these functions. In the analytic formulas below, all functions are defined on the domain $[0, 1]$, $1_{(a,b]}(x)$ represents an indicator function taking value 1 when $a < x \leq b$, $\text{sgn}(\cdot)$ is the standard sign function.

- **Example 1: burt**

$$f(x) = 20x \cos(16x^{1.2}) - 20 \cdot 1_{(0,0.5)}(x).$$

- **Example 2: cosine**

$$f(x) = \cos(5.5\pi x) - 4\text{sgn}(0.23 - x) - 2\text{sgn}(0.3 - x) - 1.75\text{sgn}(0.55 - x) + 3\text{sgn}(0.7 - x).$$

- **Example 3: heavisine**

$$f(x) = 4\sin(4\pi x) - \text{sgn}(x - 0.3) - \text{sgn}(0.72 - x).$$

- **Example 4: blip**

$$f(x) = (0.32 + 0.6x + 0.3e^{-100(x-0.3)^2})1_{[0,0.8]}(x) + (-0.28 + 0.6x + 0.3e^{-100(x-1.3)^2})1_{(0.8,1]}(x).$$

- **Example 5: cubic**

$$f(x) = 432x^3 - 540x^2 + 212.4x + 5 \cdot 1_{[0,\frac{1}{3}]}(x) - 8 \cdot 1_{(\frac{1}{3},\frac{2}{3}]}(x).$$

- **Example 6: step**

$$f(x) = 3 \cdot 1_{[0,\frac{1}{3}]}(x) + 5 \cdot 1_{(\frac{1}{3},\frac{2}{3}]}(x) + 1_{(\frac{2}{3},1]}(x).$$

The sample sizes of the synthetic data sets are $N = 256$, and all the design points are equally spaced in $[0, 1]$. In order to compare methods under different noise levels, we consider three levels of the signal-to-noise-ratio (snr), defined by $\text{snr} = \text{sd}(f)/\sigma$, where $\text{sd}(f)$ denotes the sample standard deviation of the noiseless sequence $\{f(i/N)\}_{i=1}^N$. For each example, we run 100 independent replicates with $\text{snr} = 4, 6, 8$, and report average evaluation metrics: the number of detected change points (#CP), the number of true positives (#TP), the mean square error (MSE), and the false discovery rate (FDR). Tables 1-6 present the numerical results for Examples 1-6, respectively.

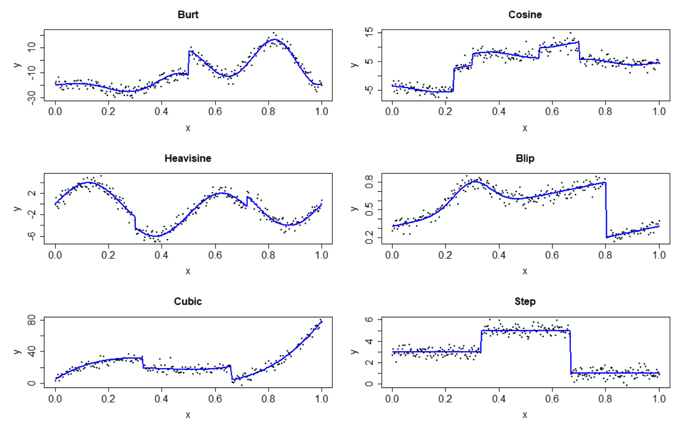

Figure 2. Plots of data sequences generated from the six mean functions with snr = 4. The true piecewise smooth functions are depicted by the blue curves and the jumps are marked by the vertical lines.

Table 1 presents the numerical results from Example 1 (**burt**). The proposed SCHACE-eBIC has the best performance among all the methods as it has the smallest MSE and highest accuracy in capturing the change-point location at $x = 0.5$. As the signal-to-ratio increases, all the methods perform better as expected. SCHACE-CV is the second best method although it tends to over-select change points. PCpluS is similar to SCHACE-CV in terms of change-point detection but produces a larger MSE in curve estimation. MDSE-QP performs worst in this example. It can estimate the number of change points reasonably well but fails to recover the change-point location accurately. The reason is that MDSE-QP can identify change points only among K prefixed knots but not in a data-adaptive way. $K = 20$ is used here, and unfortunately, none of those 20 equidistant interior knots is within the distance $h = 0.01$ of the true change point. This leads to #TP = 0 and a significantly larger MSE than other methods. We have tried MDSE-QP with $K = 50$ which produces more false positives and worse overall results. In short, the SCHACE outperforms the other two methods in both curve estimation and change-point detection.

The numerical results for Example 2 (**cosine**) are shown in Table 2. The true number of change points is 4. PCpluS and SCHACE-eBIC perform best in terms of low FDR. SCHACE-CV has the highest #TP among all the methods but tends to over-select change points. For curve estimation, SCHACE-CV and PCplus are the best, and SCHACE-eBIC is slightly worse due to slightly lower #TP. MDSE-QP has the worst MSE because it is not able to uncover change points accurately.

Tables 3-6 list the results of the rest of the examples. The overall patterns are similar to the previous examples although the performance of different methods varies. SCHACE-eBIC is the winner for change-point detection. It has constantly low FDR, and can perfectly identify change points when snr is high. When snr is low, it may miss some change points, which leads to a higher MSE than SCHACE-CV. SCHACE-CV usually selects a larger number of change points so FDR is higher. Nevertheless, it seldom misses any change points. In the low snr scenarios, it often outperforms other methods in terms of curve estimation. PCpluS is a strong competitor to our methods, performing reasonably well in most examples. But it is powerless in identifying change points for Example 5. The performance of MDSE-QP highly depends on the true locations of change points. Overall, it produces a higher number of false positives and the largest MSE in most scenarios.

Table 1. Numerical results for Example 1

Burt (snr = 4)	#CP	#TP	MSE	FDR
SCHACE-CV	1.75	0.99	0.7586	0.2032
SCHACE-eBIC	1.02	1	0.4321	0.01
PCpluS	1.79	0.98	1.06	0.3047
MDSE-QP	1.01	0	3.1672	1
Burt (snr = 6)	#CP	#TP	MSE	FDR
SCHACE-CV	1.86	1	0.3269	0.2189
SCHACE-eBIC	1	1	0.2244	0
PCpluS	2.01	1	0.5409	0.3497
MDSE-QP	1.02	0	2.9122	1
Burt (snr = 8)	#CP	#TP	MSE	FDR
SCHACE-CV	2.03	1	0.2052	0.2432
SCHACE-eBIC	1	1	0.154	0
PCpluS	1.99	1	0.3296	0.3432
MDSE-QP	1.06	0	2.8222	1

Table 2. Numerical results for Example 2

Cosine (snr = 4)	#CP	#TP	MSE	FDR
SCHACE-CV	12.31	3.74	0.2663	0.6205
SCHACE-eBIC	4.31	2.93	0.4091	0.231
PCpluS	5.07	3.45	0.2412	0.2562
MDSE-QP	4.23	0	1.0302	1
Cosine (snr = 6)	#CP	#TP	MSE	FDR
SCHACE-CV	11.85	3.97	0.1015	0.5946
SCHACE-eBIC	5.63	3.68	0.1698	0.2852
PCpluS	5.4	3.69	0.1084	0.2487
MDSE-QP	1.49	0	0.8272	1
Cosine (snr = 8)	#CP	#TP	MSE	FDR
SCHACE-CV	9.71	3.99	0.052	0.4959
SCHACE-eBIC	5.5	3.97	0.0622	0.2454
PCpluS	5.16	3.79	0.0617	0.2028
MDSE-QP	1.44	0	0.7989	1

3.4 Real Data

In this section, we apply our method to three data sets, where the time series curves show interesting patterns of abrupt changes due to unusual historical events. Since SCHACE-eBIC outperforms SCHACE-CV in the simulated

Table 3. Numerical results for Example 3

Heavisine (snr = 4)	#CP	#TP	MSE	FDR
SCHACE-CV	4.99	1.44	0.0759	0.476
SCHACE-eBIC	1.73	1.15	0.1032	0.2467
PCpluS	1.4	0.72	0.0873	0.1382
MDSE-QP	2.25	0.61	0.1043	0.6827
Heavisine (snr = 6)	#CP	#TP	MSE	FDR
SCHACE-CV	4.37	1.92	0.0262	0.3301
SCHACE-eBIC	2.33	1.96	0.0351	0.111
PCpluS	2.05	1.37	0.0429	0.1412
MDSE-QP	2.08	0.86	0.0814	0.5942
Heavisine (snr = 8)	#CP	#TP	MSE	FDR
SCHACE-CV	3.88	1.99	0.013	0.2334
SCHACE-eBIC	2.24	2	0.0193	0.0693
PCpluS	2.27	1.65	0.0252	0.1474
MDSE-QP	2.36	0.98	0.0727	0.5668

Table 4. Numerical results for Example 4

Blip (snr = 4)	#CP	#TP	MSE	FDR
SCHACE-CV	4.52	1	0.0002	0.3259
SCHACE-eBIC	2.39	1	0.0005	0.3493
PCpluS	2.25	1	0.0002	0.3486
MDSE-QP	3.1	1	0.0021	0.3354
Blip (snr = 6)	#CP	#TP	MSE	FDR
SCHACE-CV	3.93	1	0.0001	0.2155
SCHACE-eBIC	1.12	1	0.0001	0.0433
PCpluS	2.24	1	0.0001	0.3376
MDSE-QP	2.02	1	0.0019	0.3592
Blip (snr = 8)	#CP	#TP	MSE	FDR
SCHACE-CV	2.81	1	0.0001	0.154
SCHACE-eBIC	1.17	1	0.0001	0.0817
PCpluS	2.19	1	7.54e-5	0.3344
MDSE-QP	2.26	1	0.0019	0.4358

Table 5. Numerical results for Example 5

Cubic (snr = 4)	#CP	#TP	MSE	FDR
SCHACE-CV	4.04	1.2	2.5689	0.5641
SCHACE-eBIC	1.54	0.99	2.6071	0.3025
PCpluS	1.38	0.04	4.1853	0.5392
MDSE-QP	13.83	1.42	4.9809	0.8903
Cubic (snr = 6)	#CP	#TP	MSE	FDR
SCHACE-CV	4.14	1.73	0.9396	0.4384
SCHACE-eBIC	2.21	1.72	0.6255	0.21
PCpluS	2.05	0.03	2.1601	0.749
MDSE-QP	15.09	1.64	3.3408	0.8492
Cubic (snr = 8)	#CP	#TP	MSE	FDR
SCHACE-CV	3.52	1.94	0.4194	0.2938
SCHACE-eBIC	2.26	1.94	0.218	0.1067
PCpluS	2.01	0.01	1.3045	0.8475
MDSE-QP	11.93	1.8	2.2363	0.6883

Table 6. Numerical results for Example 6

Step (snr = 4)	#CP	#TP	MSE	FDR
SCHACE-CV	3.23	2	0.0088	0.2352
SCHACE-eBIC	2.09	2	0.0044	0.03
PCpluS	2.7	1.99	0.0076	0.1633
MDSE-QP	10.29	2	0.088	0.6922
Step (snr = 6)	#CP	#TP	MSE	FDR
SCHACE-CV	2.9	2	0.0037	0.1808
SCHACE-eBIC	2.01	2	0.0018	0.0033
PCpluS	2.94	1.99	0.0034	0.1969
MDSE-QP	9.32	2	0.0833	0.6459
Step (snr = 8)	#CP	#TP	MSE	FDR
SCHACE-CV	2.98	2	0.0021	0.1962
SCHACE-eBIC	2.01	2	0.001	0.0033
PCpluS	3.01	1.99	0.002	0.1932
MDSE-QP	9.17	2	0.083	0.6205

examples, we only present SCHACE-eBIC and compare it with PcPluS and MDSE-QP.

- **Revenue Passenger Miles for U.S. Domestic and International Air Carriers (2000-2022).** The revenue passenger mile (RPM) is a transportation industry measurement of airline traffic and a vital indicator of the volume of air passenger transportation. It is calculated by the number of paying passengers multiplied by the distance traveled. The data set is downloaded from web site https://fred.stlouisfed.org/series/RPM, which includes monthly RPMs (in thousands; not seasonally-adjusted) for U.S. domestic and international airlines from 2000 to 2022. Figure 3 presents the original data as well as curve fitting by three methods. We can see from Figure 3(a) that the RPM decreased greatly in March 2020, which is likely an outcome of restricted global air travel due to the COVID-19 pandemic. It starts to recover from May 2021. Plots 3(b)-(d) provide the fitted curves (blue lines) of the three methods and detected change points (orange dashed lines). SCHACE-eBIC detects two change points in RPM around March 2020 and May 2021. Interestingly, it seems that it is immune to the obvious seasonal trends and gives an excellent estimation of the overall trend. PCpluS doesn't identify any change point. The fitted curve by PCpluS demonstrates the periodicity of passenger transportation. MDSE-QP fits the curve well and detects two additional change points in 2004 and 2005.

- **Monthly average exchange rate change of the ISK per USD (2004-2015).** This data set contains the monthly average of the USD to ISK exchange rate between 2004-2015. It has been analyzed in [18]. Plot (a) in Figure 4 depicts the original data, which clearly shows the sharp depreciation of the Icelandic kr*ó*na occurred during the 2008–2011 Icelandic financial crisis.

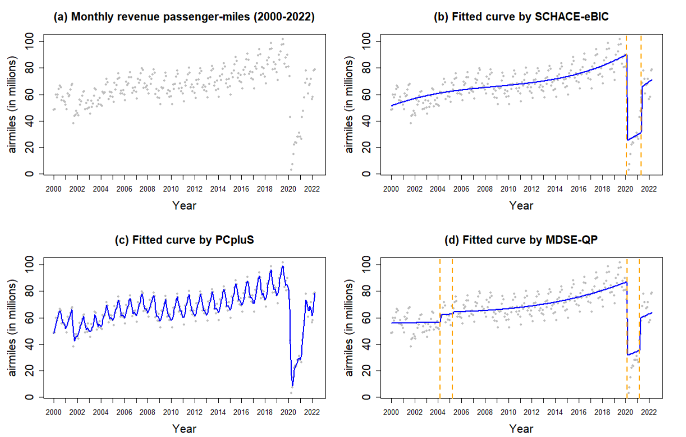

Figure 3. The trends in passenger transportation from 2000 to 2022.

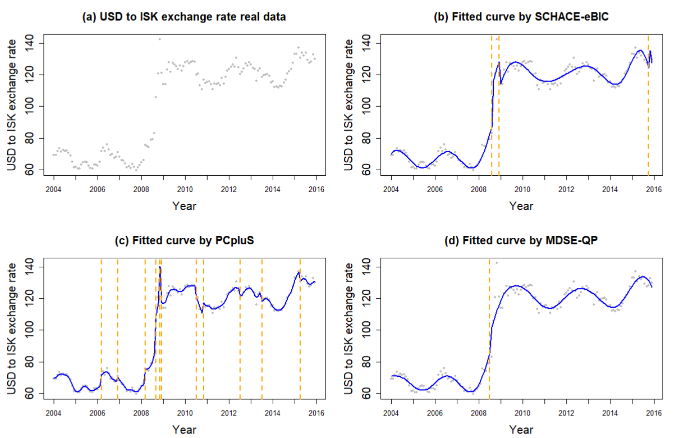

Figure 4. USD to ISK exchange rate data.

In plot (b), we see that SCHACE-eBIC detects three jumps, two of which are in 2008. Plot (c) shows that PCpluS claims a much larger number of jumps. Finally, as shown in plot (d), MDSE-QP detects one jump in 2008. In summary, SCHACE-eBIC and MDSE-QP produce similar results and identify jumps that are related to the Icelandic financial crisis, while PCpluS detects many change points that are not easy to interpret.

- **The GDP of Iran (1960-2020).** The GDP per capita serves as a crucial benchmark for assessing the progress and well-being of a nation's economy. It is calculated by the gross domestic product divided by the mid-year population. The data set has been demonstrated in [36] and can be downloaded from the authors' GitHub site. The three methods, SCHACE-eBIC, PCpluS, and MDSE-QP, find one, five, and four change points, respectively. In particular, the change point identified by SCHACE-eBIC occurred in 1978, which corresponds

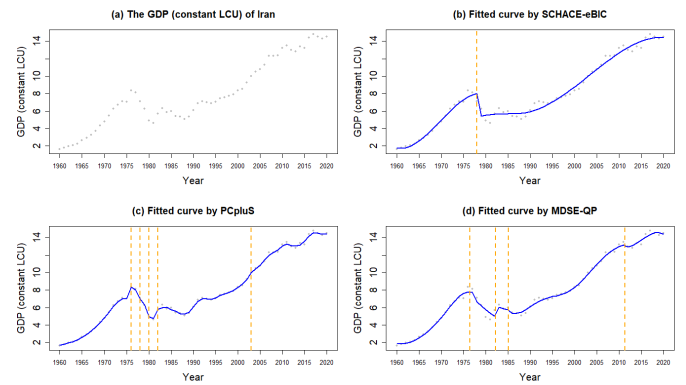

Figure 5. The GDP (constant LCU) of Iran.

to the 1978-1979 Iranian Revolution when the GDP dropped significantly.

4. DISCUSSION

In this paper, we propose a new approach called SCHACE for simultaneously detecting jumps in data and estimating the nonlinear trend between jumps. It unifies the operations of change-point detection and curve estimation into one regularization framework and avoids multi-step or iterative processes. The implementation is straightforward, and more importantly, it can be generalized to solve similar problems for multiple sequences. We develop two tuning procedures, the eBIC and CV, to select smoothing parameters. Based on our numerical experiments, we would recommend the variant SCHACE-eBIC in practice for its lower FDR and better interpretability. We developed an R package that is available on the web https://github.com/ZhaoyingLuLuLu/SCHACE. Besides the suggested variants in this paper, the users can choose the number of folds and the loss function (among squared error, trimmed squared error, or absolute error loss) in CV based on their need.

REFERENCES

[1] Abramovich, F., Antoniadis, A. and Pensky, M. (2007). Estimation of piecewise-smooth functions by amalgamated bridge regression splines. *Sankhyā: The Indian Journal of Statistics* 1–27.

[2] Antoniadis, A., Bigot, J. and Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of statistical software* **6** 1–83.

[3] Chen, J. and Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95** 759–771.

[4] Denison, D., Mallick, B. and Smith, A. (1998). Automatic Bayesian curve fitting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **60** 333–350.

[5] Donoho, D. L. and Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *biometrika* **81** 425–455.

[6] Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. CRC press.

[7] Fan, J. and Gijbels, I. (2018). *Local polynomial modelling and its applications*. Routledge.

[8] Fearnhead, P. (2005). Exact Bayesian curve fitting and signal segmentation. *IEEE Transactions on Signal Processing* **53** 2160–2166.

[9] Foygel, R. and Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. *Advances in neural information processing systems* **23**.

[10] Frick, K., Munk, A. and Sieling, H. (2014). Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76** 495–580.

[11] Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics* **42** 2243–2281.

[12] Gijbels, I., Lambert, A. and Qiu, P. (2007). Jump-preserving regression and smoothing using local linear fitting: a compromise. *Annals of the Institute of Statistical Mathematics* **59** 235–272.

[13] Hall, P. and Titterington, D. (1992). Edge-preserving and peak-preserving smoothing. *Technometrics* **34** 429–440.

[14] Hao, N., Niu, Y. S. and Zhang, H. (2013). Multiple change-point detection via a screening and ranking algorithm. *Statistica Sinica* **23** 1553.

[15] Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple Change-Point Estimation With a Total Variation Penalty. *Journal of the American Statistical Association* **105** 1480 - 1493.

[16] Huang, T., Wu, B., Lizardi, P. and Zhao, H. (2005). Detection of DNA copy number alterations using penalized least squares regression. *Bioinformatics* **21** 3811–3817.

[17] Koo, J.-Y. (1997). Spline estimation of discontinuous regression functions. *Journal of Computational and Graphical Statistics* **6** 266–284.

[18] Lee, E.-J. and Jhong, J.-H. (2021). Change Point Detection Using Penalized Multidegree Splines. *Axioms* **10** 331.

[19] Li, H., Munk, A. and Sieling, H. (2016). FDR-control in multiscale change-point segmentation. *Electronic Journal of Statistics* **10** 918–959.

[20] Liu, G.-X., Wang, M.-M., Du, X.-L., Lin, J.-G. and Gao, Q.-B. (2018). Jump-detection and curve estimation methods for discontinuous regression functions based on the piecewise B-spline function. *Communications in Statistics-Theory and Methods* **47** 5729–5749.

[21] Marron, J. S., Adak, S., Johnstone, I., Neumann, M. and Patil, P. (1998). Exact risk analysis of wavelet regression. *Journal of Computational and Graphical Statistics* **7** 278–309.

[22] McDonald, J. A. and Owen, A. B. (1986). Smoothing with split linear fits. *Technometrics* **28** 195–208.

[23] Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics & Data Analysis* **52** 374–393.

[24] Muller, H.-G. (1992). Change-points in nonparametric regression analysis. *The Annals of Statistics* **20** 737–761.

[25] Niu, Y. S., Hao, N. and Zhang, H. (2016). Multiple change-point detection: a selective overview. *Statistical Science* 611–623.

[26] Niu, Y. S. and Zhang, H. (2012). The screening and ranking algorithm to detect DNA copy number variations. *The annals of applied statistics* **6** 1306.

[27] Olshen, A. B., Venkatraman, E. S., Lucito, R. and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5** 557–572.

[28] Owrang, A., Malek-Mohammadi, M., Proutiere, A. and Jansson, M. (2017). Consistent change point detection for piecewise constant signals with normalized fused LASSO. *IEEE signal processing letters* **24** 799–803.

[29] Pein, F. (2021). Change-point regression with a smooth additive disturbance. *arXiv preprint arXiv:2112.03878*.

[30] Qiu, P. (2003). A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation. *Journal of Nonparametric Statistics* **15** 437–453.

[31] Qiu, P. (2005). *Image processing and jump regression analysis*. John Wiley & Sons.

[32] Qiu, P., Asano, C. and Li, X. (1991). Estimation of jump regression function. *Bulletin of Informatics and Cybernetics* **24** 197–212.

[33] Qiu, P. and Yandell, B. (1998). Local polynomial jump-detection algorithm in nonparametric regression. *Technometrics* **40** 141–152.

[34] Shiau, J.-J., Wahba, G. and Johnson, D. R. (1986). Partial spline models for the inclusion of tropopause and frontal boundary information in otherwise smooth two-and three-dimensional objective analysis. *Journal of Atmospheric and Oceanic Technology* **3** 714–725.

[35] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67** 91–108.

[36] Van den Burg, G. J. and Williams, C. K. (2020). An evaluation of change point detection algorithms. *arXiv preprint arXiv:2003.06222*.

[37] Vostrikova, L. Y. (1981). Detecting "disorder" in multidimensional random processes. In *Doklady akademii nauk* **259** 270–274. Russian Academy of Sciences.

[38] Wahba, G. (1990). *Spline models for observational data*. SIAM.

[39] Xia, Z. and Qiu, P. (2015). Jump information criterion for statistical inference in estimating discontinuous curves. *Biometrika* **102** 397–408.

[40] Yang, Y. and Song, Q. (2014). Jump detection in time series nonparametric regression models: a polynomial spline approach. *Annals of the Institute of Statistical Mathematics* **66** 325–344.

[41] Yin, Y. (1988). Detection of the number, locations and magnitudes of jumps. *Communications in Statistics. Stochastic Models* **4** 445–455.

[42] Zou, C., Wang, G. and Li, R. (2020). Consistent selection of the number of change-points via sample-splitting. *Annals of statistics* **48** 413.

Zhaoying Lu
Department of Mathematics
The University of Arizona
Tucson, AZ 85721–0089
USA
E-mail address: zhaoyinglu@math.arizona.edu

Ning Hao
Department of Mathematics
The University of Arizona
Tucson, AZ 85721–0089
USA
E-mail address: nhao@math.arizona.edu

Hao Helen Zhang
Department of Mathematics
The University of Arizona
Tucson, AZ 85721–0089
USA
E-mail address: hzhang@math.arizona.edu