Matouš Maciak

Bootstraping of M-smoothers

Acta Universitatis Carolinae. Mathematica et Physica, Vol. 52 (2011), No. 1, 37--46

Persistent URL: http://dml.cz/dmlcz/143666

Bootstraping of M-smoothers

MATOUŠ MACIAK

Praha

Received April 30, 2010
Revised July 26, 2010

Asymptotic distribution of local polynomial M-smoothers depends on some unknown quantities. However, a knowledge of this distribution is crucial for a hypotheses testing problem in a change-point model. Instead of using some plug-in techniques, which provide a poor approximation, a bootstrap algorithm is proposed to approximate the unknown distribution and a proper justification of this algorithm is given. Finally, some results are illustrated through a proposed simulation study.

1. Motivation and the model

Let $\{(X_i, Y_i);\ i = 1, \ldots, N \in \mathbb{N}\}$ be a finite two-dimensional random sample given from some unknown distribution function $F_{(X,Y)}(x, y)$. We want to investigate a dependence of the random variable Y given the value of a random variable X in sense of a classical nonparametric regression based on a conditional expectation function $m(x) = \mathbf{E}[Y|X = x]$. However, in our approach we would like to adopt methods that would allow us to model the unknown regression function in a less restricted way specifically, we would like to weaken some distributional assumptions (to admit also heavy tailed distributions) and some smoothness assumptions (to account for some discontinuity points). In general, we will assume the model

$$Y_i = m(X_i) + \sigma(X_i) \cdot \varepsilon_i, \qquad \varepsilon_i \sim G,\ i.i.d.,\ i = 1, \ldots, N, \tag{1}$$

where $i.i.d.$ stands for independent and identically distributed random errors, with a symmetric continuous distribution function G (i.e. $G(e) + G(-e) = 1$ for all $e \in \mathbb{R}$).

MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín

This work was kindly supported from Grant 201-090755 & Grant SVV 261315/2010. The author thanks for this support.

Key words and phrases. Nonparametric regression, local polynomial M-smoothers, change-point, smooth residual bootstrap, Mallow's metric

E-mail address: maciak@karlin.mff.cuni.cz

We also assume that $G(1) - G(-1) = \frac{1}{2}$ which is to define a scale of random errors, rather than specifying a unit variance as we want to stay free of any finite moment assumptions. One can easily simplify this heteroscedastic model into a model with a simpler, homoscedastic variance structure however, the results will be derived and proved for model (1) only.

Local polynomial M-smoothers are used to estimate the unknown regression function $m(\cdot)$ under the assumptions below and the main statistical results derived for this method are briefly summarized in Section 2. In Section 3 a model with change-points is described as a generalization of the model considered in [1] and in Section 4 a new bootstrap algorithm based on [8] will be introduced in order to obtain critical values needed to make a decision related to a hypothesis testing problem about a change-point occurrence at some given point from the domain of interest. A proper justification of this algorithm is given also in Section 4. Finally, in Section 5 a simulation study is proposed in order to see a performance of the proposed bootstrap algorithm.

2. M-smoothers

Under the model (1) we are interested in function $m(\cdot)$ which is to be estimated at some given point of interest. By adopting a local polynomial approach one will get not only an estimate for the regression function itself but also for its derivatives. Local polynomial M-smoother estimate is defined by the following minimization problem

$$\widehat{\beta}_{(x)} = \underset{(\beta_0,\ldots,\beta_p)^\top \in \mathbb{R}^{p+1}}{\text{Argmin}} \sum_{i=1}^{N} \rho\left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j\right) \cdot K_{h_N}\left(\frac{X_i - x}{h_N}\right), \tag{2}$$

where $\widehat{\beta}_{(x)} = (\widehat{\beta}_0(x), \widehat{\beta}_1(x), \ldots, \widehat{\beta}_p(x))^\top$ is a $(p+1)$-dimensional vector of parameter estimates at the given point $x \in (0,1)$, where the interval $(0,1)$ is the domain of interest for model (1) and $p \in \mathbb{N}$ is the degree of a local polynomial approximation. Function $K_{h_N}(\cdot)$ stands for a classical kernel function related to nonparametric regression approaches with an appropriate bandwidth parameter h_N and function $\rho(\cdot)$ is a general loss function which is assumed to be symmetric, Lipschitz, convex and such that its derivative exists (one-sided derivatives at least) and it holds that $\rho' = \psi$ almost everywhere (a.e.).

Under some mild assumptions one can get from the minimization problem (2) to a set of normal equations however, in case of local polynomial M-smoothers the solution to the set of these equations is not given in an explicit form. This also involves some issues related to a bandwidth parameter selection as the Asymptotic Mean Squared Error term (AMSE) which is used to determine an asymptotically optimal bandwidth parameter is not expressible in an explicit form either. Therefore, one has to implement iterative procedures or other methods in order to get close to the solution. For more details we refer to [2] or [6].

Let us now introduce the following notation:

$$\mu = \left(\int u^{0} K(u) \mathrm{d}u, \ldots, \int u^{p} K(u) \mathrm{d}u \right)^{\top}, \qquad \mathbf{H}_N = diag\{1, h_N^{-1}, \ldots, h_N^{-p}\},$$

$$\mathbf{S}_1 = \left(\int_{-1}^{1} u^{j+l} K(u) \mathrm{d}u \right)_{\substack{j=0\ldots p;\\ l=0\ldots p;}}, \qquad \mathbf{S}_2 = \left(\int_{-1}^{1} u^{j+l} K^2(u) \mathrm{d}u \right)_{\substack{j=0\ldots p;\\ l=0\ldots p;}},$$

where $K(\cdot)$ is a kernel function with a support $[-1, 1]$ and

$$\lambda_G(t, v) = -\int_{-\infty}^{\infty} \psi(v\,\varepsilon_1 - t) \mathrm{d}G(\varepsilon_1), t \in \mathbb{R}, v > 0, \text{ and } \mathsf{e}_\nu = (\underbrace{0, \ldots, 0}_{\nu\ times}, 1, \underbrace{0, \ldots, 0}_{(p-\nu)\ times})^{\top}.$$

Theorem 1 (Asymptotic normality for M-smoothers)
Under the model (1) and the assumptions A1 – A8 stated above it holds:

$$\sqrt{Nh_N} \cdot \left[\widehat{m}^{(\nu)}(x) - m^{(\nu)}(x) \right] \xrightarrow[N \to \infty]{\mathscr{D}} \mathbf{N}\left(0, \frac{\mathbf{E}\left[\psi^2(\sigma(x)\varepsilon_1)\right] \nu!^2}{[\lambda'_G(0, \sigma(x))]^2 f(x)} \mathsf{e}_\nu^{\top} \mathbf{H}_N \mathbf{S}_1^{-1} \mathbf{S}_2 \mathbf{S}_1^{-1} \mathbf{H}_N \mathsf{e}_\nu \right)$$

where $\widehat{m}^{(\nu)}(x) = \nu! \widehat{\beta}_\nu(x)$, for $\nu \in \{0, 1, \ldots, p\}$ and $\widehat{\beta}_\nu(x)$ being the elements of the vector of estimates $\widehat{\beta}_{(x)}$ defined in (2). Moreover, the expectation $\mathbf{E}\left[\psi^2(\sigma(x)\varepsilon_1)\right] = \int \psi^2(\sigma(x)\varepsilon_1) \mathrm{d}G(\varepsilon_1)$, and $\lambda'_G(0, \sigma(x)) = \frac{\partial}{\partial t} \lambda_G(t, \sigma(x))\Big|_{t=0}$ for given $x \in (0, 1)$.

In case of a model with homoscedastic variance structure the result gets slightly simpler as the scale function $\sigma(x)$ will not appear in the argument of function ψ and function $\lambda'_G(t, \sigma(x))$ in a denominator of the variance term will also become independent of $x \in (0, 1)$ implemented via the scale function $\sigma(\cdot)$.

Proof of Theorem 1. The proof of this theorem follows as a straightforward generalization of the proof for a homoscedastic model given in [7]. One just has to adopt slightly more computational effort in order to deal with function $\lambda_G(t, \sigma(x))$ which is a function of two different arguments in this case. However, the proof goes along the same lines as those in [7] therefore, it will be omitted here in this paper. □

2.1 Assumptions/conditions

For sake of completeness let us state the assumptions required for model (1) in order to assure that the formulated results hold:

- A1 The density function $f(\cdot)$ of the random variable X is absolutely continuous, positive and bounded on [0, 1] which is the support of X;
- A2 Random errors $\varepsilon_1, \ldots, \varepsilon_N$, are assumed to be i.i.d., mutually independent of X_i, for $i = 1, \ldots, N$, with a symmetric distribution given by a continuous distribution function G such that $G(1) - G(-1) = \frac{1}{2}$;
- A3 The regression function $m(\cdot)$ is assumed to be $(p + 1)$-times Lipschitz at some neighbourhood of $x \in (0, 1)$, where $p \in \mathbb{N}$ is a degree of a local polynomial approximation;
- A4 The scale function $\sigma(\cdot)$ is Lipschitz and positive on [0,1];
- A5 The loss function ρ is symmetric, convex and Lipschitz, its derivative exists (or at least one-sided derivatives exist) and $\rho' = \psi$ a.e.;

A6 We assume that function $\lambda_G(t, v) = -\int \psi(v\varepsilon_1 - t)\mathrm{d}G(\varepsilon_1)$ is Lipschitz in both arguments, $t \in \mathbb{R}$ and $v > 0$. We assume, that the partial derivative $\lambda'_G(t, v) = \frac{\partial}{\partial t}\lambda_G(t, v)$ exists and $\int \psi^2(v\varepsilon_1 - t)\mathrm{d}G(\varepsilon_1)$ is finite, both at some neighbourhoods of $t = 0$ and $v = \sigma(x)$, for $x \in (0, 1)$. Moreover, we need the following equality to be satisfied: $\lambda'_G(0, \sigma(x)) = \frac{\partial}{\partial t}\ \lambda_G(t, \sigma(x))|_{t=0} \neq 0$, for the given point $x \in (0, 1)$;

A7 Let $K_{h_N}(\cdot) = \frac{1}{h_N}K(\cdot)$ where $K(\cdot)$ is a symmetric density function with common support on $[-1, 1]$, such that $\int_{-1}^{1} u^j K(u)\mathrm{d}u < \infty$ and $\int_{-1}^{1} u^j K^2(u)\mathrm{d}u < \infty$, for $j = 0, \ldots, 2p$;

A8 The bandwidth parameter h_N satisfies: $h_N \overset{N\to\infty}{\longrightarrow} 0$, $Nh_N \overset{N\to\infty}{\longrightarrow} \infty$, more precisely $h_N \in (N^{\iota-1}, N^{-\iota-\frac{1}{2}})$, where $\iota > 0$ is small enough.

3. Change-points and bootstrapping of M-smoothers

Let us now introduce a model where the smoothness assumption A3 is relaxed a little as we allow a model with some discontinuity points in the regression function itself or in its derivatives respectively. Unlike some other papers we will stay free of any assumptions related to some ties among discontinuity points[1]. As the model with one change-point only can be easily generalized into a case with multiple change-points or even into a model with discontinuities in higher order derivatives we will discuss a model with one discontinuity point only. Let us consider the following regression model:

$$Y_i = m(X_i) + \sigma(X_i)\cdot\varepsilon_i, \qquad \varepsilon_i \sim G,\ i.i.d.,\ i = 1, \ldots, N, \tag{3}$$

where

$$m(X_i) = m_0(X_i) + \Delta\cdot\mathbb{I}_{\{X_i > x_0\}},\ \text{for some } \Delta \neq 0,$$

under the assumptions A1, A2 and A4 – A8, where $x_0 \in (0, 1)$ is some known point given in advance, function $m_0(\cdot)$ satisfies assumption A3 and $\mathbb{I}_{\{\cdot\}}$ is an identifier function for the given event of interest.

We would like to perform a test now in order to decide if there is a significant change-point occurrence at the point $x_0 \in (0, 1)$ or if there is not. We consider the following statistical test

$$\left.\begin{array}{ll} \mathsf{H}_0: & \Delta = 0 \\ \mathsf{H}_1: & \Delta \neq 0 \end{array}\right\} \text{ for given } x_0 \in (0, 1)$$

and we propose to use the following test statistic

$$T_N(x_0) = \sqrt{Nh_N}\cdot|\widehat{m}_+(x_0) - \widehat{m}_-(x_0)|, \tag{4}$$

where the quantities $\widehat{m}_+(x_0)$ and $\widehat{m}_-(x_0)$ respectively are one-sided estimates given by the minimization problem (2) respectively by Theorem 1 however, the original kernel function $K_{h_N}(\cdot)$ is replaced here by one-sided counterparts $K^+_{h_N}(\cdot) = 2K_{h_N}(\cdot)\mathbb{I}_{\{\cdot\ \geq 0\}}$ and

[1] Some authors a priori consider a kind of hierarchy: if there is a discontinuity point in a derivative there also has to be a discontinuity point at the same location in all lower order derivatives too. However, this is not the case in our paper.

$K^-_{h_N}(\cdot) = 2K_{h_N}(\cdot)\mathbb{I}_{\{\cdot\,\leq 0\}}$. One can expect of quantities $\widehat{m}_+(x_0)$ and $\widehat{m}_-(x_0)$ to be nonparametric estimates of corresponding theoretical values $m_+(x_0) = \lim_{y\searrow x_0} m(y)$ and $m_-(x_0) = = \lim_{y\nearrow x_0} m(y)$ therefore, such test statistics should be sensitive at a possible jump point and one would expect it to be large if there is a change point present at the point x_0 and it should be negligible if there is not. Let us now state the main results for this test based on statistic $T_N(x_0)$.

Theorem 2 (Distribution of the test statistics under the null hypothesis) Given the model (3) and the assumptions A1 – A8, with A3 satisfied for $m_0(\cdot)$, under the notation stated above and the null hypothesis H_0, it holds

$$\sqrt{Nh_n}\cdot\left|\widehat{m}_+(x_0)-\widehat{m}_-(x_0)\right|\xrightarrow[N\to\infty]{\mathscr{D}}\mathbf{N}\left(0,\frac{2\mathbf{E}\left[\psi^2(\sigma(x)\varepsilon_1)\right]}{[\lambda'_G(0,\sigma(x))]^2 f(x)}\mathsf{e}_0^\top\underline{\mathbf{S}}_1^{-1}\underline{\mathbf{S}}_2\underline{\mathbf{S}}_1^{-1}\mathsf{e}_0\right),$$

where the matrices $\underline{\mathbf{S}}_1$ and $\underline{\mathbf{S}}_2$ are the same as those in Theorem 1, just the kernel function $K_{h_N}(\cdot)$ is replaced by the corresponding one-sided counterpart $K^+_{h_N}(\cdot)$ or $K^-_{h_N}(\cdot)$ respectively and $\mathsf{e}_0 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^{p+1}$.

Proof of Theorem 2. Once we realize that both estimates $\widehat{m}_+(x_0)$ and $\widehat{m}_-(x_0)$ are independent of each other moreover, if appropriately normalized they both follow in asymptotic the normal distribution given by Theorem 1 even though, the one-sided kernel functions are not symmetric as originally assumed in A7 the result finally follows after some computation by combining these two facts. □

Theorem 3 (Consistency)
Under the model (3) and the assumptions as in Theorem 2, given the alternative hypothesis H_1, it holds

$$\sqrt{Nh_N}\cdot\left|\widehat{m}_+(x_0)-\widehat{m}_-(x_0)\right|\xrightarrow[N\to\infty]{\mathbf{P}}\infty.$$

Proof of Theorem 3. Under the alternative hypothesis H_1 it holds that $|\widehat{m}_+(x_0) - - \widehat{m}_-(x_0)| \xrightarrow{\mathbf{P}} |\Delta| > 0$ so the result holds under the assumption A8 as $\sqrt{Nh_N} \to \infty$ for $N \to \infty$. □

Once we know the limit distribution of the test statistic $T_N(x_0)$ we can base our decision about the hypothesis testing problem on the appropriate critical value $t_N^{x_0}(\alpha)$ and to reject the null hypothesis once $T_N(x_0) \geq t_N^{x_0}(\alpha)$ where $\alpha \in (0, 1)$ is a level of the test and $t_N^{x_0}(\alpha)$ comes from the asymptotic distribution of the test statistic $T_N(x_0)$ under the null hypothesis H_0.

4. Smooth residual bootstrap algorithm

It is easy to see that the asymptotic distribution given by Theorem 2 depends on some unknown quantities. In real situation one will never know what is the true distribution function G, the density function $f(\cdot)$ or the variability structure given by the scale function $\sigma(\cdot)$. Applying some well-known plug-in techniques could do the work however, the performance of such methods is rather poor and it involves quite much computational effort. Therefore, we introduce a new bootstrap algorithm which can be used to mimic the unknown distribution of interest.

Bootstrap algorithm for a heteroscedastic model

- B1 Compute residuals $\{\widehat{\varepsilon}_i;\ i = 1, \ldots, N\}$, where $\widehat{\varepsilon}_i = \frac{Y_i - \widehat{m}(X_i)}{\widehat{\sigma}(X_i)}$, where $\widehat{m}(X_i)$ is an estimate of $m(X_i)$ at X_i defined by (2) and the scale function $\widehat{\sigma}(X_i) = median\left\{|Y_j - \widehat{m}(X_j)|;\ j = 1, \ldots, N,\ \text{such that}\ |X_j - X_i| \leq h_N\right\}$;
- B2 Resample with replacement from the set $\{\widehat{\varepsilon}_i;\ i = 1, \ldots, N\}$ in order to obtain new residuals $\tilde{\varepsilon}_i$, for $i = 1, \ldots, N$;
- B3 Define new bootstrap residuals $\varepsilon_i^\star = V_i \cdot \tilde{\varepsilon}_i + a_N \cdot Z_i$, where $\mathbf{P}[V_i = -1] = \mathbf{P}[V_i = 1] = \frac{1}{2}$, $Z_i \sim N(0,1)$ are i.i.d. standard normal random variables and $a_N = o(1)$ is an appropriate bootstrap bandwidth parameter, such that $N h_N a_N^2 / \log N \to \infty$, as $N \to \infty$;
- B4 Define a new bootstrap data sample $\{(X_i, Y_i^\star);\ i = 1, \ldots, N\}$, where $Y_i^\star = \widehat{m}(X_i) + \widehat{\sigma}(X_i) \cdot \varepsilon_i^\star$;
- B5 Re-estimate the unknown functions $m(x_0), m_+(x_0)$ and $m_-(x_0)$ respectively based on the new data sample $\{(X_i, Y_i^\star);\ i = 1, \ldots, N\} \to$ obtain $\widehat{m}^\star(x_0), \widehat{m}_+^\star(x_0)$ and $\widehat{m}_-^\star(x_0)$;
- B6 Repeat steps $2 \to 3 \to 4 \to 5$ to get the estimates $\widehat{m}_b^\star(x_0), \widehat{m}_{+b}^\star(x_0)$ and $\widehat{m}_{-b}^\star(x_0)$, for $b = 1, \ldots, B$, for B sufficiently large;
- B7 Use the quantities produced in step 6 to mimic the unknown distribution of interest.

In order to obtain a bootstrap procedure for a model with homoscedastic variance structure one needs to slightly alter the first step of the algorithm and to consider a set of residuals without standardizing them by $\widehat{\sigma}(\cdot)$. However, in order to ensure symmetric residuals and continuous and symmetric density function of bootstrapped residuals one still has to consider a correction step involved in B3.

Before we state the main bootstrap consistency result let us mention a notation of a conditional weak convergence in probability which will become the main proving tool later on.

Definition 1 (Conditional weak convergence in probability)
Let $\{\mathbf{T}_N, \mathbf{T}_N^{'}\}_{N=1}^{\infty}$ be some random vectors. If for every real-valued and bounded continuous function f holds that

$$\mathbf{E}[f(\mathbf{T}_N^{'})|\mathbf{S}_N] - \mathbf{E}[f(\mathbf{T}_N)] \xrightarrow[N\to\infty]{\mathbf{P}} 0,$$

then $\mathbf{T}_N^{'}$ condition on $\mathbf{S}_N$ and $\mathbf{T}_N$ are said to be approaching (each other) in distribution in probability along sequences $\mathbf{S}_N$. In short we use the notation

$$\mathbf{T}_N^{'}|\mathbf{S}_N \underset{N\,\to\,\infty}{\overset{\mathscr{D}(\mathbf{P})}{\longleftrightarrow}} \mathbf{T}_N.$$

Now we can formulate the following theorem which is the main result regarding the proposed bootstrap consistency.

Theorem 4 (Bootstrap consistency)
Under the model (3) and assumptions A1 – A8 with A3 satisfied for $m_0(\cdot)$ and the additional assumption posed on the bootstrap bandwidth parameter a_N in B3, the following holds:

$$\sup_{z\in\mathbb{R}}\left\{\mathbf{P}^\star\left[\sqrt{Nh_N}\left(\widehat{m}^\star_+(x_0)-\widehat{m}^\star_-(x_0)\right)\le z\right]-\mathbf{P}\left[\sqrt{Nh_N}\left(\widehat{m}_+(x_0)-\widehat{m}_-(x_0)\right)\le z\right]\right\}\xrightarrow[N\to\infty]{\mathbf{P}}0,$$

where $\mathbf{P}^\star[\,\cdot\,]$ stands for a conditional probability given the random sequence $\mathscr{Y}=\{Y_i;\ i==1,\ldots,N\in\mathbb{N}\}$.

The result in Theorem 4 means that the bootstrap algorithm as defined in steps B1 – B7 can be used to mimic the unknown distribution of interest for the test based on statistic $T_N(x_0)$.

Proof of Theorem 4. Because of the length restriction we will not state very all details of the proof however, we will go along all important steps of this proof. Details will be given in author's Ph.D. thesis.

We will start with the definition of bootstrap residuals $\varepsilon_i^\star$, for $i=1,\ldots,N$. It was proved in [8] that resampling residuals as defined in B1 - B3 is in limit equivalent to a sampling from a continuous nonparametric estimate of the distribution function of random errors $\varepsilon_1,\ldots,\varepsilon_N$ where the distribution is moreover symmetric. Given this fact we can proceed along the same lines as those in the proof of Theorem 1 in [7]. Using some notation common for robust estimation theory and the following convergence in probability result (see e.g. [7] for more details)

$$\sup_{|t|<T}\frac{1}{\sqrt{Nh_N}}\sum_{i=1}^{N}\left[\psi\left(\widehat{\sigma}(X_i)\varepsilon_i^\star-\frac{t}{\sqrt{Nh_N}}\right)-\psi\left(\widehat{\sigma}(X_i)\varepsilon_i^\star\right)\right.$$
$$\left.-\int_{\mathbb{R}}\psi\left(\widehat{\sigma}(X_i)\varepsilon_i^\star-\frac{t}{\sqrt{Nh_N}}\right)\mathrm{d}G_N^\star(\varepsilon_i^\star)\right]\cdot K_{h_N}\left(\frac{X_i-x}{h_N}\right)\xrightarrow[N\to\infty]{\mathbf{P}}0,$$

for $T>0$, where $G_N^\star$ is a distribution function of bootstrapped random errors $\varepsilon_1^\star,\ldots,\varepsilon_N^\star$ and $t=\sqrt{Nh_N}\left[\sum_{j=0}^{p}\left(\widehat{\beta}_j(x)-\frac{m^{(j)(x)}}{j!}\right)(X_i-x)^j\right]$, going along some straightforward computation where we use the matrix notation $\mathbb{X}=\left(\left(\frac{X_i-x}{h_N}\right)^j\right)_{ij}$, for $i=1,\ldots,N$ and $j=0,\ldots,p$, to be an $(N\times(p+1))$ type matrix and $\mathbb{W}=diag\{K((X_i-x)/h_N)\}$, for $i=1,\ldots,N$ to be a diagonal $(N\times N)$ type matrix we can express the bootstrap consistency result as given in Theorem 4 in terms of a conditional weak convergence in probability in the following form:

$$\frac{\sqrt{Nh_N}}{\lambda'_{G_N^\star}(0,\widehat{\sigma}(x))}\left(\mathbb{X}^\top\mathbb{W}\mathbb{X}\right)^{-1}\cdot\mathbb{X}^\top\mathbb{W}\underline{\psi(\varepsilon^\star)}\Big|\mathscr{Y}\ \underset{N\to\infty}{\overset{\mathscr{D}(\mathbf{P})}{\longleftrightarrow}}\ \frac{\sqrt{Nh_N}}{\lambda'_{G}(0,\sigma(x))}\left(\mathbb{X}^\top\mathbb{W}\mathbb{X}\right)^{-1}\cdot\mathbb{X}^\top\mathbb{W}\underline{\psi(\varepsilon)},$$

where $\underline{\psi(\varepsilon^\star)}=(\psi(\varepsilon_1^\star),\ldots,\psi(\varepsilon_N^\star))^\mathrm{T}$, $\underline{\psi(\varepsilon)}=(\psi(\varepsilon_1),\ldots,\psi(\varepsilon_N))^\mathrm{T}$ and $\mathscr{Y}=\{Y_i;\ i=1,\ldots,N\in\mathbb{N}\}$.

Therefore, by [3] it is sufficient to prove now that

$$\underbrace{d_{M,2}\left(\frac{\sqrt{Nh_N}}{\lambda'_{G_N^\star}(0,\widehat{\sigma}(x))}\left(\mathbb{X}^\top\mathbb{W}\mathbb{X}\right)^{-1}\mathbb{X}^\top\mathbb{W}\underline{\psi(\varepsilon^\star)}\Big|\mathscr{Y},\ \frac{\sqrt{Nh_N}}{\lambda'_{G}(0,\sigma(x))}\left(\mathbb{X}^\top\mathbb{W}\mathbb{X}\right)^{-1}\mathbb{X}^\top\mathbb{W}\underline{\psi(\varepsilon)}\right)}_{d_{M,2}(\star,*)}\xrightarrow[N\to\infty]{\mathbf{P}}0$$

where $d_{M,2}(\cdot,\cdot)$ stands for a second degree Mallow's metric which can be used to metrisize a convergence in distribution and $\widehat{\sigma}(x)$ stands for the proper estimate of the scale function $\sigma(x)$ as defined in the bootstrap algorithm in step B1. We will take a look at the distribution function $G_N^\star$ of bootstrapped random errors $\varepsilon_i^\star$ for $i = 1,\ldots,N$, conditioned on sample $\mathscr{Y}$ and we will show that

$$d_{M,2}\left(\frac{1}{\sqrt{Nh_N}}\cdot\mathbb{X}^\top\mathbb{W}\underline{\psi(\varepsilon^\star)}\,|\,\mathscr{Y},\ \frac{1}{\sqrt{Nh_N}}\cdot\mathbb{X}^\top\mathbb{W}\underline{\psi(\varepsilon)}\right)\xrightarrow[N\to\infty]{\mathbf{P}}0. \tag{5}$$

Using Taylor expansion, the definition of bootstrapped random errors $\varepsilon_i^\star = V_i\cdot\tilde{\varepsilon}_i + a_N Z_i$, for $i = 1,\ldots,N$, Glivenko-Cantelli theorem, and the property $a_N = o(1)$, one can show that $\sup_{y\in\mathbb{R}}|G_N^\star(y) - G(y)| \to 0$, as $N\to\infty$ which means that the distribution of bootstrapped random errors $\varepsilon_i^\star$'s is asymptotically the same as the distribution of original random errors ε_i's. Given this fact, assumptions A1, A2 and A6, given the mutual independence of X_i's and $\varepsilon_i^\star$'s conditioned on $\mathscr{Y}$ we can apply Central limit theorem now to achieve the convergence in (5).

Now, we can use the property of Mallow's metric (see [5] for details) and to write the following:

$$\underbrace{d_{M,2}\left(\sqrt{Nh_N}\left(\mathbb{X}^\top\mathbb{W}\mathbb{X}\right)^{-1}\mathbb{X}^\top\mathbb{W}\underline{\psi(\varepsilon^\star)}\Big|\,\mathscr{Y},\ \sqrt{Nh_N}\left(\mathbb{X}^\top\mathbb{W}\mathbb{X}\right)^{-1}\mathbb{X}^\top\mathbb{W}\underline{\psi(\varepsilon)}\right)}_{\hookrightarrow d_{M,2}(\circ,\bullet)}\leq$$

$$\leq Nh_N\cdot\left\|\left(\mathbb{X}^\top\mathbb{W}\mathbb{X}\right)^{-1}\right\|_2\cdot d_{M,2}\left(\frac{1}{\sqrt{Nh_N}}\mathbb{X}^\top\mathbb{W}\underline{\psi(\varepsilon^\star)}\,|\,\mathscr{Y},\ \frac{1}{\sqrt{Nh_N}}\cdot\mathbb{X}^\top\mathbb{W}\underline{\psi(\varepsilon)}\right)\xrightarrow[N\to\infty]{\mathbf{P}}0,$$

as the matrix $\left(\mathbb{X}^\top\mathbb{W}\mathbb{X}\right)^{-1}$ can be thought of as a random linear operator on an Euclidean $\mathbb{R}^{p+1}$ space and the norm of this operator tends in probability to $\frac{\|\underline{\mathbf{S}}_2^{-1}\|_2}{Nh_N f(x)}$ where $\underline{\mathbf{S}}_2$ is the matrix defined in Theorem 2 its norm is finite and $f(x)\neq 0$, both under the assumptions A1 and A6.

Next, we need to prove the following convergence in probability

$$\frac{1}{\lambda'_{G_N^\star}(0,\widehat{\sigma}(x))}\xrightarrow[N\to\infty]{\mathbf{P}}\frac{1}{\lambda'_G(0,\sigma(x))}, \tag{6}$$

where $G_N^\star$ stands for a distribution function of bootstrapped random errors given the empirical residuals $\widehat{\varepsilon_1},\ldots,\widehat{\varepsilon_N}$ and $\sigma(x)$ in $\lambda'_G(0,\sigma(x))$ is replaced by an efficient estimate $\widehat{\sigma}(x)$ in $\lambda'_{G_N^\star}(0,\widehat{\sigma}(x))$. To show this, we will prove that

$$\sup_{|t|<\delta}\left|\lambda_{G_N^\star}(t,\widehat{\sigma}(x)) - \lambda_G(t,\sigma(x))\right|\xrightarrow[N\to\infty]{\mathbf{P}}0,$$

for some $\delta > 0$ however, it easily follows that

$$\sup_{|t|<\delta}\left|\lambda_{G_N^\star}(t,\widehat{\sigma}(x)) - \lambda_G(t,\sigma(x))\right| \leq \sup_{|t|<\delta}\left|\lambda_{G_N^\star}(t,\widehat{\sigma}(x)) - \lambda_G(t,\widehat{\sigma}(x))\right| + \sup_{|t|<\delta}\left|\lambda_G(t,\widehat{\sigma}(x)) - \lambda_G(t,\sigma(x))\right|,$$

where both terms on the right hand side tends to zero in probability, as $N\to\infty$. The first convergence is achieved using some per-partes integration and Glivenko-Cantelli theorem

SIMULATION results		Loss function ρ (L_2, L_1 or Huber)	$T_N(x_0 = 0.3)$ (under the $\mathbf{H}_0$)	$T_N(x_0 = 0.8)$ (under the $\mathbf{H}_1$)
Normal	The value of $T_N(x_0)$ *(standard error)*	*L_2 norm*	0.001 *(0.301)*	0.490 *(0.236)*
		L_1 norm	−0.049 *(0.254)*	0.452 *(0.209)*
		Huber function	−0.068 *(0.265)*	0.415 *(0.263)*
	95% Critical region *(based on $T_N(x_0)$)*	*L_2 norm*	(−0.601, 0.636)	(−0.987, −0.052)
		L_1 norm	(−0.553, 0.451)	(−0.833, −0.045)
		Huber function	(−0.703, 0.562)	(−0.906, −0.074)
Cauchy	The value of $T_N(x_0)$ *(standard error)*	*L_2 norm*	−0.028 *(3.888)*	0.157 *(3.303)*
		L_1 norm	−0.130 *(0.288)*	0.403 *(0.281)*
		Huber function	−0.136 *(0.379)*	0.521 *(0.311)*
	95% Critical region *(based on $T_N(x_0)$)*	*L_2 norm*	(−9.741, 9.885)	(−7.452, 7.061)
		L_1 norm	(−0.706, 0.462)	(−0.962, 0.109)
		Huber function	(−0.910, 0.607)	(−1.326, −0.094)

and the second convergence follows from the continuity assumption posed on $\lambda_G(t, v)$ in argument v and the fact that $\widehat{\sigma}(x)$ is a consistent estimate of the scale function $\sigma(x)$ at $x \in (0, 1)$.

Given the assumption A6 and the convergence derived above we can write that

$$\lambda'_{G_N^\star}(0, \widehat{\sigma}(x)) \xrightarrow[N \to \infty]{\mathbf{P}} \lambda'_G(0, \sigma(x)),$$

which was to be shown as the convergence in (6) follows easily. Finally, we can apply Slutzsky's theorem to show

$$\left.\begin{array}{ll} \frac{1}{\lambda'_{G_N^\star}(0,\widehat{\sigma}(x))} \longrightarrow \frac{1}{\lambda'_G(0,\sigma(x))} & \text{in probability;} \\ d_{M,2}(\circ, \bullet) \longrightarrow 0 & \text{in probability;} \end{array}\right\} \overset{(Slutzsky)}{\Longrightarrow} \quad d_{M,2}(\star, *) \xrightarrow[N \to \infty]{\mathbf{P}} 0,$$

which was about to be proved. □

5. Simulation study

To see a performance of the bootstrap algorithm we present the following simulation study: Let the regression function $m(x) = x \cdot sin(2\pi x) + 0.1 \cdot \mathbb{I}_{\{x>0.8\}}$ has one change-point located at the point $x_0 = 0.8$ with a jump size $\Delta = 0.1$ with random errors independently generated from normal distribution $N(0, 0.1)$ with 10% of outlying observations generated from distribution $N(0, 0.5)$ for the first simulation run and from the Cauchy distribution with scale equal to 0.1 (the same as for the normal case) in the second simulation run.

We estimate the regression function $m(\cdot)$ via M-smoothers approach (using L_2 norm, L_1 norm, and Huber function) at two different points ($x_0 = 0.3$ where no change occurs and $x_0 = 0.8$ where it does) and we apply the proposed bootstrap algorithm to mimic the limit distribution of the test statistic under the null hypothesis in order to obtain critical values for a statistical test to decide when the null hypothesis is true ($x_0 = 0.3$) or when the alternative hypothesis is true ($x_0 = 0.8$) respectively.

The sample size $N = 400$ and the number of bootstrap replicates was set to be 1000. The bandwidth parameter h_N was taken via an appropriate robust version of a GCV criterion given the specified loss function (see e.g. [6] for more details). One can clearly see an effect of Cauchy distributed random errors in case of L_2 norm (large standard errors and to long confidence intervals) while L_1 norm provides slightly better results also for Cauchy distribution. However, the null hypothesis is correctly rejected for Tuckey's function only in case of Cauchy distribution and $x_0 = 0.8$. For contaminated normal distributed random errors all three options provide quite similarly.

References

[1] Antoch, J., Gregoire, G., Hušková, M.: *Test for continuity of regression function*. J. Statist. Plann. Inference **137** (2006), 753–777.

[2] Antoch, J. and Janssen, P.: *Nonparametric regression M-quantiles*. Statist. Probab. Lett. **8** (1989), 355–362.

[3] Belyaev, Y. K.: Bootstrap, Resampling and Mallows Metric. Lecture Notes, No. 1, Institute for Mathematical Statistics, Umeå University, Sweden (1995).

[4] Belyaev, Y. K.: *Necessary and sufficient conditions for consistency of resampling*. Technical Report, Centre of Biostochastics, Swedish Univeristy of Agricultual Sciences, Sweden (2003).

[5] Bickel, J. P. and Freedman, D. A.: *Some asymptotic theory for the bootstrap*. Ann. Statist. **9** (6) (1981), 1196–1217.

[6] Heng, D. and Leugn, Y.: *Cross-validation in nonparametric regression with outliers*. Ann. Statist. **33** (5) (2005), 2291–2310.

[7] Maciak, M.: *M-smoothers in testing and estimating*. WDS'07 Proceedings of Contributed Papers [Part I], 73 (2007), 169–174.

[8] Neumeyer, N.: *Bootstrap procedures for empirical processes of nonparametric residuals*. Habilitationssschirft, Fakultät für Mathematik, Ruhr-Universität Bochum (2006).