

# Change-Point Detection for General Nonparametric Regression Models

Murray D. Burke\*, Gildas Bewa

Department of Mathematics and Statistics, University of Calgary, Calgary, Canada

Email: \*burke@math.ucalgary.ca

Received May 1, 2013; revised June 15, 2013; accepted June 22, 2013

Copyright © 2013 Murray D. Burke, Gildas Bewa. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## ABSTRACT

A number of statistical tests are proposed for the purpose of change-point detection in a general nonparametric regression model under mild conditions. New proofs are given to prove the weak convergence of the underlying processes which assume remove the stringent condition of bounded total variation of the regression function and need only second moments. Since many quantities, such as the regression function, the distribution of the covariates and the distribution of the errors, are unspecified, the results are not distribution-free. A weighted bootstrap approach is proposed to approximate the limiting distributions. Results of a simulation study for this paper show good performance for moderate samples sizes.

**Keywords:** Change-Point Detection; Nonparametric Regression Models; Weighted Bootstrap

## 1. Introduction

Model building is a very important task. When considering observations  $\{(X_i, Y_i) : i = 1, 2, \dots\}$  taken over time we may wish to consider the relationship between the  $X_i$  and the  $Y_i$  by some regression model, either parametric or semiparametric, linear or nonlinear. However, if, after some point  $i = k$ , there is a change in the relationship between  $X_i$  and  $Y_i$ , then a single regression model will be inappropriate and would fit the data poorly. For example, in **Figure 1** below, the scatterplot on the left is that of a combined sample of 150 points. The middle scatterplot is that of the first 60 points while the right plot is that of the remaining 90 points. Without realizing that there was a change-point, one might wrongly propose a single regression model, perhaps linear, with heteroscedastic errors. Thus, it is important to determine whether such a change has occurred before the particular form of the model is postulated. If such a change has occurred, then two regression models should be used, one for the first  $k$  pairs and another for the remaining pairs. We will propose a number of tests designed to detect whether a change has occurred in a general nonparametric model. Change-point problems arise in many areas when observations  $(X_i, Y_i)$  are taken over time  $i$ .

For example, the height of (or discharge from) a river as a function of upstream precipitation, or concentrations

\*Corresponding author.

of heavy metals in storm-water runoff as a function of suspended sediment size. We refer to the book by Csörgő and Horváth [1] for more applications and important references and to the papers [2-4] for other change-point problems.

We say that the pair  $(X, Y)$  satisfies general regression model  $M(\phi, \psi, F, \varepsilon)$  if

$$Y = \phi(X) + \psi(X)\varepsilon, P[X \leq x] = F(x),$$

for all  $x \in R^d$ , where  $\varepsilon$  is a mean zero, variance one random variable, independent of  $X$ . We write  $(X, Y) \in M(\phi, \psi, F, \varepsilon)$ .

For a sequence  $(X_1, Y_1), \dots, (X_n, Y_n)$ , consider the null hypotheses:

$$(X_i, Y_i) \in M(\phi, \psi, F, \varepsilon), i = 1, \dots, n, \quad (1)$$

where  $\phi, \psi \geq 0$ , and  $F$  are unspecified, the  $X_i$  are independent and the  $\varepsilon_i$  are independent and identically distributed. Model (1) encompasses both linear and nonlinear regression models and the case of heteroscedastic errors. We wish to test whether there is a change in the model as the observations are taken, without specifying  $\phi, \psi$  or  $F$ , that is to test the null Hypothesis (1) versus the alternative:

$$\begin{aligned} (X_i, Y_i) \in M(\phi, \psi, F, \varepsilon), 1 \leq i \leq [n\theta], \\ (X_i, Y_i) \in M(\phi^1, \psi, F, \varepsilon), [n\theta] + 1 \leq i \leq n, \end{aligned} \quad (2)$$

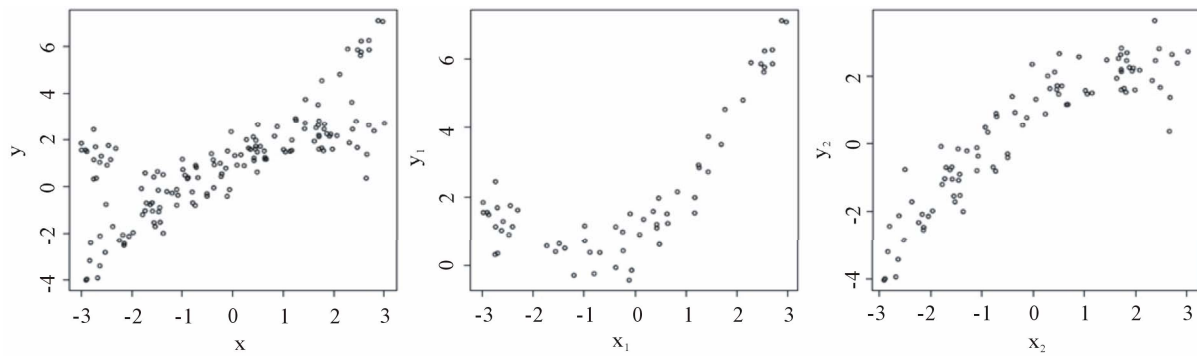


Figure 1. Scatterplot of  $(X_i, Y_i)$ .

for some  $\theta \in (0, 1)$ , where  $[s]$  denotes the integer part of  $s$ .

$$\int_{D(x)} \phi dF \neq \int_{D(x)} \phi^1 dF^1; D(x) = \{v \in R^d : v \leq x\}, \quad (3)$$

for some  $x \in R^d$ , using the usual partial order in  $R^d$ .

The alternative Hypothesis (2) includes the case where only  $\phi$  changes, that is,  $\phi \neq \phi^1, F = F^1$ , for some  $\theta \in (0, 1)$ , and when only  $F$  changes  $F \neq F^1, \phi = \phi^1$ , for some  $\theta \in (0, 1)$ .

We will base our tests on the process

$$\begin{aligned} \beta_n(x, s) &= \frac{1}{\sqrt{n}} \left( \sum_{i=1}^{[ns]} (Y_i I_{D(x)}(X_i) - \overline{Y I_{D(x)}(X)})_n \right) \\ &= \frac{1}{\sqrt{n}} \left( \sum_{i=1}^{[ns]} \left( Y_i I_{D(x)}(X_i) - \frac{[ns]}{n} \sum_{i=1}^n Y_i I_{D(x)}(X_i) \right) \right) \\ n^{-1} \leq s \leq 1 - n^{-1}, x \in R^d, \end{aligned}$$

and zero otherwise, where  $I_{D(x)}(X_i)$  is the indicator of the event

$$[X_i \in D(x)] = [X_i \leq x]$$

and

$$\overline{Y I_{D(x)}(X)}_n = n^{-1} \sum_{i=1}^n Y_i I_{D(x)}(X_i).$$

$\beta_n$  can also be written as the difference of two means

$$\sqrt{n} w \left( \frac{[ns]}{n} \right) \left( \frac{\sum_{i=1}^{[ns]} Y_i I_{D(x)}(X_i)}{[ns]} - \frac{\sum_{i=[ns]+1}^n Y_i I_{D(x)}(X_i)}{(n - [ns])} \right) \quad (4)$$

where

$$w(u) = (u(1-u))^{1/2}.$$

Note that

$$E Y I_{[X \in D(x)]} = \int_{D(x)} \phi dF.$$

We are considering the partial sums of the response variables  $Y_i$  aligned according to the  $X_i$ -observations. The advantage of  $\beta_n$  is that we can use empirical process the-

ory which has much better convergence properties, rather than trying to estimate the function  $\phi$  directly with kernel or other curve estimation methods which require a large sample size. As will be demonstrated in the simulation study, our approach works well for moderate sample sizes.

In the next section, we state the main results related to the weak convergence of the above process as well as convergence of various related statistics. These results do not assume the stringent condition of bounded variation for the regression function  $\phi$  and the variance function  $\psi$  and strong moment conditions on the error distribution, which were assumed in an earlier paper [5]. Section 3 states the asymptotic results for the weighted bootstrap version of our statistics. This is needed since the asymptotic distributions of our statistics depend on unspecified functions. A new simulation study is conducted in Section 4. Proofs of the results (different from those in [5,6]) are sketched in Section 5.

## 2. Main Results

Assume the following:

(A1) The sequences  $\{X_i\}$  and  $\{\varepsilon_i\}$ ,  $i = 1, 2, \dots$  are independent.

(A2) The  $\{X_i\}$  are independent and identically distributed random vectors with distribution function  $F$ .

(A3) The functions  $\phi, \psi : R^d \rightarrow R$ , with  $\psi \geq 0$ , satisfy

$$\int \phi^2 dF, \int \psi^2 dF < \infty.$$

(A4) The  $\{\varepsilon_i: 1 \leq i\}$  are independent and identically distributed with  $E(\{\varepsilon_i\}) = 0$  and  $E\varepsilon^2 = 1$ .

(A5)

$$\int_0^\infty \sqrt{P[|\varepsilon| > t]} dt < \infty.$$

Conditions (A1)-(A5) are much weaker than those in [5] and [6] where  $\phi$  and  $\psi$  needed to be of bounded total variation and  $\varepsilon_i$  required much higher moment conditions ( $E|\varepsilon_i|^p < \infty$ ). While the conclusions in [5] and [6] may be stronger (weighted approximations), the requirement of bounded variation for  $\phi$  and  $\psi$  limits their applicability.

Here we dispense with this requirement. For example, we permit the case where the  $X_i$  have unbounded support and the regression relationship is polynomial (unbounded), provided  $X_i$  has finite moments of a suitable degree.

For

$$x_i = (x_{i1}, \dots, x_{id}), i = 1, 2, \dots,$$

let

$$x_1 \wedge x_2 = (x_{11} \wedge x_{21}, \dots, x_{1d} \wedge x_{2d}).$$

The first main result is:

**Theorem 1** Under Assumptions (A1)-(A5) and the null Hypothesis (1),

$$\beta_n \rightarrow_D \Gamma, \text{ as } n \rightarrow \infty,$$

in the space  $D(R^d \times [0, 1])$ , equipped with the uniform norm, where  $\Gamma(x, s)$  is mean-zero Gaussian process with covariance function

$$\begin{aligned} &E\Gamma(x_1, s_1)\Gamma(x_1, s_1) \\ &= (s_1 \wedge s_2 - s_1s_2) \\ &\times (H_1(x_1 \wedge x_2) - G_1(x_1)G_1(x_2) + H_2(x_1 \wedge x_2)), \end{aligned}$$

and where

$$\begin{aligned} G_1(x) &= \int_{D(x)} \phi dF; \quad H_1(x) = \int_{D(x)} \phi^2 dF; \\ H_2(x) &= \int_{D(x)} \psi^2 dF. \end{aligned}$$

**Corollary 2** Under the conditions of Theorem 1 and the null Hypothesis (1),

$$\begin{aligned} KS_n &\rightarrow_D \sup_{x \in R^d} \sup_{0 < s < 1} |\Gamma(x, s)|; \\ MC_n &\rightarrow_D \int_0^1 \int_{R^d} (\Gamma(x, s))^2 dF ds; \quad (5) \\ KC_n &\rightarrow_D \sup_{0 < s < 1} \int_{R^d} (\Gamma(x, s))^2 dF, \end{aligned}$$

where

$$\begin{aligned} KS_n &= \sup_{x \in R^d} \sup_{0 < s < 1} |\beta_n(x, s)|; \\ MC_n &= \int_0^1 \int_{R^d} (\beta_n(x, s))^2 dF ds; \quad (6) \\ KC_n &= \sup_{0 < s < 1} \int_{R^d} (\beta_n(x, s))^2 dF. \end{aligned}$$

Corollary 2 follows immediately from Theorem 1.

**Corollary 3** Under Assumptions (A1)-(A5) and (A3) for  $\phi^1$  and  $F^1$ , the test statistics defined by (6) are consistent against the alternative Hypothesis (2).

Under the null Hypothesis (1),  $\beta_n(x, s)$  has mean zero for all  $x$  and  $s$ . The statistic  $KS_n$  in Equation (6) is a Kolmogorov-Smirnov type statistic expressing the largest deviation between the two means (before  $[ns]$  and after  $[ns]$ ). Both statistics  $MC_n$  and  $KC_n$  are Cramér-von Mises type statistics which measures the squared-integral distance between the two means. The statistic  $MC_n$  is the mean (integral over  $s$ ) such distance, while

$KC_n$  is the supremum of the squared-integral distance.

Under the alternative Hypothesis (2), when  $s \leq \theta$ , the first part of Expression (4) has mean

$$\int_{D(x)} \phi dF,$$

while the term after the minus sign has mean

$$\int_{D(x)} \phi^1 dF^1,$$

when  $s \geq \theta$ . (see inequality (3)).

### 3. Weighted Bootstrap Approximations

Since the distribution functions of the  $X_i$  and  $\varepsilon_i$ , as well as the functional forms of  $\phi$  and  $\psi$ , are not specified, the limiting distributions of all our test statistics will not be known. A weighted bootstrap is proposed here which will approximate these asymptotic distributions.

Let

$$V_1^{(j)}, V_2^{(j)}, \dots, j = 1, 2, \dots, m,$$

be  $m$  independent sequences of independent and identically distributed random variables, independent of the  $\{\varepsilon_i\}$  and  $\{X_i\}$  sequences. We will consider the processes

$$\beta_n^{(j)}(x, s) = n^{-1/2} \left( \zeta^{(j)}(x, [ns]) - \frac{[ns]}{n} \zeta^{(j)}(x, n) \right),$$

for

$$n^{-1} \leq s \leq 1 - n^{-1}, x \in R^d,$$

zero otherwise, where  $j = 1, 2, \dots, m$ , and

$$\begin{aligned} &\zeta^{(j)}(x, k) \\ &= \left( \sum_{i=1}^k Y_i V_i^{(j)} I_{D(x)}(X_i) - \left\{ \sum_{\ell=1}^k V_\ell^{(j)} \right\} \sum_{i=1}^k \frac{Y_i I_{D(x)}(X_i)}{k} \right). \end{aligned}$$

**Theorem 4** Assume Conditions (A1)-(A5) hold and that (A4)-(A5) holds for the  $\{V_i^{(j)}\}$  sequences also. Then, under the null Hypothesis (1), along almost all sequences

$$\{(X_i, Y_i), i = 1, 2, \dots\},$$

given  $(X_i, Y_i), i = 1, 2, \dots, n$ ,  $\beta_n^{(j)} \rightarrow_D \Gamma$  in the space  $D(R^d \times [0, 1])$ , equipped with the uniform norm, for each  $j$ , where  $\Gamma$  is a mean-zero Gaussian process.

Under the conditions of Theorem 4, the same limiting distributions as found in Corollary 2 for statistics based on  $\beta_n$  are obtained for each  $\beta_n^{(j)}$ . Moreover, the statistics based on  $\beta_n$  and  $\beta_n^{(j)}, j = 1, 2, \dots, m$ , are asymptotically independent.

It is also important for the bootstrap process to have a limiting distribution under the alternative hypothesis. We obtain:

**Theorem 5** If the conditions of Theorem 4 hold for the functions  $\phi^1, F^1$ , then under the alternative Hypothesis

(2), along almost all sequences  $\{(X_i, Y_i), i = 1, 2, \dots\}$ , given  $(X_i, Y_i)$ ,  $i = 1, 2, \dots, n$ ,  $\beta_n^{(j)} \rightarrow_D \Gamma^{11}$  in the space  $D(R^d \times [0, 1])$ , equipped with the uniform norm, for each  $j$ , where  $\Gamma^{11}$  is a mean-zero Gaussian process.

### 4. Simulations

The proposed tests are omnibus tests in that we do not specify the form of the regression function  $\phi$ . In order to examine how these tests perform, we conducted simulations for two quite different functions,  $\phi(x) = 1 + x + \beta x^2$  and  $\phi(x) = 2 \sin(\beta x)$ . The  $X_i$  were generated from the uniform  $U(-3, 3)$  distribution and the errors  $\varepsilon_i$  were standard normals. In each case, 1000 simulations were carried out, each with  $m = 1000$  weighted bootstrap samples. The level of the tests was  $\alpha = 0.05$  throughout. Theorem 3 allows us to choose any mean-zero, variance-one distribution for the bootstrap weights  $V_i^{(j)}$ . We chose standard normal weights for the simulation since the limiting processes are Gaussian and normal bootstrap weights have been found to work well in simulations for goodness-of-fit types of statistics [7,8]. We use the percentile method. If  $h(\beta_n)$  is one of the statistics in Corollary 2, we reject the null hypothesis if  $h(\beta_n) > c_n(\alpha)$ , where  $c_n(\alpha)$  is the  $(1 - \alpha)100$ -th percentile of the values  $h(\beta_n^{(j)})$ ,  $j = 1, 2, \dots, m$ . Under the null hypothesis,  $c_n(\alpha) \rightarrow c(\alpha)$ , where  $c(\alpha)$  is the  $(1 - \alpha)100$ -th percentile of  $h(\Gamma)$  (Theorems 1 and 3). Under the alternative hypothesis,  $c_n(\alpha) \rightarrow c^{11}(\alpha)$ , with  $c^{11}(\alpha)$ , the  $(1 - \alpha)100$ -th percentile of  $h(\Gamma^{11})$  and  $h(\beta_n) \rightarrow_{p\infty}$  (Corollary 3 and Theorem 4).

The first power study is for the alternative:

$$H_1 : Y_i = 1 + X_i + 0.4X_i^2 + \psi\varepsilon_i, \quad 1 \leq i \leq 0.4n$$

$$Y_i = 1 + X_i + \beta X_i^2 + \psi\varepsilon_i, \quad 0.4n < i \leq n,$$

where  $\psi$  does not depend on  $X_i$  and the change-point is at  $0.4n$ . The statistics  $KS_n$ ,  $KC_n$  and  $MC_n$  are defined by Equation (6).

The first three rows correspond to the empirical level of the tests under the null hypothesis ( $\beta = 0.4$ ) with different choices for  $\psi$ , the standard deviation. The break-point  $\theta$  equals 0.4. The empirical level results are reasonably good especially when  $n \geq 50$ , that is they are close to the level of 0.05. The empirical powers of the tests are in the subsequent rows for different  $\beta$ .

As can be seen in **Table 1**, the power increases as  $\beta$  gets farther from 0.4 (the null Hypothesis). Understandably the power increases in  $n$  and decreases in  $\psi$  gets larger.

In the second simulation (see **Table 2**), we consider  $\phi$  to be a sine function. We conduct our simulations for sample sizes  $n = 50, 100, 150$ . The first row of Table 2 is the empirical level of the test ( $\beta = 1$ , no change-point). Again the empirical level is good when  $n \geq 50$ .

We take  $\psi = 1$ , but vary the location of the change-

point  $\theta n$ . The power increases as  $\beta$  gets farther from 1 (the null hypothesis). The power increases with the sample size  $n$ . Except in a few instances, the power is greatest when the change-point occurs in the middle ( $\theta = 0.5$ ) and decreases as  $\theta$  is farther from 0.5. The exceptional cases are likely due to simulation error.

Considering that our tests are omnibus in nature, the power results can be seen to be quite good.

### 5. Proofs of the Main Results

Proof of Theorem 1: For  $x \in R^d, 0 \leq s \leq 1$ , let

$$\alpha_n(x, s) = \sum_{i=1}^{[ns]} (Y_i I_{D(x)}(X_i) - G_1(x)) = \alpha_{1n} + \alpha_{2n},$$

where

$$\alpha_{1n}(x, s) = \sum_{i=1}^{[ns]} (\phi(X_i) I_{D(x)}(X_i) - G_1(x));$$

$$\alpha_{2n}(x, s) = \sum_{i=1}^{[ns]} \psi(X_i) \varepsilon_i I_{D(x)}(X_i).$$

The collection  $\{D(x) : x \in R^d\}$  is a Vapnik-Cervonenkis class of sets. Consequently, the collection of indicator functions

$$\{I_{D(x)} : x \in R^d\}$$

is a Donsker class of functions. For (fixed)  $\phi$  and  $\psi$ ,

$$\{\phi I_{D(x)} : x \in R^d\}$$

and

$$\{\psi I_{D(x)} : x \in R^d\}$$

are Donsker classes (Pollard [9]). By Giné and Zinn [10, 11] (see also Corollary 2.9.3 of [12]),

$$n^{-1/2} (\alpha_{1n}(x, 1), \sum_{i=1}^n \varepsilon_i (\psi I_{D(x)}(X_i) - G_2(x))),$$

where

$$G_2(x) = \int_{D(x)} \psi dF,$$

converges in distribution to a vector of tight independent Gaussian processes,  $\eta_1$

and  $\eta_2$ , each with zero mean and covariance ( $j = 1, 2$ ):

$$E\eta_j(x_1)\eta_j(x_2) = (H_j(x_1 \wedge x_2) - G_j(x_1)G_j(x_2))$$

Consequently,

$$n^{-1/2} (\alpha_{1n}(x, 1), \alpha_{2n}(x, 1))$$

converges in distribution to a vector of tight independent Gaussian processes  $(\eta_1, \eta_2^0)$  with  $E\eta_2^0(x) = 0$  and covariance

$$E\eta_2^0(x_1)\eta_2^0(x_2) = H(x_1 \wedge x_2).$$

**Table 1. Power when  $\phi(x) = 1 + x + 0.4x^2$ ,  $\phi^1(x) = 1 + x + \psi x^2$  and breakpoint  $\theta = 0.4$ .**

$\beta$	$\psi$	$n = 30$			$n = 50$			$n = 100$		
		$KS_n$	$KC_n$	$MC_n$	$KS_n$	$KC_n$	$MC_n$	$KS_n$	$KC_n$	$MC_n$
0.4	1	0.038	0.046	0.051	0.044	0.043	0.051	0.057	0.050	0.054
	2	0.040	0.047	0.053	0.047	0.060	0.063	0.052	0.049	0.048
	3	0.043	0.040	0.046	0.056	0.048	0.050	0.056	0.049	0.048
0.2	1	0.096	0.142	0.175	0.119	0.228	0.252	0.222	0.408	0.427
	2	0.086	0.057	0.076	0.088	0.088	0.097	0.121	0.136	0.139
	3	0.063	0.059	0.071	0.090	0.087	0.088	0.112	0.093	0.098
0.0	1	0.264	0.309	0.387	0.448	0.596	0.644	0.708	0.916	0.933
	2	0.159	0.120	0.163	0.257	0.261	0.293	0.454	0.488	0.510
	3	0.104	0.068	0.092	0.169	0.154	0.166	0.287	0.266	0.280
-0.2	1	0.465	0.383	0.517	0.748	0.768	0.807	0.974	0.986	0.988
	2	0.297	0.197	0.275	0.481	0.438	0.470	0.802	0.778	0.784
	3	0.202	0.137	0.167	0.300	0.241	0.269	0.539	0.485	0.502
-0.4	1	0.663	0.443	0.591	0.916	0.826	0.888	1.000	0.998	0.999
	2	0.474	0.301	0.396	0.698	0.549	0.614	0.950	0.920	0.925
	3	0.292	0.175	0.219	0.476	0.384	0.415	0.801	0.697	0.722

**Table 2. Power when  $\phi(x) = 2 \sin(x)$ ,  $\phi^1(x) = 2 \sin(\beta x)$ , and  $\psi = 1$  with different breakpoints  $\theta$ .**

$\beta$	$\theta$	$n = 50$			$n = 100$			$n = 150$		
		$KS_n$	$KC_n$	$MC_n$	$KS_n$	$KC_n$	$MC_n$	$KS_n$	$KC_n$	$MC_n$
1.0		0.052	0.051	0.048	0.046	0.046	0.056	0.051	0.045	0.051
1.2	0.3	0.050	0.053	0.051	0.068	0.077	0.076	0.078	0.079	0.087
	0.4	0.054	0.063	0.060	0.075	0.087	0.091	0.079	0.089	0.096
	0.5	0.057	0.059	0.061	0.075	0.090	0.094	0.081	0.093	0.102
	0.6	0.058	0.055	0.054	0.065	0.073	0.071	0.086	0.092	0.098
	0.7	0.053	0.054	0.058	0.069	0.073	0.080	0.069	0.066	0.067
1.4	0.3	0.075	0.104	0.111	0.101	0.182	0.193	0.127	0.254	0.263
	0.4	0.088	0.127	0.131	0.130	0.233	0.246	0.188	0.344	0.349
	0.5	0.090	0.124	0.131	0.136	0.252	0.255	0.198	0.359	0.361
	0.6	0.092	0.131	0.139	0.140	0.246	0.248	0.203	0.368	0.359
	0.7	0.062	0.098	0.105	0.108	0.189	0.189	0.139	0.278	0.277
1.6	0.3	0.109	0.148	0.170	0.174	0.331	0.345	0.259	0.489	0.476
	0.4	0.121	0.190	0.215	0.243	0.405	0.425	0.382	0.597	0.598
	0.5	0.125	0.219	0.222	0.263	0.448	0.455	0.416	0.640	0.638
	0.6	0.115	0.193	0.197	0.254	0.423	0.429	0.383	0.603	0.592
	0.7	0.092	0.150	0.159	0.188	0.349	0.340	0.285	0.495	0.484
1.8	0.3	0.118	0.206	0.208	0.235	0.443	0.464	0.425	0.625	0.626
	0.4	0.166	0.262	0.292	0.372	0.549	0.558	0.585	0.745	0.742
	0.5	0.169	0.280	0.295	0.396	0.592	0.595	0.636	0.760	0.766
	0.6	0.148	0.256	0.260	0.380	0.580	0.583	0.588	0.733	0.729
	0.7	0.111	0.207	0.200	0.263	0.461	0.459	0.421	0.648	0.632
2.0	0.3	0.127	0.225	0.238	0.297	0.484	0.488	0.527	0.695	0.679
	0.4	0.181	0.295	0.307	0.437	0.626	0.628	0.692	0.806	0.810
	0.5	0.204	0.319	0.327	0.506	0.659	0.661	0.765	0.822	0.815
	0.6	0.190	0.284	0.319	0.446	0.622	0.620	0.705	0.802	0.790
	0.7	0.143	0.223	0.234	0.312	0.497	0.495	0.514	0.699	0.674

Lastly, (see [12], Theorem 2.12.1)

$$n^{-1/2}(\alpha_{1n}(x, s), \alpha_{2n}(x, s))$$

converges in distribution to a vector of tight Gaussian processes  $(\eta_1^1, \eta_2^0)$  on  $D(R^d \times [0, 1])$  with zero means and covariances

$$\begin{aligned} E\eta_1^1(x_1, s_1)\eta_1^1(x_2, s_2) \\ = (s_1 \wedge s_2) \times (H_1(x_1 \wedge x_2) - G_1(x_1)G_1(x_2)) \end{aligned}$$

and

$$E\eta_1^0(x_1, s_1)\eta_1^0(x_2, s_2) = (s_1 \wedge s_2)H_1(x_1 \wedge x_2).$$

It follows that  $n^{-1/2}\alpha_n$  converges in distribution to  $\Gamma^0$  with respect to the uniform metric, where

$$\Gamma^0 \equiv \eta_1^1 + \eta_2^0.$$

Since

$$\beta_n(x, s) = n^{-1/2} \left( \alpha_n(x, s) - \frac{[ns]}{n} \alpha_n(x, 1) \right),$$

We obtain  $\beta_n \rightarrow_D \Gamma$ , as  $n \rightarrow \infty$ , where

$$\Gamma =_D \Gamma^0(x, s) - s\Gamma^0(x, 1)$$

and has covariance given in the statement of Theorem 1. Hence Theorem 1 is proven.

By viewing  $\alpha_{2n}(x, [ns])$  as a strong martingale and using the functional central limit theorem for strong martingales by Ivanoff [13], one can drop the assumption (A5) and get convergence in the Shorohod topology. (see [14]).

*Proof of Corollary 3:* Assume the alternative Hypothesis (2). Then,

$$\begin{aligned} \beta_n(x, s) = n^{-1/2} \left( \alpha_n^{11}(x, s) - \frac{[ns]}{n} \alpha_n^{11}(x, 1) \right) \\ + \gamma_n(x, s), \end{aligned}$$

where, for  $s \leq \theta$ ,

$$\gamma_n(x, s) = \sqrt{n} \left( \frac{[ns]}{n} \right) \left( \frac{n - [n\theta]}{n} \right) (G_1(x) - G_1^1(x))$$

$$\alpha_n^{11}(x, s) = \alpha_n(x, s),$$

and for  $s > \theta$ ,

$$\gamma_n(x, s) = \sqrt{n} \left( \frac{[n\theta]}{n} \right) \left( \frac{n - [ns]}{n} \right) (G_1(x) - G_1^1(x)),$$

$$\alpha_n^{11}(x, s) = \alpha_n(x, \theta) + \alpha_n^1(x, s) - \alpha_n^1(x, \theta),$$

with

$$\alpha_n^1(x, s) = \sum_{i=1}^{[ns]} (Y_i I_{D(x)}(X_i) - G_1^1(x)),$$

and

$$G_1^1(x) = \int_{D(x)} \phi^1 dF^1.$$

If Condition (A3) holds for the functions  $\phi^1$  and  $F^1$ , then under the alternative Hypothesis (2),

$$n^{-1/2}\alpha_n^{11} \rightarrow_D \Gamma^{11},$$

in the space  $D(R^d \times [0, 1])$  with the uniform metric, where  $\Gamma^{11}$  is a zero-mean Gaussian process. Strong consistency of the statistics  ${}^nKS$ ,  ${}^nKC$ ,  ${}^nMC$  follows. (See the proof of Theorem 4.1 in [5]).

*Proof of Theorem 4:* Let

$$Z_i(x) = Y_i I_{D(x)}(X_i).$$

By the multivariate CLT, along almost all sequences,

$$\{(X_i, \varepsilon_i), i = 1, 2, \dots\},$$

$$\xi_n(x, s) \equiv n^{-1/2} \sum_{i=1}^{[ns]} V_i Z_i(x)$$

is a sum of iid random variables with mean zero and a resulting covariance matrix having entries of the form

$$n^{-1} \sum_{i=1}^{[(s_1 \wedge s_2)]} Y_i^2 I_{D(s_1 \wedge s_2)}(X_i),$$

which converge, uniformly in  $(x_1, s_1)$  and  $(x_2, s_2)$ ,

$$(s_1 \wedge s_2)(H_1(x_1 \wedge x_2) + H_2(x_1 \wedge x_2)).$$

Similarly, the corresponding conditional covariance of

$$\zeta^{(j)}(x, [ns])$$

converges uniformly to

$$(s_1 \wedge s_2)(H_1(x_1 \wedge x_2) - G_1(x_1 \wedge x_2) + H_2(x_1 \wedge x_2)),$$

and the corresponding conditional covariance of  $\beta^{(j)}(x, [ns])$  converges uniformly to that of  $\Gamma$ .

By the multiplier inequalities ([12], Lemma 3.6.7),

$$\begin{aligned} E_V \left\| n^{-1/2} \sum_{i=1}^n V_i Z_i \right\|_{\mathfrak{S}}^* \\ \leq 2(n_0 - 1)n^{-3/2} \sum_{i=1}^n \|Z_i\|_{\mathfrak{S}}^* E_V \left( \max_{1 \leq i \leq n} |V_i| \right) \\ + 2 \int_0^\infty \sqrt{P(V_1 > t)} dt \max_{n_0 \leq kn} E_R \left\| k^{-1/2} \sum_{i=n_0}^n Z_{R_i} \right\|_{\mathfrak{S}}^*, \end{aligned}$$

where  $(R_1, \dots, R_n)$  is uniformly distributed on the set of all permutations of  $\{1, 2, \dots, n\}$  and is independent of  $(Z_1, \dots, Z_n)$ . In our case,  $\mathfrak{S}$  is the collection of functions

$$\{I_{D(x)} : x \in R^d\}.$$

Also,

$$E \cdot \|\cdot\|_{\mathfrak{S}}^*$$

denotes (uniform) measurable covers with respect to all variables jointly. Using Theorem 1, as in the proof of Theorem 3.6.13 in [12],  $\xi_n(x, 1)$  is Donsker. It follows that ([12], Theorem 2.12.2)  $\xi_n(x, s)$  is Donsker, and Theorem 4 for  $\beta^{(j)}$  follows.

*Proof of Theorem 5:* The proof is similar to that of Theorem 4 but where the limiting covariance of is that of  $\Gamma^{11}$  instead of  $\Gamma$ .

## 6. Concluding Remarks

In this paper, we have constructed test statistics using

$$F = \left\{ I_{D(x)} : x \in R^d \right\},$$

the lower orthants. It is clear (see [12]) that other Donsker classes of functions can be used. Examples are: the collection of the closed balls, or the collection of half-spaces. The method of proof in [5,6,15] could not handle such classes. One can also obtained weak convergence of the weighted version of the process where the weight functions satisfy the Chibisov-O'Reilly conditions (see [5,6], or [16], p. 462).

The simulations show that our tests have good power considering the omnibus nature of the procedure. By avoiding curve estimation (e.g. using kernel methods to estimate  $\phi$ ), we are able to provide practical tests for moderate sample sizes.

## REFERENCES

- [1] C. Lee, M. Csörgő and L. Horváth, "Limit Theorems in Change-Point Analysis," Wiley, Chichester, 1997.
- [2] M. Csörgő, L. Horváth and B. Szyszkowicz, "Integral Tests for Suprema of Kiefer Processes with Application," *Statist, Decisions*, Vol. 15, No. 4, 1997, pp. 365-377.
- [3] M. Csörgő and B. Szyszkowicz, "Weighted Multivariate Empirical Processes and Contiguous Change-Point Analysis," *Change-Point Problems, IMS Lecture Notes Monograph Series*, Vol. 23, 1994, pp. 93-98. [doi:10.1214/lnms/1215463116](https://doi.org/10.1214/lnms/1215463116)
- [4] M. Csörgő and B. Szyszkowicz, "Applications of Multi-Time Parameter Processes to Change-Point Analysis, In: M. Fisz, Ed., *Probability Theory and Mathematical Statistics*, John Wiley & Sons, Inc., New York, 1994, pp. 159-222.
- [5] M. D. Burke, "Limit Theorems for Change-Point Detection in Nonparametric Regression Models," *Acta Mathematica (Szeged)*, Vol. 73, No. 3-4, 2007, pp. 865-882.
- [6] M. D. Burke, "Approximations for a Multivariate Hybrid Process with Applications to Change-Point Detection," *Mathematical Methods of Statistics*, Vol. 19, No. 2, 2010, pp. 121-135. [doi:10.3103/S106653071002002X](https://doi.org/10.3103/S106653071002002X)
- [7] M. D. Burke, "A Gaussian Bootstrap Approach to Estimation and Tests," In: B. Szyszkowicz, Ed., *Asymptotic Methods in Probability and Statistics*, Elsevier, Amsterdam, 1998, pp. 697-706. [doi:10.1016/B978-044450083-0/50045-9](https://doi.org/10.1016/B978-044450083-0/50045-9)
- [8] M. D. Burke, "Multivariate Tests-of-Fit and Uniform Confidence Bands Using a Weighted Bootstrap," *Statistics and Probability Letters*, Vol. 46, No. 1, 2000, pp. 13-20. [doi:10.1016/S0167-7152\(99\)00082-6](https://doi.org/10.1016/S0167-7152(99)00082-6)
- [9] D. Pollard, "A Central Limit Theorem for Empirical Processes," *Journal of the Australian Mathematical Society A*, Vol. 33, No. 2, 1982, pp. 235-248. [doi:10.1017/S1446788700018371](https://doi.org/10.1017/S1446788700018371)
- [10] E. Giné and J. Zinn, "Bootstrapping General Empirical Measures," *Annals of Probability*, Vol. 18, No. 2, 1984, pp. 851-869. [doi:10.1214/aop/1176990862](https://doi.org/10.1214/aop/1176990862)
- [11] E. Giné and J. Zinn, "Lectures on the Central Limit Theorem for Empirical Processes," *Lecture Notes in Mathematics*, Vol. 1221, 1986, pp. 50-113. [doi:10.1007/BFb0099111](https://doi.org/10.1007/BFb0099111)
- [12] A. W. Van der Vaart and J. A. Wellner, "Weak Convergence and Empirical Processes: With Applications to Statistics," Springer, New York, 1996. [doi:10.1007/978-1-4757-2545-2](https://doi.org/10.1007/978-1-4757-2545-2)
- [13] B. G. Ivanoff, "Stopping Times and Tightness for Multi-parameter martingales," *Statistics & Probability Letters*, Vol. 28, No. 2, 1996, pp. 111-114. [doi:10.1016/0167-7152\(95\)00103-4](https://doi.org/10.1016/0167-7152(95)00103-4)
- [14] M. D. Burke, "Testing Regression Models: A Strong Martingale Approach," *Fields Institute Communications*. 2004.
- [15] L. Horváth, "Approximations for hybrids of empirical and partial sums processes," *Journal of Statistical Planning and Inference*, Vol. 88, No. 1, 2000, pp. 1-18. [doi:10.1016/S0378-3758\(99\)00207-4](https://doi.org/10.1016/S0378-3758(99)00207-4)
- [16] G. R. Shorack and J. A. Wellner, "Empirical Processes with Applications to Statistics," Wiley, New York, 1986.