Electronic Journal of Statistics
Vol. 14 (2020) 2238–2271
ISSN: 1935-7524
arXiv: 1901.08491

Consistent nonparametric change point detection combining CUSUM and marked empirical processes

Maria Mohr and Natalie Neumeyer

Department of Mathematics, University of Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany, e-mail: maria.mohr@uni-hamburg.de*;* natalie.neumeyer@uni-hamburg.de

Abstract: A weakly dependent time series regression model with multivariate covariates and univariate observations is considered, for which we develop a procedure to detect whether the nonparametric conditional mean function is stable in time against change point alternatives. Our proposal is based on a modified CUSUM type test procedure, which uses a sequential marked empirical process of residuals. We show weak convergence of the considered process to a centered Gaussian process under the null hypothesis of no change in the mean function and a stationarity assumption. This requires some sophisticated arguments for sequential empirical processes of weakly dependent variables. As a consequence we obtain convergence of Kolmogorov-Smirnov and Cramér-von Mises type test statistics. The proposed procedure acquires a very simple limiting distribution and nice consistency properties, features from which related tests are lacking. We moreover suggest a bootstrap version of the procedure and discuss its applicability in the case of unstable variances.

MSC 2010 subject classifications: primary 62M10; secondary 62G08, 62G09, 62G10.
Keywords and phrases: Bootstrap, change point detection, cumulative sums, distribution-free test, heteroscedasticity, kernel estimation, nonparametric regression, sequential empirical process.

Received June 2019.

1. Introduction

Assume a finite sequence $(\boldsymbol{X}_t, Y_t)$, $t = 1, \dots, n$, of a weakly dependent $\mathbb{R}^d \times \mathbb{R}$-valued time series has been observed. Here, we interpret $\boldsymbol{X}_t$ as a covariate (which may contain past values of the process) and it is assumed that the conditional expectation of the observation Y_t, given $\boldsymbol{X}_t$ and all past values of the time series, does only depend on the covariate $\boldsymbol{X}_t$, and thus is a function $m_t(\boldsymbol{X}_t)$. We do not impose any parametric structure on the regression function. For inference on the time series it is of importance whether the regression function is time dependent or not, i.e. the hypothesis

$$H_0: \; m_t(\boldsymbol{X}_t) = m(\boldsymbol{X}_t) \text{ a.s. for all } t = 1, \dots, n$$

(for some not further specified function m) should be tested against structural changes over time such as change point alternatives.

Literature on such tests for nonparametric regression functions is rare in the time series context. Both Hidalgo (1995) and Honda (1997) suggested CUSUM tests for change points in the regression function in nonparametric time series regression models with strictly stationary and absolutely regular data. Su and Xiao (2008) extended these tests to strongly mixing and not necessarily stationary processes, allowing for heteroscedasticity, while Su and White (2010) proposed change point tests in partially linear time series models. Vogt (2015) constructed a kernel-based L_2-test for structural change in the regression function in time-varying nonparametric regression models with locally stationary regressors.

We will combine the CUSUM approach as considered by Hidalgo (1995), Honda (1997), and Su and Xiao (2008) with a marked empirical process approach. Marked empirical processes have been suggested in a seminal paper by Stute (1997) for lack-of-fit testing in nonparametric regression models with i.i.d. data. Since then they have been widely used for hypothesis testing in regression models, see Koul and Stute (1999) and Delgado and Manteiga (2001), among many others. A marked empirical process approach has been applied by Burke and Bewa (2013) for change point detection in an i.i.d. setting. In contrast to our approach they use a process of observations instead of residuals with a very complicated limit distribution, whereas we obtain a simple limit distribution and even asymptotically distribution-free tests in the case of one-dimensional covariates. To this end we show weak convergence of a sequential marked empirical process of residuals under the null hypothesis. We further demonstrate consistency under fixed alternatives of one structural break in the regression function at some time point $\lfloor ns_0 \rfloor$ for $n \to \infty$.

Moreover we suggest a wild bootstrap version of our test that can be applied to detect changes in the mean function in the case of stable variances (as alternative to using the asymptotic distribution, e.g. for multivariate covariates) as well as in the case of non-stable variances. Wild bootstrap was first introduced by Wu (1986) and Liu (1988) for linear regression with heteroscedasticity. It was used in time series context by Kreiß (1997) and Hafner and Herwartz (2000), among others. The bootstrap version of our test can detect changes in the conditional mean function, even when the conditional variance function is also not stable, but – as desired – the test does not react sensitive to the unstable variance. If no change in the mean function is detected, a test for change in the variance function can be applied, which assumes a stable mean function. The latter approach will be considered in detail in a forthcoming manuscript. Most literature assumes stationary variances of the error terms (unconditional or conditional) when testing for changes in regression. However, as Wu (2016) pointed out, non-stationary variances can occur and will most likely result in misleading inferences when not taken into account. Although this is a legitimate concern, not many results are available that deal with non-stationary variances. The CUSUM test by Su and Xiao (2008) allows for breaks in the conditional variance function. But their procedure does only seem to work for fixed breaks in the variance function that do not depend on the sample size, whereas we consider changes of the variance in some $\lfloor nt_0 \rfloor$ for $n \to \infty$. There are some approaches for testing for parameter stability in parametric time series models that consider unstable variances, see Pitarakis (2004), Perron and Zhou (2008), Kristensen (2012), Cai (2007), Xu (2015) and Wu (2016). But all the settings considered do not fit into our framework as they either do not allow for autoregression models, by assuming stationarity of the regressor variables under the null, or they do not cover heteroscedastic effects. More precisely if heteroscedasticity is considered, variance instabilities are not modeled in the conditional variance function but as a time-varying constant.

The paper is organized as follows. In section 2 we present the model and the sequential marked empirical process, on which the test statistics are based. Further the assumptions are listed. In section 3 we consider the limit distribution under the null hypothesis as well as consistency under the fixed alternative of one change point. The wild bootstrap version of the procedure is discussed in section 4, whereas simulations and a real data example are presented in section 5. Section 6 contains concluding remarks, whereas proofs are presented in the appendix.

2. The model and test statistic

Let $(Y_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ be a strongly mixing stochastic process in $\mathbb{R}\times\mathbb{R}^d$ following the regression model

$$Y_t = m_t(\boldsymbol{X}_t) + U_t, \quad t \in \mathbb{Z}.$$

The covariate $\boldsymbol{X}_t$ may include finitely many lagged values of Y_t, for instance $\boldsymbol{X}_t = (Y_{t-1}, \dots, Y_{t-d})$ such that the model includes nonparametric autoregression. The unobservable innovations $(U_t)_{t\in\mathbb{Z}}$ are assumed to fulfill $E[U_t|\mathcal{F}^t] = 0$ almost surely for the sigma-field $\mathcal{F}^t = \sigma(U_{j-1}, \boldsymbol{X}_j : j \le t)$. Our assumptions on the innovations are rather weak; in particular heteroscedastic models will be covered. To derive asymptotic critical values for the test statistic defined below we will assume strict stationarity under the following null hypothesis. Assuming $(Y_1, \boldsymbol{X}_1), \dots, (Y_n, \boldsymbol{X}_n)$ have been observed, our aim is to test the null hypothesis

$$H_0 : m_t(\cdot) = m(\cdot),\ t = 1, \dots, n,$$

for the conditional mean function $E[Y_t|\boldsymbol{X}_t = \boldsymbol{x}] = m_t(\boldsymbol{x})$, $t \in \mathbb{Z}$, and some not specified function $m : \mathbb{R}^d \to \mathbb{R}$ not depending on the time of observation t. To test H_0, we define the *sequential marked empirical process of residuals* as

$$\hat{T}_n(s, \boldsymbol{z}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (Y_i - \hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \le \boldsymbol{z}\}, \tag{2.1}$$

for $s \in [0, 1]$ and $\boldsymbol{z} \in \mathbb{R}^d$, where $\omega_n(\cdot) = I\{\cdot \in \boldsymbol{J}_n\}$ with $\boldsymbol{J}_n$ from assumption **(J)** below. Throughout $I\{\cdot\}$ denotes the indicator function, $\boldsymbol{x} \le \boldsymbol{y}$ is short for $x_j \le y_j,\ \forall\, j = 1, \dots, d$, and we use the notations $\lfloor x \rfloor = \max\{k \in \mathbb{Z} : k \le x\}$ for $x \in \mathbb{R}$ and $\boldsymbol{x} \wedge \boldsymbol{y} = (\min\{x_1, y_1\}, \dots, \min\{x_d, y_d\})$ as well as $\int_{(-\boldsymbol{\infty}, \boldsymbol{x}]} g(\boldsymbol{u}) d\boldsymbol{u} = \int_{-\infty}^{x_d} \cdots \int_{-\infty}^{x_1} g(u_1, \dots, u_d) du_1 \dots du_d$.

The regression function m is estimated by the Nadaraya-Watson estimator $\hat{m}_n$, where

$$\hat{m}_n(\boldsymbol{x}) = \frac{\sum_{j=1}^n K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_n}\right) Y_j}{\sum_{j=1}^n K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_n}\right)} \tag{2.2}$$

with kernel function K and bandwidth h_n as considered in the assumptions below.

The proposed test is a modification of the CUSUM test in Su and Xiao (2008). They consider the process

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (Y_i - \hat{m}_n(\boldsymbol{X}_i)) \hat{f}_n(\boldsymbol{X}_i) w(\boldsymbol{X}_i),$$

where $w : \mathbb{R}^d \to \mathbb{R}$ is a weighting function and $\hat{f}_n$ is the kernel density estimator. While the factor $\hat{f}_n$ has technical reasons as small random values in the denominator of $\hat{m}_n$ can be avoided, the weighting function w plays a crucial role for the power of their test (see remarks to Theorem 3.2 in Su and Xiao (2008)). Depending on the alternative, w needs to be chosen appropriately for the rejection probability to converge to one. Hence, their test (in contrast to the one based on the sequential marked process) is for fixed w not consistent against all alternatives.

Under the null hypothesis H_0 we formulate the following assumptions in order to derive the limiting distribution of $\hat{T}_n$ and corresponding test statistics in the next section.

- **(G)** Let $(Y_t, \boldsymbol{X}_t)_{t \in \mathbb{Z}}$ be strictly stationary and α-mixing with mixing coefficient $\alpha(\cdot)$ such that $\alpha(t) = O(a^{-t})$ for some $a \in (1, \infty)$.
- **(U)** For some $\gamma > 0$ and some even $Q > (d+1)(2+\gamma)$ let $E[U_t | \mathcal{F}^t] = 0$, $E[U_t^2 | \boldsymbol{X}_t] = \sigma^2(\boldsymbol{X}_t)$ and $E[|U_t|^{Q\frac{2+\gamma}{2}} | \boldsymbol{X}_t] \le c(\boldsymbol{X}_t)^Q$ a.s. for all $t \in \mathbb{Z}$, for some $c, \sigma^2 : \mathbb{R}^d \to \mathbb{R}$ with $\int \bar{c}(\boldsymbol{u}) dF(\boldsymbol{u}) \le M$ for some $M < \infty$ and $\bar{c}(\boldsymbol{u}) = \max\left\{\sigma^2(\boldsymbol{u}), c(\boldsymbol{u})^2, \ldots, c(\boldsymbol{u})^Q\right\}$, where $\mathcal{F}^t = \sigma(U_{j-1}, \boldsymbol{X}_j : j \le t)$.
- **(M)** For some $b > 2$ let $E[|Y_1|^b] < \infty$ and $\boldsymbol{X}_1$ be absolutely continuous with density $f : \mathbb{R}^d \to \mathbb{R}$ such that $\sup_{\boldsymbol{x} \in \mathbb{R}^d} E[|Y_1|^b | \boldsymbol{X}_0 = \boldsymbol{x}] f(\boldsymbol{x}) < \infty$ and $\sup_{\boldsymbol{x} \in \mathbb{R}^d} f(\boldsymbol{x}) < \infty$. Let $\sup_{\boldsymbol{x}_1, \boldsymbol{x}_j} E[|Y_1 Y_j| | \boldsymbol{X}_1 = \boldsymbol{x}_1, \boldsymbol{X}_j = \boldsymbol{x}_j] f_{1j}(\boldsymbol{x}_1, \boldsymbol{x}_j) < \infty$ for all $j \ge j^*$ and some $j^* < \infty$, where f_{1j} is the density of $(\boldsymbol{X}_1, \boldsymbol{X}_j)$.
- **(J)** Let $(c_n)_{n \in \mathbb{N}}$ be a positive sequence of real valued numbers with $c_n \to \infty$ and $c_n = O((\log n)^{1/d})$ and let $\boldsymbol{J}_n = [-c_n, c_n]^d$.
- **(F1)** Denote $\boldsymbol{I}_n = [-c_n - Ch_n, c_n + Ch_n]^d$ for some $C < \infty$ and c_n from assumption **(J)** and let for all $n \in \mathbb{N}$, $\delta_n^{-1} = \inf_{\boldsymbol{x} \in \boldsymbol{J}_n} f(\boldsymbol{x}) > 0$ and

$$p_n = \max_{\substack{\boldsymbol{k} \in \mathbb{N}_0^d \\ 1 \le |\boldsymbol{k}| \le l+1+r}} \sup_{\boldsymbol{x} \in \boldsymbol{I}_n} |D^{\boldsymbol{k}} f(\boldsymbol{x})| < \infty$$

$$0 < q_n = \max_{\substack{\boldsymbol{k} \in \mathbb{N}_0^d \\ 0 \le |\boldsymbol{k}| \le l+1+r}} \sup_{\boldsymbol{x} \in \boldsymbol{I}_n} |D^{\boldsymbol{k}} m(\boldsymbol{x})| < \infty,$$

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

(F2) For q_n from assumption **(F1)**, c_n from assumption **(J)** and C from assumption **(K)**, let $\sup_{\boldsymbol{x} \in [-c_n - 2h_n C, c_n + 2h_n C]^d} \left| D^{\boldsymbol{k}} m(\boldsymbol{x}) \right| = O(q_n)$ for all $\boldsymbol{k} \in \mathbb{N}_0^d$ with $|\boldsymbol{k}| = 2$.

(K) Let $K : \mathbb{R}^d \to \mathbb{R}$ be symmetric in each component, $l+1$ times differentiable with $\int_{\mathbb{R}^d} K(\boldsymbol{z}) d\boldsymbol{z} = 1$ and compact support $[-C, C]^d$. Additionally, let $r \geq 2$ and $\int_{\mathbb{R}^d} K(\boldsymbol{z}) \boldsymbol{z}^{\boldsymbol{k}} d\boldsymbol{z} = 0$ for all $\boldsymbol{k} \in \mathbb{N}_0^d$ with $1 \leq |\boldsymbol{k}| \leq r - 1$, where $\boldsymbol{z}^{\boldsymbol{k}} = z_1^{k_1} \cdots z_d^{k_d}$. For all $L \in \{K\} \cup \{D^{\boldsymbol{k}} K : \boldsymbol{k} \in \mathbb{N}_0^d \text{ with } 1 \leq |\boldsymbol{k}| \leq l + 1\}$ let $|L(\boldsymbol{u})| < \infty$ for all $\boldsymbol{u} \in \mathbb{R}^d$ and $|L(\boldsymbol{u}) - L(\boldsymbol{u'})| \leq \Lambda \|\boldsymbol{u} - \boldsymbol{u'}\|$ for some $\Lambda < \infty$ and for all $\boldsymbol{u}, \boldsymbol{u'} \in \mathbb{R}^d$. (Here, r, l and C are from assumption **(F1)**.)

(B1) For δ_n, p_n, q_n and r, l from assumption **(F1)** let

$$\left(\sqrt{\frac{\log n}{n h_n^{d+2(l+1)}}} + h_n^r p_n \right) p_n^{l+1} \delta_n^{l+2} = O(1),$$

and for some $\eta \in (0, 1)$ let

$$\left(\sqrt{\frac{\log n}{n h_n^{d+2(l+1)}}} + h_n^r p_n \right) p_n^{l+\eta} q_n \delta_n^{l+1+\eta} = o(1).$$

(B2) For l, p_n, q_n, δ_n from assumption **(F1)** and η from assumption **(B1)**, let h_n satisfy the following conditions

$$\frac{(\log n)^{3+\frac{d}{l+\eta}}}{\sqrt{n^{1-\frac{d}{l+\eta}} h_n^d}} q_n^2 \delta_n^2 = o(1), \quad \frac{\log h_n}{\sqrt{n h_n^d}} = o(1)$$

and

$$\sqrt{n} h_n^r p_n q_n = o(1), \quad (\log n)^3 h_n q_n^2 = o(1).$$

Remark 2.1. *Under assumption **(G)** it is not needed that the innovations $\{U_t\}_t$ are i.i.d., e.g. one can obtain a strictly stationary, but non-i.i.d., white noise by defining $U_t = \sigma_t^2 (e_t^2 - 1)$, where $\{\sigma_t e_t\}_t$ forms a strictly stationary GARCH process, see Kreiß and Neuhaus (2006), 14.2.*

*Under aforementioned assumptions, consistency properties hold for $\hat{m}_n$ uniformly on $\boldsymbol{J}_n$ from assumption **(J)** which will be shown in section A.1 of the appendix. The key tool here is an application of Theorem 2 in Hansen (2008). Assumption **(G)** implies polynomial mixing rates of the underlying process needed in Hansen (2008). Moreover, together with the first bandwidth condition in **(B2)** the bandwidth constraints in Hansen (2008) are also fulfilled. Assumptions **(M)** and parts of **(K)** are reproduced from aforementioned paper.*

*In order to satisfy the first bandwidth condition in **(B2)**, a necessary condition on the smoothness of f and m then is $l + \eta > d$, meaning that for higher dimensional covariate $\boldsymbol{X}_t$, the existence of higher order partial derivatives of f and m is needed. In order to satisfy both the first and third bandwidth condition* *in **(B2)** at the same time, the order of the kernel needs to be large, in particular* $r > \frac{d}{2}\frac{l+\eta}{l+\eta-d}$*. The second bandwidth condition in **(B2)** is implied by the first one, if the bandwidth* h_n *has a polynomial rate of decay in* n *(or slower), meaning if there exists a* $k \in (0,\infty)$ *such that* $h_n = O(n^{-k})$*. Note that* $k < \frac{1}{d} - \frac{1}{l+\eta}$ *is necessary then. Note, however, that rates of convergence that are known to be MSE-optimal for estimation, namely* $O(n^{-1/(d+2r)})$*, can not fulfill the third condition from **(B2)** as* $r/(d+2r) < 1/2$*. However, it is common that for hypothesis testing different bandwidth rates are needed than for estimation.*

3. Asymptotic results

To derive the asymptotic distribution of test statistics built from the sequential marked empirical process $\hat{T}_n$ defined in (2.1), we apply the following expansion, which uses $Y_i = m(\boldsymbol{X}_i) + U_i$ for all $i = 1, \ldots, n$ under the null hypothesis,

$$\hat{T}_n(s,\boldsymbol{z}) \quad = \quad A_{n2}(s,\boldsymbol{z}) + A_{n1}(s,\boldsymbol{z})$$

with

$$A_{n1}(s,\boldsymbol{z}) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (m(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} \tag{3.1}$$

$$A_{n2}(s,\boldsymbol{z}) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} U_i \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}. \tag{3.2}$$

Lemma A.3 in the appendix shows that $A_{n2}(s,\boldsymbol{z}) = T_n(s,\boldsymbol{z}) + o_P(1)$ uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$ with the process

$$T_n(s,\boldsymbol{z}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} U_i I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}, \quad s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d. \tag{3.3}$$

Further, Lemma A.2 in the appendix shows that

$$A_{n1}(s,\boldsymbol{z}) = s\sqrt{n} \int_{\mathbb{R}^d} (m(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{x}))\omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}\} f(\boldsymbol{x}) d\boldsymbol{x} + o_P(1)$$

holds uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$. Inserting the definition of $\hat{m}_n$ from (2.2) one obtains one term of the form

$$\frac{s}{\sqrt{n}} \sum_{i=1}^{n} \int_{(-\boldsymbol{\infty},\boldsymbol{z}]} (m(\boldsymbol{y}) - m(\boldsymbol{X}_i)) K_{h_n}(\boldsymbol{y} - \boldsymbol{X}_i)\omega_n(\boldsymbol{y}) \frac{f(\boldsymbol{y})}{\hat{f}_n(\boldsymbol{y})} d\boldsymbol{y},$$

which is negligible by Lemma A.4 and one term of the form

$$-\frac{s}{\sqrt{n}} \sum_{i=1}^{n} U_i \int_{(-\boldsymbol{\infty},\boldsymbol{z}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{X}_i)\omega_n(\boldsymbol{y}) \frac{f(\boldsymbol{y})}{\hat{f}_n(\boldsymbol{y})} d\boldsymbol{y}$$

which can further be expanded applying Lemmata A.5 and A.3 such that one obtains

$$A_{n1}(s,\boldsymbol{z}) = -sT_n(1,\boldsymbol{z}) + o_P(1)$$

uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$. From this the expansion given in the first part of Theorem 3.1 below follows. In the second part of the theorem weak convergence of T_n from (3.3) is stated.

Theorem 3.1. ***(i)*** *Suppose that **(G)**, **(U)**, **(M)**, **(J)**, **(F1)**, **(F2)**, **(K)**, **(B1)** and **(B2)** are satisfied. Then under H_0*

$$\hat{T}_n(s,\boldsymbol{z}) = T_n(s,\boldsymbol{z}) - sT_n(1,\boldsymbol{z}) + o_P(1),$$

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

(ii) *Suppose that the assumptions **(G)** and **(U)** are satisfied. Then under H_0 the process T_n converges weakly in $\ell^\infty([0,1]\times\mathbb{R}^d)$ to a centered Gaussian process G with*

$$\operatorname{Cov}\big(G(s_1,\boldsymbol{z}_1), G(s_2,\boldsymbol{z}_2)\big) = (s_1 \wedge s_2)\Sigma(\boldsymbol{z}_1 \wedge \boldsymbol{z}_2)$$

and $\Sigma : \mathbb{R}^d \to \mathbb{R}, \boldsymbol{x} \mapsto \int_{(-\boldsymbol{\infty},\boldsymbol{x}]} \sigma^2(\boldsymbol{u}) f(\boldsymbol{u}) d\boldsymbol{u}$.

The proof of the first part follows from the considerations above applying Lemmata A.2–A.5 in the appendix, while the proof of the second part is given in section A.2 of the appendix. The proof of the second part in particular makes use of a recent result on weak convergence of sequential empirical processes indexed in function classes that can be applied for strongly mixing sequences, see Mohr (2019). Note that Koul and Stute (1999) show a weak convergence result applicable to the non-sequential process $\{T_n(1,z) : z \in \mathbb{R}\}$ under less restrictive assumptions on the dependence structure and moments (see Lemma 3.1 in aforementioned reference). From Theorem 3.1 and the continuous mapping theorem one directly obtains the limit distribution of $\hat{T}_n$.

Corollary 3.2. *Suppose that the assumptions of Theorem 3.1(i) are satisfied. Then under H_0 the process $\hat{T}_n$ converges weakly in $\ell^\infty([0,1]\times\mathbb{R}^d)$ to a centered Gaussian process G_0 with*

$$\operatorname{Cov}\big(G_0(s_1,\boldsymbol{z}_1), G_0(s_2,\boldsymbol{z}_2)\big) = (s_1 \wedge s_2 - s_1 s_2)\Sigma(\boldsymbol{z}_1 \wedge \boldsymbol{z}_2)$$

and Σ as in Theorem 3.1(ii).

Remark 3.3. *There is an asymptotic effect from estimating m. If m was known, one sees by setting $\hat{m}_n = m$ in the above formulas that $\hat{T}_n(s,z) = T_n(s,z) + o_P(1)$ and therefore the following weak convergence holds,*

$$\{\hat{T}_n(s,z) : s \in [0,1], z \in \mathbb{R}^d\} \to \{G(s,z) : s \in [0,1], z \in \mathbb{R}^d\},$$

where G is the centered Gaussian process from Theorem 3.1 (ii).

Continuous functionals of the process $\hat{T}_n$ can be used as test statistics for H_0. We consider the following Kolmogorov-Smirnov and Cramér-von Mises type statistics and combinations of both,

$$T_{n1} = \sup_{s\in[0,1],\boldsymbol{z}\in\mathbb{R}^d} \left|\hat{T}_n(s,\boldsymbol{z})\right|, \quad T_{n2} = \sup_{\boldsymbol{z}\in\mathbb{R}^d} \int_0^1 \left|\hat{T}_n(s,\boldsymbol{z})\right|^2 ds,$$

$$T_{n3} = \sup_{s\in[0,1]} \int_{\mathbb{R}^d} \left|\hat{T}_n(s,\boldsymbol{z})\right|^2 v(\boldsymbol{z})d\boldsymbol{z}, \quad T_{n4} = \int_0^1 \int_{\mathbb{R}^d} \left|\hat{T}_n(s,\boldsymbol{z})\right|^2 v(\boldsymbol{z})d\boldsymbol{z}ds,$$

where $v : \mathbb{R}^d \to \mathbb{R}$ is some integrable weighting function. Applying Corollary 3.2 and the continuous mapping theorem gives convergence in distribution of those test statistics. One can obtain distribution-free tests in the case of dimension $d = 1$ as follows. Denote by $\{K_0(s,t) : s \in [0,1], t \in \mathbb{R}\}$ a Kiefer-Müller process, i.e. a centered Gaussian process with covariance function $\mathrm{Cov}(K_0(s_1,t_1), K_0(s_2,t_2)) = (s_1 \wedge s_2 - s_1 s_2)(t_1 \wedge t_2)$. Then $K_0(\cdot, \Sigma(\cdot))$ has the same distribution as $G_0(\cdot,\cdot)$. Let further $\sigma(\cdot)$ be continuous and consider the consistent estimator $\hat{c}_n = n^{-1}\sum_{i=1}^n (Y_i - \hat{m}_n(X_i))^2 \omega_n(X_i)$ for $c = \int \sigma^2(u)f(u)du$. Applying a scaling property of the process K_0 in its second component and substitution in the integrals it is easy to derive convergence in distribution as follows,

$$\frac{T_{n1}}{\hat{c}_n^{1/2}} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \sup_{s\in[0,1],t\in[0,1]} |K_0(s,t)|, \quad \frac{T_{n2}}{\hat{c}_n} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \sup_{t\in[0,1]} \int_0^1 |K_0(s,t)|^2 ds,$$

$$\frac{T_{n3}}{\hat{c}_n^2} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \sup_{s\in[0,1]} \int_0^1 |K_0(s,t)|^2 dt, \quad \frac{T_{n4}}{\hat{c}_n^2} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \int_0^1 \int_0^1 |K_0(s,t)|^2 dtds.$$

For the latter two tests however the unknown weight function $v = \sigma^2 f$ needs to be chosen to obtain the limit as stated above. To obtain feasible asymptotically distribution-free tests, T_{n3} and T_{n4} should be replaced by

$$\tilde{T}_{n3} = \sup_{s\in[0,1]} \frac{1}{n}\sum_{k=1}^n \left|\hat{T}_n(s,X_k)\right|^2 \hat{\sigma}_n^2(X_k), \ \tilde{T}_{n4} = \int_0^1 \frac{1}{n}\sum_{k=1}^n \left|\hat{T}_n(s,X_k)\right|^2 \hat{\sigma}_n^2(X_k)ds$$

applying a nonparametric estimator for the variance function such as

$$\hat{\sigma}_n^2(x) = \frac{\sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right)(Y_j - \hat{m}_n(x))^2}{\sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right)}.$$

To conclude the section we will have a closer look at the alternative of one change point. For simplicity reasons we will only consider the test based on T_{n1}. To model the alternative we assume a triangular array

$$Y_{n,t} = m_{n,t}(\boldsymbol{X}_{n,t}) + U_{n,t}, \ t = 1,\ldots,n,$$

and validity of the alternative of one change point, i.e.

$$H_1 : \exists s_0 \in (0,1) : m_{n,t}(\cdot) = \begin{cases} m_{(1)}(\cdot), & t = 1, \dots, \lfloor ns_0 \rfloor \\ m_{(2)}(\cdot), & t = \lfloor ns_0 \rfloor + 1, \dots, n \end{cases} \tag{3.4}$$

for some not further specified functions $m_{(1)} \not\equiv m_{(2)}$. Let $f_{n,t}$ denote the density of $\boldsymbol{X}_{n,t}$ and assume that for all $s \in (0,1]$ there exists a function $\bar{f}^{(s)} : \mathbb{R}^d \to \mathbb{R}$ such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{\lfloor ns \rfloor} f_{n,t}(\boldsymbol{x}) = \bar{f}^{(s)}(\boldsymbol{x}), \ \forall\, \boldsymbol{x} \in \mathbb{R}^d. \tag{3.5}$$

Even though assumption (3.5) is rather difficult to varify in practise, we refer the reader to the following example, $Y_t = a_t \cdot Y_{t-1} + \varepsilon_t$ 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 $t > \lfloor ns_0 \rfloor$, $a \neq b$ (see section 4 in Mohr and Selk (2020)).

Under some regularity conditions it can be shown by applying Kristensen's (2009) results that

$$\sup_{\boldsymbol{x} \in \boldsymbol{J}_n} |\hat{m}_n(\boldsymbol{x}) - \bar{m}_n(\boldsymbol{x})| = o_P(1), \tag{3.6}$$

where $\bar{m}_n(\boldsymbol{x}) = \sum_{i=1}^n f_{n,i}(\boldsymbol{x}) m_{n,i}(\boldsymbol{x}) / \sum_{i=1}^n f_{n,i}(\boldsymbol{x})$ converges to the mixture

$$m_{(1)}(\boldsymbol{x}) \frac{\bar{f}^{(s_0)}(\boldsymbol{x})}{\bar{f}^{(1)}(\boldsymbol{x})} + \left(1 - \frac{\bar{f}^{(s_0)}(\boldsymbol{x})}{\bar{f}^{(1)}(\boldsymbol{x})}\right) m_{(2)}(\boldsymbol{x}) \tag{3.7}$$

of the regression functions before and after the change (for details see Theorem 2.3 in Mohr (2018)). Now for fixed $\boldsymbol{z} \in \mathbb{R}^d$ and $s \in (0,1)$ with $s \le s_0$, it holds that

$$\hat{T}_n(s, \boldsymbol{z}) = \sqrt{n}\Delta(s, \boldsymbol{z}) + o_P(\sqrt{n}),$$

where

$$\Delta(s, \boldsymbol{z}) = \int_{(-\boldsymbol{\infty}, \boldsymbol{z}]} (m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u})) \left(1 - \frac{\bar{f}^{(s_0)}(\boldsymbol{u})}{\bar{f}^{(1)}(\boldsymbol{u})}\right) \bar{f}^{(s)}(\boldsymbol{u}) d\boldsymbol{u}.$$

As under H_1 this integral is non-zero for $s = s_0$ and some $\boldsymbol{z}$, convergence of T_{n1} to infinity in probability and thus the test is consistent. For a rigorous proof we refer the reader to Theorem 3.4 in Mohr (2018).

Remark 3.4. *Consider the non-marked CUSUM process $\hat{T}_n(s, \infty)$ which is analogous to Su and Xiao's (2008) procedure. Considerations as above for the fixed alternative H_1 of one change point in $\lfloor ns_0 \rfloor$ leads for $s \le s_0$ to*

$$\Delta(s, \infty) \quad = \quad \int (m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u})) \left(1 - \frac{\bar{f}^{(s_0)}(\boldsymbol{u})}{\bar{f}^{(1)}(\boldsymbol{u})}\right) \bar{f}^{(s)}(\boldsymbol{u}) d\boldsymbol{u}$$

The integral can be zero even if $m_{(1)} \neq m_{(2)}$. Then tests based on the CUSUM process will not be consistent, while tests based on the marked CUSUM process *are. We will consider some examples in section 5. Note that the above integral equals*

$$s(1-s_0)\int(m_{(1)}(\boldsymbol{u})-m_{(2)}(\boldsymbol{u}))f(\boldsymbol{u})d\boldsymbol{u},$$

in case of a stationary covariate process.

4. A bootstrap procedure and the case of non-stationary variances

As alternative to the asymptotic test considered in section 3, in this section we will suggest a wild bootstrap approach. This resampling procedure can in particular be applied in the case of multivariate covariates, where the critical values for the asymptotic tests based on Corollary 3.2 have to be estimated. Moreover, the bootstrap approach can be applied to obtain a test that detects changes in the conditional mean function, even when the conditional variance function is not stable. As desired, the test does not react sensitive to the unstable variance. In contrast to the bootstrap approach, the limiting distribution from section 3 cannot be applied in the case of changes in the variance.

We consider the model

$$Y_{n,t}=m_{n,t}(\boldsymbol{X}_{n,t})+U_{n,t},\ t=1,\dots,n,$$

with $E[U_{n,t}|\mathcal{F}_n^t]=0$ and $E[U_{n,t}^2|\boldsymbol{X}_{n,t}]=\sigma_{n,t}^2(\boldsymbol{X}_{n,t})$ a.s. for some functions $\sigma_{n,t}^2:\mathbb{R}^d\to\mathbb{R}$ and $\mathcal{F}_n^t:=\sigma(U_{n,j-1},\boldsymbol{X}_{n,j}:j\le t)$. We assume $\boldsymbol{X}_{n,t}$ to be absolutely continuous with density function $f_{n,t}$. The model considered in section 2 and the first part of section 3 is the special case where $f_{n,t}(\cdot)=f(\cdot)$ and $\sigma_{n,t}^2(\cdot)=\sigma^2(\cdot)$ for all $t=1,\dots,n$ and for some $f,\sigma^2:\mathbb{R}^d\to\mathbb{R}$ not depending on t and n. Both models allow for heteroscedasticity, but the more general model also allows for possible changes in $\sigma_{n,t}^2$, which should not effect the rejection probability of the test for

$$H_0: m_{n,t}(\cdot)=m(\cdot),\ t=1,\dots,n,$$

(for some m not depending on t and n). We again consider the procedure

$$\hat{T}_n(s,\boldsymbol{z})=\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\hat{U}_{n,i}\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}\}$$

with residuals $\hat{U}_{n,i}=Y_{n,i}-\hat{m}_n(\boldsymbol{X}_{n,i})$. Here $\hat{m}_n$ is defined as in (2.2), but replacing $(\boldsymbol{X}_j,Y_j)$ by $(\boldsymbol{X}_{n,j},Y_{n,j})$, $j=1,\dots,n$.

First define the wild bootstrap innovations as $U_{n,t}^*=\hat{U}_{n,t}\eta_t$, where $\{\eta_t\}$ are i.i.d. random variables, independent of the original sample with $E[\eta_0]=0$, $E[\eta_0^2]=1$ and $E[\eta_0^4]<\infty$. Then the bootstrap data fulfilling the null hypothesis are generated by

$$Y_{n,t}^*=\hat{m}_n(\boldsymbol{X}_{n,t})+U_{n,t}^*.$$

Note that if the original data follow an autoregression model, say $d=1$ and $X_{n,t}=Y_{n,t-1}$, by the above choice the resulting bootstrap data does not follow the same structure. As was pointed out by Kreiß and Lahiri (2012) this bootstrap data generation is still a reasonable choice in particular if the dependence structure of the underlying process does not show up in the asymptotic distribution. Another possibility might be a dependent wild bootstrap as suggested in Shao (2010).

The bootstrap residuals are defined as $\hat{U}^*_{n,t} = Y^*_{n,t} - \hat{m}^*_n(\boldsymbol{X}_{n,t})$, where $\hat{m}^*_n$ is defined as $\hat{m}_n$ in (2.2), but replacing $(\boldsymbol{X}_j, Y_j)$ by $(\boldsymbol{X}_{n,j}, Y^*_{n,j})$, $j = 1, \ldots, n$. The bootstrap process is defined as

$$\hat{T}^*_n(s, \boldsymbol{z}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \hat{U}^*_{n,i} \omega_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\}.$$

Bootstrap versions $T^*_{n\ell}$, $\ell = 1, \ldots, 4$, are defined analogous to the test statistics $T_{n\ell}$, $\ell = 1, \ldots, 4$, but based on $\hat{T}^*_n$ instead of $\hat{T}_n$. Then H_0 is rejected if $T_{n\ell}$ is larger than the $(1-\alpha)$-quantile of the conditional distribution of $T^*_{n\ell}$, given the original data.

To motivate that we obtain a valid procedure (which holds the level asymptotically and is consistent) even in the case of changing variances, we will consider the limiting process G_0 of the original process $\hat{T}_n$ and the conditional limiting process G^*_0 of the bootstrap version $\hat{T}^*_n$ in subsections 4.1 and 4.2 below. We will see that the processes G_0 and G^*_0 coincide under the null hypothesis. Note that some steps of the derivation can be proven rigorously, see Mohr (2018) for details, but deriving the weak convergence (see assumption (ii) in the next paragraph) would require a limit theorem for sequential empirical processes indexed in function classes for weakly dependent non-stationary triangular arrays. Such a result is, to the best of our knowledge, not yet available in the literature and thus a rigorous proof is beyond the scope of the paper (see Mohr (2019) for a related limit theorem that requires stationarity).

4.1. Asymptotics for non-homogeneous variances

Heuristically under H_0 one can proceed as in the proof of the first part of Theorem 3.1 in the beginning of section 3. Under some regularity assumptions similar to those in section 2 and the assumption that the limit $\bar{f}^{(s)}$ as in (3.5) exists one obtains the expansion

$$\hat{T}_n(s, \boldsymbol{z}) = \Gamma_n(s, 1, \boldsymbol{z}) - \Gamma_n(1, s, \boldsymbol{z}) + o_P(1)$$

with the process

$$\Gamma_n(s, t, \boldsymbol{z}) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} U_{n,i} \bar{g}^{(t)}(\boldsymbol{X}_{n,i}) \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\} : s, t \in [0, 1], \boldsymbol{z} \in \mathbb{R}^d,$$

where $\bar{g}^{(t)} := \bar{f}^{(t)} / \bar{f}^{(1)}$. Now assume that

(i) the limit $\bar{h}^{(s)}(\cdot) := \lim_{n\to\infty} n^{-1} \sum_{i=1}^{\lfloor ns \rfloor} \sigma^2_{n,i}(\cdot) f_{n,i}(\cdot)$ exists for all $s \in (0, 1]$,

(ii) the process Γ_n converges weakly to a centered Gaussian process Γ.

The limiting covariance then is

$$E[\Gamma(s_1,t_1,\boldsymbol{z}_1)\Gamma(s_2,t_2,\boldsymbol{z}_2)] = \int_{(-\boldsymbol{\infty},\boldsymbol{z}_1\wedge\boldsymbol{z}_2]} \bar{h}^{(s_1\wedge s_2)}(\boldsymbol{u})\bar{g}^{(t_1)}(\boldsymbol{u})\bar{g}^{(t_2)}(\boldsymbol{u})d\boldsymbol{u}.$$

Then with the continuous mapping theorem the weak convergence of $\hat{T}_n$ to a centered Gaussian process $\{G_0(s,\boldsymbol{z}) : s\in[0,1], \boldsymbol{z}\in\mathbb{R}^d\}$ follows with covariances

$$\begin{aligned}\operatorname{Cov}(G_0(s_1,\boldsymbol{z}_1),G_0(s_2,\boldsymbol{z}_2)) = \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}_1\wedge\boldsymbol{z}_2]} &\Big(\bar{h}^{(s_1\wedge s_2)}(\boldsymbol{u}) - \bar{h}^{(s_1)}(\boldsymbol{u})\bar{g}^{(s_2)}(\boldsymbol{u}) \\ &-\bar{h}^{(s_2)}(\boldsymbol{u})\bar{g}^{(s_1)}(\boldsymbol{u}) + \bar{h}^{(1)}(\boldsymbol{u})\bar{g}^{(s_1)}(\boldsymbol{u})\bar{g}^{(s_2)}(\boldsymbol{u})\Big)\, d\boldsymbol{u}.\end{aligned}$$

Note that this is consistent with the stationary case as then $\bar{h}^{(s)}(\cdot) = s\sigma^2(\cdot)f(\cdot)$ and $\bar{g}^{(s)}(\cdot) = s$ and the same covariance function as in Corollary 3.2 is obtained. The convergence of the test statistics $T_{n\ell}$, $\ell = 1,\dots,4$, in distribution follows again from the continuous mapping theorem.

Under the change point alternative H_1 from (3.4) with $m_{(1)} \not\equiv m_{(2)}$, analogous to the considerations in section 3 it holds that the test statistic T_{n1} converges to infinity in probability.

4.2. Derivations for the bootstrap process

Concerning the weak convergence of the bootstrap process $\hat{T}_n^*$, conditionally on the sample, we have again a look at the expansion in the beginning of section 3 for the derivation of the first part of the proof of Theorem 3.1. In what follows let P^* denote the conditional probability and E^* the conditional expectation, given the observations. Further let $Z_n = o_{P^*}(1)$ be short for $P^*(|Z_n| > \epsilon) = o_P(1)$ for all $\epsilon > 0$. Here we obtain

$$\hat{T}_n(s,\boldsymbol{z}) = A^*_{n2}(s,\boldsymbol{z}) + A^*_{n1}(s,\boldsymbol{z})$$

with

$$A^*_{n2}(s,\boldsymbol{z}) := \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor} U^*_{n,i}\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}\}$$

and (similar to Lemma A.2 in the appendix)

$$\begin{aligned}A^*_{n1}(s,\boldsymbol{z}) :=& \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}(\hat{m}_n(\boldsymbol{X}_{n,i}) - \hat{m}^*_n(\boldsymbol{X}_{n,i}))\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}\} \\ =& \sqrt{n}\int_{(-\boldsymbol{\infty},\boldsymbol{z}]}(\hat{m}_n(\boldsymbol{x}) - \hat{m}^*_n(\boldsymbol{x}))\omega_n(\boldsymbol{x})\bar{f}^{(s)}(\boldsymbol{x})d\boldsymbol{x} + o_{P^*}(1)\end{aligned}$$

with $\bar{f}^{(s)}$ as in (3.5). Inserting the definition of $\hat{m}_n^*$ this leads to a term (similar to Lemma A.4) of the form

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\int_{(-\boldsymbol{\infty},\boldsymbol{z}]}\left(\hat{m}_n(\boldsymbol{x})-\hat{m}_n(\boldsymbol{X}_{n,j})\right)K_{h_n}(\boldsymbol{x}-\boldsymbol{X}_{n,j})\omega_n(\boldsymbol{x})\frac{\bar{f}^{(s)}(\boldsymbol{x})}{\bar{f}^{(1)}(\boldsymbol{x})}d\boldsymbol{x},$$

which is negligible, and a term (similar to Lemma A.5) of the form

$$\begin{aligned}&-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}U_{n,j}^*\int_{(-\boldsymbol{\infty},\boldsymbol{z}]}K_{h_n}(\boldsymbol{x}-\boldsymbol{X}_{n,j})\omega_n(\boldsymbol{x})\frac{\bar{f}^{(s)}(\boldsymbol{x})}{\bar{f}^{(1)}(\boldsymbol{x})}d\boldsymbol{x}\\&=-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}U_{n,j}^*\omega_n(\boldsymbol{X}_{n,j})\frac{\bar{f}^{(s)}(\boldsymbol{X}_{n,j})}{\bar{f}^{(1)}(\boldsymbol{X}_{n,j})}I\{\boldsymbol{X}_{n,j}\le\boldsymbol{z}\}+o_{P^*}(1).\end{aligned}$$

Thus one obtains (under suitable regularity conditions) the expansion

$$\hat{T}_n^*(s,\boldsymbol{z})=\Gamma_n^*(s,1,\boldsymbol{z})-\Gamma_n^*(1,s,\boldsymbol{z})+o_{P^*}(1),$$

where

$$\Gamma_n^*(s,t,\boldsymbol{z}):=\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}U_{n,i}^*\omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t)}(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}\},\ \ s,t\in[0,1],\boldsymbol{z}\in\mathbb{R}^d,$$

and $\bar{g}^{(t)}$ is defined as in section 4.1. In what follows we will assume that the process Γ_n^*, conditionally on the sample, converges weakly to a centered Gaussian process, in probability. Then, by the continuous mapping theorem, $\hat{T}_n^*$, conditionally converges weakly to a centered Gaussian process, say G_0^*. We will calculate the asymptotic variances in order to show that under H_0 those coincide with the covariances of G_0 as in section 4.1. First note $E^*[U_{n,i}^*U_{n,j}^*]=\hat{U}_{n,i}^2I\{i=j\}$ holds almost surely. Under H_0 it holds that $\hat{U}_{n,t}=m(\boldsymbol{X}_{n,t})-\hat{m}_n(\boldsymbol{X}_{n,t})+U_{n,t}$ and $\hat{m}_n$ consistently estimates m, and thus

$$\begin{aligned}&E^*\left[\Gamma_n^*(s_1,t_1,\boldsymbol{z}_1)\Gamma_n^*(s_2,t_2,\boldsymbol{z}_2)\right]\\&=\frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor\wedge\lfloor ns_2\rfloor}\hat{U}_{n,i}^2\omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}_1\wedge\boldsymbol{z}_2\}\\&=\frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor\wedge\lfloor ns_2\rfloor}U_{n,i}^2\omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}_1\wedge\boldsymbol{z}_2\}+o_P(1)\\&=E\left[\Gamma(s_1,t_1,\boldsymbol{z}_1)\Gamma(s_2,t_2,\boldsymbol{z}_2)\right]+o_P(1)\end{aligned}$$

under H_0, where Γ is the limiting distribution of Γ_n in section 4.1. Thus, under H_0, $\hat{T}_n^*$ indeed (presumably) converges weakly to G_0 in probability, and thus the test statistic $T_{n\ell}^*$ converges conditionally in distribution, to the same limits as $T_{n\ell}$ (respectively for $\ell=1,\dots,4$).

Under the alternative H_1 as in (3.4), $\hat{U}_{n,i} = m_{n,i}(\boldsymbol{X}_{n,i}) - \hat{m}_n(\boldsymbol{X}_{n,i}) + U_{n,i}$ and thus it holds that

$$\begin{aligned}
&E^*\left[\Gamma_n^*(s_1, t_1, \boldsymbol{z}_1)\Gamma_n^*(s_2, t_2, \boldsymbol{z}_2)\right] \\
&= \frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor \wedge \lfloor ns_2\rfloor} \hat{U}_{n,i}^2 \omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}_1 \wedge \boldsymbol{z}_2\} \\
&= \frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor \wedge \lfloor ns_2\rfloor} U_{n,i}^2 \omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}_1 \wedge \boldsymbol{z}_2\} \\
&+ r_{n1} + r_{n2}
\end{aligned}$$

for fixed $s_1, s_2, t_1, t_2 \in [0, 1]$ and $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{R}^d$. The first term again converges in probability to $E\left[\Gamma(s_1, t_1, \boldsymbol{z}_1)\Gamma(s_2, t_2, \boldsymbol{z}_2)\right]$. It can further be shown that

$$\begin{aligned}
r_{n1} = \frac{2}{n}\sum_{i=1}^{\lfloor ns_1\rfloor \wedge \lfloor ns_2\rfloor} & U_{n,i}(m_{n,i}(\boldsymbol{X}_{n,i}) - \hat{m}_n(\boldsymbol{X}_{n,i}))\omega_n(\boldsymbol{X}_{n,i}) \\
&\cdot \bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}_1 \wedge \boldsymbol{z}_2\}
\end{aligned}$$

converges to zero in probability. However,

$$\begin{aligned}
r_{n2} &= \frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor \wedge \lfloor ns_2\rfloor} (m_{n,i}(\boldsymbol{X}_{n,i}) - \hat{m}_n(\boldsymbol{X}_{n,i}))^2\omega_n(\boldsymbol{X}_{n,i}) \\
&\qquad \cdot \bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}_1 \wedge \boldsymbol{z}_2\} \\
&= \frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor \wedge \lfloor ns_2\rfloor} (m_{n,i}(\boldsymbol{X}_{n,i}) - \bar{m}_n(\boldsymbol{X}_{n,i}))^2\omega_n(\boldsymbol{X}_{n,i}) \\
&\qquad \cdot \bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}_1 \wedge \boldsymbol{z}_2\} \\
&+ o_P(1),
\end{aligned}$$

with the same $\bar{m}_n$ as in (3.6), which converges to

$$(m_{(1)}(\boldsymbol{x}) - m_{(2)}(\boldsymbol{x}))\bar{g}^{(s_0)}(\boldsymbol{x}) + m_{(2)}(\boldsymbol{x})$$

(see (3.7)). Thus, it can be shown that

$$\begin{aligned}
r_{n2} = \frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor \wedge \lfloor ns_2\rfloor \wedge \lfloor ns_0\rfloor} & \Big(m_{(1)}(\boldsymbol{X}_{n,i}) - m_{(2)}(\boldsymbol{X}_{n,i})\Big)^2 \omega_n(\boldsymbol{X}_{n,i})\Big(1 - \bar{g}^{(s_0)}(\boldsymbol{X}_{n,i})\Big)^2 \\
&\cdot \bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}_1 \wedge \boldsymbol{z}_2\} \\
+ \frac{1}{n}\sum_{i=\lfloor ns_1\rfloor \wedge \lfloor ns_2\rfloor \wedge \lfloor ns_0\rfloor + 1}^{\lfloor ns_1\rfloor \wedge \lfloor ns_2\rfloor} & \Big(m_{(1)}(\boldsymbol{X}_{n,i}) - m_{(2)}(\boldsymbol{X}_{n,i})\Big)^2 \omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(s_0)}(\boldsymbol{X}_{n,i})^2 \\
&\cdot \bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}_1 \wedge \boldsymbol{z}_2\}
\end{aligned}$$

$$+ o_P(1).$$

It can be seen that these terms do not vanish but converge to some limit in probability. Thus the limiting distribution G_0^* under H_1 is not equal to G_0 and in particular depends on the changepoint s_0. As seen before under H_1 the original test statistic T_{n1} converges in probability to infinity. On the other hand, the bootstrap test statistic T_{n1}^* conditionally converges in distribution to some non-degenerated limit, in probability. Thus the bootstrap test is consistent.

5. Finite sample properties

A small Monte Carlo study is conducted in order to compare the results for T_{n1} and T_{n2} from section 3 with those of the traditional CUSUM versions denoted by $KS := \sup_{s\in[0,1]} |\hat{T}_n(s,\infty)|$ and $CM := \int |\hat{T}_n(s,\infty)|^2 ds$. Note that the results for $\tilde{T}_{n3}$ and $\tilde{T}_{n4}$ are similar and omitted for reasons of brevity. Asymptotic tests are applied to data satisfying models 1 and 2, while the bootstrap versions are applied to model 3 and 4 explained below. All simulations are carried out with a level of 5%, 500 replications and 200 bootstrap replications and for sample sizes $n \in \{100, 300, 500\}$. For the nonparametric estimators we use a fourth order Epanechnikov kernel (and the corresponding product kernel in model 4 below) and the bandwidth is chosen by the cross validation method. For simplicity we set $\omega_n \equiv 1$. The data is simulated from the following models.

$$\text{(model 1)} \quad Y_t = m_t(X_t) + \sqrt{1+0.5X_t^2}\varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0,1),$$
$$m_t(x) = \begin{cases} 0.5x, & t = 1,\dots,\lfloor n/2\rfloor \\ (0.5+\Delta_0 e^{-0.8x^2})x, & t = \lfloor n/2\rfloor+1,\dots,n \end{cases},$$

where X_i is an exogenous variable following the AR(1) model $X_t = 0.4X_{t-1}+\xi_t$ with ξ_i being i.i.d. $\sim \mathcal{N}(0,1)$ and $\Delta_0 \in \{0, 0.5, 1, 1.5, 1.5, 2, 2.5, 3, 3.5, 4\}$.

$$\text{(model 2)} \quad Y_t = m_t(Y_{t-1}) + \sigma(Y_{t-1})\varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0,1),$$
$$m_t(x) = \begin{cases} -0.9x, & t = 1,\dots,\lfloor n/2\rfloor \\ (-0.9+\Delta_0)\,x, & t = \lfloor n/2\rfloor+1,\dots,n \end{cases},$$

with $\Delta_0 \in \{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8\}$. Consider the homoscedastic case, where $\sigma^2(x) = 1$ and the heteroscedastic case, where $\sigma^2(x) = 1 + 0.1x^2$.

$$\text{(model 3)} \quad Y_t = m_t(Y_{t-1}) + \sigma_t(Y_{t-1})\varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0,1),$$
$$\sigma_t^2(x) = \begin{cases} 1+0.1x^2, & t = 1,\dots,\lfloor nt_0\rfloor \\ 1+0.8x^2, & t = \lfloor nt_0\rfloor+1,\dots,n \end{cases},$$
$$m_t(x) = \begin{cases} 0.9x, & t = 1,\dots,\lfloor n/2\rfloor \\ (0.9-\Delta_0)\,x, & t = \lfloor n/2\rfloor+1,\dots,n \end{cases},$$

with $\Delta_0 \in \{0, 1.3\}$ and $t_0 \in \{0.25, 0.5, 0.75\}$.

$$\text{(model 4)} \quad Y_t = m_t(Y_{t-1}, Y_{t-2}) + \sigma(Y_{t-1}, Y_{t-2})\varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0,1),$$
$$m_t(x_1, x_2) = \begin{cases} 0.9x_1 - 0.4x_2, & t = 1, \dots, \lfloor n/2 \rfloor \\ (0.9 - \Delta_0)\, x_1 - 0.4x_2, & t = \lfloor n/2 \rfloor + 1, \dots, n \end{cases},$$

with $\Delta_0 \in \{0, 1.3\}$. We consider three different choices for the conditional variance function, namely $\sigma^2(x_1, x_2) = 1$ for an AR(2) model, $\sigma^2(x_1, x_2) = 1 + 0.4x_1^2$ for an AR(2)-ARCH(1) model and $\sigma^2(x_1, x_2) = 1 + 0.2x_1^2 + 0.2x_2^2$ for an AR(2)-ARCH(2) model.

Model 1 is a regression model with autoregressive covariables. In model 2 we consider both a homoscedastic and heteroscedastic autoregression model, while model 3 is a heteroscedastic autoregression with non-homogeneous variances. In model 4 we consider both homoscedastic and heteroscedastic second order autoregression models. All models fulfill H_0 for $\Delta_0 = 0$ and H_1 for $\Delta_0 \neq 0$ with a change in regression function occurring in $\lfloor n/2 \rfloor$. Further, note that for model 1 the covariate process $\{X_t\}$ fulfills the required stationarity and mixing conditions (see 2.6.1 (iii) in Fan and Yao (2003)) which are inherited to $\{Y_t\}$ under H_0. Model 2 also fulfills the stationarity and mixing conditions under the null (see 2.6.1 (iii) in Fan and Yao (2003) and Theorem 1 in Lu (1998) for $\sigma^2(x) = 1$ and $\sigma^2(x) = 1 + 0.1x^2$, respectively). Note that model 3 is not stationary as one change occurs in the conditional variance function under both H_0 and H_1. For model 4 the limiting distribution from Corollary 3.2 does not result in an asymptotically distribution-free test as the covariate is multivariate. Thus for both model 3 and 4 we apply the bootstrap procedure from section 4.

Figures 1, 2 and 3 are visualizations of the performance of T_{n1} and T_{n2}, as well as KS and CM in model 1 and 2. Under the null the rejection frequencies for all tests are near the nominal level. For model 1 the CUSUM tests are not consistent against H_1, while the tests based on the marked process are. In model 2 the rejection frequencies of all tests increase with increasing break size. Note however that the increase is much faster for T_{n1} and T_{n2} than for the CUSUM tests. Also note that the influence of the conditional variance is rather small resulting in a similar performance in both the homoscedastic and heteroscedastic case.

Table 1 shows the rejection frequencies of the bootstrap procedure for model 3 using T_{n1}^* and T_{n2}^*, as well as the bootstrap version of the CUSUM tests KS and CM under both the null and the alternative hypothesis. The level simulations show that all tests perform reasonably well under H_0, approximately holding the level indicating that the bootstrap test is – as desired – not sensitive to changes in the conditional variance function. Furthermore, it can be seen that for all models and all tests the rejection frequency under H_1 exceeds the level, indicating that the change point is detected. With increasing sample size, the number of rejections increases rapidly for T_{n1}^* and T_{n2}^*, while it stays approximately constant for the bootstrap versions of KS and CM. This is presumably due to the fact that the test statistics based on $\hat{T}_n(s,\infty)$ estimate some integral that might be small under H_1. As was pointed out in subsection 4.2, this integral not vanishing is essential for the consistency property for the bootstrap tests.

Table 2 shows the rejection frequencies for model 4 in all three cases of variance function, when using the tests based on T^*_{n1} and T^*_{n2}, as well as the bootstrap versions of KS and CM under both H_0 and H_1. It can be seen that under H_0 the tests reject a little more often than in the models considered in section 5, overestimating the level of 5% sometimes for finite sample sizes. Under the alternative the number of rejections increases rapidly for T^*_{n1} and T^*_{n2} with increasing n, while it stays relatively low for KS and CM. In summary, the bootstrap tests perform reasonably well and are therefore an acceptable alternative to the tests using critical values of the limiting distribution, which are here not known due to multidimensional covariates. Furthermore in these models, they outperform the bootstrap versions of the CUSUM tests.

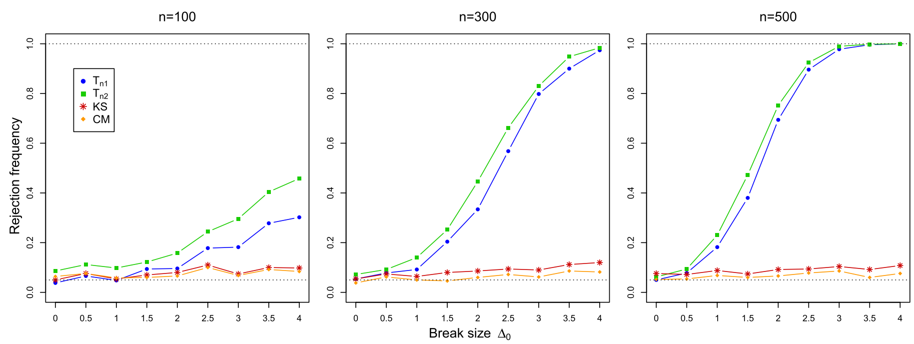

FIG 1. *Rejection frequencies in model 1*

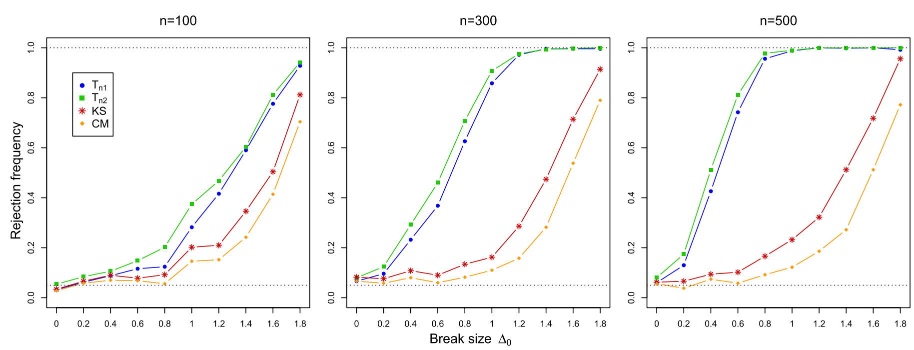

FIG 2. *Rejection frequencies in model 2 with $\sigma^2(x) = 1$*

Finally, we apply the asymptotic test based on T_{n1} to 36 measurements of the annual flow volume of the small Czech river Ráztoka recorded between 1954

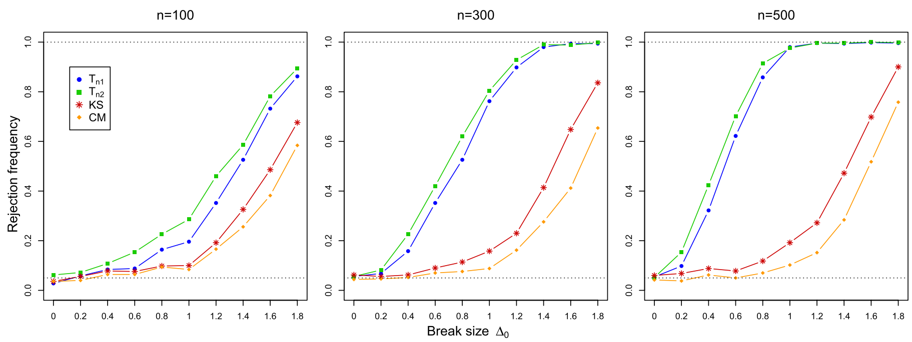

FIG 3. *Rejection frequencies in model 2 with* $\sigma^2(x) = 1 + 0.1x^2$

TABLE 1
Rejection frequencies in model 3 ($\alpha = 0.05$)

		under H_0				under H_1			
t_0	n	T^*_{n1}	T^*_{n2}	KS	CM	T^*_{n1}	T^*_{n2}	KS	CM
0.25	100	0.030	0.046	0.030	0.054	0.286	0.270	0.192	0.168
	300	0.068	0.064	0.080	0.052	0.652	0.644	0.248	0.172
	500	0.060	0.052	0.058	0.046	0.878	0.868	0.264	0.194
0.50	100	0.068	0.048	0.068	0.056	0.420	0.438	0.316	0.256
	300	0.066	0.050	0.056	0.046	0.868	0.894	0.378	0.292
	500	0.046	0.040	0.058	0.040	0.994	0.996	0.434	0.324
0.75	100	0.060	0.056	0.072	0.070	0.404	0.388	0.332	0.266
	300	0.048	0.048	0.050	0.056	0.830	0.848	0.382	0.250
	500	0.034	0.040	0.046	0.056	0.986	0.988	0.350	0.202

TABLE 2
Rejection frequencies in model 4 ($\alpha = 0.05$)

		under H_0				under H_1			
model	n	T^*_{n1}	T^*_{n2}	KS	CM	T^*_{n1}	T^*_{n2}	KS	CM
AR(2)	100	0.082	0.068	0.082	0.074	0.124	0.110	0.080	0.070
	300	0.064	0.070	0.054	0.048	0.284	0.308	0.096	0.070
	500	0.076	0.058	0.068	0.060	0.480	0.532	0.098	0.070
AR(2)-	100	0.076	0.060	0.094	0.068	0.098	0.106	0.070	0.058
ARCH(1)	300	0.084	0.098	0.086	0.096	0.252	0.282	0.088	0.074
	500	0.098	0.078	0.080	0.074	0.476	0.484	0.120	0.078
AR(2)-	100	0.076	0.064	0.064	0.044	0.096	0.104	0.072	0.050
ARCH(2)	300	0.100	0.082	0.092	0.076	0.226	0.236	0.108	0.074
	500	0.082	0.068	0.076	0.056	0.392	0.420	0.094	0.068

and 1989. It was considered by Hušková and Antoch (2003). We set X_t as the annual rainfall and Y_t as the annual flow volume. The asymptotic test clearly rejects H_0 with a p-value of 0.0006. The possible change point is estimated by $\hat{s}_n$ from section 6 and suggests a change in 1979. Note that this is consistent with the literature. As was pointed out by Hušková and Antoch (2003) deforestation had started around that time, which is a possible explanation. Figure 4 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, $\sup_{z\in\mathbb{R}} |\hat{T}_n(\cdot, z)|$, as well as the critical value (red horizontal line) and the estimated change (green vertical line).

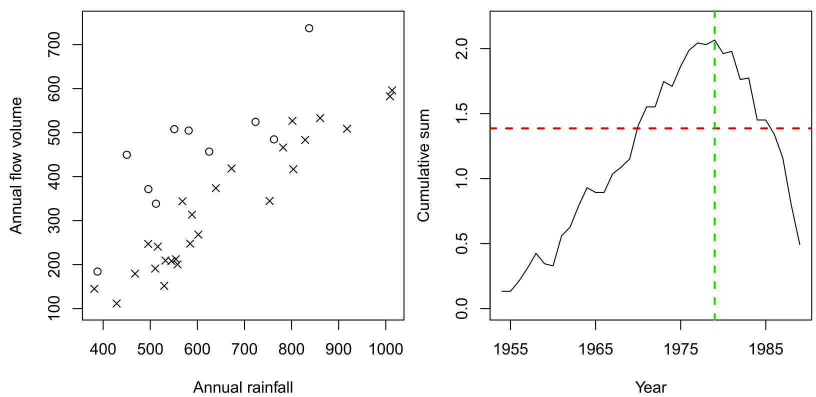

Fig 4. *Ráztoka data: scatterplot (left) and CUSUM (right)*

6. Concluding remarks

We suggested a new test for structural breaks in the regression function in nonparametric time series (auto-)regression. Our approach combines CUSUM statistics with the marked empirical process approach from goodness-of-fit testing. The considered model is rather general. It requires strict stationarity under the null, but no independence of the innovations, nor homoscedasticity. We show favorable asymptotic properties and demonstrate that the new testing procedures are consistent against fixed alternatives, while the traditional CUSUM tests are not. An estimator for the change point is given by $\hat{s}_n := \arg\max_{s\in[0,1]} \sup_{\boldsymbol{z}\in\mathbb{R}^d} |\hat{T}_n(s, \boldsymbol{z})|$. Asymptotic properties of this estimator will be considered in future research.

Moreover we have suggested a bootstrap version that can also be applied to detect changes in the regression function in the presence of changing variance functions. In a forthcoming paper we will consider testing for changes in the variance function.

Appendix A: Proofs and derivations

In subsection A.1 we give some auxiliary results for the proof of Theorem 3.1. The proof of the first part of the theorem was given in the main text, while the proof of the second part can be found in subsection A.2. Lemmata are proved in subsection A.3, while more detailed proofs can also be found in Mohr (2018).

A.1. Auxiliary results

The following assumptions are formulated for the first lemma that gives uniform rates of convergence for the regression estimator $\hat{m}_n$ from (2.2) and its derivatives. They hold under the assumptions of Theorem 3.1.

- **(P)** Let $(Y_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ be a strictly stationary and strongly mixing process with mixing coefficient $\alpha(\cdot)$. For some $b > 2$ let $\alpha(t) = O(t^{-\beta})$ for $t \to \infty$ with some $\beta > (1 + (b-1)(1+d))/(b-2)$.
- **(B3)** With b and β from assumption **(P)** let $(\log n)/(n^\theta h_n^d) = o(1)$ hold for $\theta = (\beta - 1 - d - (1+\beta)/(b-1))/(\beta + 3 - d - (1+\beta)/(b-1))$.

Lemma A.1. *Under the assumptions **(P)**, **(M)**, **(J)**, **(F1)**, **(K)**, **(B1)** and **(B3)** the following rates of convergence can be obtained for the Nadaraya-Watson estimator $\hat{m}_n$,*

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

(b) $\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right)\right| = O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^{d+2|\boldsymbol{k}|}}} + h_n^r p_n\right) p_n^{|\boldsymbol{k}|} q_n \delta_n^{|\boldsymbol{k}|+1}\right)$ *for all* $\boldsymbol{k} \in \mathbb{N}_0^d$ *with* $1 \le |\boldsymbol{k}| \le l+1$,

(c) $\sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{J}_n \\ \boldsymbol{x}\neq\boldsymbol{y}}} \frac{\left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right) - D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{y}) - m(\boldsymbol{y})\right)\right|}{\|\boldsymbol{x}-\boldsymbol{y}\|^\eta} = o_P(1)$ *for all* $\boldsymbol{k} \in \mathbb{N}_0^d$ *with* $|\boldsymbol{k}| = l$.

The proof of Lemma A.1 is analogous to the proof of Theorem 8 of Hansen (2008) and omitted for the sake of brevity. The proofs of the following lemmata are given in subsection A.3.

Lemma A.2. *Under the assumptions of Theorem 3.1 (i) and under H_0 we have for A_{n1} from (3.1)*

$$A_{n1}(s,\boldsymbol{z}) = s\sqrt{n}\int_{\mathbb{R}^d} (m(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{x}))\omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z}\} f(\boldsymbol{x}) d\boldsymbol{x} + o_P(1)$$

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

Lemma A.3. *Under the assumptions of Theorem 3.1 (i) and under H_0 we have for A_{n2} from (3.2) and T_n from (3.3)*

$$A_{n2}(s,\boldsymbol{z}) = T_n(s,\boldsymbol{z}) + o_P(1),$$

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

Lemma A.4. *Under the assumptions of Theorem 3.1 (i) and under* H_0

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{(-\boldsymbol{\infty},\boldsymbol{z}]}\left(m(\boldsymbol{y})-m(\boldsymbol{X}_i)\right)K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\omega_n(\boldsymbol{y})\frac{f(\boldsymbol{y})}{\hat{f}_n(\boldsymbol{y})}d\boldsymbol{y}=o_P(1)$$

holds uniformly in $\boldsymbol{z}\in\mathbb{R}^d$.

Lemma A.5. *Under the assumptions of Theorem 3.1 (i) and under* H_0

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i\bigg(\int_{(-\boldsymbol{\infty},\boldsymbol{z}]}K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\omega_n(\boldsymbol{y})\frac{f(\boldsymbol{y})}{\hat{f}_n(\boldsymbol{y})}d\boldsymbol{y}-\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq\boldsymbol{z}\}\bigg)=o_P(1)$$

holds uniformly in $\boldsymbol{z}\in\mathbb{R}^d$.

A.2. Proof of Theorem 3.1(ii)

For the proof of the second part of Theorem 3.1 we use a recent result on weak convergence of sequential empirical processes indexed in function classes that can be applied for strongly mixing sequences, see Mohr (2019). It is stated in Lemma A.7 and uses the following notion of bracketing number.

Definition A.6 (Bracketing number)**.** *Let* $\mathcal{X}$ *be a measure space,* $\mathcal{F}$ *some class of functions* $\mathcal{X}\to\mathbb{R}$ *and* ρ *some semi norm on* $\mathcal{F}$*. For all* $\varepsilon>0$*, let* $N=N(\varepsilon)$*, be the smallest integer, for which there exist a class of functions* $\mathcal{X}\to\mathbb{R}$*, denoted by* $\mathcal{B}$ *and called bounding class and a function class* $\mathcal{A}\subset\mathcal{F}$ *called approximating class such that* $|\mathcal{B}|=|\mathcal{A}|=N$*,* $\rho(b)<\varepsilon$*,* $\forall\ b\in\mathcal{B}$ *and for all* $\varphi\in\mathcal{F}$ *there exist* $a^*\in\mathcal{A}$ *and* $b^*\in\mathcal{B}$ *such that* $|\varphi-a^*|\leq b^*$*. Then* $N(\varepsilon)$ *is called the bracketing number and denoted by* $\tilde{N}_{[\,]}(\varepsilon,\mathcal{F},\rho)$*.*

Note that the usual notion for bracketing number (as in Definition 2.1.6 in van der Vaart and Wellner (1996)) will be referred to as $N_{[\,]}(\varepsilon,\mathcal{F},\rho)$.

Lemma A.7 (Corollary 2.7 in Mohr (2019))**.** *Let* $\{X_t:t\in\mathbb{Z}\}$ *be a strictly stationary sequence of random variables with values in some measure space* $\mathcal{X}$*. Let* $\mathcal{F}$ *be a class of measurable functions* $\mathcal{X}\to\mathbb{R}$*. Let furthermore the following assumptions hold.*

- ***(A1)*** *Let* $\{X_t:t\in\mathbb{Z}\}$ *be strongly mixing, such that* $\sum_{t=1}^{\infty}t^{Q-2}\alpha(t)^{\gamma/(2+\gamma)}<\infty$ *for some* $\gamma>0$ *and even* $Q>2$*.*
- ***(A2)*** *Let* $\int_0^1 x^{-\gamma/(2+\gamma)}(\tilde{N}_{[\,]}(x,\mathcal{F},\|\cdot\|_{L_2(P)}))^{1/Q}dx<\infty$ *for* Q *and* γ *from assumption **(A1)**, where* $X_1\sim P$*. Furthermore, assume that each* $\varepsilon>0$ *allows for a choice of bounding class* $\mathcal{B}$*, such that* $E\left[|b(X_1)|^{i(2+\gamma)/2}\right]^{1/2}\leq\varepsilon$ *for all* $b\in\mathcal{B}$ *and for all* $i=2,\ldots,Q$*.*
- ***(A3)*** *Let* $\mathcal{F}$ *possess an envelope function* F*, with* $E[|F(X_1)|^Q]<\infty$ *and let there exist a constant* $L<\infty$*, such that* $\sup_{\varphi\in\mathcal{F}}E\left[|\varphi(X_1)|^{Q(2+\gamma)/2}\right]\leq L$*.*

Furthermore, let for all $K\in\mathbb{N}$ *and all finite collections* $\varphi_k\in\mathcal{F}$*,* $s_k\in[0,1]$*,* $k=1,\ldots,K$*,* $(G_n(s_k,\varphi_k))_{k=1,\ldots,K}\overset{\mathcal{D}}{\to}(G(s_k,\varphi_k))_{k=1,\ldots,K}$*, where we denote*

$G_n(s,\varphi) = n^{-1/2}\sum_{i=1}^{\lfloor ns\rfloor}(\varphi(X_i) - E[\varphi(X_i)])$ *for* $s \in [0,1], \varphi \in \mathcal{F}$ *and where* $G = \{G(s,\varphi) : s \in [0,1], \varphi \in \mathcal{F}\}$ *is a centered Gaussian process.*

Then $\{G_n(s,\varphi) : s \in [0,1], \varphi \in \mathcal{F}\}$ *converges weakly to* G *in* $\ell^\infty([0,1]\times\mathcal{F})$.

Proof of Theorem 3.1(ii). First notice that due to assumption **(G)** and under the null hypothesis $(U_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a strictly stationary sequence of random variables with values in $\mathbb{R}\times\mathbb{R}^d$. Denote by P the common marginal distribution of $(U_1, \boldsymbol{X}_1)$ and define $\mathcal{F} := \{(u,\boldsymbol{x}) \mapsto uI\{\boldsymbol{x} \le \boldsymbol{z}\} : \boldsymbol{z} \in \mathbb{R}^d\}$. The convergence of T_n is then implied by the weak convergence of

$$G_n := \left\{G_n(s,\varphi) := \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\left(\varphi(U_i,\boldsymbol{X}_i) - \int\varphi dP\right) : s \in [0,1], \varphi \in \mathcal{F}\right\}$$

in $\ell^\infty([0,1]\times\mathcal{F})$. We apply Lemma A.7. Condition **(A1)** on the mixing coefficient of $(U_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is implied by assumption **(G)** on the mixing coefficient of $(Y_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ and the null hypothesis as measurable functions maintain mixing properties. To show condition **(A2)** on the function class $\mathcal{F}$, the choice of approximating functions and bounding functions, as in Definition A.6, will be discussed in more detail. Denote with $\bar{c}$ from assumption **(U)**, $h(\boldsymbol{x}) = \bar{c}(\boldsymbol{x})f(\boldsymbol{x})$ and $H(\boldsymbol{x}) = \int_{(-\boldsymbol{\infty},\boldsymbol{x}]} h(\boldsymbol{t})d\boldsymbol{t}$ for $\boldsymbol{x} \in \mathbb{R}^d$ and for all $i = 1,\ldots,d$ and $x \in \mathbb{R}$,

$$h_i(x) = \int\ldots\int h(x_1,\ldots,x_{i-1},x,x_{i+1},\ldots,x_d)dx_1\ldots dx_{i-1}dx_{i+1}\ldots dx_d$$

and $H_i(x) = \int_{-\infty}^{x} h_i(t)dt$. Let $\varepsilon > 0$ and choose for all $i = 1,\ldots,d$ some $N_i = N_i(\varepsilon) \in \mathbb{N}$ and $-\infty = z_{0,i} < \cdots < z_{N_i,i} = \infty$, such that

$$H_i\left(z_{j_i,i}\right) - H_i\left(z_{j_i-1,i}\right) \le \frac{\varepsilon^2}{d},\ \forall\ j_i = 1,\ldots,N_i,\ i = 1,\ldots,d. \tag{A.1}$$

Since H_i is continuous and $H_i(-\infty) = H(-\boldsymbol{\infty}) = 0$ and $H_i(\infty) = H(\boldsymbol{\infty}) \le M$ for $M < \infty$ from assumption **(U)**, N_i can be chosen to be smaller than $2dM\varepsilon^{-2}$ for all $i = 1,\ldots,d$. By using cartesian products, a partition of $\mathbb{R}^d$ is obtained. For simplicity reasons the following notation will be used. For $\boldsymbol{j} = (j_1,\ldots,j_d) \in \mathbb{N}^d$ let $\boldsymbol{z_j} := (z_{j_1,1},\ldots,z_{j_d,d})$, and $\boldsymbol{j-1} := (j_1-1,\ldots,j_d-1) \in \mathbb{N}^d$. For all $\boldsymbol{j} \in \times_{i=1}^d\{1,\ldots,N_i\}$ define approximating functions

$$a_{\boldsymbol{j}}(u,\boldsymbol{x}) := uI\left\{\boldsymbol{x} \le \boldsymbol{z_j}\right\}$$

and bounding functions

$$b_{\boldsymbol{j}}(u,\boldsymbol{x}) := |u|\left(I\left\{\boldsymbol{x} \le \boldsymbol{z_j}\right\} - I\left\{\boldsymbol{x} \le \boldsymbol{z_{j-1}}\right\}\right).$$

Notice that $a_{\boldsymbol{j}} \in \mathcal{F}$ while $b_{\boldsymbol{j}} \notin \mathcal{F}$ for all $\boldsymbol{j} \in \times_{i=1}^d\{1,\ldots,N_i\}$. For each $\boldsymbol{z} \in \mathbb{R}^d$ there exists a $\boldsymbol{j} \in \times_{i=1}^d\{1,\ldots,N_i\}$ such that $\boldsymbol{z} \in (\boldsymbol{z_{j-1}},\boldsymbol{z_j}]$. Therefore for each $\varphi \in \mathcal{F}$ there exists a $\boldsymbol{j} \in \times_{i=1}^d\{1,\ldots,N_i\}$ such that $|\varphi - a_{\boldsymbol{j}}| \le b_{\boldsymbol{j}}$. Further for $\boldsymbol{j} \in \times_{i=1}^d\{1,\ldots,N_i\}$ it holds that

$$\|b_{\boldsymbol{j}}\|_{L_2(P)}^2 = E\left[|U_t|^2\left(I\left\{\boldsymbol{X}_t \le \boldsymbol{z_j}\right\} - I\left\{\boldsymbol{X}_t \le \boldsymbol{z_{j-1}}\right\}\right)\right]$$

$$\leq H(\boldsymbol{z_j}) - H(\boldsymbol{z_{j-1}}) \leq \sum_{i=1}^{d} \left(H_i\left(z_{j_i,i}\right) - H_i\left(z_{j_i-1,i}\right)\right) \leq \varepsilon^2$$

due to (A.1). Furthermore for all $i = 2, \dots, Q$ by Jensen's inequality and **(U)**, it holds that

$$E\left[|U_t|^{i\frac{2+\gamma}{2}}|\boldsymbol{X}_t\right] \leq E\left[|U_t|^{i\frac{2+\gamma}{2}}|\boldsymbol{X}_t\right]^{\frac{i}{Q}} \leq (c(\boldsymbol{X}_t)^Q)^{\frac{i}{Q}} = c(\boldsymbol{X}_t)^i \text{ a.s.},$$

and thus

$$\begin{aligned}\int |b_{\boldsymbol{j}}|^{i\frac{2+\gamma}{2}}dP &= E\left[|U_t|^{i\frac{2+\gamma}{2}}\left(I\left\{\boldsymbol{X}_t \leq \boldsymbol{z_j}\right\} - I\left\{\boldsymbol{X}_t \leq \boldsymbol{z_{j-1}}\right\}\right)\right] \\ &\leq H(\boldsymbol{z_j}) - H(\boldsymbol{z_{j-1}}) \leq \varepsilon^2.\end{aligned}$$

Since $N_i = O(\varepsilon^{-2})$ for all $i = 1, \dots, d$, we have $\tilde{N}_{[\,]}(\varepsilon, \mathcal{F}, \|\cdot\|_{L_2(P)}) = O\left(\varepsilon^{-2d}\right)$. As $Q > d(2+\gamma)$ holds, assumption **(A2)** from Lemma A.7 is therefore satisfied. Assumption **(A3)** is also satisfied as $\bar{F} : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R},\ (u, \boldsymbol{x}) \mapsto u$ is an envelope function of $\mathcal{F}$ such that

$$\int |\bar{F}|^Q dP = E\left[|U_t|^Q\right] \leq E\left[|U_t|^{Q\frac{2+\gamma}{2}}\right]^{\frac{2}{2+\gamma}} \leq \left(\int c(\boldsymbol{u})^Q f(\boldsymbol{u}) d\boldsymbol{u}\right)^{\frac{2}{2+\gamma}} < \infty,$$

and additionally, it holds that

$$\sup_{\varphi \in \mathcal{F}} \int |\varphi|^{Q\frac{2+\gamma}{2}} dP = \sup_{\boldsymbol{z} \in \mathbb{R}^d} E\left[|U_t|^{Q\frac{2+\gamma}{2}} I\{\boldsymbol{X}_t \leq \boldsymbol{z}\}\right] \leq \int c(\boldsymbol{u})^Q f(\boldsymbol{u}) d\boldsymbol{u} < \infty.$$

What is left to show, is the convergence of all finite dimensional distributions of T_n. To this end we will apply Cramér-Wold's device. Let $\lambda_1, \dots, \lambda_K \in \mathbb{R} \setminus \{0\}$ and consider

$$\sum_{j=1}^{K} \lambda_j T_n(s_j, \boldsymbol{z}_j) = \sum_{i=1}^{n} \xi_{n,i},$$

where $\xi_{n,i} := \frac{1}{\sqrt{n}} U_i \sum_{j=1}^{K} \lambda_j I\{\boldsymbol{X}_i \leq \boldsymbol{z}_j\} I\left\{\frac{i}{n} \leq s_j\right\}$. Now Corollary 1 in Rio (1995) can be applied, which is a central limit theorem for strongly mixing triangular arrays. Following the notations in Rio (1995) define $V_{n,l} := Var(\sum_{i=1}^{l} \xi_{n,i})$ for all $l = 1, \dots, n$, and $n \in \mathbb{N}$. Let furthermore $Q_{n,i}$ be the càdlàg inverse function of $t \mapsto P(|\xi_{n,i}| > t)$, i.e.

$$Q_{n,i}(u) := \sup\{t > 0 : P(|\xi_{n,i}| > t) > u\},\ \forall\ u > 0,$$

with the convention that $\sup \emptyset := 0$. Let $\{\tilde{\alpha}_n(t) : t \in \mathbb{N}\}$ be the sequence of mixing coefficients of $\{\xi_{n,i} : 1 \leq i \leq n, n \in \mathbb{N}\}$. For $t \in (0, \infty)$ define $\tilde{\alpha}_n(t) := \tilde{\alpha}_n(\lfloor t \rfloor)$. Let its càdlàg inverse function be defined by

$$\tilde{\alpha}_n^{-1}(u) := \sup\{t > 0 : \tilde{\alpha}_n(t) > u\},\ \forall\ u > 0.$$

Condition (a) in Corollary 1 in Rio (1995) is easy to verify. Concerning condition (b) in aforementioned corollary note that by Markov's inequality, it holds that for all $t > 0$ and with $q := Q\frac{2+\gamma}{2}$

$$P(|\xi_{n,i}| > t) \le t^{-q} n^{-\frac{q}{2}} \tilde{M},$$

where $\tilde{M} := (\sum_{j=1}^{K} |\lambda_j|)^q M$ for $M < \infty$ from assumption **(U)**. Hence, for all $u > 0$ we have $Q_{n,i}(u) \le u^{-\frac{1}{q}} n^{-\frac{1}{2}} \tilde{M}^{\frac{1}{q}}$. By similar arguments, we obtain $\tilde{\alpha}_n^{-1}(u) \le \tilde{A} - \log_a(u)$ for all $u > 0$, where $\tilde{A} := \log_a(A)$. Furthermore, $V_{n,n}$ converges to $\sum_{j_1=1}^{K} \sum_{j_2=1}^{K} \lambda_{j_1} \lambda_{j_2} (s_{j_1} \wedge s_{j_2}) \Sigma(\boldsymbol{z}_{j_1} \wedge \boldsymbol{z}_{j_2}) > 0$. Putting the results together, it can be obtained that

$$V_{n,n}^{-\frac{3}{2}} \sum_{i=1}^{n} \int_0^1 \tilde{\alpha}_n^{-1}\left(\tfrac{x}{2}\right) Q_{i,n}^2(x) \inf\left\{\tilde{\alpha}_n^{-1}\left(\tfrac{x}{2}\right) Q_{i,n}(x), \sqrt{V_{n,n}}\right\} dx$$

$$\le \frac{1}{\sqrt{n}} \tilde{M}^{\frac{2}{q}} V_{n,n}^{-\frac{3}{2}} \int_0^1 \left(\tilde{A} - \log_a\left(\tfrac{x}{2}\right)\right) x^{-\frac{2}{q}} \inf\left\{\left(\tilde{A} - \log_a\left(\tfrac{x}{2}\right)\right) x^{-\frac{1}{q}} \tilde{M}^{\frac{1}{q}}, \sqrt{n}\sqrt{V_{n,n}}\right\} dx$$

converges to zero, and thus condition (b) is satisfied.

Applying Corollary 1 in Rio (1995), it holds that $\sum_{i=1}^{n} \xi_{n,i}/(V_{n,n})^{1/2}$ converges to the standard normal distribution and thus the assertion follows. □

A.3. Proofs of lemmata

Note that within the proofs we will make use of the following notations

$$(g(\cdot))^+ := g(\cdot) I\{g(\cdot) \ge 0\}, \ (g(\cdot))^- := g(\cdot) I\{g(\cdot) < 0\}.$$

Proof of Lemma A.2. For some l-times differentiable function $h : \boldsymbol{J}_n \to \mathbb{R}$ define the norm

$$\|h\|_{l+\eta} := \max_{\substack{\boldsymbol{k} \in \mathbb{N}_0^d \\ 1 \le |\boldsymbol{k}| \le l}} \sup_{\boldsymbol{x} \in \boldsymbol{J}_n} \left|D^{\boldsymbol{k}} h(\boldsymbol{x})\right| + \max_{\substack{\boldsymbol{k} \in \mathbb{N}_0^d \\ |\boldsymbol{k}| = l}} \sup_{\substack{\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{J}_n \\ \boldsymbol{x} \neq \boldsymbol{y}}} \frac{\left|D^{\boldsymbol{k}} h(\boldsymbol{x}) - D^{\boldsymbol{k}} h(\boldsymbol{y})\right|}{\|\boldsymbol{x} - \boldsymbol{y}\|^\eta}$$

and $\mathcal{H} := \mathcal{C}_{1,n}^{l+\eta}(\boldsymbol{J}_n) := \{h : \boldsymbol{J}_n \to \mathbb{R} : \|h\|_{l+\eta} \le 1, \sup_{\boldsymbol{x} \in \boldsymbol{J}_n} |h(\boldsymbol{x})| \le z_n \sqrt{\log n}\}$ with $z_n := q_n \delta_n ((\log n)/(n h_n^d))^{1/2}$. The third bandwidth condition in **(B2)** implies

$$\left(\sqrt{\frac{\log n}{n h_n^d}} + h_n^r p_n\right) q_n \delta_n = O\left(\sqrt{\frac{\log n}{n h_n^d}} q_n \delta_n\right)$$

and thus Lemma A.1 implies that $P(\hat{h}_n \in \mathcal{C}_{1,n}^{l+\eta}(\boldsymbol{J}_n)) \to 1$ as $n \to \infty$ holds for $\hat{h}_n(\boldsymbol{x}) = (m(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{x})) \omega_n(\boldsymbol{x})$. Let furthermore $\mathcal{F} := \{\boldsymbol{x} \mapsto I\{\boldsymbol{x} \le \boldsymbol{z}\} : \boldsymbol{z} \in \mathbb{R}^d\}$

and $\boldsymbol{X}_t \sim P$. Then the assertion of the lemma follows if we show

$$\sup_{s\in[0,1]}\sup_{\varphi\in\mathcal{F}}\sup_{h\in\mathcal{H}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\left(h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i)-\int h\varphi dP\right)\right| = o_P(1).$$

To this end let $\varepsilon_{n1}=\varepsilon_{n2}=n^{-1/2}$ and $\varepsilon_{n3}=n^{-1/2}/(\log n)$ and let further $0=s_1<\cdots<s_{K_n}=1$ partition $[0,1]$ in intervals of length $2\varepsilon_{n1}$ such that $K_n=O(\varepsilon_{n1}^{-1})$. Furthermore, let $J_n := N_{[\,]}\left(\varepsilon_{n2},\mathcal{F},\|\cdot\|_{L_2(P)}\right)$ and $M_n := N_{[\,]}\left(\varepsilon_{n3},\mathcal{H},\|\cdot\|_\infty\right)$, where $\|\cdot\|_\infty$ is the supremum norm on $\boldsymbol{J}_n$. Let $[\varphi_1^l,\varphi_1^u],\dots,[\varphi_{J_n}^l,\varphi_{J_n}^u]$ denote the brackets needed to cover $\mathcal{F}$. Note that they can be chosen to be indicator functions and therefore non negative. Let furthermore $[h_1^l,h_1^u],\dots,[h_{M_n}^l,h_{M_n}^u]$ define the brackets needed to cover $\mathcal{H}$. It can be shown that $J_n=O\left(\varepsilon_{n2}^{-2d}\right)$ and $M_n=O(\exp(c_n^d\varepsilon_{n3}^{-d/(l+\eta)}))$ and further

$$\begin{aligned}
&\sup_{s\in[0,1]}\sup_{\varphi\in\mathcal{F}}\sup_{h\in\mathcal{F}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\left(h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i)-\int h\varphi dP\right)\right|\\
&\le \max_{\substack{1\le k\le K_n\\1\le j\le J_n\\1\le m\le M_n}}\sup_{\varphi\in[\varphi_j^l,\varphi_j^u]}\sup_{h\in[h_m^l,h_m^u]}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_k\rfloor}\left(h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i)-\int h\varphi dP\right)\right|\\
&+\max_{\substack{1\le k\le K_n\\1\le j\le J_n\\1\le m\le M_n}}\sup_{\substack{s\in[0,1]\\|s-s_k|\le\varepsilon_{n1}}}\sup_{\varphi\in[\varphi_j^l,\varphi_j^u]}\sup_{h\in[h_m^l,h_m^u]}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i)-\int h\varphi dP\right)\right.\\
&\qquad\left.\cdot\left(I\left\{\tfrac{i}{n}\le s\right\}-I\left\{\tfrac{i}{n}\le s_k\right\}\right)\right|\\
&\le \max_{\substack{1\le k\le K_n\\1\le j\le J_n\\1\le m\le M_n}}\left\{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_k\rfloor}\left((h_m^u(\boldsymbol{X}_i))^+\varphi_j^u(\boldsymbol{X}_i)-\int(h_m^u)^+\varphi_j^u dP\right)\right|\right.\\
&\qquad+\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_k\rfloor}\left((h_m^u(\boldsymbol{X}_i))^-\varphi_j^l(\boldsymbol{X}_i)-\int(h_m^u)^-\varphi_j^l dP\right)\right|\\
&\qquad+\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_k\rfloor}\left((h_m^l(\boldsymbol{X}_i))^+\varphi_j^l(\boldsymbol{X}_i)-\int(h_m^l)^+\varphi_j^l dP\right)\right|\\
&\qquad\left.+\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_k\rfloor}\left((h_m^l(\boldsymbol{X}_i))^-\varphi_j^u(\boldsymbol{X}_i)-\int(h_m^l)^-\varphi_j^u dP\right)\right|\right\}\\
&+o(1).
\end{aligned}$$

In what follows we only consider the first line on the right hand side, while the other ones can be treated similarly. We apply Theorem 2.1 of Liebscher (1996) to the random variable (for m, j, k fixed)

$$Z_i := \left((h_m^u(\boldsymbol{X}_i))^+ \varphi_j^u(\boldsymbol{X}_i) - \int (h_m^u)^+ \varphi_j^u dP \right) I\left\{ \frac{i}{n} \leq s_k \right\}.$$

The mixing coefficient of $\{Z_t : 1 \leq t \leq n\}$ can be bounded by the mixing coefficient of $\{\boldsymbol{X}_t : t \in \mathbb{Z}\}$ due to Bradley (1985), Section 2, remark (iv). Further, the variables are centered and have a bound of order $O(z_n \log n)$. Applying Theorem 2.1 to $\sum_{i=1}^n Z_i$ yields for all $\epsilon > 0$ and $n \in \mathbb{N}$ large enough

$$\begin{aligned}
&P\left(\max_{\substack{1\leq k\leq K_n \\ 1\leq j\leq J_n \\ 1\leq m\leq M_n}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k \rfloor} \left((h_m^u(\boldsymbol{X}_i))^+ \varphi_j^u(\boldsymbol{X}_i) - \int (h_m^u)^+ \varphi_j^u dP \right) \right| > \epsilon \right) \\
&\leq \sum_{\substack{1\leq k\leq K_n \\ 1\leq j\leq J_n \\ 1\leq m\leq M_n}} P\left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k \rfloor} \left((h_m^u(\boldsymbol{X}_i))^+ \varphi_j^u(\boldsymbol{X}_i) - \int (h_m^u)^+ \varphi_j^u dP \right) \right| > \epsilon \right) \\
&\leq K_n J_n M_n 4 \exp\left(- \frac{n\epsilon^2}{64n \left\lfloor (nh_n^d)^{1/2} \right\rfloor z_n^2 \log(n) + \frac{8}{3}\sqrt{n}\epsilon \left\lfloor (nh_n^d)^{1/2} \right\rfloor z_n \log(n)^{1/2}} \right) \\
&+ K_n J_n M_n 4 \frac{n}{\left\lfloor (nh_n^d)^{1/2} \right\rfloor} \alpha\left(\left\lfloor (nh_n^d)^{1/2} \right\rfloor \right) \\
&= o(1),
\end{aligned}$$

where the first and second bandwidth constraints in **(B2)** were used. Details are omitted for the sake of brevity. □

Proof of Lemma A.4. First, using the uniform rates of convergence results in Lemma A.1 applied to $(m(\boldsymbol{X}_t), \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ together with the first and the last bandwidth condition in assumption **(B2)** as well as the second condition in assumption **(B1)**, it can be shown that

$$\sup_{\boldsymbol{z}\in\mathbb{R}^d} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{(-\boldsymbol{\infty},\boldsymbol{z}]} (m(\boldsymbol{y}) - m(\boldsymbol{X}_i))\, K_{h_n}(\boldsymbol{y} - \boldsymbol{X}_i)\omega_n(\boldsymbol{y}) \left(\frac{f(\boldsymbol{y})}{\hat{f}_n(\boldsymbol{y})} - 1 \right) d\boldsymbol{y} \right| = o_P(1).$$

Defining the function class

$$\mathcal{F}_{n,1} := \left\{ \boldsymbol{x} \mapsto \int_{(-\boldsymbol{\infty},\boldsymbol{z}]} (m(\boldsymbol{y}) - m(\boldsymbol{x})) K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) \omega_n(\boldsymbol{y}) d\boldsymbol{y} : \boldsymbol{z} \in \mathbb{R}^d \right\},$$

and imposing $\boldsymbol{X}_t \sim P$, the assertion of the lemma follows if we show

$$\sup_{\varphi\in\mathcal{F}_{n,1}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\varphi(\boldsymbol{X}_i) - \int \varphi dP \right) \right| = o_P(1), \tag{A.2}$$

$$\sup_{\varphi\in\mathcal{F}_{n,1}}\left|\int \varphi dP\right| = o\left(n^{-1/2}\right). \tag{A.3}$$

For the proof of (A.2) let $\varepsilon_n := n^{-1/2}/(\log n)$ and $J_n := N_{[\,]}\left(\varepsilon_n, \mathcal{F}_{n,1}, \|\cdot\|_{L_1(P)}\right)$. Then there exists a partition $\boldsymbol{z}_1,\dots,\boldsymbol{z}_{J_n}$ of $\mathbb{R}^d$ such that $\|\varphi_j^u - \varphi_j^l\|_{L_1(P)} \le \varepsilon_n$ for all $j\in\{1,\dots,J_n\}$, where

$$\varphi_j^u(\boldsymbol{x}) := \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}_j]} \big((m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\big)^+\omega_n(\boldsymbol{y})d\boldsymbol{y} + \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}_{j-1}]} \big((m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\big)^-\omega_n(\boldsymbol{y})d\boldsymbol{y}$$

and

$$\varphi_j^l(\boldsymbol{x}) := \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}_{j-1}]} \big((m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\big)^+\omega_n(\boldsymbol{y})d\boldsymbol{y} + \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}_j]} \big((m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\big)^-\omega_n(\boldsymbol{y})d\boldsymbol{y}.$$

It then holds that $J_n = O(\varepsilon_n^{-d})$. Using these brackets of $\mathcal{F}_{n,1}$, it can be shown that

$$\begin{aligned}
&\sup_{\varphi\in\mathcal{F}_{n,1}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\varphi(\boldsymbol{X}_i)-\int\varphi dP\right)\right| \\
&= \max_{1\le j\le J_n}\sup_{\varphi\in[\varphi_j^l,\varphi_j^u]}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\varphi(\boldsymbol{X}_i)-\int\varphi dP\right)\right| \\
&\le \max_{1\le j\le J_n}\max\left\{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\varphi_j^u(\boldsymbol{X}_i)-\int\varphi_j^u dP\right)\right|,\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\varphi_j^l(\boldsymbol{X}_i)-\int\varphi_j^l dP\right)\right|\right\} \\
&+ o(1),
\end{aligned}$$

where it is sufficient to discuss the proof of

$$\max_{1\le j\le J_n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\varphi_j^u(\boldsymbol{X}_i)-\int\varphi_j^u dP\right)\right| = o_P(1)$$

as the other assertion works analogously. By defining

$$\varphi_{j,1}^u(\boldsymbol{x}) := \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}_j]} \big((m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\big)^+\omega_n(\boldsymbol{y})d\boldsymbol{y}$$

and

$$\varphi_{j,2}^u(\boldsymbol{x}) := \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}_{j-1}]} \big((m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\big)^-\omega_n(\boldsymbol{y})d\boldsymbol{y}$$

it holds that $\varphi_j^u(\boldsymbol{x}) = \varphi_{j,1}^u(\boldsymbol{x}) + \varphi_{j,2}^u(\boldsymbol{x})$. Thus, again the problem is reduced to showing

$$\max_{1\le j\le J_n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\varphi_{j,1}^u(\boldsymbol{X}_i)-\int\varphi_{j,1}^u dP\right)\right| = o_P(1),$$

the other assertion works analogously. Similar to before we apply Theorem 2.1 of Liebscher (1996) to the random variable $Z_i := \varphi_{j,1}^u(\boldsymbol{X}_i) - \int \varphi_{j,1}^u dP$ for fixed j. Note that the mixing properties are the same as the ones of the original process. Further the variables are centered and posses a bound of order $O(h_n q_n)$. For all $\epsilon > 0$ and $n \in \mathbb{N}$ large enough it can then be shown that

$$\begin{aligned}
&P\left(\max_{1\le j\le J_n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\varphi_{j,1}^u(\boldsymbol{X}_i)-\int\varphi_{j,1}^u dP\right)\right| > \epsilon\right)\\
&\le \sum_{j=1}^{J_n} P\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\varphi_{j,1}^u(\boldsymbol{X}_i)-\int\varphi_{j,1}^u dP\right)\right| > \epsilon\right)\\
&\le J_n 4\exp\left(-\frac{n\epsilon^2}{64n(\lfloor(\log n)^2\rfloor+1)h_n^2q_n^2+\frac{8}{3}\sqrt{n}\epsilon(\lfloor(\log n)^2\rfloor+1)h_nq_n}\right)\\
&+ J_n 4\frac{n}{\lfloor(\log n)^2\rfloor+1}\alpha\left(\lfloor(\log n)^2\rfloor+1\right)\\
&= o(1),
\end{aligned}$$

where eventually the last bandwidth condition in assumption **(B3)** was used. The assertion in (A.3) can be shown by using Taylor's expansion for both m and f up to order $r-1$ and the assumptions in **(F1)**. Thus

$$\begin{aligned}
&\sup_{\varphi\in\mathcal{F}_{n,1}}\left|\int\varphi dP\right|\\
&= \sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}(m(\boldsymbol{y})-m(\boldsymbol{x}))\frac{1}{h_n^d}K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\omega_n(\boldsymbol{y})I\{\boldsymbol{y}\le\boldsymbol{z}\}d\boldsymbol{y}f(\boldsymbol{x})d\boldsymbol{x}\right|\\
&\le \int_{\mathbb{R}^d}\left|\int_{\mathbb{R}^d}(m(\boldsymbol{y})-m(\boldsymbol{y}-\boldsymbol{t}h_n))K(\boldsymbol{t})\omega_n(\boldsymbol{y})f(\boldsymbol{y}-\boldsymbol{t}h_n)d\boldsymbol{t}\right|d\boldsymbol{y}\\
&= O(h_n^r p_n q_n) = o\left(n^{-1/2}\right),
\end{aligned}$$

where the last equality holds by the third condition in **(B3)**. □

Proof of Lemma A.5. First, using the uniform rates of convergence results in Lemma A.1 applied to $(U_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ together with assumptions on the bandwidth, it can be shown that

$$\sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n U_i\int_{\mathbb{R}^d}K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\omega_n(\boldsymbol{y})I\{\boldsymbol{y}\le\boldsymbol{z}\}\left(\frac{f(\boldsymbol{y})}{\hat{f}_n(\boldsymbol{y})}-1\right)d\boldsymbol{y}\right| = o_P(1).$$

Furthermore, it can be shown that uniformly in $\boldsymbol{z} \in \mathbb{R}^d$ and for $q := Q\frac{2+\gamma}{2} > 2$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \bigg(\int_{(-\boldsymbol{\infty}, \boldsymbol{z}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{X}_i) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \le \boldsymbol{z}\} \bigg)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Bigg(U_i I\{|U_i| \le n^{1/q}\} \Big(\int\limits_{(-\boldsymbol{\infty}, \boldsymbol{z}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{X}_i) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \le \boldsymbol{z}\} \Big)$$
$$- E\Bigg[U_i I\{|U_i| \le n^{1/q}\} \Big(\int\limits_{(-\boldsymbol{\infty}, \boldsymbol{z}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{X}_i) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \le \boldsymbol{z}\} \Big) \Bigg] \Bigg)$$
$$+ o_P(1).$$

Defining the function class

$$\mathcal{F}_{n,2} := \Big\{ (u, \boldsymbol{x}) \mapsto u I\{|u| \le n^{1/q}\} \Big(\int_{(-\boldsymbol{\infty}, \boldsymbol{z}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z}\} \Big) : \boldsymbol{z} \in \mathbb{R}^d \Big\},$$

and imposing $(U_t, \boldsymbol{X}_t) \sim P$, the assertion of the lemma follows if we show

$$\sup_{\varphi \in \mathcal{F}_{n,2}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\varphi(U_i, \boldsymbol{X}_i) - \int \varphi dP \Big) \right| = o_P(1). \tag{A.4}$$

To this end let $\varepsilon_n := n^{-1/2}/(\log n)$ and $J_n := N_{[\,]}\big(\varepsilon_n, \mathcal{F}_{n,2}, \|\cdot\|_{L_1(P)}\big)$. Then there exists a partition $\boldsymbol{z}_1, \dots, \boldsymbol{z}_{J_n}$ of $\mathbb{R}^d$ such that $\|\varphi_j^u - \varphi_j^l\|_{L_1(P)} \le \varepsilon_n$ for all $j \in \{1, \dots, J_n\}$, where

$$\varphi_j^u(u, \boldsymbol{x}) := \varphi_{j,1}^u(u, \boldsymbol{x}) + \varphi_{j,2}^u(u, \boldsymbol{x}) + \varphi_{j,3}^u(u, \boldsymbol{x}),$$

where

$$\varphi_{j,1}^u(u, \boldsymbol{x}) := u I\{u < 0\} I\{|u| \le n^{1/q}\} \Bigg(\int_{(-\boldsymbol{\infty}, \boldsymbol{z}_j]} K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z}_j\} \Bigg),$$

$$\varphi_{j,2}^u(u, \boldsymbol{x}) := u I\{u \ge 0\} I\{|u| \le n^{1/q}\} \Bigg(\int_{(-\boldsymbol{\infty}, \boldsymbol{z}_{j-1}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z}_{j-1}\} \Bigg),$$

$$\varphi_{j,3}^u(u, \boldsymbol{x}) := |u| I\{|u| \le n^{1/q}\} \Bigg(\int_{(-\boldsymbol{\infty}, \boldsymbol{z}_j]} \big(K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) \big)^+ \omega_n(\boldsymbol{y}) d\boldsymbol{y}$$

$$- \int_{(-\boldsymbol{\infty},\boldsymbol{z}_{j-1}]} \big(K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\big)^+ \omega_n(\boldsymbol{y}) d\boldsymbol{y} \Bigg),$$

and similarly,

$$\varphi_j^l(u,\boldsymbol{x}) := \varphi_{j,1}^l(u,\boldsymbol{x}) + \varphi_{j,2}^l(u,\boldsymbol{x}) + \varphi_{j,3}^l(u,\boldsymbol{x}),$$

where

$$\varphi_{j,1}^l(u,\boldsymbol{x}) := uI\{u \geq 0\} I\{|u| \leq n^{1/q}\} \Bigg(\int_{(-\boldsymbol{\infty},\boldsymbol{z}_j]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}_j\} \Bigg),$$

$$\varphi_{j,2}^l(u,\boldsymbol{x}) := uI\{u < 0\} I\{|u| \leq n^{1/q}\} \Bigg(\int_{(-\boldsymbol{\infty},\boldsymbol{z}_{j-1}]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}_{j-1}\} \Bigg),$$

$$\varphi_{j,3}^l(u,\boldsymbol{x}) := -|u| I\{|u| \leq n^{1/q}\} \Bigg(\int_{(-\boldsymbol{\infty},\boldsymbol{z}_j]} \big(K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\big)^+ \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \int_{(-\boldsymbol{\infty},\boldsymbol{z}_{j-1}]} \big(K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\big)^+ \omega_n(\boldsymbol{y}) d\boldsymbol{y} \Bigg).$$

It then holds that $J_n = O(\varepsilon_n^- d)$. Using these brackets of $\mathcal{F}_{n,2}$, it can be shown that

$$\begin{aligned} &\sup_{\varphi \in \mathcal{F}_{n,2}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\varphi(U_i, \boldsymbol{X}_i) - \int \varphi dP \right) \right| \\ &\leq \max_{1 \leq j \leq J_n} \left\{ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\varphi_j^u(U_i, \boldsymbol{X}_i) - \int \varphi_j^u dP \right) \right| \right. \\ &\qquad \left. + \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\varphi_j^l(U_i, \boldsymbol{X}_i) - \int \varphi_j^l dP \right) \right| \right\} + o(1). \end{aligned}$$

We will only consider the first term, as the second one is treated analogously. Similar to before we apply Theorem 2.1 of Liebscher (1996) to the random variable $Z_i := \varphi_{j,1}^u(U_i, \boldsymbol{X}_i) - \int \varphi_{j,1}^u dP$ for fixed j. Note that the mixing properties are the same as the ones of the original process. Further, the variables are centered and posses a bound of order $O(n^{1/q})$. For all $\epsilon > 0$ and $n \in \mathbb{N}$ large enough it can then be shown that

$$P\left(\max_{1 \leq j \leq J_n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\varphi_{j,1}^u(U_i, \boldsymbol{X}_i) - \int \varphi_{j,1}^u dP \right) \right| > \epsilon \right)$$

$$\leq \sum_{j=1}^{J_n} P\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varphi_{j,1}^u(U_i,\boldsymbol{X}_i)-\int\varphi_{j,1}^u dP\right)\right|>\epsilon\right)$$
$$\leq J_n 4\exp\left(-\frac{n\epsilon^2}{64n(\lfloor\log(n)^2\rfloor+1)h_n+\frac{8}{3}\sqrt{n}\epsilon(\lfloor\log(n)^2\rfloor+1)n^{1/q}}\right)$$
$$+J_n 4\frac{n}{\lfloor\log(n)^2\rfloor+1}\alpha\left(\lfloor\log(n)^2\rfloor+1\right)$$
$$=o(1),$$

due to the third bandwidth condition in assumption **(B3)** and as $q > 2$. ☐

Proof of Lemma A.3. It will be shown that uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor} U_i I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} I\{\boldsymbol{X}_i \notin [-c_n,c_n]^d\} = o_P(1). \tag{A.5}$$

To this end define the function class

$$\mathcal{F} := \left\{(u,\boldsymbol{x}) \mapsto uI\{\boldsymbol{x}\leq\boldsymbol{z}\}I\{\boldsymbol{x}\notin[-a,a]^d\} : \boldsymbol{z}\in\mathbb{R}^d, a\in\mathbb{R}_+\right\}$$

and for $s \in [0,1]$ and $\varphi \in \mathcal{F}$

$$G_n(s,\varphi) := \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\left(\varphi(U_i,\boldsymbol{X}_i)-\int\varphi dP\right),$$

where $(U_t,\boldsymbol{X}_t) \sim P$ and $\int \varphi dP = 0$. Similarly to the proof of Theorem 3.1 (ii) an application of Theorem 2.5 in Mohr (2019) yields for all $\delta_n \searrow 0$ and with $d(\varphi,\psi) := \|\varphi-\psi\|_{L_{Q\frac{2+\gamma}{2}}(P)}$,

$$\sup_{\substack{\{s,t\in[0,1],\varphi,\psi\in\mathcal{F}:\\ |s-t|+d(\varphi,\psi)<\delta_n\}}} |G_n(s,\varphi)-G_n(t,\psi)| = o_P(1). \tag{A.6}$$

Note that $Q > (d+1)(2+\gamma)$ is needed here as $\tilde{N}_{[\,]}(\varepsilon,\mathcal{F},\|\cdot\|_{L_2(P)}) = O(\varepsilon^{-2(d+1)})$. Defining $\varphi_n(u,\boldsymbol{x}) := uI\{\boldsymbol{x}\leq\boldsymbol{z}\}I\{\boldsymbol{x}\notin[-c_n,c_n]^d\}$ for some fixed $\boldsymbol{z}\in\mathbb{R}^d$, it thens holds that $\varphi_n \in \mathcal{F}$ for all $n \in \mathbb{N}$ and

$$d(\varphi_n,0) = \|\varphi_n\|_{L_{Q\frac{2+\gamma}{2}}(P)} \leq \left(\int c(\boldsymbol{x})^Q I\{\boldsymbol{x}\notin[-c_n,c_n]^d\}f(\boldsymbol{x})d\boldsymbol{x}\right)^{\frac{1}{Q}\frac{2}{2+\gamma}} \overset{n\to\infty}{\longrightarrow} 0,$$

where the convergence holds by the dominated convergence theorem because $\int c(\boldsymbol{x})^Q f(\boldsymbol{x})d\boldsymbol{x} < \infty$. With

$$\left(\int c(\boldsymbol{x})^Q I\{\boldsymbol{x}\notin[-c_n,c_n]^d\}f(\boldsymbol{x})d\boldsymbol{x}\right)^{\frac{1}{Q}\frac{2}{2+\gamma}} =: \delta_n \searrow 0$$

it can therefore be concluded that

$$\sup_{s\in[0,1]} \sup_{\boldsymbol{z}\in\mathbb{R}^d} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} U_i I\{\boldsymbol{X}_i \le \boldsymbol{z}\} I\{\boldsymbol{X}_i \notin [-c_n, c_n]^d\} \right| \le \sup_{\substack{\{s,t\in[0,1],\varphi,\psi\in\mathcal{F}:\\ |s-t|+d(\varphi,\psi)<\delta_n\}}} |G_n(s,\varphi) - G_n(t,\psi)| .$$

With (A.6) the last term is $o_P(1)$ which proves the assertion in (A.5) and therefore the assertion of the lemma. □

Acknowledgements

The authors are grateful to two referees for their constructive suggestions on the first version of the manuscript.

References

Bradley, R. C. (1985). Basic properties of strong mixing conditions. In Eberlein, E. and Taqqu, M. S., editors, *Dependence in Probability and Statistics*, pages 165–192. Birkhäuser, Boston. MR0899990

Burke, M. D. and Bewa, G. (2013). Change-point detection for general nonparametric regression models. *Open Journal of Statistics*, 3:261–267.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *J. Econom.*, 136:163–188. MR2328589

Delgado, M. A. and Manteiga, W. G. (2001). Significance testing in nonparametric regression based on the bootstrap. *Ann. Statist.*, 29:1469–1507. MR1873339

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York. MR1964455

Hafner, C. M. and Herwartz, H. (2000). Testing for linear autoregressive dynamics under heteroskedasticity. *Econom. J.*, 3:177–197. MR1820413

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econom. Theory*, 24:726–748. MR2409261

Hidalgo, J. (1995). A nonparametric conditional moment test for structural stability. *Econom. Theory*, 1:671–698. MR1363629

Honda, T. (1997). The CUSUM tests with nonparametric regression residuals. *J. Japan Statist. Soc.*, 27:45–63. MR1463356

Hušková, M. and Antoch, J. (2003). Detection of structural changes in regression. *Tatra Mt. Math. Publ.*, 26:201–215. MR2055177

Koul, H. L. and Stute, W. (1999). Nonparametric model checks for time series. *Ann. Statist.*, 27:204–236. MR1701108

Kreiß, J.-P. (1997). Asymptotic properties of residual bootstrap for autoregressions. Technical report, Technical University of Braunschweig, Germany. `https://www.tu-braunschweig.de/Medien-DB/stochastik/kreiss-1997.pdf`.

Kreiß, J.-P. and Lahiri, S. N. (2012). Bootstrap methods for time series. In T. Subba Rao, S. S. R. and Rao, C., editors, *Handbook of Statistics*, pages 3–26. Elsevier, Amsterdam. MR3295420

Kreiß, J.-P. and Neuhaus, G. (2006). *Einführung in die Zeitreihenanalyse*. Springer, Berlin. MR0833744

Kristensen, D. (2009). Uniform convergence rates of kernel estimators with heterogeneous dependent data. *Econom. Theory*, 25:1433–1445. MR2540506

Kristensen, D. (2012). Non-parametric detection and estimation of structural change. *Econom. J.*, 15:420–461. MR3000054

Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Process. Appl.*, 65:69–80. MR1422880

Liu, R. Y. (1988). Bootstrap procedure under some non-i.i.d. models. *Ann. Statist.*, 16:1696–1708. MR0964947

Lu, Z. (1998). On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Stat. Sinica*, 8:1205–1217. MR1666249

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. (2019). A weak convergence result for sequential empirical processes under weak dependence. *Stochastics*, 92:140–164. MR4037089

Mohr, M. and Selk, L. (2020). Estimating change points in nonparametric time series regression models. *Stat. Papers*. https://doi.org/10.1007/s00362-020-01162-8.

Perron, P. and Zhou, J. (2008). Testing Jointly for Structural Changes in the Error Variance and Coefficients of a Linear Regression Model. *Working paper, Boston University*.

Pitarakis, J.-Y. (2004). Least squares estimation and tests of breaks in mean and variance under misspecification. *Econom. J.*, 7:32–54. MR2076626

Rio, E. (1995). About the Lindeberg method for strongly mixing sequences. *ESAIM Probab. Statist.*, 1:35–61. MR1382517

Shao, X. (2010). The dependent wild bootstrap. *J. Am. Stat. Assoc.*, 105:218–235. MR2656050

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.*, 25:613–641. MR1439316

Su, L. and White, H. (2010). Testing structural change in partially linear models. *Econom. Theory*, 26:1761–1806. MR2738016

Su, L. and Xiao, Z. (2008). Testing structural change in time-series nonparametric regression models. *Stat. Interface*, 1:347–366. MR2476751

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

Vogt, M. (2015). Testing for structural change in time-varying nonparametric regression models. *Econom. Theory*, 31:811–859. MR3377270

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.*, 14:1261–1295. MR0868303

Wu, J. (2016). Detecting structural changes under nonstationary volatility. *Econom. Lett.*, 146:151–154. MR3542608

Xu, K.-L. (2015). Testing for structural change under non-stationary variances. *Econom. J.*, 18:274–305. MR3370843