SCIENCE CHINA
Mathematics

• ARTICLES •

December 2016 Vol. 59 No. 12: 2463–2484
doi: 10.1007/s11425-016-0138-x

On nonparametric change point estimator based on empirical characteristic functions

In memory of Professor Xiru Chen (1934–2005)

TAN ChangChun[1], SHI XiaoPing[2], SUN XiaoYing[3] & WU YueHua[3,*]

[1]*School of Mathematics, Hefei University of Technology, Hefei* 230009*, China;*
[2]*Department of Mathematics and Statistics, Thompson Rivers University, Kamloops, BC V*2*C* 0*C*8*, Canada;*
[3]*Department of Mathematics and Statistics, York University, Toronto, ON M*3*J* 1*P*3*, Canada*
Email: cctan@ustc.edu.cn, xshi@tru.ca, sunying@mathstat.yorku.ca, wuyh@mathstat.yorku.ca

Received May 5, 2016; accepted May 26, 2016; published online July 20, 2016

Abstract We propose a nonparametric change point estimator in the distributions of a sequence of independent observations in terms of the test statistics given by Hušková and Meintanis (2006) that are based on weighted empirical characteristic functions. The weight function $\omega(t;a)$ under consideration includes the two weight functions from Hušková and Meintanis (2006) plus the weight function used by Matteson and James (2014), where a is a tuning parameter. Under the local alternative hypothesis, we establish the consistency, convergence rate, and asymptotic distribution of this change point estimator which is the maxima of a two-side Brownian motion with a drift. Since the performance of the change point estimator depends on a in use, we thus propose an algorithm for choosing an appropriate value of a, denoted by $a_{\rm s}$ which is also justified. Our simulation study shows that the change point estimate obtained by using $a_{\rm s}$ has a satisfactory performance. We also apply our method to a real dataset.

Keywords change point estimator, empirical characteristic function, tuning parameter, convergence rate, asymptotic distribution

MSC(2010) 62F12, 62G20

Citation: Tan C C, Shi X P, Sun X Y, et al. On nonparametric change point estimator based on empirical characteristic functions. Sci China Math, 2016, 59: 2463–2484, doi: 10.1007/s11425-016-0138-x

1 Introduction

Change point problems are common in many research areas including medical and health sciences, financial econometrics and risk management (see [3, 5, 8]). If there exists a change point in a data sequence, the results derived from the statistical analysis without taking it into account might be misleading. Many methods have been proposed in the literature to test or estimate change points in mean, variance, regression parameters, etc. For example, a local comparison method based statistic has been proposed in [6] to test whether there is a change point in mean of a data sequence. Note that these change point problems are special cases of the problem of change point in distributions of a sequence of random variables that will be considered in this paper.

Nonparametric methods play an important role in tackling the problem of a change point in distributions of a data sequence. Most of the nonparametric methods are based on either empirical distributions,

*Corresponding author

U-statistics or quantile functions (see [4,8,14,22]). Another nonparametric tool is the empirical characteristic function (ECF). The definition of the ECF was given in [21]. Kent [18] studied the weak convergence theorem of the ECF. Since then, the ECF has been applied to solve various statistical problems such as hypothesis testing for symmetry about the origin, dependence or normality (see [10, 11, 17, 19, 24]).

A class of test statistics based on the ECF has been proposed in [15] to test if there is a change point in distributions of a sequence of independent random variables. They gave two choices of the weight function for their proposed statistics. They studied the limiting behaviour of the test statistics under both null and alternative hypotheses. Built upon their statistics, a change point estimator is given in this paper for the same change point problem. The weight function $\omega(t;a)$ under consideration includes the two weight functions from [15] plus the weight function used in [20], where a is a tuning parameter. We will study the consistency, convergence rate, and asymptotic distribution of this estimator when the difference between the distributions before and after the change point tends to zero as the sample size goes to infinity.

Simulation results in [15] showed that the test statistics are robust with respect to the value of the tuning parameter a in the weight function, which, however, is selected from 1 to 4 increased by 1 each time in their simulation study. It is noted that the domain of a in their weight functions ranges from 0 to infinity. The real data example reveals that the change point estimate may be influenced significantly by the value of the tuning parameter a (see Table 1 of Section 4). Thus, accuracy of the change point estimate is in question. To tackle this problem, we propose an algorithm for selecting an appropriate value of a, $a_{\rm s}$, in order to obtain a change point estimate with a satisfactory accuracy.

The rest of the paper is organized as follows: In Section 2, we propose a nonparametric change point estimator in the distributions of a sequence of independent observations in terms of the test statistics given in [15] that are based on weighted empirical characteristic functions. In Section 3, we investigate the asymptotic properties of this estimator assuming that there exists one change point in the data sequence. We also give an example there. We present an algorithm for selecting $a_{\rm s}$ which is also justified in Section 4. We carry out simulation study to evaluate the performance of the change point estimation with the use of $a_{\rm s}$ in Section 5. A real data example is also given there. The proofs of all the theorems are given in Appendix.

The following notation are used throughout the rest of this paper. $I_A(\cdot)$ denotes the indicator function of the set A. $\lfloor a \rfloor$ represents the largest integer not greater than the real number a. "$\rightarrow_P$" stands for the convergence in probability. "$\Rightarrow$" means the weak convergence. $\Phi(\cdot)$ denotes the cumulative distribution function (cdf) of a standard normal distribution.

2 The change point estimator based on the ECF

Let $Y_{n,1}, Y_{n,2}, \dots, Y_{n,n}$ be a sequence of independent random variables, where $Y_{n,j}$ has a distribution function $F_{n,j}$, $j = 1, 2, \dots, n$. Consider the testing problem

$$H_0 : F_1 = F_{n,1} = F_{n,2} = \cdots = F_{n,n},$$

against

$$H_1 : F_1 = F_{n,1} = \cdots = F_{n,k_0^{(n)}} \neq F_{n,k_0^{(n)}+1} = \cdots = F_{n,n} = F_n, \quad \text{for} \quad k_0^{(n)} < n, \tag{2.1}$$

where $k_0^{(n)}$, F_1 and F_n are unknown. $k_0^{(n)}$ is called the change point. For the sake of convenience, the subscript n in $Y_{n,j}$ and $F_{n,j}$ and the superscript n in $k_0^{(n)}$ are all suppressed if there is no confusion.

The following class of test statistics has been developed in [15] based on the empirical characteristic function and a non-negative weight function $\omega(\cdot)$ with a non-negative tuning parameter a:

$$T_{\omega,\gamma}(k) = \left(\frac{k(n-k)}{n^2}\right)^{\gamma} \frac{k(n-k)}{n} \int_{-\infty}^{\infty} |\phi_k(t) - \phi_k^0(t)|^2 \omega(t) dt, \tag{2.2}$$

where $\gamma \in (0,1]$, $\omega(\cdot)$ satisfies that $\omega(t) > 0, t \in R$ and $0 < \int \omega(t)dt < \infty$, $\phi_k(t)$ and $\phi_k^0(t)$ are ECFs based on $Y_1, \ldots, Y_k$ and $Y_{k+1}, \ldots, Y_n$, respectively, i.e.,

$$\phi_k(t) = \frac{1}{k}\sum_{j=1}^{k} \exp\{\mathrm{i}tY_j\}, \quad \phi_k^0(t) = \frac{1}{n-k}\sum_{j=k+1}^{n} \exp\{\mathrm{i}tY_j\}, \quad j = 1, 2, \ldots, n.$$

Under the alternative hypothesis, we propose the change point estimator for k_0 as

$$\hat{k}_n = \arg\max_{1\leqslant k<n} T_{\omega,\gamma}(k). \tag{2.3}$$

Some choices of $\omega(\cdot)$ are

$$\omega_1(t;a) = \frac{1}{2a}\exp\{-a|t|\}, \quad t \in \mathbb{R}^1, \quad a > 0, \tag{2.4}$$

$$\omega_2(t;a) = \frac{\sqrt{a}}{\sqrt{\pi}}\exp\{-at^2\}, \quad t \in \mathbb{R}^1, \quad a > 0, \tag{2.5}$$

or

$$\omega_3(t;a) = \frac{a2^a\Gamma(\frac{1+a}{2})}{2\sqrt{\pi}\Gamma(1-\frac{a}{2})}|t|^{-a-1}, \quad t \in \mathbb{R}^1, \quad a \in (0,2). \tag{2.6}$$

We remark that $\omega_1(t;a)$ and $\omega_2(t;a)$ were given in [15] while $\omega_3(t;a)$ was used as the weight function in [20] for obtaining their nonparametric change point estimator in distributions of a sequence of multivariate random variables.

We assume that k_0 satisfies

$$k_0 = \lfloor n\tau_0 \rfloor, \quad \tau_0 \in [\kappa_1, \kappa_2] \quad \text{for some} \quad 0 < \kappa_1 \leqslant \kappa_2 < 1. \tag{2.7}$$

This is a conventional assumption made in change point detection problems [8]. The estimator for τ_0 is given by

$$\hat{\tau}_n = \frac{\hat{k}_n}{n} = \frac{1}{n}\arg\max_{1\leqslant k<n} T_{\omega,\gamma}(k). \tag{2.8}$$

3 Consistency and asymptotic distribution of the change point estimator

Define

$$\begin{aligned}\Delta_n &= \int\left\{\left(\int \cos(tx)d(F_1(x) - F_n(x))\right)^2 + \left(\int \sin(tx)d(F_1(x) - F_n(x))\right)^2\right\}\omega(t)dt \\ &= E[h(Y_1, Y_2)] - 2E[h(Y_1, Y_{k_0+1})] + E[h(Y_{k_0+1}, Y_{k_0+2})],\end{aligned} \tag{3.1}$$

and $h(x,y) = \int \cos(t(x-y))\omega(t)dt$. In this section, we will study consistency, convergence rate and asymptotic distribution of the change point estimator $\hat{\tau}_n$ under the assumption that $\Delta_n \to 0$. Its convergence rate not only describes how fast $\hat{\tau}_n$ converges to τ_0 but also is necessary in order to derive its asymptotic distribution that will enable us to calculate its mean square error (MSE). The following two theorems are given in sequel, and their proofs are given in Appendix.

Theorem 3.1. *Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of independent random variables, where $Y_1, \ldots, Y_{k_0}$ have a common distribution function F_1, and $Y_{k_0+1}, \ldots, Y_n$ have a common distribution function F_n. Assume that k_0 satisfies* (2.7) *and $\gamma \in (0,1]$. If Δ_n defined in* (3.1) *satisfies that $\Delta_n \to 0$ and*

$$n\Delta_n^2 \to \infty, \quad as \quad n \to \infty. \tag{3.2}$$

Then, as $n \to \infty$,

(1) $\hat{\tau}_n \to_P \tau_0$;

(2) $|\hat{\tau}_n - \tau_0| = O_P(\frac{1}{n\Delta_n^2})$.

Remark 3.2. If Δ_n is a constant not depending on n, from the proof of Theorem 3.1, we can make the conclusion that $\forall\,\varepsilon > 0$, $\lim_{n\to\infty} P(n^{\frac{1}{2}} l^{-1}(n)|\hat{\tau}_n - \tau_0| > \varepsilon) = 0$, where $l(n)$ is a slow varying function satisfying $\lim_{n\to+\infty} l(n) = +\infty$. Furthermore, similar to [23], we can obtain the stronger a.s. convergence rate: $|\hat{\tau}_n - \tau_0| = O(\frac{M(n)}{n})$ a.s. for any $M(n)$ satisfying that $M(n) \to \infty$ as $n \to \infty$.

We now consider the asymptotic distribution of $\hat{\tau}_n$ assuming that $\Delta_n \to 0$ as $n \to \infty$.

Theorem 3.3. *Under the same conditions as in Theorem* 3.1*, we have, as $n \to \infty$,*

$$\frac{(1+\gamma-2\gamma\tau_0)^2}{\lambda_1^2}\Delta_n^2(\hat{k}_n - k_0) \Rightarrow \arg\max_{-\infty<s<+\infty} G(s),$$

where

$$G(s) = \begin{cases} W_1(-s) + \dfrac{s}{2}, & s \leqslant 0, \\ \dfrac{\lambda_2}{\lambda_1} W_2(s) - \dfrac{s}{2}\left(\dfrac{1-\gamma+2\gamma\tau_0}{1+\gamma-2\gamma\tau_0}\right), & s > 0, \end{cases} \tag{3.3}$$

$W_1(s)$ and $W_2(s)$ are mutually independent standard Brownian motion processes defined respectively on $[0,\infty)$, and

$$\begin{aligned} \lambda_1 &= (E\{[E[h(Y_{k_0+1}, Y_1) \mid Y_1] - Eh(Y_{k_0+1}, Y_1)] \\ &\quad - [E[h(Y_1, Y_2) \mid Y_1] - Eh(Y_1, Y_2)]\}^2)^{\frac{1}{2}}, \end{aligned} \tag{3.4}$$

$$\begin{aligned} \lambda_2 &= (E\{[E[h(Y_1, Y_{k_0+1}) \mid Y_{k_0+1}] - Eh(Y_1, Y_{k_0+1})] \\ &\quad - [E[h(Y_{k_0+1}, Y_{k_0+2}) \mid Y_{k_0+2}] - Eh(Y_{k_0+1}, Y_{k_0+2})]\}^2)^{\frac{1}{2}}. \end{aligned} \tag{3.5}$$

Remark 3.4. As commented in [16, Remark 2.3], λ_1 and λ_2 can be estimated using the observations $Y_1, Y_2, \ldots, Y_n$.

Remark 3.5. When Δ_n is a constant not depending on n, i.e., F_n does not vary with n, we can obtain the asymptotic distribution of $\hat{\tau}_n$ similarly as in [12, 13], which is a two-way random walk depending on the underlying distribution F_1, F_n and Δ_n in a quite intricate way.

In the following, we evaluate the MSE of $\hat{k}_n$ by applying the above theoretical results via an example. Let F_1 and F_n be respectively $N(0,1)$ and $N(\mu_0, 1)$ with $\mu_0 \neq 0$. For illustration purpose, $\omega_2(t;a)$ is chosen as the weight function. It is easy to derive that

$$\Delta_n = \frac{2\sqrt{a}}{\sqrt{2+a}}\left[1 - \exp\left\{-\frac{\mu_0^2}{4(2+a)^2}\right\}\right]. \tag{3.6}$$

To find the MSE of $\hat{k}_n$, we need first to calculate both λ_1 and λ_2 in order to use (3.3) which, by (3.4) and (3.5), are equal and have the following expression,

$$\begin{aligned} \lambda_1^2 &= \lambda_2^2 \\ &= \frac{a}{\sqrt{a^2+2a+2}}\left[1 + \exp\left\{-\frac{\mu_0^2}{2(a^2+2a+2)}\right\} - 2\exp\left\{-\frac{(2a^2+4a+3)\mu_0^2}{8(1+a)^2(a^2+2a+2)}\right\}\right] \\ &\quad - \frac{2a\sqrt{1+a}}{\sqrt{2+a}\sqrt{2a^2+4a+3}}\left[1 - \exp\left\{-\frac{\mu_0^2}{4(2+a)^2}\right\}\right]\left[1 - \exp\left\{-\frac{\mu_0^2}{2(2a^2+4a+3)}\right\}\right] \\ &\quad + \frac{a}{2+a}\left[1 - \exp\left\{-\frac{\mu_0^2}{4(2+a)^2}\right\}\right]^2. \end{aligned} \tag{3.7}$$

By Theorem 3.3,

$$E_a(\hat{k}_n - k_0)^2 \approx \frac{1}{(1+\gamma-2\gamma\tau_0)^4}\frac{\lambda_1^4}{\Delta_n^4} E\Big[\arg\max_{-\infty<s<+\infty} G(s)\Big]. \tag{3.8}$$

Note that in this example, $G(s)$ does not depend on λ_1. By (3.6) and (3.7),

$$\frac{\lambda_1^2}{\Delta_n^2} \approx \frac{1}{4} - \frac{(1+a)^{\frac{1}{2}}(2+a)^{\frac{5}{2}}}{(2a^2+4a+3)^{\frac{3}{2}}} + \frac{(2+a)^5}{(a+1)^2(a^2+2a+2)^{\frac{3}{2}}\mu_0^2},$$

which, jointly with (3.8), implies that $E_a(\hat{k}_n - k_0)^2$ depends on μ_0 and the tuning parameter a via λ_1^2/Δ_n^2. Thus for a given μ_0, an appropriate value of a should be chosen in order to have a small MSE of $\hat{k}_n$, which will be dealt with in the next section.

4 An algorithm for selecting an appropriate tuning parameter a

The example given in Section 3 shows that it is important to select an appropriate a. We now present a real data example to further demonstrate how the change point estimate can be affected by the choice of a. Consider the Nile data, a time series of the annual flow of the river Nile at Aswan from 1871 to 1970 (see [2,7,9]), which has a change in year 1898 corresponding to the 28th observation in the data sequence detected in [25]. The data is depicted in Figure 1. For the purpose of illustration, we assume that the observations are independent as in [7]. We use (2.3) with respective weight functions $\omega_1(t;a)$, $\omega_2(t;a)$, and $\omega_3(t;a)$ for different values of a to estimate the change point. The resulted change point estimates are reported in Table 1.

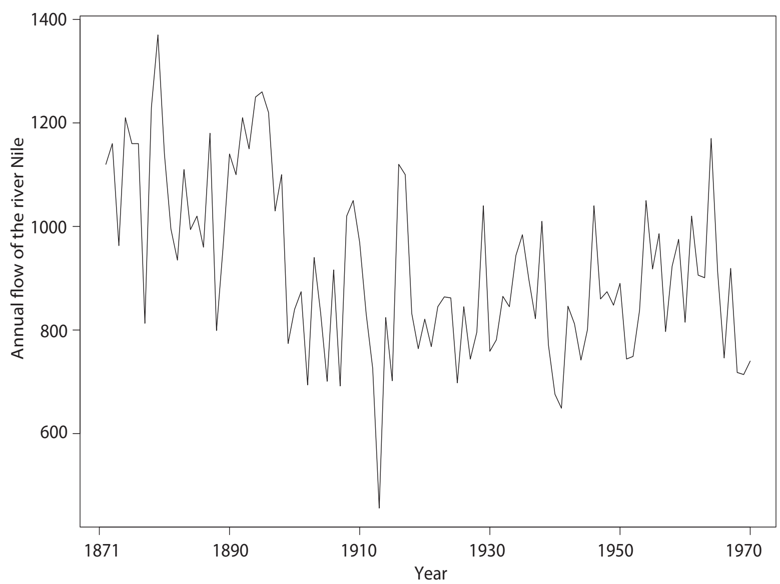

Figure 1 The Nile data

Table 1 The estimated change point $\hat{k}_n$ using different weight function $\omega(t;a)$ with different values of a and a fixed $\gamma = 0.5$

$\omega_1(t;a)$	a	1	2	3	4	5	6	7	$\cdots$	100
	$\hat{k}_n$	47	48	48	48	48	28	28	$\cdots$	28
$\omega_2(t;a)$	a	1	2	3	$\cdots$	22	23	24	$\cdots$	100
	$\hat{k}_n$	48	48	48	48	48	28	28	$\cdots$	28
$\omega_3(t;a)$	a	0.001	0.002	$\cdots$	0.009	0.01	0.02	0.03	$\cdots$	2
	$\hat{k}_n$	47	47	$\cdots$	48	28	28	28	$\cdots$	28

It can be seen from Table 1 that the value of a has a large impact on the accuracy of the change point estimate. An inappropriate a may result in a misleading estimate. In practice, we have no information about the change point in a given data sequence. However, a needs to be prechosen in order to find the change point estimate by (2.3). As shown above, different values of a might result in different change point estimates. Thus it is important to select a value from a set of possible values of a such that the resulted change point estimate has a satisfactory performance. Such an appropriate choice of a is denoted as a_{s} in this paper, where the subscript "s" is taken from the first letter of "selection". We propose the following algorithm for finding a_{s}.

Step 1. Let $Y_1, Y_2, \ldots, Y_{k_0}, Y_{k_0+1}, \ldots, Y_n$ be a given data sequence with the change point located at k_0 and

$$\mathcal{A} = \{a_1, a_2, \ldots, a_\ell\}$$

be a set of possible values for a. For each a_i from the set $\mathcal{A}$, we obtain

$$\hat{k}_{a_i} = \arg\max_k T_{\gamma,w}(k).$$

Table 2 Number of the change point estimate $\hat{k}_a$ fell into the interval $(k_0 - \delta, k_0 + \delta)$ for $\delta = 5$ (top entry), 10 (middle entry) and 15 (bottom entry) by using the weight function ω_1 (upper part), ω_2 (middle part) and ω_3 (lower part) when F_1 is $N(0,1)$ and F_n is $N(1,1)$

	a	1	2	3	4	5	6	7	8	9	10	11	a_{s}
	$k_0 = 30$	706	733	743	752	755	761	762	762	760	762	761	752
		873	899	904	907	905	908	909	909	907	908	909	906
		927	941	944	949	951	951	951	951	950	951	951	951
ω_1	$k_0 = 50$	725	749	763	771	772	770	770	772	772	772	773	773
		895	915	928	931	934	931	930	930	930	930	930	935
		964	970	970	971	972	973	973	973	973	973	973	972
	$k_0 = 70$	691	730	742	744	742	744	745	745	745	745	745	740
		856	872	879	887	888	891	891	891	891	891	891	888
		926	935	942	944	942	942	940	937	937	936	937	942
	$k_0 = 30$	734	745	753	754	762	761	762	762	762	760	760	760
		899	906	906	905	909	908	909	909	909	908	907	909
		940	948	949	950	952	951	951	951	951	950	950	952
ω_2	$k_0 = 50$	754	765	769	772	770	771	768	770	772	772	772	770
		915	927	930	934	932	931	930	930	930	930	930	932
		970	969	971	972	972	972	973	973	973	973	973	972
	$k_0 = 70$	725	741	741	742	743	744	745	746	745	745	745	744
		868	880	885	887	890	891	891	892	891	890	890	891
		933	944	942	942	941	940	939	937	937	937	937	942
	a	0.2	0.4	0.5	0.7	0.9	1.1	1.3	1.5	1.6	1.8	2	a_{s}
	$k_0 = 30$	676	703	710	721	731	738	747	740	748	750	745	733
		825	847	854	864	867	874	881	880	880	884	889	872
		894	911	915	919	919	926	934	936	933	937	936	928
ω_3	$k_0 = 50$	773	775	786	792	796	797	802	803	805	807	809	799
		931	934	938	946	950	953	952	950	950	951	952	954
		976	974	974	976	977	980	981	980	981	981	980	981
	$k_0 = 70$	680	706	715	718	721	733	745	748	742	743	744	734
		836	850	861	866	870	878	886	890	891	891	892	879
		912	925	934	937	938	942	945	950	949	949	950	944

Table 3 Number of the change point estimate $\hat{k}_a$ fell into the interval $(k_0-\delta, k_0+\delta)$ for $\delta=5$ (top entry), 10 (middle entry) and 15 (bottom entry) by using the weight function ω_1 (upper part), ω_2 (middle part) and ω_3 (lower part) when F_1 is $N(0,1)$ and F_n is $N(1,2)$

	a	1	2	3	4	5	6	7	8	9	10	11	$a_{\rm s}$
	$k_0=30$	636	643	630	613	599	580	566	549	542	533	528	600
		819	824	809	794	784	771	753	735	729	719	711	777
		890	895	878	863	850	841	825	812	805	797	790	848
ω_1	$k_0=50$	727	735	717	706	688	667	653	648	638	625	615	686
		895	904	890	881	867	853	846	835	828	814	806	879
		953	956	955	948	938	932	925	919	915	909	901	942
	$k_0=70$	730	762	767	755	742	724	714	703	681	675	673	743
		901	921	915	902	887	873	867	856	848	842	840	896
		950	962	963	952	945	942	936	929	921	917	917	947
	$k_0=30$	647	638	624	610	596	580	578	568	560	550	546	596
		827	817	803	791	778	771	766	754	745	737	735	781
		897	887	868	858	846	841	836	825	821	813	811	848
ω_2	$k_0=50$	740	718	712	701	683	680	661	651	651	644	644	682
		907	894	885	875	865	862	851	845	843	835	832	865
		958	959	950	943	938	936	930	926	925	920	919	939
	$k_0=70$	767	769	763	752	737	727	721	712	706	698	691	738
		925	920	908	899	886	876	874	866	859	856	853	890
		964	965	958	950	946	944	942	936	934	929	926	947
	a	0.2	0.4	0.5	0.7	0.9	1.1	1.3	1.5	1.6	1.8	2.0	$a_{\rm s}$
	$k_0=30$	617	636	629	617	609	588	564	551	525	502	470	580
		789	799	789	779	775	755	733	716	689	666	633	753
		867	868	857	849	844	831	811	796	772	745	716	831
ω_3	$k_0=50$	749	743	739	725	708	701	687	670	642	605	567	698
		910	905	903	885	877	871	856	849	822	786	752	869
		966	960	958	946	939	934	921	915	896	870	843	936
	$k_0=70$	743	757	762	762	760	748	728	717	686	665	636	753
		884	900	903	906	905	900	891	881	859	841	810	902
		938	948	949	954	957	954	954	949	928	916	898	956

Step 2. Compute the mean of $\hat{k}_{a_i}$, $i=1,2,\ldots,\ell$ as

$$\bar{\hat{k}} = \frac{1}{\ell}\sum_{i=1}^{\ell}\hat{k}_{a_i}.$$

Then

$$a_{\rm s} = \arg\min_{a_i} |\hat{k}_{a_i} - \bar{\hat{k}}|.$$

From the proposed algorithm, it can be seen that $a_{\rm s}$ is dependent on the data sequence and hence random. $a_{\rm s}$ might not give us the best change point estimate but it will provide an improved performance over a fixed one, which is not only justified in Proposition 4.1, but also confirmed by the simulation study in the next section.

Proposition 4.1. *Given a data sequence $Y_1, Y_2, \ldots, Y_{k_0}, Y_{k_0+1}, \ldots, Y_n$ with the change point located at k_0 and $\mathcal{A} = \{a_1, a_2, \ldots, a_\ell\}$ be a set of possible values for a. Then there exists at least one point $a^* \neq a_s$ in $\mathcal{A}$ such that $|\hat{k}_{a_s} - k_0| \leqslant |\hat{k}_{a^*} - k_0|$.*

The proof is given in Appendix.

Table 4 Number of the change point estimate $\hat{k}_a$ fell into the interval $(k_0-\delta, k_0+\delta)$ for $\delta = 5$ (top entry), 10 (middle entry) and 15 (bottom entry) by using the weight function ω_1 (upper part), ω_2 (middle part) and ω_3 (lower part) when F_1 is $L(0,1)$ and F_n is $L(1,1)$, the distribution of $Y+1$ with $Y \sim L(0,1)$

	a	1	2	3	4	5	6	7	8	9	10	11	a_{s}
	$k_0=30$	676	680	667	652	646	639	630	622	616	614	613	640
		830	841	834	830	823	819	812	805	801	798	798	823
		896	906	900	900	897	893	889	882	880	877	876	900
ω_1	$k_0=50$	702	718	702	689	682	674	673	667	664	663	659	687
		885	890	880	868	862	855	853	851	847	846	844	870
		952	947	938	935	934	930	933	930	925	925	923	942
	$k_0=70$	658	670	666	670	662	653	653	649	643	639	634	661
		829	835	827	822	818	814	813	811	804	801	801	820
		904	903	896	895	895	889	887	885	881	878	879	899
	$k_0=30$	674	665	646	645	642	638	635	629	623	621	618	639
		835	836	824	823	819	815	813	808	804	803	801	819
		902	901	898	900	895	891	890	887	883	881	879	896
ω_2	$k_0=50$	708	697	696	687	678	677	674	668	670	666	666	681
		882	880	868	865	859	856	855	853	852	848	848	862
		941	938	930	932	932	931	931	931	932	927	924	935
	$k_0=70$	658	663	665	664	660	653	650	651	648	646	647	655
		824	825	821	815	816	812	810	810	807	806	806	814
		898	895	891	893	893	888	887	884	883	883	883	893
	a	0.2	0.4	0.5	0.7	0.9	1.1	1.3	1.5	1.6	1.8	2.0	a_{s}
	$k_0=30$	657	673	670	664	670	667	659	647	634	615	586	661
		813	829	826	824	829	825	820	811	801	786	763	824
		885	903	899	894	900	892	888	881	875	857	841	892
ω_3	$k_0=50$	719	717	711	700	685	679	663	650	639	630	610	682
		907	895	886	879	866	859	847	834	822	814	795	864
		960	950	949	946	937	938	934	924	918	911	896	940
	$k_0=70$	656	676	683	679	671	666	651	639	626	605	585	664
		829	841	842	841	842	839	840	831	819	802	789	839
		897	902	911	913	912	910	910	900	890	878	867	909

5 Simulation studies

In this section, we carry out a simulation study to investigate the performance of $\hat{k}_n$ obtained via (2.3) when using different values of a including a_{s} in terms of accuracy of the change point estimate. In addition, we apply (2.3) with $a = a_{\mathrm{s}}$ to the Nile data.

5.1 Simulation study

We perform a simulation study to compare the change point estimate obtained via (2.3) using a set of fixed values of a and a_{s}. The following is the details of the simulation study.

(1) Generate data $Y_1, Y_2, \ldots, Y_{k_0}$ from the distribution F_1 and $Y_{k_0+1}, \ldots, Y_n$ from the distribution F_n with one change point located at $k_0 = 30, 50$, or 70, where $n = 100$. Three cases of F_1 are considered:

Case 1. The normal distribution $N(0,1)$.

Case 2. The laplace distribution $L(0,1)$.

Table 5 Number of the change point estimate $\hat{k}_a$ fell into the interval $(k_0-\delta, k_0+\delta)$ for $\delta=5$ (top entry), 10 (middle entry) and 15 (bottom entry) by using the weight function ω_1 (upper part), ω_2 (middle part) and ω_3 (lower part) when F_1 is $L(0,1)$ and F_n is $L(1,\sqrt{2})$, the distribution of $\sqrt{2}Y+1$ with $Y \sim L(0,1)$

	a	1	2	3	4	5	6	7	8	9	10	11	$a_{\rm s}$
	$k_0=30$	587	594	575	568	560	548	533	528	517	506	501	552
		773	775	763	754	739	733	715	710	704	699	689	740
		859	860	846	837	827	820	808	801	796	786	776	831
ω_1	$k_0=50$	676	677	667	653	647	642	624	620	605	602	591	655
		872	860	854	847	839	832	819	809	800	797	788	844
		945	937	928	921	915	912	907	897	889	882	876	921
	$k_0=70$	633	642	637	636	633	618	610	609	601	596	594	632
		814	818	815	811	808	793	780	775	772	770	768	810
		896	896	896	895	889	876	865	863	861	863	858	891
	$k_0=30$	590	582	569	563	556	556	544	534	533	529	524	557
		766	765	755	744	736	734	728	716	717	713	710	737
		854	849	834	830	824	822	818	811	809	806	801	828
ω_2	$k_0=50$	666	671	662	648	646	639	628	626	621	617	614	653
		860	857	848	840	837	831	821	818	814	809	803	840
		937	931	921	913	912	911	907	907	903	898	893	919
	$k_0=70$	623	626	632	633	625	622	617	613	609	608	607	627
		812	809	806	808	798	793	789	782	781	777	776	805
		893	891	891	891	880	878	875	869	865	864	865	887
	a	0.2	0.4	0.5	0.7	0.9	1.1	1.3	1.5	1.6	1.8	2.0	$a_{\rm s}$
	$k_0=30$	597	611	606	597	587	571	547	529	508	489	466	562
		759	768	765	751	748	735	713	693	678	660	638	728
		840	844	847	840	839	822	801	782	764	746	731	816
ω_3	$k_0=50$	692	690	667	656	639	618	603	583	559	526	504	620
		879	876	853	842	823	804	783	768	742	709	686	808
		952	945	933	923	915	906	895	885	864	836	816	910
	$k_0=70$	627	649	649	657	650	639	623	611	581	552	526	637
		805	818	817	825	822	817	810	800	772	755	725	816
		888	900	903	905	902	900	894	885	868	850	823	902

Case 3. The gamma distribution $G(1,1)$. Correspondingly, we consider $F_n(x)=F_1((x-b)/d)$ for $b=1$, and $d=1$ or $\sqrt{2}$.

(2) For a chosen weight function $\omega(t;a)$ and a given set of possible values of a, say $\mathcal{A}$, first execute the Step 1 of the algorithm given in Section 4 and obtain $\{\hat{k}_a,\ a\in\mathcal{A}\}$, and then execute the Step 2 of this algorithm to obtain $a_{\rm s}$. Compute the change point estimate $\hat{k}_{a_{\rm s}}$.

(3) Repeat (1)–(2) for 1,000 times and then compute the number of times that the change point estimate falls into the interval $[k_0-\delta, k_0+\delta]$ for $\delta=5,10,15$.

In this simulation study, γ is set as 0.5, and $\mathcal{A}$ is chosen as $\{1,2,3,\ldots,15\}$ for both ω_1 and ω_2 but $\{0.2,0.4,\ldots,2\}$ for ω_3. The simulation results are reported in Tables 2–7, which show that the value of a has a large impact on the accuracy of the change point estimate for all three weight functions. From these tables, it can been seen that the change point estimate obtained by using $a_{\rm s}$ always outperforms the change point estimates obtained by using some values of a, and has the best performance in some cases. It can also be observed that the weight function ω_3 performed better than both ω_1 and ω_2 in terms of the accuracy of change point estimation overall.

Table 6 Number of the change point estimate $\hat{k}_a$ fell into the interval $(k_0 - \delta, k_0 + \delta)$ for $\delta = 5$ (top entry), 10 (middle entry) and 15 (bottom entry) by using the weight function ω_1 (upper part), ω_2 (middle part) and ω_3 (lower part) when F_1 is $G(1,1)$ and F_n is $G(4, \frac{1}{2})$, the distribution of $Y+1$ with $Y \sim G(1,1)$

	a	1	2	3	4	5	6	7	8	9	10	11	a_{s}
	$k_0 = 30$	957	925	898	873	857	841	824	815	809	808	802	864
		993	985	973	958	949	940	934	930	926	924	920	949
		999	994	990	986	981	974	969	967	964	963	960	979
ω_1	$k_0 = 50$	929	903	889	870	853	838	826	821	816	817	815	862
		981	975	974	969	965	960	953	950	944	944	943	970
		997	994	991	991	990	987	983	980	977	977	976	991
	$k_0 = 70$	845	840	835	825	809	804	794	786	784	778	775	816
		939	938	936	931	919	914	905	900	898	896	892	919
		973	973	973	970	962	957	952	946	946	942	939	964
	$k_0 = 30$	920	887	873	857	845	833	822	817	814	809	808	844
		981	967	959	949	942	937	932	931	928	926	925	941
		993	988	985	981	976	972	967	967	965	964	963	976
ω_2	$k_0 = 50$	895	880	860	847	839	832	825	821	822	815	815	839
		973	970	967	966	960	956	953	950	951	944	944	964
		990	991	990	990	987	985	984	982	981	977	977	990
	$k_0 = 70$	835	830	821	809	805	800	794	788	785	783	784	808
		937	939	930	920	914	911	905	900	898	898	900	915
		970	974	969	963	957	956	953	947	945	944	946	958
	a	0.2	0.4	0.5	0.7	0.9	1.1	1.3	1.5	1.6	1.8	2.0	a_{s}
	$k_0 = 30$	956	953	946	930	922	907	874	849	803	767	724	906
		995	995	992	988	985	979	965	955	937	919	894	979
		998	998	996	996	995	994	991	989	977	965	945	994
ω_3	$k_0 = 50$	917	918	915	913	908	899	881	864	844	826	795	898
		982	981	979	978	975	973	971	965	956	948	939	974
		998	996	996	994	992	992	989	987	984	983	979	993
	$k_0 = 70$	831	841	841	835	829	829	818	815	797	786	777	832
		929	937	937	933	925	920	916	907	895	891	879	923
		972	977	977	976	971	969	965	958	950	948	934	969

The selection of $\mathcal{A}$ for ω_1, ω_2, and ω_3 is important for the performance of the change point estimation. The selection of $\mathcal{A}$ for the weight function ω_3 can simply be chosen as the equally spaced points between 0 to 2 which is the domain of a in $\omega_3(t;a)$. By [15], the role of the tuning parameter a is to control the rate of decay of the weight function ω_1 and ω_2. Thus in our simulation studies, we have only presented the simulation results for using $a \leqslant 11$. As a matter of fact, the accuracy of the change point estimate using $a > 11$ is almost the same as the one using $a = 11$ for the weight function being ω_1 or ω_2, and the change point estimates using either ω_1 or ω_2 perform similarly when a goes to infinity. From this experience and the effect of a on the weight function in theory, we recommend to increase the value of a from 1 by 1 each time to estimate the change point until the change point estimate remains the same. Then $\mathcal{A}$ can be chosen as a collection of all the values of a that has been tried.

We also conduct a simulation study to compare the empirical distribution of the change point estimator $\hat{\tau}_n$ defined in (2.8) with its asymptotic distribution. In light of Appendix B in [1] or [16], the cumulative distribution function $H(s)$ of

$$\arg\max_s G(s)$$

has the following expression:

Table 7 Number of the change point estimate $\hat{k}_a$ fell into the interval $(k_0-\delta, k_0+\delta)$ for $\delta=5$ (top entry), 10 (middle entry) and 15 (bottom entry) by using the weight function ω_1 (upper part), ω_2 (middle part) and ω_3 (lower part) when F_1 is $G(1,1)$ and F_n is $G(\frac{3+2\sqrt{2}}{2}, 2\sqrt{2}-2)$, the distribution of $\sqrt{2}Y+1$ with $Y \sim G(1,1)$

	a	1	2	3	4	5	6	7	8	9	10	11	a_{s}
	$k_0=30$	952	941	909	890	872	857	842	834	827	823	819	875
		991	987	974	965	955	948	941	935	931	929	929	953
		997	994	990	985	979	975	971	968	965	963	962	979
ω_1	$k_0=50$	937	931	912	902	894	885	877	873	868	866	866	904
		983	985	982	981	979	976	974	972	971	969	969	985
		996	996	996	996	995	993	993	992	990	989	989	996
	$k_0=70$	863	870	875	876	872	871	867	864	864	866	866	882
		950	957	958	962	957	958	956	956	957	957	956	961
		981	983	984	987	986	989	986	987	987	988	988	990
	$k_0=30$	931	906	888	873	865	856	846	840	836	833	831	863
		986	972	965	956	952	947	943	938	935	935	933	951
		993	990	986	980	978	974	972	970	968	968	967	978
ω_2	$k_0=50$	918	906	896	892	887	885	878	874	874	870	868	890
		982	981	978	979	978	975	972	972	972	971	971	980
		994	995	995	995	993	992	992	992	992	991	991	993
	$k_0=70$	861	876	876	871	868	868	866	866	864	865	865	869
		954	960	962	958	956	956	954	957	956	956	956	958
		981	986	987	987	986	987	984	987	987	987	986	986
	a	0.2	0.4	0.5	0.7	0.9	1.1	1.3	1.5	1.6	1.8	2.0	a_{s}
	$k_0=30$	947	944	934	920	916	893	868	833	806	783	748	898
		993	990	987	987	983	973	963	945	932	923	898	975
		996	996	995	995	994	992	990	981	973	968	946	993
ω_3	$k_0=50$	936	940	939	936	932	920	917	912	894	882	860	928
		988	989	992	990	988	987	988	984	978	975	970	989
		999	999	999	998	998	997	996	997	997	996	995	998
	$k_0=70$	861	870	879	879	885	886	884	881	874	872	871	887
		951	954	956	952	957	955	953	955	953	950	950	957
		987	987	987	985	985	982	981	981	978	978	981	983

Table 8 The quantiles of the empirical (ED) and asymptotic (AD) distributions of $\hat{\tau}_n$

Sample size	Quantiles	5%	10%	50%	90%	95%
300	ED	0.473	0.483	0.500	0.520	0.533
	AD	0.490	0.494	0.500	0.506	0.510
500	ED	0.486	0.490	0.500	0.510	0.514
	AD	0.494	0.496	0.500	0.504	0.506

$$H(s)=\begin{cases} -\dfrac{1}{\sqrt{2\pi}}|s|^{\frac{1}{2}}\exp\left\{-\dfrac{|s|}{8}\right\}-c_1\exp\{a_1|s|\}\Phi(-b_1|s|^{\frac{1}{2}})+\left(d_1-2+\dfrac{1}{2}|s|\right)\Phi\left(-\dfrac{1}{2}|s|^{\frac{1}{2}}\right), & s<0,\\ 1+\dfrac{\theta_0}{\lambda_0}\dfrac{1}{\sqrt{2\pi}}s^{\frac{1}{2}}\exp\left\{-\dfrac{\theta_0^2}{8\lambda_0}s\right\}+c_2\exp\{a_2 s\}\Phi(-b_2 s^{\frac{1}{2}}) & \\ \quad -\left(d_2-2+\dfrac{\theta_0^2}{2\lambda_0}s\right)\Phi\left(-\dfrac{\theta_0}{2\lambda_0}s^{\frac{1}{2}}\right), & s\geqslant 0,\end{cases}$$

Table 9 The true values and estimates of the parameters

	True value			Mean of estimates		
Sample size	τ_0	Δ_n	$\lambda_1^2 = \lambda_2^2$	$\hat{\tau}_n$	$\widehat{\Delta}_n$	$\widehat{\lambda}_1^2 = \widehat{\lambda}_2^2$
300	0.5	0.032	0.020	0.501	0.031	0.029
500	0.5	0.032	0.020	0.500	0.030	0.029

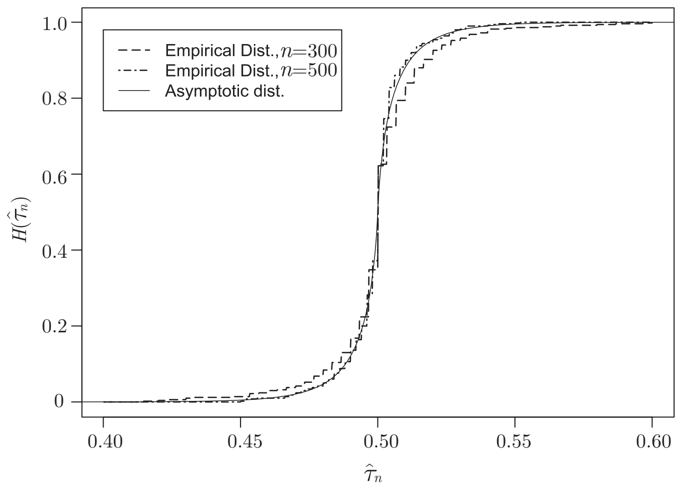

Figure 2 The empirical cumulative distributions of the change point estimator $\hat{\tau}_n$ with $n = 300$, $n = 500$, and the asymptotic cumulative distribution $H(\cdot)$

where $\lambda_0 = \frac{\lambda_2}{\lambda_1}$, $\theta_0 = \frac{1-\gamma+2\gamma\tau_0}{1+\gamma-2\gamma\tau_0}$, $\alpha = \frac{\theta_0}{\lambda_0^2}$, $\beta = \frac{\theta_0}{\lambda_0}$, $a_1 = \frac{\theta_0}{2\lambda_0^2}(1 + \frac{\theta_0}{\lambda_0^2}), b_1 = \frac{1}{2} + \frac{\theta_0}{\lambda_0^2}$, $c_1 = \frac{\lambda_0^2(\lambda_0^2+2\theta_0)}{\theta_0(\lambda_0^2+\theta_0)}$, $d_1 = \frac{(\lambda_0^2+2\theta_0)^2}{\theta_0(\lambda_0^2+\theta_0)}$, $a_2 = \frac{\lambda_0^2+\theta_0}{2}$, $b_2 = \frac{2\lambda_0^2+\theta_0}{2\lambda_0}$, $c_2 = \frac{\theta_0(2\lambda_0^2+\theta_0)}{\lambda_0^2(\lambda_0^2+\theta_0)}$, and $d_2 = \frac{(2\lambda_0^2+\theta_0)^2}{\lambda_0^2(\lambda_0^2+\theta_0)}$. Here, λ_1 and λ_2 are defined in (3.4) and (3.5), respectively.

For the purpose of demonstration, we consider the case that F_1 and F_n are respectively $N(0,1)$ and $N(\mu_0,1)$ with $\mu_0 = 1$. For simple presentation, we only consider the weight function $\omega_2(t;a)$. Let the tuning parameter a be 1, and γ be 0.5. The sample size n is set as 300 and 500, respectively, and τ_0 is chosen as 0.5, which implies that the true change point is located at $k_0 = 150$ for $n = 300$ and at $k_0 = 250$ for $n = 500$. The true values of Δ_n, λ_1^2 and λ_2^2 are calculated using (3.6) and (3.7) respectively and are shown in Table 9.

We generate 500 samples for each parameter setting and each sample size. For each sample generated, we find the change point estimate $\hat{\tau}_n$. Then we compute the quantiles of the empirical distribution of $\hat{\tau}_n$ that are shown in Table 8. The means of the estimates for Δ_n, λ_1 and λ_2 based on 500 samples are given in Table 9, which shows that they are very close to their true values. Similar to [16], we obtain the quantiles of the asymptotic distribution of $\hat{\tau}_n$ that are displayed in Table 8. The quantiles of both empirical and asymptotic distributions of $\hat{\tau}_n$ are very close to each other. For graphical comparison, we display the empirical cumulative distribution of $\hat{\tau}_n$ with $n = 300$, or $n = 500$ and its asymptotic cumulative distribution in Figure 2. It further confirms the good approximation of the empirical cumulative distribution of $\hat{\tau}_n$ to its asymptotic distribution.

5.2 A real data example

In this subsection, we revisit the Nile data discussed in Section 4. We employ all three weight functions with a_{s} chosen from $\{1, 2, \ldots, 100\}$ for both ω_1 and ω_2 but $\{0.2, 0.4, \ldots, 2\}$ for ω_3. We set γ to be either $0, 0.5$, or 1. They have all detected that the change point is located at the 28th observation, corresponding to the year 1898, which is the same as that detected in [25].

Acknowledgements This work was supported by Natural Sciences and the Engineering Research Council of Canada (Grant No. 105557-2012), National Natural Science Foundation for Young Scientists of China (Grant No. 11201108), the National Statistical Research Plan Project (Grant No. 2012LZ009) and the Humanities and Social Sciences Project from Ministry of Education of China (Grant No. 12YJC910007). This paper is dedicated to the memory of Professor Xiru Chen. The corresponding author is greatly indebted to him for his invaluable advice, guidance, inspiration and encouragement.

References

1 Bai J. Estimation of a change point in multiple regression models. Rev Econ Stat, 1997, 79: 551–63
2 Balke N S. Detecting level shifts in time series. J Bus Econom Statist, 1993, 11: 81–92
3 Brodskij B E, Darchovskij B S. Non-parametric Statistical Diagnosis: Problems and Methods. Dordrecht: Kluwer Academic Publishers, 2000
4 Carlstein E. Nonparametric change point estimation. Ann Stat, 1988, 16: 188–197
5 Chen J, Gupta A K. Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance. New York: Springer Science & Business Media, 2011
6 Chen X R. Testing and interval estimation in a change-point model allowing at most one change. Sci China Ser A, 1988, 30: 817–827
7 Cobb G W. The problem of the Nile: Conditional solution to a change point problem. Biometrika, 1978, 65: 243–251
8 Csörgő M, Horváth L. Limit Theorems in Change-Point Analysis. New York: John Wiley & Sons Inc, 1997
9 Dumbgen L. The asymptotic behavior of some nonparametric change point estimators. Ann Stat, 1991, 19: 1471–1495
10 Epps T W. Limiting behaviour of the ICF test for normality under Gram-Charlier alternatives. Statist Probab Lett, 1999, 42: 175–184
11 Feuerverger A, Mureika R A. The empirical characteristic function and its applications. Ann Stat, 1977, 5: 88–97
12 Gombay E. U-statistics for change under alternatives. J Multivariate Anal, 2001, 78: 139–158
13 Hinkley D. Inference about the change-point in a sequence of random variables. Biometrika, 1970, 57: 1–17
14 Holmes M, Kojadinovic I, Quessy J. Nonparametric tests for change-point detection á la Gombay and Horváth. J Multivar Anal, 2013, 115: 16–32
15 Hušková M, Meintanis S G. Change point analysis based on empirical characteristic functions. Metrika, 2006, 6: 145–168
16 Jin B S, Dong C L, Tan C C, et al. Estimator of a change point in single index models. Sci China Math, 2014, 57: 1701–1712
17 Kankainen A, Ushakov N G. A consistent modification of a test for independence based on the empirical characteristic function. J Math Sci, 1998, 89: 1486–1494
18 Kent J T. A weak convergence theorem for the empirical characteristic function. J Appl Probab, 1975, 12: 515–523
19 Koutrouvelis I A, Meintanis S G. Testing for stability based on the empirical characteristic function with applications to financial data. J Stat Comput Simul, 1999, 64: 275–300
20 Matteson D S, James N A. A nonparametric approach for multiple change point analysis of multivariate data. J Amer Statist Assoc, 2014, 109: 334–345
21 Parzen E. On estimation of a probability density function and mode. Ann Math Stat, 1962, 33: 1065–1076
22 Rafajlowicz E, Pawlak M, Steland A. Nonparametric sequential change-point detection by a vertically trimmed box method. IEEE Trans Inform Theory, 2010, 56: 3621–3634
23 Shi X P, Wu Y H, Miao B Q. Strong convergence rate of estimators of change point and its application. Comput Stat Data Anal, 2009, 53: 990–998
24 Ushakov N G. Selected Topics in Characteristic Functions. Berlin: Walter de Gruyter, 1999
25 Zeileis A, Kleiber C, Krämer W, et al. Testing and dating of structural changes in practice. Comput Statist Data Anal, 2003, 44: 109–123

Appendix

The proofs of the theorems in this paper are technically involved, so in order to give the idea, we focus on the main steps of the proofs only. Denote

$$\begin{aligned}
&\tilde{h}(Y_r, Y_s) = h(Y_r, Y_s) - E[h(Y_r, Y_s) \mid Y_r] - E[h(Y_r, Y_s) \mid Y_s] + E[h(Y_r, Y_s)],\\
&\overline{h}(Y_r, Z_1) = E[h(Y_r, Z_1) \mid Y_r] - E[h(Y_r, Z_1)],\\
&\overline{h}(Y_r, Z_2) = E[h(Y_r, Z_2) \mid Y_r] - E[h(Y_r, Z_2)], \qquad \text{(A.1)}
\end{aligned}$$

where Z_1 and Z_2 are independent of $Y_1, Y_2, \dots, Y_n$ and follow the distributions F_1 and F_n, respectively.

The following Hájiek-Rényi-Chow inequality is needed in the proofs of theorems.

Lemma A.1 (Hájek-Rényi-Chow inequality). *Suppose that $\{X_n, n \geqslant m\}, 1 \leqslant m \leqslant n$, is a martingale difference sequence. Let $\sigma_n^2 = EX_n^2$ and $c_1 \geqslant c_2 \geqslant \cdots \geqslant c_n > 0$. Define*

$$S_n = \sum_{j=1}^{n} X_j.$$

Then for any $x > 0$, we have

$$P\Big(\max_{m \leqslant j \leqslant n} c_j |S_j| \geqslant x\Big) \leqslant \frac{1}{x^2}\bigg[mc_m^2\sigma_m^2 + \sum_{j=m+1}^{n} c_j^2\sigma_j^2\bigg].$$

A.1 Proof of Theorem 3.1

To simplify the notation, $T_{\omega,\gamma}(k)$ is abbreviated by $T(k)$. Without loss of generality, we assume that $k = \lfloor n\tau \rfloor$. Since

$$T(k) \leqslant |T(k) - ET(k)| + ET(k),$$

and

$$ET(k_0) \leqslant |ET(k_0) - T(k_0)| + T(k_0),$$

by the triangle inequality, it is easy to show that

$$ET(k_0) - ET(k) \leqslant 2 \max_{1 \leqslant k < n} |T(k) - ET(k)| + T(k_0) - T(k). \tag{A.2}$$

Let

$$c_{k,n}(\gamma) = \bigg(\frac{k(n-k)}{n^2}\bigg)^{\gamma} \frac{k(n-k)}{n}, \quad k = 1, 2, \dots, n-1.$$

Then $T(k) = c_{k,n}(\gamma)Q_k$, where

$$Q_k = \frac{1}{k^2}\sum_{r,s=1}^{k} h(Y_r, Y_s) + \frac{1}{(n-k)^2}\sum_{r,s=k+1}^{n} h(Y_r, Y_s) - \frac{2}{k(n-k)}\sum_{r=1}^{k}\sum_{s=k+1}^{n} h(Y_r, Y_s). \tag{A.3}$$

For $k \leqslant k_0$, Q_k can be decomposed as follows:

$$\begin{aligned} Q_k = {} & \frac{1}{k^2}\sum_{r=1}^{k} h(Y_r, Y_r) + \frac{1}{(n-k)^2}\sum_{r=k+1}^{n} h(Y_r, Y_r) + \frac{1}{k^2}\sum_{r=1}^{k}\sum_{s=1, s\neq r}^{k} h(Y_r, Y_s) \\ & + \frac{1}{(n-k)^2}\bigg[\sum_{r=k+1}^{k_0}\sum_{s=k+1, s\neq r}^{k_0} + \sum_{r=k_0+1}^{n}\sum_{s=k_0+1, s\neq r}^{n} + 2\sum_{r=k+1}^{k_0}\sum_{s=k_0+1}^{n}\bigg] h(Y_r, Y_s) \\ & - \frac{2}{k(n-k)}\sum_{r=1}^{k}\bigg[\sum_{s=k+1}^{k_0} + \sum_{s=k_0+1}^{n}\bigg] h(Y_r, Y_s). \end{aligned}$$

So

$$\begin{aligned} EQ_k = {} & \frac{n}{k(n-k)}\int \omega(t)dt + \frac{(n-k_0)^2}{(n-k)^2}[E[h(Y_1, Y_2)] - 2E[h(Y_1, Y_{k_0+1})] + E[h(Y_{k_0+1}, Y_{k_0+2})]] \\ & + \bigg[\frac{k-k_0}{(n-k)^2} - \frac{1}{k}\bigg]E[h(Y_1, Y_2)] - \frac{n-k_0}{(n-k)^2}E[h(Y_{k_0+1}, Y_{k_0+2})], \end{aligned} \tag{A.4}$$

where

$$E[h(Y_1, Y_2)] = \int \bigg\{\bigg(\int \cos(tx)dF_1(x)\bigg)^2 + \bigg(\int \sin(tx)dF_1(x)\bigg)^2\bigg\}\omega(t)dt,$$

and

$$E[h(Y_{k_0+1}, Y_{k_0+2})] = \int \left\{ \left(\int \cos(tx) dF_n(x) \right)^2 + \left(\int \sin(tx) dF_n(x) \right)^2 \right\} \omega(t) dt.$$

Then we have, as $k \leqslant k_0$,

$$\begin{aligned} ET(k) - ET(k_0) = & \left[\left(\frac{k(n-k)}{n^2} \right)^\gamma - \left(\frac{k_0(n-k_0)}{n^2} \right)^\gamma \right] \int \omega(t) dt \\ & + \left[\left(\frac{k(n-k)}{n^2} \right)^\gamma \frac{k(n-k_0)}{n-k} - \left(\frac{k_0(n-k_0)}{n^2} \right)^\gamma k_0 \right] \frac{(n-k_0)}{n} \Delta_n \\ & + \left[\left(\frac{k(n-k)}{n^2} \right)^\gamma \left(\frac{k(k-k_0)}{n(n-k)} - \frac{n-k}{n} \right) + \left(\frac{k_0(n-k_0)}{n^2} \right)^\gamma \frac{n-k_0}{n} \right] E[h(Y_1, Y_2)] \\ & - \left[\left(\frac{k(n-k)}{n^2} \right)^\gamma \frac{k(n-k_0)}{n(n-k)} - \left(\frac{k_0(n-k_0)}{n^2} \right)^\gamma \frac{k_0}{n} \right] E[h(Y_{k_0+1}, Y_{k_0+2})]. \end{aligned} \tag{A.5}$$

It is easy to conclude that from (3.2) the second term is the dominating one in (A.5). Using the mean value theorem, we obtain that

$$ET(k) - ET(k_0) = g_1'(\xi_1)(\tau - \tau_0) n\Delta_n + o_p(n\Delta_n), \tag{A.6}$$

where $g_1'(\cdot)$ is the first order derivative of $g_1(\cdot)$ with

$$g_1(x) = (1-\tau_0)^2 x^{\gamma+1} (1-x)^{\gamma-1},$$

and $\tau \leqslant \xi_1 \leqslant \tau_0$. Similar arguments yield that, as $k > k_0$

$$ET(k) - ET(k_0) = g_2'(\xi_2)(\tau - \tau_0) n\Delta_n + o_p(n\Delta), \tag{A.7}$$

where $g_2'(\cdot)$ is the first order derivative of $g_2(\cdot)$ with

$$g_2(x) = \tau_0^2 x^{\gamma-1} (1-x)^{\gamma+1},$$

and $\tau_0 \leqslant \xi_2 \leqslant \tau$. Combining (A.2) and (A.5)–(A.7), we obtain that

$$\begin{aligned} n\Delta_n |\tau - \tau_0| \delta + o_p(n\Delta_n) & \leqslant ET(k_0) - ET(k) \\ & \leqslant 2 \max_{1 \leqslant k < n} |T(k) - ET(k)| + T(k_0) - T(k), \end{aligned} \tag{A.8}$$

where $\delta = \min\{g_1'(\xi_1), g_2'(\xi_2)\}$. Since $\hat{\tau}_n = \hat{k}_n/n$, $T(\hat{k}_n) \geqslant T(k_0)$, and T is nonnegative, by replacing τ by $\hat{\tau}_n$ in (A.8), we have

$$n\Delta_n |\hat{\tau}_n - \tau_0| \delta + o_p(n\Delta_n) \leqslant 2 \max_{1 \leqslant k < n} |T(k) - ET(k)|. \tag{A.9}$$

In order to show the consistency of change point estimator $\hat{\tau}_n$, we consider the probability $P(|\hat{\tau}_n - \tau_0| > \varepsilon)$, $\forall \varepsilon > 0$. It is easily to see from (A.9) that

$$\begin{aligned} P(|\hat{\tau}_n - \tau_0| > \varepsilon) \leqslant & P\left(\max_{1 \leqslant k < k_0} |T(k) - ET(k)| > \frac{n\varepsilon\delta\Delta_n}{2} \right) \\ & + P\left(\max_{k_0 < k < n} |T(k) - ET(k)| > \frac{n\varepsilon\delta\Delta_n}{2} \right). \end{aligned} \tag{A.10}$$

Because of the symmetry, we only show

$$P\left(\max_{1 \leqslant k \leqslant k_0} |T(k) - ET(k)| > \frac{n\varepsilon\delta\Delta_n}{2} \right) \to 0$$

as $n \to \infty$. The remaining part is analogous and thus is omitted.

We start with that

$$P\biggl(\max_{1\leqslant k\leqslant k_0}|T(k)-ET(k)|>\frac{n\varepsilon\delta\Delta_n}{2}\biggr).$$

If $k\leqslant k_0$, by (A.3),

$$T(k)-ET(k)=A_1+A_2+\cdots+A_{12}, \tag{A.11}$$

with

$$A_1=\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{1}{k}\sum_{r=1}^{k}\sum_{s=1,s\neq r}^{k}\tilde{h}(Y_r,Y_s),\quad A_2=\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{1}{n-k}\sum_{r=k+1}^{n}\sum_{s=k+1,s\neq r}^{n}\tilde{h}(Y_r,Y_s),$$

$$A_3=\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{1}{n}\sum_{r=1}^{n}\sum_{s=1,s\neq r}^{n}\tilde{h}(Y_r,Y_s),\quad A_4=\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{2(n-k_0)}{n}\sum_{r=1}^{k}\overline{h}(Y_r,Z_1),$$

$$A_5=-\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{2(n-k)}{nk}\sum_{r=1}^{k}\overline{h}(Y_r,Z_1),\quad A_6=-\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{2k(n-k_0)}{n(n-k)}\sum_{r=k+1}^{k_0}\overline{h}(Y_r,Z_1),$$

$$A_7=-\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{2k}{n(n-k)}\sum_{r=k+1}^{k_0}\overline{h}(Y_r,Z_1),\quad A_8=\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{2k(n-k_0)}{n(n-k)}\sum_{r=k_0+1}^{n}\overline{h}(Y_r,Z_2),$$

$$A_9=-\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{2k}{n(n-k)}\sum_{r=k_0+1}^{n}\overline{h}(Y_r,Z_2),\quad A_{10}=\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{2k(n-k_0)}{n(n-k)}\sum_{r=k+1}^{k_0}\overline{h}(Y_r,Z_2),$$

$$A_{11}=-\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{2k(n-k_0)}{n(n-k)}\sum_{r=k_0+1}^{n}\overline{h}(Y_r,Z_1),\quad A_{12}=-\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{2(n-k_0)}{n}\sum_{r=1}^{k}\overline{h}(Y_r,Z_2),$$

where Z_1 and Z_2 have the distribution functions F_1 and F_n, respectively, and are independent of $Y_1,Y_2,\ldots,Y_n$.

Next we investigate each term in (A.11). Towards this end, we consider the following statistics

$$S_k(\tilde{h})=\sum_{1\leqslant i<j\leqslant k}\tilde{h}(Y_i,Y_j),\quad k=1,2,\ldots,n,$$

where $\tilde{h}$ is defined in (A.1). Since $E[S_{k+1}(\tilde{h})\mid Y_1,Y_2,\ldots,Y_k]=S_k(\tilde{h})$ for $k=1,2,\ldots,n-1$, $\{S_k,\sigma(Y_1,\ldots,Y_k);k=1,2,\ldots,n\}$ is a martingale, where $\sigma(Y_1,\ldots,Y_k)$ denotes the σ-field generated by $Y_1,\ldots,Y_k$. Then by the Hájek-Rényi-Chow inequality

$$\begin{aligned}P\biggl(\max_{1\leqslant k\leqslant k_0}|A_1|>\frac{n\varepsilon\delta\Delta_n}{2}\biggr)&\leqslant P\biggl(\max_{1\leqslant k\leqslant k_0}\frac{|S_k(\tilde{h})|}{k^{1-\gamma}}>\frac{n^{1+\gamma}\varepsilon\delta\Delta_n}{4}\biggr)\\&\leqslant\frac{c}{n^2\varepsilon^2\delta^2\Delta_n^2}\biggl\{\frac{1+I_{\{\gamma=1/2\}}\log n}{n^{\min(2\gamma,1)}}\biggr\}\leqslant\frac{c}{n^2\varepsilon^2\delta^2\Delta_n^2}.\end{aligned}$$

Similar arguments yield that

$$P\biggl(\max_{1\leqslant k\leqslant k_0}|A_2|>\frac{n\varepsilon\delta\Delta_n}{2}\biggr)\leqslant\frac{c}{n^2\varepsilon^2\delta^2\Delta_n^2},$$

and

$$P\biggl(\max_{1\leqslant k\leqslant k_0}|A_3|>\frac{n\varepsilon\delta\Delta_n}{2}\biggr)\leqslant\frac{c}{n^2\varepsilon^2\delta^2\Delta_n^2}.$$

Since each of $\{E(h(Y_r,Z_1)\mid Y_r)-Eh(Y_r,Z_1),r=1,2,\ldots,k_0\}$, $\{E(h(Y_r,Z_1)\mid Y_r)-Eh(Y_r,Z_1),r=k_0+1,\ldots,n\}$, $\{E(h(Y_r,Z_2)\mid Y_r)-Eh(Y_r,Z_2),r=1,2,\ldots,k_0\}$, and $\{E(h(Y_r,Z_2)\mid Y_r)-Eh(Y_r,Z_2),r=k_0+1,\ldots,n\}$ is an identically distributed and independent sequence of random variables with zero mean and finite variance, the application of the Hájiek-Rényi-Chow inequality leads to

$$P\biggl(\max_{1\leqslant k\leqslant k_0}|A_4|>\frac{n\varepsilon\delta\Delta_n}{2}\biggr)\leqslant\frac{c}{n\varepsilon^2\delta^2\Delta_n^2},$$

$$P\left(\max_{1\leqslant k\leqslant k_0}|A_5|>\frac{n\varepsilon\delta\Delta_n}{2}\right)\leqslant\frac{c}{n^{2+2\gamma}\varepsilon^2\delta^2\Delta_n^2}\sum_{k=1}^m\frac{1}{k^{2-2\gamma}}\leqslant\frac{c}{n^2\varepsilon^2\delta^2\Delta_n^2}.$$

Similarly, we can obtain that

$$P\left(\max_{1\leqslant k\leqslant k_0}|A_6|>\frac{n\varepsilon\delta\Delta_n}{2}\right)\leqslant\frac{c}{n\varepsilon^2\delta^2\Delta_n^2},\quad P\left(\max_{1\leqslant k\leqslant k_0}|A_7|>\frac{n\varepsilon\delta\Delta_n}{2}\right)\leqslant\frac{c}{n^2\varepsilon^2\delta^2\Delta_n^2},$$
$$P\left(\max_{1\leqslant k\leqslant k_0}|A_8|>\frac{n\varepsilon\delta\Delta_n}{2}\right)\leqslant\frac{c}{n\varepsilon^2\delta^2\Delta_n^2},\quad P\left(\max_{1\leqslant k\leqslant k_0}|A_9|>\frac{n\varepsilon\delta\Delta_n}{2}\right)\leqslant\frac{c}{n^2\varepsilon^2\delta^2\Delta_n^2},$$
$$P\left(\max_{1\leqslant k\leqslant k_0}|A_{10}|>\frac{n\varepsilon\delta\Delta_n}{2}\right)\leqslant\frac{c}{n\varepsilon^2\delta^2\Delta_n^2},\quad P\left(\max_{1\leqslant k\leqslant k_0}|A_{11}|>\frac{n\varepsilon\delta\Delta_n}{2}\right)\leqslant\frac{c}{n\varepsilon^2\delta^2\Delta_n^2},$$
$$P\left(\max_{1\leqslant k\leqslant k_0}|A_{12}|>\frac{n\varepsilon\delta\Delta_n}{2}\right)\leqslant\frac{c}{n\varepsilon^2\delta^2\Delta_n^2}.$$

Thus, we have

$$P\left(\max_{1\leqslant k\leqslant k_0}|T(k)-ET(k)|>\frac{n\varepsilon\delta\Delta_n}{2}\right)\leqslant\frac{c_0}{\varepsilon^2\delta^2 n\Delta_n^2}.\tag{A.12}$$

By (3.2), (A.10) and (A.12), it follows that

$$\lim_{n\to\infty}P(|\hat{\tau}_n-\tau_0|>\varepsilon)=0,$$

i.e., $\hat{\tau}_n\to_P\tau_0$.

We now prove that

$$|\hat{\tau}_n-\tau_0|=O_P\left(\frac{1}{n\Delta_n^2}\right).$$

It follows from (A.4) that

$$\begin{aligned}&2n^2\Delta_n^2|\tau-\tau_0|\tilde{\delta}\\&\quad\leqslant(ET(k_0))^2-(ET(k))^2\\&\quad=\begin{cases}2n^2\Delta_n^2|\tau-\tau_0|\xi_1^{2\gamma+1}(1-\xi_1)^{2\gamma-3}(1-\tau_0)^4(1+\gamma-2\gamma\xi_1)+o(n^2\Delta_n^2), & k\leqslant k_0,\\ 2n^2\Delta_n^2|\tau-\tau_0|\xi_2^{2\gamma-3}(1-\xi_2)^{2\gamma+1}\tau_0^4(1-\gamma+2\gamma\xi_2)+o(n^2\Delta_n^2), & k>k_0,\end{cases}\end{aligned}\tag{A.13}$$

where $\tau<\xi_1<\tau_0$ for the case that $k<k_0$, and $\tau_0<\xi_2<\tau$ for the case that $k>k_0$, and

$$\tilde{\delta}=\min\{\xi_1^{2\gamma+1}(1-\xi_1)^{2\gamma-3}(1-\tau_0)^4(1+\gamma-2\gamma\xi_1),\xi_2^{2\gamma-3}(1-\xi_2)^{2\gamma+1}\tau_0^4(1-\gamma+2\gamma\xi_2)\}.$$

Thus the convergence rate of $\hat{\tau}_n$ may be found via the limiting behaviour of $(ET(k_0))^2-(ET(k))^2$. We decompose it as follows:

$$\begin{aligned}(ET(k_0))^2-(ET(k))^2&=[(T(k)-ET(k))-(T(k_0)-ET(k_0))]^2\\&\quad+2(T(k_0)-ET(k_0))(T(k)-ET(k)-(T(k_0)-ET(k_0)))\\&\quad+2(T(k)-ET(k))(ET(k)-ET(k_0))\\&\quad+2ET(k_0)(T(k)-ET(k)-(T(k_0)-ET(k_0)))\\&\quad+(T(k_0))^2-(T(k))^2.\end{aligned}\tag{A.14}$$

By the definition of $T(k)$, it can be decomposed as

$$T(k)-ET(k)=B_1+B_2+\cdots+B_6,$$

where

$$B_1=\left[\left(\frac{k(n-k)}{n^2}\right)^\gamma\frac{(n-k)}{nk}\right]\sum_{r\neq s,1}^k[h(Y_r,Y_s)-Eh(Y_r,Y_s)],$$

$$B_2 = \left[\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{k}{n(n-k)}\right]\sum_{r\neq s,k+1}^{k_0}[h(Y_r,Y_s)-Eh(Y_r,Y_s)],$$

$$B_3 = \left[\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{k}{n(n-k)}\right]\sum_{r\neq s,k_0+1}^{n}[h(Y_r,Y_s)-Eh(Y_r,Y_s)],$$

$$B_4 = -\left[\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{2}{n}\right]\sum_{r=1}^{k}\sum_{s=k+1}^{k_0}[h(Y_r,Y_s)-Eh(Y_r,Y_s)],$$

$$B_5 = -\left[\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{2}{n}\right]\sum_{r=1}^{k}\sum_{s=k_0+1}^{n}[h(Y_r,Y_s)-Eh(Y_r,Y_s)],$$

$$B_6 = \left[\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{2k}{n(n-k)}\right]\sum_{r=k+1}^{k_0}\sum_{s=k_0+1}^{n}[h(Y_r,Y_s)-Eh(Y_r,Y_s)].$$

We first deal with B_1, and we have

$$\begin{aligned}
B_1 &= \frac{1}{n}g_3(\tau)\sum_{r\neq s,1}^{k}[h(Y_r,Y_s)-Eh(Y_r,Y_s)]\\
&= n^{\frac{1}{2}}g_3(\tau)\frac{\sqrt{k}(k-1)}{n^{\frac{3}{2}}}\frac{\sqrt{k}}{k(k-1)}\sum_{r\neq s,1}^{k}[h(Y_r,Y_s)-Eh(Y_r,Y_s)]\\
&= n^{\frac{1}{2}}g_3(\tau)O(1)O_p(1)\\
&= O_P(n^{\frac{1}{2}}),
\end{aligned}$$

where $g_3(x) = x^{\gamma-1}(1-x)^{\gamma+1}$. The third equality is implied by the asymptotic normality of U-statistic. Similarly, we obtain $B_2 = O_P(n^{\frac{1}{2}}|\tau-\tau_0|^{\frac{3}{2}})$, and $B_3 = O_P(n^{\frac{1}{2}})$. The fourth term B_4 can be decomposed further as

$$B_4 = B_{4.1} + B_{4.2} + B_{4.3},$$

where

$$B_{4.1} = -\frac{2}{n}\left[\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\right]\sum_{r=1}^{k}\sum_{s=k+1}^{k_0}\tilde{h}(Y_r,Y_s),$$

$$B_{4.2} = -\frac{2}{n}\left[\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\right]\sum_{r=1}^{k}\sum_{s=k+1}^{k_0}[E[h(Y_r,Y_s)\mid Y_r]-Eh(Y_r,Y_s)],$$

$$B_{4.3} = -\frac{2}{n}\left[\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\right]\sum_{r=1}^{k}\sum_{s=k+1}^{k_0}[E[h(Y_r,Y_s)\mid Y_s]-Eh(Y_r,Y_s)].$$

The application of the central limit theory yields that $B_{4.1} = O_p(1)$, $B_{4.2} = O_p(n^{\frac{1}{2}})$, and $B_{4.3} = O_p(n^{\frac{1}{2}})$. Thus we obtain $B_5 = O_p(n^{\frac{1}{2}})$. Similar arguments yield also $B_4 = O_p(n^{\frac{1}{2}})$, and $B_6 = O_p(n^{\frac{1}{2}})$. Combining the above results, we immediately obtain

$$T(k) - ET(k) = O_P(n^{\frac{1}{2}}). \tag{A.15}$$

Similar arguments yield that

$$T(k_0) - ET(k_0) = O_P(n^{\frac{1}{2}}). \tag{A.16}$$

Thus,

$$(T(k) - ET(k)) - (T(k_0) - ET(k_0)) = O_P(n^{\frac{1}{2}}|\tau-\tau_0|^{\frac{1}{2}}). \tag{A.17}$$

Denote $A = [(T(k) - ET(k)) - (T(k_0) - ET(k_0))]^2 + 2(T(k_0) - ET(k_0))[(T(k) - ET(k)) - (T(k_0) - ET(k_0))] + 2(T(k) - ET(k))(ET(k) - ET(k_0)) + 2ET(k_0)[(T(k) - ET(k)) - (T(k_0) - ET(k_0))]$. It can be easily derived from the proof of Theorem 3.1 that $ET(k) - ET(k_0) = O(n\Delta_n|\tau - \tau_0|)$ and $ET(k_0) = O(n\Delta_n)$. Thus by (3.2), and (A.15)–(A.17), we obtain that

$$\begin{aligned} A &= O_P(n|\tau - \tau_0|) + O_P(n|\tau - \tau_0|^{1/2}) + O_P(n^{\frac{3}{2}}\Delta_n|\tau - \tau_0|) + O_P(n^{\frac{3}{2}}\Delta_n|\tau - \tau_0|^{\frac{1}{2}}) \\ &= O_P(n^{\frac{3}{2}}\Delta_n|\tau - \tau_0|^{\frac{1}{2}}), \end{aligned} \tag{A.18}$$

for $k \leqslant k_0$. Similar arguments yield that for $k > k_0$,

$$A = O_P(n^{\frac{3}{2}}\Delta_n|\tau - \tau_0|^{\frac{1}{2}}). \tag{A.19}$$

Since $T^2(\hat{k}_n) \geqslant T^2(k_0)$, by combining (A.14) and (A.13), and replacing τ by $\hat{\tau}_n$ in (A.13), (A.18) and (A.19), we obtain that

$$2n^2\Delta_n^2|\hat{\tau}_n - \tau_0|\tilde{\delta} \leqslant O_P(n^{\frac{3}{2}}\Delta_n|\hat{\tau}_n - \tau_0|^{\frac{1}{2}}).$$

Hence, we have that

$$P\left(\frac{2n^2\Delta_n^2|\hat{\tau}_n - \tau_0|\delta}{n^{\frac{3}{2}}\Delta_n|\hat{\tau}_n - \tau_0|^{\frac{1}{2}}} > M_0\right) \leqslant P\left(\frac{A}{n^{\frac{3}{2}}\Delta_n|\hat{\tau}_n - \tau_0|^{\frac{1}{2}}} > M_0\right) < \varepsilon,$$

for every small number $\varepsilon > 0$, and every large number $M_0 > 0$, i.e.,

$$P(\sqrt{n}\Delta_n|\hat{\tau}_n - \tau_0|^{\frac{1}{2}} > M_0) = P(n\Delta_n^2|\hat{\tau}_n - \tau_0| > M) < \varepsilon,$$

which yields that $|\hat{\tau}_n - \tau_0| = O_p(n^{-1}\Delta_n^{-2})$. The proof is complete.

A.2 Proof of Theorem 3.3

To show that

$$\frac{(1+\gamma-2\gamma\tau_0)^2}{\lambda_1^2}\Delta_n^2(\hat{k}_n - k_0) \Rightarrow \arg\max_u G(u),$$

where

$$G(u) = \begin{cases} W_1(-u) + \dfrac{u}{2}, & u \leqslant 0, \\ \dfrac{\lambda_2}{\lambda_1}W_2(u) - \dfrac{1}{2}\left(\dfrac{1-\gamma+2\gamma\tau_0}{1+\gamma-2\gamma\tau_0}\right)u, & u > 0, \end{cases}$$

and $W_1(\cdot)$ and $W_2(\cdot)$ are two mutually independent standard Brownian motion processes on $[0, \infty)$, it is equivalent to prove that

$$\Delta_n^2(\hat{k}_n - k_0) \Rightarrow \arg\max_s V(s), \tag{A.20}$$

where

$$V(s) = \begin{cases} \lambda_1 W_1(-s) + \dfrac{1+\gamma-2\gamma\tau_0}{2}s, & s \leqslant 0, \\ \lambda_2 W_2(s) - \dfrac{1-\gamma+2\gamma\tau_0}{2}s, & s > 0, \end{cases}$$

since

$$\arg\max_s V(s) = \frac{\lambda_1^2}{(1+\gamma-2\gamma\tau_0)}\arg\max_u G(u)$$

by a change in variable

$$u = \frac{(1+\gamma-2\gamma\tau_0)^2}{\lambda_1^2}s.$$

In the following, we show that (A.20) holds true.

By Theorem 3.1, $\hat{k}_n = k_0 + O_P(\Delta_n^{-2})$. Hence, to derive the asymptotic distribution of the change point estimator $\hat{k}_n$, we only need to consider the behaviour of k satisfying $k = k_0 + \lfloor s\Delta_n^{-2} \rfloor$ and $s \in [-M, M]$, for any given $M > 0$. Define

$$V_n(s) = \frac{1}{4n\tau_0^{2\gamma+1}(1-\tau_0)^{2\gamma+1}}(T^2(k) - T^2(k_0)),$$

where $k = k_0 + \lfloor s\Delta_n^{-2} \rfloor$. Since

$$\hat{k}_n = \arg\max_{1\leqslant k<n} T(k),$$

we can easily derive that

$$\arg\max_s V_n(s) = \Delta_n^2(\hat{k}_n - k_0).$$

If we can show that $V_n(s) \Rightarrow V(s)$, by the continuous mapping theorem, we have

$$\Delta_n^2(\hat{k}_n - k_0) \Rightarrow \arg\max_s V(s).$$

To achieve $V_n(s) \Rightarrow V(s)$, we decompose $V_n(s)$ into five terms as follows:

$$V_n(s) \equiv D_1 + D_2 + D_3 + D_4 + D_5, \tag{A.21}$$

where

$$\begin{aligned}
D_1 &= [4n\tau_0^{2\gamma+1}(1-\tau_0)^{2\gamma+1}]^{-1}[|ET(k)|^2 - |ET(k_0)|^2],\\
D_2 &= [4n\tau_0^{2\gamma+1}(1-\tau_0)^{2\gamma+1}]^{-1}[|(T(k) - ET(k)) - (T(k_0) - ET(k_0))|^2],\\
D_3 &= [4n\tau_0^{2\gamma+1}(1-\tau_0)^{2\gamma+1}]^{-1}[2(T(k_0) - ET(k_0))(T(k) - ET(k) - (T(k_0) - ET(k_0)))],\\
D_4 &= [4n\tau_0^{2\gamma+1}(1-\tau_0)^{2\gamma+1}]^{-1}[2(T(k) - ET(k))(ET(k) - ET(k_0))],\\
D_5 &= [4n\tau_0^{2\gamma+1}(1-\tau_0)^{2\gamma+1}]^{-1}[2ET(k_0)(T(k) - ET(k) - (T(k_0) - ET(k_0)))].
\end{aligned}$$

We first consider the case that $s \leqslant 0$, or equivalently $k \leqslant k_0$. Since $k = k_0 + \lfloor s\Delta_n^{-2} \rfloor$, by using the mean value theorem, we obtain that

$$\begin{aligned}
D_1 &= [4n\tau_0^{2\gamma+1}(1-\tau_0)^{2\gamma+1}]^{-1}[ET(k) - ET(k_0)][ET(k) + ET(k_0)]\\
&= [4n\tau_0^{2\gamma+1}(1-\tau_0)^{2\gamma+1}]^{-1}\\
&\quad \times \left\{ n\Delta_n \left[\left(\frac{k(n-k)}{n^2}\right)^{\gamma} \frac{k(n-k_0)}{(n-k)n} - \left(\frac{k_0(n-k_0)}{n^2}\right)^{\gamma} \frac{k_0}{n} \right] \frac{(n-k_0)}{n} + o_P(n\Delta_n) \right\}\\
&\quad \times \left\{ n\Delta_n \left[\left(\frac{k(n-k)}{n^2}\right)^{\gamma} \frac{k(n-k_0)}{(n-k)n} + \left(\frac{k_0(n-k_0)}{n^2}\right)^{\gamma} \frac{k_0}{n} \right] \frac{(n-k_0)}{n} + o_P(n\Delta_n) \right\}\\
&= \frac{1+\gamma-2\gamma\tau_0}{2}\Delta_n^2(k-k_0) + o_P(\Delta_n^2(k-k_0))\\
&\to \frac{1+\gamma-2\gamma\tau_0}{2}s.
\end{aligned} \tag{A.22}$$

In view of the proof of Theorem 3.1, we obtain that, as $n \to \infty$,

$$\begin{aligned}
\max_{|k-k_0|\leqslant M/\Delta_n^2} \frac{D_2}{\Delta_n^2|k-k_0|} &= o_P(1),\\
\max_{|k-k_0|\leqslant M/\Delta_n^2} \frac{D_3}{\Delta_n^2|k-k_0|} &= o_P(1),\\
\max_{|k-k_0|\leqslant M/\Delta_n^2} \frac{D_4}{\Delta_n^2|k-k_0|} &= o_P(1).
\end{aligned}$$

Consequently, it suffices to investigate only the limiting behaviours of D_1 and D_5, as $n \to \infty$. Since each of $\{E(h(Y_r, Z_1) \mid Y_r) - Eh(Y_r, Z_1),\ r = 1, 2, \ldots, k_0\}$, $\{E(h(Y_r, Z_1) \mid Y_r) - Eh(Y_r, Z_1),\ r = k_0 + 1, \ldots, n\}$,

$\{E(h(Y_r, Z_2) \mid Y_r) - Eh(Y_r, Z_2), r = 1, 2, \dots, k_0\}$, and $\{E(h(Y_r, Z_2) \mid Y_r) - Eh(Y_r, Z_2),\ r = k_0 + 1, \dots, n\}$ is an independently and identically distributed sequence with zero mean and finite variance, where Z_1 and Z_2 are independent of $Y_1, Y_2, \dots, Y_n$ and follow the distributions F_1 and F_n respectively, by the weak convergence of partial sums, we can prove that

$$\begin{aligned}
D_5 &= [4n\tau_0^{2\gamma+1}(1-\tau_0)^{2\gamma+1}]^{-1}[n\Delta_n\tau_0^{\gamma}(1-\tau_0)^{\gamma}(1-\tau_0)\tau_0 + o_P(n\Delta_n)] \\
&\quad \times \Bigg\{ -\frac{2}{n}\Bigg[\left(\frac{k(n-k)}{n^2}\right)^{\gamma} + \left(\frac{k_0(n-k_0)}{n^2}\right)^{\gamma}\frac{(n-k_0)}{k_0}\Bigg]\sum_{l=1}^{k}\sum_{r=k+1}^{k_0}[E[h(Y_l, Y_r) \mid Y_r] - Eh(Y_l, Y_r)] \\
&\quad + \frac{2}{n}\Bigg[\left(\frac{k(n-k)}{n^2}\right)^{\gamma}\frac{k}{n-k} + \left(\frac{k_0(n-k_0)}{n^2}\right)^{\gamma}\Bigg]\sum_{l=k_0+1}^{n}\sum_{r=k+1}^{k_0}[E[h(Y_l, Y_r) \mid Y_r] - Eh(Y_l, Y_r)]\Bigg\} \\
&\quad + [4n\tau_0^{2\gamma+1}(1-\tau_0)^{2\gamma+1}]^{-1}[n\Delta_n\tau_0^{\gamma}(1-\tau_0)^{\gamma}(1-\tau_0)\tau_0 + o_P(n\Delta_n)]o_P(n^{\frac{1}{2}}|\tau-\tau_0|^{\frac{1}{2}}) \\
&= [n\tau_0^{2\gamma+1}(1-\tau_0)^{2\gamma+1}]^{-1}[n\Delta_n\tau_0^{\gamma}(1-\tau_0)^{\gamma}(1-\tau_0)\tau_0 + o_P(n\Delta_n)]\tau_0^{\gamma}(1-\tau_0)^{\gamma} \\
&\quad \times \Bigg\{ -\sum_{r=k+1}^{k_0}[E[h(Y_r, Z_1) \mid Y_r] - Eh(Y_1, Y_2)] \\
&\quad + \sum_{r=k+1}^{k_0}[E[h(Z_2, Y_r) \mid Y_r] - Eh(Y_1, Y_{k_0+1})] + o_P(n^{\frac{1}{2}}|\tau-\tau_0|^{\frac{1}{2}})\Bigg\} \\
&= \Delta_n\sum_{r=k+1}^{k_0}\{[E[h(Z_2, Y_r) \mid Y_r] - Eh(Y_1, Y_{k_0+1})] - [E[h(Y_r, Z_1) \mid Y_r] - Eh(Y_1, Y_2)]\} \\
&\quad + o_P(n^{\frac{1}{2}}|\tau-\tau_0|^{\frac{1}{2}}\Delta_n) \\
&\Rightarrow \lambda_1 W_1(-s),
\end{aligned} \tag{A.23}$$

where $W_1(\cdot)$ is a standard Brownian motion process on $[0, \infty)$ and

$$\begin{aligned}
\lambda_1 &= (E\{[E[h(Z_2, Y_r) \mid Y_r] - Eh(Y_1, Y_{k_0+1})] - [E[h(Y_r, Z_1) \mid Y_r] - Eh(Y_1, Y_2)]\}^2)^{\frac{1}{2}} \\
&= (E\{[E[h(Y_{k_0+1}, Y_1) \mid Y_1] - Eh(Y_1, Y_{k_0+1})] - [E[h(Y_1, Y_2) \mid Y_1] - Eh(Y_1, Y_2)]\}^2)^{\frac{1}{2}}.
\end{aligned}$$

Therefore, by (A.21)–(A.23), we obtain that, for $s \leqslant 0$,

$$V_n(s) \Rightarrow \lambda_1 W_1(-s) + \frac{1+\gamma-2\gamma\tau_0}{2}s.$$

For the case that $s > 0$, similar arguments yield that

$$V_n(s) \Rightarrow \lambda_2 W_2(s) - \frac{1-\gamma+2\gamma\tau_0}{2}\ s,$$

where $W_2(\cdot)$ is another standard Brownian motion process on $[0, \infty)$ independent of $W_1(\cdot)$, and

$$\begin{aligned}
\lambda_2 = (E\{&[E[h(Y_1, Y_{k_0+1}) \mid Y_{k_0+1}] - Eh(Y_1, Y_{k_0+1})] \\
&- [E[h(Y_{k_0+1}, Y_{k_0+2}) \mid Y_{k_0+1}] - Eh(Y_{k_0+1}, Y_{k_0+2})]\}^2)^{\frac{1}{2}}.
\end{aligned}$$

In summary,

$$V_n(s) = \frac{1}{4n\tau_0^{2\gamma+1}(1-\tau_0)^{2\gamma+1}}(T^2(k_0 + \lfloor s\Delta_n^{-2}\rfloor) - T^2(k_0)) \Rightarrow V(s),$$

where

$$V(s) = \begin{cases} \lambda_1 W_1(-s) + \dfrac{1+\gamma-2\gamma\tau_0}{2}s, & s \leqslant 0, \\ \lambda_2 W_2(s) - \dfrac{1-\gamma+2\gamma\tau_0}{2}s, & s > 0. \end{cases}$$

The proof is completed.

A.3 Proof of Proposition 4.1

Suppose that $k_0 \geqslant \bar{\hat{k}}$,

$$|\hat{k}_{a_{\mathrm{s}}} - k_0| = |\hat{k}_{a_{\mathrm{s}}} - \bar{\hat{k}} + \bar{\hat{k}} - k_0| \leqslant |\hat{k}_{a_{\mathrm{s}}} - \bar{\hat{k}}| + |\bar{\hat{k}} - k_0| \leqslant |\hat{k}_{a_i} - \bar{\hat{k}}| + |\bar{\hat{k}} - k_0|.$$

The last inequality holds true for any $a_i \in \mathcal{A}$ by the definition of a_{s}. There always exists at least one point $a^* \neq a_{\mathrm{s}}$ in $\mathcal{A}$ such that $\hat{k}_{a^*} \leqslant \min(\hat{k}_{a_{\mathrm{s}}}, \bar{\hat{k}})$. Therefore,

$$|\hat{k}_{a_{\mathrm{s}}} - k_0| \leqslant |\hat{k}_{a^*} - \bar{\hat{k}}| + |\bar{\hat{k}} - k_0| \leqslant \bar{\hat{k}} - \hat{k}_{a^*} + k_0 - \bar{\hat{k}} = |\hat{k}_{a^*} - k_0|. \tag{A.24}$$

Similarly, we can show (A.24) for the case that $k_0 < \bar{\hat{k}}$. The proof is completed.