Statistical Papers (2020) 61:1437–1463
https://doi.org/10.1007/s00362-020-01162-8

REGULAR ARTICLE

Estimating change points in nonparametric time series regression models

Maria Mohr[1] · Leonie Selk[1]

Received: 16 September 2019 / Revised: 13 January 2020 / Published online: 27 January 2020

Abstract
In this paper we consider a regression model that allows for time series covariates as well as heteroscedasticity with a regression function that is modelled nonparametrically. We assume that the regression function changes at some unknown time $\lfloor ns_0 \rfloor$, $s_0 \in (0, 1)$, and our aim is to estimate the (rescaled) change point s_0. The considered estimator is based on a Kolmogorov-Smirnov functional of the marked empirical process of residuals. We show consistency of the estimator and prove a rate of convergence of $O_P(n^{-1})$ which in this case is clearly optimal as there are only n points in the sequence. Additionally we investigate the case of lagged dependent covariates, that is, autoregression models with a change in the nonparametric (auto-) regression function and give a consistency result. The method of proof also allows for different kinds of functionals such that Cramér-von Mises type estimators can be considered similarly. The approach extends existing literature by allowing nonparametric models, time series data as well as heteroscedasticity. Finite sample simulations indicate the good performance of our estimator in regression as well as autoregression models and a real data example shows its applicability in practise.

Keywords Change point estimation · Time series · Nonparametric regression · Autoregression · Conditional heteroscedasticity · Consistency · Rates of convergence

Mathematics Subject Classification Primary 62G05; Secondary 62G08 · 62G20 · 62M10

1 Introduction

Change point analysis has gained attention for decades in mathematical statistics. There is a vast literature on testing for structural breaks when the possible timing of

✉ Leonie Selk
leonie.selk@math.uni-hamburg.de

[1] Department of Mathematics, University of Hamburg, Hamburg, Germany

such a break, the *change point*, is unknown, see for instance Kirch and Kamgaing (2012) and reference mentioned therein. This paper, however, is concerned with the estimation of the change point when assuming its existence.

The most simple set of models can be described as follows

$$Y_t = \mu_1 I\{t \leq \lfloor ns_0 \rfloor\} + \mu_2 I\{t > \lfloor ns_0 \rfloor\} + \varepsilon_t,\ t = 1, \ldots, n,$$

where $s_0 \in (0, 1)$ is the (rescaled) change point, μ_1 and μ_2 the signal before and after the break, respectively, and $(\varepsilon_t)_t$ being stationary and centred errors. These models are often referred to as AMOC-models (at most one change). The problem naturally moved from the standard case with independent errors (see Ferger and Stute (1992) among others) to the time series context. Both Bai (1994) and Antoch et al. (1997) allow for linear processes and Hušková and Kirch (2008) more generally for dependent errors.

Additional information on the form of the signal can be expressed through a process of covariates $(X_t)_t$ resulting in linear regression models with a change in the regression parameter, such as

$$Y_t = \beta_1 X_t I\{t \leq \lfloor ns_0 \rfloor\} + \beta_2 X_t I\{t > \lfloor ns_0 \rfloor\} + \varepsilon_t,\ t = 1, \ldots, n,$$

where β_1 and β_2 are the regression coefficients before and after the break, respectively. Bai (1997), Horváth et al. (1997) and Aue et al. (2012) among others consider the estimation of a change point in (multiple) linear regression models making use of least squares estimation. Considering $X_t = Y_{t-1}$ in the linear regression model from above, one obtains autoregressive models with one change in the autoregressive parameter. The estimation of the parameters and the unknown change point in AR(1) models was for instance considered by Chong (2001), Pang et al. (2014) and Pang and Zhang (2015).

Our aim is to propose an estimator for the change point s_0 in a nonparametric version of the regression model from above, namely

$$Y_t = m_{(1)}(X_t) I\{t \leq \lfloor ns_0 \rfloor\} + m_{(2)}(X_t) I\{t > \lfloor ns_0 \rfloor\} + \varepsilon_t,\ t = 1, \ldots, n,$$

for some nonparametric regression functions $m_{(1)}, m_{(2)}$ (before and after the break) and in addition also investigate the autoregressive case where $X_t = Y_{t-1}$. While the investigation of points of discontinuity in (nonparametric) regression functions has been studied to some extend (see for instance Döring and Jensen (2015) for an overview), not that much research has been devoted to change point analysis in nonparametric models as the one above, where the change occurs in time. Delgado and Hidalgo (2000) propose estimators for the location and size of structural breaks in a nonparametric regression model imposing scalar breaks in time or values taken by some regressors, as in threshold models. Their rates of convergence and limiting distribution depends on a bandwidth, chosen for the kernel estimation. Chen et al. (2005) estimate the time of a scalar change in the conditional variance function in nonparametric heteroscedastic regression models using a hybrid procedure that combines the least squares and nonparametric methods.

The paper at hand extends existing literature, on the one hand by allowing for nonparametric heteroscedastic regression models with a general change in the unknown regression function where both errors and covariates are allowed to be time series, and on the other hand by investigating the autoregressive case. The achieved rate of convergence for the proposed estimator of $O_P(n^{-1})$ is optimal as described in Hariz et al. (2007).

The remainder of the paper is organized as follows. The model and the considered estimator are introduced in Sect. 2. Section 3 contains the regularity assumptions as well as the asymptotic results for the proposed estimator. Section 4 is concerned with the special case of lagged dependent covariates, that is the autoregressive case. In Sect. 5 we describe a simulation study and discuss a real data example, whereas Sect. 6 concludes the paper. Proofs of the main results as well as auxiliary lemmata can be found in the Appendix.

2 The model and estimator

Let $\{(Y_t, \boldsymbol{X}_t) : t \in \mathbb{N}\}$ be a weakly dependent stochastic process in $\mathbb{R} \times \mathbb{R}^d$ following the regression model

$$Y_t = m_t(\boldsymbol{X}_t) + U_t,\ t \in \mathbb{N}. \tag{2.1}$$

The unobservable innovations are assumed to fulfill $E[U_t|\mathcal{F}^t] = 0$ almost surely for the sigma-field $\mathcal{F}^t = \sigma(U_{j-1}, \boldsymbol{X}_j : j \leq t)$. We assume there exists a change point in the regression function such that

$$m_{n,t}(\cdot) = m_t(\cdot) = \begin{cases} m_{(1)}(\cdot), & t = 1, \ldots, \lfloor ns_0 \rfloor \\ m_{(2)}(\cdot), & t = \lfloor ns_0 \rfloor + 1, \ldots, n \end{cases}, \qquad m_{(1)} \not\equiv m_{(2)} \tag{2.2}$$

where $\lfloor ns_0 \rfloor$ with $s_0 \in (0, 1)$ is the unknown time the change occurs. Note that we keep above notations for simplicity reasons, however, the considered process is in fact a triangular array process $\{(Y_{n,t}, \boldsymbol{X}_{n,t}) : 1 \leq t \leq n, n \in \mathbb{N}\}$ and will be treated appropriately.

Assuming $(Y_1, \boldsymbol{X}_1), \ldots, (Y_n, \boldsymbol{X}_n)$ have been observed, the aim is to estimate s_0. The idea is to base the estimator on the *sequential marked empirical process of residuals*, namely

$$\hat{T}_n(s, \boldsymbol{z}) := \frac{1}{n} \sum_{i=1}^{\lfloor ns \rfloor} (Y_i - \hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\},$$

for $s \in [0, 1]$ and $\boldsymbol{z} \in \mathbb{R}^d$, where $\boldsymbol{x} \leq \boldsymbol{y}$ is short for $x_j \leq y_j$ for all $j = 1, \ldots, d$, $\omega_n(\cdot) = I\{\cdot \in \boldsymbol{J}_n\}$ being from assumption (V) below and $\hat{m}_n$ being the Nadaraya-

Watson estimator, that is

$$\hat{m}_n(\boldsymbol{x}) = \frac{\sum_{j=1}^n K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_n}\right) Y_j}{\sum_{j=1}^n K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_n}\right)},$$

with kernel function K and bandwidth h_n as considered in the assumptions below. Then we want to estimate s_0 by

$$\hat{s}_n := \min\left\{ s : \sup_{z\in\mathbb{R}^d} |\hat{T}_n(s,z)| = \sup_{\bar{s}\in[0,1]} \sup_{z\in\mathbb{R}^d} |\hat{T}_n(\bar{s},z)| \right\}. \tag{2.3}$$

Note that $\hat{s}_n = \lfloor n\hat{s}_n \rfloor / n$.

Remark The advantage of using marked residuals in comparison to using the classical CUSUM $\hat{T}_n(s,\infty)$ to estimate the change point is that in the first case the estimator is consistent for all changes of the form (2.2) whereas there are several examples in which the use of $\hat{T}_n(s,\infty)$ leads to a non-consistent estimator. To this end see the remark below the proof of Theorem 3.1 and compare to Mohr and Neumeyer (2019).

Remark Mohr and Neumeyer (2019) constructed procedures based on functionals of $\hat{T}_n$, e.g. a Kolmogorov-Smirnov test statistic $\sup_{s\in[0,1]}\sup_{z\in\mathbb{R}^d}|\hat{T}_n(s,z)|$, to test the null hypothesis of no changes in the unknown regression function against change point alternatives as in (2.2). Given that such a test has rejected the null, the use of an M-estimator as in (2.3) seems natural. Furthermore, Cramér-von Mises type test statistics of the form $\sup_{s\in[0,1]}\int_{\mathbb{R}^d}|\hat{T}_n(s,z)|^2\nu(z)dz$ for some integrable $\nu:\mathbb{R}^d\to\mathbb{R}$ were also considered by Mohr and Neumeyer (2019). Assuming strict stationarity of the covariates and the existence of a density f such that $\boldsymbol{X}_t \sim f$ for all t, as in (IX.1) below, the Cramér-von Mises approach from above with $\nu \equiv f$ leads to an alternative estimator for s_0, namely

$$\tilde{s}_n := \min\left\{ s : \left(\int_{\mathbb{R}^d} |\hat{T}_n(s,z)|^2 f(z)dz\right)^{1/2} = \sup_{\bar{s}\in[0,1]} \left(\int_{\mathbb{R}^d} |\hat{T}_n(\bar{s},z)|^2 f(z)dz\right)^{1/2} \right\}.$$

However, to obtain a feasible estimator one needs to replace the integral $\int_{\mathbb{R}^d}|\hat{T}_n(s,z)|^2 f(z)dz$ by its empirical counterpart $\frac{1}{n}\sum_{k=1}^n|\hat{T}_n(s,\boldsymbol{X}_k)|^2$ in practise as f is not known.

3 Asymptotic results

In this section we will derive asymptotic properties for $\hat{s}_n$. To this end we introduce the following assumptions.

(I) For all $t\in\mathbb{Z}$ let $E[U_t|\mathcal{F}^t]=0$ a.s. for $\mathcal{F}^t=\sigma(U_{j-1},\boldsymbol{X}_j : j\le t)$ and $E[|U_t|^q]\le C_U$ for some $C_U<\infty$ and $q>2$.

(II) For all $t \in \mathbb{Z}$ let $E[|m_{(1)}(\boldsymbol{X}_t) - m_{(2)}(\boldsymbol{X}_t)|^r] \leq C_m$ for some $C_m < \infty$ and $r > 2$.

(III) Let $\{(Y_t, \boldsymbol{X}_t) : 1 \leq t \leq n, n \in \mathbb{N}\}$ be strongly mixing with mixing coefficient $\alpha(\cdot)$. For q, r from assumptions (I) and (II) and $b := \min(q, r)$ let $\alpha(t) = O(t^{-\bar{\alpha}})$ with some $\bar{\alpha} > \max\big((1 + (b-1)(1+d))/(b-2), (b+2)/(b-2)\big)$.

(IV) For b from assumption (III) let $E[|Y_t|^b] < \infty$ and let $\boldsymbol{X}_t$ be absolutely continuous with density function $f_t : \mathbb{R}^d \to \mathbb{R}$ that satisfies $\sup_{\boldsymbol{x} \in \mathbb{R}^d} E[|Y_t|^b | \boldsymbol{X}_t = \boldsymbol{x}] f_t(\boldsymbol{x}) < \infty$ and $\sup_{\boldsymbol{x} \in \mathbb{R}^d} f_t(\boldsymbol{x}) < \infty$ for all $t \in \{1, \ldots, n\}$ and $n \in \mathbb{N}$. Let there exist some $L \geq 0$ such that $\sup_{|i-j| \geq L} \sup_{\boldsymbol{x}_i, \boldsymbol{x}_j} E[|Y_i Y_j| | \boldsymbol{X}_i = \boldsymbol{x}_i, \boldsymbol{X}_j = \boldsymbol{x}_j] f_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j) < \infty$ for all $n \in \mathbb{N}$, where f_{ij} is the density function of $(\boldsymbol{X}_i, \boldsymbol{X}_j)$.

(V) Let $(c_n)_{n \in \mathbb{N}}$ be a positive sequence of real valued numbers satisfying $c_n \to \infty$ and $c_n = O((\log n)^{1/d})$ and let $\boldsymbol{J}_n = [-c_n, c_n]^d$.

(VI) For some $C < \infty$ and c_n from assumption (V) let $\boldsymbol{I}_n = [-c_n - Ch_n, c_n + Ch_n]^d$ and let $\delta_n^{-1} = \inf_{\boldsymbol{x} \in \boldsymbol{J}_n} \inf_{1 \leq t \leq n} f_t(\boldsymbol{x}) > 0$ for all $n \in \mathbb{N}$. Further, let for all $n \in \mathbb{N}$

$$p_n = \max_{|\boldsymbol{k}|=1} \sup_{\boldsymbol{x} \in \boldsymbol{I}_n} \sup_{1 \leq t \leq n} |D^{\boldsymbol{k}} f_t(\boldsymbol{x})| < \infty$$
$$0 < q_n = \max_{0 \leq |\boldsymbol{k}| \leq 1} \sup_{\boldsymbol{x} \in \boldsymbol{I}_n} \max_{j=1,2} |D^{\boldsymbol{k}} m_{(j)}(\boldsymbol{x})| < \infty,$$

where $|\boldsymbol{i}| = \sum_{j=1}^d i_j$ and $D^{\boldsymbol{i}} = \frac{\partial^{|\boldsymbol{i}|}}{\partial x_1^{i_1} \ldots \partial x_d^{i_d}}$ for $\boldsymbol{i} = (i_1, \ldots, i_d) \in \mathbb{N}_0^d$.

(VII) Let $K : \mathbb{R}^d \to \mathbb{R}$ be symmetric in each component with $\int_{\mathbb{R}^d} K(\boldsymbol{z}) d\boldsymbol{z} = 1$ and compact support $[-C, C]^d$. Additionally let $|K(\boldsymbol{u})| < \infty$ for all $\boldsymbol{u} \in \mathbb{R}^d$ and $|K(\boldsymbol{u}) - K(\boldsymbol{u'})| \leq \Lambda \|\boldsymbol{u} - \boldsymbol{u'}\|$ for some $\Lambda < \infty$ and for all $\boldsymbol{u}, \boldsymbol{u'} \in \mathbb{R}^d$, where $\|\boldsymbol{x}\| = \max_{i=1,\ldots,d} |x_i|$.

(VIII) With b and $\bar{\alpha}$ from assumption (III) let

$$\frac{\log(n)}{n^\theta h_n^d} = o(1) \text{ for } \theta = \frac{\bar{\alpha} - 1 - d - \frac{1+\bar{\alpha}}{b-1}}{\bar{\alpha} + 3 - d - \frac{1+\bar{\alpha}}{b-1}}.$$

For δ_n, p_n, q_n from assumption (VI) let

$$\left(\sqrt{\frac{\log(n)}{nh_n^d}} + h_n p_n\right) p_n q_n \delta_n = o(n^{-\zeta})$$

for some $\zeta > 0$.

(IX.1) For all $1 \leq t \leq n, n \in \mathbb{N}$ let $f_t(\cdot) = f(\cdot)$, for some density f.

(IX.2) For all $1 \leq t \leq n, n \in \mathbb{N}$ let $f_t(\cdot) = f_{(1)}(\cdot)$ for all $t = 1, \ldots, \lfloor ns_0 \rfloor$ and $f_t(\cdot) = f_{(2)}(\cdot)$ for all $t = \lfloor ns_0 \rfloor + 1, \ldots, n$, for some densities $f_{(1)}, f_{(2)}$.

Remark The assumptions on the error terms and the mixing assumptions particularly allow for conditional heteroscedasticity. Assumptions (I), (II) and (III) are a trade off between the existence of moments and the rate of decay of the mixing coefficient. Assumptions (III), (IV), (VII) and the first part of (VIII) are reproduced from Kristensen (2012). Together with (V) and (VI), they are used to obtain uniform rates of convergence for $\hat{m}_n$ stated in Lemma A.1 in the Appendix. In (IX.1), we assume stationarity of the covariates for the whole observation period, while in the case of (IX.2) we assume stationarity before and right after the change occurs. Nevertheless both assumptions rule out general autoregressive effects such as $\boldsymbol{X}_t = (Y_{t-1}, \ldots, Y_{t-d})$. We will address this issue separately in Sect. 4.

Theorem 3.1 *Assume (I), (II), (III), (IV), (V), (VI), (VII) and (VIII). Furthermore let either (IX.1) or (IX.2) hold. Then the change point estimator $\hat{s}_n$ is consistent, i. e.*

$$|\hat{s}_n - s_0| = o_P(1).$$

Theorem 3.2 *Under the assumptions of Theorem 3.1 for the change point estimator $\hat{s}_n$ it holds that*

$$|\hat{s}_n - s_0| = O_P(r_n^{-1}),$$

where $r_n = n$.

The proofs of the theorems can be found in Appendix A.2. We state both theorems seperately since we need Theorem 3.1 to prove Theorem 3.2.

Remark To obtain the rates of convergence we make use of the fact that $\hat{s}_n$ can be expressed using the sup norm on $l^\infty(\mathbb{R}^d)$, i.e.

$$N : l^\infty(\mathbb{R}^d) \to \mathbb{R},\ g \mapsto N(g) := \sup_{z \in \mathbb{R}^d} |g(z)|,$$

where $l^\infty(\mathbb{R}^d)$ is the space of all uniformly bounded real valued functions on $\mathbb{R}^d$. Note that similarly $\tilde{s}_n$ can be expressed using the $L_2(P)$ norm, when $(\boldsymbol{X}_t)_t$ is strictly stationary with marginal distribution P, namely

$$\tilde{N} : l^\infty(\mathbb{R}^d) \to \mathbb{R},\ g \mapsto \tilde{N}(g) := \left(\int_{\mathbb{R}^d} |g(z)|^2 f(z) dz \right)^{1/2}.$$

Using $\tilde{N}(g) \le N(g)$ for all $g \in l^\infty(\mathbb{R}^d)$, corresponding results for $\tilde{s}_n$ as in Theorem 3.1 and Theorem 3.2 can be proven in a similar matter.

4 The autoregressive case

In this section we will consider the case where the exogenous variables include finitely many lagged values of the endogenous variable, we will refer to this model as the *autoregressive case*. We will focus on one dimensional covariates, however, the results do not depend on the dimension and can also be formulated for higher order autoregression models. Consider the nonparametric autoregression

$$Y_t = m_t(Y_{t-1}) + U_t,\ t = 1, \dots, n, \tag{4.1}$$

with unobservable innovations U_t and one change in the regression function occurring at some unknown time $\lfloor ns_0 \rfloor$ as in (2.2).

Furthermore assume the following.

(IX.3) For all $1 \le t \le n, n \in \mathbb{N}$ let $X_t := Y_{t-1}$ be absolutely continuous with density f_t. Let there exist densities $f_{(1)}$ and $f_{(2)}$ such that $f_t(\cdot) = f_{(1)}(\cdot)$ for all $t = 1, \dots, \lfloor ns_0 \rfloor$ and $R_n(x) := \frac{1}{n}\sum_{j=\lfloor ns_0 \rfloor+1}^{n} f_j(x) - \frac{n-\lfloor ns_0 \rfloor}{n} f_{(2)}(x) \to 0$ for all $x \in \mathbb{R}$ and $n \to \infty$.

Remark Note that (IX.3) requires on the one hand strict stationarity up to the time of change $\lfloor ns_0 \rfloor$. On the other hand the time series needs to reach its (new) stationary distribution fast enough after the change. This is a generalization of (IX.2) where we assumed stationarity both before and right after the change point, which can not be fulfilled in the model (4.1). A necessary condition then is that there exists a stationary solution of equation (4.1) under both $m_{(1)}(\cdot)$ and $m_{(2)}(\cdot)$ as regression functions.

Example Consider the AR(1)-model

$$Y_t = a_t \cdot Y_{t-1} + \varepsilon_t$$

with standard normally distributed innovations $(\varepsilon_t)_t$ and $a_t = a \in (-1, 1)$ for $t \le \lfloor ns_0 \rfloor$, $a_t = b \in (-1, 1)$ for $t > \lfloor ns_0 \rfloor$, $a \neq b$. Then assumption (IX.3) is fulfilled. Note to this end that $X_t := Y_{t-1} \sim \mathcal{N}(0, 1/(1-a^2))$ for $t \le \lfloor ns_0 \rfloor$. The distribution after the change point is given by $X_{\lfloor ns_0 \rfloor+1+k} \sim \mathcal{N}\big(0, b^{2(k+1)}/(1-a^2) + \sum_{i=0}^{k} b^{2i}\big)$ for all $k > 0$. Thus with $\sigma_j^2 := b^{2(j-\lfloor ns_0 \rfloor)}/(1-a^2) + \sum_{i=0}^{j-\lfloor ns_0 \rfloor-1} b^{2i}$ by the mean value theorem it holds for some ξ_j between σ_j^2 and $(1-b^2)^{-1}$ that

$$\begin{aligned} R_n(x) &= \frac{1}{n}\sum_{j=\lfloor ns_0 \rfloor+1}^{n}\left(\frac{1}{\sqrt{2\pi\sigma_j^2}}\exp\left(-\frac{x^2}{2\sigma_j^2}\right) - \frac{1}{\sqrt{2\pi(1-b^2)^{-1}}}\exp\left(-\frac{x^2}{2(1-b^2)^{-1}}\right)\right) \\ &= \frac{1}{n}\sum_{j=\lfloor ns_0 \rfloor+1}^{n}\left(\sigma_j^2 - \frac{1}{1-b^2}\right)\exp\left(-\frac{x^2}{2\xi_j}\right)\left(-\frac{1}{2}\cdot\frac{1}{\sqrt{2\pi}\xi_j^{\frac{3}{2}}} + \frac{1}{(2\pi\xi_j)^{\frac{1}{2}}}\cdot\frac{x^2}{2\xi_j^2}\right) \\ &\le C\frac{1}{n}\sum_{j=\lfloor ns_0 \rfloor+1}^{n}\left|\sigma_j^2 - \frac{1}{1-b^2}\right| \end{aligned}$$

for some constant $C < \infty$ for all $x \in \mathbb{R}$. Further we can conclude

$$\frac{1}{n}\sum_{j=\lfloor ns_0\rfloor+1}^{n}\left|\sigma_j^2-\frac{1}{1-b^2}\right| = \frac{1}{n}\sum_{j=\lfloor ns_0\rfloor+1}^{n}\left|\frac{b^{2(j-\lfloor ns_0\rfloor)}}{1-a^2}+\frac{1-b^{2(j-\lfloor ns_0\rfloor)}}{1-b^2}-\frac{1}{1-b^2}\right|$$
$$= \left|\frac{1}{1-a^2}-\frac{1}{1-b^2}\right|\frac{1}{n}\sum_{j=\lfloor ns_0\rfloor+1}^{n} b^{2(j-\lfloor ns_0\rfloor)}$$

and thus $R_n(x) \xrightarrow[n\to\infty]{} 0$ for all $x \in \mathbb{R}$.

In general verifying assumption (IX.3) for model (4.1) means to compare the distribution of a stochastic process that is not yet in balance with its stationary distribution. A well known technique to deal with this task is coupling, see e. g. Franke et al. (2002).

Under (IX.3) we get the following consistency result for our change point estimator in the autoregressive case.

Theorem 4.1 *Assume model* (4.1) *under (I), (II), (III), (IV), (V), (VI), (VII), (VIII) and (IX.3). Then the change point estimator $\hat{s}_n$ is consistent, i. e.*

$$|\hat{s}_n - s_0| = o_P(1).$$

The proof can be found in Appendix A.2.

Remark Another possibility to handle the autoregressive case would be to model the change in a different way, namely

$$Y_t = \begin{cases} Y_t^{(1)} = m_{(1)}\big(Y_{t-1}^{(1)}\big) + U_t^{(1)}, & t = 1, \ldots, \lfloor ns_0\rfloor \\ Y_t^{(2)} = m_{(2)}\big(Y_{t-1}^{(2)}\big) + U_t^{(2)}, & t = \lfloor ns_0\rfloor + 1, \ldots, n \end{cases}, \qquad m_{(1)} \not\equiv m_{(2)},$$

for two stationary processes $\big(Y_t^{(1)}\big)_t$, $\big(Y_t^{(2)}\big)_t$, see e. g. Kirch et al. (2015). In this case assumption (IX.2) is fulfilled and thus Theorems 3.1 and 3.2 apply.

5 Finite sample properties

5.1 Simulations

To investigate the finite sample performance of our estimator, we generate data from two different basic models, namely

(IID) $Y_t = m_t(X_t) + \sigma(X_t)\varepsilon_t$, where the observations $(X_t)_t$ are i.i.d., univariate and standard normally distributed, just as the errors $(\varepsilon_t)_t$.

(TS) $Y_t = m_t(X_t) + \sigma(X_t)\varepsilon_t$, where $(\varepsilon_t)_t$ i.i.d. $\sim N(0, 1)$ and the univariate observations $(X_t)_t$ stem from a time series $X_t = 0.4X_{t-1} + \eta_t$ with standard normal innovations $(\eta_t)_t$.

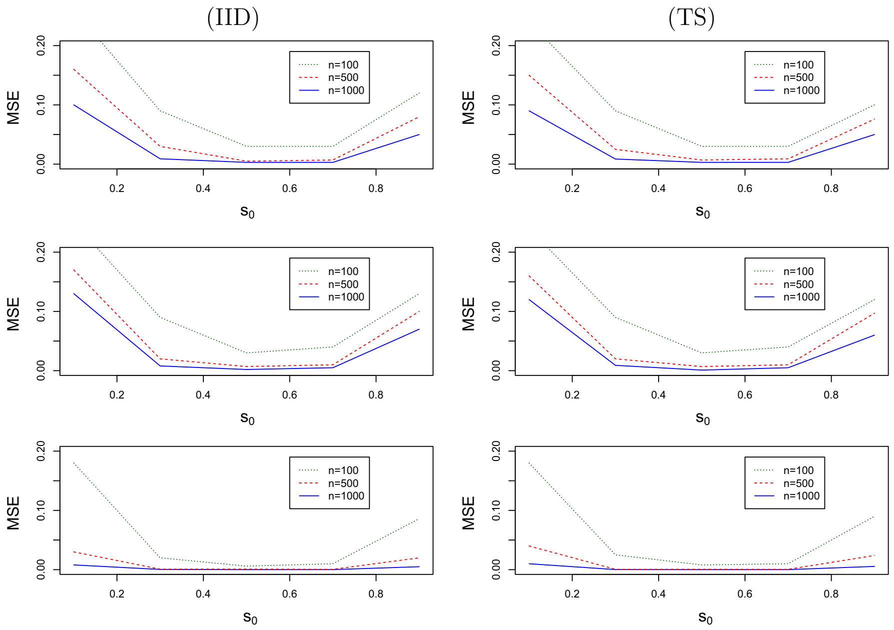

Fig. 1 Simulation results for model (IID) (left), model (TS) (right) and change point scenario (C1) (top), change point scenario (C2) (middle), change point scenario (C3) (bottom)

For both models we generate data both for the homoscedastic case $\sigma \equiv 1$ as well as for the heteroscedastic case $\sigma(x) = \sqrt{1 + 0.5x^2}$. The results for both are very similar in all situations, thus we only present the results for the heteroscedastic case. To model the change in the regression function we use three different scenarios

$$\text{(C1)}\ \ m_t = \begin{cases} -0.5x, & t = 1, \ldots, \lfloor ns_0 \rfloor \\ 0.5x & t = \lfloor ns_0 \rfloor + 1, \ldots, n, \end{cases}$$

$$\text{(C2)}\ \ m_t = \begin{cases} 0.1x, & t = 1, \ldots, \lfloor ns_0 \rfloor \\ 0.9x & t = \lfloor ns_0 \rfloor + 1, \ldots, n, \end{cases}$$

$$\text{(C3)}\ \ m_t = \begin{cases} 0.5x, & t = 1, \ldots, \lfloor ns_0 \rfloor \\ (0.5 + 3\exp(-0.8x^2))x & t = \lfloor ns_0 \rfloor + 1, \ldots, n, \end{cases}$$

where we let s_0 range from 0.1 to 0.9. In Fig. 1 the results for 1000 replications and sample sizes $n = 100, 500, 1000$ are shown, where we plot s_0 against the estimated mean squared error of our estimator $\hat{s}_n$. The kernel for $\hat{m}_n$ is chosen as the Epanechnikov kernel of order four and the bandwidth is determined by a cross-validation method. It can be seen that our estimator performs quite well even for the smallest sample size $n = 100$ when s_0 is 0.5 or close to it whereas for a change point that lies closer to the boundaries of the observation interval a larger sample size is needed to get satisfying results. This is due to the fact that if $s_0 = 0.1$ or $s_0 = 0.9$ there are only 10 observations before and after the change point respectively for $n = 100$ and thus the estimation of $m_{(1)}$ and $m_{(2)}$ respectively are poor. Moreover an asymmetry

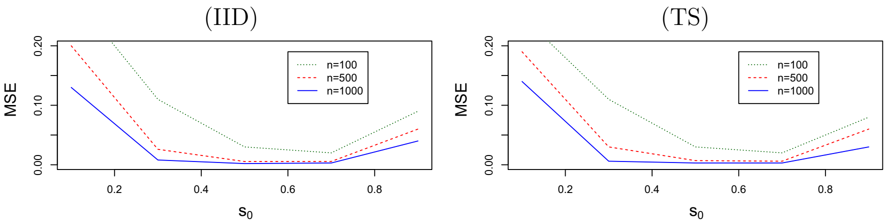

Fig. 2 Simulation results for model (IID) (left) and model (TS) (right) with change point scenario (C1) and an additional change in the variance function

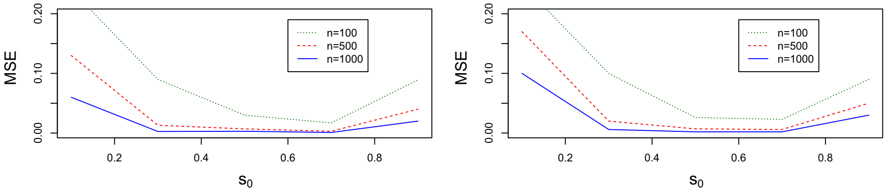

Fig. 3 Simulation results for model (AR) and change point scenario (C1) (left), change point scenario (C2) (right)

in the results is striking. This stems from the CUSUM type statistic that our estimator is based on. For $s_0 = 0.1$ e. g. the sum consists of only $0.1n$ summands and thus the estimation of e. g. $E[U_t]$ is worse than if $s_0 = 0.9$ and the estimation is based on $0.9n$ summands. The effect of a decreasing performance of the estimators the closer s_0 gets to the boundaries is typical for change point estimators based on CUSUM statistics and can be antagonized by the use of appropriate weights, see e. g. Ferger (2005).

To stress our estimator a little further we simulate the scenario that there is also a change in the variance function σ at a different time point than the change in the regression function m. In this situation the estimator should still be able to detect s_0, the change point in the regression function. The results are shown in Fig. 2 for model (IID) and model (TS) with change point scenario (C1) where $\sigma_t(x) = \sqrt{1 + 0.1x^2}$ for $t \leq 0.4n$ and $\sigma_t(x) = \sqrt{1 + 0.8x^2}$ for $t > 0.4n$. They confirm the good performance of our estimator even in this more difficult situation.

As discussed in Sect. 4 our estimator can also be applied to the autoregressive case. To investigate the finite sample performance in this situation we generate data according to the model

$$\text{(AR)} \quad Y_t = m_t(Y_{t-1}) + \sigma(Y_{t-1})\varepsilon_t, \text{ where } (\varepsilon_t)_t \text{ i.i.d.} \sim N(0, 1).$$

For $\sigma \equiv 1$ and change point scenario (C1) as well as (C2) assumption (IX.3) is fulfilled, see the example in Sect. 4. Simulation results for these cases are shown in Fig. 3 where the setting is the same as described above. They look very similar to the results of model (IID) and (TS) and thus confirm the theoretical result of Theorem 4.1. Even for examples where assumption (IX.3) can not be verified easily the performance of our estimator is satisfying, see Fig. 4 for model (AR) with $\sigma \equiv 1$ and change point scenario (C3) as well as the heteroscedastic model (AR) with $\sigma = \sqrt{1 + 0.5x^2}$ and change point scenario (C1).

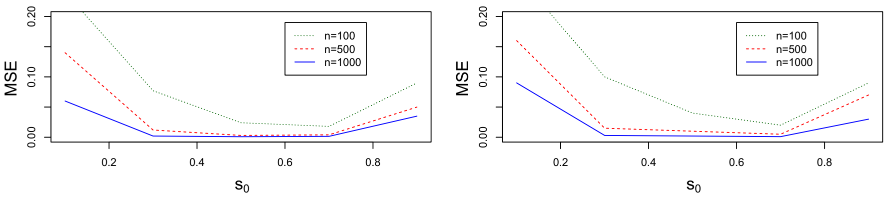

Fig. 4 Simulation results for homoscedastic model (AR) with change point scenario (C3) (left) and heteroscedastic model (AR) with change point scenario (C1) (right)

As stated in the remark in Sect. 2 it is also possible to base the estimator on a Cramér-von Mises type functional of the marked empirical process of residuals. The simulation results for this type of estimator are very similar to those presented here for the Kolmogorov-Smirnov type estimator $\hat{s}_n$ and are omitted for the sake of brevity.

5.2 Data example

Finally, we will consider a real data example. The data at hand contains 36 measurements of the annual flow volume of the small Czech river, Ráztoka, recorded between 1954 and 1989 as well as the annual rainfall during that time. It was considered by Hušková and Antoch (2003) to investigate the effect of controlled deforestation on the capability for water retention of the soil. To this end it is of interest if and when the relationship between rainfall and flow volume changes. We set X_t as the annual rainfall and Y_t as the annual flow volume. Mohr and Neumeyer (2019) applied their Kolmogorov-Smirnov test to this data set, which clearly rejects the null of no change in the conditional mean function, indicating the existence of a change in the relationship between rainfall and flow volume. Using $\hat{s}_n$ to estimate the unknown time of change suggests a change in 1979. Note that this is consistent with the literature. As was pointed out by Hušková and Antoch (2003) large scale deforestation had started around that time. Figure 5 shows on the left-hand side the scatterplot X_t against Y_t using dots for the observations after the estimated change and crosses for the observations before the estimated change. On the right-hand side the figure shows the cumulative sum, $n^{1/2} \sup_{z\in\mathbb{R}} |\hat{T}_n(\cdot, z)|$, as well as the critical value of the test used in Mohr and Neumeyer (2019) (red horizontal line) and the estimated change (green vertical line). Note that $\tilde{s}_n$ leads to the same result.

6 Concluding remarks

In this paper we consider nonparametric regression models with a change in the unknown regression function that allows for time series data as well as conditional heteroscedasticity. We propose an estimator for the rescaled change point that is based on the sequential marked empirical process of residuals and show consistency as well as a rate of convergence of $O_P(n^{-1})$. In an autoregressive setting we additionally give a consistency result for the proposed estimator.

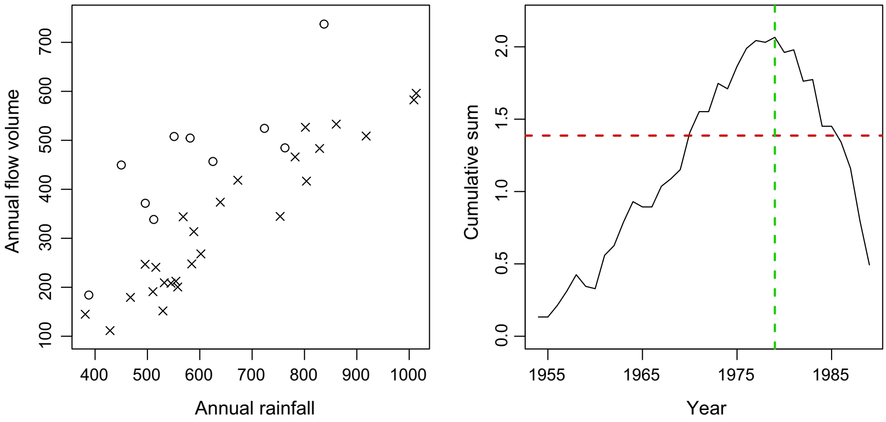

Fig. 5 Ráztoka data: scatterplot (left) and CUSUM (right)

If more than one change occurs, the proposed estimator is not consistent for one of the changes in some situations. For detecting multiple changes we refer the reader to alternative procedures such as the MOSUM procedure proposed by Eichinger and Kirch (2018) or the wild binary segmentation procedure by Fryzlewicz (2014) (see also Fryzlewicz (2019)).

Investigating the asymptotic distribution of the proposed estimator is a subsequent issue. Certainly, it is of great interest as it can be used to obtain confidence intervals. However, this subject goes beyond the scope of the paper at hand and is postponed to future research.

Acknowledgements Open Access funding provided by Projekt DEAL. Financial support by the DFG (Research Unit FOR 1735 *Structural Inference in Statistics: Adaptation and Efficiency*) is gratefully acknowledged.

Appendix: Proofs

A.1 Auxiliary results

Lemma A.1 *Under the assumptions (III), (IV), (V), (VI), (VII) and (VIII), it holds that*

$$\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} |\hat{m}_n(\boldsymbol{x}) - \bar{m}_n(\boldsymbol{x})| = O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^d}} + h_n p_n\right)\delta_n p_n q_n\right),$$

where

$$\bar{m}_n(\boldsymbol{x}) = \frac{\sum_{i=1}^n f_i(\boldsymbol{x}) m_i(\boldsymbol{x})}{\sum_{i=1}^n f_i(\boldsymbol{x})}.$$

The proof is similar to the proof of Lemma 2.2 in Mohr (2018). The key tool is an application of Theorem 1 in Kristensen (2009). Details are omitted for the sake of brevity.

Remark Under (IX.1) we have

$$\bar{m}_n(\boldsymbol{x}) = \frac{\lfloor ns_0 \rfloor}{n} m_{(1)}(\boldsymbol{x}) + \frac{n - \lfloor ns_0 \rfloor}{n} m_{(2)}(\boldsymbol{x}),$$

under (IX.2) and (IX.3) we have

$$\bar{m}_n(\boldsymbol{x}) = \frac{\frac{\lfloor ns_0 \rfloor}{n} f_{(1)}(\boldsymbol{x})}{\bar{f}_n(\boldsymbol{x})} (m_{(1)}(\boldsymbol{x}) - m_{(2)}(\boldsymbol{x})) + m_{(2)}(\boldsymbol{x}),$$

where

$$\bar{f}_n(\boldsymbol{x}) := \frac{1}{n} \sum_{i=1}^n f_i(\boldsymbol{x}) = \begin{cases} \frac{\lfloor ns_0 \rfloor}{n} f_{(1)}(\boldsymbol{x}) + \frac{n - \lfloor ns_0 \rfloor}{n} f_{(2)}(\boldsymbol{x}), & \text{for (IX.2)} \\ \frac{\lfloor ns_0 \rfloor}{n} f_{(1)}(\boldsymbol{x}) + \frac{n - \lfloor ns_0 \rfloor}{n} f_{(2)}(\boldsymbol{x}) + R_n(\boldsymbol{x}), & \text{for (IX.3)} \end{cases}$$

with $R_n(\cdot)$ from assumption (IX.3).

Lemma A.2 *Under the assumptions of Theorem 3.1 as well as under those of Theorem 4.1 there exists a constant $\bar{C} = \bar{C}(C) < \infty$ such that*

$$P\left(\sup_{s \in [0,1]} \sup_{z \in \mathbb{R}^d} \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s \rfloor} U_i \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \le z\} \right| > C\kappa_n \right) \le \bar{C} \kappa_n^{\frac{1}{q}-1}$$

for all $L = 0, 1, \ldots, n - \kappa_n$, $1 \le \kappa_n \le n$, $n \in \mathbb{N}$ and all $C > 0$ with q from assumption (I).

Proof of Lemma A.2 The proof follows along similar lines as the proof of Lemma A.3 in Mohr (2018). Throughout the proof the values of C and $\bar{C}$ may vary from line to line but they are always positive, finite and independent of n. Further note that deterministic terms that are of order $O(\kappa_n)$ can be omitted as we can choose constants appropriately. It holds that

$$\begin{aligned} &\sup_{s \in [0,1]} \sup_{z \in \mathbb{R}^d} \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s \rfloor} U_i \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \le z\} \right| \\ &= \sup_{s \in [0,1]} \sup_{z \in \mathbb{R}^d} \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s \rfloor} U_i \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \le z\} - E\left[\sum_{i=L+1}^{L+\lfloor \kappa_n s \rfloor} U_i \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \le z\} \right] \right| \end{aligned}$$

$$\leq \sup_{s\in[0,1]}\sup_{z\in\mathbb{R}^d}\left|\sum_{i=L+1}^{L+\lfloor\kappa_n s\rfloor}U_i I\{|U_i|>\kappa_n^{\frac{1}{q}}\}\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq\boldsymbol{z}\}-E\left[\sum_{i=L+1}^{L+\lfloor\kappa_n s\rfloor}U_i I\{|U_i|>\kappa_n^{\frac{1}{q}}\}\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq\boldsymbol{z}\}\right]\right| \tag{A.1}$$

$$+\sup_{s\in[0,1]}\sup_{z\in\mathbb{R}^d}\left|\sum_{i=L+1}^{L+\lfloor\kappa_n s\rfloor}U_i I\{|U_i|\leq\kappa_n^{\frac{1}{q}}\}\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq\boldsymbol{z}\}-E\left[\sum_{i=L+1}^{L+\lfloor\kappa_n s\rfloor}U_i I\{|U_i|\leq\kappa_n^{\frac{1}{q}}\}\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq\boldsymbol{z}\}\right]\right| \tag{A.2}$$

where (A.1) is of the desired rate in probability since

$$P\left(\sum_{i=L+1}^{L+\kappa_n}|U_i|I\{|U_i|>\kappa_n^{\frac{1}{q}}\}>C\kappa_n\right)\leq C^{-1}C_U\kappa_n^{\frac{1}{q}-1}$$

by Markov's inequality with

$$\begin{aligned}E\left[|U_i|I\{|U_i|>\kappa_n^{\frac{1}{q}}\}\right]&=E\left[|U_i|^q|U_i|^{-(q-1)}I\{|U_i|>\kappa_n^{\frac{1}{q}}\}\right]\\&\leq\kappa_n^{-\frac{q-1}{q}}E[|U_i|^q]\\&\leq C_U\kappa_n^{\frac{1}{q}-1}\qquad\text{for all } i \text{ and for } C_U<\infty \text{ from assumption (I).}\end{aligned}$$

Considering the term (A.2) we define the function class

$$\mathcal{F}_n:=\left\{(u,\boldsymbol{x})\mapsto uI\{|u|\leq\kappa_n^{\frac{1}{q}}\}\omega_n(\boldsymbol{x})I\{\boldsymbol{x}\leq\boldsymbol{z}\}:\boldsymbol{z}\in\mathbb{R}^d\right\}$$

to rewrite the assertion as

$$P\left(\sup_{s\in[0,1]}\sup_{\varphi\in\mathcal{F}_n}\left|\sum_{i=L+1}^{L+\lfloor\kappa_n s\rfloor}\left(\varphi(U_i,\boldsymbol{X}_i)-\int\varphi\, dP\right)\right|>C\kappa_n\right)\leq\bar{C}\kappa_n^{\frac{1}{q}-1}.$$

Now we will cover [0, 1] by finitely many intervals and $\mathcal{F}_n$ by finitely many brackets to replace the supremum by a maximum. Let therefore

$$0=s_1<\cdots<s_{K_n}=1$$

part the interval [0, 1] in K_n subintervals of length $\bar{\epsilon}_n$ with $\bar{\epsilon}_n = \kappa_n^{-\frac{1}{q}}$. Then

$$
\begin{aligned}
&\sup_{s\in[0,1]} \sup_{\varphi\in\mathcal{F}_n} \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s\rfloor} \left(\varphi(U_i, \boldsymbol{X}_i) - \int \varphi dP \right) \right| \\
&= \max_k \sup_{\substack{s\in[0,1] \\ |s-s_k|\le\bar{\epsilon}_n}} \sup_{\varphi\in\mathcal{F}_n} \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s\rfloor} \left(\varphi(U_i, \boldsymbol{X}_i) - \int \varphi dP \right) \right| \\
&\le \max_k \sup_{\varphi\in\mathcal{F}_n} \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s_k\rfloor} \left(\varphi(U_i, \boldsymbol{X}_i) - \int \varphi dP \right) \right| \\
&\quad + \max_k \sup_{\substack{s\in[0,1] \\ |s-s_k|\le\bar{\epsilon}_n}} \sup_{\varphi\in\mathcal{F}_n} \sum_{i=L+1}^{L+\kappa_n} \underbrace{\left| \varphi(U_i, \boldsymbol{X}_i) - \int \varphi dP \right|}_{\le 2\kappa_n^{\frac{1}{q}}} \left| I\left\{ \frac{i-L}{\kappa_n} \le s \right\} \right. \\
&\quad \left. - I\left\{ \frac{i-L}{\kappa_n} \le s_k \right\} \right| \\
&\le \max_k \sup_{\varphi\in\mathcal{F}_n} \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s_k\rfloor} \left(\varphi(U_i, \boldsymbol{X}_i) - \int \varphi dP \right) \right| + 2\kappa_n^{\frac{1}{q}} \left(\kappa_n \bar{\epsilon}_n + 1 \right)
\end{aligned}
$$

and $2\kappa_n^{\frac{1}{q}} (\kappa_n\bar{\epsilon}_n + 1) = 2(\kappa_n + \kappa_n^{\frac{1}{q}}) = O(\kappa_n)$. Further let

$$
\begin{aligned}
\varphi_{\boldsymbol{j}}^u(u, \boldsymbol{x}) &:= uI\{|u| \le \kappa_n^{\frac{1}{q}}\} I\{u \ge 0\} \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z_j}\} \\
&\quad + uI\{|u| \le \kappa_n^{\frac{1}{q}}\} I\{u < 0\} \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z_{j-1}}\}
\end{aligned}
$$

and

$$
\begin{aligned}
\varphi_{\boldsymbol{j}}^l(u, \boldsymbol{x}) &:= uI\{|u| \le \kappa_n^{\frac{1}{q}}\} I\{u \ge 0\} \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z_{j-1}}\} \\
&\quad + uI\{|u| \le \kappa_n^{\frac{1}{q}}\} I\{u < 0\} \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z_j}\}
\end{aligned}
$$

form the brackets $[\varphi_{\boldsymbol{j}}^l, \varphi_{\boldsymbol{j}}^u]_{\boldsymbol{j}\in\times_{i=1}^d\{1,\dots,N_i\}}$ of $\mathcal{F}_n$, where $\boldsymbol{z_j} := (z_{j_1,1}, \dots, z_{j_d,d})$ and

$$
-\infty = z_{0,i} < \cdots < z_{N_i,i} = \infty
$$

gives a partition of $\mathbb{R}$ for all $i = 1, \dots, d$. Then for all $\boldsymbol{z} \in \mathbb{R}^d$ there exists a $\boldsymbol{j} \in \times_{i=1}^d\{1, \dots, N_i\}$ such that $\boldsymbol{z_{j-1}} < \boldsymbol{z} \le \boldsymbol{z_j}$ where $\boldsymbol{j-1} := (j_1 - 1, \dots, j_d - 1)$. Thus every element φ of $\mathcal{F}_n$ lies in one of the brackets $[\varphi_{\boldsymbol{j}}^l, \varphi_{\boldsymbol{j}}^u]$, i.e. $\varphi_{\boldsymbol{j}}^l(u, \boldsymbol{x}) \le \varphi(u, \boldsymbol{x}) \le \varphi_{\boldsymbol{j}}^u(u, \boldsymbol{x})$ for all $(u, \boldsymbol{x})$. We say a bracket $[\varphi_{\boldsymbol{j}}^l, \varphi_{\boldsymbol{j}}^u]$ is of size ϵ_n if $\int(\varphi_{\boldsymbol{j}}^u - \varphi_{\boldsymbol{j}}^l) dP \le \epsilon_n$. The total number of brackets of size ϵ_n needed to cover $\mathcal{F}_n$ is denoted by $J_n :=$

$N_{[\,]}(\epsilon_n, \mathcal{F}_n, \|\cdot\|_{L_1(P)})$ and is of order $J_n = O(\epsilon_n^{-d})$, which follows analogously to but easier than the proof of Lemma A.7 in Mohr (2018).

For all $\varphi \in \mathcal{F}_n$ there exists a $\boldsymbol{j}$ with $\varphi^l_{\boldsymbol{j}} \le \varphi \le \varphi^u_{\boldsymbol{j}}$ and thus

$$\varphi - \int \varphi dP \le \varphi^u_{\boldsymbol{j}} - \int \varphi^u_{\boldsymbol{j}} dP + \int (\varphi^u_{\boldsymbol{j}} - \varphi^l_{\boldsymbol{j}}) dP$$

and

$$\varphi - \int \varphi dP \ge \varphi^l_{\boldsymbol{j}} - \int \varphi^l_{\boldsymbol{j}} dP - \int (\varphi^u_{\boldsymbol{j}} - \varphi^l_{\boldsymbol{j}}) dP.$$

Therefore for all $s \in [0, 1]$

$$\begin{aligned}
&\sup_{\varphi\in\mathcal{F}_n} \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s\rfloor} \left(\varphi(U_i, \boldsymbol{X}_i) - \int \varphi dP \right) \right| \\
&= \max_{\boldsymbol{j}} \sup_{\varphi\in[\varphi^l_{\boldsymbol{j}}, \varphi^u_{\boldsymbol{j}}]} \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s\rfloor} \left(\varphi(U_i, \boldsymbol{X}_i) - \int \varphi dP \right) \right| \\
&\le \max_{\boldsymbol{j}} \max \left\{ \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s\rfloor} \left(\varphi^u_{\boldsymbol{j}}(U_i, \boldsymbol{X}_i) - \int \varphi^u_{\boldsymbol{j}} dP \right) \right|, \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s\rfloor} \left(\varphi^l_{\boldsymbol{j}}(U_i, \boldsymbol{X}_i) \right.\right.\right. \\
&\quad \left.\left.\left. - \int \varphi^l_{\boldsymbol{j}} dP \right) \right| \right\} \\
&\quad + \kappa_n \max_{\boldsymbol{j}} \underbrace{\int (\varphi^u_{\boldsymbol{j}} - \varphi^l_{\boldsymbol{j}}) dP}_{\le \epsilon_n}
\end{aligned}$$

and $\kappa_n \epsilon_n = O(\kappa_n)$ if we choose ϵ_n constant. Thus it remains to show that

$$P\left(\max_{\boldsymbol{j},k} \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s_k\rfloor} \left(\varphi^u_{\boldsymbol{j}}(U_i, \boldsymbol{X}_i) - \int \varphi^u_{\boldsymbol{j}} dP \right) \right| > C\kappa_n \right) \le \bar{C} \kappa_n^{\frac{1}{q}-1}$$

and the same with $\varphi^u_{\boldsymbol{j}}$ replaced by $\varphi^l_{\boldsymbol{j}}$. Recall that

$$\begin{aligned}
&\max_{\boldsymbol{j},k} \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s_k\rfloor} \left(\varphi^u_{\boldsymbol{j}}(U_i, \boldsymbol{X}_i) - \int \varphi^u_{\boldsymbol{j}} dP \right) \right| \\
&\le \max_{\boldsymbol{j},k} \left| \sum_{i=L+1}^{L+\lfloor \kappa_n s_k\rfloor} \left(U_i I\{|U_i| \le \kappa_n^{\frac{1}{q}}\} I\{U_i \ge 0\} \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \le \boldsymbol{z}_{\boldsymbol{j}}\} \right.\right. \\
&\quad \left.\left. - E\left[U_i I\{|U_i| \le \kappa_n^{\frac{1}{q}}\} I\{U_i \ge 0\} \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \le \boldsymbol{z}_{\boldsymbol{j}}\} \right] \right) \right| \qquad \text{(A.3)}
\end{aligned}$$

$$+\max_{\boldsymbol{j},k}\Bigg|\sum_{i=L+1}^{L+\lfloor\kappa_n s_k\rfloor}\Big(U_i I\{|U_i|\le\kappa_n^{\frac{1}{q}}\}I\{U_i<0\}\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}_{\boldsymbol{j-1}}\}$$
$$-E\left[U_i I\{|U_i|\le\kappa_n^{\frac{1}{q}}\}I\{U_i<0\}\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}_{\boldsymbol{j-1}}\}\right]\Big)\Bigg|.$$

We will only consider the first summand in more detail since the rest works analogously. To prove that (A.3) is stochastically of the desired rate we apply a Bernstein type inequality for α-mixing processes, see Liebscher (1996) Therorem 2.1. Following his notation we define

$$Z_i:=\Big(U_{i+L}I\{|U_{i+L}|\le\kappa_n^{\frac{1}{q}}\}I\{U_{i+L}\ge 0\}\omega_n(\boldsymbol{X}_{i+L})I\{\boldsymbol{X}_{i+L}\le\boldsymbol{z}\}$$
$$-E\left[U_{i+L}I\{|U_{i+L}|\le\kappa_n^{\frac{1}{q}}\}I\{U_{i+L}\ge 0\}\omega_n(\boldsymbol{X}_{i+L})I\{\boldsymbol{X}_{i+L}\le\boldsymbol{z}\}\right]\Big)I\left\{\frac{i}{\kappa_n}\le s_k\right\}$$

for fixed $\boldsymbol{z}\in\mathbb{R}^d$ and $s\in[0,1]$. Note that $|Z_i|\le 2\kappa_n^{\frac{1}{q}}=:S(\kappa_n)$, Z_i is centered and

$$D(\kappa_n,N):=\sup_{0\le T\le\kappa_n-1}E\left[\left(\sum_{j=T+1}^{(T+N)\wedge\kappa_n}Z_j\right)^2\right]\le N^2E[Z_i^2]\le C_U N^2$$

by assumption (I). Thus Liebscher's Theorem can be applied with $N=\lfloor\kappa_n^{1-\frac{2}{q}}\rfloor$. This means that

$$P\Bigg(\max_{\boldsymbol{j},k}\Bigg|\sum_{i=L+1}^{L+\lfloor\kappa_n s_k\rfloor}U_iI\{|U_i|\le\kappa_n^{\frac{1}{q}}\}I\{U_i\ge 0\}\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}_{\boldsymbol{j}}\}$$
$$-E\left[U_iI\{|U_i|\le\kappa_n^{\frac{1}{q}}\}I\{U_i\ge 0\}\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}_{\boldsymbol{j}}\}\right]\Bigg|>C\kappa_n\Bigg)$$
$$\le\sum_{\boldsymbol{j},k}P\Bigg(\Bigg|\sum_{i=L+1}^{L+\lfloor\kappa_n s_k\rfloor}\Big(U_iI\{|U_i|\le\kappa_n^{\frac{1}{q}}\}I\{U_i\ge 0\}\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}_{\boldsymbol{j}}\}$$
$$-E\left[U_iI\{|U_i|\le\kappa_n^{\frac{1}{q}}\}I\{U_i\ge 0\}\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}_{\boldsymbol{j}}\}\right]\Big)\Bigg|>C\kappa_n\Bigg)$$
$$\le J_nK_n\left(4\exp\left(-\frac{C^2\kappa_n^2}{64\frac{\kappa_n}{N}D(\kappa_n,N)+\frac{8}{3}C\kappa_nNS(\kappa_n)}\right)+4\frac{\kappa_n}{N}\alpha(N)\right)$$
$$\le J_nK_n\left(4\exp\left(-\frac{C^2\kappa_n^2}{64C_U\kappa_n^{2-\frac{2}{q}}+\frac{16}{3}C\kappa_n^{2-\frac{1}{q}}}\right)+4\kappa_n^{\frac{2}{q}}\alpha(\kappa_n^{1-\frac{2}{q}})\right)$$
$$\le J_nK_n\left(4\exp\left(-C_1\kappa_n^{\frac{1}{q}}\right)+4\kappa_n^{\frac{2}{q}}\alpha(\kappa_n^{1-\frac{2}{q}})\right)$$

$$\leq C_2\kappa_n^{\frac{1}{q}}\left((C_1\kappa_n^{\frac{1}{q}})^{-q}+\kappa_n^{\frac{2}{q}-\bar{\alpha}+\frac{2\bar{\alpha}}{q}}\right)$$
$$\leq \bar{C}\kappa_n^{\frac{1}{q}-1}$$

for some constants C_1, C_2, $\bar{C}$ where the second to last inequality follows from the fact that $\exp(-x) < x^{-k}k!$ for all $k \in \mathbb{N}$ and $x \in \mathbb{R}_+$ and the last inequality is true by assumption (III) which implies $\bar{\alpha} > (q+2)/(q-2)$. This completes the proof. □

Lemma A.3 *Under the assumptions of Theorem 3.1 as well as under those of Theorem 4.1 there exists a constant $\bar{C} = \bar{C}(C) < \infty$ such that*

$$P\Bigg(\sup_{s\in[0,1]}\sup_{z\in\mathbb{R}^d}\Bigg|\sum_{i=L+1}^{(L+\lfloor\kappa_n s\rfloor)\wedge\lfloor ns_0\rfloor}\Big((m_{(1)}(\boldsymbol{X}_i)-\bar{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq z\}$$
$$-E\left[(m_{(1)}(\boldsymbol{X}_i)-\bar{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq z\}\right]\Big)\Bigg| > C\kappa_n\Bigg) \leq \bar{C}\kappa_n^{\frac{1}{r}-1}$$

and

$$P\Bigg(\sup_{s\in[0,1]}\sup_{z\in\mathbb{R}^d}\Bigg|\sum_{i=L\vee\lfloor ns_0\rfloor+1}^{L+\lfloor\kappa_n s\rfloor}\Big((m_{(2)}(\boldsymbol{X}_i)-\bar{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq z\}$$
$$-E\left[(m_{(2)}(\boldsymbol{X}_i)-\bar{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq z\}\right]\Big)\Bigg| > C\kappa_n\Bigg) \leq \bar{C}\kappa_n^{\frac{1}{r}-1}$$

for all $L = 0, 1, \ldots, n-\kappa_n$, $1 \leq \kappa_n \leq n$, $n \in \mathbb{N}$ and all $C > 0$ with r from assumption (II).

Proof of Lemma A.3 First we will distinguish between the cases $L + \lfloor\kappa_n s\rfloor \leq \lfloor ns_0\rfloor$ and $L + \lfloor\kappa_n s\rfloor > \lfloor ns_0\rfloor$. In the first case we can write

$$\sum_{i=L+1}^{L+\lfloor\kappa_n s\rfloor}(m_{(1)}(\boldsymbol{X}_i)-\bar{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq z\}$$
$$=\sum_{i=L+1}^{L+\lfloor\kappa_n s\rfloor}\left(m_{(1)}(\boldsymbol{X}_i)-\frac{m_{(1)}(\boldsymbol{X}_i)\sum_{j=1}^{\lfloor ns_0\rfloor}f_j(\boldsymbol{X}_i)}{\sum_{j=1}^{n}f_j(\boldsymbol{X}_i)}-\frac{m_{(2)}(\boldsymbol{X}_i)\sum_{j=\lfloor ns_0\rfloor+1}^{n}f_j(\boldsymbol{X}_i)}{\sum_{j=1}^{n}f_j(\boldsymbol{X}_i)}\right)$$
$$\cdot\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq z\}$$
$$=\sum_{i=L+1}^{L+\lfloor\kappa_n s\rfloor}(m_{(1)}(\boldsymbol{X}_i)-m_{(2)}(\boldsymbol{X}_i))\frac{\sum_{j=\lfloor ns_0\rfloor+1}^{n}f_j(\boldsymbol{X}_i)}{\sum_{j=1}^{n}f_j(\boldsymbol{X}_i)}\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq z\}$$

and analogously for the second case

$$\sum_{i=L\vee\lfloor ns_0\rfloor+1}^{L+\lfloor\kappa_n s\rfloor}(m_{(2)}(\boldsymbol{X}_i)-\bar{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le \boldsymbol{z}\}$$
$$=\sum_{i=L\vee\lfloor ns_0\rfloor+1}^{L+\lfloor\kappa_n s\rfloor}(m_{(2)}(\boldsymbol{X}_i)-m_{(1)}(\boldsymbol{X}_i))\frac{\sum_{j=1}^{\lfloor ns_0\rfloor} f_j(\boldsymbol{X}_i)}{\sum_{j=1}^{n} f_j(\boldsymbol{X}_i)}\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le \boldsymbol{z}\}.$$

We will only examine the case $L+\lfloor\kappa_n s\rfloor\le\lfloor ns_0\rfloor$ in detail since the other case works analogously.

The remainder of the proof is similar to the proof of Lemma A.2. With $g(\boldsymbol{X}_i):=(m_{(1)}(\boldsymbol{X}_i)-m_{(2)}(\boldsymbol{X}_i))$ and $\bar{f}_n^{(s_0)}(\boldsymbol{X}_i)=\frac{\sum_{j=\lfloor ns_0\rfloor+1}^{n} f_j(\boldsymbol{X}_i)}{\sum_{j=1}^{n} f_j(\boldsymbol{X}_i)}$ it holds

$$\sup_{s\in[0,1]}\sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\sum_{i=L+1}^{L+\lfloor\kappa_n s\rfloor} g(\boldsymbol{X}_i)I\{|g(\boldsymbol{X}_i)|>\kappa_n^{\frac{1}{r}}\}\bar{f}_n^{(s_0)}(\boldsymbol{X}_i)\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le \boldsymbol{z}\}\right|$$
$$\le\sum_{i=L+1}^{L+\kappa_n}|g(\boldsymbol{X}_i)|I\{|g(\boldsymbol{X}_i)|>\kappa_n^{\frac{1}{r}}\}$$

and further

$$P\left(\sum_{i=L+1}^{L+\kappa_n}|g(\boldsymbol{X}_i)|I\{|g(\boldsymbol{X}_i)|>\kappa_n^{\frac{1}{r}}\}>C\kappa_n\right)\le C^{-1}\kappa_n^{-1}C_m\kappa_n\kappa_n^{\frac{1}{r}-1}$$

by the Markov inequality with

$$\begin{aligned}E\left[|g(\boldsymbol{X}_i)|I\{|g(\boldsymbol{X}_i)|>\kappa_n^{\frac{1}{r}}\}\right]&=E\left[|g(\boldsymbol{X}_i)|^r|g(\boldsymbol{X}_i)|^{-(r-1)}I\{|g(\boldsymbol{X}_i)|>\kappa_n^{\frac{1}{r}}\}\right]\\&\le\kappa_n^{-\frac{r-1}{r}}E[|g(\boldsymbol{X}_i)|^r]\\&\le C_m\kappa_n^{\frac{1}{r}-1}\end{aligned}$$

for all i and for some $C_m<\infty$ by assumption (II). Thus we can rewrite our assertion as

$$P\left(\sup_{s\in[0,1]}\sup_{\varphi\in\mathcal{F}_n}\left|\left(\sum_{i=L+1}^{L+\lfloor\kappa_n s\rfloor}\varphi(\boldsymbol{X}_i)-\int\varphi\,dP\right)\right|>C\kappa_n\right)\le\bar{C}\kappa_n^{\frac{1}{r}-1},$$

with the function class

$$\mathcal{F}_n:=\left\{\boldsymbol{x}\mapsto g(\boldsymbol{x})I\{|g(\boldsymbol{x})|\le\kappa_n^{\frac{1}{r}}\}\bar{f}_n^{(s_0)}(\boldsymbol{x})\omega_n(\boldsymbol{x})I\{\boldsymbol{x}\le \boldsymbol{z}\}:\boldsymbol{z}\in\mathbb{R}^d\right\}.$$

To replace the supremum over φ by a maximum we cover $\mathcal{F}_n$ by finitely many brackets $[\varphi^l_{\boldsymbol{j}}, \varphi^u_{\boldsymbol{j}}]_{\boldsymbol{j}\in\times_{i=1}^d\{1,\dots,N_i\}}$ where

$$\varphi^u_{\boldsymbol{j}}(\boldsymbol{x}) := g(\boldsymbol{x})I\{|g(\boldsymbol{x})| \le \kappa_n^{\frac{1}{r}}\}I\{g(\boldsymbol{x}) \ge 0\}\bar{f}_n^{(s_0)}(\boldsymbol{x})\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z_j}\} \\ + g(\boldsymbol{x})I\{|g(\boldsymbol{x})| \le \kappa_n^{\frac{1}{r}}\}I\{g(\boldsymbol{x}) < 0\}\bar{f}_n^{(s_0)}(\boldsymbol{x})\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z_{j-1}}\}$$

and

$$\varphi^u_{\boldsymbol{j}}(\boldsymbol{x}) := g(\boldsymbol{x})I\{|g(\boldsymbol{x})| \le \kappa_n^{\frac{1}{r}}\}I\{g(\boldsymbol{x}) \ge 0\}\bar{f}_n^{(s_0)}(\boldsymbol{x})\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z_{j-1}}\} \\ + g(\boldsymbol{x})I\{|g(\boldsymbol{x})| \le \kappa_n^{\frac{1}{r}}\}I\{g(\boldsymbol{x}) < 0\}\bar{f}_n^{(s_0)}(\boldsymbol{x})\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z_j}\}$$

and $\boldsymbol{j}$, $\boldsymbol{z_j}$ are defined as in the proof of Lemma A.2. The total number of brackets $J_n := N_{[\,]}(\epsilon_n, \mathcal{F}_n, \|\cdot\|_{L_1(P)})$ needed to cover $\mathcal{F}_n$ is again of order $J_n = O(\epsilon_n^{-d})$, which follows analogously to but easier than the proof of Lemma A.7 in Mohr (2018). Now we proceed completely analogously to the proof of Lemma A.2 by replacing the supremum over s by a maximum as well and applying Liebscher's Theorem. Since the arguments are the same as in the aforementioned proof we omit this part for the sake of brevity. □

Lemma A.4 *Under the assumptions of Theorem 3.1 as well as under those of Theorem 4.1 it holds*

$$P\left(\sup_{s\in[0,1]}\sup_{z\in\mathbb{R}^d}\left|\sum_{i=L+1}^{L+\lfloor\kappa_n s\rfloor}(\bar{m}_n(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \le \boldsymbol{z}\}\right| > C\kappa_n\right) \le C^{-1}\kappa_n^{-\zeta}$$

for all $L = 0, 1, \dots, n-\kappa_n$, $1 \le \kappa_n \le n$, $n \in \mathbb{N}$ and all $C > 0$ with $\zeta > 0$ from assumption (VIII).

Proof of Lemma A.4 It holds

$$P\left(\sup_{s\in[0,1]}\sup_{z\in\mathbb{R}^d}\left|\sum_{i=L+1}^{L+\lfloor\kappa_n s\rfloor}(\bar{m}_n(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \le \boldsymbol{z}\}\right| > C\kappa_n\right)$$
$$\le P\left(\sum_{i=L+1}^{L+\kappa_n}|\bar{m}_n(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{X}_i)|\omega_n(\boldsymbol{X}_i) > C\kappa_n\right)$$
$$\le P\left(\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}|\bar{m}_n(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{x})| > C\right)$$
$$\le C^{-1}E[\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}|\bar{m}_n(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{x})|]$$

by the Markov inequality. Further by Lemma A.1 with assumption (VIII) it holds that

$$\frac{\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} |\bar{m}_n(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{x})|}{n^{-\zeta}} \xrightarrow[n\to\infty]{P} 0$$

which implies

$$\frac{E[\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} |\bar{m}_n(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{x})|]}{n^{-\zeta}} \xrightarrow[n\to\infty]{} 0$$

and thus for sufficiently large n

$$\begin{aligned} E[\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} |\bar{m}_n(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{x})|] &\leq n^{-\zeta} \\ &\leq \kappa_n^{-\zeta} \end{aligned}$$

for $\kappa_n \leq n$. This completes the proof. □

A.2 Proof of main results

We will proof Theorem 3.1 under the assumption (IX.1) and simply make a note on the parts that change under (IX.2).

Proof of Theorem 3.1 First note that for all $s \in [0, 1]$ and $z \in \mathbb{R}^d$

$$\hat{T}_n(s, z) = A_n(s, z) + \Delta_{n,1}(s)\Delta_{n,2}(z), \tag{A.4}$$

where $A_n(s, z) = A_{n,1}(s, z) + A_{n,2}(s, z) + A_{n,3}(s, z) + A_{n,4}(s, z)$ with

$$A_{n,1}(s, z) := \frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor} U_i \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq z\} \tag{A.5}$$

$$\begin{aligned} A_{n,2}(s, z) := \frac{1}{n}\sum_{i=1}^{\lfloor n(s\wedge s_0)\rfloor} \Big(&(m_{(1)}(\boldsymbol{X}_i) - \bar{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq z\} \\ &- E\left[(m_{(1)}(\boldsymbol{X}_i) - \bar{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq z\}\right]\Big) \end{aligned} \tag{A.6}$$

$$\begin{aligned} A_{n,3}(s, z) := I\{s > s_0\}\frac{1}{n}\sum_{i=\lfloor ns_0\rfloor+1}^{\lfloor ns\rfloor} \Big(&(m_{(2)}(\boldsymbol{X}_i) - \bar{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq z\} \\ &- E\left[(m_{(2)}(\boldsymbol{X}_i) - \bar{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq z\}\right]\Big) \end{aligned} \tag{A.7}$$

$$A_{n,4}(s, z) := \frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor} (\bar{m}_n(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq z\} \tag{A.8}$$

and

$$\Delta_{n,1}(s) := I\{s \leq s_0\}\frac{n-\lfloor ns_0\rfloor}{n}\frac{\lfloor ns\rfloor}{n} + I\{s > s_0\}\frac{n-\lfloor ns\rfloor}{n}\frac{\lfloor ns_0\rfloor}{n}$$
$$\Delta_{n,2}(\boldsymbol{z}) := \int_{(-\infty,\boldsymbol{z}]}(m_{(1)}(\boldsymbol{x}) - m_{(2)}(\boldsymbol{x}))f(\boldsymbol{x})\omega_n(\boldsymbol{x})d\boldsymbol{x},$$

since by inserting the definition of $\bar{m}_n$ we obtain for $s \leq s_0$

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor} E\left[(m_{(1)}(\boldsymbol{X}_i) - \bar{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}\right]\\
&\quad= \frac{n-\lfloor ns_0\rfloor}{n}\frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor} E\left[(m_{(1)}(\boldsymbol{X}_i) - m_{(2)}(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}\right]\\
&\quad= \frac{n-\lfloor ns_0\rfloor}{n}\frac{\lfloor ns\rfloor}{n}\Delta_{n,2}(\boldsymbol{z})
\end{aligned}$$

and for $s > s_0$

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{\lfloor ns_0\rfloor} E\left[(m_{(1)}(\boldsymbol{X}_i) - \bar{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}\right]\\
&\qquad+ \frac{1}{n}\sum_{i=\lfloor ns_0\rfloor+1}^{\lfloor ns\rfloor} E\left[(m_{(2)}(\boldsymbol{X}_i) - \bar{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}\right]\\
&= \frac{n-\lfloor ns_0\rfloor}{n}\frac{1}{n}\sum_{i=1}^{\lfloor ns_0\rfloor} E\left[(m_{(1)}(\boldsymbol{X}_i) - m_{(2)}(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}\right]\\
&\qquad- \frac{\lfloor ns_0\rfloor}{n}\frac{1}{n}\sum_{i=\lfloor ns_0\rfloor+1}^{\lfloor ns\rfloor} E\left[(m_{(1)}(\boldsymbol{X}_i) - m_{(2)}(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}\right]\\
&= \frac{n-\lfloor ns\rfloor}{n}\frac{\lfloor ns_0\rfloor}{n}\Delta_{n,2}(\boldsymbol{z}).
\end{aligned}$$

Note that we use the notation $\int_{(\infty,\boldsymbol{z}]} g(\boldsymbol{x})d\boldsymbol{x} = \int_{-\infty}^{z_d}\cdots\int_{-\infty}^{z_1} g(x_1,\ldots,x_d)dx_1\ldots dx_d$ here. Due to the dominated convergence theorem and assumption (II), it holds that

$$\Delta_{n,1}(s)\Delta_{n,2}(\boldsymbol{z}) = \Delta_1(s)\Delta_2(\boldsymbol{z}) + o(1),$$

uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$, where

$$\Delta_1(s) := I\{s \leq s_0\}(1-s_0)s + I\{s > s_0\}(1-s)s_0,$$
$$\Delta_2(\boldsymbol{z}) := \int_{(-\infty,\boldsymbol{z}]}(m_{(1)}(\boldsymbol{x}) - m_{(2)}(\boldsymbol{x}))f(\boldsymbol{x})d\boldsymbol{x}.$$

Note that under (IX.2) the same assertion holds with

$$\Delta_{n,2}(\boldsymbol{z}) := \int_{(-\infty,z]} (m_{(1)}(\boldsymbol{x}) - m_{(2)}(\boldsymbol{x})) \frac{f_{(1)}(\boldsymbol{x}) f_{(2)}(\boldsymbol{x})}{\frac{\lfloor ns_0 \rfloor}{n} f_{(1)}(\boldsymbol{x}) + \frac{n-\lfloor ns_0 \rfloor}{n} f_{(2)}(\boldsymbol{x})} \omega_n(\boldsymbol{x}) d\boldsymbol{x}$$

and

$$\Delta_2(\boldsymbol{z}) := \int_{(-\infty,z]} (m_{(1)}(\boldsymbol{x}) - m_{(2)}(\boldsymbol{x})) \frac{f_{(1)}(\boldsymbol{x}) f_{(2)}(\boldsymbol{x})}{s_0 f_{(1)}(\boldsymbol{x}) + (1 - s_0) f_{(2)}(\boldsymbol{x})} d\boldsymbol{x}.$$

By Lemmata A.2, A.3 and A.4 with $\kappa_n = n$, it holds that $A_n(s, z) = o_P(1)$ uniformly in $s \in [0, 1]$ and $z \in \mathbb{R}^d$. Hence, we have shown that

$$\sup_{z \in \mathbb{R}^d} |\hat{T}_n(s, z)| = \Delta_1(s) \sup_{z \in \mathbb{R}^d} |\Delta_2(z)| + o_P(1)$$

uniformly in $s \in [0, 1]$ under both cases (IX.1) and (IX.2). The assertion then follows by Theorem 2.12 in Kosorok (2008) as s_0 is well-separated maximum of $[0, 1] \to \mathbb{R},\ s \mapsto \Delta_1(s)$. □

Remark Note that there are examples of $m_{(1)}, m_{(2)}$ and f resp. $f_{(1)}, f_{(2)}$ that lead to $\Delta_2(\infty) = 0$. In those cases a change point estimator based on the classical CUSUM $\hat{T}_n(s, \infty)$ is not consistent.

Proof of Theorem 3.2 First note that $s_0 = \frac{\lfloor ns_0 \rfloor}{n} + O(n^{-1})$ and $\hat{s}_n = \frac{\lfloor n\hat{s}_n \rfloor}{n}$. Thus we can consider $\left| \frac{\lfloor n\hat{s}_n \rfloor}{n} - \frac{\lfloor ns_0 \rfloor}{n} \right|$ instead of $|\hat{s}_n - s_0|$. The proof follows mainly along the same lines as the proof of Theorem 1 in Hariz et al. (2007). Consider the norm $N : l^\infty(\mathbb{R}^d) \to \mathbb{R},\ g \mapsto \sup_{z \in \mathbb{R}^d} |g(z)|$ and let $M > 0$. We will show below that for all $\eta > 0$ and $b, c > 0$ it holds

$$\begin{aligned} P\left(r_n \left| \frac{\lfloor n\hat{s}_n \rfloor}{n} - \frac{\lfloor ns_0 \rfloor}{n} \right| > 2^M \right) &= P\left(r_n^{-1} 2^M < \left| \frac{\lfloor n\hat{s}_n \rfloor}{n} - \frac{\lfloor ns_0 \rfloor}{n} \right| \le \eta \right) \\ &\quad + P\left(\left| \frac{\lfloor n\hat{s}_n \rfloor}{n} - \frac{\lfloor ns_0 \rfloor}{n} \right| > \eta \right) \\ &\le E_{n,1} + E_{n,2} + E_{n,3} + E_{n,4}, \end{aligned} \tag{A.9}$$

where

$$\begin{aligned} E_{n,1} &:= P\Bigg(r_n^{-1} 2^M < \left| \frac{\lfloor n\hat{s}_n \rfloor}{n} - \frac{\lfloor ns_0 \rfloor}{n} \right| \le \eta, \\ &\qquad N(A_n(\hat{s}_n, \cdot) - A_n(s_0, \cdot)) \ge C \left| \frac{\lfloor n\hat{s}_n \rfloor}{n} - \frac{\lfloor ns_0 \rfloor}{n} \right| \Bigg) \\ E_{n,2} &:= P(N(A_n(s_0, \cdot)) > c) \end{aligned}$$

$$E_{n,3} := P(\Delta_{n,1}(s_0)N(\Delta_{n,2}(\cdot)) \leq b)$$
$$E_{n,4} := P\left(|\hat{s}_n - s_0| > \eta\right),$$

with $C := b - 2c$. Now it holds that $E_{n,4} \to 0$ for all $\eta > 0$, due to Theorem 3.1. Further, $E_{n,2} \to 0$ for all $c > 0$ as $A_n(s_0, z) = o_P(1)$ holds uniformly in $z \in \mathbb{R}^d$. Finally choose $b > 0$ and $n' = n'(b) \in \mathbb{N}$ such that $E_{n,3} = 0$ for all $n \geq n'$, which exists as $\Delta_1(s_0)N(\Delta_2(\cdot)) > 0$ and $\Delta_{n,1}(s_0)N(\Delta_{n,2}(\cdot)) = \Delta_1(s_0)N(\Delta_2(\cdot)) + o(1)$. We then choose $c > 0$ such that $b - 2c > 0$. To see the validity of (A.9) first note that for all $s \in [0, 1]$

$$\begin{aligned}\hat{T}_n(s, \cdot) &= A_n(s, \cdot) + \Delta_{n,1}(s)\Delta_{n,2}(\cdot)\\ &= A_n(s, \cdot) - A_n(s_0, \cdot) + A_n(s_0, \cdot)\left(1 - \frac{\Delta_{n,1}(s)}{\Delta_{n,1}(s_0)}\right) + \frac{\Delta_{n,1}(s)}{\Delta_{n,1}(s_0)}\hat{T}_n(s_0, \cdot).\end{aligned}$$

Applying the norm and triangular inequality we obtain for all $s \in [0, 1]$

$$\begin{aligned}N(\hat{T}_n(s, \cdot)) \leq N(A_n(s, \cdot) - A_n(s_0, \cdot)) + \left(1 - \frac{\Delta_{n,1}(s)}{\Delta_{n,1}(s_0)}\right)N(A_n(s_0, \cdot))\\ + \left(\frac{\Delta_{n,1}(s)}{\Delta_{n,1}(s_0)}\right)N(\hat{T}_n(s_0, \cdot))\end{aligned}$$

which is equivalent to

$$\begin{aligned}N(\hat{T}_n(s, \cdot)) - N(\hat{T}_n(s_0, \cdot)) \leq N(A_n(s, \cdot) - A_n(s_0, \cdot))\\ + \left(\frac{\Delta_{n,1}(s)}{\Delta_{n,1}(s_0)} - 1\right)\left(N(\hat{T}_n(s_0, \cdot)) - N(A_n(s_0, \cdot))\right).\end{aligned}$$

Due to the definition of $\hat{s}_n$ it holds that $N(\hat{T}_n(\hat{s}_n, \cdot)) - N(\hat{T}_n(s_0, \cdot)) \geq 0$. Additionally using the specific definition of $\Delta_{n,1}$ we obtain

$$\begin{aligned}&N(A_n(\hat{s}_n, \cdot) - A_n(s_0, \cdot))\\ &\geq \left(1 - \frac{\Delta_{n,1}(\hat{s}_n)}{\Delta_{n,1}(s_0)}\right)\left(N(\hat{T}_n(s_0, \cdot)) - N(A_n(s_0, \cdot))\right)\\ &\geq \underbrace{\min\left(\frac{n}{\lfloor ns_0\rfloor}, \frac{n}{n - \lfloor ns_0\rfloor}\right)}_{>1}\left|\frac{\lfloor n\hat{s}_n\rfloor}{n} - \frac{\lfloor ns_0\rfloor}{n}\right|\left(N(\hat{T}_n(s_0, \cdot)) - N(A_n(s_0, \cdot))\right)\\ &\geq \left|\frac{\lfloor n\hat{s}_n\rfloor}{n} - \frac{\lfloor ns_0\rfloor}{n}\right|\left(\Delta_{n,1}(s_0)N(\Delta_{n,2}(\cdot)) - 2N(A_n(s_0, \cdot))\right),\end{aligned}$$

where we again make use of the triangular inequality in the last step. Putting the results together we obtain

$$
\begin{aligned}
&P\left(r_n^{-1}2^M < \left|\frac{\lfloor n\hat{s}_n\rfloor}{n} - \frac{\lfloor ns_0\rfloor}{n}\right| \leq \eta\right)\\
&\leq P\left(r_n^{-1}2^M < \left|\frac{\lfloor n\hat{s}_n\rfloor}{n} - \frac{\lfloor ns_0\rfloor}{n}\right| \leq \eta,\ \Delta_{n,1}(s_0)N(\Delta_{n,2}(\cdot)) > b,\ N(A_n(s_0,\cdot)) \leq c\right)\\
&\quad + P(\Delta_{n,1}(s_0)N(\Delta_{n,2}(\cdot)) \leq b) + P(N(A_n(s_0,\cdot)) > c)\\
&\leq P\left(r_n^{-1}2^M < \left|\frac{\lfloor n\hat{s}_n\rfloor}{n} - \frac{\lfloor ns_0\rfloor}{n}\right| \leq \eta,\right.\\
&\quad \left. N(A_n(\hat{s}_n,\cdot) - A_n(s_0,\cdot)) \geq C\left|\frac{\lfloor n\hat{s}_n\rfloor}{n} - \frac{\lfloor ns_0\rfloor}{n}\right|\right)\\
&\quad + P(\Delta_{n,1}(s_0)N(\Delta_{n,2}(\cdot)) \leq b) + P(N(A_n(s_0,\cdot)) > c).
\end{aligned}
$$

Finally we will investigate $E_{n,1}$. To do this we define shells

$$
S_{n,l} = \left\{t \in [0,1] : 2^l < r_n\left|t - \frac{\lfloor ns_0\rfloor}{n}\right| \leq 2^{l+1}\right\}
$$

and choose $L_n = L_n(\eta)$ such that $2^{L_n} < r_n\eta \leq 2^{L_n+1}$ for some $\eta \leq \frac{1}{2}$. Then

$$
\begin{aligned}
E_{n,1} &\leq \sum_{l=M}^{L_n} P\left(\frac{\lfloor n\hat{s}_n\rfloor}{n} \in S_{n,l},\ N(A_n(\hat{s}_n,\cdot) - A_n(s_0,\cdot)) \geq C\left|\frac{\lfloor n\hat{s}_n\rfloor}{n} - \frac{\lfloor ns_0\rfloor}{n}\right|\right)\\
&\leq \sum_{l=M}^{L_n} P\left(\sup_{s:\left|\frac{\lfloor ns\rfloor}{n} - \frac{\lfloor ns_0\rfloor}{n}\right| \leq 2^{l+1}r_n^{-1}} N(A_n(s,\cdot) - A_n(s_0,\cdot)) \geq C2^l r_n^{-1}\right)\\
&\leq \sum_{l=M}^{L_n}\sum_{i=1}^{4} P\left(\sup_{s:\left|\frac{\lfloor ns\rfloor}{n} - \frac{\lfloor ns_0\rfloor}{n}\right| \leq 2^{l+1}r_n^{-1}} N(A_{n,i}(s,\cdot) - A_{n,i}(s_0,\cdot)) \geq \frac{C}{4}2^l r_n^{-1}\right)\\
&\leq \tilde{C}\left(\left(\frac{n}{r_n}\right)^{\frac{1}{q}-1}\sum_{l=M}^{L_n}(2^{\frac{1}{q}-1})^l + \left(\frac{n}{r_n}\right)^{\frac{1}{r}-1}\sum_{l=M}^{L_n}(2^{\frac{1}{r}-1})^l + \left(\frac{n}{r_n}\right)^{-\zeta}\sum_{l=M}^{L_n}(2^{-\zeta})^l\right)
\end{aligned}
$$

for some constant $\tilde{C} < \infty$ by Lemmata A.2, A.3 and A.4 with $\kappa_n = \left\lfloor 2^{l+1}\frac{n}{r_n}\right\rfloor$ with q from assumption (I), r from assumption (II) and $\zeta > 0$ from assumption (VIII). Now choosing $r_n = n$ and letting n and thus L_n tend to infinity and then M to infinity, the assertion of Theorem 3.2 follows. □

Proof of Theorem 4.1 Under (IX.3) we have for all $s \in [0,1]$ and $z \in \mathbb{R}$

$$\hat{T}_n(s,z) = A_n(s,z) + \Delta_{n,1}(s)\Delta_{n,2}(z) + \tilde{\Delta}_n(s,z),$$

with $A_n(s,z)$ and $\Delta_{n,1}(s)$ from the proof of Theorem 3.1, and with

$$\Delta_{n,2}(z) := \int_{(-\infty,z]} (m_{(1)}(x) - m_{(2)}(x)) \frac{f_{(1)}(x) f_{(2)}(x)}{\frac{\lfloor ns_0 \rfloor}{n} f_{(1)}(x) + \frac{n - \lfloor ns_0 \rfloor}{n} f_{(2)}(x) + R_n(x)} \omega_n(x) dx$$

and

$$\tilde{\Delta}_n(s,z) := \int_{(-\infty,z]} (m_{(1)}(x) - m_{(2)}(x)) I\{s \le s_0\} \frac{\lfloor ns \rfloor}{n} \cdot \frac{f_{(1)}(x) R_n(x)}{\frac{\lfloor ns_0 \rfloor}{n} f_{(1)}(x) + \frac{n - \lfloor ns_0 \rfloor}{n} f_{(2)}(x) + R_n(x)} \omega_n(x) dx.$$

Now it holds that

$$\Delta_{n,2}(z) \to \int_{(-\infty,z]} (m_{(1)}(x) - m_{(2)}(x)) \frac{f_{(1)}(x) f_{(2)}(x)}{s_0 f_{(1)}(x) + (1 - s_0) f_{(2)}(x)} dx =: \Delta_2(z)$$

and $\tilde{\Delta}_n(s,z) \to 0$ uniformly in $s \in [0,1]$ and $z \in \mathbb{R}$, due to dominated convergence and assumption (II). Hence we have uniformly in s and z

$$\hat{T}_n(s,z) = A_n(s,z) + \Delta_1(s)\Delta_2(z) + o(1),$$

with $\Delta_1(s)$ as in the proof of Theorem 3.1. The rest goes analogously to the proof of Theorem 3.1. □

Remark Note that for finite $n \in \mathbb{N}$ we do not get the decomposition of $\hat{T}_n$ as in (A.4) in the proof of Theorem 3.1. We only obtain this kind of decomposition when letting n tend to infinity. The decomposition for finite n, however, is essential for the proof of the rates of convergence in Theorem 3.2.

References

Antoch J, Hušková M, Prášková Z (1997) Effect of dependence on statistics for determination of change. J Stat Plan Inference 60:291–310

Aue A, Horváth L, Hušková M (2012) Segmenting mean-nonstationary time series via trending regressions. J Econ 168:367–381

Bai J (1994) Least squares estimation of a shift in linear processes. J Time Ser Anal 15:435–472

Bai J (1997) Estimation of a change point in multiple regression models. Rev Econ Stat 79:551–563

Chen G, Choi YK, Zhou Y (2005) Nonparametric estimation of structural change points in volatility models for time series. J Econ 126:79–114

Chong TT-L (2001) Structural change in AR(1) models. Econ Theory 17:87–155
Delgado MA, Hidalgo J (2000) Nonparametric inference on structural breaks. J Econ 96:113–144
Döring M, Jensen U (2015) Smooth change point estimation in regression models with random design. Ann Inst Stat Math 67:595–619
Eichinger B, Kirch C (2018) A MOSUM procedure for the estimation of multiple random change points. Bernoulli 24:526–564
Ferger D (2005) Weighted least squares estimators for a change-point. Econ Qual Control 20:255–270
Ferger D, Stute W (1992) Convergence of changepoint estimators. Stoch Process Appl 42:345–351
Franke J, Kreiss J-P, Mammen E, Neumann M (2002) Properties of the nonparametric autoregressive bootstrap. J Time Ser Anal 23:555–585
Fryzlewicz P (2014) Wild binary segmentation for multiple change-point detection. Ann Stat 42:2243–2281
Fryzlewicz P (2019) Detecting possibly frequent change-points: wild binary segmentation 2 and steepest-drop model selection. Preprint on arXiv:1812.06880
Hariz SB, Wylie JJ, Zhang Q (2007) Optimal rate of convergence for nonparametric change-point estimators for nonstationary sequences. Ann Stat 35:1802–1826
Horváth L, Hušková M, Serbinowska M (1997) Estimators for the time of change in linear models. Statistics 29:109–130
Hušková M, Antoch J (2003) Detection of structural changes in regression. Tatra Mt Math Publ 26:201–215
Hušková M, Kirch C (2008) Bootstrapping confidence intervals for the change? Point of time series. J Time Ser Anal 29:947–972
Kirch C, Kamgaing JT (2012) Testing for parameter stability in nonlinear autoregressive models. J Time Ser Anal 33:365–385
Kirch C, Muhsal B, Ombao H (2015) Detection of changes in multivariate time series with application to EEG data. J Am Stat Assoc 110:1197–1216
Kosorok MR (2008) Introduction to empirical processes and semiparametric inference. Springer, New York
Kristensen D (2009) Uniform convergence rates of kernel estimators with heterogeneous dependent data. Econ Theory 25:1433–1445
Kristensen D (2012) Non-parametric detection and estimation of structural change. Econ J 15:420–461
Liebscher E (1996) Strong convergence of sums of α-mixing random variables with applications to density estimation. Stoch Process. Appl. 65:69–80
Mohr M (2018) Changepoint detection in a nonparametric time series regression model. PhD thesis, University of Hamburg. http://ediss.sub.uni-hamburg.de/volltexte/2018/9416/
Mohr M, Neumeyer N (2019) Consistent nonparametric change point detection combining CUSUM and marked empirical processes. Preprint on arXiv:1901.08491
Pang T, Zhang D (2015) Asymptotic inferences for an AR(1) model with a change point and possibly infinite variance. Commun Stat Theory Methods 44:4848–4865
Pang T, Zhang D, Chong TT-L (2014) Asymptotic inferences for an AR(1) model with a change point: stationary and nearly non-stationary cases. J Time Ser Anal 35:133–150