

# Multiple change-point estimation in spectral representation

Jaehee Kim<sup>1,a</sup>

<sup>a</sup>Duksung Women's University

---

## Abstract

We discuss multiple change-point estimation as edge detection in piecewise smooth functions with finitely many jump discontinuities. In this paper we propose change-point estimators using concentration kernels with Fourier coefficients. The change-points can be located via the signal based on Fourier transformation system. This method yields location and amplitude of the change-points with refinement via concentration kernels. We prove that, in an appropriate asymptotic framework, this method provides consistent estimators of change-points with an almost optimal rate. In a simulation study the proposed change-point estimators are compared and discussed. Applications of the proposed methods are provided with Nile flow data and daily won-dollar exchange rate data.

**Keywords:** change-point model, concentration kernel, edge detection, Fourier series, piecewise smoothness, sample Fourier coefficients, spectral expansion

---

## 1. Introduction

We are concerned with the change-point problem as discontinuity detection or jump detection. If one acquires a signal from a piece of equipment and its changes are signaled by discontinuities, it is important to locate change-points in the signal. Not only the location but also the amplitude of jump discontinuities of piecewise smooth data is of interest for many scientific applications, for example, functional magnetic resonance imaging (fMRI) segmentation. We are interested in detecting change-points in piecewise smooth functions realized by spectral representation. We discuss a framework for discovering change-points in piecewise smooth functions with finitely many jump discontinuities. Fields of application include econometrics, biostatistics, and signal processing.

Consider observations  $Y_1, \dots, Y_n$  with finite fourth moment and obeying the model

$$Y_i = f(x_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

where  $x_i = i/n$ ,  $i = 1, \dots, n$ . The errors  $\epsilon_1, \dots, \epsilon_n$  are independent and each distributed symmetric continuous distribution with mean 0 and variance  $\sigma^2$ . For example,  $N(0, \sigma^2)$ . The function  $f$  is continuous on  $(0, 1)$  except at some unknown change-points or discontinuity points  $\tau_1, \dots, \tau_K \in (0, 1)$ , at which it has left and right limits. For definiteness, we suppose that  $\tau_l$  is an event time, i.e.,  $\tau_l = x_i$

---

This research was supported by Mid-career Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2018R1A2B6001664). It is also supported by Basic Research Lab (No. 2021R1A4A5028907). and Basic Science Research (No. 2021R1F1A1054968).

<sup>1</sup> Department of Statistics, Duksung Women's University, Samyang-ro 144-gil 33, Dobong-gu, Seoul 01369, Korea.  
E-mail: [jaehee@duksung.ac.kr](mailto:jaehee@duksung.ac.kr)

Published 31 January 2022 / journal homepage: <http://csam.or.kr>

© 2022 The Korean Statistical Society, and Korean International Statistical Society. All rights reserved.

for some  $i$ . If  $f$  is assumed to be constant except at the change-points, then the model (1.1) reduces to the mean shift model. If  $f$  is assumed to be linear with discontinuity at the change-points, then the model (1.1) reduces to the piecewise linear model.

Considerable attention has also been devoted to testing the hypothesis of continuity against the alternative of one or more change-points. For dependent data Kim and Hart (1998) proposed a test for change using sample Fourier coefficients. When the null hypothesis of no change is rejected, then estimating the change-point(s) is of interest.

Nonparametric change-point estimation research has been done since precise knowledge of the discontinuous locations is essential for underlying functions. McDonald and Owen (1986) and Hall and Titterton (1992) investigated change-point estimation based on three smoothed estimates of the function at each point. Müller (1992) proposed estimators for the location and size of a discontinuity in a smooth regression model, including the case of discontinuities in derivatives. Wang (1995) proposed to detect jumps and sharp cusps in a function using wavelet transformation. Loader (1996) suggested a change-point estimator based on one-sided local polynomials. Kim and Hart (2011) used local Fourier coefficients for one change-point detection. Change-point algorithms either estimate all change-points concurrently or hierarchically. Concurrent methods generally optimize a single objective function.

For example, given that there are  $k$  change-points, Hawkins (2001) estimated change-point locations by maximizing a likelihood function. Lavielle and Teyssi re (2006) accomplished the same task by minimizing a loss function. Guralnik and Srivastava (1999) used sequential methods to generally estimate change-points one at a time.

The issue of edge or jump detection with Fourier series was treated in some literature. Kvernadze (1998) identified the jumps of a function with generalized bounded variation. Gelb and Tadmor (1999, 2000, 2002) discussed a general framework for recovering edges in piecewise smooth functions with finitely many jump discontinuities. Engelberg (2008) discussed techniques for using a function's Fourier coefficients to determine the locations and sizes of the jump discontinuities of the function since the existence of the discontinuity is indeed encoded in the coefficients. From operational calculus Fleiss *et al.* (2010) proposed a new approach for change-point detection based on algebraic theory.

Our main justification for considering concentration kernels and Hilbert transformation is that Fourier series methods have played a fundamental role in functional approach in statistics. The distribution of change-points is empirically determined, and its theoretical properties are examined as well. Our goal is to detect multiple change-points as discontinuity points without imposing any parametric structure. Virtually any salient measure at each point would ensure detection of each change-point.

The proposed method can detect any change including discontinuities, and functional changes within an independent or dependent error sequence. Estimation is performed in a manner that simultaneously identifies both the number and locations of change-points even though there needs to be some cutoff values.

The rest of the paper is organized as follows. In Section 2, the Hilbert transformation and Fourier series are defined and their properties are investigated, and consistency of the resulting change-point estimator is established. Proofs of all results are given in Section 3. More precisely, we provide rates of convergence for the change-point estimators and show their almost optimal rates of convergence with the underlying discontinuous functions. A simulation study illustrating good finite sample properties of the new procedure is summarized in Section 4. In Section 5 two examples with Nile River data and won-dollar exchange rate data are given as applications. Concluding remarks are made in Section 6.

## 2. Change-point estimators based on Fourier series

A function is called piecewise continuous if it has only a finite number of discontinuities in any interval, and its left and right limits exist at each discontinuity point. A function is called piecewise differentiable if the function and its derivatives are both piecewise continuous. Piecewise differentiable functions are composed of differentiable functions that have been pasted together like constant values and waves.

We consider the problem of detecting change-points in a function. We use techniques for using a function's Fourier coefficients to determine the locations and size of the jump discontinuities of the function. Our approach is based on localization using appropriate concentration kernels and separation of scales by nonlinear enhancement using Fourier projection. Tolstov (1962) provides theory about Fourier series and Fourier coefficients.

We assume that the underlying function  $f$  has the following Fourier series representation based on the orthogonal system  $\{b_j(x) : j = 0, 1, \dots\}$ ,

$$f(x) = \phi_0 + \sum_{j=1}^{\infty} \phi_j b_j(x), \quad x \in (0, 1), \quad (2.1)$$

where  $\phi_j = \int_0^1 f(x) b_j(x) dx$ ,  $j = 0, 1, \dots$ , are the Fourier coefficients of  $f$ .

Consider the trigonometric bases  $\{b_j(x) = \sqrt{2} \cos \pi j x\}$ , The sample Fourier coefficients are defined by,

$$\hat{\phi}_j = \frac{1}{n} \sum_{i=1}^n Y_i \sqrt{2} \cos \pi j x_i, \quad j = 0, 1, \dots, n-1.$$

Let's consider the convergence of the Fourier series. Tolstov (1962) provides a theory about Fourier series. In many cases, it is possible to reconstruct a function from its Fourier coefficients. Consider a piecewise differentiable periodic function,  $f(x)$ . At discontinuity points, the convergence is to the mean of values to which the function tends from the left and the right of the discontinuity. Therefore, the convergence of the Fourier series to the function is pointwise whenever the function is continuous and is to the average values of the function at the jumps in the function's value.

A little change in the values of a function even on a small interval can change its Fourier coefficients considerably. However, an absolutely integrable function  $f(x)$  has a right-hand and a left-hand derivative in a neighborhood of  $x$ , then, its Fourier series remains convergent no matter what the value of  $f(x)$ . This is known as the localization principle of the Fourier series.

For square integrable functions Parseval's equation states that

$$\sum_j \hat{\phi}_j^2 = \int_0^1 |f(t)|^2 dt.$$

It means that Fourier coefficients are square summable if a function is square integrable. Suppose that the Fourier coefficients of  $f(t)$  are absolutely summable

$$\sum_j \hat{\phi}_j^2 < \infty,$$

then, it establishes the uniform convergence of the Fourier series.

At  $t_i = i/n$ ,  $1 \leq i \leq n$ . We define the sample Fourier coefficients as

$$\hat{\phi}_j = \frac{1}{n} \sum_{1 \leq i \leq n} Y_i \sqrt{2} \cos \pi j t_i, \quad 0 \leq j \leq n-1.$$

The Fourier series estimator is defined as

$$\hat{f}_J(x) = \hat{\phi}_0 + \sum_{j=1}^J \hat{\phi}_j \sqrt{2} \cos \pi j x, \quad (2.2)$$

where  $J$  is the number of Fourier coefficients. At the change-point  $\tau$ , the change size (or amplitude) is measured by

$$\Delta = \Delta_\tau = f(\tau+) - f(\tau-).$$

We need to split a piecewise differentiable function into its continuous and discontinuous parts. One way to perform this decomposition is the followings. Suppose that  $f(t)$  has  $m$  jumps at the locations  $\tau_1, \dots, \tau_m$  with the heights  $h_1, \dots, h_m$ , respectively. It can be expressed with

$$w(t) = f(t) - \sum_{i=1}^m h_i K(t - \tau_i), \quad (2.3)$$

where  $K$  is the kernel function and  $\tau_i$  is change-point. Here,  $w(t)$  is continuous and piecewise differentiable with no jumps. Then, we have

$$f(t) = w(t) + \sum_{i=1}^m h_i K(t - \tau_i) \quad (2.4)$$

where  $w(t)$  is continuous and piecewise differentiable. The sum  $\sum_{i=1}^m h_i k(t - \tau_i)$  is discontinuous and piecewise differentiable. By the linearity of the Fourier coefficients, the Fourier coefficients of  $f(t)$ , denoted by  $\phi_j$ , are

$$\phi_j = c_j + d_j. \quad (2.5)$$

The Fourier coefficients of  $w(t)$ , which is denoted by  $c_j$  are absolutely summable.

Using the linearity of the Fourier coefficients, the Fourier coefficients of the sum, denoted by  $d_j$  are

$$d_j = \begin{cases} \sum_{u=1}^m h_u \psi_j, & \text{at } \tau_i \\ 0, & \text{otherwise} \end{cases} \quad (2.6)$$

where

$$\psi_j = \int K(t - \tau_i) \sqrt{2} \cos \pi j t dt. \quad (2.7)$$

And its estimator is

$$\hat{\psi}_j = \frac{1}{n} \sum_{i=1}^n K(t - \tau_i) \sqrt{2} \cos \pi j x_i,$$

and these coefficients are not summable.

The piecewise smooth function  $f$  can be recovered from its spectral coefficients and retain the spectral accuracy. Edges or discontinuity points are separated by pieces of smoothness. For piecewise differentiable functions the convergence of the Fourier series to the function is pointwise wherever the function is continuous and is to the average value of the function at the jumps in the function's value.

In order to determine where the edges of the data are, we use the Hilbert transform of the Fourier coefficients. Because it changes the Fourier coefficients as little as possible, but causes the partial sums of the Fourier series of a discontinuous function to grow at the discontinuities but not elsewhere.

At the edges in piecewise smooth functions with finitely many jump discontinuities, where

$$[f](x) \equiv f(x+) - f(x-) \neq 0. \quad (2.8)$$

Consider a concentration kernel  $K_\delta$  that is properly scaled and admissible. It recovers both the locations and the amplitudes of the jumps (Gelb and Tadmor, 1999). If a discontinuity is characterized by a ‘‘phenomenon’’, the existence of the discontinuity is indeed encoded in the coefficients. The question is how to effectively decode the discontinuity. To concentrate the function about the discontinuity, a different spectral representation can be used.

In practice one encounters the function  $f(x)$  with finitely many jump discontinuities and the differential of  $f(x)$  on each side of the discontinuity to have bounded variation. Such kernels are characterized by

$$K_\delta * f(x) \rightarrow [f](x) \quad \text{as} \quad \delta \rightarrow 0. \quad (2.9)$$

Thus, the support of  $K_\delta * f(x)$  tends to concentrate near the edges of  $f(x)$ . Also

$$\int t K_\delta(t) d_j = \begin{cases} \sum_{u=1}^m h_u \sum_l \hat{\psi}_l \cos \pi l x, \\ 0. \end{cases} \quad (2.10)$$

The derivatives provide information about discontinuities and smoothness. From equation (2.1), consider a derivative such as

$$\begin{aligned} f'(x) &= -\sqrt{2\pi} \sum_{j=1}^{\infty} j \phi_j \sin \pi j x \\ &= -\sqrt{2\pi} \sum_{j=1}^{\infty} g(j) \phi_j \sin \pi j x, \end{aligned} \quad (2.11)$$

where  $g(\cdot)$  is the admissible concentration factor.

To detect such edges, consider concentration kernels  $K_\delta(\cdot)$ , depending on the small scale  $\delta$  such as

$$K_\delta(\cdot) * f(x) = [f](x) + O(\delta). \quad (2.12)$$

Thus,  $K_\delta$  tends to concentrate near the singular support of  $f$ . It is shown that the concentration kernel is properly scaled and admissible to recover both the locations and the amplitudes of the jumps.

A real function  $f(t)$  and its Hilbert transform  $H[f(t)]$  are related to each other in such a way that they create a so-called strong analytic signal together. The strong analytic signal can be written with an amplitude and a phase where the derivative of the phase can be identified as the instantaneous frequency. It is not hard to see that a function and its Hilbert transform are also orthogonal. This

orthogonality is not always realized in applications because of truncations in numerical calculations. However, a function and its Hilbert transform have the same energy and therefore the energy can be used to measure the calculation accuracy of the approximated Hilbert transform.

If we create a Fourier series of a function  $f(t)$  and change the sine functions to cosine functions and vice versa, we get the Hilbert transform of  $f(t)$ . We try to suggest concentration kernels to detect edges satisfying (2.12). The key to locating the discontinuities lies in the relationship between the conjugate Fourier partial sums and the jump discontinuities.

Given first  $n$  Fourier modes, the concentration kernel is written as

$$K(t) = - \sum_{j=1}^n g\left(\frac{j}{n}\right) \sqrt{2} \sin \pi jt. \quad (2.13)$$

Here,  $g(\xi)$  is the concentration factor satisfying

$$\frac{g(\xi)}{\xi} \in C^2[0, 1],$$

where  $C^2$  is the class of twice differentiable and continuous functions. This admissibility condition is for odd concentration kernels. If the concentration factor  $g(\xi)$  is normalized so that

$$\int_0^1 \frac{g(\xi)}{\xi} d\xi = 1. \quad (2.14)$$

### 3. Asymptotic theory for the proposed change-point estimator

In the context of change-point estimation, it is well known that the points around a true change-point cannot be distinguished asymptotically with a fixed change magnitude. For example, in the least-squares fitting, the total variation with perfect segmentation is asymptotically equivalent to that with an estimate of change-point in a neighborhood of the true change-point. The proposed method has nonlinear enhancement that is easily implemented to pinpoint finitely many change-points in piecewise smooth  $f$ 's.

Let  $\mathcal{G}_n(M) = \{\hat{\tau}_1, \dots, \hat{\tau}_M\}$  denote the set of change-point estimates using the proposed method. Theorem 3.1 establishes the desirable property for the proposed edge detector when  $M$  is prespecified and  $\mathcal{G}_n(M)$  is asymptotically closed to the true change-point set. Let  $C_M(\delta_n)$  contain all the sets in the  $\delta_n$  neighborhood of the true change-point locations,

$$C_M(\delta_n) = \{(\tau'_1, \dots, \tau'_M) : 0 < \tau'_1 < \dots < \tau'_M \leq 1, \quad |\tau'_s - \tau_s| \leq \delta_n, \text{ for } 1 \leq s \leq M\}, \quad (3.1)$$

where  $\delta_n$  is some positive sequence.

To establish the consistency of the proposed edge detector, the following assumptions are imposed,

- (A1) (Piecewise smoothness)  $f(x)$  is piecewise smooth in the sense of having a finite number of jump discontinuities such that for all  $x$

$$\frac{f(x+t) - f(x-t) - [f](x)}{t} \in L^1[0, \pi] \quad (3.2)$$

where  $[f](x) = f(x+) - f(x-)$  is a jump function.

- (A2) Let  $\lambda_n = \min_{1 \leq k \leq M+1} n(\tau_k - \tau_{k-1})$ ,  $\lambda_n \rightarrow \infty$  as  $n \rightarrow \infty$ .

- (A3) There are  $M$  finite number of change-points with the corresponding edge.
- (A4)  $\eta_{\min} = \min_{1 \leq r \leq M} d(x_{\tau_r})$  is positive constant, where  $d(x)$  is the change-point detection statistic.

Assumption A1 provides the piecewise smoothness of underlying functions. A2 is a standard requirement for the theoretical development of a multiple change-points problem. A3 means that there is a finite number of change-points. And A4 assumes that the smallest signal amount among all the change-points is bounded away from zero. This assumption allows change-points to be asymptotically distinguishable. When the number of change-points is known as  $M_n = M$ , and the change magnitudes are fixed we can establish the same rate of convergence as  $O_p(1)$ .

Gelb and Tadmor (1999, 2000, 2002) provide some odd concentration kernels. We propose and compare the methods with the following kernels. Since the difference

$$\hat{f}'(x) = \hat{f}(x+) - \hat{f}(x-) \rightarrow \Delta, \quad x \in (0, 1) \quad (3.3)$$

has some information about the local discontinuity and may shoot off at the edge. Therefore, the estimator of the derivative of  $f$  can be a candidate for edge detection. Consider the underlying function estimation with the sample Fourier coefficients and then take the difference since the Gibbs phenomenon appears at the edge. The Gibbs phenomenon occurs as overshoot in which the Fourier series of a piecewise continuously differentiable periodic function behaves at a jump discontinuity. Also concentration kernels are used for edge detection. With the sample Fourier coefficients  $\hat{\phi}_j$ , we have

$$\hat{f}(x) = \hat{\phi}_0 + \sum_{j=1}^J \hat{\phi}_j \sqrt{2} \cos \pi j x, \quad x \in (0, 1) \quad (3.4)$$

where  $J$  is the truncated number of the sample Fourier coefficients. In practice, a data-driven choice of  $J$  is desirable. For example, Eubank and Hart (1992) proposed to choose  $J$  to minimize

$$r(k) = \sum_{j=1}^k \phi_j^2 - \frac{k c_\alpha \hat{\sigma}^2}{n}, \quad k = 1, 2, \dots, n-1,$$

where  $c_\alpha$  is controlled to detect the change-points, and  $\hat{\sigma}^2$  is the variance estimator of the underlying function.

Take its derivative such as

$$\hat{f}'(x) = -\sqrt{2}\pi \sum_{j=1}^J j \hat{\phi}_j \sin \pi j x, \quad x \in (0, 1). \quad (3.5)$$

Appropriate concentration kernels are considered as edge detector statistics. The proposed estimators are as follows considering the kernels in Gelb and Tamor (2000). These concentration kernels amplify the gap at the discontinuity points. Let  $\xi_j = j/n$ .

- (i) With a kernel  $g(t) = t$

$$d_1(x) = \left| -\sqrt{2}\pi \sum_{j=1}^J \xi_j \hat{\phi}_j \sin \pi j x \right|, \quad x \in (0, 1). \quad (3.6)$$

(ii) Using a polynomial kernel  $g(t) = pt^p$  with  $p = 1/2$

$$d_2(x) = \left| -\sqrt{2\pi} \sum_{j=1}^J p(\xi_j)^p \hat{\phi}_j \sin \pi jx \right|, \quad x \in (0, 1). \quad (3.7)$$

(iii) With a trigonometric kernel  $g(t) = \sin \pi t$

$$d_3(x) = \left| -\sqrt{2\pi} \sum_{j=1}^J \sin(\pi \xi_j) \hat{\phi}_j \sin \pi jx \right|, \quad x \in (0, 1). \quad (3.8)$$

(iv) With an exponential kernel  $g(t) = \exp(t^2/t^2 - 1)$

$$d_4(x) = \left| -\sqrt{2\pi} \sum_{j=1}^J \exp\left(\frac{\xi_j^2}{\xi_j^2 - 1}\right) \hat{\phi}_j \sin \pi jx \right|, \quad x \in (0, 1). \quad (3.9)$$

**Theorem 3.1** Under assumptions A1 and A2,  $|\hat{\tau}_s - \tau_s| = o_p(1)$  for  $s = 1, \dots, M$ .

**proof.** Consider  $d_1(\cdot)$  function when there is only one change-point  $\tau$ . Then we extend to the case when there is a fixed known number of change-points.

The idea of proof is as follows: When the estimated number of change points  $\hat{M}$  is less than  $M_0$ , the effect of goodness of fit term  $g_n$  dominates, which results in  $P(\hat{M} < M) \rightarrow 0$ . On the other hand, the effect of the penalty term  $q_n$  dominates and leads to  $P(\hat{M} > M) \rightarrow 0$ . Therefore, we have  $P(\hat{M} = M) \rightarrow 1$ .

Consider the change-point is estimated where the maximum of  $|d(\cdot)|$  is obtained. Assume that the change-point occurs at the design point such as  $\tau = c/n$  for simplicity.

$$f(x) = \begin{cases} 0, & 0 \leq x < \tau \\ 1, & \tau \leq x < 1. \end{cases}$$

We will show that  $d(\cdot)$  has the maximum at the true change-point.

Since  $\hat{\phi}_j \rightarrow \phi_j$  in probability for the sample Fourier coefficients (Kim and Hart, 2011), we have

$$d(x) = \sum_{j=1}^J \frac{j}{n} \hat{\phi}_j \sin \pi jx \rightarrow \sum_{j=1}^J \frac{j}{n} [\phi_j + o_p(1)] \sin \pi jx = d_\phi(x) + o_p(1),$$

with the corresponding true Fourier coefficients  $\phi_j$ 's. Likewise

$$d(\tau) = \sum_{j=1}^J \frac{j}{n} \hat{\phi}_j \sin \pi j\tau \rightarrow \sum_{j=1}^J \frac{j}{n} \phi_j \sin \pi j\tau = d_\phi(\tau) + o_p(1),$$

where  $d_\phi(\tau) = \sum_{j=1}^J j/n \phi_j \sin \pi j\tau$ .

Let  $d_1(x) = |d(x)|$ . We should show, for some  $\eta > 0$ ,  $\tau - \eta < x < \tau + \eta$ ,

$$P(|d_1(x) - d_1(\tau)| > \delta) \rightarrow 0.$$



Since  $d(x) \rightarrow d_\phi(x)$  and  $d(\tau) \rightarrow d_\phi(\tau)$  in probability by the functional property, we have

$$\begin{aligned} P\left(|d(x) - d_\phi(x)| > \frac{\delta}{3}\right) &< \frac{\epsilon}{3}, \\ P\left(|d_\phi(\tau) - d(\tau)| > \frac{\delta}{3}\right) &< \frac{\epsilon}{3}. \end{aligned}$$

For  $\epsilon > 0$ ,  $x \in (\tau - \eta, \tau + \eta)$  in some neighborhood of  $\tau$ , and  $x^* = \arg \max_{(\tau - \eta, \tau + \eta)} |d(x)|$ .

$$\begin{aligned} P(|d_1(x) - d_1(\tau)| > \delta) &= P(|\max |d(x)| - |d(\tau)|| > \delta) \\ &= P(|d(x^*)| - |d(\tau)| > \delta) \\ &\leq P(|d(x^*) - d(\tau)| > \delta) \\ &\leq P(|d(x^*) - d_\phi(x^*)| + |d_\phi(x^*) - d_\phi(\tau)| + |d_\phi(\tau) - d(\tau)| > \delta) \\ &\leq P(|d(x^*) - d_\phi(x^*)| > \delta/3) + P(|d_\phi(x^*) - d_\phi(\tau)| > \delta/3) \\ &\quad + P(|d_\phi(\tau) - d(\tau)| > \delta/3) \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned} \tag{3.10}$$

Since we have  $x \in (\tau - \eta, \tau + \eta)$ ,

$$P\left(|d_\phi(x) - d_\phi(\tau)| > \frac{\delta}{3}\right) < \frac{\epsilon}{3},$$

in the  $\eta$ -neighborhood of  $\tau$ .

Take  $J = O(n^{1/3})$  since

$$\begin{aligned} \left| \sum_{j=1}^J \frac{j}{n} \phi_j (\sin \pi j x - \sin \pi j \tau) \right| &\leq \sum_{j=1}^J \frac{j}{n} |\phi_j \pi j (x - \tau) \cos \pi j x^*| \\ &\leq \eta \sum_{j=1}^J \frac{\pi j^2}{n} |\phi_j| \\ &= \eta \cdot O\left(\frac{J^3}{n}\right) < \epsilon. \end{aligned} \tag{3.11}$$

Similarly, for  $x > \tau$  and  $x$  is in the neighborhood of  $\tau$ , it holds. Therefore, using (3.10)

$$\begin{aligned} P(|\hat{\tau} - \tau| > \delta_\tau) &= P\left(\left| \max_{\tau - \eta < x < \tau + \eta} d_1(x) - d_1(\tau) \right| > \delta\right) < \epsilon, \\ |\hat{\tau} - \tau| &= o_p(1). \end{aligned}$$

Likewise for each change-point, under assumptions A1, A2 and A3, we have

$$|\hat{\tau}_s - \tau_s| = o_p(1) \quad \text{for } s = 1, \dots, M. \tag{3.12}$$

To prove equation (3.12), it suffices to show that for any  $\epsilon > 0$ , there exists  $\delta > 0$ ,

$$P(|\hat{\tau}_s - \tau_s| < \delta, s = 1, \dots, M) > 1 - \epsilon, \quad \text{as } n \rightarrow \infty. \tag{3.13}$$

Or for  $G_n(M) = \{(\hat{\tau}_1, \dots, \hat{\tau}_M)\}$  the set of change-point estimators in  $C_M(\delta)$  in (3.1)

$$P(G_n(M) \in C_M(\delta)) > 1 - \epsilon.$$

Let

$$D_r(M, \delta) = \{(\tau'_1, \dots, \tau'_M) : 0 < \tau'_1 < \dots < \tau'_M \leq 1, \quad |\tau'_s - \tau'_r| > \delta, \text{ for } 1 \leq s \leq M\}.$$

By the assumptions, for some  $r$

$$P(G_n(M) \in D_r(M, \delta)) < \frac{\epsilon}{M}.$$

For every  $(\tau'_1, \dots, \tau'_M) \in D_r(M, \delta)$ ,

$$\begin{aligned} P(G_n(M) \in C_M(\delta_n)) &= 1 - P\left(\bigcup_r [G_n(M) \in D_r(M, \delta)]\right) \\ &\geq 1 - \sum_{r=1}^M P(G_n(M) \in D_M(\delta)) \\ &> 1 - \epsilon \end{aligned}$$

which completes the result.  $\square$

**Theorem 3.2** The true number of change-points  $M$  is fixed and finite. Then

$$P(\hat{M} = M) \rightarrow 1 \quad \text{as } n \rightarrow \infty \quad (3.14)$$

where  $\hat{M}$  is the estimator of  $M$ .

**proof.** From Theorem 3.1

$$P(\hat{\tau}_l \in B_l) > 1 - \delta,$$

where

$$\begin{aligned} B_l &= \{\tau'_l : 0 < \tau'_1 < \dots < \tau'_M \leq 1, \quad |\tau'_1 - \tau'_s| < \delta_n \text{ for } 1 \leq s \leq M\} \\ P(\hat{M}_n < M) &= P\left(\bigcup_{l=1}^{M-1} [\hat{\tau}_l \notin B_l]\right) \\ &\leq \sum_{l=1}^{M-1} P(\hat{\tau}_l \notin B_l) \\ &\leq (M-1)\delta < \epsilon. \end{aligned}$$

Therefore,

$$P(\hat{M} \geq M) \rightarrow 1 \quad \text{as } n \rightarrow \infty. \quad \square \quad (3.15)$$

For  $M < M^*$  for some finite  $M^*$ , consider

$$\begin{aligned}
P(\hat{M} > M) &= \{ \exists \hat{\tau}_l \text{ such that } |\hat{\tau}_l - \tau_s| > \delta_n, \text{ for } 1 \leq s \leq M_n \} \\
&\leq P\left(\bigcup_{l=1}^{M^*} [\hat{\tau}_l \notin B_l]\right) \\
&\leq \sum_{l=1}^{M^*} P(\hat{\tau}_l \notin B_l). \\
&\leq M^* \delta < \epsilon.
\end{aligned}$$

That is,

$$\begin{aligned}
P(\hat{M} > M) &\rightarrow 0 \quad \text{as } n \rightarrow \infty. \\
P(\hat{M} = M) &\rightarrow 1 \quad \text{as } n \rightarrow \infty. \quad \square
\end{aligned} \tag{3.16}$$

#### 4. Simulation

We conducted simulations to investigate the behavior of the proposed change-point estimators. The data were generated from various change-point models with i.i.d. normal errors having mean 0 and variance  $\sigma^2$ , and the design points  $x_i = i/n$ ,  $i = 1, \dots, n$  in the corresponding range. The models considered are as follows,

$$Y_i = f(x_i) + \epsilon_i, \quad i = 1, \dots, n.$$

where  $\epsilon_i \sim N(0, \sigma^2)$ . The change-points are assumed to occur at  $x_i \in (0, 1)$ .

(i) One change-point

$$f(t) = \begin{cases} -\text{sign}(t) \times \cos((t/2) \times (2 + \text{sign}(t))), & -1 \leq t < 0 \\ -1, & t = 0 \\ -\text{sign}(t) \times \cos((t/2) \times (2 + \text{sign}(t))), & 0 < t \leq 1 \end{cases}$$

where

$$\text{sign}(t) = \begin{cases} -1, & t \leq 0 \\ 0, & t = 0 \\ 1, & t > 0. \end{cases}$$

(ii) Periodic discontinuities

$$f(t) = t - \frac{1}{2}, \quad t \in (0, 1).$$

(iii) Two change-points at  $t = 1, 2$  in the step function

$$f(t) = \begin{cases} 0, & 0 \leq t < 1 \\ 1, & 1 \leq t < 2 \\ 0, & t \geq 2 \end{cases}$$

(iv) Two change-points at  $t = 1, 2$  in linear trend with  $\Delta = 0.5$

$$f(t) = \begin{cases} \beta t, & 0 \leq t < 1 \\ (\beta + \Delta)t + 1, & 1 \leq t < 2 \\ \beta t, & t \geq 2 \end{cases}$$

(v) Two change-points at  $t = 1, 2$  in trigonometric function

$$f(t) = \begin{cases} \cos(\pi t), & 0 \leq t < 1 \\ \cos(\pi t) + 1, & 1 \leq t < 2 \\ \cos(\pi t), & t \geq 2 \end{cases}$$

(vi) Two edges at  $t = -1, 1$  in piecewise smooth function

$$f(t) = \begin{cases} \cos\left(t - \frac{t}{2} \times (\text{sign}(|t| - 1))\right), & -2 \leq t < 0 \\ \cos(2.5t + t(\text{sign}(t - 1))), & 0 \leq t < 2. \end{cases}$$

where

$$\text{sign}(t) = \begin{cases} -1, & t < 0 \\ 0, & t = 0 \\ 1, & t > 0. \end{cases}$$

A sample size of  $n = 200$  is used for Case (i) and  $n = 300$  is used for the other cases, and 1,000 repetitions were performed for each edge detection estimator.

To detect the peaks as change-points, the following measures

$$\begin{aligned} d_i(x_i) - d_i(x_{i-1}) &> 0, \quad d_i(x_i) - d_i(x_{i+1}) > 0, \\ [d_i(x_i) - d_i(x_{i-1})] \times [d_i(x_i) - d_i(x_{i+1})] &> 0, \end{aligned}$$

are checked for the possible candidate estimates, for  $i = 1, 2, 3, 4$ .

Six types of functions including change-points are considered in the simulation. The change-point estimates are provided in Figures 1, 2, and 3, respectively. In these figures, the dotted (red) lines are true underlying functions with data (dot) and the solid lines (green) are  $d_1, d_2, d_3, d_4$  statistics used for change-point estimation. In these figures the input data have small variance  $\sigma = 0.1$  and the output of change-point detection statistics seem to have peaks to find the change-points. We intend to show how the change-point detection statistics behave and move according to the underlying functions.

We finally note via simulation that the results remain satisfactory even with a higher level. It was difficult to detect the change-points in piecewise polynomial functions when the variances were bigger. We do not need any a priori knowledge of the upper bound of the number of change-points for applying it. Table 1 provides the simulation results with three evaluation columns divided by a vertical bar. In the first column, the probability at least one true edge are calculated. The second column provides the probability that both edges are detected with top 10 candidates. The third column shows the probability that both edges are detected with top 5 candidates. Inside each column, four change-point estimators  $d_1, d_2, d_3, d_4$  are listed and computations are done with  $\sigma = 0.1$  for low noise and  $\sigma = 0.5$  for moderate noise. Overall, the proposed edge detectors work fine in the sense

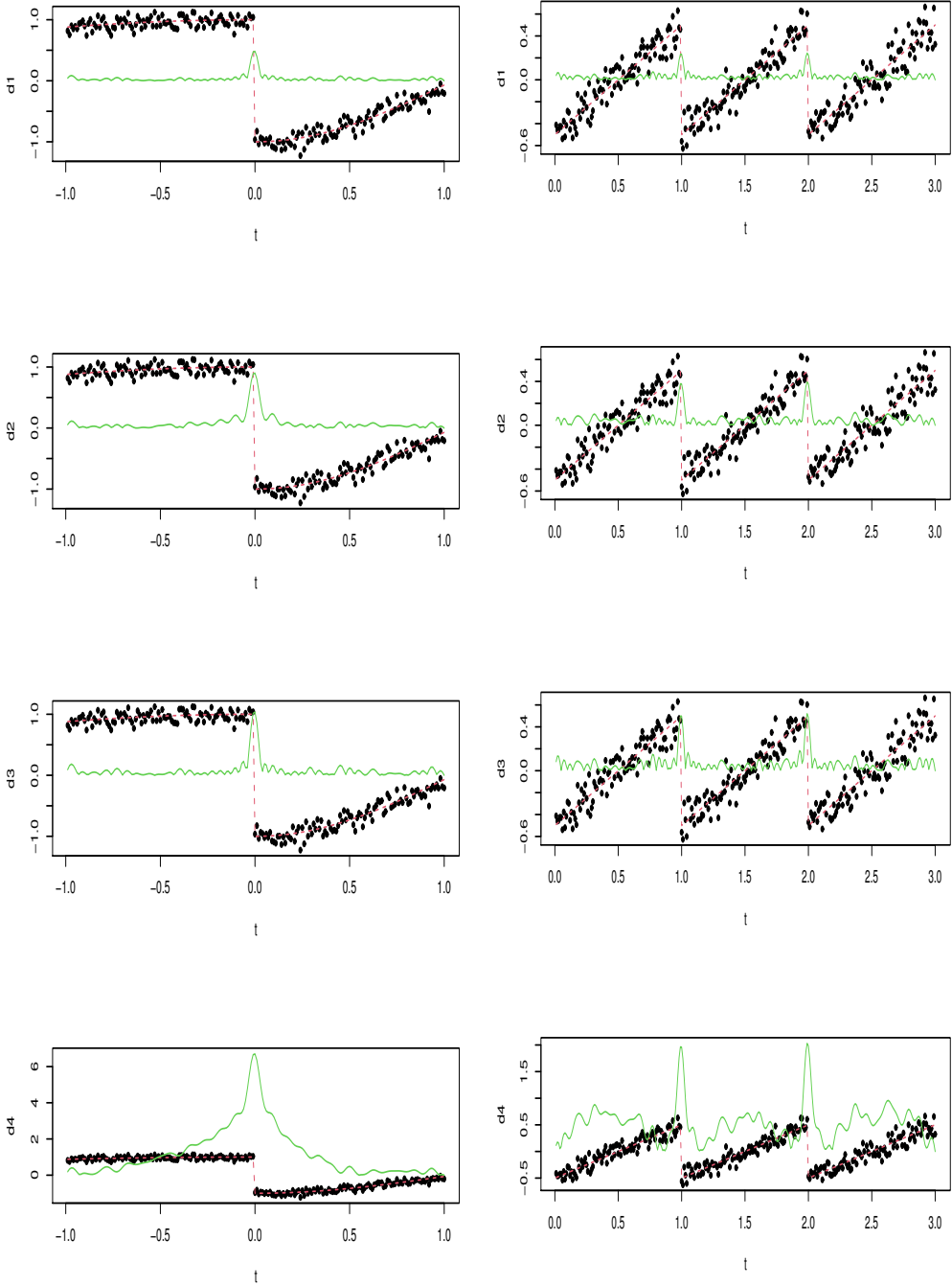


Figure 1: Estimates  $d_1, d_2, d_3$  and  $d_4$  (green line) from the top for Case (i) (left) and Case (ii) (right). Data (dot) with  $\sigma = 0.1$  have true underlying function (red line).

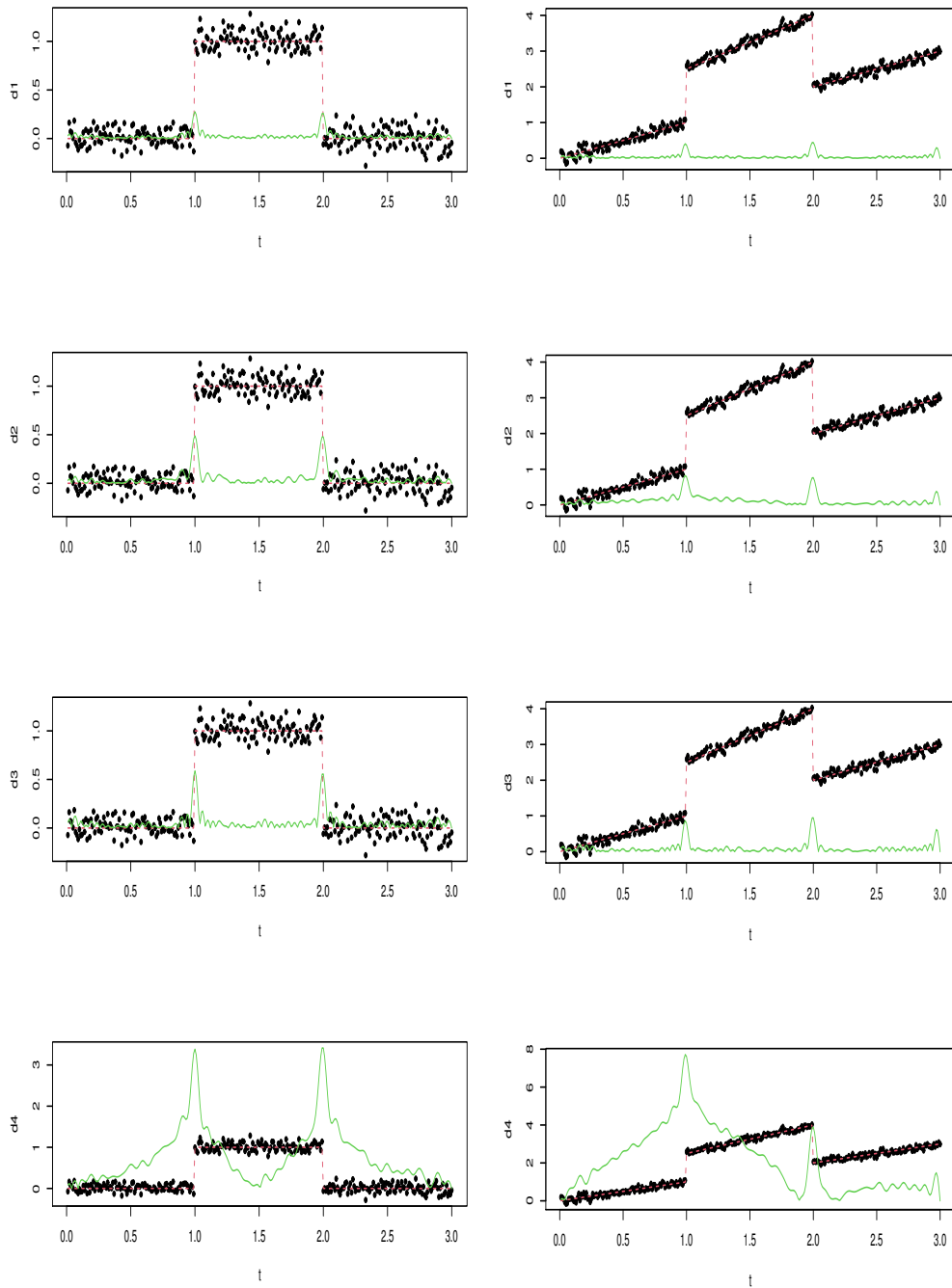


Figure 2: Estimates  $d_1, d_2, d_3$  and  $d_4$  (green line) from the top for Case (iii) (left) and Case (iv) (right). Data (dot) with  $\sigma = 0.1$  have true underlying function (red line).

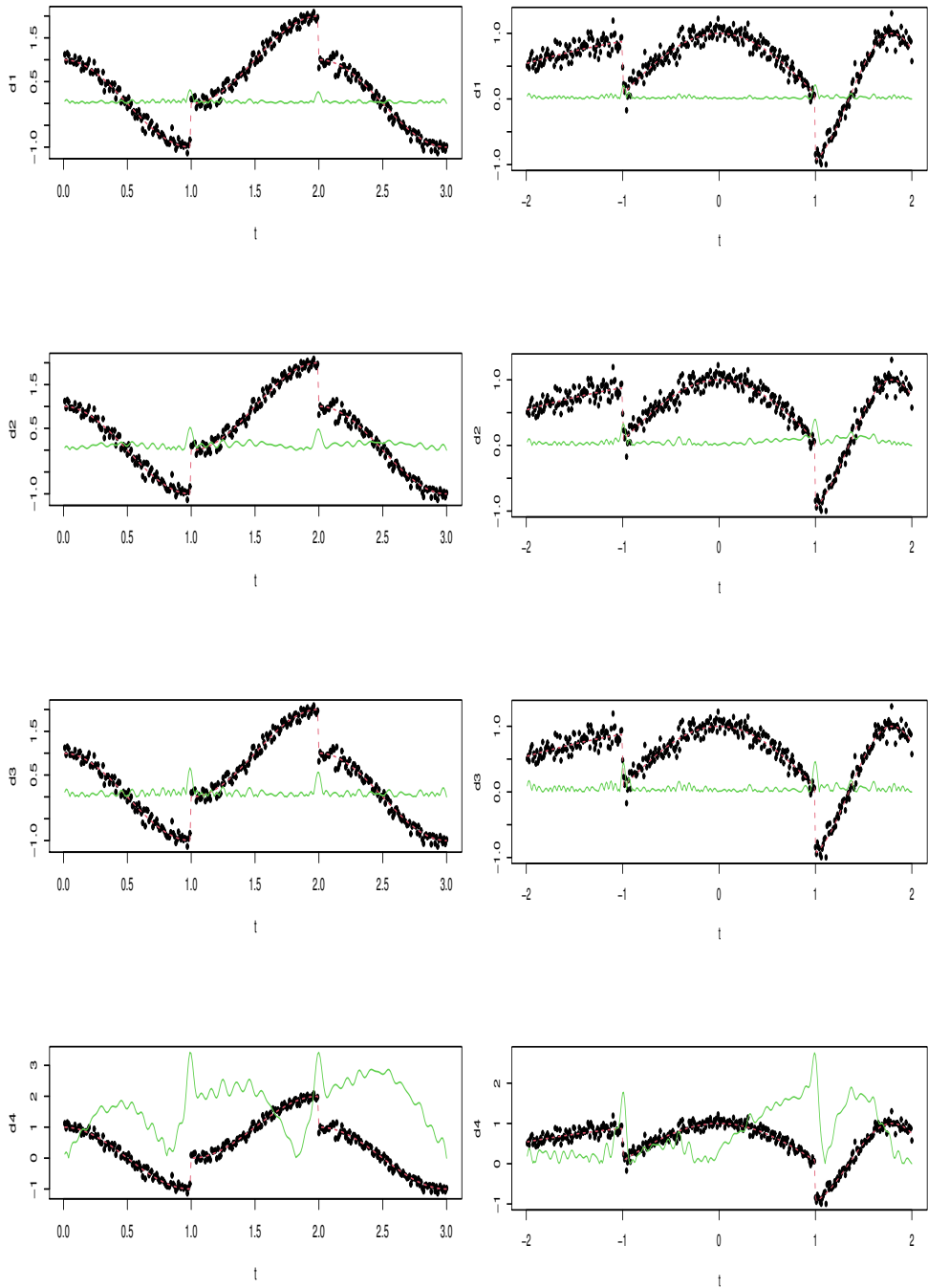


Figure 3: Estimates  $d_1, d_2, d_3$  and  $d_4$  (green line) from the top for Case (v) (left) and Case (vi) (right). Data (dot) with  $\sigma = 0.1$  have true underlying function (red line).

Table 1: Table 1. Comparison of change-point estimators in 1,000 repetitions with  $J = \lfloor n/4 \rfloor$  : proportions of estimated intervals containing true change-points

Model	$\sigma$	At least one true edge				Both edges with top 10				Both edges with top 5			
		$d_1$	$d_2$	$d_3$	$d_4$	$d_1$	$d_2$	$d_3$	$d_4$	$d_1$	$d_2$	$d_3$	$d_4$
(i)	0.1	1.0	1.0	1.0	1.0	0.998	1	0.998	1.0	0.998	1.0	1.0	0.998
$n = 200$	0.5	0.782	0.944	0.806	0.937	0.399	0.687	0.420	0.661	0.227	0.440	0.250	0.937
(ii)	0.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.999	1.0	1.0	1.0
$n = 300$	0.5	0.789	0.946	0.802	0.95	0.394	0.662	0.412	0.625	0.233	0.414	0.243	0.34
(iii)	0.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
$n = 300$	0.5	0.841	0.988	0.857	0.999	0.458	0.851	0.489	0.928	0.278	0.601	0.306	0.542
(iv)	0.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
$n = 300$	0.5	0.996	0.999	0.996	1.0	0.922	0.997	0.933	0.937	0.758	0.909	0.776	0.783
(v)	0.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
$n = 300$	0.5	0.818	0.922	0.837	0.643	0.434	0.619	0.459	0.247	0.264	0.393	0.282	0.111
(vi)	0.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.936	0.969	0.511	0.963
$n = 300$	0.5	0.693	0.889	0.720	0.827	0.254	0.471	0.271	0.054	0.133	0.225	0.141	0.011

of detection proportions. The method is not limited to the normal noise distributions. The noise distributions can be considered from normal, uniform and double exponential distributions.

Figures 4 to 6 provide the simulation results in barplots of change-points in the left column and plots of estimates of  $d_1$ ,  $d_2$ ,  $d_3$ , and  $d_4$  in one example data set in the right column. These barplots show that the estimates concentrate on the true change-points. The plots with the one data set show that the behavior of the considered statistics tend to have peaks around the true change-points. Each estimator can detect the edge type change-points for different underlying change patterns.

In Figures 7 to 9 with the simulation results,  $d_4$  does not detect both change-points well according to shape of underlying functions and variance. Therefore appropriate kernel functions might be useful for change-point detection of some types of underlying functions.

## 5. Examples

We applied our methodology to a classic data set consisting of the annual volumes of the Nile River from 1871 to 1970, as shown in Figure 10(a). Cobb (1978) estimated one change-point as  $t = 28$  (year 1898). Cobb (1978) used an approximation to the conditional distribution of the maximum likelihood estimator of the change-point. The multiple change-points problem can be replaced by a model selection one. A Bayesian information criterion (BIC) can be used as a model comparison measure, since a BIC can measure the efficiency of the parameterized model in terms of predicting the data. It penalizes the complexity of the model where complexity involves the number of parameters in the model as

$$\text{BIC} = -2 \log(\text{Likelihood}) + M \log n = n \log \hat{\sigma}_\epsilon^2 + M \log n$$

where  $\hat{\sigma}_\epsilon^2$  is an estimator of error variance,  $M$  is the number of free parameters to be estimated such as the number of change-points.

By the BIC criterion, one change-point model is considered in Kim and Hart (2011) and results in change-point estimate as  $t = 28$ . All four edge estimators pick  $t = 28$  as the same as the change-point estimate in Kim and Hart (2011). Especially edge4 estimates two change-points at  $t = 28$  and  $t = 68$ . The vertical line denotes the change-point estimate as  $t = 28$  in Figure 10(b). The edge measures and the maximum are chosen from  $t = 20$  to  $t = 80$ , that is, 20% is truncated for change-point search.

We also applied our methodology to detect change-points in the daily won-dollar exchange rate from January 3, 1995 to December 21, 2011 as shown in Figure 11(a). The three methods provide



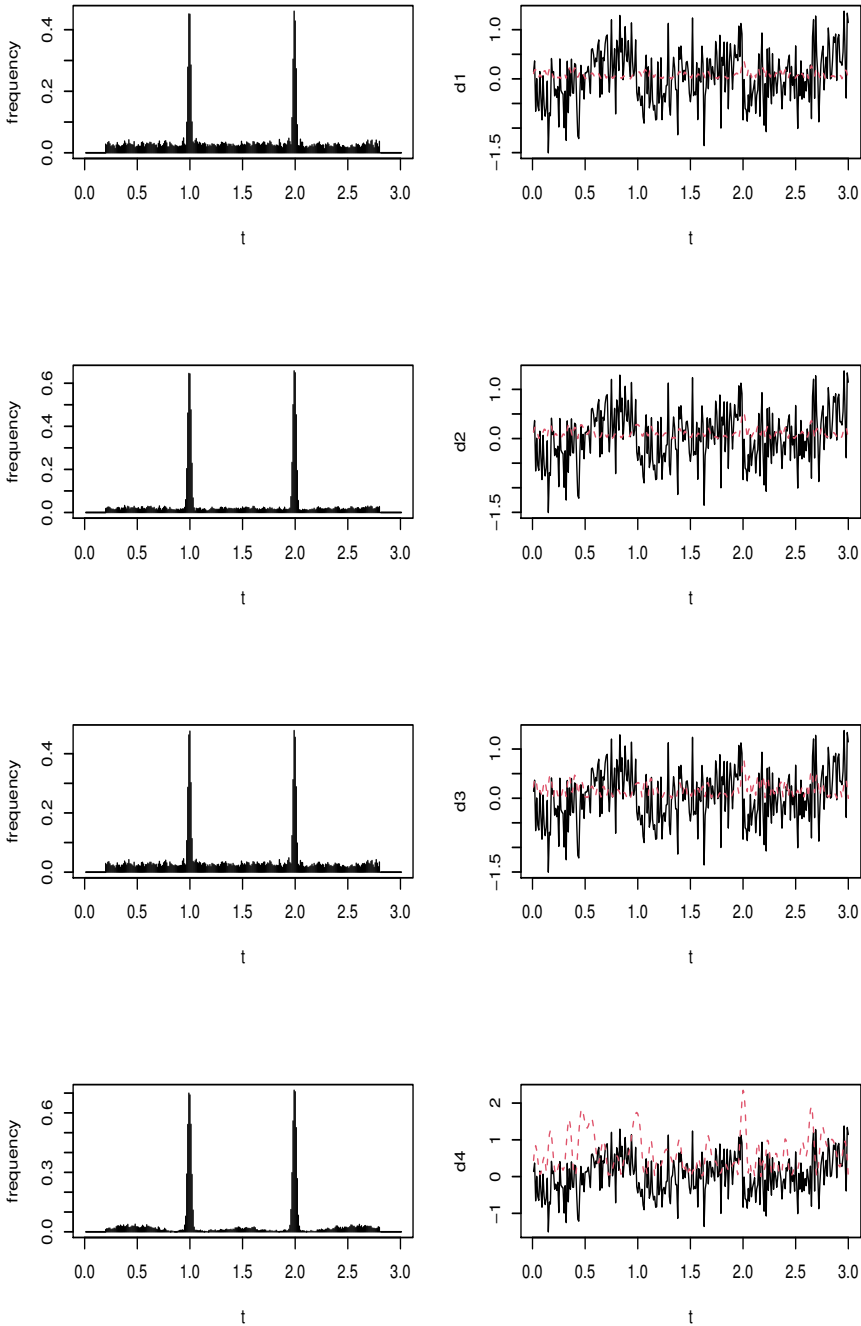


Figure 4: Case (i) Barplots (left) in 1,000 repetitions and plots (right) of change-point estimates  $d_1, d_2, d_3$  and  $d_4$  (from top to bottom) in one data set (data: black solid, estimates: red dotted) with  $\sigma = 0.5, n = 300$ .

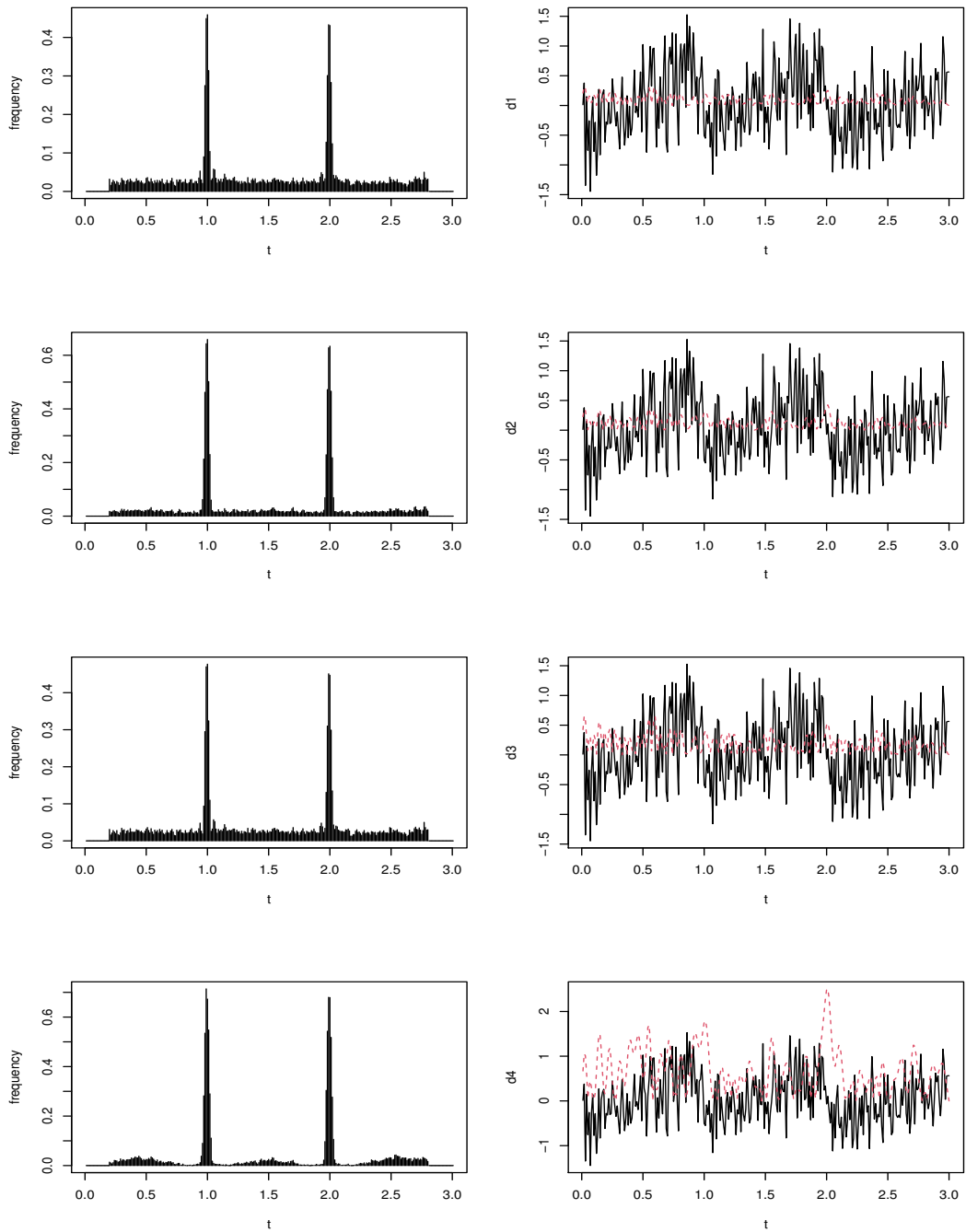


Figure 5: Case (ii) Barplots (left) in 1,000 repetitions and plots (right) of change-point estimates  $d_1$ ,  $d_2$ ,  $d_3$  and  $d_4$  (from top to bottom) in one data set (data: black solid, estimates: red dotted) with  $\sigma = 0.5$ ,  $n = 300$ .

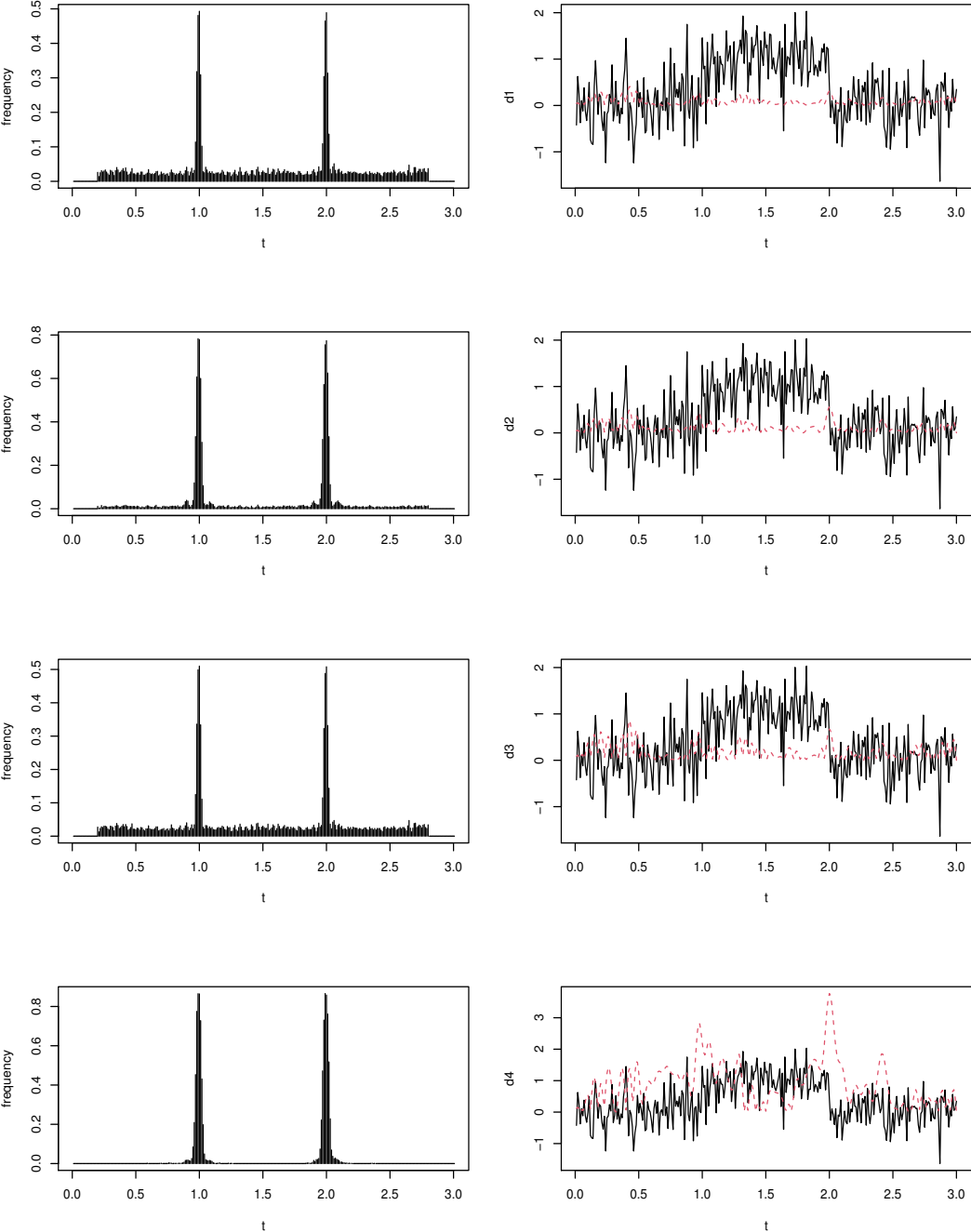


Figure 6: Case (iii) Barplots (left) in 1,000 repetitions and plots (right) of change-point estimates  $d_1, d_2, d_3$  and  $d_4$  (from top to bottom) in one data set (data: black solid, estimates: red dotted) with  $\sigma = 0.5, n = 300$ .

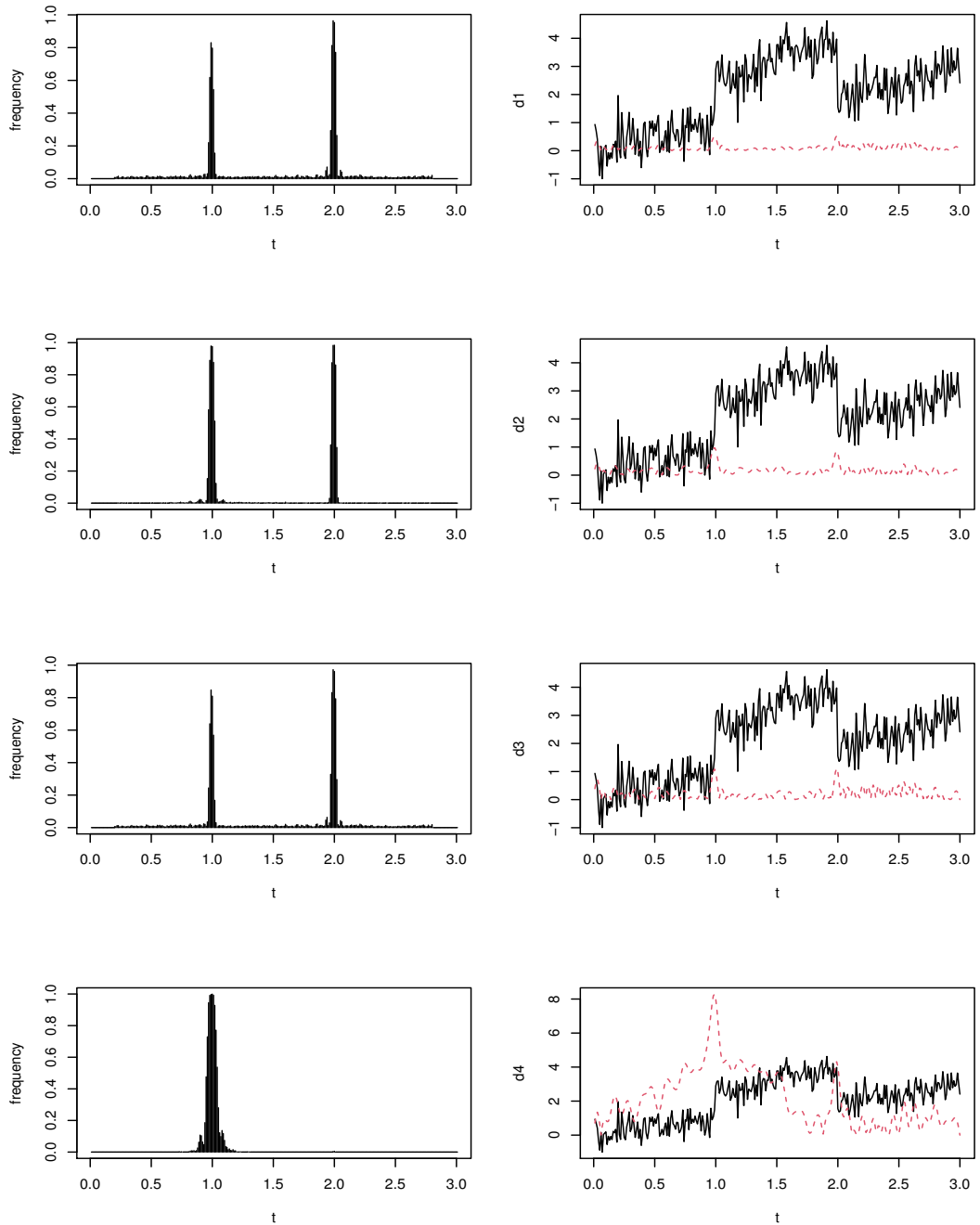


Figure 7: Case (iv) Barplots (left) in 1,000 repetitions and plots (right) of change-point estimates  $d_1$ ,  $d_2$ ,  $d_3$  and  $d_4$  (from top to bottom) in one data set (data: black solid, estimates: red dotted) with  $\sigma = 0.5$ ,  $n = 300$ .

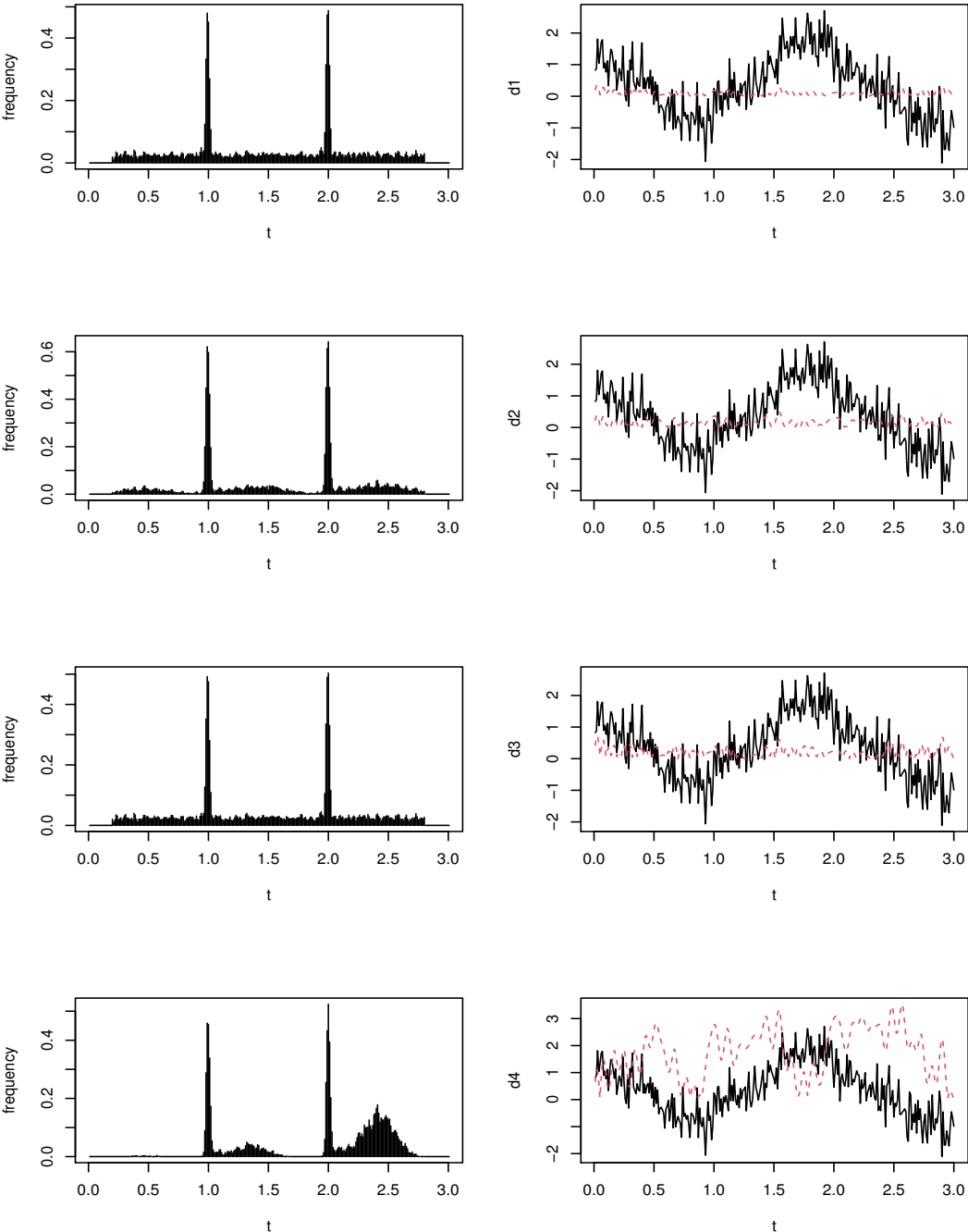


Figure 8: Case (v) Barplots (left) in 1,000 repetitions and plots (right) of change-point estimates  $d_1$ ,  $d_2$ ,  $d_3$  and  $d_4$  (from top to bottom) in one data set (data: black solid, estimates: red dotted) with  $\sigma = 0.5$ ,  $n = 300$ .

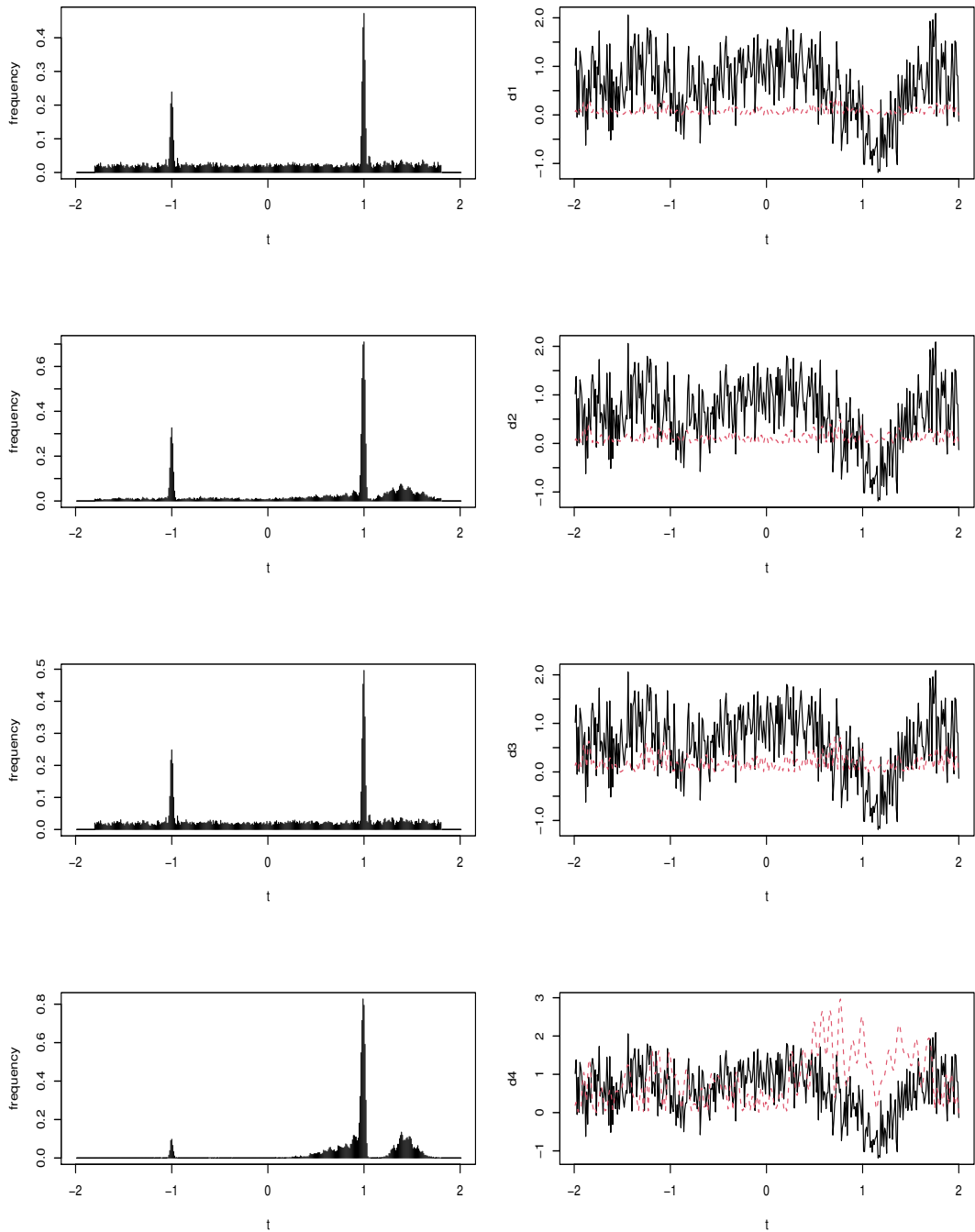


Figure 9: Case (vi) Barplots (left) in 1,000 repetitions and plots (right) of change-point estimates  $d_1, d_2, d_3$  and  $d_4$  (from top to bottom) in one data set (data: black solid, estimates: red dotted) with  $\sigma = 0.5, n = 300$ .

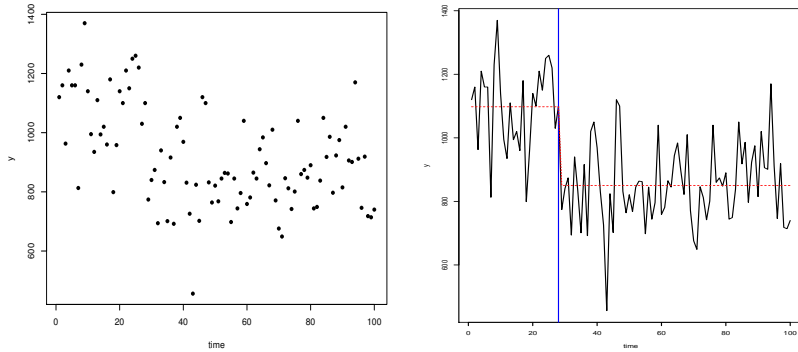


Figure 10: (a) Nile data plot (b) Change-point estimates as the vertical line at  $t = 28$  by  $d_1, d_2, d_3$ .

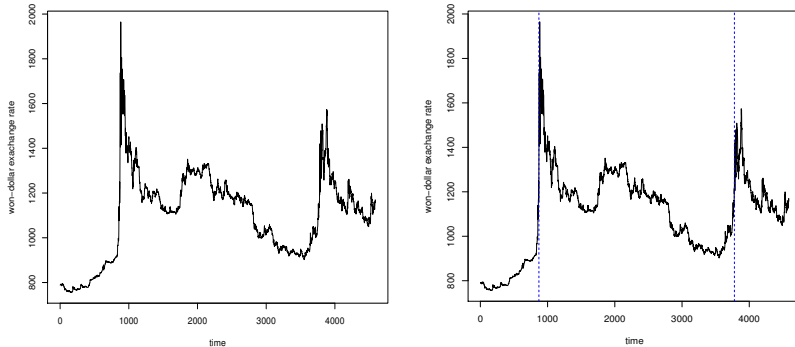


Figure 11: (a) Exchange rate data (b) Change-point estimates as the vertical line at  $t = 871(870), 3779$  by  $d_1, d_3, d_4$ .

two change-points at 871 (Dec. 9, 1997) and 3779 (Oct. 6, 2008) as shown in Figure 11(b). Edge2 estimates change-points at  $t = 870, 3779$ .

## 6. Concluding remarks

In the multiple change-point (MCP) problem, we have proposed estimators based on Fourier series and concentration kernels from Hilbert transformation for detection of multiple change-points as edges or discontinuities. We were able to consistently detect any discontinuity or abrupt functional changes without making any distributional assumptions. The proposed method is elegantly able to estimate both the number of change-points and their locations, thus eliminating the need for prior knowledge. The consistency of the proposed procedure is established under mild conditions. The true number of change-points is assumed to be unknown, and the distribution of change-points is estimated and used to choose the number of change-points. To facilitate the implementation of the method, we suggest giving a possible range for the number of change-points, which has been shown to work faster.

The proposed method is based on the assumption that there exists at least one change-point. In practical applications, we need to use some tests for change or discontinuity to verify this assumption. This edge detection method can be extended to the multidimensional case with the choice of an ap-

appropriate concentration factor. Our proposed method is an omnibus method, and thus cannot diagnose whether a change occurs in location, scale, or shape. Also the method can be extended to the case with dependent noises such as AR(p) for future applications. However our methods have limitation that the large variability of data makes it hard to detect the change-points. In such cases, local kernel functions are expected to improve the detection ability.

## References

- Cobb GW (1978). The problem of the Nile: Conditional solution to a change-point problem, *Biometrika*, **65**, 243–251.
- Engelberg S (2008). Edge detection using Fourier coefficients, *The American Mathematical Monthly*, **115**, 499–513.
- Eubank RL and Hart J (1992). Testing goodness-of-fit in regression via order selection criteria, *Annals of Statistics*, **20**, 1412–1425.
- Fliess M, Cedric J, and Mboup M (2010). Algebraic change-point detection, *Applicable Algebra in Engineering, Communication and Computing*, **21**, 131–143.
- Gelb A and Tadmor E (1999). Detection of edges in spectral data, *Applied and Computational Harmonic Analysis*, **7**, 101–135.
- Gelb A and Tadmor E (2000). Detection of edges in spectral data II. Nonlinear enhancement, *SIAM Journal on Numerical Analysis*, **38**, 1389–1408.
- Gelb A and Tadmor E (2002). Spectral reconstruction of piecewise smooth functions from their discrete data, *Mathematical Modelling and Numerical Analysis*, **2**, 155–175.
- Guralnik V and Srivastava J (1999). Event detection from time series data. In *Proceedings of the Fifth ACM SIGKDD*.
- Hall P and Titterton DM (1992). Edge-preserving and peak-preserving smoothing, *Technometrics*, **34**, 429–440.
- Hawkins DM (2001). Fitting multiple change-point models to data, *Computational Statistics & Data Analysis*, **37**, 323–341.
- Kim J and Hart JD (1998). Tests for change in a mean function when the data are dependent, *Journal of Time Series Analysis*, **19**, 399–424.
- Kim J and Hart JD (2011). A change-point estimator using local fourier series, *Journal of Nonparametric Statistics*, **23**, 83–98.
- Kvernadze G (1998). Determination of the jumps of a bounded function by its Fourier series, *Journal of Approximation Theory*, **92**, 167–190.
- Lavielle M and Teyssière G (2006). Detection of multiple change-points in multivariate time series, *Lithuanian Mathematical Journal*, **46**, 287–306.
- McDonald JA and Owen AB (1986). Smoothing with split linear fits, *Technometrics* **28**, 195–208.
- Müller HG (1992). Change-points in nonparametric regression analysis, *The Annals of Statistics*, **20**, 737–761.
- Tolstov GP (1962). *Fourier Series*, Dover, New York.
- Wang Y (1995). Jump and sharp cusp detection by wavelets, *Biometrika*, **82**, 385–397.

Received July 31, 2021; Revised October 15, 2021; Accepted October 28, 2021