The Canadian Journal of Statistics
Vol. 48, No. 3, 2020, Pages 417–446
La revue canadienne de statistique

Robust multivariate change point analysis based on data depth

Shojaeddin CHENOURI[1*], Ahmad MOZAFFARI[1], and Gregory RICE[1]

[1]*Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada*

Key words and phrases: Change point detection; CUSUM; data depth.

MSC 2010: Primary 62G05; secondary 62H15.

Abstract: Modern methods for detecting changes in the scale or covariance of multivariate distributions rely primarily on testing for the constancy of the covariance matrix. These depend on higher-order moment conditions, and also do not work well when the dimension of the data is large or even moderate relative to the sample size. In this paper, we propose a nonparametric change point test for multivariate data using rankings obtained from data depth measures. As the data depth of an observation measures its centrality relative to the sample, changes in data depth may signify a change of scale of the underlying distribution, and the proposed test is particularly responsive to detecting such changes. We provide a full asymptotic theory for the proposed test statistic under the null hypothesis that the observations are stable, and natural conditions under which the test is consistent. The finite sample properties are investigated by means of a Monte Carlo simulation, and these along with the theoretical results confirm that the test is robust to heavy tails, skewness and high dimensionality. The proposed methods are demonstrated with an application to structural break detection in the rate of change of pollutants linked to acid rain measured in Turkey lake, a lake in central Ontario, Canada. Our test suggests a change in the rate of acid rain in the late 1980s/early 1990s, which coincides with clean air legislation in Canada and the US. *The Canadian Journal of Statistics* 48: 417–446; 2020 © 2020 Statistical Society of Canada

Résumé: Les méthodes modernes de détection de changements dans l'échelle ou la covariance des distributions multivariées se fondent d'abord sur des tests pour la constance de la matrice de covariance. Elles dépendent ainsi de conditions sur les moments d'ordre supérieur et fonctionnent mal lorsque la dimension des données est grande, voire modérée, par rapport à la taille d'échantillon. Les auteurs proposent un test de rupture pour des données multivariées basé sur des rangs obtenus à l'aide d'une mesure de profondeur. Puisque la profondeur d'une observation mesure sa centralité par rapport à l'échantillon, un changement à sa profondeur peut indiquer un changement d'échelle de la distribution sous-jacente, et le test proposé est particulièrement sensible à ce type de changements. Les auteurs développent une théorie asymptotique complète pour la statistique de test proposée sous l'hypothèse nulle que les observations sont stables, en autant que quelques conditions naturelles garantissant la convergence du test soient respectées. Ils explorent les propriétés du test sur des échantillons finis à l'aide d'études de Monte Carlo qui confirment, avec les résultats théoriques, que le test est robuste aux queues lourdes, à l'asymétrie et à la grande dimensionnalité. Ils illustrent les méthodes proposées avec une application à la détection de points de rupture structurels dans le taux de changement des polluants liés aux pluies acides mesurées au lac Turkey situé en Ontario au Canada. Le test suggère un changement dans le taux lié aux pluies acides à la fin des années 1980 ou au début des années 1990, ce qui coincide avec l'entrée en vigueur de lois sur la qualité de l'air au Canada et aux États-Unis. *La revue canadienne de statistique* 48: 417–446; 2020 © 2020 Société statistique du Canada

* *Author to whom correspondence may be addressed.*
E-mail: schenouri@uwaterloo.ca

1. INTRODUCTION

Frequently when considering a sample of ordered observations, it is of interest to determine whether the sample contains one or more change points at which the underlying stochastic structure of the observations changes. Methods used for detecting and estimating such changes are the focus of the field of change point analysis, which enjoys an enormous literature, and we refer the reader to Brodsky & Darkhovsky (1993), Csörgő & Horváth (1997) and Horváth & Rice (2014) for an overview of the subject. When it comes to scalar observations, change point theory is quite comprehensive. In particular, it includes considerations for robust change point analysis, as well as tests and estimators for changes in the scale or variance of observations; see Dehling & Fried (2012), Gombay (1994), Hušková (2013) and Bandyopadhyay & Mukherjee (2007) for robust change point methods, and Inclán & Tiao (1994), Wied, Krämer & Dehling (2012) and Dette, Wu & Zhou (2015) for changes in the variance and autocorrelation structure of time series.

In extending such methodology to multivariate data, a number of difficulties are encountered. The most natural extension of the notion of variance for data in $\mathbb{R}^p$, $p \geq 2$, is the covariance matrix, which captures the second-order properties of the distribution. Methodology for testing the constancy of the covariance matrix is developed in Galeano & Peña (2007), Aue et al. (2009) and Kao, Trapani & Urga (2018). These require fourth-order moment conditions, which is no different than similar procedures for scalar data, but they also rely on the dimension being rather small, as the calculation of critical values and/or normalizing sequences of the test statistics proposed involve estimating the fourth-order structure of the data. Moreover, as noted in Zhou (2013) and Dette, Wu & Zhou (2015), these methods are also sensitive to misidentification of the mean, in the sense that if the mean were to change during the observation period whilst the covariance structure remains fixed, then the tests tend to spuriously reject the hypothesis of constant covariance. Robust methods in multivariate change point analysis are quite limited in the literature, perhaps due to the difficulty in defining appropriate robust measures in higher dimensions. Lung-Yut-Fong, Lévy-Leduc & Cappé (2015) develop a robust change point test to detect changes in the location in multivariate distributions based on a coordinate-wise generalization of the Wilcoxon/Mann–Whitney statistic. Bickel (1964, 1965) discuss the potential drawbacks of coordinate-wise generalizations of univariate procedures to the multivariate setting.

In this paper, we propose a nonparametric change point procedure for multivariate data based on ranks obtained from data depths. A multivariate data depth is a measure of how central, or deep, a given point in $\mathbb{R}^p$ is with respect to a multivariate distribution or data cloud. We refer the reader to Small (1990), Liu (1990), Donoho & Gasko (1992), Liu, Parelius & Singh (1999), Zuo & Serfling (2000), and the references therein for more detailed descriptions and properties. As data depth provides a measure of centrality of individual observations within a multivariate dataset, testing for changes in the distribution of depths provides a means of determining if the scale of the underlying distribution of the observations changes. See Liu & Singh (1993), Liu & Singh (2006), Chenouri, Small & Farrar (2011) and Chenouri & Small (2012), for multivariate nonparametric tests. The proposed methodology, in this paper, has a number of advantages. Since the proposed test statistic is ultimately based on ranks, it is robust to heavy tailed and skewed distributions, and establishing its asymptotic properties does not rely on any moment conditions. Moreover, the asymptotic properties, and corresponding critical values used to determine the significance of the test statistic are derived using the exchangeability of data depth measures, and hence does not depend on the data dimensionality. This allows the test to have good size properties even when the dimensionality is large compared to the sample size; a fact that is demonstrated via a comprehensive simulation study. We derive and study conditions under which the test is asymptotically consistent which show that the proposed test is not necessarily sensitive to changes in the mean, but is sensitive to changes in the scale of the distribution.

The test is also remarkably simple to implement, as the test statistic is computed by maximizing a cumulative sum (CUSUM) type process based on data depth ranks.

The remainder of the paper is organized as follows. In Section 2, we formally define several data depth measures that are used in subsequent sections. Section 3 contains the formal statement of the change-point problem, and introduces the proposed test statistic along with a full asymptotic theory. We study the efficacy of these asymptotic results in finite samples, and compare our test to the covariance change point test of Aue et al. (2009) by means of a Monte Carlo simulation study in Section 4. In Section 5, we present an application of our test to acid rain data obtained from Turkey Lake, a lake in central Ontario, taken the early 1980s until the early 2000s. The formal proofs of the asymptotic results as well as some additional simulation results are provided in Appendices A and B following the references.

2. DATA DEPTH AND MULTIVARIATE DEPTH RANKING

Let $\mathbf{X}_1, \dots, \mathbf{X}_N$ be a random sample of size N from a p-dimensional distribution F and suppose that $\widehat{F}_N$ represents the corresponding empirical distribution, taken as a nonparametric estimate of F. In this section, $D(\mathbf{x}; F)$ shall denote, for any given depth function D, the depth of the point $\mathbf{x} \in \mathbb{R}^p$ with respect to the distribution F. In particular, $D(\mathbf{x}; \widehat{F}_N)$ shall denote the depth function with respect to the empirical distribution $\widehat{F}_N$ or dataset $\mathbb{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_N\}$. Quite a large number of depth functions have been introduced in the literature, each having the basic intention of providing a measure of relative centrality of the data points. We now formally define a few depth functions that we will use for the purpose of change-point detection.

The *halfspace depth* or *Tukey depth* at a point $\mathbf{x} \in \mathbb{R}^p$ is the minimal proportion of data contained in a closed half-space whose boundary, a $(p-1)$-dimensional hyperplane, passing through $\mathbf{x}$. More formally

$$\mathrm{HSD}\left(\mathbf{x}\,;\,\widehat{F}_N\right) = \frac{\min_{\|\mathbf{u}\|=1}\ \#\{i\ :\ \mathbf{u}^\top \mathbf{X}_i \le \mathbf{u}^\top \mathbf{x},\, i = 1, 2, \dots, N\}}{N}.$$

The notion of halfspace depth in $\mathbb{R}$ dates back to Hotelling (1929) and Chamberlin (1937), while its extension to $\mathbb{R}^2$ was given by Hodges (1955) in his bivariate sign test, and Tukey (1975) formally defined the halfspace depth as a tool for visualizing bivariate data. Donoho (1982) and Donoho & Gasko (1992) extended the notion to $\mathbb{R}^p$, and studied several properties of the theoretical and empirical halfspace depth.

Another multivariate generalization of the notion discussed in Hotelling (1929) and Chamberlin (1937) is the simplicial depth (Liu, 1990). The simplicial depth of a given point $\mathbf{x} \in \mathbb{R}^p$ with respect to the empirical distribution $\widehat{F}_N$ is defined to be

$$\mathrm{SD}(\mathbf{x}; \widehat{F}_N) = \binom{N}{p+1}^{-1} \sum \mathbb{1}(S[X_{i_1}, X_{i_2}, \dots, X_{i_{p+1}}] \ni \mathbf{x}),$$

where $S[X_{i_1}, X_{i_2}, \dots, X_{i_{p+1}}]$ is the closed simplex with vertices $X_{i_1}, X_{i_2}, \dots, X_{i_{p+1}}$, and where $\sum$ runs over all possible subsets of $\mathbb{X}$ of size $p+1$, and $\mathbb{1}(\cdot)$ is the indicator function. Oja depth (Oja, 1983; Zuo & Serfling, 2000) is defined similarly but instead considers the volume of simplices based on subsets of data.

The *spatial depth* of a point $\mathbf{x} \in \mathbb{R}^p$ with respect to $\widehat{F}_N$ is

$$\mathrm{SPD}(\mathbf{x}\,;\,\widehat{F}_N) = \frac{1}{1 + N^{-1} \sum_{i=1}^N \|\mathbf{X}_i - \mathbf{x}\|},$$

where $\|\cdot\|$ is the usual Euclidean norm; see Gower (1974), Brown (1983) and Zuo & Serfling (2000). An affine equivariant version of this depth function may be obtained by replacing the Euclidean norm with the generalized Euclidean norm $\|\cdot\|_{\mathbf{M}}$ as

$$\|\mathbf{x}\|_{\mathbf{M}} = \sqrt{\mathbf{x}^\top \mathbf{M}\,\mathbf{x}}, \qquad \text{for all } \mathbf{x} \in \mathbb{R}^p,$$

where $\mathbf{M}$ is a $p \times p$ positive definite matrix, such as a covariance matrix; see Rao (1988). This generalized distance is known as the Mahalanobis distance (Mahalanobis, 1936) when the matrix $\mathbf{M}$ is the inverse of either the theoretical or empirical covariance matrices. The Mahalanobis distance gives rise to the *Mahalanobis depth* of $\mathbf{x}$, which is defined by

$$\text{MHD}(\mathbf{x}\,;\widehat{F}_N) = \frac{1}{1 + \|\mathbf{x} - \widehat{\boldsymbol{\mu}}\|^2_{\widehat{\Sigma}^{-1}}},$$

where $\widehat{\boldsymbol{\mu}}$ is the centroid of the data and $\widehat{\boldsymbol{\Sigma}}$ is the empirical covariance matrix. See Zuo & Serfling (2000). Modified robust versions of the affine equivariant SPD and MHD can be obtained by replacing $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ with appropriate robust estimates such as the reweighted MCD estimator in Rousseeuw & van Zomeren (1990). In this paper we will use Mahalanobis depth with the reweighted MCD estimates with the asymptotic breakdown values of 25% and 50%, denoted by MHD75 and MHD50, indicating 75% and 50% of the observations are subsampled in order to calculate the estimators, respectively.

The population versions of these depth functions are defined similarly. Given a depth function D, we can compute the depth values of the sample points $\mathbf{X}_1, \ldots, \mathbf{X}_N$ with respect to a given (either theoretical or empirical) distribution F, that is

$$D(\mathbf{X}_1\,; F), \ldots, D(\mathbf{X}_N\,; F),$$

which may subsequently be ordered into an increasing list. This defines what we shall call the *depth ranking* of the data. In particular, for a sample point $\mathbf{X}_i$

$$\widehat{R}_i = \#\{\mathbf{X}_j\ ;\ D(\mathbf{X}_j\,;\widehat{F}_N) \le D(\mathbf{X}_i\,;\widehat{F}_N),\ j = 1, \ldots N\}, \tag{2.1}$$

is the empirical depth rank of $\mathbf{X}_i$, and where $\#A$ denotes the cardinality of any set A.

3. MULTIVARIATE CHANGE POINT PROBLEM AND DEPTH BASED WILCOXON TYPE CUSUM PROCESSES

Consider multivariate observations $\mathbf{X}_1, \ldots, \mathbf{X}_N$ in $\mathbb{R}^p$, which we assume are independent. In the cases when these data are naturally ordered, for instance if they are observations, or model residuals, from a multivariate time series, it is often of interest to determine whether the distribution of the observations remains stable throughout the sample, or if instead the distribution seems to change at one or more change points. In light of the assumed independence, we may cast this problem within the framework of hypothesis testing by considering a test of

$$H_0\ :\quad \mathbf{X}_i,\ \ 1 \le i \le N \text{ are identically distributed,}$$

namely that the distribution of the observations is homogenous throughout the sample, versus

$$H_{A,\theta}\ :\ \mathbf{X}_1, \ldots, \mathbf{X}_{\lfloor N\theta \rfloor} \sim F_1,\ \text{and } \mathbf{X}_{\lfloor N\theta \rfloor + 1}, \ldots, \mathbf{X}_N \sim F_2,\ \text{where } \theta \in (0, 1), \tag{3.1}$$

where $F_1 \neq F_2$. We call this alternative the "at-most-one change" alternative, and we discuss later how one might adapt the proposed method to more general alternatives allowing for multiple changes. The parameter θ denotes the unknown break fraction that defines when the sample changes distribution from F_1 to F_2, and we call $k^* = \lfloor N\theta \rfloor$ the change point.

We propose to test H_0 versus $H_{A,\theta}$ based on the depth ranks of the observations. The basic idea is that under $H_{A,\theta}$ and if the nature of the difference between F_1 and F_2 is a change in scale, then one would expect that the depths of the observations $\mathbf{X}_1, \dots, \mathbf{X}_{\lfloor N\theta \rfloor}$ would be on average larger/smaller than the depths of $\mathbf{X}_{\lfloor N\theta \rfloor+1}, \dots, \mathbf{X}_N$, and hence the depth ranks $\widehat{R}_1, \dots, \widehat{R}_{\lfloor N\theta \rfloor}$ defined by (2.1) would tend to be larger/smaller than $\widehat{R}_{\lfloor N\theta \rfloor+1}, \dots, \widehat{R}_N$. On the other hand, under H_0 the ranks $\widehat{R}_i$ are uniformly distributed on the integers between 1 and N, regardless of the underlying common distribution of $\mathbf{X}_1, \dots, \mathbf{X}_N$, and hence for each $i = 1, \dots, N$

$$\mathrm{E}[\,\widehat{R}_i\,] = \frac{N+1}{2}, \quad \text{and} \quad \mathrm{Var}[\,\widehat{R}_i\,] = \frac{N^2-1}{12}.$$

This suggests basing a test of H_0 on functionals of the rank-CUSUM process

$$Z_N(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{\lfloor Nt \rfloor} \frac{\widehat{R}_i - (N+1)/2}{\sqrt{(N^2-1)/12}} \qquad t \in (0,1).$$

A similar process to $Z_N(t)$ is studied in the context of univariate nonparametric location change point analysis based on linear ranks in Bhattacharya & Frierson (1981) and Carlstein, Müller & (1994). Under $H_{A,\theta}$, and if θ were known, then $Z_N(\theta)$ coincides with the two sample Wilcoxon test statistic, and hence a natural functional of $Z_N(t)$ to consider is

$$T_N = \sup_{0 \le t \le 1} |\, Z_N(t)\, |,$$

which is the maximally selected two sample Wilcoxon test statistic over all possible break fractions applied to the data-depths. We are able to obtain the following large sample results for Z_N and T_N under H_0.

Theorem 1. *Under H_0, the process $\{Z_N(t),\ 0 \le t \le 1\}$ converges weakly in $\mathcal{D}[0,1]$ endowed with the Skorokhod topology (see Chapter 3 of Billingsley, 1968) to a standard Brownian bridge*

$$\{B(t),\ 0 \le t \le 1\}.$$

The proof of Theorem 1 is provided in Appendix A. As a consequence of the continuous mapping theorem, we obtain the following corollary:

Corollary 1. *Under H_0, $T_N \xrightarrow{\mathcal{D}} \sup_{0 \le t \le 1} |B(t)|$.*

Corollary 1 shows that a test of H_0 with asymptotic size α is to reject if $T_N > \Xi_{1-\alpha}$, where Ξ_q is the $100q\%$ quantile of $\sup_{0 \le t \le 1} |\, B(t)\, |$. This is the test we will study further.

3.1. Consistency of T_N under $H_{A,\theta}$, and Estimation of Change Point

We now turn to the asymptotic consistency of the test under the at-most-one change alternative. Under $H_{A,\theta}$ and the assumed independence of the observations, it is evident that the empirical distribution function $\widehat{F}_N$ based on the whole sample is converging in sup-norm by the

Glivenko-Cantelli theorem to

$$F_* = \theta F_1 + (1-\theta)F_2.$$

Since the test statistic T_N is based on the depth-ranks, the condition for consistency naturally depends on the distribution of the depths themselves. We define $H_1(u) = \Pr(D(\mathbf{Y}, F_*) \le u)$ and $H_2(u) = \Pr(D(\mathbf{Z}, F_*) \le u)$, where $\mathbf{Y} \sim F_1$, and $\mathbf{Z} \sim F_2$. We assume the following continuity conditions on the distributions H_1 and H_2, and the depth function D.

Assumption 1. *$H_1(u)$ and $H_2(u)$ are each Lipschitz with constant C, meaning that for all $u, v \in \mathbb{R}$ we have*

$$|\, H_i(u) - H_i(v)\, | \le C\, |u - v| \quad i = 1,\, 2\,.$$

Assumption 2. $\mathrm{E}\left[\sup_{x\in\mathbb{R}^p} |\, D(\mathbf{x}\,;\, \widehat{F}_N) - D(\mathbf{x}\,;\, F_*)\,|\right] = O(N^{-1/2}).$

Assumptions 1 and 2 coincide with Assumptions A1 and A3 in Zuo & He (2006). Assumption 2 is satisfied for many depth functions, see Section 4 of Zuo & He (2006), as long as

$$\sup_{x\in\mathbb{R}^p} |\widehat{F}_N(\mathbf{x}) - F_*(\mathbf{x})| = O_P(N^{-1/2}). \tag{3.2}$$

This follows from the multivariate Dvoretzky-Kiefer-Wolfowitz inequality when the data satisfy $H_{A,\theta}$ and are independent. See Alexander (1984) and Massart (1986).

Theorem 2. *Suppose $H_{A,\theta}$, and Assumptions 1 and 2 hold. If*

$$\int_{-\infty}^{\infty} \left\{\theta H_1(u) + (1-\theta)H_2(u)\right\} dH_1(u) \ne \frac{1}{2}, \tag{3.3}$$

then $T_N \xrightarrow{P} \infty$.

The consistency condition (3.3) comes as a result of the Chernoff-Savage theorem (Chernoff & Savage, 1958), which implies that a necessary and sufficient condition for $Z_N(\theta)$ to have an asymptotically nonzero mean is (3.3) under $H_{A,\theta}$. Terms akin to the left-hand side of (3.3) are often used in power and sample size calculations for tests based on linear rank statistics, see Shieh, Jan & Randles (2006) for an example.

The proof of Theorem 2 is provided in Appendix A. Next, we provide an analysis of the condition (3.3) in the simple case of normal observations and Mahalonobis depth.

Example 1. *Suppose that*

$$\boldsymbol{Y}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_{\lfloor N\theta \rfloor} \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \text{ and } \mathbf{Z}, \boldsymbol{X}_{\lfloor N\theta \rfloor+1}, \ldots, \boldsymbol{X}_N \sim N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), \tag{3.4}$$

where $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a p-variate normal random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. A simple calculation shows that

$$MHD(\mathbf{x}\,;\, F_*) = \frac{1}{1 + \|\,\mathbf{x} - \boldsymbol{\mu}_*\|_{\Sigma_*^{-1}}},$$

where $\boldsymbol{\mu}_* = \theta\, \boldsymbol{\mu}_1 + (1-\theta)\, \boldsymbol{\mu}_2$, and

$$\boldsymbol{\Sigma}_* = \theta\, \boldsymbol{\Sigma}_1 + (1-\theta)\, \boldsymbol{\Sigma}_2 + \theta\, (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_*)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_*)^\top + (1-\theta)\, (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_*)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_*)^\top.$$

First we consider a simple location shift alternative, i.e. when

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}, \quad \textit{and} \quad \boldsymbol{\mu}_1 = \mathbf{0}, \;\; \boldsymbol{\mu}_2 = \boldsymbol{\mu}.$$

In this case, $\boldsymbol{\mu}_* = (1-\theta)\,\boldsymbol{\mu}$, *and* $\boldsymbol{\Sigma}_* = \boldsymbol{\Sigma} + \theta(1-\theta)\,\boldsymbol{\mu}\,\boldsymbol{\mu}^\top$. *It follows that* $\| \boldsymbol{Y} - \boldsymbol{\mu}_* \|_{\Sigma_*^{-1}}$ *and* $\| \boldsymbol{Z} - \boldsymbol{\mu}_* \|_{\Sigma_*^{-1}}$ *are each weighted noncentral* χ^2 *random variables with equal weights, and noncentrality parameters* $(1-\theta)^2\boldsymbol{\mu}^\top\boldsymbol{\mu}$ *and* $\theta^2\boldsymbol{\mu}^\top\boldsymbol{\mu}$, *respectively. In particular, if* $\theta = 1-\theta = 1/2$, *then* $H_1 = H_2$, *and (3.3) is not satisfied.*

With regard to a change in the scale parameter, consider $\boldsymbol{Y} \sim N_p(\mathbf{0}, \sigma_1^2\boldsymbol{I})$, *and* $\boldsymbol{Z} \sim N_p(\mathbf{0}, \sigma_2^2\boldsymbol{I})$, *where* $\mathbf{0}$ *is the zero vector and* $\boldsymbol{I}$ *is the identity matrix. Then* $\boldsymbol{\mu}_* = \mathbf{0}$, *and* $\boldsymbol{\Sigma}_* = \sigma_*^2\boldsymbol{I}$, *where* $\sigma_*^2 = \theta\sigma_1^2 + (1-\theta)\sigma_2^2$. *It follows that*

$$\| \boldsymbol{Y} - \boldsymbol{\mu}_* \|_{\Sigma_*^{-1}} \overset{D}{=} \frac{\sigma_1^2}{\sigma_*^2}\,\chi_p^2, \quad \textit{and} \quad \| \boldsymbol{Z} - \boldsymbol{\mu}_* \|_{\Sigma_*^{-1}} \overset{D}{=} \frac{\sigma_2^2}{\sigma_*^2}\,\chi_p^2.$$

Elementary calculations then show that in this case, (3.3) holds if and only if

$$\int_0^\infty F_{\chi^2(p)}\left(\frac{\sigma_1^2}{\sigma_2^2}\,u\right) f_{\chi^2(p)}(u)\,du \neq \frac{1}{2},$$

where $F_{\chi^2(p)}$ *and* $f_{\chi^2(p)}$ *denote the cumulative distribution function and density function of a* χ_p^2 *random variable, respectively. This holds if and only if* $\sigma_1^2 \neq \sigma_2^2$, *and hence the test is consistent under this assumption.*

Under $H_{A,\theta}$, it is often of interest to estimate the change point k^*, or equivalently the break fraction θ. This may be because the change point might indicate an important event related to the observed data. In addition, one may use an estimator $\widehat{k^*}$ of k^* to test for the presence of additional change points through the binary segmentation procedure, in which after an initial change point $\widehat{k^*}$ is estimated, one may segment the data into two subsamples $\mathbf{X}_1, \ldots, \mathbf{X}_{\widehat{k^*}}$ and $\mathbf{X}_{\widehat{k^*}+1}, \ldots, \mathbf{X}_N$, and further apply change point testing and estimation to these samples. See Csörgő & Horváth (1997) and the introduction of Fryzlewicz (2014) for a review of the procedure. We utilize this procedure in our analysis of acid rain data in Section 5. Natural estimators of k^* and θ are

$$\widehat{k^*} = \min\left\{ k \;:\; | Z_N(\lfloor k/N \rfloor) | = \sup_{0\le t\le 1} | Z_N(t) | \right\}, \;\text{ and }\; \widehat{\theta} = \frac{\widehat{k^*}}{N},$$

the latter of which is remarkably a consistent estimator under Assumptions 1, 2 and Equation (3.3).

Theorem 3. *Under the conditions of Theorem 2,* $|\widehat{\theta} - \theta| = o_P(1)$.

The proof of Theorem 3 is provided in Appendix A.

4. SIMULATION STUDY

We now present the results of a Monte Carlo simulation study in which we investigated the finite sample properties of T_N. All simulations were carried out in version 3.2.3 of the `R` programming language (`R` Core Team, 2019). From each data generating process (DGP) considered, we generated 1,000 independent samples, and from each sample we computed the test statistic T_N using the depths MHD, MHD75, MHD50, HSD, SPD, and, in the case when the dimension

$p = 2$, the simplicial depth of Liu (Liu, 1990), SLD, and the Oja depth (Oja, 1983), OD. See also Zuo & Serfling (2000). We also compared these results to the test proposed by Aue et al. (2009) to test for changes in the covariance matrix, which is based on maximizing a CUSUM process based on vectorized versions of $\mathbf{X}_i\mathbf{X}_i^\top$. We call this method simply "MCUSUM". The percentage of trials in which T_N exceeded the 5% critical value of the distribution of $\sup_{0\le t\le 1} |B(t)|$ are reported in the following, as well as the percentage of times the MCUSUM statistic of Aue et al. (2009) exceeds the 5% critical value of its asymptotic distribution. We now turn to the specific definitions of the DGP's, starting with examples satisfying the null hypothesis of no change, and a study of the empirical size.

4.1. Empirical Size

We let $\mathbf{0}$ and $\mathbf{I}$ denote the zero vector and identity matrix, respectively. We further let $\mathbf{1}$ denote a vector of 1's. We consider the following six DGP's in this case:

- $N_p(\mathbf{0}, \mathbf{I})$: $\mathbf{X}_i$ has a p-variate normal distribution with mean vector $\mathbf{0}$, and covariance matrix $\mathbf{I}$.
- $C_p(\mathbf{0}, \mathbf{I})$: $\mathbf{X}_i$ has a p-variate Cauchy distribution with location vector $\mathbf{0}$, and dispersion matrix $\mathbf{I}$.
- $U_p(1)$: $\mathbf{X}_i$ is uniformly distributed on the interior of the unit sphere in p dimensions.
- $SN_p(\mathbf{0}, \mathbf{I}, 3 \times \mathbf{1})$: $\mathbf{X}_i$ has a p-variate skewed normal distribution with mean vector $\mathbf{0}$, covariance matrix $\mathbf{I}$, and skewness vector $3 \times \mathbf{1}$.
- $SN_p(\mathbf{0}, \mathbf{I}, 10 \times \mathbf{1})$: $\mathbf{X}_i$ has a p-variate skewed normal distribution with mean vector $\mathbf{0}$, covariance matrix $\mathbf{I}$, and skewness vector $10 \times \mathbf{1}$.
- $ST_p(\mathbf{0}, \mathbf{I}, 5 \times \mathbf{1}, 4)$: $\mathbf{X}_i$ has a p-variate skewed t-distribution with 4 degrees of freedom, mean vector $\mathbf{0}$, covariance matrix $\mathbf{I}$, and skewness vector $5 \times \mathbf{1}$.

We considered two basic simulations:

(1) With the dimension p fixed and small ($p = 2$) and increasing N from 25 to 200, and
(2) with increasing p from 5 to 25 for $N = 25$, $N = 100$, and $N = 200$.

Due to the computational complexity of HSD, SLD and OD in high dimensions, we omitted those depth functions from the simulations when the dimension exceeded two. For computing HSD, there is an exact algorithm for $p \le 3$ and an approximate algorithm for higher dimensions; see Rousseeuw and Ruts (1998) and Struyf & Rousseeuw (2000). Algorithms also exist to calculate Oja depth for dimension $p \le N$, however these are prohibitively slow. For SLD, computational algorithms have only been implemented for dimension $p = 2$. Nice discussions of these limitations are given in the documentations of the `R` packages `depth` of Genest, Masse & Plante (2017) and `mrfDepth` of Segaert et al. (2018). The results for $p = 2$ and increasing N are reported in Table 1, and for larger values of p the results are reported in Table 2. We summarize these results as follows:

- With the dimension $p = 2$, the depth-based tests exhibited a good size for all values of N and for all DGP's, including the heavy tailed and skewed distributions, with the exception of the HSD and SLD for small values of N ($N = 25$), which is due to producing many ties in ranks.
- In low-dimension $p = 2$, the MCUSUM test tended to be undersized, and this coincides with the simulations presented in Aue et al. (2009) for small to moderate values of N. The MCUSUM test relies on, among other conditions, the assumption that $E\|\mathbf{X}_i\|^4 < \infty$, and so is not expected to perform well in the case of $C_p(\mathbf{0}, \mathbf{I})$. In this particular case we observed that the test was quite undersized.
- Increasing the dimension has some effect on the depth based tests that rely on robustly estimating the covariance matrix (MHD50, MHD75), and in these cases the test was oversized for large p ($p = 20$).

TABLE 1: Empirical size based on 1,000 trials with the nominal level of 5% for 2-dimensional DGP's, and sample sizes 25, 100 and 200.

DGP	Methods	$N=25$	$N=100$	$N=200$	DGP	Methods	$N=25$	$N=100$	$N=200$
	MHD	0.045	0.058	0.039		MHD	0.044	0.058	0.043
	MHD75	0.048	0.047	0.039		MHD75	0.046	0.053	0.039
	MHD50	0.048	0.047	0.040		MHD50	0.046	0.052	0.039
$N_p(\mathbf{0},\mathbf{I})$	SPD	0.045	0.055	0.036	$C_p(\mathbf{0},\mathbf{I})$	SPD	0.046	0.053	0.040
	HSD	0.071	0.054	0.039		HSD	0.066	0.057	0.041
	SLD	0.067	0.050	0.033		SLD	0.062	0.056	0.045
	OD	0.043	0.058	0.035		OD	0.038	0.054	0.042
	MCUSUM	0.002	0.022	0.047		MCUSUM	0.001	0.005	0.002
	MHD	0.038	0.042	0.043		MHD	0.046	0.042	0.037
	MHD75	0.038	0.050	0.038		MHD75	0.046	0.052	0.044
	MHD50	0.038	0.050	0.041		MHD50	0.046	0.052	0.043
$U_p(1)$	SPD	0.041	0.042	0.042	$SN_p(\mathbf{0},\mathbf{I},3\mathbf{1})$	SPD	0.048	0.039	0.046
	HSD	0.070	0.043	0.043		HSD	0.072	0.036	0.035
	SLD	0.066	0.043	0.042		SLD	0.063	0.034	0.035
	OD	0.036	0.044	0.045		OD	0.046	0.042	0.037
	MCUSUM	0.005	0.032	0.038		MCUSUM	0.001	0.027	0.029
	MHD	0.037	0.042	0.043		MHD	0.044	0.050	0.038
	MHD75	0.037	0.046	0.043		MHD75	0.061	0.055	0.054
	MHD50	0.037	0.046	0.044		MHD50	0.061	0.055	0.054
$SN_p(\mathbf{0},\mathbf{I},10\mathbf{1})$	SPD	0.040	0.039	0.044	$ST_p(\mathbf{0},\mathbf{I},5\mathbf{1},4)$	SPD	0.053	0.049	0.048
	HSD	0.075	0.040	0.044		HSD	0.074	0.050	0.038
	SLD	0.068	0.041	0.046		SLD	0.070	0.049	0.041
	OD	0.042	0.041	0.046		OD	0.046	0.053	0.046
	MCUSUM	0.007	0.022	0.032		MCUSUM	0.003	0.009	0.010

- In contrast, the test based on spatial depth SPD had good size for all dimensions and DGPs.
- The MCUSUM test is sensitive to the dimension of the underlying data, and in contrast to the low dimensional case tends to be oversized. The reason for this seems to be that this test relies on estimating the covariance matrix of vech($\mathbf{X}_i\mathbf{X}_i^\top$), which is often singular or nearly singular in high dimensions.
- In summary, in low-dimensions the test of H_0 based on T_N exhibited good size for all DGPs and depths considered. In high dimensions ($p > 10$) we recommend using T_N with spatial depth in order to perform the test.

4.2. Empirical Power

For evaluating the empirical power of the proposed method, we generated data with the same underlying distributions as in the previous section, but with a change in scale introduced at $\theta = 0.5$ so that the data follow $H_{A,0.5}$. In particular, we generated data as in (3.1) with:

- $\mathbf{X}_1,\dots,\mathbf{X}_{\lfloor N\theta\rfloor} \sim N_p(\mathbf{0},\mathbf{I})$ and $\mathbf{X}_{\lfloor N\theta\rfloor+1},\dots,\mathbf{X}_N \sim N_p(\mathbf{0},\sigma^2\mathbf{I})$.
- $\mathbf{X}_1,\dots,\mathbf{X}_{\lfloor N\theta\rfloor} \sim C_p(\mathbf{0},\mathbf{I})$ and $\mathbf{X}_{\lfloor N\theta\rfloor+1},\dots,\mathbf{X}_N \sim C_p(\mathbf{0},\sigma^2\mathbf{I})$.
- $\mathbf{X}_1,\dots,\mathbf{X}_{\lfloor N\theta\rfloor} \sim U_p(1)$ and $\mathbf{X}_{\lfloor N\theta\rfloor+1},\dots,\mathbf{X}_N \sim U_p(\sigma^2)$.
- $\mathbf{X}_1,\dots,\mathbf{X}_{\lfloor N\theta\rfloor} \sim SN_p(\mathbf{0},\mathbf{I},3\times\mathbf{1})$ and $\mathbf{X}_{\lfloor N\theta\rfloor+1},\dots,\mathbf{X}_N \sim SN_p(\mathbf{0},\sigma^2\mathbf{I},3\times\mathbf{1})$.

TABLE 2: Empirical size based on 1,000 trials with the nominal level of 5% for 5, 10 and 20 dimensional DGP's, and sample sizes 25, 100 and 200.

		$N=25$			$N=100$			$N=200$		
DGP	Methods	$p=5$	$p=10$	$p=20$	$p=5$	$p=10$	$p=20$	$p=5$	$p=10$	$p=20$
	MHD	0.041	0.045	0.046	0.041	0.059	0.054	0.032	0.039	0.045
	MHD75	0.048	0.079	0.037	0.043	0.074	0.142	0.046	0.058	0.071
$N_p(\mathbf{0},\mathbf{I})$	MHD50	0.044	0.073	0.037	0.044	0.078	0.159	0.051	0.051	0.081
	SPD	0.043	0.054	0.051	0.038	0.050	0.048	0.034	0.038	0.044
	MCUSUM	0.000	*	*	0.013	0.000	*	0.026	0.005	*
	MHD	0.052	0.050	0.043	0.041	0.047	0.040	0.041	0.047	0.043
	MHD75	0.057	0.057	0.047	0.042	0.043	0.048	0.042	0.047	0.045
$C_p(\mathbf{0},\mathbf{I})$	MHD50	0.057	0.052	0.047	0.042	0.045	0.047	0.041	0.048	0.041
	SPD	0.052	0.054	0.058	0.042	0.037	0.045	0.045	0.048	0.043
	MCUSUM	0.000	*	*	0.001	0.000	*	0.000	0.000	*
	MHD	0.050	0.044	0.038	0.053	0.046	0.042	0.052	0.044	0.047
	MHD75	0.055	0.072	0.043	0.052	0.048	0.069	0.054	0.054	0.049
$U_p(1)$	MHD50	0.055	0.067	0.043	0.047	0.045	0.068	0.055	0.047	0.052
	SPD	0.045	0.040	0.047	0.053	0.050	0.040	0.056	0.047	0.049
	MCUSUM	0.000	*	*	0.008	0.000	*	0.018	0.000	*
	MHD	0.044	0.044	0.043	0.054	0.036	0.042	0.043	0.040	0.042
	MHD75	0.046	0.079	0.043	0.061	0.072	0.134	0.058	0.049	0.068
$SN_p(\mathbf{0},\mathbf{I},3\mathbf{1})$	MHD50	0.045	0.079	0.043	0.049	0.069	0.137	0.059	0.048	0.077
	SPD	0.047	0.051	0.038	0.047	0.037	0.048	0.039	0.041	0.041
	MCUSUM	0.000	*	*	0.007	0.000	*	0.018	0.006	*
	MHD	0.034	0.038	0.047	0.056	0.053	0.046	0.047	0.029	0.049
	MHD75	0.052	0.091	0.046	0.060	0.079	0.113	0.050	0.043	0.065
$SN_p(\mathbf{0},\mathbf{I},10\mathbf{1})$	MHD50	0.051	0.084	0.045	0.062	0.077	0.114	0.055	0.045	0.079
	SPD	0.034	0.050	0.042	0.047	0.055	0.043	0.046	0.033	0.043
	MCUSUM	0.000	*	*	0.010	0.000	*	0.016	0.003	*
	MHD	0.040	0.059	0.050	0.046	0.041	0.044	0.043	0.040	0.040
	MHD75	0.046	0.093	0.043	0.042	0.061	0.070	0.054	0.041	0.047
$ST_p(\mathbf{0},\mathbf{I},5\mathbf{1},4)$	MHD50	0.045	0.093	0.043	0.044	0.066	0.077	0.052	0.045	0.049
	SPD	0.048	0.054	0.052	0.047	0.050	0.044	0.047	0.039	0.042
	MCUSUM	0.000	*	*	0.003	0.000	*	0.002	0.000	*

* The test is infeasible due to the fact that the covariance matrix estimates of $\text{vech}(\mathbf{X}_i\mathbf{X}_i^\top)$ are not invertible.

- $\mathbf{X}_1,\ldots,\mathbf{X}_{\lfloor N\theta\rfloor} \sim SN_p(\mathbf{0},\mathbf{I},10\times\mathbf{1})$ and $\mathbf{X}_{\lfloor N\theta\rfloor+1},\ldots,\mathbf{X}_N \sim SN_p(\mathbf{0},\sigma^2\mathbf{I},10\times\mathbf{1})$.
- $\mathbf{X}_1,\ldots,\mathbf{X}_{\lfloor N\theta\rfloor} \sim ST_p(\mathbf{0},\mathbf{I},5\times\mathbf{1},4)$ and $\mathbf{X}_{\lfloor N\theta\rfloor+1},\ldots,\mathbf{X}_N \sim ST_p(\mathbf{0},\sigma^2\mathbf{I},5\times\mathbf{1},4)$.

In each of these cases, the parameter σ^2 defines the magnitude by which the scale of the data changes. Similarly as above, we evaluated the empirical power when $p=2$ and sample sizes $N=25$, $N=100$, and $N=200$ for all depths considered and for increasing values of σ^2. Also, in the case of higher dimensions we only considered the Mahalonobis-type depths and spatial depths. These results were again compared to the MCUSUM method of Aue et al. (2009). For

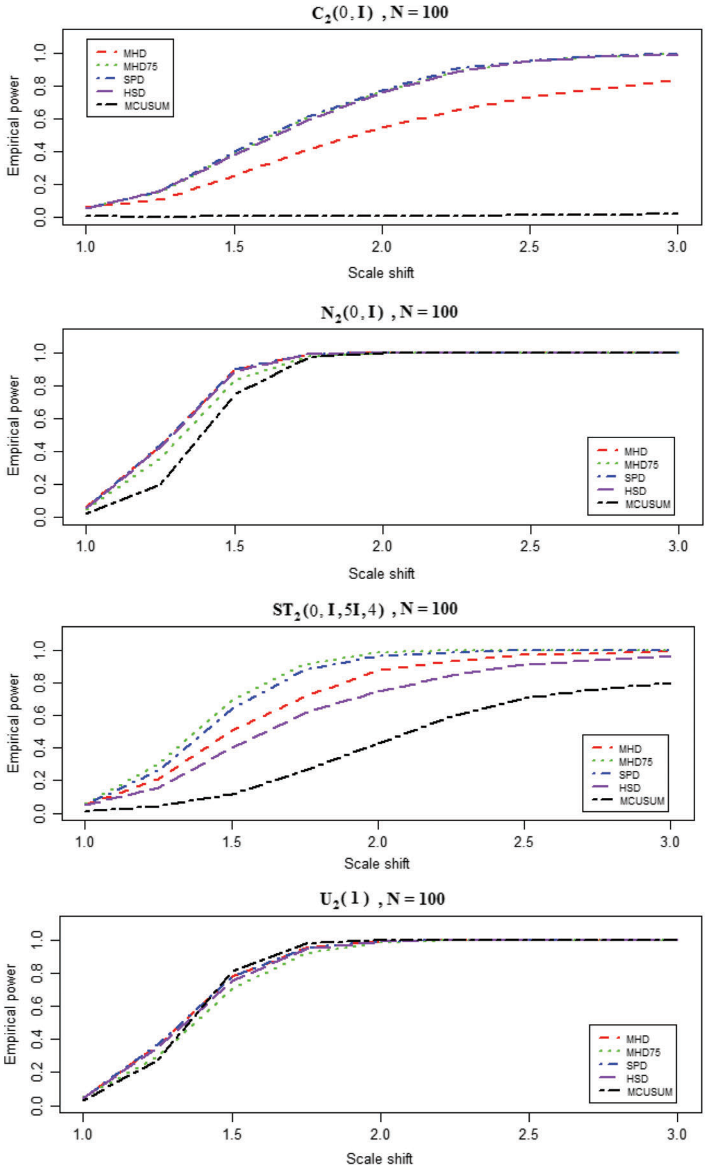

FIGURE 1: The empirical power based on 1,000 independent simulations with nominal level of 5% as a function of σ^2 when $N = 100$ for Gaussian, Cauchy, Skewed-t and Uniform data.

the sake of brevity, in the case of $p = 2$ we have summarized the results with power curves in Figure 1, and complete simulation results are reported in the tables in Appendix B.1. For increasing p the results are reported in Table 3 when the scale shift is $\sigma^2 = 1.25$. We summarize the results as follows:

- When $p = 2$ and for Gaussian data the tests based on T_N exhibited power that exceeded that of the MCUSUM method in the case of normal data. We found this surprising since the MCUSUM method is related to the maximally selected likelihood ratio test for a change in the covariance matrix under the assumption of normality, and hence is in a sense optimal in

TABLE 3: Empirical power based on 1,000 trials with nominal level of 5% for 5, 10 and 20 dimensional DGPs, sample sizes ranging from 25 to 200, and scale shift $\sigma^2 = 1.25$.

		$N = 25$			$N = 100$			$N = 200$		
DGP	Methods	$p = 5$	$p = 10$	$p = 20$	$p = 5$	$p = 10$	$p = 20$	$p = 5$	$p = 10$	$p = 20$
	MHD	0.301	0.398	0.261	0.852	0.986	**1**	0.988	**1**	**1**
	MHD75	0.185	0.244	0.151	0.738	0.937	0.993	0.975	**1**	**1**
$N_p(\mathbf{0}, \mathbf{I})$	MHD50	0.192	0.232	0.151	0.729	0.940	0.997	0.973	**1**	**1**
	SPD	**0.333**	**0.559**	**0.859**	**0.864**	**0.990**	**1**	**0.991**	**1**	**1**
	MCUSUM	0.000	*	*	0.158	0.001	*	0.676	0.532	*
	MHD	0.076	0.095	0.118	0.156	0.191	0.202	0.235	0.301	0.332
	MHD75	0.078	0.087	0.096	**0.174**	0.200	0.212	0.287	0.308	0.336
$C_p(\mathbf{0}, \mathbf{I})$	MHD50	0.078	0.091	0.096	**0.174**	**0.204**	0.204	0.290	0.309	0.338
	SPD	**0.084**	**0.096**	**0.120**	0.173	**0.204**	**0.214**	**0.297**	**0.311**	**0.341**
	MCUSUM	0.000	*	*	0.001	0.000	*	0.002	0.000	*
	MHD	0.146	0.155	0.129	0.379	0.371	0.357	0.618	0.627	0.643
	MHD75	0.125	0.124	0.109	0.320	0.342	0.318	0.593	0.612	0.625
$U_p(1)$	MHD50	0.125	0.120	0.110	0.313	0.333	0.321	0.585	0.610	0.616
	SPD	**0.171**	**0.175**	**0.170**	**0.388**	**0.391**	**0.384**	**0.624**	**0.649**	**0.663**
	MCUSUM	0.000	*	*	0.035	0.000	*	0.188	0.004	*
	MHD	0.238	0.401	0.253	**0.829**	0.978	**1**	**0.989**	**1**	**1**
	MHD75	0.199	0.237	0.163	0.744	0.929	0.989	0.969	**1**	**1**
$SN_p(\mathbf{0}, \mathbf{I}, 3\mathbf{1})$	MHD50	0.202	0.246	0.163	0.722	0.937	0.993	0.971	**1**	**1**
	SPD	**0.285**	**0.540**	**0.841**	0.827	**0.990**	**1**	0.983	**1**	**1**
	MCUSUM	0.000	*	*	0.137	0.002	*	0.641	0.488	*
	MHD	0.291	**0.999**	0.277	0.802	0.982	0.999	**0.981**	**1**	**1**
	MHD75	0.227	0.365	0.151	0.716	0.925	**1**	0.976	0.999	**1**
$SN_p(\mathbf{0}, \mathbf{I}, 10\mathbf{1})$	MHD50	0.225	0.238	0.151	0.715	0.928	0.994	0.973	0.999	**1**
	SPD	**0.332**	0.242	**0.846**	**0.817**	**0.989**	0.998	**0.981**	**1**	**1**
	MCUSUM	0.000	*	*	0.137	0.000	*	0.652	0.473	*
	MHD	**0.112**	0.105	0.153	0.172	**0.203**	0.174	**0.313**	**0.283**	**0.312**
	MHD75	0.105	**0.109**	0.110	0.180	0.202	**0.183**	0.312	0.281	0.304
$ST_p(\mathbf{0}, \mathbf{I}, 5\mathbf{1}, 4)$	MHD50	0.105	**0.109**	0.110	0.180	0.202	**0.183**	0.312	0.281	0.304
	SPD	0.105	0.104	**0.252**	**0.181**	**0.203**	0.180	**0.313**	0.282	0.307
	MCUSUM	0.000	*	*	0.011	0.000	*	0.069	0.018	*

* The test is infeasible due to the fact that the covariance matrix estimates of $\mathrm{vech}(\mathbf{X}_i\mathbf{X}_i^\top)$ are not invertible. Bold values indicate the tests with the highest empirical power.

this situation. We investigated this further and think this can be attributed to the phenomenon of "nonmonotonic power" (Vogelsang, 1999), which arises due to estimating the covariance matrix of $\mathrm{vech}(\mathbf{X}_i\mathbf{X}_i^\top)$ under $H_{A,\theta}$. With simulations reported in Appendix B.2, we show that when this covariance matrix estimate is replaced by its theoretical value, we call this modified procedure MCUSUM*, then the MCUSUM test is indeed the best among those considered in this situation.

- For skewed and/or heavy tailed data the test based on SPD and MHD75 performed best, and all methods performed very similarly in terms of detecting scale changes in uniformly distributed data.

- In higher-dimensions, SPD often exhibited the highest power (see Table 2).
- In summary, the tests which seemed to balance size with power the best in both low and high dimensions were the tests based on T_N derived from the spatial depth SPD and robust Mahalonobis depth MHD75.

To evaluate the computational time required to calculate the test statistic for each of the considered depth functions, we carried out a numerical study in two extreme scenarios with Normal data: $p = 2$ and $N = 25$, and $p = 20$ and $N = 200$. The run time for the calculation of test statistics for MHD, MHD75, MHD50, SPD, HSD, SLD and OD are (in seconds) 1.274, 0.189, 0.173, 1.426, 0.290, 0.294 and 0.247 s respectively, when $p = 2$ and $N = 25$. When $p = 20$ and $N = 200$, the computation times for the same depths except SLD and OD, which are infeasible to compute in this dimension, are 11.954, 0.236, 0.241, 23.130 1.176, respectively. Generally we found that the simulations based on MHD and SPD depths required somewhat more time to complete, whereas the other depths considered generally required similar amounts of time to compute.

It is important, especially in practice, to address how the choice of depth function impacts the detection of change points. From the simulation results in dimension $p = 2$, as can be seen from Table 2 that all depth functions perform generally more or less the same in terms of type I error of detection. There is a slight inflation seen in the cases of HSD, SLD for small samples ($N = 25$) that is due to the ranks of data points with tied depth values been averaged rather than randomly broken. This disappears as the sample increases to $N = 100$ and $N = 200$. In terms of power of detection, Figure 1 indicates that both SPD and MHD75 perform the best overall distributional settings for $p = 2$. This observation holds true for higher dimensions in the case of SPD for all sample sizes.

4.3. Changes in the Direction of Variability

In this subsection we intend to illustrate one weakness of the proposed method. The simulations in Section 4.2 show that the depth-based change point statistic T_N is capable of detecting changes in the variability of vector-valued data that can be generally described as “expansions/contractions" of the distribution. However, if the nature of the change of the variability of the distribution is that the scale remains the same but the direction changes, then in general we expect the depths before and after the change to be similar, and hence we do not expect T_N to be capable of detecting the change in distribution. To further investigate and illustrate this issue, we conducted a simulation with normal data following $H_{A,0.5}$ in which $\mathbf{X}_1, \dots, \mathbf{X}_{\lfloor N\theta \rfloor} \sim N_2(\mathbf{0}, \mathbf{\Sigma}_1)$ and $\mathbf{X}_{\lfloor N\theta \rfloor+1}, \dots, \mathbf{X}_N \sim N_2(\mathbf{0}, \mathbf{\Sigma}_2)$, where $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are defined as

$$\mathbf{\Sigma}_1 = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & \sigma^2 \end{bmatrix}.$$

An illustration of this data is given in Figure 2. Table 4 contains the results from applying the tests under study to the simulated data generated as above. As expected, in this case MCUSUM and MCUSUM* outperform the depth-based methods. Among different variants of the depth measures, it seems that MHD50 and MHD75 can pick up the change in the direction of variability to some extent.

5. APPLICATION TO ACID RAIN DATA

In this section we consider an application of the proposed change point detection procedure to analyze measurements of pollutants associated with acid rain. Acid rain is an offensive natural phenomenon that can adversely affect plants, fresh water lakes and soils, and can also harm

TABLE 4: Empirical power based on 1,000 trials with the nominal level of 5% for change in the direction of variability test for bivariate distributions with different scale shifts, and sample sizes 25, 100 and 200.

Methods	$\sigma = 1.25$	$\sigma = 1.5$	$\sigma = 1.75$	$\sigma = 2$	$\sigma = 2.25$	$\sigma = 2.5$	$\sigma = 2.75$	$\sigma = 3$
				$N = 25$				
MHD	0.042	0.042	0.021	0.030	0.018	0.018	0.018	0.018
MHD75	0.035	0.062	0.059	0.096	0.105	0.107	0.124	0.146
MHD50	0.035	0.062	0.059	0.096	0.105	0.107	0.124	0.146
SPD	0.044	0.048	0.034	0.049	0.043	0.049	0.040	0.048
HSD	0.066	0.070	0.047	0.057	0.035	0.040	0.031	0.044
SLD	0.061	0.061	0.042	0.051	0.031	0.036	0.031	0.038
OD	0.042	0.046	0.025	0.031	0.020	0.020	0.016	0.017
MCUSUM	0.009	0.011	0.026	0.040	0.062	0.072	0.102	0.119
MCUSUM*	0.142	0.413	0.656	0.824	0.925	0.944	0.975	0.989
				$N = 100$				
MHD	0.051	0.033	0.034	0.020	0.023	0.021	0.009	0.016
MHD75	0.061	0.102	0.112	0.169	0.206	0.241	0.286	0.327
MHD50	0.062	0.106	0.112	0.167	0.205	0.239	0.283	0.330
SPD	0.050	0.043	0.042	0.034	0.049	0.043	0.031	0.044
HSD	0.052	0.033	0.031	0.017	0.025	0.018	0.010	0.011
SLD	0.049	0.039	0.031	0.023	0.028	0.022	0.016	0.017
OD	0.050	0.037	0.035	0.020	0.026	0.022	0.010	0.017
MCUSUM	0.199	0.724	0.966	1	1	1	1	1
MCUSUM*	0.458	0.946	0.998	1	1	1	1	1
				$N = 200$				
MHD	0.041	0.040	0.029	0.025	0.017	0.020	0.016	0.011
MHD75	0.060	0.095	0.146	0.182	0.202	0.264	0.327	0.354
MHD50	0.057	0.098	0.139	0.184	0.205	0.267	0.325	0.351
SPD	0.046	0.047	0.045	0.042	0.041	0.054	0.039	0.033
HSD	0.038	0.036	0.032	0.025	0.021	0.017	0.011	0.007
SLD	0.042	0.036	0.032	0.020	0.023	0.024	0.014	0.010
OD	0.040	0.038	0.030	0.023	0.020	0.023	0.012	0.009
MCUSUM	0.568	0.997	1	1	1	1	1	1
MCUSUM*	0.775	0.998	1	1	1	1	1	1

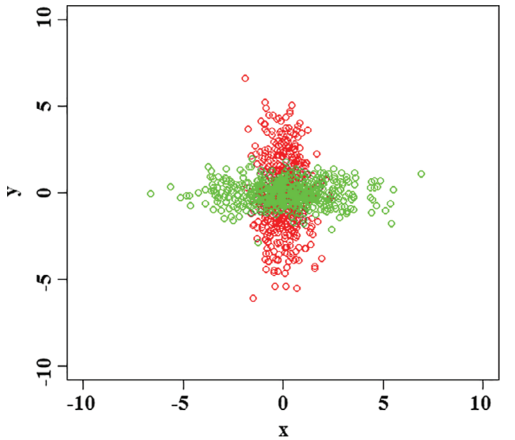

FIGURE 2: A plot of change in the direction of variability for 400 data points, $\theta = 0.5$ and $\sigma = 3$.

insects and aquatic animals. Due to these adverse effects, acid rain has been the focus of a number of policies adopted in Canada and the US that aim to limit emissions of sulphur dioxide and nitrogen oxide, including the Eastern Canada Acid Rain Program, which was established in 1985, and major amendments to the US Clean Air Act in 1990. The goal of this analysis is to determine if the rate of change in acid rain levels remained stable following these efforts to curb emissions, or if they changed one or more times.

The specific data that we consider came from Turkey Lake, which is a large freshwater lake located in central Ontario (Algoma District) Canada, and consists of approximately biweekly (twice per month) measurements of pH, sulphate concentration (SO_4 (meq/L)), calcium concentration (Ca (mg/L)) and alkalinity (meq/L) taken between 15 February 1980 and 22 December 2003. We linearly interpolated the original measurements and evaluated this interpolation biweekly in order to obtain a biweekly time series of length 573. Figure 3 shows the resulting first differenced time series. One notable feature of these series is that they each contain a few outlying points.

We note again here that the test proposed above is developed assuming the data under the null hypothesis are independent and identically distributed, and should be used with caution when applied to time series that might be serially dependent. The autocorrelation functions of each first differenced series are shown in Appendix B.3, which suggest that these series are at least approximately weak white noises.

To this multivariate sequence we applied the change point test based on T_N using the depth measures SPD and MHD75 as well as the MCUSUM method of Aue et al. (2009). A plot of the process $|Z_N(t)|$ against the 95% quantile of $\sup_{0 \le t \le 1} |B(t)|$ for both depth measures as well as the MCUSUM process defined in Aue et al. (2009) is plotted in Figures 4 and 5, from which it is clear that each test provides strong evidence against the hypothesis that the scale/variance of these series remains constant throughout the sample. The fact that the tests agree suggests that the MCUSUM test is not strongly affected by the outliers present in the series. Using $\widehat{k}^*$, we estimate that the change in scale in these series occurred in late 1989, which is consistent with the adoption of clean air policies and reduced emissions in Canada and the US at the end of the 1980s. In order to identify potential additional change points, we applied binary segmentation with Bonferroni correction, which entailed applying the change point tests again to the data before and after this estimated change point. Using T_N for each subsample we could not reject H_0 at level 0.05, indicating the rate of change of these series appears to be stable

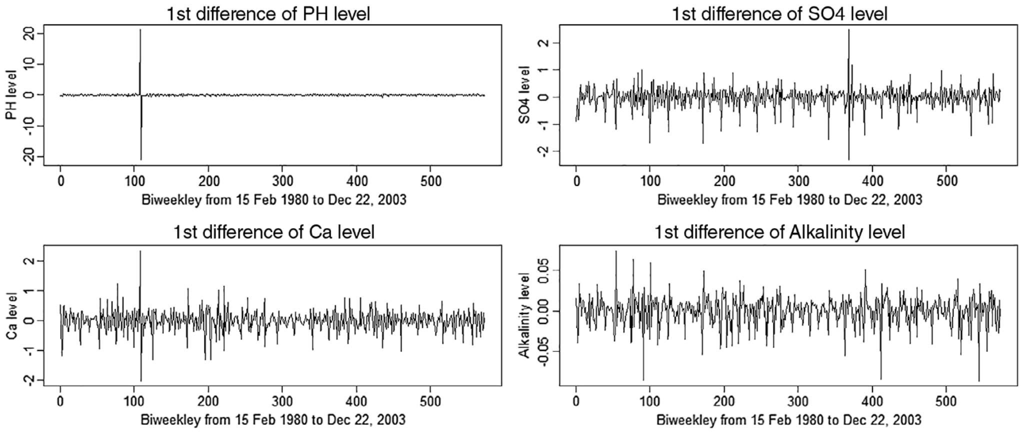

FIGURE 3: First differenced pH, SO_4, Ca and alkalinity series.

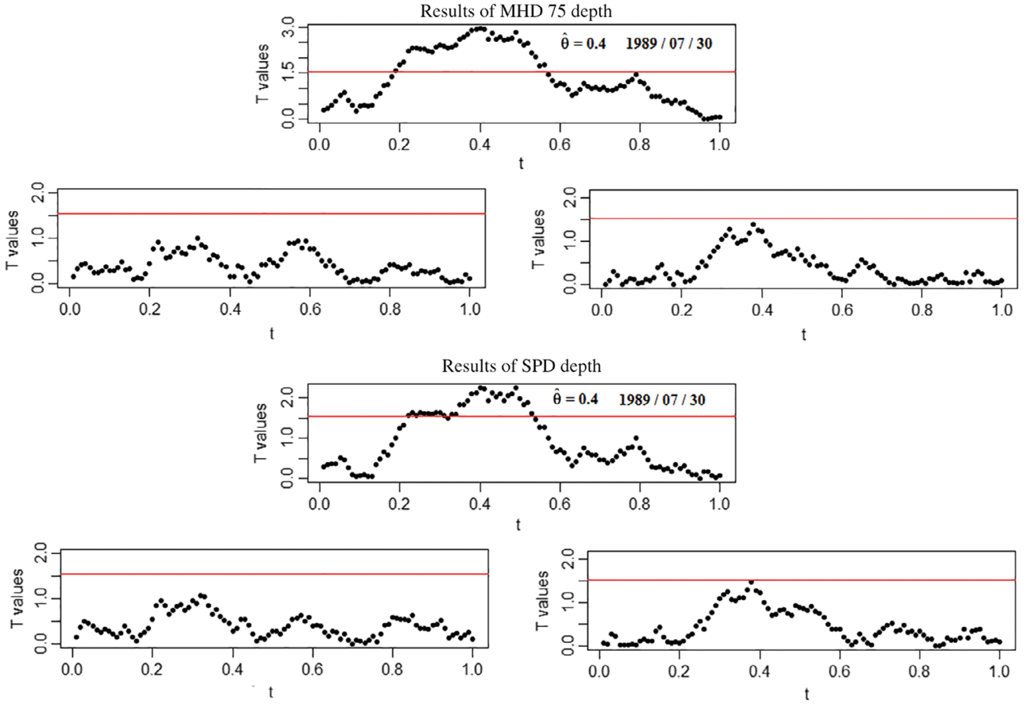

FIGURE 4: Plots of $| Z_N(t) |$ based on MHD75 and SPD depth measures. The top panels show $| Z_N(t) |$ computed from the whole sample, and the bottom panels show $| Z_N(t) |$ calculated from the subsamples after binary segmentation. The horizontal lines indicate the Bonferroni corrected confidence limits.

before and after 1989/1990. Also, when we applied the binary segmentation procedure with Bonferroni correction using the MCUSUM method, a similar result was observed, which shows the agreement between the outcome of the proposed method with that of Aue et al. (2009). Given that each of these procedures has power when the nature of the change is a change in scale, these

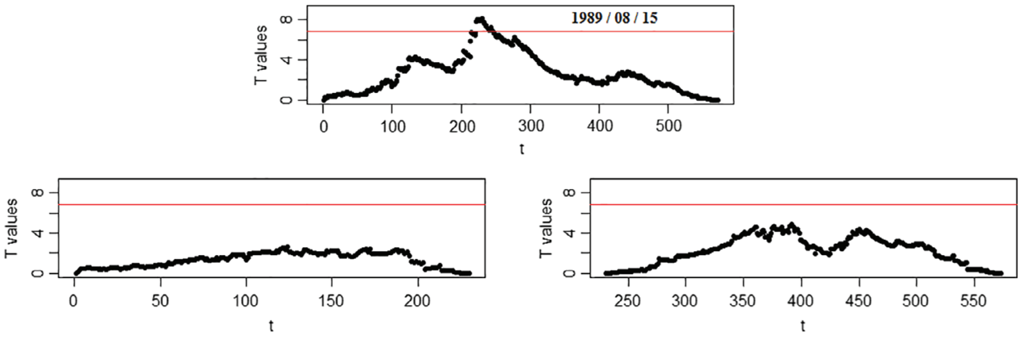

FIGURE 5: Plots of the MCUSUM process defined in Aue et al. (2009) applied to the whole sample as well as subsamples determined using binary segmentation. The horizontal lines indicate the Bonferroni corrected confidence limits.

results suggest that the distribution of the rate of change of these series contracts after the change point in 1989/1990.

6. CONCLUSION

In the current investigation, a multivariate nonparametric change point detection test was proposed based on the ranks of data depths. Through a comprehensive comparative study, it was demonstrated that the method can work very well with different types of depth functions, and also detect changes in the stochastic structure of data with various dimensions, and sample sizes, especially when the nature of the change is an expansion or contraction of the distribution. Also, by theoretical investigation of the asymptotic properties of the proposed test, it was proven that the changes can be consistently detected, and estimated based on data depth ranking. Simulations and a data application indicate that the proposed method improves over standard approaches based on covariance matrix estimation when the underlying distribution is heavy tailed or skewed, and/or when the dimensionality of the data is large.

ACKNOWLEDGEMENTS

The authors would like to thank two anonymous referees, an Associate Editor, and the Acting Editor Richard Lockhart for their helpful comments. The research of Drs. Chenouri and Rice was partially supported by the Natural Sciences and Engineering Research Council of Canada. Dr. Chenouri also would like to thank Christopher G. Small and Abdel H. El-Shaarawi for the fruitful discussion on an earlier version of this paper. The acid rain dataset was kindly provided by Abdel H. El-Shaarawi.

BIBLIOGRAPHY

Alexander, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Annals of Probability*, 12, 1041–1067.

Aue, A., Hörmann, S., Horváth, L., & Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *Annals of Statistics*, 37, 4046–4087.

Bandyopadhyay, U. & Mukherjee, A. (2007). Nonparametric partial sequential test for location shift at an unknown time point. *Sequential Analysis*, 26, 99–113.

Bhattacharya, P. K. & Frierson, D. (1981). A nonparametric control chart for detecting small disorders. *Annals of Statistics*, 9, 544–554.

Bickel, P. J. (1964). On some alternative estimates for shift in the p-variate one sample problem. *Annals of Mathematical Statistics*, 35, 1079–1090.

Bickel, P. J. (1965). On some asymptotically non-parametric competitors of Hotelling's T^2. *Annals of Mathematical Statistics*, 36, 160–173.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Brodsky, B. & Darkhovsky, B. (1993). *Nonparametric Methods in Change-Point Problems*. Kluwer, Berlin.

Brown, B. M. (1983). Statistical use of spatial median. *Journal of Royal Statistical Society, B*, 45, 23–30.

Chamberlin, E. (1937). *The Theory of Monopolistic Competition*. Harvard University Press.

Carlstein, E. G., Müller, H. -G., & Siegmund, D. (1994). *Change-Point Problems, IMS Lecture Notes*, Vol. 23.

Chenouri, S. & Small, C. G. (2012). A multivariate nonparametric multi-sample test based on data depth. *Electronic Journal of Statistics*, 6, 760–782.

Chenouri, S., Small, C. G., & Farrar, T. J. (2011). Data depth-based nonparamertic scale tests. *Canadian Journal of Statistics*, 39, 356–369.

Chernoff, H. & Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Annals of Mathematical Statistics*, 29, 972–994.

Csörgő, M. & Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, Chichester.

Dette, H., Wu, W., & Zhou, Z. (2015). Change point analysis of second order characteristics in non-stationary time series, Technical report.

Dehling, H. & Fried, R. (2012). Asymptotic distribution of two-sample empirical U-quantiles with applications to robust tests for shifts in location. *Journal of Multivariate Analysis*, 105, 124–140.

Donoho, D. L. (1982). Breakdown properties of multivariate location estimators, Ph.D. Thesis, Department of Statistics, Harvard University.

Donoho, D. L. & Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlingness. *Annals of Statistics*, 20, 1803–1827.

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

Galeano, P. & Peña, D. (2007). Covariance changes detection in multivariate time series. *Journal of Statistical Planning and Inference*, 137, 194–211.

Genest, M., Masse, J. -C., & Plante, J. -F. (2017). *depth: Nonparametric depth functions for multivariate analysis. R package version 2.1-1*.

Gombay, E. (1994). Testing for change-points with rank and sign statistics. *Statistics & Probability Letters*, 20, 49–56.

Gombay, E. & Hušková, M. (1997). Rank based estimators of the change-point. *Journal of Statistical Planning and Inference*, 67, 137–154.

Gower, J. C. (1974). Algorithm as 78: The mediancentre. *Applied Statistics*, 23, 466–470.

Hájek, J. & Šidák, Z. V. (1967). *Theory of Rank Test*. Academic Press, New York.

Hodges, J. L. (1955). A bivariate sign test. *Annals of Mathematical Statistics*, 26, 523–527.

Horváth, L. & Rice, G. (2014). Extensions of some classical methods in change point analysis. *TEST*, 23, 219–255.

Hotelling, H. (1929). Stability in competition. *The Economic Journal*, 39, 41–57.

Hušková, M. (2013). Robust change point analysis. *Robustness and Complex Data Structures*, Springer, 171–190.

Inclán, C. & Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of change of variance. *Journal of the American Statistical Association*, 89, 913–923.

Liu, R. Y. (1990). On a notion of data depth based on random simplices. *Annals of Statistics*, 18, 405–414.

Liu, R. Y., Parelius, J. M., & Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Annals of Statistics*, 27, 783–858.

Liu, R. Y. & Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, 88, 252–260.

Liu, R. Y. & Singh, K. (2006). Rank tests for multivariate scale difference based on data depth. *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications, DIMACS Series*, AMS, 17–36.

Lung-Yut-Fong, A., Lévy-Leduc, C., & Cappé, O. (2015). Homogeneity and change-point detection tests for multivariate data using rank statistics. *Journal of the French Statistical Society*, 156, 133–162.

Kao, C., Trapani, L., & Urga, G. (2018). Testing for instability in covariance structures. *Bernoulli*, 24, 740–771.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceeding of National Academy of India*, 12, 49–55.

Massart, P. (1986). Rates of convergence in the central limit theorem for empirical processes. *Annales de l'Institut Henri Poincaré*, 22, 381–423.

R Development Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rao, C. R. (1988). Methodology based on the L1 norm in statistical inference. *Sankhya Series A*, 50, 289–313.

Rousseeuw, P. J. & Ruts, I. (1998). Constructing the bivariate Tukey median. *Statistica Sinica*, 8, 828–839.

Rousseeuw, P. J. & van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of American Statistical Association*, 85, 633–639.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics Probability Letters*, 1, 327–333.

Segaert, P., Hubert, M., Rousseeuw, P., & Raymaekers, J. (2018). *mrfDepth: Depth measures in multivariate, regression and functional settings. R package version 1.0.10*.

Shieh, G., Jan, S., & Randles, R. H. (2006). On power and sample size determinations for the Wilcoxon-Mann-Whitney test. *Journal of Nonparametric Statistics*, 18, 33–43.

Small, C. G. (1990). A survey of multidimensional medians. *International Statistical Review*, 58, 263–277.

Struyf, A. & Rousseeuw, P. J. (2000). High-dimensional computation of the deepest location. *Computational Statistics and Data Analysis*, 34, 415–426.

Tukey, J. W. (1975). Mathematics and picturing data. *Proceedings of the International Congress of Mathematics Vancouver 1974*; pp. 523–531.

Zhou, Z. (2013). Heteroscedasticity and Autocorrelation Robust Structural Change Detection. *Journal of the American Statistical Association*, 108, 726–740.

Zuo, Y. & He, X. (2006). On the limiting distributions of multivariate depth-based rank sum statistics and related tests. *Annals of Statistics*, 34, 2879–2896.

Zuo, Y. & Serfling, R. (2000). General notions of statistical depth function. *Annals of Statistics*, 28, 461–482.

Vogelsang, T. J. (1999). Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *Journal of Econometrics*, 88, 283–299.

Wied, D., Krämer, W., & Dehling, H. (2012). Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory*, 28, 570–589.

APPENDIX

A. Proofs of Theorems in Section 3

Proof of Theorem 1. Define

$$U_i = \frac{\widehat{R}_i - (N+1)/2}{\sqrt{N(N^2-1)/12}}, \qquad i = 1, \ldots, N\,.$$

Evidently $Z_N(t) = \sum_{i=1}^{\lfloor Nt \rfloor} U_i$. Under H_0, the random variables $U_1, \ldots, U_N$ are exchangeable, and $\sum_{i=1}^{N} U_i = 0$ while also $\sum_{i=1}^{N} U_i^2 = 1$. In addition,

$$\max_{1 \le i \le N} |\, U_i \,| = \frac{N - (N+1)/2}{\sqrt{N(N^2-1)/12}} = \frac{N-1}{\sqrt{N(N^2-1)/3}} \to 0\,, \quad \text{as} \quad N \to \infty\,.$$

This shows that the conditions of Theorem 24.1 of Billingsley (1968) are satisfied for the process $Z_N(t) = \sum_{i=1}^{\lfloor Nt \rfloor} U_i$, from which the theorem follows. ■

Proof of Theorem 2. Throughout the proof we let c_i, $i \geq 0$, denote unimportant numerical constants. Clearly $T_N \geq |\, Z_N(\theta) \,|$, and so the theorem follows if $|\, Z_N(\theta) \,| \overset{P}{\to} \infty$, as $N \to \infty$. For

$0 \leq t \leq 1$, let

$$\tilde{Z}_N(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{\lfloor Nt \rfloor} \frac{R_i - (N+1)/2}{\sqrt{(N^2-1)/12}},$$

where

$$R_i = \#\{\mathbf{X}_j \ ; \ D(\mathbf{X}_j\,;F_*) \leq D(\mathbf{X}_i\,;F_*), \ j = 1, \ldots N\}\,. \tag{A1}$$

Evidently

$$Z_N(\theta) = G_N + \tilde{Z}_N(\theta), \tag{A2}$$

with $G_N = Z_N(\theta) - \tilde{Z}_N(\theta)$. By the triangle inequality

$$|\,G_N\,| \leq \frac{1}{\sqrt{N}} \left(\frac{N^2-1}{12}\right)^{-1/2} \sum_{i=1}^{\lfloor N\theta \rfloor} |\,\widehat{R}_i - R_i\,| \leq c_0\, N^{-3/2} \sum_{i=1}^{\lfloor N\theta \rfloor} |\,\widehat{R}_i - R_i\,|\,.$$

Let

$$A_{i,j} = \{D(\mathbf{X}_j, F_*) \leq D(\mathbf{X}_i, F_*)\} \cap \{D(\mathbf{X}_j, \widehat{F}_N) > D(\mathbf{X}_i, \widehat{F}_N)\}\,,$$
$$B_{i,j} = \{D(\mathbf{X}_j, F_*) > D(\mathbf{X}_i, F_*)\} \cap \{D(\mathbf{X}_j, \widehat{F}_N) \leq D(\mathbf{X}_i, \widehat{F}_N)\}\,.$$

Then clearly

$$|\,\widehat{R}_i - R_i\,| \leq \sum_{j=1}^{N} \mathbb{1}(A_{i,j}) + \sum_{j=1}^{N} \mathbb{1}(B_{i,j}), \tag{A3}$$

where $\mathbb{1}(A)$ denotes the indicator of the event A. It follows according to the definition of $A_{i,j}$ that if $\gamma_N = \max_{1 \leq j \leq N} |\,D(\mathbf{X}_j, F_*) - D(\mathbf{X}_j, \widehat{F}_N)\,|$, then

$$A_{i,j} \subset \{|\,D(\mathbf{X}_i, F_*) - D(\mathbf{X}_j, F_*)\,| \leq 2\gamma_N\}\,.$$

Moreover, by the definitions of $D(\mathbf{X}_i, F_*)$ and $D(\mathbf{X}_i, \widehat{F}_N)$, it follows that

$$\gamma_N \leq \sup_{\mathbf{x} \in \mathbb{R}^p} |\,D(\mathbf{x}, \widehat{F}_N) - D(\mathbf{x}, F_*)\,|\,,$$

and hence

$$\mathbb{1}(A_{i,j}) \leq \mathbb{1}(\{|\,D(\mathbf{X}_i, F_*) - D(\mathbf{X}_j, F_*)\,| \leq 2 \sup_{\mathbf{x} \in \mathbb{R}^p} |\,D(\mathbf{x}, \widehat{F}_N) - D(\mathbf{x}, F_*)\,|\}). \tag{A4}$$

According to Assumption 1, and since $D(\mathbf{X}_i, F_*)$ and $D(\mathbf{X}_j, F_*)$ are independent for $i \neq j$, the random variable $|\,D(\mathbf{X}_i, F_*) - D(\mathbf{X}_j, F_*)\,|$ has a Lipschitz distribution function for $i \neq j$, from which it follows that for any $1 \leq i, j \leq N$,

$$\mathrm{E}[\mathbb{1}(\{|\,D(\mathbf{X}_i, F_*) - D(\mathbf{X}_j, F_*)\,| \leq 2 \sup_{\mathbf{x} \in \mathbb{R}^p} |\,D(\mathbf{x}, \widehat{F}_N) - D(\mathbf{x}, F_*)\,|\})]$$
$$= \mathrm{E}[\mathrm{E}[\mathbb{1}(\{|\,D(\mathbf{X}_i, F_*) - D(\mathbf{X}_j, F_*)\,|$$

$$\leq 2 \sup_{\mathbf{x}\in\mathbb{R}^p} |D(\mathbf{x}, \widehat{F}_N) - D(\mathbf{x}, F_*)|\})| \sup_{\mathbf{x}\in\mathbb{R}^p} |D(\mathbf{x}, \widehat{F}_N) - D(\mathbf{x}, F_*)|]]$$

$$\leq c_1 \mathrm{E}[\sup_{\mathbf{x}\in\mathbb{R}^p} |D(\mathbf{x}, \widehat{F}_N) - D(\mathbf{x}, F_*)|] = O(N^{-1/2}),$$

where in the last line we applied Assumption 1. This combined with (A4) gives that

$$\mathrm{E}\left(\sum_{j=1}^{N} \mathbb{1}(A_{i,j})\right) = O(N^{1/2}).$$

The same arguments imply that the above bound also holds for the sum of the $B_{i,j}$ terms in (A3), which gives that

$$\mathrm{E}|\widehat{R}_i - R_i| = O(N^{1/2}). \tag{A5}$$

This along with (A3) implies that $\mathrm{E}|G_N| = O(1)$, and hence $G_N = O_P(1)$ by Markov's inequality. The random variables $D(\mathbf{X}_i, F_*)$, $1 \leq i \leq N$ satisfy the conditions of the Chernoff and Savage Theorem on page 234 of Hájek & Šidák (1967), from which it follows that $\tilde{Z}_N(\theta)$ is asymptotically Gaussian with bounded variance and asymptotic mean

$$\sqrt{N}\theta\left[\int_{-\infty}^{\infty} \theta H_1(u) + (1-\theta)H_2(u)dH_1(u) - \frac{1}{2}\right],$$

which tends in absolute value to positive infinity under (3.3). This implies that $|\tilde{Z}_N(\theta)| \xrightarrow{P} \infty$, which, along with (A2) and the fact that $G_N = O_P(1)$, proves the result. ■

Proof of Theorem 3. Let

$$Z_k = Z_N(k) = \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \frac{\widehat{R}_i - (N+1)/2}{\sqrt{(N^2-1)/12}},$$

$$\tilde{Z}_k = \tilde{Z}_N(k) = \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \frac{R_i - (N+1)/2}{\sqrt{(N^2-1)/12}},$$

$$G_k = Z_k - \tilde{Z}_k = \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \frac{R_i - \widehat{R}_i}{\sqrt{(N^2-1)/12}}.$$

Under these definitions, it is clear that according to the definition of $\widehat{k^*}$,

$$\widehat{k^*} = \min\{k \ : \ |Z_k| = \max_{1\leq i\leq N} |Z_i|\}$$

$$= \min\{k \ : \ |Z_k|^2 = \max_{1\leq i\leq N} |Z_i|^2\}$$

$$= \min\{k \ : \ |Z_k|^2 - |\tilde{Z}_{k^*}|^2 = \max_{1\leq i\leq N} |Z_i|^2 - |\tilde{Z}_{k^*}|^2\}.$$

From the definition of $\widehat{\theta}$, the statement of Theorem 3 is equivalent with the existence of a real-valued sequence a_N satisfying $a_N \to \infty$, $a_N/N \to 0$, and

$$\Pr(|\widehat{k^*} - k^*| > a_N) = \Pr(\widehat{k^*} < k^* - a_N) + \Pr(\widehat{k^*} > k^* + a_N) \to 0, \quad N \to \infty. \tag{A6}$$

We now show that for the sequence $a_N = N^{\kappa}$, $1/2 < \kappa < 1$,

$$\Pr(\widehat{k^*} < k^* - a_N) \to 0\,, \quad as \quad N \to \infty\,.$$

The second term on the right-hand side of (A6) can be handled in a similar way, and so the details are omitted. Evidently

$$\Pr(\widehat{k^*} < k^* - a_N) \le \Pr\left(\max_{1 \le i \le k^* - 2a_N} |Z_i|^2 - |\tilde{Z}_{k^*}|^2 > \max_{k^* - a_N \le i \le k^*} |Z_i|^2 - |\tilde{Z}_{k^*}|^2 \right). \tag{A7}$$

Since $Z_i = \tilde{Z}_i + G_i$, we have by some simple algebra that for $1 \le i \le k^*$,

$$|Z_i|^2 - |\tilde{Z}_{k^*}|^2 = A_i + B_i + C_i + G_i^2 + 2\tilde{Z}_i G_i,$$

where

$$A_i = \left(\tilde{Z}_i - \frac{i}{k^*}\tilde{Z}_{k^*}\right)^2, \quad B_i = 2\left(\tilde{Z}_i - \frac{i}{k^*}\tilde{Z}_{k^*}\right)\frac{i}{k^*}\tilde{Z}_{k^*},$$

and

$$C_i = \tilde{Z}_{k^*}^2 \left(\left(\frac{i}{k^*}\right)^2 - 1 \right).$$

It follows under the conditions of Theorem 2 and the Chernoff-Savage theorem (Chernoff & Savage, 1958) that there exists nonzero constants c_0 and $c_1 > 0$ so that $\tilde{Z}_{k^*}^2 = N\,c_0^2\,(1 + o_P(1))$, and hence for each $i \le k^*$,

$$C_i = c_0^2\,N\left(\frac{i^2 - (k^*)^2}{(k^*)^2}\right)(1 + o_P(1)) = c_1\left(\frac{i^2 - (k^*)^2}{k^*}\right)(1 + o_P(1)). \tag{A8}$$

Moreover, it follows by the definition of C_i that for all N, $C_i < C_j$ for all $i < j$, and by the above property and the fact that $k^* = \lfloor N\theta \rfloor$, if $a_N = N^{\kappa}$ for $1/2 < \kappa < 1$, we get with a little algebra that $C_{k^*-a_N} - C_{k^*-2\,a_N} = 2\,a_N(1 + o_P(a_N)) \overset{P}{\to} \infty$, as $N \to \infty$. We now aim to show the following four results:

$$\max_{1 \le k \le k^* - 2a_N} \frac{A_k}{|C_k|} = o_P(1)\,, \tag{A9}$$

$$\max_{1 \le k \le k^* - 2a_N} \frac{B_k}{|C_k|} = o_P(1)\,, \tag{A10}$$

$$\max_{1 \le k \le k^* - 2a_N} \frac{G_k^2}{|C_k|} = o_P(1)\,, \tag{A11}$$

$$\max_{1 \le k \le k^* - 2a_N} \frac{\tilde{Z}_k G_k}{|C_k|} = o_P(1)\,. \tag{A12}$$

Towards establishing (A9), we note that for all $k \leq k^*$,

$$\left(\frac{k^*}{(k^*)^2 - k^2}\right) \leq \frac{k^*}{k(k^* - k)}.$$

Since the ranks R_k, $k < k^*$, are the ranks of independent and identically distributed $D(\mathbf{X}_1, F_*), \ldots, D(\mathbf{X}_k, F_*)$, and $\tilde{Z}_k$ is the partial sum of these ranks, it follows from equation (2.22) in Gombay & Hušková (1997) that for all $\varepsilon > 0$,

$$\begin{aligned}
&\Pr\left(\max_{1 \leq k \leq k^* - 2a_N} \frac{k^*}{k(k^* - k)} A_k < \varepsilon\right) \\
&= \Pr\left(\max_{1 \leq k \leq k^* - 2a_N} \frac{k^*}{k(k^* - k)} \left(\sum_{i=1}^{k} \frac{R_i - (N+1)/2}{\sqrt{(N^2-1)/12}} - \frac{k}{k^*} \sum_{i=1}^{k^*} \frac{R_i - (N+1)/2}{\sqrt{(N^2-1)/12)}}\right)^2 < \varepsilon N\right) \\
&\longrightarrow 1 \text{ as } N \to \infty,
\end{aligned}$$

proving (A9). Equation (A10) can be established using a similar argument as that used to establish (2.18) in Gombay & Hušková (1997), so the details are omitted here. In order to show (A11), we note that according to the definition of G_k and the triangle inequality that

$$\begin{aligned}
\max_{1 \leq k \leq k^* - 2a_N} |G_k| &= \max_{1 \leq k \leq k^* - 2a_N} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \frac{R_i - \widehat{R}_i}{\sqrt{(N^2-1)/12}} \right| \\
&\leq \max_{1 \leq k \leq k^* - 2a_N} \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \left| \frac{R_i - \widehat{R}_i}{\sqrt{(N^2-1)/12}} \right| \\
&\leq \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left| \frac{R_i - \widehat{R}_i}{\sqrt{(N^2-1)/12}} \right| = O_P(1),
\end{aligned} \tag{A13}$$

where the last line follows from (A5) and Chebyshev's inequality. Evidently then $\max_{1 \leq k \leq k^* - 2\, a_N} |G_k|^2 = O_P(1)$, and also for some positive constant c_2 and for $k < k^*$,

$$\left| c_0^2 N \left(\frac{k^2 - (k^*)^2}{(k^*)^2}\right) \right| \geq c_2 N,$$

from which with (A8), (A11) now follows. Equation (A12) follows similarly. In addition to (A9)–(A12), we also have that

$$\max_{k^* - a_N \leq k \leq k^*} A_k = O_P(1), \tag{A14}$$

$$\max_{k^* - a_N \leq k \leq k^*} B_k = O_P(1), \tag{A15}$$

$$\max_{k^* - a_N \leq k \leq k^*} G_k^2 = O_P(1), \tag{A16}$$

$$\max_{k^* - a_N \leq k \leq k^*} \tilde{Z}_k\, G_k = o_P(a_N). \tag{A17}$$

Equations (A14) and (A15) follow from Lemma 2.1 in Gombay & Hušková (1997). Equation (A11) follows from (A13), and the crude approximation that, with some positive constant c_5,

$$\begin{aligned}
\max_{1\le k\le N} |\tilde{Z}_k| &= \max_{1\le k\le N}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{k}\frac{R_i-(N+1)/2}{\sqrt{(N^2-1)/12}}\right| \\
&\le \max_{1\le k\le N}\frac{1}{\sqrt{N}}\sum_{i=1}^{k}\left|\frac{R_i-(N+1)/2}{\sqrt{(N^2-1)/12}}\right| \\
&\le \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left|\frac{R_i-(N+1)/2}{\sqrt{(N^2-1)/12}}\right| \\
&\le c_5\sqrt{N},
\end{aligned}$$

where in the last line we used that $|R_i-(N+1)/2|\le N$. Now combining the above results, we have for all $\epsilon>0$ and N sufficiently large,

$$\begin{aligned}
&\Pr\left(\max_{1\le k\le k^*-2a_N}|Z_k|^2-|\tilde{Z}_{k^*}|^2>\max_{k^*-a_N\le k\le k^*}|Z_k|^2-|\tilde{Z}_{k^*}|^2\right) \\
&\quad\le \Pr\left(\max_{1\le k\le k^*-2a_N} C_k\left(1+\frac{A_k}{|C_k|}+\frac{B_k}{|C_k|}+\frac{G_k^2}{|C_k|}+\frac{\tilde{Z}_k G_k}{|C_k|}\right)\right. \\
&\qquad\qquad \left. >\max_{k^*-a_N\le k\le k^*} A_k+B_k+C_k+G_k^2+2\tilde{Z}_k G_k\right) \\
&\quad\le \Pr(C_{k^*-a_N}\Lambda_N-C_{k^*-2a_N}<\max_{k^*-a_N\le k\le k^*}A_k+B_k+G_k^2+2\tilde{Z}_kG_k) \\
&\quad\to 0,\quad as\quad N\to\infty,
\end{aligned}$$

since $\Lambda_N=\max_{1\le k\le k^*-2a_N}(1+\frac{A_k}{|C_k|}+\frac{B_k}{|C_k|}+\frac{G_k^2}{|C_k|})\xrightarrow{P}1$ as $N\to\infty$ by (A9)–(A12), which gives the result. ∎

B1. Additional Simulation Results

TABLE B.1: Empirical power based on 1,000 trials with the nominal level of 5% for 2-dimensional DGP's with different scale shifts and sample size $N = 25$.

DGP	Methods	$\sigma = 1.25$	$\sigma = 1.5$	$\sigma = 1.75$	$\sigma = 2$	$\sigma = 2.25$	$\sigma = 2.5$	$\sigma = 2.75$	$\sigma = 3$
	MHD	0.171	0.376	0.540	0.716	0.794	0.889	0.934	0.962
	MHD75	0.140	0.277	0.438	0.598	0.701	0.803	0.870	0.924
	MHD50	0.140	0.277	0.438	0.598	0.701	0.803	0.870	0.924
$N_p(\mathbf{0}, \mathbf{I})$	**SPD**	0.163	0.385	0.574	**0.744**	**0.830**	**0.908**	**0.947**	**0.969**
	HSD	**0.218**	**0.408**	**0.579**	0.741	0.807	0.880	0.918	0.954
	SLD	0.189	0.373	0.542	0.706	0.769	0.860	0.908	0.942
	OD	0.173	0.379	0.549	0.723	0.813	0.901	0.939	0.968
	MCUSUM	0.010	0.023	0.058	0.100	0.153	0.191	0.223	0.252
	MHD	0.097	0.131	0.161	0.194	0.250	0.324	0.360	0.391
	MHD75	0.090	0.144	0.194	0.283	0.352	0.434	0.468	0.532
	MHD50	0.090	0.144	0.194	0.283	0.352	0.434	0.468	0.532
$C_p(\mathbf{0}, \mathbf{I})$	**SPD**	0.104	0.149	0.213	0.302	0.372	0.445	0.488	0.571
	HSD	**0.123**	**0.170**	**0.242**	**0.330**	**0.378**	**0.455**	**0.505**	**0.579**
	SLD	0.116	0.156	0.216	0.305	0.349	0.429	0.473	0.551
	OD	0.098	0.156	0.200	0.283	0.359	0.426	0.480	0.541
	MCUSUM	0.000	0.002	0.002	0.002	0.003	0.003	0.007	0.007
	MHD	0.152	0.307	0.432	0.559	0.664	0.753	0.789	0.837
	MHD75	0.133	0.256	0.356	0.490	0.624	0.710	0.750	0.808
	MHD50	0.133	0.256	0.356	0.490	0.624	0.710	0.750	0.808
$U_p(1)$	**SPD**	0.165	0.321	**0.449**	0.554	0.684	**0.775**	**0.814**	**0.852**
	HSD	**0.190**	**0.337**	0.446	**0.563**	0.672	0.739	0.785	0.824
	SLD	0.177	0.316	0.431	0.514	0.651	0.707	0.748	0.804
	OD	0.160	0.304	0.433	0.549	**0.685**	0.759	0.801	0.843
	MCUSUM	0.011	0.025	0.055	0.098	0.139	0.140	0.185	0.238
	MHD	0.128	0.261	0.412	0.581	0.638	0.692	0.751	0.799
	MHD75	0.138	0.285	0.448	0.600	0.739	0.824	**0.886**	0.931
	MHD50	0.138	0.285	0.448	0.600	0.739	0.824	**0.886**	0.931
$SN_p(\mathbf{0}, \mathbf{I}, 3\mathbf{1})$	**SPD**	0.141	**0.312**	**0.475**	0.637	**0.747**	**0.826**	0.885	**0.933**
	HSD	**0.162**	0.267	0.383	0.501	0.556	0.587	0.624	0.664
	SLD	0.142	0.245	0.355	0.480	0.525	0.548	0.608	0.637
	OD	0.147	0.292	0.457	**0.641**	0.736	0.810	0.881	0.925
	MCUSUM	0.005	0.010	0.023	0.032	0.052	0.082	0.096	0.109
	MHD	0.133	0.250	0.411	0.546	0.637	0.712	0.727	0.786
	MHD75	0.152	0.269	0.454	0.622	0.747	**0.833**	0.873	0.916
	MHD50	0.152	0.269	0.453	0.622	0.747	**0.833**	0.873	0.916
$SN_p(\mathbf{0}, \mathbf{I}, 10\mathbf{1})$	**SPD**	**0.171**	**0.287**	**0.469**	**0.649**	**0.749**	0.832	**0.876**	**0.923**
	HSD	0.163	0.230	0.354	0.429	0.512	0.519	0.533	0.559
	SLD	0.147	0.203	0.326	0.401	0.474	0.489	0.516	0.539
	OD	0.155	0.271	0.461	0.633	0.736	0.828	0.863	0.922
	MCUSUM	0.008	0.011	0.028	0.033	0.043	0.070	0.086	0.092
	MHD	0.115	0.294	0.421	0.550	0.666	0.748	0.778	0.865
	MHD75	0.106	0.259	0.381	0.508	0.631	0.700	0.787	0.862
	MHD50	0.106	0.259	0.381	0.508	0.631	0.700	0.787	0.862
$ST_p(\mathbf{0}, \mathbf{I}, 5\mathbf{1}, 4)$	**SPD**	0.119	0.301	0.440	0.585	**0.711**	**0.783**	**0.841**	**0.906**
	HSD	**0.151**	**0.326**	**0.473**	**0.599**	0.692	0.767	0.822	0.878
	SLD	0.146	0.310	0.442	0.564	0.667	0.747	0.790	0.859
	OD	0.112	0.292	0.441	0.568	0.696	0.775	0.831	0.891
	MCUSUM	0.002	0.006	0.008	0.019	0.021	0.034	0.043	0.042

TABLE B.2: Empirical power based on 1,000 trials with the nominal level of 5% for 2-dimensional DGP's with different scale shifts and sample size $N = 100$.

DGP	Methods	$\sigma = 1.25$	$\sigma = 1.5$	$\sigma = 1.75$	$\sigma = 2$	$\sigma = 2.25$	$\sigma = 2.5$	$\sigma = 2.75$	$\sigma = 3$
	MHD	0.431	0.896	0.990	**1**	**1**	0.999	**1**	**1**
	MHD75	0.362	0.830	0.977	0.999	**1**	**1**	**1**	**1**
	MHD50	0.366	0.829	0.977	0.999	**1**	**1**	**1**	**1**
$N_p(\mathbf{0}, \mathbf{I})$	SPD	**0.438**	**0.898**	0.989	**1**	**1**	**1**	**1**	**1**
	HSD	0.426	0.883	**0.991**	**1**	**1**	**1**	**1**	**1**
	SLD	0.416	0.873	0.989	**1**	**1**	**1**	**1**	**1**
	OD	0.431	0.896	0.990	**1**	**1**	**1**	**1**	**1**
	MCUSUM	0.203	0.743	0.970	0.999	**1**	0.803	**1**	**1**
	MHD	0.106	0.251	0.410	0.545	0.654	0.734	0.787	0.836
	MHD75	0.152	0.388	0.603	0.763	0.886	0.949	0.974	**0.988**
	MHD50	0.151	0.387	0.599	0.763	0.886	0.950	0.974	**0.988**
$C_p(\mathbf{0}, \mathbf{I})$	SPD	**0.162**	**0.401**	**0.611**	**0.770**	**0.902**	**0.953**	**0.983**	**0.988**
	HSD	0.157	0.381	0.589	0.756	0.887	0.948	0.979	0.983
	SLD	0.155	0.383	0.593	0.761	0.879	0.950	0.981	0.984
	OD	0.149	0.379	0.594	0.757	0.890	0.948	0.978	0.985
	MCUSUM	0.002	0.005	0.007	0.009	0.009	0.017	0.012	0.021
	MHD	0.362	0.779	0.953	0.994	0.998	**1**	**1**	**1**
	MHD75	0.295	0.708	0.921	0.987	0.997	**1**	0.999	**1**
	MHD50	0.295	0.707	0.919	0.988	0.997	**1**	0.999	**1**
$U_p(1)$	SPD	**0.371**	0.781	0.955	0.993	0.997	**1**	**1**	**1**
	HSD	0.348	0.756	0.947	0.993	0.997	**1**	0.999	**1**
	SLD	0.337	0.752	0.946	0.990	0.997	**1**	0.999	**1**
	OD	0.367	0.776	0.953	0.992	0.997	**1**	**1**	**1**
	MCUSUM	0.279	**0.812**	**0.980**	**0.999**	**1**	**1**	**1**	**1**
	MHD	0.382	0.785	0.965	0.991	**1**	**1**	**1**	**1**
	MHD75	0.399	0.839	0.977	**0.999**	**1**	**1**	**1**	**1**
	MHD50	0.401	**0.840**	0.977	**0.999**	**1**	**1**	**1**	**1**
$SN_p(\mathbf{0}, \mathbf{I}, 3\mathbf{1})$	SPD	**0.404**	0.822	0.979	0.997	**1**	**1**	**1**	**1**
	HSD	0.344	0.693	0.918	0.961	0.995	0.996	0.999	**1**
	SLD	0.329	0.678	0.916	0.960	0.994	0.995	0.999	**1**
	OD	0.403	0.824	**0.982**	0.996	**1**	**1**	**1**	**1**
	MCUSUM	0.164	0.594	0.905	0.977	0.998	0.999	0.999	0.998
	MHD	0.331	0.743	0.947	0.990	0.998	**1**	0.999	**1**
	MHD75	0.343	**0.813**	**0.971**	**0.999**	**1**	**1**	**1**	**1**
	MHD50	0.345	**0.813**	**0.971**	**0.999**	**1**	**1**	**1**	**1**
$SN_p(\mathbf{0}, \mathbf{I}, 10\mathbf{1})$	SPD	0.335	0.799	0.968	0.996	**1**	**1**	**1**	**1**
	HSD	0.258	0.618	0.833	0.932	0.965	0.986	0.991	0.994
	SLD	0.252	0.610	0.829	0.933	0.960	0.986	0.989	0.995
	OD	**0.358**	0.805	**0.971**	0.998	**1**	**1**	**1**	**1**
	MCUSUM	0.159	0.565	0.894	0.971	0.992	**1**	0.998	0.999
	MHD	0.213	0.512	0.717	0.878	0.937	0.976	0.983	0.995
	MHD75	**0.306**	**0.692**	**0.912**	**0.987**	**0.998**	**1**	**1**	**1**
	MHD50	0.305	0.688	**0.912**	**0.987**	**0.998**	**1**	**1**	**1**
$ST_p(\mathbf{0}, \mathbf{I}, 5\mathbf{1}, 4)$	SPD	0.266	0.640	0.880	0.967	0.988	**1**	0.999	**1**
	HSD	0.160	0.400	0.615	0.745	0.850	0.916	0.943	0.968
	SLD	0.156	0.398	0.603	0.735	0.837	0.906	0.939	0.965
	OD	0.264	0.632	0.873	0.965	0.986	**1**	0.999	**1**
	MCUSUM	0.041	0.114	0.265	0.430	0.595	0.707	0.764	0.803

TABLE B.3: Empirical power based on 1,000 trials with the nominal level of 5% for 2-dimensional DGP's with different scale shifts and sample size $N = 200$.

DGP	Methods	$\sigma = 1.25$	$\sigma = 1.5$	$\sigma = 1.75$	$\sigma = 2$	$\sigma = 2.25$	$\sigma = 2.5$	$\sigma = 2.75$	$\sigma = 3$
	MHD	0.692	**0.997**	**1**	**1**	**1**	**1**	**1**	**1**
	MHD75	0.642	0.991	**1**	**1**	**1**	**1**	**1**	**1**
	MHD50	0.640	0.991	**1**	**1**	**1**	**1**	**1**	**1**
$N_p(\mathbf{0}, \mathbf{I})$	SPD	**0.693**	**0.997**	**1**	**1**	**1**	**1**	**1**	**1**
	HSD	0.680	0.996	**1**	**1**	**1**	**1**	**1**	**1**
	SLD	0.683	0.996	**1**	**1**	**1**	**1**	**1**	**1**
	OD	**0.693**	**0.997**	**1**	**1**	**1**	**1**	**1**	**1**
	MCUSUM	0.564	0.994	**1**	**1**	**1**	**1**	**1**	**1**
	MHD	0.175	0.423	0.628	0.781	0.849	0.910	0.935	0.918
	MHD75	0.240	0.641	0.861	0.969	**0.993**	**1**	**1**	**1**
	MHD50	0.240	0.641	0.860	0.969	**0.993**	**1**	**1**	**1**
$C_p(\mathbf{0}, \mathbf{I})$	SPD	0.243	**0.658**	**0.880**	**0.971**	**0.993**	**1**	**1**	**1**
	HSD	0.246	0.648	0.877	0.966	**0.993**	**1**	**1**	**1**
	SLD	0.247	0.648	0.877	0.967	0.992	**1**	**1**	**1**
	OD	**0.250**	0.649	0.870	0.967	**0.993**	**1**	**1**	**1**
	MCUSUM	0.002	0.002	0.005	0.016	0.007	0.020	0.022	0.027
	MHD	0.637	0.977	**0.999**	**1**	**1**	**1**	**1**	**1**
	MHD75	0.558	0.955	**0.999**	**1**	**1**	**1**	**1**	**1**
	MHD50	0.557	0.954	**0.999**	**1**	**1**	**1**	**1**	**1**
$U_p(1)$	SPD	0.632	0.980	**0.999**	**1**	**1**	**1**	**1**	**1**
	HSD	0.617	0.976	**0.999**	**1**	**1**	**1**	**1**	**1**
	SLD	0.605	0.976	**0.999**	**1**	**1**	**1**	**1**	**1**
	OD	0.633	0.978	**0.999**	**1**	**1**	**1**	**1**	**1**
	MCUSUM	**0.690**	**0.999**	0.998	**1**	**1**	**1**	**1**	**1**
	MHD	0.578	0.981	**1**	**1**	**1**	**1**	**1**	**1**
	MHD75	**0.638**	0.993	**1**	**1**	**1**	**1**	**1**	**1**
	MHD50	0.636	**0.994**	**1**	**1**	**1**	**1**	**1**	**1**
$SN_p(\mathbf{0}, \mathbf{I}, 3\mathbf{1})$	SPD	0.605	0.988	**1**	**1**	**1**	**1**	**1**	**1**
	HSD	0.513	0.958	0.998	**1**	**1**	**1**	**1**	**1**
	SLD	0.511	0.954	0.998	**1**	**1**	**1**	**1**	**1**
	OD	0.613	0.989	**1**	**1**	**1**	**1**	**1**	**1**
	MCUSUM	0.503	0.982	**1**	**1**	**1**	**1**	**1**	**1**
	MHD	0.538	0.964	0.999	**1**	**1**	**1**	**1**	**1**
	MHD75	**0.653**	**0.992**	**1**	**1**	**1**	**1**	**1**	**1**
	MHD50	**0.653**	**0.992**	**1**	**1**	**1**	**1**	**1**	**1**
$SN_p(\mathbf{0}, \mathbf{I}, 10\mathbf{1})$	SPD	0.588	0.980	**1**	**1**	**1**	**1**	**1**	**1**
	HSD	0.440	0.900	0.988	**1**	**1**	**1**	**1**	**1**
	SLD	0.426	0.886	0.988	**1**	**1**	**1**	**1**	**1**
	OD	0.590	0.985	**1**	**1**	**1**	**1**	**1**	**1**
	MCUSUM	0.471	0.973	**1**	**1**	**1**	**1**	**1**	**1**
	MHD	0.295	0.731	0.947	0.994	0.998	**1**	**1**	**1**
	MHD75	**0.465**	**0.935**	**0.997**	**1**	**1**	**1**	**1**	**1**
	MHD50	**0.465**	0.934	0.996	**1**	**1**	**1**	**1**	**1**
$ST_p(\mathbf{0}, \mathbf{I}, 5\mathbf{1}, 4)$	SPD	0.398	0.864	0.992	**1**	**1**	**1**	**1**	**1**
	HSD	0.240	0.626	0.883	0.970	0.990	0.999	0.998	**1**
	SLD	0.224	0.614	0.867	0.959	0.987	0.998	0.998	**1**
	OD	0.391	0.859	0.988	0.999	**1**	**1**	**1**	**1**
	MCUSUM	0.091	0.342	0.665	0.830	0.920	0.955	0.946	0.972

TABLE B.4: Empirical power based on 1,000 trials with the nominal level of 5% for $p = 5, 10$ and 20 dimensional DGP's, sample sizes ranging from $N = 25$ to $N = 200$, and a scale shift of 2.

		$N = 25$			$N = 100$			$N = 200$		
DGP	Methods	$p = 5$	$p = 10$	$p = 20$	$p = 5$	$p = 10$	$p = 20$	$p = 5$	$p = 10$	$p = 20$
	MHD	0.968	0.997	0.935	**1**	**1**	**1**	**1**	**1**	**1**
	MHD75	0.871	0.936	0.668	**1**	**1**	**1**	**1**	**1**	**1**
$N_p(\mathbf{0}, \mathbf{I})$	MHD50	0.868	0.948	0.668	**1**	**1**	**1**	**1**	**1**	**1**
	SPD	**0.994**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
	MCUSUM	0.000	*	*	**1**	0.863	*	**1**	**1**	*
	MHD	0.288	0.360	0.347	0.815	0.889	0.910	0.967	0.995	0.994
	MHD75	0.304	0.381	0.291	0.878	0.912	0.914	**0.994**	0.998	0.995
$C_p(\mathbf{0}, \mathbf{I})$	MHD50	0.304	0.367	0.291	0.873	0.910	**0.928**	0.993	0.998	**0.996**
	SPD	**0.364**	**0.407**	**0.417**	**0.886**	**0.914**	0.927	**0.994**	**0.999**	**0.996**
	MCUSUM	0.000	*	*	0.002	0.001	*	0.006	0.007	*
	MHD	0.546	0.533	0.457	0.993	0.989	0.989	**1**	**1**	**1**
	MHD75	0.498	0.487	0.372	0.993	0.987	0.982	**1**	**1**	**1**
$U_p(1)$	MHD50	0.498	0.501	0.372	0.993	0.987	0.987	**1**	**1**	**1**
	SPD	**0.585**	**0.595**	**0.602**	**0.996**	0.989	**0.991**	**1**	**1**	**1**
	MCUSUM	0.000	*	*	0.617	0.000	*	**1**	0.421	*
	MHD	0.961	0.993	0.944	**1**	**1**	**1**	**1**	**1**	**1**
	MHD75	0.869	0.936	0.672	**1**	**1**	**1**	**1**	**1**	**1**
$SN_p(\mathbf{0}, \mathbf{I}, 3\mathbf{1})$	MHD50	0.873	0.942	0.672	**1**	**1**	**1**	**1**	**1**	**1**
	SPD	**0.981**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
	MCUSUM	0.000	*	*	**1**	0.823	*	**1**	**1**	*
	MHD	0.941	0.997	0.947	**1**	**1**	**1**	**1**	**1**	**1**
	MHD75	0.853	0.951	0.698	**1**	**1**	**1**	**1**	**1**	**1**
$SN_p(\mathbf{0}, \mathbf{I}, 10\mathbf{1})$	MHD50	0.852	0.940	0.698	**1**	**1**	**1**	**1**	**1**	**1**
	SPD	**0.981**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
	MCUSUM	0.000	*	*	**1**	0.839	*	**1**	**1**	*
	MHD	0.321	0.316	0.744	**0.883**	**1**	**1**	**1**	**1**	**1**
	MHD75	**0.329**	0.320	0.563	0.882	**1**	**1**	**1**	**1**	**1**
$ST_p(\mathbf{0}, \mathbf{I}, 5\mathbf{1}, 4)$	MHD50	**0.329**	0.320	0.563	0.882	**1**	**1**	**1**	**1**	**1**
	SPD	0.327	**0.324**	**0.936**	0.881	**1**	**1**	**1**	**1**	**1**
	MCUSUM	0.000	*	*	0.489	0.052	*	0.963	0.924	*

* The test is not feasible in that dimension.

TABLE B.5: Empirical power based on 1,000 trials with the nominal level of 5% for $p = 5$, 10 and 20 dimensional DGP's, sample sizes ranging from $N = 25$ to $N = 200$, and a scale shift of 3.

		$N = 25$			$N = 100$			$N = 200$		
DGP	Methods	$p = 5$	$p = 10$	$p = 20$	$p = 5$	$p = 10$	$p = 20$	$p = 5$	$p = 10$	$p = 20$
	MHD	0.999	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
	MHD75	0.997	**1**	0.957	**1**	**1**	**1**	**1**	**1**	**1**
$N_p(\mathbf{0}, \mathbf{I})$	MHD50	0.997	**1**	0.957	**1**	**1**	**1**	**1**	**1**	**1**
	SPD	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
	MCUSUM	0.000	*	*	**1**	0.999	*	**1**	**1**	*
	MHD	0.601	0.664	0.639	0.997	**1**	0.986	**1**	**1**	**1**
	MHD75	0.638	0.634	0.562	0.997	**1**	**0.999**	**1**	**1**	**1**
$C_p(\mathbf{0}, \mathbf{I})$	MHD50	0.639	0.639	0.562	0.997	**1**	**0.999**	**1**	**1**	**1**
	SPD	**0.692**	**0.712**	**0.729**	**0.998**	**1**	**0.999**	**1**	**1**	**1**
	MCUSUM	0.000	*	*	0.025	0.010	*	0.037	0.117	*
	MHD	0.835	0.815	0.749	**1**	**1**	**1**	**1**	**1**	**1**
	MHD75	0.815	0.792	0.662	**1**	**1**	**1**	**1**	**1**	**1**
$U_p(1)$	MHD50	0.813	0.787	0.662	**1**	**1**	**1**	**1**	**1**	**1**
	SPD	**0.863**	**0.847**	**0.864**	**1**	**1**	**1**	**1**	**1**	**1**
	MCUSUM	0.000	*	*	0.970	0.006	*	**1**	0.963	*
	MHD	**1**	**1**	0.999	**1**	**1**	**1**	**1**	**1**	**1**
	MHD75	**1**	**1**	0.955	**1**	**1**	**1**	**1**	**1**	**1**
$SN_p(\mathbf{0}, \mathbf{I}, 3\mathbf{1})$	MHD50	**1**	**1**	0.955	**1**	**1**	**1**	**1**	**1**	**1**
	SPD	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
	MCUSUM	0.000	*	*	**1**	**1**	*	**1**	**1**	*
	MHD	**1**	**1**	0.999	**1**	**1**	**1**	**1**	**1**	**1**
	MHD75	0.998	0.998	0.952	**1**	**1**	**1**	**1**	**1**	**1**
$SN_p(\mathbf{0}, \mathbf{I}, 10\mathbf{1})$	MHD50	0.998	0.999	0.952	**1**	**1**	**1**	**1**	**1**	**1**
	SPD	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
	MCUSUM	0.000	*	*	**1**	**1**	*	**1**	**1**	*
	MHD	0.586	0.617	0.981	**1**	**1**	**1**	**1**	**1**	**1**
	MHD75	**0.652**	0.672	0.896	**1**	**1**	**1**	**1**	**1**	**1**
$ST_p(\mathbf{0}, \mathbf{I}, 5\mathbf{1}, 4)$	MHD50	**0.652**	0.672	0.896	**1**	**1**	**1**	**1**	**1**	**1**
	SPD	0.640	**0.685**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
	MCUSUM	0.000	*	*	0.954	0.668	*	**1**	**1**	*

* The test is not feasible in that dimension.

TABLE B.6: Comparison of MCUSUM and MCUSUM* on 1,000 trials with the nominal level of 5% for $N_2(\mathbf{0}, \mathbf{I})$ with different scale shifts, and sample sizes $N = 25,\ 100$, and 200.

Methods	$\sigma = 1$	$\sigma = 1.25$	$\sigma = 1.5$	$\sigma = 1.75$	$\sigma = 2$	$\sigma = 2.25$	$\sigma = 2.5$	$\sigma = 2.75$	$\sigma = 3$
					$N = 25$				
MCUSUM	0.002	0.010	0.023	0.058	0.100	0.153	0.191	0.223	0.252
MCUSUM*	0.052	0.127	0.293	0.514	0.699	0.808	0.898	0.954	0.961
					$N = 100$				
MCUSUM	0.022	0.203	0.743	0.970	0.999	1	0.803	1	1
MCUSUM*	0.064	0.353	0.866	0.994	1	1	1	1	1
					$N = 200$				
MCUSUM	0.047	0.564	0.994	1	1	1	1	1	1
MCUSUM*	0.042	0.659	0.995	1	1	1	1	1	1

B2. Nonmonotonic power of the MCUSUM method

One possibility that might explain the poor performance of the MCUSUM method even for normal data could be the inaccuracy of estimated covariance of vech($\mathbf{X}_i\mathbf{X}_i^\top$) under $H_{A,\theta}$. To examine this, we implemented an idealized version, denoted by MCUSUM*, which replaces the estimator of the covariance of vech($\mathbf{X}_i\mathbf{X}_i^\top$) with its theoretical value calculated under the known distribution of the data in the Gaussian case. A comparison between MCUSUM and MCUSUM* is provided in Table B.6, in which we observe that this adjustment seems to explain both why the MCUSUM test tends to be undersized, and why it underperforms the depth based tests even under Gaussianity of the underlying data generating process.

B3. Autocorrelation plots of first differenced acid rain series

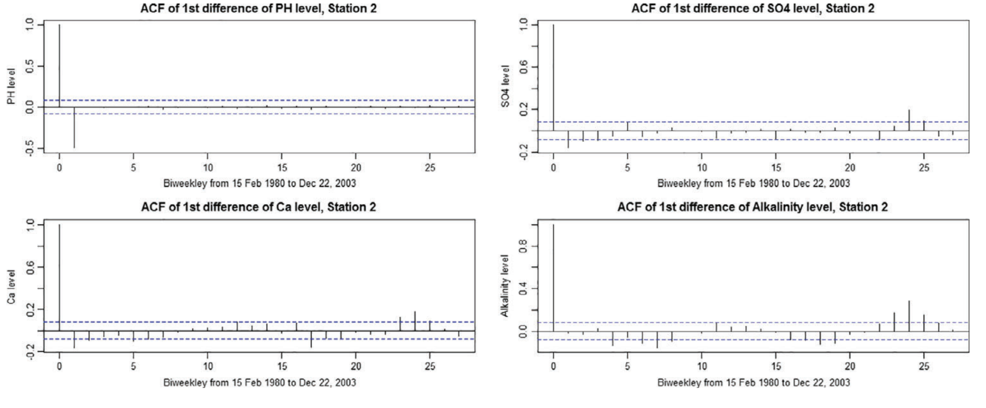

FIGURE B.1: ACF plots of first differenced pH, SO_4, Ca and alkalinity series.

Received 17 March 2018
Accepted 12 May 2019