LSTA #298484 VOL 37, ISS 17

U-Statistic Based Modified Information Criterion for Change Point Problems

JIANMIN PAN[1] AND JIAHUA CHEN[2]

QUERY SHEET

This page lists questions we have about your paper. The numbers displayed at left can be found in the text of the paper for reference. In addition, please review your paper as a whole for corrections.

There are no Editor Queries for this paper.

TABLE OF CONTENTS LISTING

The table of contents for the journal will list your paper exactly as it appears below:

Communications in Statistics—Theory and Methods, 37: 1–26, 2008

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920801983185

Taylor & Francis
Taylor & Francis Group

Inference

U-Statistic Based Modified Information Criterion for Change Point Problems

JIANMIN PAN[1] AND JIAHUA CHEN[2]

[1]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, Tennessee, USA
[2]Department of Statistics & Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

The U-statistic based modified information criterion (MIC) is proposed and applied to detect the change point in a sequence of independent random variables. In this article, we show that the method is consistent in selecting the correct model, and the resulting test statistic has a simple limiting distribution. We investigate the method based on both symmetric and anti-symmetric kernel functions. The simulation results indicate that the new method has better power in detecting the changes compared to other methods, such as the likelihood based MIC (Chen et al., 2006) and the Bayesian information criterion of Schwarz (BIC, Schwarz, 1978).

Keywords Change point; Consistency; Limiting distribution; Model complexity; Nonparametric model; *U*-statistic.

Mathematics Subject Classification 62G10; 62G20.

1. Introduction

In applications such as in quality control, we are often interested in knowing whether a sequence of observations $x_1, x_2, \ldots, x_n$ can be modeled as a random sample from a single distribution $f(x)$, or it should be divided into two subsequences $x_1, x_2, \ldots, x_k$ and $x_{k+1}, \ldots, x_n$ with some k such that they can be viewed as two random samples, one is from $f_1(x)$ and the other is from $f_2(x)$. When the $f(x)$, $f_1(x)$, and $f_2(x)$ are chosen from a parametric family, we make parametric inference on change point detection. The change point problem has been given considerable attention over the years; see Page (1954, 1955), Hinkley (1971), Picard (1985), Zacks (1983), Inclán and Tiao (1994), Kim et al. (2000) and Lee and Park (2001).

Due to their simplicity, the parametric methods are often more efficient. In general, their effectiveness relies on correctly specifying the parametric

Received April 3, 2007; Accepted February 12, 2008
Address correspondence to Jianmin Pan, Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN 38105, USA; E-mail: jianmin.pan@stjude.org

distribution of the population. When we do not have sufficient knowledge about the physical background of the sample, we may not be able to propose a defensible parametric model. In addition, some parametric methods can have poor behavior when the true distribution of the sample differs from the assumed model. For example, if the $f(x)$, $f_1(x)$, and $f_2(x)$ are assumed to be normal distributions but the data are from some stable distribution with a small index of stability, we may end up detecting a change point due to the random occurrence of some unusually large observations. To avoid this problem, nonparametric methods are often considered. Nonparametric *methods* can also be useful when we are only interested in detecting changes in some aspects of the underlying distribution, for instance, whether the sequence of the observations has gone through a location shift or a scale change. A U-statistic can be constructed to reflect the change in these specific aspects.

In general, a U-statistic is the average of a simple m-variate function over every possible subset of m observations from a sample of n observations. Many commonly used statistics such as sample mean and sample variance are U-statistics. Two U-statistics can be defined based on two sub-samples, one consists of $x_1, \ldots, x_k$ and the other $x_{k+1}, \ldots, x_n$. Their difference after proper scaling reflects a possible change in the designated aspect. Since a statistic is defined for each k, a stochastic process indexed by k is the result.

Csörgö and Horváth (1988) first applied the U-statistic to change point problems. Gombay and Horváth (1995), Gombay (2000, 2001), and others studied the large sample behaviors of the process and the change point estimator. Under the null model, the process converges to a Gaussian process after proper normalization. We can hence test the existence of a change point based on the maximum of the process with its critical value *determined* by the percentile of the supremum of the limiting Gaussian process. However, the computation of the percentiles for the supremum of the Gaussian process is usually not easy. See more details in Csörgö and Horváth (1997). In this article, we propose and investigate the use of an MIC principle (Chen et al., 2006) to U-statistic approach. We show that the statistic of the new method has a simple limiting distribution so that its asymptotic critical values can be easily computed.

The article is organized as follows. In Sec. 2, we give a brief review about the modified information criterion in Chen et al. (2006). In Sec. 3, we introduce U-statistics based MIC for both symmetric and anti-symmetric kernels and obtain the null limiting distributions of the corresponding statistics when there exist no change points in the sequence. We conduct simulation studies in Sec. 4, and the new method is compared to several existing methods and found to have good finite sample properties. For the convenience of presentation, the proofs of main results are deferred to the Appendix.

2. Modified Information Criterion

When the null hypothesis of no-change is rejected, a more complex model with two distributions $f_1(x)$ and $f_2(x)$ plus the location of change, k, is preferred than a simple model $f(x)$. The change point problem may hence be regarded as a special case of the model selection problem (Csörgö and Horváth, 1997). In the context of model selection, Akaike information criterion and Bayesian information criterion are routinely used; see Konishi and Kitagawa (1996), Volinsky and Raftery (2000), Bogdan et al. (2004), and Bengtsson and Cavanaugh (2006). Since the change point models are nonregular, these criteria are no-longer optimal and lose some useful properties. Chen et al. (2006) refined the notion of model complexity in change point models, and proposed the modified information criterion. We briefly review this concept in this section. Its application to nonparametric method will be presented in the next section.

Suppose we have a sequence of independent observations $X_1, \ldots, X_n$. It is suspected that X_i has density function $f(x, \theta_1)$ for $i \leq \tau$ and density $f(x, \theta_2)$ for $i > \tau$, and $f(x, \theta_1)$ and $f(x, \theta_2)$ belong to the same parametric distribution family $\{f(x, \theta); \theta \in \Theta\}$ with $\Theta \subset \mathscr{R}^d$. The problem is to test whether this change has indeed occurred and if so, find the location of the change k. Hence, the null hypothesis is:

$$H_0 : X_i \sim f(x, \theta), \quad \theta = \theta_1 = \theta_2, \quad \text{for } 1 \leq i \leq n$$

and the alternative is:

$$H_1 : X_i \sim f(x, \theta_1) \quad \text{for } i \leq k \quad \text{and} \quad X_i \sim f(x, \theta_2) \quad \text{for } i \geq k,$$
$$\theta_1 \neq \theta_2 \quad \text{and} \quad 1 \leq k < n.$$

For regular parametric (not change point) models with log likelihood function $\ell_n(\theta)$, the Bayesian information criterion (Schwarz, 1978) is defined as:

$$BIC = -2\ell_n(\hat{\theta}) + d\log(n),$$

where $\hat{\theta}$ is the maximum point of $\ell_n(\theta)$, and d is the dimension of parameter θ. The best model according to this criterion is the one which minimizes BIC.

The log likelihood function for the change point problem has the form:

$$\ell_n(\theta_1, \theta_2, k) = \sum_{i=1}^{k} \log f(X_i, \theta_1) + \sum_{i=k+1}^{n} \log f(X_i, \theta_2).$$

The Bayesian information criterion for the change point problem becomes

$$BIC(k) = -2\ell_n(\hat{\theta}_{1k}, \hat{\theta}_{2k}, k) + [2d + 1]\log(n)$$

where $\hat{\theta}_{1k}$, $\hat{\theta}_{2k}$ maximize $\ell_n(\theta_1, \theta_2, k)$ for given k.

Chen et al. (2006) suggested that the model is the least complex when the change point τ is located in the middle of the sequence because both parameters θ_1 and θ_2 are effective in this case. The model is particularly unappealing when τ is near 1 or n but does not equal one of them. When this happens, an additional set of parameters is introduced just for a small proportion of observations. Hence, the model complexity is increased when τ moves away from the middle of the sequence. Based on this consideration, the modified information criterion was proposed as, for $1 \leq k < n$:

$$MIC(k) = -2\ell_n(\hat{\theta}_{1k}, \hat{\theta}_{2k}, k) + \left[2d + \left(\frac{2k}{n} - 1\right)^2\right]\log(n). \tag{1}$$

Under the null model, they defined:

$$MIC(n) = -2\ell_n(\hat{\theta}, \hat{\theta}, n) + d\log(n),$$

where $\hat{\theta}$ maximizes $\ell_n(\theta, \theta, n)$ or $\ell_n(\theta)$. If $MIC(n) > \min_{1\leq k<n} MIC(k)$, the model with a change point is selected and the change point is estimated by $\hat{\tau}$ such that:

$$MIC(\hat{\tau}) = \min_{1\leq k<n} MIC(k).$$

The penalty term in (1) can be motivated as follows. If the change point is at k, the variance of $\hat{\theta}_{1k}$ would be proportional to k^{-1} and the variance of $\hat{\theta}_{2k}$ would be proportional to $(n-k)^{-1}$. Thus, the total variance is proportional to:

$$\frac{1}{k} + \frac{1}{n-k} = 4n^{-1}\left[1 - \left(\frac{2k}{n} - 1\right)^2\right]^{-1}.$$

The specific form in (1) reflects this important fact. Thus, a larger elevation in the U-statistic is needed to justify a change when k is near 1 or n. This notion is shared by many researchers. The method in Inclán and Tiao (1994) scales down the statistics heavier when the suspected change point is near 1 or n. The U-statistic method in Gombay and Horváth (1995) is scaled down by multiplying the factor $k(n-k)$.

Let

$$S_n = MIC(n) - \min_{1\leq k<n} MIC(k) + d\log n,$$

then $S_n \to \chi^2(d)$ in distribution under null hypothesis, and $S_n \to \infty$ in probability under alternative when there exists one change point in the sequence; see Theorem 1 in Chen et al. (2006). The inference based on S_n will be called the likelihood based MIC in this article.

3. *U*-Statistic Based MIC Method

We now introduce a U-statistic based nonparametric MIC method. Without specific parametric models, the null hypothesis becomes

$$H_0 : X_1, \ldots, X_n \text{ i.i.d.} \sim F(x)$$

and the alternative hypothesis is:

$$H_1 : X_1, \ldots, X_\tau \text{ i.i.d.} \sim F(x), \quad X_{\tau+1}, \ldots, X_n \text{ i.i.d.} \sim G(x)$$
$$\text{and} \quad F(x) \neq G(x) \quad \text{for some } x.$$

The distribution functions F, G, and the change point τ are unknown. We assume $\tau = [n\lambda]$ for some λ with $0 < \lambda < 1$ under the alternative, where $[x]$ is the largest integer no larger than x.

Let $h : \mathscr{R}^2 \to \mathscr{R}$ be a Borel measurable function. A U-statistic with order 2 based on n independent observations $X_1, \ldots, X_n$ is defined as:

$$U_n(X) = \binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} h(X_i, X_j).$$

A U-statistic of order m replaces h by a m-variate function, and the summation is taken over all subsets of size m.

As usual in the theory of U-statistics, we investigate the change point problems based on both cases of symmetric kernels:

$$h(y, x) = h(x, y), \quad -\infty < x, y < \infty,$$

and anti-symmetric kernels

$$h(y, x) = -h(x, y), \quad -\infty < x, y < \infty$$

in this section.

3.1. *Symmetric Kernel Case*

Let h be a symmetric kernel function. Define $\theta_1 = E_F h(X_1, X_2)$ and $\theta_2 = E_G h(X_1, X_2)$, which are the expected values of $h(X_1, X_2)$ under the distributions $F(x)$ and $G(x)$, respectively. When using U-statistics based on the kernel function h, we give up the possibility of detecting all changes in from F to G, but detecting the change in the expected value of $h(X_1, X_2)$. The expected value of $h(x, y)$ could be mean, variance of the distribution or whatever. Hence, we need to decide what change we want to detect in the distribution and then select an appropriate kernel.

To apply U-statistic method to change point problems, we define:

$$\hat{\theta}_1(k) = \binom{k}{2}^{-1} \sum_{1 \le i < j \le k} h(X_i, X_j) \quad \text{and} \quad \hat{\theta}_2(k) = \binom{n-k}{2}^{-1} \sum_{k < i < j \le n} h(X_i, X_j). \tag{2}$$

These estimators are unbiased estimators of θ_1 and θ_2 based on the first k and the remaining $n - k$ observations if the change point is located at k for $k = 2, \ldots, n-2$. For convenience, we define both $\hat{\theta}_1(k) = 0$ and $\hat{\theta}_2(k) = 0$ for $k = 1, n-1$ and n.

It is now very natural to examine the size of the difference between $\hat{\theta}_1(k)$ and $\hat{\theta}_2(k)$. For each k, $|\hat{\theta}_1(k) - \hat{\theta}_2(k)|$ compares the means of h based on the first k and the last $n - k$ observations. When the difference is large for some k, there are some evidences to reject the null model in favor of the alternative model. However, the evidences are not of the same importance for different choices of k. Thus, it is important to assign a proper weight for each k. One obvious choice is related to the variance of $\hat{\theta}_1(k) - \hat{\theta}_2(k)$, which can be written as:

$$\mathrm{Var}[\hat{\theta}_1(k) - \hat{\theta}_2(k)] = \frac{4n\sigma^2}{k(n-k)} + O\left[\frac{1}{k^2} + \frac{1}{(n-k)^2}\right]$$

under H_0. Thus, we define

$$V_n^{(1)}(k) = \left(4n\hat{\sigma}_{k1}^2\right)^{-1} k(n-k)[\hat{\theta}_1(k) - \hat{\theta}_2(k)]^2, \tag{3}$$

where $\hat{\sigma}_{k1}^2$ is an estimator of $\sigma^2 = Var\{E[h(X_1, X_2)|X_2]\}$ which is defined by:

$$\hat{\sigma}_{k1}^2 = \frac{1}{n}\left\{\sum_{j=1}^{k}[h_{k1}(X_j) - \hat{\theta}_1(k)]^2 + \sum_{j=k+1}^{n}[h_{k2}(X_j) - \hat{\theta}_2(k)]^2\right\}$$

for each k, where for $j = 1, \ldots, k$:

$$h_{k1}(X_j) = \begin{cases} \frac{1}{k-1} \sum_{1 \le i \le k,\ i \ne j} h(X_j, X_i), & \text{if } k = 2, \ldots, n \\ 0, & \text{if } k = 1 \end{cases} \tag{4}$$

and for $j = k+1, \ldots, n$,

$$h_{k2}(X_j) = \begin{cases} \frac{1}{n-k-1} \sum_{k < i \le n,\ i \ne j} h(X_j, X_i), & \text{if } k = 1, \ldots, n-2 \\ 0, & \text{if } k = n-1 \text{ and } n \end{cases}. \tag{5}$$

The following proposition indicates that $\hat{\sigma}^2_{k1}$ is a consistent estimator of σ^2 under the null hypothesis H_0, and still has some nice properties under the alternative hypothesis H_1. If $\tilde{h}(t) = Eh(X_1, t)$, then $\sigma^2 = Var[\tilde{h}(X_1)]$.

Proposition 3.1. (1) *Assume that* $Eh^2(X_1, X_2) < \infty$, $\sigma^2 > 0$ *and* $E\tilde{h}^4(X_1) < \infty$. *Then we have, under the null hypothesis* H_0, *as* $n \to \infty$:

$$\max_{1 \le k \le n} |\hat{\sigma}^2_{k1} - \sigma^2| = o_p(1).$$

(2) *Let* $\sigma_1^2 = Var[\tilde{h}_1(X_1)]$ *and* $\sigma_2^2 = Var[\tilde{h}_2(X_{\tau+1})]$ *with* $\tilde{h}_1(t) = Eh(X_1, t)$ *and* $\tilde{h}_2(t) = Eh(X_{\tau+1}, t)$. *Assume that there exists a change point at* $\tau = [n\lambda]$ *with* $0 < \lambda < 1$. *Then, as* $n \to \infty$:

$$\hat{\sigma}^2_{k1} \to \lambda\sigma_1^2 + (1-\lambda)\sigma_2^2 \mathrel{\hat{=}} \sigma_0^2$$

in probability uniformly for all k *such that* $|k - \tau| \le n(\log n)^{-1}$.

We now take the main idea for the modified information criterion in Chen et al. (2006) into consideration, we finally define the test statistic as:

$$U_n^{(1)} = \max_{\mathbf{1 \le k < n}} \left\{ V_n^{(1)}(k) - \left(\frac{2k}{n-1}\right)^2 \log n \right\}.$$

When the alternative model is favored, the location of the change point can be estimated as follows. Let

$$U_n^{(1)}(k) = V_n^{(1)}(k) - \left(\frac{2k}{n} - 1\right)^2 \log n$$

and define $\hat{\tau}$ as the value of k such that:

$$U_n^{(1)}(\hat{\tau}) = \max_{\mathbf{1 \le k < n}} U_n^{(1)}(k). \tag{6}$$

Compared to the parametric inference in Chen et al. (2006), the role of $V_n^{(1)}(k)$ is similar to that of $\ell_n(\hat{\theta}_{1k}, \hat{\theta}_{2k}, k) - \ell_n(\hat{\theta}, \hat{\theta}, n)$, and the role of $U_n^{(1)}$ is similar to that of S_n, accordingly.

One significant advantage of using the MIC is its simpler large sample behavior (see Chen et al., 2006). The key difference between MIC and other information criteria such as Akaike Information Criterion (AIC; Akaike, 1973) and Bayesian information criterion (BIC; Schwarz, 1978) is that the test statistic based on the MIC has a simple chi-square limiting distribution. This is particularly appealing when designing a test with correct asymptotic significance level. At the same time, the MIC based procedures have higher or comparable powers to many other methods (see Chen et al., 2006). The hypothetical change point is forced to the middle of the sequence by the MIC which does not really matter when θ_1 is the same as θ_2 (under H_0). Ideally, the estimated location of the change point is close to the true value, rather than being pushed to the middle of the sequence under the alternative model.

Theorem 3.1.

(1) *Assume that the null hypothesis H_0 is true, and $E|h(X_1, X_2)|^4 < \infty$ and $\sigma^2 > 0$ are satisfied. Then, as $n \to \infty$:*

$$U_n^{(1)} \to \chi_1^2$$

in distribution.

(2) *Assume that the alternative hypothesis H_1 is true and the change point $\tau = [n\lambda]$ with $\lambda \in (0, 1)$. Then:*

$$U_n^{(1)} \to \infty$$

in probability.

From Theorem 3.1, we conclude that the method based on test statistic $U_n^{(1)}$ is consistent in the sense that we will choose the model with a change point with probability approaching 1 when there exists indeed one change point at τ such that $\tau/n \to \lambda \in (0, 1)$.

The proofs of Proposition 3.1 and Theorem 3.1 will be presented in Appendix.

3.2. *Anti-Symmetric Kernel Case*

For any anti-symmetric kernel h, it is obvious that $Eh(X_1, X_{\tau+1}) = 0$ under the null hypothesis. We assume that:

$$\mu = Eh(X_1, X_{\tau+1}) \neq 0 \tag{7}$$

under the alternative H_1, and

$$Eh^2(X_i, X_j) < \infty \quad \text{for all } i < j \tag{8}$$

and

$$\sigma^2 = Var\{\tilde{h}(X_{\tau+1})\} > 0, \tag{9}$$

where $\tilde{h}(t) = Eh(t, X_1)$ is the projection. Condition (8) implies that $\sigma < \infty$. We will rely on the following generalized U-statistic for the kernel $h(x, y)$ to detect the change in the sequence $X_1, \ldots, X_n$. Let

$$Z_k = \sum_{1 \le i \le k} \sum_{k < j \le n} h(X_i, X_j), \quad \text{for } 1 \le k \le n-1,$$

and $Z_k = 0$ if $k = n$. Since $EZ_k = 0$ under the null hypothesis H_0 and $EZ_k = k(n-k)\mu \neq 0$ if k is the true change point, it is natural to examine the size of Z_k. We will have evidence to reject the null hypothesis H_0 in favor of the alternative hypothesis H_1 if $|Z_k|$ is significantly large for some k. Also, we will assign a proper weight for each k when considering the size of Z_k. Obviously, it is reasonable to assume that the weight is inversely proportional to the approximate standard deviation of Z_k under the null hypothesis H_0. Notice that:

$$\mathrm{Var}(Z_k) = EZ_k^2 = nk(n-k)\sigma^2 + O[k(n-k)]$$

under H_0. Similarly, we adopt the idea of MIC in Chen et al. (2006). Denote

$$\hat{\sigma}_{k2}^2 = \frac{1}{n}\left\{ \sum_{j=1}^{k} [h_{k1}(X_j)]^2 + \sum_{j=k+1}^{n} [h_{k2}(X_j)]^2 \right\},$$

where $h_{k1}(X_j)$ and $h_{k2}(X_j)$ are defined in (4) and (5), and

$$V_n^{(2)}(k) = \frac{Z_k^2}{\hat{\sigma}_{k2}^2 nk(n-k)},$$

$$U_n^{(2)}(k) = V_n^{(2)}(k) - \left(\frac{2k}{n-1}\right)^2 \log n,$$

then we define

$$U_n^{(2)} = \max_{\mathbf{1 \le k < n}} U_n^{(2)}(k)$$

as the test statistic.

As in symmetric kernel case, $V_n^{(2)}(k)$ plays a similar role to $\ell_n(\hat{\theta}_{1k}, \hat{\theta}_{2k}, k) - \ell_n(\hat{\theta}, \hat{\theta}, n)$, and the role of $U_n^{(2)}$ is similar to S_n compared to the parametric inference in Chen et al. (2006).

Proposition 3.2. (1) *Assume that* (7)–(9) *hold and* $E\tilde{h}^4(X_1) < \infty$, *then we have under the null hypothesis* H_0, *as* $n \to \infty$,

$$\max_{1 \le k \le n} \left|\hat{\sigma}_{k2}^2 - \sigma^2\right| = o_p(1).$$

(2) *Let* $\sigma_1^2 = Var[\tilde{h}_1(X_1)]$ *and* $\sigma_2^2 = Var[\tilde{h}_2(X_{\tau+1})]$ *with* $\tilde{h}_1(t) = Eh(X_1, t)$ *and* $\tilde{h}_2(t) = Eh(X_{\tau+1}, t)$. *Under the alternative* H_1 *there exists a change point at* $\tau = [n\lambda]$ *with* $0 < \lambda < 1$, *then we have as* $n \to \infty$:

$$\hat{\sigma}_{k2}^2 \to \lambda\sigma_1^2 + (1-\lambda)\sigma_2^2 \hat{=} \sigma_0^2$$

in probability uniformly for all k *such that* $|k - \tau| \le n(\log n)^{-1}$.

If the null hypothesis H_0 is rejected, we define $\hat{\tau}$, the estimator of change point τ, as the value of k such that:

$$U_n^{(2)}(\hat{\tau}) = \max_{\mathbf{1 \le k < n}} U_n^{(2)}(k). \tag{10}$$

Theorem 3.2.

(1) *Assume that* (7)–(9) *hold, and*:

$$E|h(X_1, X_2)|^4 < \infty$$

and

$$E\{\tilde{h}^2(X_1)\log\log[|\tilde{h}(X_1)| + 1]\} < \infty.$$

Then, we have, as $n \to \infty$:

$$U_n^{(2)} \to \chi_1^2$$

in distribution under the null hypothesis H_0.

(2) *If there is a change at* τ *such that* $\frac{\tau}{n} \to \lambda \in (0, 1)$, *as* $n \to \infty$, *then*

$$U_n^{(2)} \to \infty$$

in probability.

Theorem 3.2 implies that the test based on statistic $U_n^{(2)}$ is consistent. We will also prove Proposition 3.2 and Theorem 3.2 in Appendix.

3.3. *Examples of Kernel Functions*

It is certain that the choice of the kernels in the proposed method plays a crucial role. We now take a moment to examine possibilities to detect changes in some aspects of underlying distribution by choosing a specific kernel $h(x, y)$.

- Symmetric Kernels:
  1. Let $h(x, y) = x + y$. It follows that $\theta = 2EX$ and $\sigma^2 = Var(X)$. This kernel can be used to detect the change in the mean.
  2. To detect a change in variance, we could choose $h(x, y) = (x - y)^2$. It follows that $\theta = 2Var(X)$ and $\sigma^2 = E(X - EX)^4 - (Var(X))^2$. The statistic $V_n^{(1)}(k)$ is essentially the difference between two sample variances.
  3. Gini's mean difference: Let $h(x, y) = |x - y|$, then $\theta = E|X_1 - X_2|$ and $\sigma^2 = E\tilde{h}^2(X_1) - \theta^2$ with $\tilde{h}(t) = E|X_1 - t|$. This kernel can be used to detect the change in the average difference. It might be a more robust procedure in determining the change in scale than using the kernel $(x - y)^2$.
- Anti-symmetric Kernels:
  1. To detect a change in mean, define $h(x, y) = x - y$. It follows that $\mu = EX_1 - EX_{\tau+1}$ and $\sigma^2 = Var(X_1)$. The $V_n^{(2)}(k)$ is essentially constructed

by the difference between two sample means based on the first k observations and the last $n-k$ observations.

2. Let $h(x, y) = \text{sgn}(x-y)$. It follows that $\mu = P(X_1 > X_{\tau+1}) - P(X_1 < X_{\tau+1})$ and $\sigma^2 = 4Var(F(X_1)) = \frac{1}{3}$. Hence, it can used to detect the change in the probability whether the random variables have the tendency to increase or decrease.
3. Let $h(x, y) = x^m - y^m$, where m is any an integer. It follows that $\mu = EX_1^m - EX_{\tau+1}^m$ and $\sigma^2 = Var(X_1^m)$. We can use this kernel to detect the change in the mth moment.

We do not have a single rule that fits all situations in general to select a kernel function in applications. The problem of choosing an appropriate kernel for detecting changes in moment is simple. If the robustness is of concern, $h(x, y) = \text{sgn}(x-y)$ can be a good choice for location change. We may let $h(x, y) = \text{sgn}(x-y)\min\{|x-y|, M\}$ with a large constant M to better compromise between the efficiency and robustness. In general, the applicant must choose a kernel function in conjunction with his or her scientific objection.

In the following simulation study, we choose $h(x, y) = x - y$ and $x^2 - y^2$ to detect the change in the mean or change in the second moment, respectively.

4. Simulation Study

In this section, we use simulation to investigate finite sample properties and assess the performance of the U-statistic based MIC method. Firstly, we conduct a simulation to compare the estimators of change point and then the powers of this method to others, such as the likelihood based MIC, BIC, and the (unmodified) U-statistic methods.

Both simulation experiments were done by generating data from following five models:

- Model 1: Normal model with a change 0.5 in the mean;
- Model 2: Normal model with a change of factor 2 in the variance;
- Model 3: Exponential model with a change of factor $\sqrt{2}$ in the mean;
- Model 4: Normal model with a change 0.5 in the mean, and a change of factor 2 in the variance;
- Model 5: Gamma model with a change $\sqrt{2}-1$ in the mean, and a change of factor 2 in the variance.

These models are denoted as M1–M5 in Tables 1–4. The sample sizes are chosen to be $n = 60$, $n = 100$, and $n = 200$. Under the alternative hypothesis, the change points are placed at 10%n, 15%n, 20%n, 25%n, and 50%n in the sequence, respectively. As discussed in Sec. 3.3, we choose the kernel function $h(x, y) = x - y$ for the first, third, and fifth models, and $h(x, y) = x^2 - y^2$ for the second and fourth models. Both $h(x, y) = x - y$ and $h(x, y) = x^2 - y^2$ seem appropriate for Model 5 if the shape parameter of the model is fixed. Because $\sum X_i$ is a complete sufficient statistic in this case, the choice of $h(x, y) = x - y$ is most efficient. This is confirmed by our unreported simulation that the choice of $h(x, y) = x^2 - y^2$ is less efficient. T1–T4

The nominal levels α are chosen to be 0.05 and 0.10. The simulation is repeated 5,000 times for each combination of the sample size, location of change, nominal level, and model.

Table 1
The comparison of $\widehat{P}(|\hat{\tau} - \tau| < n\delta)$ for $\tau = 50\%n$ in U-statistic based MIC and others by using $h(x, y) = x - y$ in Models 1, 3, and 5, $h(x, y) = x^2 - y^2$ in Models 2 and 4

		$n = 60$					$n = 100$					$n = 200$				
	δ	0.10	0.15	0.20	0.25	0.30	0.10	0.15	0.20	0.25	0.30	0.10	0.15	0.20	0.25	0.30
M1	MIC	62.1	73.8	81.5	86.8	90.8	70.1	80.8	87.5	91.5	94.3	81.5	90.2	94.8	97.0	98.5
	BIC	45.0	54.3	61.5	67.8	73.3	56.2	65.8	72.5	77.4	81.6	74.1	82.7	87.9	91.0	93.5
	U_{MIC}	60.7	72.7	80.7	86.1	90.6	68.8	79.7	86.3	91.1	94.0	81.7	89.7	93.8	96.5	97.9
	U	45.4	55.2	62.3	68.5	74.9	56.0	65.8	72.2	78.2	82.7	74.5	82.6	87.2	90.3	92.7
M2	MIC	57.8	69.7	78.7	84.8	88.7	67.5	78.6	85.4	90.0	93.2	79.9	88.6	93.1	95.6	97.3
	BIC	39.6	49.1	57.3	63.6	69.3	53.0	63.1	69.7	75.1	79.7	71.4	79.8	84.8	88.4	91.0
	U_{MIC}	55.0	66.7	75.5	81.7	86.3	62.2	72.7	79.8	84.3	88.3	72.4	81.6	86.9	90.1	92.7
	U	41.2	51.2	60.0	67.0	74.0	49.6	59.4	66.3	71.7	77.5	64.0	72.8	78.5	82.6	86.3
M3	MIC	50.0	63.4	73.1	80.3	85.8	55.9	68.6	77.7	84.0	88.6	68.2	79.3	86.4	90.8	93.6
	BIC	28.8	38.1	45.4	52.4	59.7	35.3	44.1	52.0	58.9	64.8	51.0	60.5	67.4	72.4	76.9
	U_{MIC}	49.1	62.9	72.3	79.4	86.1	56.1	68.1	76.5	82.2	86.9	66.4	77.1	84.2	88.6	91.7
	U	30.4	40.6	48.8	56.6	65.4	37.8	46.9	54.9	60.9	67.9	50.7	60.1	67.1	72.5	77.1
M4	MIC	53.4	63.5	70.3	75.6	79.4	68.9	77.9	83.4	86.9	89.1	87.2	93.2	95.3	96.8	97.7
	BIC	40.0	47.6	53.6	58.4	62.8	57.9	65.8	71.0	74.5	77.3	81.9	88.1	90.3	92.2	93.6
	U_{MIC}	58.7	70.1	77.5	83.5	88.2	66.4	76.3	81.9	86.3	90.0	78.4	85.4	89.3	92.0	94.0
	U	46.8	57.0	64.5	71.6	78.4	56.3	65.6	71.6	77.0	81.8	72.3	79.7	84.0	87.0	89.6
M5	MIC	58.6	70.7	80.0	86.2	90.3	68.6	79.7	86.7	91.1	94.0	81.3	89.3	93.5	96.2	97.6
	BIC	42.2	51.6	60.0	66.7	72.6	54.1	63.7	70.9	75.8	80.0	73.5	81.4	86.0	89.9	92.0
	U_{MIC}	58.2	70.2	78.6	85.1	89.2	66.1	76.9	84.2	89.3	92.5	79.5	87.1	92.6	95.0	96.8
	U	43.7	53.5	61.8	68.9	74.9	53.7	63.1	70.2	76.4	81.2	71.7	79.1	84.8	87.9	90.5

Table 2
The comparison of $\widehat{P}(|\hat{\tau} - \tau| < n\delta)$ for $\tau = 25\%n$ in U-statistic based MIC and others by using $h(x, y) = x - y$ in Models 1, 3, and 5, $h(x, y) = x^2 - y^2$ in Models 2 and 4

		$n = 60$					$n = 100$					$n = 200$				
	δ	0.10	0.15	0.20	0.25	0.30	0.10	0.15	0.20	0.25	0.30	0.10	0.15	0.20	0.25	0.30
M1	MIC	46.1	57.2	66.3	73.2	78.7	56.9	68.2	75.2	80.8	85.1	73.2	81.0	85.9	89.9	92.3
	BIC	44.5	55.7	66.1	72.0	74.3	55.1	65.9	73.7	80.1	81.6	72.8	80.7	86.0	90.1	91.3
	U_{MIC}	44.3	55.0	64.5	70.8	76.9	56.0	66.7	74.8	80.4	84.7	72.7	81.3	87.2	91.2	93.7
	U	44.8	56.4	68.6	71.1	73.9	55.9	66.3	74.6	79.8	81.6	72.9	81.5	87.0	91.1	92.2
M2	MIC	42.3	51.9	61.3	69.1	75.1	54.2	63.9	71.5	78.4	82.5	71.4	79.1	84.5	88.5	91.3
	BIC	40.8	51.4	62.3	69.6	72.0	51.9	61.4	70.6	78.6	80.3	70.3	78.3	83.8	88.5	89.6
	U_{MIC}	48.9	61.3	75.5	79.9	84.3	56.0	67.4	78.7	86.6	89.6	70.4	79.7	87.2	94.5	96.0
	U	48.6	62.3	79.5	81.3	83.4	54.9	67.1	79.9	88.7	89.8	66.5	77.2	85.6	95.2	95.7
M3	MIC	32.6	43.2	53.5	63.0	69.4	38.6	49.3	58.5	67.4	74.2	51.8	62.3	70.1	76.8	81.8
	BIC	31.4	42.3	54.6	62.1	64.5	36.5	46.8	56.9	67.1	69.2	49.1	59.0	66.6	76.6	78.4
	U_{MIC}	36.0	48.2	61.6	68.3	74.6	41.5	53.3	64.3	74.0	79.8	54.0	65.3	74.5	84.0	88.4
	U	36.2	49.5	66.2	68.9	71.4	40.4	52.8	65.3	75.5	77.5	51.7	63.2	73.3	85.2	86.5
M4	MIC	45.9	54.8	63.4	72.9	76.0	60.3	68.0	73.8	82.3	85.1	81.9	88.3	91.6	94.6	95.7
	BIC	40.0	49.6	59.0	71.0	72.5	55.1	62.6	69.2	80.4	81.8	79.4	85.3	88.9	93.7	94.3
	U_{MIC}	55.4	67.4	81.0	85.0	88.3	62.7	74.3	84.0	91.8	94.0	77.5	86.1	92.2	97.8	98.7
	U	54.6	68.5	84.8	86.5	88.0	60.8	72.9	83.9	93.0	93.9	74.6	83.7	90.7	98.1	98.5
M5	MIC	43.7	54.5	63.7	70.9	76.8	53.4	63.7	72.1	78.6	83.4	71.9	80.4	85.7	89.6	92.3
	BIC	42.7	52.9	63.9	70.6	72.8	51.9	62.0	71.0	78.2	80.4	70.6	79.5	84.9	89.5	90.9
	U_{MIC}	46.7	58.9	71.1	76.0	80.8	56.1	67.3	76.6	83.4	87.5	72.4	82.1	87.9	93.2	94.9
	U	46.2	58.9	74.4	76.7	78.9	54.4	65.9	77.1	84.6	86.1	69.8	80.0	86.6	93.5	94.3

Table 3
The comparison of $\widehat{P}(|\hat{\tau} - \tau| < n\delta)$ for $\tau = 15\%n$ in U-statistic based MIC and others by using $h(x, y) = x - y$ in Models 1, 3, and 5, $h(x, y) = x^2 - y^2$ in Models 2 and 4

		$n = 60$					$n = 100$					$n = 200$				
	δ	0.10	0.15	0.20	0.25	0.30	0.10	0.15	0.20	0.25	0.30	0.10	0.15	0.20	0.25	0.30
M1	MIC	31.3	38.1	43.9	50.4	57.3	41.3	48.9	54.1	59.7	65.4	59.1	66.7	71.2	75.4	79.8
	BIC	46.8	55.0	58.4	61.1	63.7	53.5	64.2	67.1	69.2	71.2	69.8	79.0	81.4	83.1	84.7
	U_{MIC}	34.8	40.3	46.3	52.6	60.0	41.6	48.3	54.4	59.6	64.9	59.0	66.0	70.7	75.6	79.8
	U	52.5	56.4	59.8	63.0	65.9	54.7	63.5	67.0	69.6	71.6	69.9	78.8	81.2	83.4	84.7
M2	MIC	31.7	38.5	44.1	50.0	56.3	39.6	47.2	52.5	57.8	63.8	56.0	63.3	68.4	73.2	77.7
	BIC	45.3	55.1	58.0	60.7	62.8	53.5	66.0	68.4	70.5	72.0	67.1	78.2	80.1	81.7	82.9
	U_{MIC}	50.0	54.3	58.6	63.2	68.2	53.2	64.3	67.9	71.3	75.7	66.4	80.0	82.6	85.0	87.4
	U	64.3	67.7	70.4	72.8	74.6	63.0	76.7	78.6	80.0	81.6	71.4	87.9	88.9	90.0	90.8
M3	MIC	22.3	28.4	34.5	41.2	48.9	26.2	33.6	39.5	45.7	52.6	37.7	46.0	51.2	57.0	63.1
	BIC	37.0	46.3	49.7	52.6	55.1	39.2	52.6	55.4	58.1	60.3	50.5	65.1	67.3	69.3	71.1
	U_{MIC}	34.4	38.9	44.3	50.1	56.7	37.6	46.6	51.7	57.1	62.9	45.3	56.9	61.3	65.7	70.6
	U	50.1	53.7	57.0	60.1	62.6	49.7	62.1	64.9	67.2	69.0	55.0	71.6	73.7	75.7	77.3
M4	MIC	39.1	51.3	55.1	58.9	62.7	48.2	61.1	64.5	67.9	71.0	70.8	79.6	82.7	85.1	87.3
	BIC	44.7	59.5	61.5	63.4	64.9	52.0	68.8	70.3	71.6	72.7	74.0	85.3	86.7	87.6	88.2
	U_{MIC}	58.1	61.5	65.3	69.2	73.2	61.3	72.9	75.5	78.6	81.5	74.8	88.1	89.9	91.4	92.9
	U	69.7	72.3	74.6	76.4	78.2	69.1	82.8	84.0	85.3	86.3	78.2	93.4	94.2	94.8	95.2
M5	MIC	31.8	38.2	43.9	50.0	57.0	40.0	46.8	52.9	58.0	64.1	56.9	64.7	69.5	74.0	77.9
	BIC	46.0	54.8	57.6	60.2	62.9	52.5	63.3	66.6	68.8	70.9	67.9	78.2	80.4	82.1	83.2
	U_{MIC}	43.6	48.5	53.5	59.0	65.0	48.4	56.9	61.3	66.3	71.3	63.4	73.8	77.4	80.3	83.8
	U	58.4	62.0	65.0	67.8	70.1	59.5	70.6	73.2	75.6	77.4	70.8	84.9	86.7	87.8	89.0

Table 4
The powers comparison between U-statistic based MIC and others for $\alpha = 0.05$ by using $h(x, y) = x - y$ in Models 1, 3, and 5, $h(x, y) = x^2 - y^2$ in Models 2 and 4

		$n = 60$					$n = 100$					$n = 200$				
	k	6	9	12	15	30	10	15	20	25	50	20	30	40	50	100
M1	MIC	10.3	14.6	19.8	24.6	37.8	14.5	22.2	30.6	38.4	55.2	27.4	44.4	57.8	69.8	87.1
	BIC	11.2	15.9	18.7	22.3	30.6	18.2	25.0	31.9	36.9	48.9	32.3	46.0	55.7	66.1	79.4
	U_{MIC}	9.7	14.3	19.4	23.5	34.5	14.5	23.0	30.1	38.7	55.5	27.9	44.7	59.5	69.1	88.0
	U	10.5	15.1	19.1	21.7	28.5	16.4	23.3	29.6	35.1	46.2	32.0	46.3	57.4	64.6	81.1
M2	MIC	10.3	14.8	19.9	22.2	31.8	16.0	23.9	29.9	37.8	51.7	30.0	45.1	57.5	67.4	84.6
	BIC	11.7	15.9	19.8	20.6	25.0	17.3	24.6	28.3	34.6	41.9	33.5	46.4	56.3	63.6	75.6
	U_{MIC}	18.6	22.9	26.2	26.6	22.3	28.0	34.1	37.5	42.1	40.5	44.9	55.7	63.3	67.9	75.5
	U	21.2	24.2	26.8	26.6	18.2	29.4	34.8	35.8	38.8	28.2	46.7	55.5	60.4	62.9	57.8
M3	MIC	7.1	9.0	10.7	13.1	18.4	9.1	12.7	16.0	19.8	29.6	14.8	22.8	31.4	37.3	55.9
	BIC	7.7	9.1	10.5	12.2	14.4	10.3	13.0	16.0	18.1	23.4	16.2	23.7	30.0	33.4	43.9
	U_{MIC}	12.9	14.7	15.9	17.1	16.5	15.5	18.3	21.2	23.1	23.8	23.0	29.1	35.9	39.8	48.0
	U	15.0	15.6	16.7	17.5	13.8	16.9	19.8	21.6	21.6	17.5	25.7	30.5	35.3	35.9	35.8
M4	MIC	10.2	13.1	15.4	19.9	25.0	17.9	25.7	34.5	40.1	54.9	36.5	54.1	68.0	78.3	91.4
	BIC	9.7	11.7	13.0	16.1	17.0	16.2	22.4	27.8	31.9	38.8	34.9	49.6	61.3	70.6	81.6
	U_{MIC}	24.7	29.6	33.0	36.8	31.0	35.2	43.9	48.4	51.8	51.2	55.6	68.2	76.3	82.4	88.8
	U	26.7	31.0	32.9	35.4	24.1	37.8	45.7	47.6	48.8	38.3	58.8	68.9	75.2	79.1	77.6
M5	MIC	9.9	15.7	20.6	24.6	34.7	14.5	22.3	29.9	35.9	53.0	28.6	45.5	58.0	69.4	85.3
	BIC	12.2	17.6	21.2	24.3	31.7	17.3	24.5	31.5	36.5	46.3	33.6	46.5	56.8	65.9	79.0
	U_{MIC}	16.8	21.4	25.8	26.7	31.8	22.5	28.7	35.3	39.5	47.7	38.3	53.1	62.1	70.7	84.1
	U	18.6	22.9	26.7	25.9	25.9	25.4	31.1	35.6	37.3	38.9	41.2	53.4	59.8	66.4	73.1

The corresponding results for the U-statistic based MIC, U-statistic, likelihood-based MIC and BIC methods in percentages are placed in the columns of U_{MIC}, U, MIC, and BIC in Tables 1–4.

4.1. *Comparison of Estimator of Change Point*

The modified information criterion is expected to have better efficiency at estimating the change point τ than other methods if it is close to the middle of the sequence. It is important to investigation its efficiency when τ is at the beginning or end of the sequence.

We calculated the corresponding proportions of $|\hat{\tau} - \tau| \leq n\delta$ in 5,000 repetitions for a number of choices of δ, denoted as $\widehat{P}(|\hat{\tau} - \tau| < n\delta)$. We present the results for $\delta = 50\%,\ 25\%$, and 15% in Tables 1–3. We use $\hat{\tau}_{U_M}$, $\hat{\tau}_{MIC}$, $\hat{\tau}_{BIC}$, and $\hat{\tau}_U$ for estimators based on modified U, MIC, BIC, and unmodified U methods, respectively. From these results, we conclude that:

1. The probability $P\{|\hat{\tau} - \tau| \leq n\delta\}$ increases as n increases in all cases;
2. when $\tau = 50\%n$, we have in all models

$$\begin{aligned} &\widehat{P}\{|\hat{\tau}_{U_M} - \tau| \leq n\delta\} \quad \text{and} \quad \widehat{P}\{|\hat{\tau}_{MIC} - \tau| \leq n\delta\} \\ &\quad \geq \widehat{P}\{|\hat{\tau}_{BIC} - \tau| \leq n\delta\} \quad \text{and} \quad \widehat{P}\{|\hat{\tau}_U - \tau| \leq n\delta\}. \end{aligned} \tag{11}$$

   That is, the modified U and the MIC are more efficient estimators compared to the unmodified U and the BIC in almost all cases.
3. When $\tau = 25\%n$, Eq. (11) is true, or there is no difference among the four methods in model 1. That is, the modified U and the MIC are more efficient or comparable estimators to other two methods in Model 1. However, in Models 2–5, we find:

$$\begin{aligned} &\widehat{P}\{|\hat{\tau}_{U_M} - \tau| \leq n\delta\} \approx \widehat{P}\{|\hat{\tau}_U - \tau| \leq n\delta\} \\ &\quad \geq \widehat{P}\{|\hat{\tau}_{MIC} - \tau| \leq n\delta\} \approx \widehat{P}\{|\hat{\tau}_{BIC} - \tau| \leq n\delta\}. \end{aligned} \tag{12}$$

   That is, the modified and unmodified U estimators are more efficient.
4. When $\tau = 15\%n$, the outcomes are mixed. The unmodified U seems to out perform, and the modified U is comparable to other methods in Models 2–5.

4.2. *Power Comparison*

Under the same simulation setup described above, the powers are calculated for each method. However, we only present the results for nominal level 0.05 in Table 4.

The results in Table 4 provide some additional information on the methods considered. First, all methods seem to be consistent, and their powers increase significantly as the sample size increases. Second, all methods have better powers in detecting the change when the change point is located around the middle of the sequence. Third, the performance comparison between the U-statistic based MIC and the likelihood based MIC is not always in favor of the likelihood based MIC (see Models 4 and 5 in Table 1) even though it is often so as expected. In detail, the U-statistic based MIC has better powers compared to the likelihood based MIC when the change appears early or late in the sequence. When the change is located in the middle of the sequence, the likelihood based MIC has marginally better or comparable powers. Finally, the U-statistic based MIC has comparable powers for change appearing early or late, and has significant better powers for change appearing around the middle compared to U-statistic. It is similar when comparing the likelihood-based MIC to BIC method. This is expected because the main difference between the MIC and other traditional information criteria is the preference of the MIC for the model with change located in the middle of the sequence. We also notice that the U-statistics based MIC method has consistently better powers compared to BIC method in all cases in Models 3, 4, and 5 and most of the cases in Models 1 and 2.

We conclude that the U-statistic based MIC method is comparable to or sometimes better than the likelihood-based MIC and U-statistic methods when some suitable kernels are identified, and better than the BIC method in most of the cases. Hence, we suggest using the U-statistic based MIC rather than the likelihood-based MIC, the BIC, and the (unmodified) U-statistic methods when we do not have sufficient knowledge about the physical background of the sample.

Appendix: Proofs of the Main Results

A.1 Existing Results

One commonly used approach in large sample theory is to link the statistic under investigation to a summation of independent random variables. In the literature of U-statistics, it is known as the projection method.

Let h be a symmetric kernel function of order 2 (the general result is also true) and $X_1, \ldots, X_n$ be an iid sample. Assume that $E[h(X_1, X_2)]^2 < \infty$ and $Eh(X_1, X_2) = 0$. Define

$$T_n = \sum_{1 \le i < j \le n} h(X_i, X_j)$$

and the projection of $h(X_1, X_2)$ in the σ-algebra of X_1 as:

$$\tilde{h}(X_1) = E[h(X_1, X_2) \mid X_1].$$

Let

$$P_n = \sum_{1 \le i < j \le n} [\tilde{h}(X_i) + \tilde{h}(X_j)] = (n-1) \sum_{i=1}^{n} \tilde{h}(X_i).$$

Note that P_n is a summation of independent random variables, which is regarded as a projection of T_n.

It turns out that the difference between P_n and T_n is not large compared to the values of P_n or T_n as $n \to \infty$. More precisely, we have the following theorem by Hall (1979).

Lemma A.1. *With the notation and assumptions stated in the Appendix, we have*:

$$\max_{\mathbf{1<k\le n}} |T_k - P_k| = O_p(n).$$

Based on this result, it becomes possible for us to study the property of the U-statistics through that of sum of independent random variables. The next result from Gombay and Horváth (1995) further approximates a U-statistic based stochastic process with a well-known Brownian bridge.

For each given k, let $\hat{\theta}_1(k)$ and $\hat{\theta}_2(k)$ be defined as in (2). Define, as in Gombay and Horváth (1995), for $\frac{2}{n+1} \leq t \leq \frac{n-2}{n+1}$:

$$Q_n(t) = \frac{n^{1/2}}{2\sigma} t(1-t)\left\{\hat{\theta}_1([(n+1)t]) - \hat{\theta}_2([(n+1)t])\right\}, \tag{13}$$

and $Q_n(t) = 0$, otherwise. We have the following result from Gombay and Horváth (1995).

Lemma A.2. *Assume that* $E|h(X_1, X_2)|^v < \infty$ *for some* $v > 2$ *and* $\sigma^2 = Var[\tilde{h}(X)] > 0$. *Then there exists a sequence of Brownian bridges* $\{B_n(t), 0 \leq t \leq 1\}$ *such that*:

$$\sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \frac{|Q_n(t) - B_n(t)|}{[t(1-t)]^{1/2-\delta}} = O_p(n^{-\delta})$$

for all $0 \leq \delta < \frac{1}{2} - \frac{1}{v}$.

Obviously, we have

$$\sup_{c_1 \leq t \leq c_2} |Q_n(t) - B_n(t)| = O_p(n^{-\delta}) \tag{14}$$

and

$$\sup_{\frac{1}{n} \leq t \leq \frac{n-1}{n}} \frac{|Q_n(t) - B_n(t)|}{[t(1-t)]^{1/2}} = O_p(1), \tag{15}$$

where $0 < c_1 \leq c_2 < 1$ are two constants. The results enable us to assess the order of $V_n^{(1)}(k)$ defined in (3) conveniently with the help of the next result which is from Csörgö and Révész (1981).

Lemma A.3. *Let* ϵ_n *be a decreasing sequence of numbers such that* $\epsilon_n \to 0$. *Then, for all real* y:

$$\lim_{n\to\infty} P\left\{ \sup_{\epsilon_n < t < 1-\epsilon_n} \frac{B(t)}{\sqrt{t(1-t)}} \leq a\left(y, 2\log\frac{1-\epsilon_n}{\epsilon_n}\right)\right\} = \exp(-e^{-y}),$$

$$\lim_{n\to\infty} P\left\{ \sup_{\epsilon_n < t < 1-\epsilon_n} \frac{|B(t)|}{\sqrt{t(1-t)}} \leq a\left(y, 2\log\frac{1-\epsilon_n}{\epsilon_n}\right)\right\} = \exp(-2e^{-y}),$$

where $\{B(t), 0 \leq t \leq 1\}$ *is a sequence of Brownian bridges, and*

$$a(y, T) = \left(y + 2\log T + \frac{1}{2}\log\log T - \frac{1}{2}\log\pi\right)(2\log T)^{-1/2}.$$

By taking $\epsilon_n = \frac{\mathbf{1}}{\mathbf{n+1}}$, we will be able to show that:

$$\sup_{\frac{1}{n}\le t\le \frac{n-1}{n}} \frac{|B(t)|}{\sqrt{t(1-t)}} = O_p[(\log\log n)^{1/2}]. \tag{16}$$

This with the following classical result from Darling and Erdös (1956) are very handy in our future proof.

Lemma A.4. *Let $X_1, \ldots, X_n$ be independent random variables with mean 0 and variance 1, and a uniformly bounded third absolute moment. Put $R_n = \sum_{i=1}^n X_i$ and let*

$$U_n = \max_{1\le k\le n} \frac{R_k}{\sqrt{k}}.$$

Then:

$$\lim_{n\to\infty} P\{U_n < b(y, \log n)\} = \exp(-e^{-y}/2\pi^{1/2}),$$

for any $-\infty < y < \infty$, where

$$b(y, T) = (2\log T)^{1/2} + \frac{\log\log T}{2(2\log T)^{1/2}} + \frac{y}{(2\log T)^{1/2}}.$$

By taking $y = \log\log n$, we have:

$$\max_{1\le k\le n} \frac{R_k}{\sqrt{k}} = O_p[(\log\log n)^{1/2}] \tag{17}$$

which will be used to prove the consistency of $\hat{\sigma}^2_{k1}$.

In the following lemmas, we assume that h is an anti-symmetric kernel. Theorem 3.2 can be proved with the help of the following Lemmas A.5–A.7 from Csörgö and Horváth (1997). Lemma A.5 implies that the penalty is a prominent term in $U_n^{(2)}$ if the null model is true, which is the key to prove the limiting distribution of test statistic.

Lemma A.5. *Under the null hypothesis H_0, assume that* (8) *and* (9) *hold, and*

$$E\{\tilde{h}^2(X_1)\log\log[|\tilde{h}(X_1)| + 1]\} < \infty,$$

then we have:

$$\lim_{n\to\infty} P\left\{A(\log n)\max_{1\le k<n} \frac{Z_k}{\sigma\sqrt{nk(n-k+1)}} \le y + D(\log n)\right\} = \exp(-e^{-y})$$

and

$$\lim_{n\to\infty} P\left\{A(\log n)\max_{1\le k<n} \frac{|Z_k|}{\sigma\sqrt{nk(n-k+1)}} \le y + D(\log n)\right\} = \exp(-2e^{-y})$$

for all real y, where

$$A(x) = \sqrt{2 \log x}$$

and

$$D(x) = 2 \log x + \frac{1}{2} \log \log x - \frac{1}{2} \log \pi.$$

By taking $y = \log \log \log n$, Lemma A.5 implies that:

$$\max_{1 \le k \le n-1} \frac{|Z_k|}{\sqrt{nk(n-k)}} = O_p[(\log \log n)^{1/2}]. \tag{18}$$

Lemma A.6. *Under the conditions of Lemma* A.5, *there exists a sequence of Brownian bridges* $\{B_n(t), 0 \le t \le 1\}$ *such that*:

$$\sup_{0<t<1} \left| \frac{Z_{[(n+1)t]}}{\sigma n^{3/2}} - B_n(t) \right| = o_p(1).$$

Lemma A.7. *Assume that* (7)–(8) *hold, then we have under the alternative hypothesis* H_1 *there exists one change point at* $\tau = [n\lambda]$,

$$\frac{1}{n^2} Z_\tau \to \lambda(1-\lambda)\mu$$

in probability.

A.2 The Consistency of $\hat{\sigma}_{k1}^2$ *and* $\hat{\sigma}_{k2}^2$

In this subsection, we present the proofs for the consistency of $\hat{\sigma}_{k1}^2$ and $\hat{\sigma}_{k2}^2$ (Propositions 3.1 and 3.2) with the help of Lemmas in Sec. A.1.

The Proof of Proposition 3.1.

Part 1. To show that

$$\max_{\mathbf{1 \le k \le n}} |\hat{\sigma}_{k1}^2 - \sigma^2| = o_p(1) \tag{19}$$

under the null hypothesis H_0. Let

$$I_1(k) = \sum_{j=1}^{k} [h_{k1}(X_j) - \hat{\theta}_1(k)]^2, \quad I_2(k) = \sum_{j=k+1}^{n} [h_{k2}(X_j) - \hat{\theta}_2(k)]^2,$$

then, $\hat{\sigma}_{k1}^2 = \frac{1}{n}[I_1(k) + I_2(k)]$.

It is obvious that

$$\max_{2 \le k \le n-2} |\hat{\sigma}_{k1}^2 - \sigma^2| = o_p(1) \tag{20}$$

implies (19). We now prove (20) by considering $k \le \sqrt{n}(\log n)^{-1}$ and $k > \sqrt{n}(\log n)^{-1}$, separately. Note that:

$$I_1(k) = \sum_{j=1}^{k} \left\{ \frac{1}{k-1} \sum_{1 \le i \le k, i \ne j} [h(X_i, X_j) - \theta_1] \right\}^2 - k[\hat{\theta}_1(k) - \theta_1]^2.$$

By Kolmogorov Maximal Inequality, we have:

$$\max_{k \le \sqrt{n}(\log n)^{-1}} \left| \sum_{i=1}^{k} [\tilde{h}(X_i) - \theta_1] \right| = O_p[n^{1/4}(\log n)^{-1/2}]. \tag{21}$$

Also, it is obvious that $k[\hat{\theta}_1(k) - \theta_1]^2$ is $O_p(1)$ if k is finite. Hence, we assume that k is large enough, then we have by Lemma A.1 and Eq. (21):

$$\begin{aligned}
&\max_{k \le \sqrt{n}(\log n)^{-1}} \{k[\hat{\theta}_1(k) - \theta_1]^2\} \\
&\le \max_{k \le \sqrt{n}(\log n)^{-1}} \frac{C}{(k-1)^3} \left\{ \sum_{1 \le i < j \le k} [h(X_i, X_j) - \theta_1] \right\}^2 \\
&\le \max_{k \le \sqrt{n}(\log n)^{-1}} \frac{C}{(k-1)^3} \left\{ (k-1) \sum_{i=1}^{k} [\tilde{h}(X_i) - \theta_1] + O_p[\sqrt{n}(\log n)^{-1}] \right\}^2 \\
&\le 2C \max_{k \le \sqrt{n}(\log n)^{-1}} \left\{ \sum_{i=1}^{k} [\tilde{h}(X_i) - \theta_1] \right\}^2 + O_p[n(\log n)^{-2}] \\
&= O_p[n(\log n)^{-2}].
\end{aligned} \tag{22}$$

For $j = 1, \ldots, k$, denote that

$$W_{jk} = \frac{1}{k-1} \sum_{1 \le i \le k, i \ne j} \{[h(X_i, X_j) - \theta_1] - [\tilde{h}(X_j) - \theta_1]\}.$$

Then from (22), we have uniformly for $k \le \sqrt{n}(\log n)^{-1}$:

$$\begin{aligned}
|I_1(k) - k\sigma^2| &\le \left| \sum_{j=1}^{k} \left\{ W_{jk} + [\tilde{h}(X_j) - \theta_1] \right\}^2 - k\sigma^2 \right| + O_p[n(\log n)^{-2}] \\
&\le \left| \sum_{j=1}^{k} \{[\tilde{h}(X_j) - \theta_1]^2 - \sigma^2\} \right| + \sum_{j=1}^{k} W_{jk}^2 \\
&\quad + 2 \left| \sum_{j=1}^{k} W_{jk} [\tilde{h}(X_j) - \theta_1] \right| + O_p[n(\log n)^{-2}] \\
&\le 2 \sum_{j=1}^{k} W_{jk}^2 + O_p[n(\log n)^{-2}],
\end{aligned} \tag{23}$$

the last equality is due to, for $k \le \sqrt{n}(\log n)^{-1}$,

$$2\left|\sum_{j=1}^{k} W_{jk}[\tilde{h}(X_j)-\theta_1]\right| \le \sum_{j=1}^{k} W_{jk}^2 + \sum_{j=1}^{k}[\tilde{h}(X_j)-\theta_1]^2$$
$$= \sum_{j=1}^{k} W_{jk}^2 + O_p[\sqrt{n}(\log n)^{-1}].$$

We now claim that:

$$\max_{k\le\sqrt{n}(\log n)^{-1}} \sum_{j=1}^{k} W_{jk}^2 = O_p[n(\log n)^{-1}]. \tag{24}$$

Since $\{h(X_i, X_j), i = 1, \ldots, k, i \neq j\}$ are conditionally independent given X_j and $EW_{jk}^4 < \infty$, we have by Kolmogorov inequality:

$$P\left\{\max_{k\le\sqrt{n}(\log n)^{-1}} \sum_{j=1}^{k} W_{jk}^2 > n(\log n)^{-1}\right\} \le \sum_{k=1}^{\sqrt{n}(\log n)^{-1}} \sum_{j=1}^{k} P\{W_{jk}^2 > n(k\log n)^{-1}\}$$
$$\le C(\log n)^2/n^2 \sum_{k=1}^{\sqrt{n}(\log n)^{-1}} k^3 = C(\log n)^{-2} \to 0.$$

Hence, (24) follows. Equations (23) and (24) imply that:

$$\max_{k\le\sqrt{n}(\log n)^{-1}} |I_1(k) - k\sigma^2| = O_p[n(\log n)^{-1}]. \tag{25}$$

For $k > \sqrt{n}(\log n)^{-1}$, we have by the Extension of the Kolmogorov Maximal Inequality for the reverse martingale (see Sen and Singer, 1993),

$$\max_{k>\sqrt{n}(\log n)^{-1}} |W_{jk}| = O_p[n^{-1/4}\log n], \tag{26}$$

uniformly for j, and by (17) in Lemma A.4:

$$\max_{k>\sqrt{n}(\log n)^{-1}} \frac{1}{\sqrt{k}}\left|\sum_{i=1}^{k}[\tilde{h}(X_i)-\theta_1]\right| = O_p[(\log\log n)^{1/2}]. \tag{27}$$

Hence, by Lemma A.1 and (27):

$$\max_{k>\sqrt{n}(\log n)^{-1}} \{k[\hat{\theta}_1(k)-\theta_1]^2\} \le \max_{k>\sqrt{n}(\log n)^{-1}} \frac{C}{k^3}\left\{k\sum_{i=1}^{k}[\tilde{h}(X_i)-\theta_1] + O_p(n)\right\}^2$$
$$\le C \max_{k>\sqrt{n}(\log n)^{-1}} \left\{\frac{1}{\sqrt{k}}\sum_{i=1}^{k}[\tilde{h}(X_i)-\theta_1]\right\}^2$$
$$+ O_p(n^2)[\sqrt{n}(\log n)^{-1}]^{-3}$$
$$= O_p[n^{1/2}(\log n)^3]. \tag{28}$$

Similar to the discussion in (23), by using (28), we have for $k > \sqrt{n}(\log n)^{-1}$:

$$\begin{aligned}|I_1(k) - k\sigma^2| &\le \left|\sum_{j=1}^{k}\left\{W_{jk} + [\tilde{h}(X_j) - \theta_1]\right\}^2 - k\sigma^2\right| + O_p[n^{1/2}(\log n)^3] \\ &\le \left|\sum_{j=1}^{k}\{[\tilde{h}(X_j) - \theta_1]^2 - \sigma^2\}\right| + \sum_{j=1}^{k} W_{jk}^2 \\ &\quad + 2\sqrt{\sum_{j=1}^{k} W_{jk}^2 \cdot \sum_{j=1}^{k}[\tilde{h}(X_j) - \theta_1]^2} + O_p[n^{1/2}(\log n)^3]. \end{aligned} \tag{29}$$

(26) implies that

$$\max_{k>\sqrt{n}(\log n)^{-1}} \sum_{j=1}^{k} W_{jk}^2 = O_p[n^{1/2}(\log n)^2]. \tag{30}$$

It is obvious that

$$\max_{k>\sqrt{n}(\log n)^{-1}} \sum_{j=1}^{k}[\tilde{h}(X_j) - \theta_1]^2 = O_p(n). \tag{31}$$

Thus, (29)–(31) and Kolmogorov Maximal Inequality indicate that:

$$\max_{k>\sqrt{n}(\log n)^{-1}} |I_1(k) - k\sigma^2| = O_p[n^{3/4}(\log n)]. \tag{32}$$

Hence, we have from (25) and (32):

$$\max_{2\le k\le n-2} |I_1(k) - k\sigma^2| = O_p[n(\log n)^{-1}].$$

Similarly,

$$\max_{2\le k\le n-2} |I_2(k) - (n-k)\sigma^2| = O_p[n(\log n)^{-1}].$$

Thus, we complete the proof of Part 1 because

$$\max_{2\le k\le n-2} |\hat{\sigma}_{k1}^2 - \sigma^2| \le \frac{1}{n}\max_{2\le k\le n-2} |I_1(k) - k\sigma^2| + \frac{1}{n}\max_{2\le k\le n-2} |I_2(k) - (n-k)\sigma^2| = o_p(1).$$

Part 2. The proof,

$$\hat{\sigma}_{k1}^2 \to \lambda\sigma_1^2 + (1-\lambda)\sigma_2^2 \mathrel{\hat{=}} \sigma_0^2,$$

uniformly for all k such that $|k - \tau| \le n(\log n)^{-1}$, is similar to the proof in the first part. We only need note that in the current case:

$$\frac{1}{n}I_1(k) = \lambda\sigma_1^2 + o_p(1) \quad \text{and} \quad \frac{1}{n}I_2(k) = (1-\lambda)\sigma_2^2 + o_p(1)$$

uniformly for k such that $|k - \tau| \le n(\log n)^{-1}$.

The proof of Proposition 3.2 is almost the same, hence we will not repeat the proof here.

A.3 The Null Limiting Distributions of Test Statistics

Now we are ready to prove Theorems 3.1 and 3.2.

The Proof of Theorem 3.1. The proof is divided into several small steps. We proceed as follows.

Part 1. To show that $U_n^{(1)} \to \chi_1^2$ in distribution under the null hypothesis H_0.

Step 1. First we show that

$$\max_{\mathbf{1 \le k \le n-1}} V_n^{(1)}(k) = O_p(\log\log n)$$

where $V_n^{(1)}(k)$ is defined by (3).

By the definition of $Q_n(t)$ in (13) and Proposition 3.1, for some constant C:

$$\begin{aligned}
\max_{\mathbf{1 \le k \le n-1}} V_n^{(1)}(k) &= \max_{\mathbf{1 \le k \le n-1}} \left\{ \frac{n}{4\sigma^2} \frac{k}{n} \left(1 - \frac{k}{n}\right) [\hat{\theta}_1(k) - \hat{\theta}_2(k)]^2 \right\} [1 + o_p(1)] \\
&\le C \sup_{\frac{1}{n} \le t \le 1 - \frac{1}{n}} \frac{[Q_n(t)]^2}{t(1-t)} \\
&\le C \sup_{\frac{1}{n} \le t \le 1 - \frac{1}{n}} \left[\frac{Q_n(t) - B_n(t)}{\sqrt{t(1-t)}} \right]^2 + C \sup_{\frac{1}{n} \le t \le 1 - \frac{1}{n}} \left[\frac{B_n(t)}{\sqrt{t(1-t)}} \right]^2 \\
&= O_p(\log\log n),
\end{aligned}$$

where we have utilized the results (15) and (16) in Lemmas A.2 and A.3.

Step 2. To show that $\hat{\tau}/n \to \frac{1}{2}$ in probability, where $\hat{\tau}$ is defined by (6). For any $\epsilon > 0$, define

$$\Delta = \{k : |2k - n| < n\epsilon\}. \tag{33}$$

It is seen that:

$$\begin{aligned}
P\{\hat{\tau} \in \Delta\} &\ge P\left\{ U_n^{(1)}(n/2) \ge \max_{k \notin \Delta} U_n^{(1)}(k) \right\} \\
&\ge P\left\{ 4\epsilon^2 \log n \ge \max_{k \notin \Delta} V_n^{(1)}(k) - V_n^{(1)}(n/2) \right\} \to 1
\end{aligned}$$

since $\max_{k \notin \Delta} V_n^{(1)}(k) - V_n^{(1)}(n/2) = O_p(\log\log n)$.

Step 3. To derive an upper bound on the size of $U_n^{(1)}$. Since $\hat{\tau}/n \to \frac{1}{2}$, we have by noting the relationship of $Q_n(t)$ and $V_n^{(1)}(k)$:

$$\begin{aligned}
U_n^{(1)} &\le \max_{k \in \Delta} V_n^{(1)}(k) + o_p(1) \\
&\le \sup_{|t - \frac{1}{2}| \le \epsilon} \frac{[Q_n(t)]^2}{t(1-t)} + o_p(1)
\end{aligned}$$

$$\leq \frac{4}{1-4\epsilon^2} \sup_{|t-\frac{1}{2}|\leq\epsilon} [Q_n(t)]^2 + o_p(1)$$
$$= \frac{4}{1-4\epsilon^2} \sup_{|t-1/2|\leq\epsilon} [B_n(t)]^2 + o_p(1),$$

where the last equality comes from (14) in Lemma A.2.

Since the sample path function of Brownian bridges $\{B_n(t), 0 \leq t \leq 1\}$ is continuous in t with probability 1 and $2B_n(1/2) \sim N(0, 1)$, we have shown that $U_n^{(1)}$ is bounded by a random quantity whose limiting distribution is chi-square with 1 degree of freedom because ϵ can be taken arbitrarily small.

It is obvious that this upper bound is also a lower bound, since

$$\begin{aligned} U_n^{(1)} &\geq V_n^{(1)}(n/2) = 4[Q_n(1/2)]^2 + o_p(1) \\ &= 4[B_n(1/2)]^2 + o_p(1) \to \chi_1^2. \end{aligned}$$

Hence, the result under the null model is proved.

Part 2: To show that $U_n^{(1)} \to \infty$ in probability under H_1.

When the alternative model is true such that $\theta_1 \neq \theta_2$ and $\tau = [n\lambda]$ with $\lambda \in (0, 1)$, we have:

$$\begin{aligned} U_n^{(1)} &\geq V_n^{(1)}(\tau) - (2\lambda - 1)^2 \log n \\ &= \frac{n\lambda(1-\lambda)}{4\sigma_0^2}[\hat{\theta}_1(\tau) - \hat{\theta}_2(\tau)]^2[1 + o_p(1)] - (2\lambda - 1)^2 \log n. \end{aligned}$$

By Lemma A.1, we have:

$$\hat{\theta}_1(\tau) - \theta_1 = \frac{2}{\tau}\sum_{i=1}^{\tau}[\tilde{h}_1(X_i) - \theta_1] + O_p(n^{-1}) = o_p(1)$$

and

$$\hat{\theta}_2(\tau) - \theta_2 = \frac{2}{n-\tau}\sum_{i=\tau+1}^{n}[\tilde{h}_2(X_i) - \theta_2] + O_p(n^{-1}) = o_p(1).$$

Hence,

$$\hat{\theta}_1(\tau) - \hat{\theta}_2(\tau) = \theta_1 - \theta_2 + o_p(1).$$

Consequently,

$$U_n^{(1)} \geq \frac{n\lambda(1-\lambda)}{4\sigma_0^2}(\theta_1 - \theta_2)^2 + o_p(n) \to \infty.$$

Thus we complete the proof.

The Proof of Theorem 3.2. (1) To show that $\frac{\hat{\tau}}{n} \to \frac{1}{2}$ in probability, where $\hat{\tau}$ is defined by (10).

For any $\epsilon > 0$:

$$P\{\hat{\tau} \in \Delta\} \geq P\left\{U_n^{(2)}\left(\frac{n}{2}\right) \geq \max_{k \notin \Delta} U_n^{(2)}(k)\right\}$$
$$\geq P\left\{\max_{k \notin \Delta} \frac{Z_k^2}{nk(n-k)} - \frac{4Z_{\frac{n}{2}}^2}{n^3} \leq \sigma^2 \epsilon^2 \log n\right\},$$

where Δ is defined in (33). Due to

$$\max_{k \notin \Delta} \frac{Z_k^2}{nk(n-k)} - \frac{4Z_{\frac{n}{2}}^2}{n^3} = O_p(\log \log n)$$

from (18), we have $P\{\hat{\tau} \in \Delta\} \to 1$ as $n \to \infty$.

(2) By Lemma A.6 and Proposition 3.2, we have, for any $\epsilon > 0$:

$$\begin{aligned} U_n^{(2)} &\leq \max_{k \in \Delta} \left\{\frac{Z_k^2}{\sigma^2 nk(n-k)}\right\}[1 + o_p(1)] \\ &\leq \max_{k \in \Delta} \left\{\frac{Z_k^2}{\sigma^2 n^3}\left[\frac{k}{n}\left(1 - \frac{k}{n}\right)\right]^{-1}\right\}[1 + o_p(1)] \\ &\leq 4 \sup_{|t - \frac{1}{2}| < \epsilon} \frac{Z_{[(n+1)t]}^2}{\sigma^2 n^3}[1 + o_p(1)] \\ &= 4 \sup_{|t - \frac{1}{2}| < \epsilon} B_n^2(t)[1 + o_p(1)], \end{aligned}$$

where $\{B_n(t), 0 \leq t \leq 1\}$ is a sequence of Brownian Bridges, which have continuous sample path functions in t with probability one. Note that:

$$2B_n\left(\frac{1}{2}\right) \sim N(0, 1).$$

Hence, as $n \to \infty$:

$$U_n^{(2)} \leq 4B_n^2\left(\frac{1}{2}\right) + o_p(1) \to \chi_1^2,$$

since ϵ can be made arbitrarily small. On the other side, we have:

$$U_n^{(2)} \geq V_n^{(2)}\left(\frac{n}{2}\right) = \frac{4Z_{\frac{n}{2}}^2}{\sigma^2 n^3} + o_p(1) = 4B^2\left(\frac{1}{2}\right) + o_p(1) \to \chi_1^2.$$

Hence, $U_n^{(2)} \to^{\mathscr{D}} \chi_1^2$ as $n \to \infty$.

(3) When $\frac{\tau}{n} \to \lambda \in (0, 1)$, we have by Lemma A.7:

$$\begin{aligned} U_n^{(2)} &\geq \frac{Z_\tau^2}{n\tau(n-\tau)\sigma_0^2}[1 + o_p(1)] - (2\lambda - 1)^2 \log n \\ &= \frac{\lambda(1-\lambda)\mu^2}{\sigma_0^2} n + o_p(n) \to \infty \end{aligned}$$

in probability. Thus, we complete the proof of the theorem.

Acknowledgment

The authors thank the referees for insightful suggestions on the simulation which led to a substantially improved article.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: Petrov, B. N., Csaki, E., eds. *2nd Int. Symp. Inf. Theory*. Budapest: Akademiai Kiado, pp. 267–281.

Bengtsson, T., Cavanaugh, J. E. (2006). An improved Akaike information criterion for state-space model selection. *Computat. Statist. Data Anal.* 50:2635–2654.

Bogdan, M., Ghosh, J. K., Doerge, R. W. (2004). Modifying the Schwarz Bayesian information criterion to locate multiple interacting quantitative trait loci. *Genetics* 167:989–999.

Chen, J., Gupta, A. K., Pan, J. (2006). Information criterion and change point problem for regular models. *Sankhyā* 68:252–282.

Csörgö, M., Horváth, L. (1988). Invariance principles for change-point problems. *J. Multivariate Anal.* 27:151–168.

Csörgö, M., Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. New York: John Wiley & Sons.

Csörgö, M., Révész, P. (1981). *Strong Approximations in Probability and Statistics*. New York: Academic Press.

Darling, D. A., Erdös, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.* 23:143–155.

Gombay, E., Horváth, L. (1995). An application of U-statistics to change-point analysis. *Acta. Sci. Math.* 60:345–357.

Gombay, E. (2000). U-statistics for sequential change detection. *Metrika* 54:133–145.

Gombay, E. (2001). U-statistics for change under alternatives. *J. Multivariate Anal.* 78: 139–158.

Hall, P. (1979). On the invariance principle for U-statistics. *Stoch. Proc. Appl.* 9:163–174.

Hinkley, D. V. (1971). Inference about the change-point from cumulative sum tests. *Biometrika* 58:509–523.

Inclán, C., Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Amer. Statist. Assoc.* 89:913–923.

Kim, S., Cho, S., Lee, S. (2000). On the cusum test for parameter changes in GARCH(1,1) models. *Commun. Statist. Theor. Meth.* 29:445–462.

Konishi, S., Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika* 83:875–890.

Lee, S., Park, S. (2001). The cusum of squares test for scale changes in infinite order moving average processes. *Scand. J. Statist.* 28:625–644.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika* 41:100–116.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* 42:523–527.

Picard, D. (1985). Testing and estimating change-points in time series. *Adv. Appl. Probab.* 17:841–867.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* 6:461–464.

Sen, P. K., Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. New York: Chapman and Hall.

Volinsky, C. T., Raftery, A. E. (2000). Bayesian information criterion for censored survival models. *Biometrics* 56:256–262.

Zacks, S. (1983). Survey of classical and Bayesian approaches to the change-point problem. In: Rivzi, M. H., ed. *Recent Advances in Statistics*. New York: Academic Press, pp. 245–269.